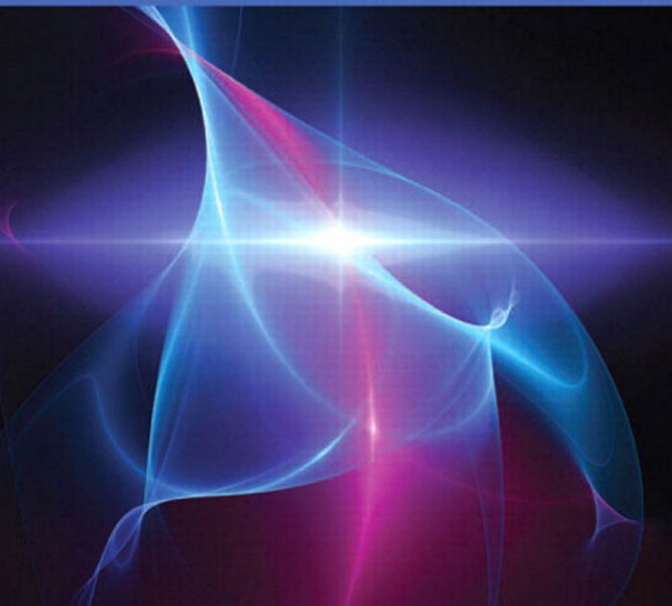


Seventh Edition

# The Economics of Health and Health Care

Sherman Folland Allen C. Goodman Miron Stano



*Seventh Edition*

# THE ECONOMICS OF HEALTH AND HEALTH CARE

**Sherman Folland**

*Professor of Economics, Oakland University*

**Allen C. Goodman**

*Professor of Economics, Wayne State University*

**Miron Stano**

*Professor of Economics and Management, Oakland University*

**PEARSON**

Boston Columbus Indianapolis New York San Francisco Upper Saddle River  
Amsterdam Cape Town Dubai London Madrid Milan Munich Paris Montréal Toronto  
Delhi Mexico City São Paulo Sydney Hong Kong Seoul Singapore Taipei Tokyo

**Executive Editor:** David Alexander  
**Editorial Director:** Sally Yagan  
**Editor in Chief:** Donna Battista  
**Assistant Editor:** Lindsey Sloan  
**Editorial Assistant:** Emily Brodeur  
**Director of Marketing:** Patrice Jones  
**Senior Managing Editor:** Nancy Fenton  
**Production Project Manager:** Nancy Freihofer  
**Permissions Specialist:** Estelle Simpson  
**Senior Manufacturing Buyer:** Carol Melville  
**Cover Photo:** Yang MingQi/Fotolia  
**Cover Design:** Jayne Conte  
**Full-Service Project Management and Composition:** Integra  
**Printer/Binder:** Edwards Brothers  
**Cover Printer:** Lehigh Phoenix  
**Text Font:** 10/12 Times

Credits and acknowledgments borrowed from other sources and reproduced, with permission, in this textbook appear on appropriate page within text.

---

**Copyright © 2013, 2010, 2007, 2004, 2001 Pearson Education, Inc., publishing as Prentice Hall, One Lake Street, Upper Saddle River, New Jersey 07458.** All rights reserved. Manufactured in the United States of America. This publication is protected by Copyright, and permission should be obtained from the publisher prior to any prohibited reproduction, storage in a retrieval system, or transmission in any form or by any means, electronic, mechanical, photocopying, recording, or likewise. To obtain permission(s) to use material from this work, please submit a written request to Pearson Education, Inc., Permissions Department, One Lake Street, Upper Saddle River, New Jersey 07458.

Many of the designations by manufacturers and sellers to distinguish their products are claimed as trademarks. Where those designations appear in this book, and the publisher was aware of a trademark claim, the designations have been printed in initial caps or all caps.

#### **Library of Congress Cataloging-in-Publication Data**

Folland, Sherman.

The economics of health and health care/Sherman Folland, Allen C. Goodman, Miron Stano.—7th ed.  
p. cm.

Includes bibliographical references and index.

ISBN-13: 978-0-13-277369-0

ISBN-10: 0-13-277369-4

I. Goodman, Allen C. II. Stano, Miron. III. Title.

[DNLN: 1. Economics, Medical. 2. Government Regulation. 3. Health Care Reform—economics.

4. Health Policy—economics. 5. Insurance, Health—economics. W 74.1]

LC Classification not assigned

338.4'73621—dc23

2011053121

10 9 8 7 6 5 4 3 2 1

**PEARSON**

[www.pearsonhighered.com](http://www.pearsonhighered.com)

ISBN 13: 978-0-13-277369-0

ISBN 10: 0-13-277369-4

# BRIEF CONTENTS

## **PART I Basic Economics Tools**

- Chapter 1 Introduction 1
- Chapter 2 Microeconomic Tools for Health Economics 20
- Chapter 3 Statistical Tools for Health Economics 48
- Chapter 4 Economic Efficiency and Cost-Benefit Analysis 63

## **PART II Supply and Demand**

- Chapter 5 Production of Health 85
- Chapter 6 The Production, Cost, and Technology of Health Care 105
- Chapter 7 Demand for Health Capital 129
- Chapter 8 Demand and Supply of Health Insurance 148
- Chapter 9 Consumer Choice and Demand 173

## **PART III Information and Insurance Markets**

- Chapter 10 Asymmetric Information and Agency 195
- Chapter 11 The Organization of Health Insurance Markets 210
- Chapter 12 Managed Care 236
- Chapter 13 Nonprofit Firms 265

## **PART IV Key Players in the Health Care Sector**

- Chapter 14 Hospitals and Long-Term Care 283
- Chapter 15 The Physician's Practice 302
- Chapter 16 Health Care Labor Markets and Professional Training 319
- Chapter 17 The Pharmaceutical Industry 344

## **PART V Social Insurance**

- Chapter 18 Equity, Efficiency, and Need 367
- Chapter 19 Government Intervention in Health Care Markets 390
- Chapter 20 Government Regulation: Principal Regulatory Mechanisms 407
- Chapter 21 Social Insurance 435
- Chapter 22 Comparative Health Care Systems 466
- Chapter 23 Health System Reform 492

## **PART VI Special Topics**

- Chapter 24 The Health Economics of Bads 513
- Chapter 25 Epidemiology and Economics: HIV/AIDS in Africa 531

# CONTENTS

Preface *xix*

## Part I Basic Economics Tools

### Chapter 1 Introduction 1

What Is Health Economics?	2
The Relevance of Health Economics	3
The Size and Scope of the Health Economy	3
Health Care's Share of GDP in the United States	3
Health Care Spending in Other Countries	4
Importance of the Health Economy in Personal Spending	5
Importance of Labor and Capital in the Health Economy	6
Time—The Ultimate Resource	7
The Importance Attached to Economic Problems of Health Care Delivery	8
Inflation	9
Access	9
Quality	9
The Economic Side to Other Health Issues	9
Economic Methods and Examples of Analysis	10
Features of Economic Analysis	10
Examples of Health Economics Analysis	11
Does Economics Apply to Health and Health Care?	12
An Example: Does Price Matter?	12
Is Health Care Different?	13
Presence and Extent of Uncertainty	13
Prominence of Insurance	13
Problems of Information	14
Large Role of Nonprofit Firms	15
Restrictions on Competition	15
Role of Equity and Need	16
Government Subsidies and Public Provision	16
Conclusions	16
<i>Summary</i>	17
• <i>Discussion Questions</i>	18
• <i>Exercises</i>	18

### Chapter 2 Microeconomic Tools for Health Economics 20

Scarcity and the Production Possibilities Frontier	21
Practice with Supply and Demand	23
The Demand Curve and Demand Shifters	23
The Supply Curve and Supply Shifters	24
Equilibrium	25
Comparative Statics	25

Functions and Curves	26
Linear Functions	27
Demand Functions	27
Derived Demand	28
Consumer Theory: Ideas Behind the Demand Curve	28
Utility	29
Indifference Curves	30
Budget Constraints	31
Consumer Equilibrium	32
Individual and Market Demands	33
Elasticities	33
Production and Market Supply	35
The Production Function	35
Production Functions	36
Isocost Curves	38
Cost Minimization or Output Maximization	39
Marginal and Average Cost Curves	40
The Firm Supply Curve Under Perfect Competition	41
Monopoly and Other Market Structures	43
Conclusions	45
<i>Summary</i>	45
• <i>Discussion Questions</i>	46
• <i>Exercises</i>	47
<b>Chapter 3 Statistical Tools for Health Economics</b>	<b>48</b>
Hypothesis Testing	49
Difference of Means	49
The Variance of a Distribution	50
Standard Error of the Mean	51
Hypotheses and Inferences	53
Regression Analysis	54
Ordinary Least Squares (OLS) Regressions	55
A Demand Regression	56
Estimating Elasticities	57
Multiple Regression Analysis	58
Interpreting Regression Coefficients	58
Dummy Variables	59
Statistical Inference in the Sciences and Social Sciences	60
Conclusions	61
<i>Summary</i>	61
• <i>Discussion Questions</i>	61
• <i>Exercises</i>	62
<b>Chapter 4 Economic Efficiency and Cost-Benefit Analysis</b>	<b>63</b>
Economic Efficiency	63
Cost-Benefit Analysis: Background	66
Cost-Benefit Analysis: Basic Principles	66
Measuring Costs	67

Risk Equity Versus Equality of Marginal Costs per Life Saved	67
Marginal Analysis in CBA	68
Discounting	70
Risk Adjustment and CBA	71
Distributional Adjustments	72
Inflation	72
Valuing Human Life	72
Willingness to Pay and Willingness to Accept	73
Contingent Valuation	73
How Valuable Is the Last Year of Life?	73
Cost-Benefit Analyses of Heart Care Treatment	75
Cost-Effectiveness Analysis	76
Advantages of CEA	77
Cost-Utility Analysis, QALYs, and DALYs	77
An ACE Inhibitor Application of Cost-Effectiveness Analysis	78
QALYs Revisited: Praise and Criticism	78
Are QALYs Consistent with Standard Welfare Economics?	78
Extra-Welfarism	79
Sen's Capability Approach and QALYs	79
Linearity Versus What People Think	79
The Ageism Critique of QALYs	80
Conclusions	80
<i>Summary</i>	81
• <i>Discussion Questions</i>	81
• <i>Exercises</i>	82

## Part II Supply and Demand

### Chapter 5 Production of Health 85

The Production Function of Health	85
The Historical Role of Medicine and Health Care	88
The Rising Population and the Role of Medicine	88
What Caused the Mortality Rate Declines? Was It Medicine?	89
What Lessons Are Learned from the Medical Historian?	93
The Production of Health in the Modern Day	93
Preliminary Issues	93
The Contribution of Health Care to Population Health:	
The Modern Era	94
Is Health Care Worth It?	95
Prenatal Care	96
The World's Pharmacies	97
How Does Health Care Affect Other Measures of Health?	97
On the Importance of Lifestyle and Environment	98
Cigarettes, Exercise, and a Good Night's Sleep	99
The Family as Producer of Health	100
Social Capital and Health	100
Environmental Pollution	101
Income and Health	101

The Role of Schooling	101
Two Theories About the Role of Schooling	102
Empirical Studies on the Role of Schooling in Health	102
Conclusions	102
<i>Summary</i>	103
• <i>Discussion Questions</i>	103
• <i>Exercises</i>	104

## **Chapter 6 The Production, Cost, and Technology of Health Care 105**

Production and the Possibilities for Substitution	106
Substitution	106
What Degree of Substitution Is Possible?	107
Elasticity of Substitution	107
Estimates for Hospital Care	109
Costs in Theory and Practice	109
Deriving the Cost Function	110
Cost Minimization	111
Economies of Scale and Scope	112
Why Would Economies of Scale and Scope Be Important?	113
Empirical Cost-Function Studies	114
Difficulties Faced by All Hospital Cost Studies	114
Modern Results	116
Summary: Empirical Cost Studies and Economies of Scale	116
Technical and Allocative Inefficiency	116
Technical Inefficiency	116
Allocative Inefficiency	117
Frontier Analysis	118
The Uses of Hospital Efficiency Studies	119
For-Profit Versus Nonprofit Hospitals	120
Efficiency and Hospital Quality	120
Are Hospital Frontier Efficiency Studies Reliable?	120
Performance-Based Budgeting	121
Technological Changes and Costs	121
Technological Change: Cost Increasing or Decreasing?	121
Health Care Price Increases When Technological Change Occurs	122
Diffusion of New Health Care Technologies	124
Who Adopts and Why?	124
Other Factors That May Affect Adoption Rates	125
Diffusion of Technology and Managed Care	126
Conclusions	126
<i>Summary</i>	126
• <i>Discussion Questions</i>	127
• <i>Exercises</i>	127

## **Chapter 7 Demand for Health Capital 129**

The Demand for Health	129
The Consumer as Health Producer	129
Time Spent Producing Health	130
Labor–Leisure Trade-Offs	131
Trading Leisure for Wages	132
Preferences Between Leisure and Income	132



The Investment/Consumption Aspects of Health	133
Production of Healthy Days	133
Production of Health and Home Goods	134
Investment over Time	135
The Cost of Capital	135
The Demand for Health Capital	136
Marginal Efficiency of Investment and Rate of Return	136
The Decreasing <i>MEI</i>	136
Changes in Equilibrium: Age, Wage, and Education	137
Age	137
Wage Rate	138
Education	139
Empirical Analyses Using Grossman's Model	139
Obesity—The Deterioration of Health Capital	141
An Economic Treatment of Obesity	142
Economic Effects	143
Why Has Obesity Increased?	144
Conclusions	145
<i>Summary</i>	146
• <i>Discussion Questions</i>	146
• <i>Exercises</i>	146

## **Chapter 8 Demand and Supply of Health Insurance 148**

What Is Insurance?	148
Insurance Versus Social Insurance	149
Insurance Terminology	149
Risk and Insurance	150
Expected Value	150
Marginal Utility of Wealth and Risk Aversion	151
Purchasing Insurance	152
The Demand for Insurance	153
How Much Insurance?	153
Changes in Premiums	155
Changes in Expected Loss	155
Changes in Wealth	156
The Supply of Insurance	157
Competition and Normal Profits	157
The Case of Moral Hazard	159
Demand for Care and Moral Hazard	159
Effects of Coinsurance and Deductibles	162
Health Insurance and the Efficient Allocation of Resources	162
The Impact of Coinsurance	162
The Demand for Insurance and the Price of Care	166
The Welfare Loss of Excess Health Insurance	167
The Income Transfer Effects of Insurance	168
Conclusions	170
<i>Summary</i>	170
• <i>Discussion Questions</i>	171
• <i>Exercises</i>	171

**Chapter 9 Consumer Choice and Demand 173**

Applying the Standard Budget Constraint Model	174
The Consumer's Equilibrium	175
Demand Shifters	176
Health Status and Demand	178
Two Additional Demand Shifters—Time and Coinsurance	178
The Role of Time	178
The Role of Coinsurance	180
Issues in Measuring Health Care Demand	182
Individual and Market Demand Functions	182
Measurement and Definitions	183
Differences in the Study Populations	183
Data Sources	183
Experimental and Nonexperimental Data	184
Empirical Measurements of Demand Elasticities	184
Price Elasticities	184
Individual Income Elasticities	185
Income Elasticities Across Countries	186
Insurance Elasticities	187
Impacts of Insurance on Aggregate Expenditures	189
Other Variables Affecting Demand	189
Ethnicity and Gender	189
Urban Versus Rural	190
Education	190
Age, Health Status, and Uncertainty	191
Conclusions	191
<i>Summary</i>	192
• <i>Discussion Questions</i>	193
• <i>Exercises</i>	193

**Part III Information and Insurance Markets****Chapter 10 Asymmetric Information and Agency 195**

Overview of Information Issues	196
Asymmetric Information	196
On the Extent of Information Problems in the Health Sector	196
Asymmetric Information in the Used-Car Market: The Lemons Principle	198
Application of the Lemons Principle: Health Insurance	199
Inefficiencies of Adverse Selection	200
Experience Rating and Adverse Selection	201
The Agency Relationship	202
Agency and Health Care	202
Consumer Information, Prices, and Quality	203
Consumer Information and Prices	203

Consumer Information and Quality 204  
Other Quality Indicators 206  
Conclusions 207  
    *Summary* 208 • *Discussion Questions* 208 • *Exercises* 209

**Chapter 11 The Organization of Health Insurance Markets 210**

Loading Costs and the Behavior of Insurance Firms 210  
    Impacts of Loading Costs 211  
    Insurance for Heart Attacks and Hangnails 212  
    Loading Costs and the Uninsured 212  
Employer Provision of Health Insurance: Who Pays? 212  
    Spousal Coverage: Who Pays? 215  
    How the Tax System Influences Health Insurance Demand 215  
    Who Pays the Compensating Differentials?—Empirical Tests 217  
    Other Impacts of Employer Provision of Health Insurance 218  
Employer-Based Health Insurance and Labor Supply 219  
    Health Insurance and Retirement 219  
    Health Insurance and Mobility 219  
The Market for Insurance 221  
    The Market for Private Insurance 221  
    Insurance Practices 222  
    The Past 30 Years 223  
The Uninsured: An Analytical Framework 224  
    The Working Uninsured 227  
    The Impacts of Mandated Coverage 229  
Technological Change, Higher Costs, and Inflation 230  
    The Cost-Increasing Bias Hypothesis 231  
    Goddeeris's Model—Innovative Change over Time 231  
    Evidence on Technological Change and Inflation 233  
Conclusions 233  
    *Summary* 234 • *Discussion Questions* 234 • *Exercises* 235

**Chapter 12 Managed Care 236**

What Is the Organizational Structure? 237  
What Are the Economic Characteristics? 238  
The Emergence of Managed Care Plans 239  
    Employer-Sponsored Managed Care 240  
    Medicaid and Medicare Managed Care Plans 242  
    Managed Care Contracts with Physicians 242  
    Managed Care Contracts with Hospitals 243  
Development and Growth of Managed Care—Why Did It Take So Long? 244  
    Federal Policy and the Growth of Managed Care 244  
    The Economics of Managed Care 245

Modeling Managed Care	246
Modeling Individual HMOs	246
How Much Care?	247
What Types of Care?	247
Framework for Prediction	248
Where Managed Care Differs from FFS—Dumping, Creaming, and Skimping	248
Equilibrium and Adverse Selection in a Market with HMOs	249
How Does Managed Care Differ?—Empirical Results	251
Methodological Issues—Selection Bias and Quality of Care	251
Comparative Utilization and Costs	252
The RAND Study—A Randomized Experiment	252
Recent Evidence	252
Growth in Spending	255
Competitive Effects	256
Theoretical Issues	256
Managed Care Competition in Hospital Markets	257
Managed Care Competition in Insurance Markets	258
Managed Care and Technological Change	259
The Managed Care Backlash	259
Conclusions	262
<i>Summary</i>	262
• <i>Discussion Questions</i>	263
• <i>Exercises</i>	264

## **Chapter 13 Nonprofit Firms 265**

An Introduction to Nonprofits	265
Why Nonprofits Exist and Why They Are Prevalent in Health Care	266
Nonprofits as Providers of Unmet Demands for Public Goods	266
The Public Good–Private Good Aspect of Donations	267
Relevance to Health Care Markets	268
Nonprofits as a Response to Contract Failure	269
Applications of Contract Failure to Health Care	269
Financial Matters and the Nonprofit	270
Summary of the Reasons for the Prevalence of Nonprofits	270
Models of Nonprofit Hospital Behavior	270
The Quality–Quantity Nonprofit Theory	271
The Profit-Deviating Nonprofit Hospital	272
The Hospital as a Physicians’ Cooperative	274
Maximizing Net Revenue per Physician	274
A Comparison of the Quantity–Quality and the Physicians’ Cooperative Theories	275
The Evidence: Do Nonprofit Hospitals Differ from For-Profit Hospitals?	276

Summary of Models of Hospital Behavior	279
What Causes Conversion of Nonprofits into For-Profits?	279
The Relative Efficiency of Nonprofits Versus For-Profits	280
Are Nonprofit Health Care Firms Less Technically or Allocatively Efficient?—Hospital and Nursing Home Studies	280
Conclusions	281
<i>Summary</i>	281
• <i>Discussion Questions</i>	282
• <i>Exercises</i>	282

## **Part IV Key Players in the Health Care Sector**

### **Chapter 14 Hospitals and Long-Term Care 283**

Background and Overview of Hospitals	283
History	284
Organization	285
Regulation and Accreditation	286
Hospital Utilization and Costs	286
Competition and Costs	287
Closures, Mergers, and Restructuring	291
Quality of Care	293
Nursing Homes	294
Background and Costs	294
Quality of Care	295
Excess Demand	296
Financing Long-Term Care	297
Hospice, Home Health, and Informal Care	298
Conclusions	299
<i>Summary</i>	299
• <i>Discussion Questions</i>	300
• <i>Exercises</i>	301

### **Chapter 15 The Physician's Practice 302**

A Benchmark Model of the Physician's Practice	303
Do Physicians Respond to Financial Incentives?	305
Physician Agency and Supplier-Induced Demand	305
Modeling Supplier-Induced Demand	306
The Target Income Hypothesis	306
The Benchmark Model as a Synthesis	308
The Parallel Between Inducement and Marketing	309
What Do the Data Say About Supplier-Induced Demand?	309
Physician Fees, Fee Tests, and Fee Controls	309
Diffusion of Information and Small Area Variations	311
Contributions to These Variations	311
The Physician Practice Style Hypothesis	312
Multiple Regression Approaches	313
SAV and the Social Cost of Inappropriate Utilization	313
Other SAV Applications	315

Other Physician Issues and Policy Puzzles	315
Malpractice	315
Direct-to-Consumer Advertising of Drugs	316
Paying for Outcomes	316
Conclusions	317
<i>Summary</i>	317 • <i>Discussion Questions</i> 318 • <i>Exercises</i> 318

## **Chapter 16 Health Care Labor Markets and Professional Training 319**

The Demand for and Supply of Health Care Labor	319
Production Functions and Isoquants	320
Marginal Productivity of Labor	321
Factor Substitution and Labor Demand	322
The Supply of Labor	322
Factor Productivity and Substitution Among Factors	324
Measurement of Physician Productivity	324
The Efficient Utilization of Physician Assistants: Substitution Among Inputs	324
Health Manpower Availability and the Meaning of Shortages	325
Availability of Physicians	326
Economic Definitions of Shortages of Health Professionals	327
The Role of Monopsony Power: Shortages of Registered Nurses	330
Medical Education Issues and the Question of Control	332
Sources of Medical School Revenues	332
Capital Market Imperfections Justify Subsidies	332
Teaching Hospitals, Medical Schools, and Joint Production	333
Foreign Medical School Graduates	334
The Control of Medical Education	335
Control over Entry	335
Licensure and Monopoly Rents	336
Licensure and Quality	338
Other Physician Labor Issues	339
Specialization	339
Physician Income by Gender	340
Conclusions	341
<i>Summary</i>	341 • <i>Discussion Questions</i> 342 • <i>Exercises</i> 343

## **Chapter 17 The Pharmaceutical Industry 344**

Structure and Regulation	346
Competition	346
Barriers to Entry	347
Regulation	349
The Production of Health and Substitutability	350
Least-Cost Production	351

Insurance and Substitutability	352
Technological Change	353
Drug Pricing and Profits	354
Monopoly Pricing	354
Price Discrimination	355
Monopsony Pricing and Price Controls	356
Competition and Generic Entry	357
Research and Development (R&D) and Innovation	357
Investment Decisions	358
R&D Spending	359
Firm Size and Innovation	360
Prices, Price Regulation, and Innovation	360
Cost Containment	361
Copayments	361
Generic Substitutes	362
Drug Formularies	363
New Drugs and Health Care Spending	363
Conclusions	364
<i>Summary</i>	365
• <i>Discussion Questions</i>	365
• <i>Exercises</i>	366

## **Part V Social Insurance**

### **Chapter 18 Equity, Efficiency, and Need 367**

Efficiency and Competitive Markets	368
The Concept of Pareto Efficiency (Optimality)	368
Trading Along the Budget Line	369
The Competitive Equilibrium	370
The First Fundamental Theorem of Welfare Economics	371
Redistribution of the Endowment	371
Price Discrimination	372
Trade-offs Between Equity and Efficiency	372
Deviations from the Competitive Model in the Health Care Sector	373
The Assumptions Under Perfect Competition	373
Promoting Competition in the Health Care Sector	374
The Theorem of the Second Best	374
An Economic Efficiency Rationale for Social Health Insurance	375
Need and Need-Based Distributions	376
Health Care Needs and the Social Welfare Function	377
Norman Daniels's Concept of Health Care Need	381
Economic Criticisms of Need-Based Distributions	382
Horizontal Equity and Need	382
Theories of Social Justice	385

Utilitarianism	385
Rawls and Justice as Fairness	385
Liberalism, Classical and Modern	386
Conclusions	387
<i>Summary</i>	388 • <i>Discussion Questions</i> 388 • <i>Exercises</i> 389

## **Chapter 19 Government Intervention in Health Care Markets 390**

Economic Rationale for Government Intervention	390
Monopoly Power	391
Public Goods	392
Externalities	394
Other Rationales for Government Intervention	394
Forms of Government Intervention	395
Commodity Taxes and Subsidies	395
Public Provision	396
Transfer Programs	398
Regulation	398
Government Involvement in Health Care Markets	398
Support of Hospitals	399
The Hill-Burton Act	399
Department of Veterans Affairs and Department of Defense	400
Food and Drug Administration	400
Mandated Health Insurance Benefits	400
Tax Policy	401
Public Health	401
Other Government Programs	401
Government Failure	402
Who Does the Regulator Represent?	403
Bureaucracy and Efficiency	404
Conclusions	405
<i>Summary</i>	405 • <i>Discussion Questions</i> 406 • <i>Exercises</i> 406

## **Chapter 20 Government Regulation: Principal Regulatory Mechanisms 407**

Do the Laws of Supply and Demand Apply?	407
Objectives of Regulation	409
Regulatory Policy	409
Regulatory Instruments in Health Care	409
Regulation of the Hospital Sector	410
Empirical Findings on Regulation	410
Prospective Payment	412
Description of PPS	412
The Theory of Yardstick Competition and DRGs	413
On the Effects of Medicare's Prospective Payment System	417



- Regulation of Physician Payment 423
  - UCR Reimbursement, Assignment, and Alternative Payment Mechanisms 424
  - Relative Value Scales 424
- Antitrust 426
  - Enforcement 426
  - Exemptions 426
  - Measuring Monopoly Power 427
  - Antitrust Procedures 429
  - The Elzinga-Hogarty (EH) Criterion 430
- Conclusions 432
  - Summary* 432 • *Discussion Questions* 433 • *Exercises* 433

**Chapter 21 Social Insurance 435**

- Social Insurance Policies and Social Programs 435
  - Program Features 436
- Historical Roots of Social Insurance 437
  - European Beginnings 437
  - Early Experience in the United States 438
  - The Establishment of Medicare and Medicaid 438
- The Patient Protection and Affordable Care Act (PPACA) of 2010 439
- Medicare and Medicaid in the United States 440
  - Medicare 440
    - Part D Prescription Drug Insurance 442
  - Medicaid 445
    - Medicaid Eligibility 446
    - The Medicaid-Medicare Relationship 449
    - Medicare and Medicaid: Conflicting Incentives for Long-Term Care 449
    - Children’s Health Insurance Program—CHIP 450
- Public Insurance and Health 451
- The Effects of Medicare and Medicaid 454
  - Costs and Inflation 455
  - Health Status 458
  - Medicare: Recent Changes and Future Prospects 459
- Criticisms of the U.S. Health Care System 462
- Conclusions 463
  - Summary* 464 • *Discussion Questions* 464 • *Exercises* 465

**Chapter 22 Comparative Health Care Systems 466**

- Contemporary Health Care Systems 466
  - A Typology of Contemporary Health Care Systems 466
- The United Kingdom—The National Health Service 467
  - The National Health Service 469

China—An Emerging System	473
The Canadian Health Care System	476
Background	476
Physician Fees and Quantity	478
Why Are Fees and Hospital Costs Lower in Canada?	479
Administrative Costs	479
A Comparison	481
Different Systems: The Public's Evaluation	484
Differences in Health Care Spending Across Countries	487
A Model of Health Expenditure Shares	487
Conclusions	489
<i>Summary</i>	490
• <i>Discussion Questions</i>	491
• <i>Exercises</i>	491

## **Chapter 23 Health System Reform 492**

Goals of Reform	492
Basic Issues in Reform	493
The Costs of Universal Coverage	494
Ensuring Access to Care	495
Employer Versus Individual Mandates	495
Separation of Health Insurance from Employment	496
Single Payer Versus Multiple Insurers	497
Competitive Strategies	498
Development of Alternative Delivery Systems	499
Consumer-Directed Health Plans and Health Savings Accounts	499
Other Market Reforms	500
Representation of the Competitive Approach	501
Health System Reform and International Competitiveness	502
Quality of Care	504
The Patient Protection and Affordable Care Act (PPACA) of 2010	505
Conclusions	510
<i>Summary</i>	510
• <i>Discussion Questions</i>	511
• <i>Exercises</i>	512

## **Part VI Special Topics**

### **Chapter 24 The Health Economics of Bads 513**

An Introduction to Bads	514
Models of Addiction	516
Imperfectly Rational Addiction Models	516
Myopic Addiction Models	516
Rational Addiction	517
Rationales for Public Intervention	519
Other Interventions	520
Advertising Restrictions on Cigarettes and Alcohol	520
The Possible Effects of Brand Switching	522

Increased Demand or Brand Switching?	522
Advertising and Alcohol Consumption	523
Excise Taxes and Consumption of Cigarettes and Alcohol	523
The Consumption-Reducing Effects of Excise Taxes in Theory	524
Excise Taxes and Cigarette Consumption in Practice	524
Excise Taxes and Alcohol Consumption	527
Conclusions	528
<i>Summary</i>	529
• <i>Discussion Questions</i>	529
• <i>Exercises</i>	530

**Chapter 25 Epidemiology and Economics: HIV/AIDS in Africa 531**

Concepts from Epidemiology	531
Economic Epidemiology	534
Rational Epidemics	534
The Prevalence Elasticity of Demand for Prevention	535
The Economic Consequences of Epidemics	535
The Difficulty of Eradicating Diseases	537
Information	537
The Role of Government in Battling Epidemics	538
Case Study: HIV/AIDS in Africa	539
HIV/AIDS	539
Costs of AIDS	540
Fighting AIDS	543
Economic Theory and African Reality	545
Conclusions	547
<i>Summary</i>	547
• <i>Discussion Questions</i>	548
• <i>Exercises</i>	548

*Glossary* 549

*References* 556

*Name Index* 590

*Subject Index* 595

# PREFACE

This is the seventh edition of a book conceived in 1990 and first published in 1993. The world of health economics has changed since 1990. Our first (1993) edition made but a single reference to “managed care.” Until this seventh edition, we spoke of the United States as the only country without a universal health insurance plan. This, of course, changed with the passage of the Patient Protection and Affordable Care Act (PPACA) in March 2010.

## NEW TO THE SEVENTH EDITION

- The most important change from the previous edition is the passage of the 2010 Patient Protection and Affordable Care Act (PPACA). While legal challenges are still occurring, PPACA has had major impacts on health programs at the federal and state levels. We discuss PPACA in passing in several chapters, and in detail in Chapter 23.
- Chapter 7 on health capital includes a new section on obesity. It looks at the reasons for the recent explosion in obesity and overweight status in the United States and elsewhere, and the economic costs that it imposes. Chapter 19 also includes a feature on a “junk food” tax directed toward obesity.
- Chapter 17 provides major updates on changes in the pharmaceutical industry. The changes in drug copayments, as well as the movements of several blockbuster drugs into generic status, have led to considerable impacts on firms in the industry.
- Chapter 21 (Social Insurance) continues with the most up-to-date evaluations of the Medicare Part D drug benefit. It also updates research and policy work on the inherent conflicting incentives between the Medicare and Medicaid programs that jointly lead to inefficiencies in the provision of health services for the elderly and the poor.
- Chapter 22 introduces a new section on the Chinese economy. The Chinese health economy has moved away from the command system of the 1950s through the 1970s, instituting a more incentive-based system, with its conflicting impacts. Chapter 22 also provides updated comparative survey information on comparative satisfaction across eleven countries, looking in particular at differences in access and in costs.
- Chapter 23, health care reform, looks carefully at PPACA. The analysis suggests that there will be major improvements in access to care, but it cautions that cost control incentives are yet incomplete.
- Finally, Chapter 25, “Epidemiology and Economics: HIV/AIDS in Africa” updates a popular chapter that we introduced in the fourth edition and we have continued to revise. The constant revision shows the enormous changes in treatment cost and delivery that have occurred since our initial coverage in 2003.

As in all previous editions, we have sought to provide the most current data for student and their instructors. We have inserted major data updates into Chapters 1, 11, 12, 14, 21, 22, 24, and 25, and we have included numerous new features.

From the beginning, we have sought to assist instructors in conveying a clear, step-by-step understanding of health economics to their students. We have also believed it important for instructors to demonstrate what health economics researchers are doing in theory and in empirical work. The book synthesizes contemporary developments around a set of economic principles including maximization of consumer utility and economic profit, and it makes these principles accessible to undergraduate as well as to graduate students. Rather than focusing on institutions specific to the health care economy, we have chosen to emphasize core economics themes as basic as supply and demand, as venerable as technology or labor issues, and as modern as the economics of information. We have

constantly tried to improve accessibility to the book for the wide range of health services students and practitioners.

Students must have a working knowledge of the analytical tools of economics and econometrics to appreciate the field of health economics. Some students may be ready to plunge directly into Chapter 5, “The Production of Health” upon completion of the introductory Chapter 1. However, Chapters 2 through 4 help the students and their instructors to develop or to review needed analytical concepts before tackling the core subject matter. In Chapter 2, students with as little as one semester of microeconomics may review and study how economists analyze problems, using relevant health economics examples. Chapter 3 provides a review of core statistical tools that characterize modern economic and health services analyses. Chapter 4 completes the core economic concepts by reviewing the concept of economic efficiency, and showing how cost-benefit, cost-effectiveness, and cost-utility analyses fit into the general economic framework.

Consistent with an emphasis on clear exposition, the book makes extensive use of graphs, tables, and charts. As in all previous editions, we require no calculus. Discussion questions and exercises help students master the basics and prompt them to think about the issues. We also include features on up-to-date applications of theory and policy developments, as well as occasional tidbits containing purely background information.

We caution that some chapters, such as those on insurance and on regulation, although developed without advanced mathematics, are logically complex and will require considerable effort. No painless way is available to appreciate the scope of the contributions that scholars have made in recent years. More advanced students of the health care economy who seek further challenges can utilize a comprehensive references section, with 1,100 sources, to enrich their (and our) work through referral to the original sources.

## ADDITIONAL SOURCES

The Internet now contains tables and charts that were once available only in book form and then only after several years. We have chosen to focus on those sites that we believe to be both long lasting and reliable.

- Bureau of the Census, for health insurance, (<http://www.census.gov/hhes/www/hlthins/hlthins.html>)
- Centers for Disease Control and Prevention ([www.cdc.gov](http://www.cdc.gov))
- Centers for Medicare and Medicaid Services, for research, statistics, data, and systems ([www.cms.hhs.gov/home/rsds.asp](http://www.cms.hhs.gov/home/rsds.asp))
- Kaiser Family Foundation ([www.kff.org](http://www.kff.org)), specializing in studies of health insurance
- National Institutes of Health ([www.nih.gov](http://www.nih.gov))
- Organization for Economic Cooperation and Development, for international data ([www.oecd.org](http://www.oecd.org))
- Social Security Administration, for research and analysis ([www.ssa.gov/policy/research\\_subject.html](http://www.ssa.gov/policy/research_subject.html))

Health-related material is increasingly accessible both in print and on the Internet. University-affiliated professionals and their students will most often discover that their libraries have extensive electronic access to a wide range of journals. Most health economists make a point of browsing *Health Affairs*, an up-to-date policy journal. *Health Economics*, *Journal of Health Economics*, and *International Journal of Health Care Finance and Economics* have emerged as the leading technical journals that specialize in health economics. As one can see from our comprehensive reference section, many other specialized journals, including health services and medical journals not often referenced by economists, also address readers’ needs. In the popular press, the *New York Times* and the *Wall Street Journal* also provide excellent health economics coverage.

*The Handbook of Health Economics* (Elsevier), a two-volume set published in 2000, emerged as an invaluable source for specific topics, with more detail and more mathematic rigor than any text, including this one. The *Elgar Companion to Health Economics*, published in 2006, provides

both useful updates and important new topics. They are not texts, however. Our book, with its graphical analysis, discussion questions, and problem sets, provides a valuable complement to both *Handbook* and *Companion* offerings.

## **ALTERNATIVE COURSE DESIGNS**

The economics of health and health care encompasses an evolving literature with no single “correct” order for the course design. U.S. economists typically organize topics through markets, and include the roles of government much later. International health economists and population and public health students and scholars often assign the governmental sector far more importance; it is “public” health, after all. No matter how we construct it, a text is necessarily linear in that one chapter must follow another.

Our text offers instructors considerable flexibility. We have divided the 25 chapters into six parts:

- I. Basic Economics Tools (Chapters 1–4)
- II. Supply and Demand (Chapters 5–9)
- III. Information and Insurance Markets (Chapters 10–13)
- IV. Key Players in the Health Care Sector (Chapters 14–17)
- V. Social Insurance (Chapters 18–23)
- VI. Special Topics (Chapters 24–25)

The categories are not entirely exclusive. Chapter 8, looking at the demand and supply of insurance, is as important to Part III on insurance as it is to Part II on demand and supply of goods.

From front to back, we follow an “economics” model in which we first examine consumers and firms in a world without government and governmental policies. As a result, explicit discussions of government policies do not come until Chapter 19, although we examine regulation, licensing, and mandates in reference to other topics much earlier. Many economics instructors may wish to follow the chapters in the book’s order.

Instructors with population health, public health, or policy interests may wish to “tool up” on some of the earlier analyses and then skip directly to Part V, in which we look at social insurance, health care regulation, and health care reform. After that, they may wish to browse selected topics. Although some analyses build on one another within chapters, we seek to minimize cross-referencing among chapters.

## **Instructor Resources**

*The Economics of Health and Health Care* is connected to the Instructor Resource Center available at [www.pearsonhighered.com/folland](http://www.pearsonhighered.com/folland). Registration is simple and gives you immediate access to new titles and new editions. As a registered faculty member, you can download resource files. The following supplements are available to adopting instructors:

- Instructor’s Manual with Test Item File
- Power Point Slides

*The International Handbook On Teaching And Learning Economics* (Elgar, 2012) has a section by Allen Goodman on the teaching of health economics, along with over 70 chapters on general course content, specific fields, pedagogic techniques, and the scholarship of the teaching enterprise.

## ACKNOWLEDGMENTS

This book emerges from years of classroom and professional interaction, and we thank those students and colleagues who have discussed the economics of health and health care with us and who have challenged our ideas. Annie Todd suggested the endeavor back in 1989 and introduced us to each other, and Tom McGuire convinced us that we could succeed. We have been blessed with outstanding editors—Jill Leckta, Leah Jewell, Rod Banister, Gladys Soto, Marie McHale, Michael Dittamo, Chris Rogers, Susie Abraham, and for this edition, David Alexander and Lindsey Sloan.

We have received outstanding support from our publishers, with John Travis, Maureen Wilson, Kelly Warsak, Melissa Feimer, Clara Bartunek and Kerri Tomasso managing the production of previous editions. Nancy Freihofer is our most capable manager for this edition.

We extend our gratitude to instructors at hundreds of universities in the United States and around the world, who have helped us pass the market test by adopting the book. Numerous professional colleagues generously offered their time and energy to read and critique various chapters. We thank:

Carson Bays, East Carolina University	James Hamilton, Wayne State University	Martin Pernick, University of Michigan
James Burgess, Veterans Administration	Janet Hankin, Wayne State University	Uwe E. Reinhardt, Princeton University
Frank Chaloupka, University of Illinois–Chicago	Govind Hariharan, Kennesaw State University	Joann Richards, Oakland University
Hope Corman, Rider College	Ronald Horwitz, Oakland University	John Rizzo, Stony Brook University
Merton Finkler, Lawrence University	Harold Hotelling, late of Lawrence Technological University	T. Paul Schultz, Yale University
Colleen Flood, University of Toronto	Kathy A. Kelly, University of Texas at Arlington	Ya-chen Tina Shih, MEDTAP International, Inc.
Howard Forman, Yale University	Paul Langley, University of Colorado	Kosali Simon, Indiana University
Andrew D. Foster, Brown University	William Low, University of Northern British Columbia	Robert J. Sokol, Wayne State University
Gary Fournier, Florida State University	Timothy McBride, Washington University	John M. Tilford, University of Arkansas
H. E. Frech, III, University of California–Santa Barbara	Tom McCarthy, National Economic Research Associates	Scott Vander Linde, Calvin College
Martin Gaynor, Carnegie Mellon University	Catherine McLaughlin, University of Michigan	Rob Verner, Kent State University
John Goddeeris, Michigan State University	Eric Nauenberg, University of Toronto	Jian Wang, Peking University
Warren Greenberg, George Washington University	Eleanor Nishiura, Wayne State University	Kenneth Warner, University of Michigan
Michael Grossman, City University of New York	John Nyman, University of Minnesota	Gerard Wedig, University of Pennsylvania
Jack Hadley, Center for Health Policy Studies	Albert A. Okunade, University of Memphis	Burton Weisbrod, Northwestern University
		William White, Cornell University
		Barbara Wolfe, University of Wisconsin–Madison
		Michael Woolley, University of Southern California

We also thank the following colleagues who, with their classes, test-taught and test-learned parts of the text:

Randall Ellis, Boston University	Lee R. Mobley, Research Triangle Institute	Richard Scheffler, University of California–Berkeley
Gail Jensen, Wayne State University	Kathleen Possai, Wayne State University	Nancy Wolff, Rutgers University
Thomas McGuire, Harvard University		

In the 22 years since we began this project, our wives have patiently supported our efforts and our young daughters have grown into talented young women. All have watched their husbands and fathers learn about health economics. We remain eternally grateful.

**S. F.**

*Rochester, Michigan*

**A. C. G.**

*Detroit, Michigan*

**M. S.**

*Rochester, Michigan*

# Introduction

- What Is Health Economics?
- The Relevance of Health Economics
- Economic Methods and Examples of Analysis
- Does Economics Apply to Health and Health Care?
- Is Health Care Different?
- Conclusions

Health care accounts for over one-sixth of the U.S. economy! In 2009, the year before passage of the Patient Protection and Affordable Care Act (PPACA), the Obama reform legislation, about 50 million Americans at any moment in time lacked health insurance. Health, health care costs, and health insurance have dominated the economic and political landscape in the United States and many other countries. Health economics studies the allocation of resources to and within the health economy. Because the health care sector has become the largest sector of the U.S. economy, and its share of gross domestic product (GDP) is expected to grow well into the twenty-first century, we should not be surprised that health economics has emerged as a distinct specialty within economics.

Our table of contents provides an overview of the scope of health economics and the emphasis of this text. Demand and supply of specific health services are prominent. Private health insurance markets critically define the U.S. workplace, so we carefully examine these markets. Government, through its social programs and power to regulate, receives close attention. Because a hospitalization is in many ways different from a trip to a supermarket, we also concentrate on issues such as information, quality of care, and equity of access. Finally, we look to the health care systems of other countries for information on their practices and for potential insights on the policy issues that dominate the political landscape.

In this first chapter, we provide further background information on health economics and health economists. We follow with a broad overview of the magnitude and importance of the health care sector and with an introduction to some major policy concerns. As our final goal, we seek to promote the theme that economics helps explain how health care markets function. We focus on methods used in economic analyses and address two recurring questions: Is health care different, and does economics apply? Despite stressing the distinctive features of health care services and markets, we answer both in the affirmative. With appropriate modifications



to conventional analytical tools, economics is relevant and useful. As we shall see throughout the book, although there is continuing controversy on many major issues, health economists have provided insight and solutions to most problems of academic and policy interest.

## WHAT IS HEALTH ECONOMICS?

Health economics is defined by *who* health economists are, and *what* they do! Morrisey and Cawley (2008) examine the field of health economics and those who shape the discipline. Their 2005 survey of U.S. health economists describes who they are.

1. **Training:** Almost all (96 percent) held academic doctorate degrees. Nearly three-quarters of those with doctorates received their degrees in economics.
2. **Employment:** 64 percent worked in university settings; 15 percent worked for nonprofit organizations; and 12 percent worked in government, mainly the federal government.
3. **Academic Unit:** Only 24 percent of health economists held their appointments in economics departments. More (26 percent) were found in schools of public health, and another 18 percent were employed in schools of medicine.

Second, and of greater relevance to us, Morrisey and Cawley's survey gathered information on what health economists actually do. Health economists draw on various sub-disciplines of training within economics, including labor economics, industrial organization, public finance, and cost-benefit analyses.

Throughout this book, we describe many specific research studies. At this time, consider that the United States devotes by far the largest share of GDP to health care spending (over one-sixth), and its per-capita health care spending (over \$8,000) greatly exceeds that of any other country. Most health economists agree that these spending patterns reflect the rapid rate of adoption of new technology in the United States. The United States does not have a very impressive record in terms of broad health outcomes indicators such as life expectancy and infant mortality. Critics of the U.S. health care system often wonder what Americans are getting for their money. Policymakers and health economists seek to determine whether spending on new technology is worth it. Arguably, there is no more important issue.

Consider, for example, a new surgical procedure for a patient with acute myocardial infarction (heart attack). It is not enough to estimate the immediate cost impact of the new procedure and the expected benefit to the patient in terms of short-term survival. By impacting the patient's health for many years, the new treatment will affect spending well into the future. David Cutler (2007) develops a framework to address these complex interrelationships in "The Lifetime Costs and Benefits of Medical Technology." He analyzes *revascularization*, a set of surgical procedures such as coronary bypass and angioplasty that restore blood flow. He looks at a group of Medicare patients who have had heart attacks and he tracks them for up to 17 years. Chapter 4 devotes considerable attention to Cutler's work, but here we highlight his conclusion that revascularization costs \$33,000 for an extra year of life. Is this worth it? Most would agree that it is!

Health care costs in general, and technology-related costs in particular, are relevant to all countries (Box 1-1 provides an international perspective). Health economics is still a relatively new discipline with an evolving scope and pedagogy, and neither it, nor we, will provide answers to all the health system questions that nations face. Despite this caveat, we cannot think of any field of study that is more relevant and has more urgently needed solutions. The remainder of this chapter pursues this theme, but as a more practical illustration of the relevance of health economics, we paraphrase a saw that students and teachers have often heard at their freshman convocations: "Turn to the left and then turn to the right. Before too long, chances are pretty good that one of you will be working in a health-related field."

**BOX 1-1****Technological Change and Health Care Costs—Why Rising Health Care Costs Affect All Nations**

In a March 2005 speech to the National Association of Business Economics, then-Chairman of the Council of Economic Advisers Harvey Rosen noted that over the last several decades, the quality of health care has improved—diagnostic techniques, surgical procedures, and therapies for a wide range of medical problems have all continually improved. Treatment of a heart attack today is simply not the same “commodity” as treatment of a heart attack in 1970. Although innovations like coronary bypass surgery and cardiac catheterization have raised expenditures per heart patient, they have actually reduced the prices of obtaining various health outcomes, such as surviving hospitalization due to a heart attack.

Rosen noted that some improvements in medical technique were quite inexpensive. Prescribing aspirin for heart attack victims leads to a substantial improvement in their survival probabilities. But new medical technologies were often costly. For example, it cost about \$2 million to acquire a PET (positron emission tomography) machine, which can detect changes in cells before they form a tumor large enough to be spotted by X-rays or MRI. Such costly improvements lead medical expenditures to grow.

This technology-based theory also helps explain why countries as different as the United States, the United Kingdom, or Japan have all experienced increases in health care expenditures. Rosen argued that these societies have at least one thing in common—they all have access to the same expensive innovations in technology. The technology-based explanation puts any debate over cost containment in a new light. Is it a bad thing if costs are rising mostly because of quality improvements? A key question in this context is whether people value these innovations at their incremental social cost. No one knows for sure, but economist Dana Goldman reiterates a provocative insight: “If you had the choice between buying 1960s medicine at 1960s prices or today’s medicine at today’s prices, which would you prefer?” A vote for today’s medicine is validation of the improvement and willingness to pay for improved quality!

*Sources:* Rosen, Harvey S. Chairman, Council of Economic Advisers, Comments on Health Care at the Conference of the National Association of Business Economics, March 21, 2005, <http://georgebush-whitehouse.archives.gov/cea/nabe-healthcare032105.pdf>; Goldman, Dana P., “Pressure from Rising Health-Care Costs: How Can Consumers Get Relief?” [www.rand.org/commentary/102305PE.html](http://www.rand.org/commentary/102305PE.html).

**THE RELEVANCE OF HEALTH ECONOMICS**

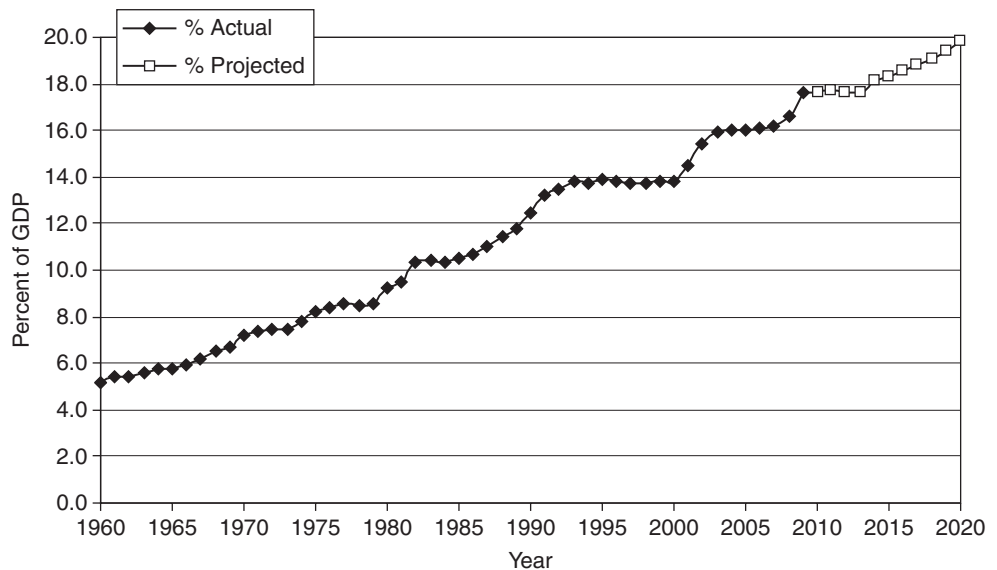
The study of health economics is important and interesting in three related ways: (1) the size of the contribution of the health sector to the overall economy, (2) the national policy concerns resulting from the importance many people attach to the economic problems they face in pursuing and maintaining their health, and (3) the many health issues that have a substantial economic element.

**The Size and Scope of the Health Economy**

The health economy merits attention for its sheer size, constituting a large share of GDP in the United States, as well as in other countries. It also represents a substantial capital investment and a large and growing share of the labor force.

**Health Care’s Share of GDP in the United States**

By the second decade of the twenty-first century, more than \$1 out of every \$6 spent on final goods and services in the U.S. economy went to the health sector. As recently as 1980, the share of GDP (the market value of final goods and services produced within the borders of a country in a year) devoted to health care was \$1 in \$11, and in 1960 it was just \$1 in \$20. Figure 1-1 tracks the health economy’s share of GDP from 1960 to 2009, with projections through 2020. The conclusion? The health care sector is a large and growing portion of our economy.



**FIGURE 1-1** U.S. Health Expenditure Shares, 1960–2020 *Source:* Centers for Medicare and Medicaid Services: [http://www.cms.gov/NationalHealthExpendData/25\\_NHE\\_Fact\\_Sheet.asp](http://www.cms.gov/NationalHealthExpendData/25_NHE_Fact_Sheet.asp), accessed August 8, 2011.

In calculating the share of GDP spent on health care, we net out the effects of general inflation. Therefore, only three major possibilities exist to explain the substantially increased ratios shown in Figure 1-1:

1. People may be buying more health services. Patients may be consulting with health care providers more frequently, doctors may be ordering more tests, or they may be prescribing more drugs.
2. People may be buying higher-quality health services, including products and services that previously were not available. Laser surgery, organ transplants, measles vaccines, and new treatments for burn victims, unavailable in 1960, have raised the quality of care. Economic theory suggests that people are willing to pay more for better quality.
3. Health care inflation may be higher than the general inflation rate. Higher incomes and the increased prevalence of insurance, including large government programs such as Medicare and Medicaid, may have led to increased health care prices over time.

We seek to understand these phenomena and their contributions to total spending. The study of demand, insurance, production, technology, and labor supply, among other topics, will help meet this challenge.

### Health Care Spending in Other Countries

Examining the health economies of other countries enhances our understanding of the U.S. health economy. Many countries have large health care sectors and face the same major issues. Table 1-1 shows how health care spending as a share of GDP grew rapidly in most countries between 1960 and 1980. A more mixed picture emerges after 1980. The health care share in the United States continued to grow in each period after 1980 shown in Table 1-1, but growth was more modest in most other countries.

The data also indicate the relative size of the U.S. health economy compared to that of other countries. For example, health care's share of GDP in the United States is nearly twice as large as the share in the United Kingdom—a country with national health insurance. Is care costlier in the United States? Is it higher quality care, or are we simply consuming more?

**TABLE 1-1 Health Expenditures as Percent of GDP in Selected OECD Countries**

Country	1960	1970	1980	1990	2000	2009 <sup>a</sup>
Australia			6.3	6.9	8.3	8.5
Austria	4.3	5.2	7.5	8.4	9.9	11.0
Belgium	3.9	6.3	7.2	8.6	10.4	10.9
Canada	5.4	6.9	7.0	8.9	8.8	11.4
Czech Republic				4.7	6.5	8.2
Denmark			8.9	8.3	8.3	11.5
Finland	3.8	5.5	6.3	7.7	7.0	9.2
France	3.8	5.4	7.0	8.4	9.6	11.8
Germany		6.0	8.4	8.3	10.3	11.6
Greece		5.4	5.9	6.6	7.8	9.7
Hungary					6.9	7.4
Iceland	3.0	4.7	6.3	7.8	9.5	9.7
Ireland	3.7	5.1	8.3	6.1	6.3	9.5
Italy				7.7	8.1	9.5
Japan	3.0	4.6	6.5	6.0	7.7	8.1
Korea			3.4	4.0	4.6	6.9
Luxembourg		3.1	5.2	5.4	5.8	7.8
Mexico				4.8	5.6	6.4
Netherlands			7.4	8.0	8.0	12.0
New Zealand		5.2	5.9	6.9	7.7	10.3
Norway	2.9	4.4	7.0	7.6	8.4	9.6
Poland				4.8	5.5	7.4
Portugal		2.5	5.3	5.9	8.8	10.2
Slovak Republic					5.5	9.1
Spain	1.5	3.5	5.3	6.5	7.2	9.5
Sweden		6.8	8.9	8.2	8.2	10.0
Switzerland	4.9	5.4	7.3	8.2	10.3	11.4
Turkey			3.3	3.6	4.9	6.2
United Kingdom	3.9	4.5	5.6	6.0	7.2	9.8
United States	5.1	7.0	8.7	11.9	13.2	17.4

<sup>a</sup>2009 or most recent year. OECD data for the United States may differ slightly from values reported by the Centers for Medicare and Medicaid Services.

Source: Organization for Economic Cooperation and Development (OECD) Health Care Data, June 2011.

## Importance of the Health Economy in Personal Spending

Because it accounts for such a large share of the domestic product, the size of the health economy is also reflected through other key indicators. Two of these are especially easy to relate to at the personal level: (1) share of income spent on medical care and (2) number of jobs in the health economy.

Table 1-2 provides data on how U.S. consumers spend their disposable incomes. It shows that in 2009, consumers spent 17.9 percent of their budgets on health care, as opposed to 13.8 percent on food, and 18.8 percent on housing. When we add drug expenditures, we find that slightly more than \$1 in \$5 of consumption expenditures goes to medical care plus drugs and sundries. These figures represent a major shift in spending patterns. As recently as 1960, food represented about 25 percent of spending, housing about 15 percent, and medical care only 5 percent.

**TABLE 1-2 Total Consumption Expenditures (in \$ Billions) by Type, 2009**

	2009	% of Total
<b>Personal consumption expenditures</b>	10,001.3	<b>100.0</b>
<b>Durable goods</b>	1,026.5	<b>10.3</b>
<b>Nondurable goods</b>	2,807.8	<b>28.1</b>
Food	1,381.5	13.8
Clothing and shoes	322.2	3.2
Gasoline, fuel oil, and other energy goods	303.7	3.0
Other	800.4	8.0
<b>Services</b>	6,167.0	<b>61.7</b>
Housing	1,876.3	18.8
Household operation	294.7	2.9
Transportation	290.1	2.9
Health care including health insurance	1,786.2	17.9
Recreation	378.8	3.8
Other	937.3	9.4
Health Care including Health Insurance	1,786.2	17.9
Physicians	396.2	4.0
Dentists	105.0	1.0
Other professional services	271.1	2.7
Hospitals and nursing homes	850.9	8.5
Health insurance	163.0	1.6

Source: U.S Department of Commerce, Bureau of Economic Analysis: <http://www.bea.gov/national/nipaweb/TableView.asp?SelectedTable=70&Freq=Year&FirstYear=2008&LastYear=2009>, accessed May 19, 2011.

### Importance of Labor and Capital in the Health Economy

The flip side of spending is reflected through the jobs that have been created in the health economy. As shown in Table 1-3A, 15.5 million people—11.1 percent of all employed civilians—worked at various health services sites in 2009. These numbers continued to grow despite the loss of over 5 million jobs in the U.S economy between 2007 and 2009. Hospitals dominate, employing 40.5 percent of health care workers. Other major employers include offices and clinics of physicians (10 percent), nursing care facilities (12.1 percent), and offices and clinics of dentists (5.2 percent).

Table 1-3B provides information on specific health care occupations and their growth since 1970. In 2009, there were over 972,000 physicians and almost 268,000 pharmacists. The nursing sector alone consisted of over 3 million people with over three-quarters of them trained as registered nurses.

The considerable growth in health care personnel is evident. In 1970, there were 334,000 physicians, or 164 physicians per 100,000 people. By 2009, the number of physicians had increased by 191 percent to 972,000 or 317 per 100,000 population. The number of registered nurses had more than tripled by 2009, with their number per 100,000 population more than doubling from 369 to 842.

Reflecting the increases in spending, the health care sector serves increasingly as a source of employment. Thus, cutbacks in spending on health care, if proposed and implemented, would typically mean cutbacks in employment opportunities.

In addition to labor, a substantial amount of capital has been drawn to the U.S. health care system. The number of nursing home beds increased from about 1.3 million in 1976 to about 1.7 million in 2009 (beds per capita, however, decreased slightly). The number of short-term

**TABLE 1-3A Persons Employed at Health Services Sites, 2009**

Category	Employment (in Thousands)
All Employed Civilians	139,877
At Health Services Sites	15,478
Hospitals	6,265
Offices and clinics of physicians	1,555
Nursing care facilities	1,869
Home health care services	967
Outpatient care centers	1,102
Offices and clinics of dentists	801
All other sites	2,919

Source: U.S. Department of Health and Human Services, *Health United States*, 2010, Table 105.

**TABLE 1-3B Active Health Personnel and Number per 100,000 Population (in Parentheses)**

Occupation	1970 Total	1980 Total	1990 Total	2000 Total	2009 Total
Physicians	334,028 (164)	467,679 (206)	615,421 (247)	813,770 (289)	972,400 (317)
Licensed Practical and Licensed Vocational Nurses	—	—	—	679,470 (241)	728,670 (237)
Registered Nurses	750,000 (369)	1,272,900 (562)	1,789,600 (720)	2,189,670 (778)	2,583,770 (842)
Pharmacists	112,750 (55)	142,780 (63)	161,900 (65)	212,660 (76)	267,860 (87)
U.S. Population	203,302,031	226,542,199	248,709,873	281,421,906	306,964,822

Source: U.S. Department of Health and Human Services, *Health United States*, Various Years.

hospital beds (as distinguished from nursing homes) peaked in the late 1970s, at almost 1.5 million, but the total number has since leveled at approximately 950,000. There are also considerable and growing amounts of other capital—such as diagnostic equipment—per bed.

### Time—The Ultimate Resource

Data on health care expenses and labor and capital inputs reflect only some of the items used by people to produce health. Inputs that are not bought and sold in the marketplace are also important. These include peoples' own contributions of time and effort in producing health care and entail real costs to society.

For example, when people use their own time to produce better health for themselves, or for loved ones, as caregivers, the cost to the individuals and society is the value of the leisure that they forego. Adults who are taking care of their elderly parents for two hours per day, seven days per week, provide care that might otherwise have to be purchased in the market for \$15 per hour or more. In this simple illustration, the caregivers provide care worth over \$10,500 per year. Though such examples are not necessarily the population norm, these time costs must be added to our measured health care costs.

We have stressed inputs, but the contribution of health resources to the economy is ultimately a measure of the value of the output—health itself. We measure the values of improvements to our health in both consumption and production. We value health both for its own sake and for its contribution to the production of other goods. The intrinsic value of being healthy is ultimately the value we attach to life and limb, which people commonly describe as infinite in certain circumstances, and at least substantial in others. The value of health in the production of other goods is exemplified not just in reduced absenteeism rates but also in output per worker on the job. In both its consumption and production aspects, the output of the health sector makes a substantial contribution to the economy.

### The Importance Attached to Economic Problems of Health Care Delivery

The health sector receives attention from policymakers because of its widely perceived problems. The substantial resources devoted to health care are reflected in a more meaningful way through the average level of this nation's spending for health care. Table 1-4 provides various measures of health care spending and its growth since 1960.

Table 1-4 shows how national health expenditures (NHE) grew from \$27 billion in 1960 to \$2,486 billion in 2009. During that period, the U.S. population grew by 65 percent from 186 million to 307 million. Thus, NHE per capita rose by a multiple of 55, from \$147 in 1960 to \$8,086 in 2009.

However, the real increase is what matters most. Prices, as measured by the broad-based consumer price index (CPI), rose 625 percent over the same period. After deflating by the CPI, we find that real expenditures per capita in 2009 were 7.6 times the 1960 level—still a hefty increase.<sup>1</sup>

**TABLE 1-4 National Health Expenditures and Other Data for Selected Years**

Year	NHE (\$bil)	% Growth in NHE over Previous Year	GDP (\$bil)	NHE per Capita	NHE % GDP	CPI	Price Indices	
							Hospital + Related Services	Physician Services
1960	27.3		526	147	5.2	29.6		21.9
1970	74.8	10.6	1,038	356	7.2	38.8		34.5
1980	255.7	13.1	2,788	1,110	9.2	82.4	69.2	76.5
1990	724.0	11.0	5,801	2,853	12.5	130.7	178.0	160.8
2000	1,378.0	7.1	9,952	4,878	13.8	172.2	317.3	244.7
2005	2,021.0	6.7	12,638	6,827	16.0	195.3	439.9	287.5
2006	2,152.1	6.5	13,399	7,198	16.1	201.6	468.1	291.9
2007	2,283.5	6.1	14,062	7,561	16.2	207.3	498.9	303.2
2008	2,391.4	4.7	14,369	7,845	16.6	215.3	534.0	311.3
2009	2,486.3	4.0	14,119	8,086	17.6	214.5	567.9	320.8

Sources: The NHE and GDP data are available from the Centers for Medicare and Medicaid Services, NHE Tables, Table 1: [www.cms.gov/NationalHealthExpendData/25\\_NHE\\_Fact\\_Sheet.asp](http://www.cms.gov/NationalHealthExpendData/25_NHE_Fact_Sheet.asp) (accessed May 17, 2011). Price indexes are found in U.S. Department of Commerce, *Statistical Abstract of the United States*, 2011 (and earlier years).

<sup>1</sup> The 7.6 multiple is determined by dividing \$8,086 (2009 spending) by 2.145 (2009 CPI divided by 100) and dividing the result by the corresponding ratio for 1960. National health spending updates are available at the Centers for Medicare and Medicaid Services Web site, [www.cms.hhs.gov](http://www.cms.hhs.gov).

## **Inflation**

Although we have deflated the spending values using the CPI, medical care prices have grown faster historically than prices overall. Table 1-4 also shows the pattern of health care inflation since 1960. Note that hospital and physician care prices have risen much faster than the CPI—a phenomenon that is typical of other health care services and commodities as well. In 2009, the medical care price index (not shown in Table 1-4) rose by 3.2 percent even as the U.S. economy experienced deflation.

Medical price inflation is a common problem for maintaining health programs, and it has spurred numerous cost-containment efforts by the government. Understanding and evaluating the effects of such measures are important tasks for the health economist.

## **Access**

For many, the rising costs significantly reduce accessibility to health care. Financial affordability influences demand for most goods and services, and there are many reasons why some people do not have health insurance. What is clear is that the number of uninsured has risen and that in 2009 and 2010 approximately 50 million people in the United States lacked insurance. Many more are considered to have inadequate insurance. The problems of cost, inflation, and uninsured people have led to numerous proposals for some form of national health insurance. Later in this book, we will examine several broad groups of plans, the national health insurance programs that exist in other countries, and the newly established Patient Protection and Affordable Care Act, passed in 2010.

## **Quality**

Increases in the quality of care contribute to spending increases. Often, the focus is on ensuring quality through professional licensure and certification and, especially for hospitals, through quality-assurance programs. At the same time, concerns arise about access to high-quality care, and they are not limited to those without insurance or with minimal insurance. Other observers, however, express concerns that the quality of care in the United States is often excessive, especially for some “high-tech” treatments. For such treatments, the costs of resources may exceed the benefits to patients. The interplay among insurance, technology, and consumption is of major interest to economists.

## **The Economic Side to Other Health Issues**

Production, costs, and insurance are naturally issues that involve economics, but many other health issues have economic components, even though they may seem to be purely medical concerns. A few examples illustrate this point.

The choice of a health care treatment seems purely medical to many people, but physicians and other providers increasingly believe in evaluating and comparing alternative treatments on economic grounds. It is necessary to examine the costs of alternative techniques. Physicians are also increasingly sensitive to the economic side of the patient-physician relationship. The patient’s preferences are considered valid in determining the appropriateness of a given treatment.

We also must explore the economic reasons behind people’s health choices. People take care of themselves well at some times and poorly at other times. People’s desired health status can be understood as a meaningful economic choice. Even addiction to a relatively benign substance such as caffeine or a harmful substance such as methamphetamine can be understood better when analyzed as a possibly rational economic choice. Other health issues clearly have an economic aspect: What role should the government play in health? What health care investments should a developing country make? Should cigarette advertising be banned? Questions like these are not solely economic; but they have an economic side.



## ECONOMIC METHODS AND EXAMPLES OF ANALYSIS

We have already provided a formal definition of health economics as “the study of the allocation of resources to and within the health economy.” From another perspective, however, health economics is what economists actually do and how they apply economics to health. Economists in practice use certain characteristic approaches to their analyses of the world.

### Features of Economic Analysis

Many distinctive features of economics might be exhaustively identified, but we emphasize four:

1. Scarcity of societal resources
2. Assumption of rational decision making
3. Concept of marginal analysis
4. Use of economic models

**SCARCITY OF RESOURCES** Economic analysis is based on the premise that individuals must give up some of one resource in order to get some of another. At the national level, this means that increasing shares of GDP going to health care ultimately imply decreasing shares available for other uses. The “opportunity cost” of (what we give up to get) health care may be substantial.

While most people will recognize the money costs of goods and services, economists view time as the ultimate scarce resource. Individuals sell their time for wages, and many individuals will refuse overtime work even if offered more than their normal wage rate—because “it’s not worth it.” Similarly, many will pass up “free” health care because the travel and waiting time costs are too high.

**RATIONAL DECISION MAKING** Economists typically approach problems of human economic behavior by assuming that the decision maker is a rational being. Rationality is effectively defined as “making choices that best further one’s own ends given one’s resource constraints.” Some behaviors may appear irrational. However, when disputes over rationality arise, economists often attempt to point out, perhaps with some delight, that so-called irrational behavior often makes sense when the incentives facing the decision maker are properly understood.

**MARGINAL ANALYSIS** Mainstream economic analyses feature reasoning at the margin. To make an appropriate choice, decision makers must understand the cost as well as the benefit of the next, or marginal, unit. Marginal analysis often entails the mental experiment of trading off the incremental costs against the incremental benefits at the margin.

A prime example involves the purchase of brand-name drugs. Patients’ decisions to buy brand-name drugs, particularly for elective treatments, may depend critically on whether they must pay \$2 or \$3 per pill, or, instead, a fraction of those amounts if prescription drug insurance is available.

**USE OF MODELS** Finally, economics characteristically develops models to depict its subject matter. The models may be described in words, graphs, or mathematics. This text features words and graphs. These models may be understood as metaphors for reality. We say, “This is the market for physician care,” meaning “This is like the market for physician care.” Any model can be pushed too far and must be tested against a sense of reality and ultimately against the facts. Nonetheless, they can be apt, and we can learn from them.

In economic analysis, the models are often abstract. Abstract models help to make sense of the world, in economics as in everyday life. A young child asking what the solar system is like will undoubtedly be shown the familiar drawing of the Sun and planets in their orbits—an abstract model. The drawing is quickly grasped, yet no one supposes that the sky really looks like this.

## Examples of Health Economics Analysis

What do economists do? Economists see, on the one hand, that physicians, clinics, and hospitals use labor and machines just like other segments of the economy. These health care providers charge prices for their services, and they hope to earn at least enough to keep themselves from wanting to switch to other types of work. In this sense, the economic theories that tell us what firms do, and how they combine resources to maximize profits, provide insights into how the health care sector works.

On the other hand, the health sector is complicated. Consumers must often depend on health care providers to tell them whether they are ill, and, if so, what they should do. Providers often prescribe treatments seemingly irrespective of their costs. Physicians may not only be entrepreneurs themselves, but they may also manage their own laboratories, control hospitals, or influence insurance companies that also affect the health care sector.

Some of the earliest work in health economics addressed several of these issues. Milton Friedman and Simon Kuznets, both later Nobel laureates, studied the so-called physician shortage of the 1930s. They discovered that although physicians earned 32 percent per year more than dentists, their training costs were 17 percent higher. The remaining difference was still large, however, and Friedman and Kuznets (1945) attributed part of the higher returns on investment enjoyed by physicians to barriers to entry into the medical profession. Barriers to entry tend to lower the supply of physicians, hence raising physicians' incomes and raising the rate of return on their investment in their own education. The situation also meant that the value to society of producing one more physician exceeded the costs of producing one more physician.

In 1958 Reuben Kessel addressed the practice, then common, for physicians to charge different fees to different patients. The practice was often interpreted by the medical profession as an act of charity, for it was thought that physicians were charging lower fees to poorer patients. Kessel (1958) showed, however, that the practice of charging different fees might also be understood as a form of price discrimination that allowed physicians to increase their profits by charging more to those patients with greater ability to pay.

Although such pricing fits neatly into a model of a price-discriminating monopolist, the medical profession consists of a large number of practitioners. Kessel understood the unique institutional characteristics that nevertheless made the monopoly model applicable. For example, using historical analysis to support his case, Kessel argued that physicians used the system of hospital privileges to perpetuate price discrimination. Physicians, in order to practice, are usually affiliated with hospitals. Historically, those who adopted competitive economic practices were often branded as “unethical” and could be denied membership in county medical societies, access to specialty certification, or hospital privileges.

These early contributions also illustrate two features of health economics. The first is the interdisciplinary nature of health research. Economists must know how health care is delivered. This information comes from health care providers, as well as from members of other disciplines, such as public health, sociology, and psychology, who study how organizations operate.

The second feature concerns the institutions in the health care system, including hospitals, insurers, or regulators. Examples of these include the following:

- relations of doctors to hospitals
- the organization and practices of the health insurance industry
- licensing and certification of health care providers

Such institutional concerns do not necessarily distinguish health economics from other forms of economic analysis, but the particular history and present form of health care institutions do set health economics apart from other fields. We must take care to describe the institutional characteristics of the health sector in order to devise appropriate models.

## DOES ECONOMICS APPLY TO HEALTH AND HEALTH CARE?

Many observers complain that economics is irrelevant to the study of health. This issue is raised often enough in serious contexts to require consideration. The complaint suggests a model of health care in which health is primarily a technical issue of medical science, better left to experts. One gets sick and one sees a doctor, who provides the medically appropriate treatment.

If economics studies how scarce resources are used to produce goods and services and then how these goods and services are distributed, then clearly economics applies. Certainly health care resources are scarce; in fact, their cost concerns most people. There is no question that health care is produced and distributed.

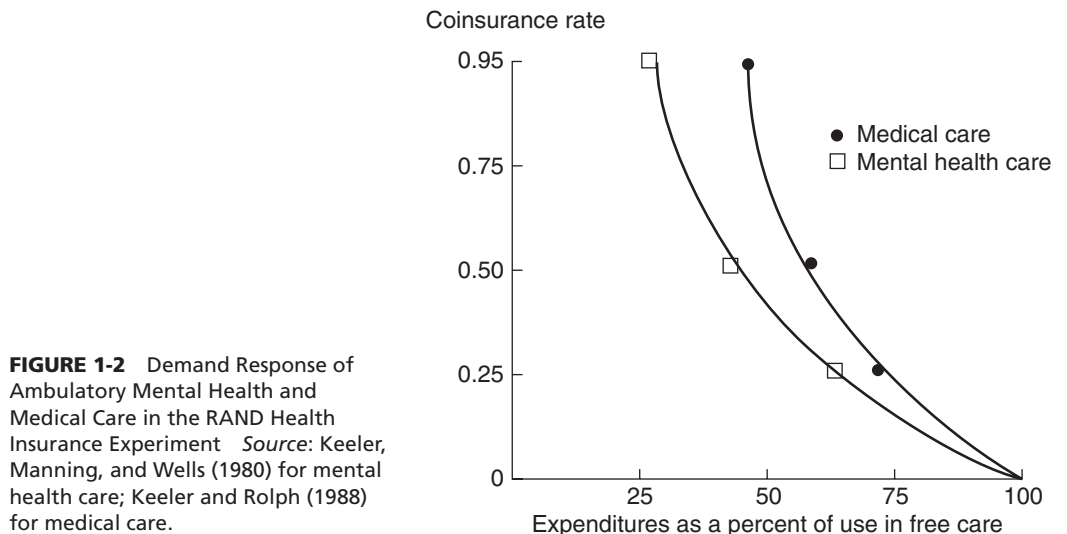
Nevertheless, one can question whether the characteristic approaches of economics apply to health care. Are health care consumers rational? Do they calculate optimally at the margin? Imagine a loved one suffering cardiac arrest. Is there time or reasoning power left to calculate? Would anyone question the price of emergency services under such circumstances?

However, much of health care simply does not fit this emergency image. A considerable amount of health care is elective, meaning that patients have and will perceive some choice over whether and when to have the diagnostics or treatment involved. Much health care is even routine, involving problems such as upper respiratory infections, back pain, and diagnostic checkups. The patient often has prior experience with these concerns. Furthermore, even in a real emergency, consumers have agents to make or help make decisions on their behalf. Traditionally physicians have served as agents and more recently, care managers have also entered the process. Thus, rational choices can be made.

### An Example: Does Price Matter?

Does price matter? Many have argued that health care is so different from other goods that consumers do not respond to financial incentives. These views have been justified by arguments that demand is based on need, or arguments that patients leave decisions entirely to their providers, who are concerned with their own interests rather than how much patients have to pay.

Data from the RAND Health Insurance Experiment, a pioneering project of the 1970s that examined consumer choices and health outcomes resulting from alternative insurance arrangements, give an unequivocal answer to this question: Yes, economic incentives matter. Figure 1-2 examines the use of ambulatory mental health and medical care where amounts of health care consumed are measured along the horizontal axis. These amounts are scaled in percentage terms



**FIGURE 1-2** Demand Response of Ambulatory Mental Health and Medical Care in the RAND Health Insurance Experiment *Source:* Keeler, Manning, and Wells (1980) for mental health care; Keeler and Rolph (1988) for medical care.

from zero to 100 percent, where 100 percent reflects the average level of care consumed by the group that used the most care on average. This group, not surprisingly, is the group with “free” care. The vertical axis measures the economic incentives as indicated by the coinsurance rate—the percentage of the bill paid out directly by the consumer. Thus, a higher coinsurance rate reflects a higher price to the consumer.

The curve shown in Figure 1-2 is similar to an economist’s demand curve in that it shows people consuming more care as the care becomes less costly in terms of dollars paid out-of-pocket. More importantly, the curve demonstrates that economic incentives do matter. Those facing higher prices demand less care.

## IS HEALTH CARE DIFFERENT?

Although economics certainly applies to health care, it is more challenging to answer the question of how directly and simply it applies. Is economic theory so easily applicable that a special field of health economics is not even necessary? Is health care so special as to be unique? Or is the truth somewhere in between?

We argue that health care has many distinctive features, but that it is not unique in any of them. What is unique, perhaps, is the combination of features and even the sheer number of them. We review these distinctive features to alert students as to those salient features of health care that require special attention. In each case where health is distinctive in economic terms, a body of economic theory and empirical work illuminates the issue.

### Presence and Extent of Uncertainty

When Nobel laureate Kenneth Arrow (1963) directed his attention to the economics of health, he helped establish health economics as a field. He stressed the prevalence of uncertainty in health care, on both the demand side and the supply side. Consumers are uncertain of their health status and need for health care in any coming period. This means that the demand for health care is irregular in nature from the individual’s perspective; likewise, the demand facing a health care firm is irregular.

Uncertainty is also prevalent on the supply side. Standard economic analysis often assumes that products, and the pleasures that they bring, are well understood by the purchasers. The purchase of steak, milk, new clothes, or a ticket to a basketball game provides expected well-being that is easily known and understood. In contrast, several cases of product uncertainty exist in the health field. Consumers often do not know the expected outcomes of various treatments without physicians’ advice, and in many cases physicians themselves cannot predict the outcomes of treatments with certainty.

Arrow argued that uncertainty on both the supply and demand sides might lead to the result that insurance markets for various risks would often fail to arise. These factors suggested a role for government. Although others dispute the conclusion, the point stands that uncertainty is a prominent feature in the field of health, and the tools of economic analysis of uncertainty and risk will need to be used to analyze health issues fully. In a special 2001 edition of the *Journal of Health Politics, Policy, and Law* (Peterson, 2001), 27 health policy experts revisited Arrow’s 1963 article to explore its impact on the evolution of health economics and health policy research.

### Prominence of Insurance

Consumers purchase insurance to guard against this uncertainty and risk. Because we have health insurance, neither most Americans nor citizens of other countries pay directly for the full costs of their health care. Rather, the costs are paid indirectly through coinsurance and through insurance premiums that are often, although not always, purchased through participation in the labor force.

**TABLE 1-5 Personal Health Care Spending, Selected Years (in \$ Billions)**

Year	Total	Total Health Insurance	Private Health Insurance	Medicare	Medicaid	Other Programs/ 3rd Party	Out-of-Pocket	% Out-of-Pocket
1960	23.4		5.0			5.5	12.9	55.1
1970	63.1	29.6	14.0	7.3	5.0	11.8	25.0	39.6
1980	217.1	131.9	61.4	36.3	24.7	36.3	58.4	26.9
1990	616.6	403.0	204.8	107.3	69.7	96.0	138.8	22.5
2000	1164.4	843.5	405.8	215.9	186.9	153.7	202.1	17.4
2005	1692.6	1278.3	603.8	326.4	287.7	210.9	263.8	15.6
2006	1798.8	1367.6	636.4	381.7	283.7	225.0	272.1	15.1
2007	1904.3	1444.7	663.8	407.4	302.5	241.3	289.4	15.2
2008	1997.2	1528.1	692.7	440.8	316.5	249.1	298.2	14.9
2009	2089.9	1615.0	712.2	471.3	345.7	261.5	299.3	14.3

Source: Centers for Medicare and Medicaid Services: [www.cms.hhs.gov/NationalHealthExpendData/downloads/tables.pdf](http://www.cms.hhs.gov/NationalHealthExpendData/downloads/tables.pdf), Table 6. Accessed May 17, 2011. Values for 1960 from *Health United States*, 2005, Table 123.

Table 1-5 provides data on the sources of payment for personal health care services for selected years since 1960. In addition to out-of-pocket costs, these payment sources include private insurance; Medicare and Medicaid (the major government programs for the elderly and certain lower income households); and other public and private programs. In 1960, 55 percent of all personal health care expenditures were paid out-of-pocket, meaning that 45 percent was paid by third-party payers (either private or government). Out-of-pocket costs dropped dramatically following the introduction of Medicare and Medicaid in 1966, the continued growth of private insurance, and the introduction of new programs such as the Childrens' Health Insurance Program (CHIP) established in 1997.

By 2009, 86 percent of personal health care spending was paid by third parties. We will carefully study this phenomenon and its effects for both private and public insurance. It should be clear, even prior to our focused analyses, that the separation of spending from the direct payment for care must weaken some of the price effects that might be expected in standard economic analysis. Insurance changes the demand for care, and it potentially also changes the incentives facing providers.

Changed incentives that face providers concern us more as the insurance portion of the bill increases. How the insurers pay the health care firm thus becomes a critical fact of economic life. Whether insurers cover a procedure, or a professional's services, may determine whether providers use the procedure.

Furthermore, changes in insurance payment procedures can substantially change provider behavior and provider concerns. In the 1980s Medicare, faced with rapidly increasing expenditures, changed its hospital payment system from one based largely on costs (i.e., retrospective reimbursement) to one with fixed payments per admission determined by the resources typically used to treat the medical condition (as classified by Diagnosis Related Groups, or DRGs). With a prospective DRG payment system, an extra day of care suddenly added to the hospital's costs, rather than to its revenues. This reimbursement system, still used today, led to shorter stays, reduced demand for hospital beds, and ultimately the reduction in size and/or closing down of many hospitals.

### Problems of Information

Uncertainty can in part be attributed to lack of information. Actual and potential information problems in health care markets raise many economic questions. Sometimes information is unavailable to all parties concerned. For example, neither gynecologists nor their patients may

recognize the early stages of cervical cancer without Pap smears. At other times, the information in question is known to some parties but not to all, and then it is the asymmetry of information that is problematic.

The problems of information mean that careful economic analysts must modify their methods. Standard analyses often assume that consumers have the necessary knowledge about the quality of the food or the clothing that they purchase. People purchase beef as opposed to fish, or cotton as opposed to nylon fabrics, basing their decisions on the characteristics of the goods, their prices, and the goods' abilities to bring pleasure.

Health goods and services depart substantively from this model. Consumers may not know which physicians or hospitals are good, capable, or even competent. Consumers may not know whether they themselves are ill or what should be done if they are. This lack of information often makes an individual consumer, sometimes referred to as the *principal*, dependent on the provider, as an *agent*, in a particular way. The provider offers both the information and the service, leading to the possibility of conflicting interests. Newhouse (2002), for example, speaks of a health care "quality chasm" that may be traced to both inadequate consumer information and to inadequate financial incentives. Health economics must address the provision of health services in this context.

### **Large Role of Nonprofit Firms**

Economists often assume that firms maximize profits. Economic theory provides models that explain how businesses allocate resources in order to maximize profits. Yet many health care providers, including many hospitals, insurers, and nursing homes, have nonprofit status.

What, then, motivates these nonprofit institutions if they cannot enjoy the profits of their endeavors? The economist must analyze the establishment and perpetuation of nonprofit institutions, and understand the differences in their behaviors from for-profit firms. This problem has recently emerged in the context of academic medical centers in the United States. Many current college students, and most certainly their parents and grandparents, know of the prominent roles of great hospitals affiliated with great universities such as Harvard or Johns Hopkins. The public and the larger medical community are aware of the major hospitals as centers of health care, teaching, and research. Yet with the changing health economics of the twenty-first century, the organization of these hospitals and the funding of their activities are continuously evolving.

### **Restrictions on Competition**

Economists and policymakers generally laud the competitive market because the entry of firms or providers in the face of high prices and/or profits will cause the other firms or providers to lower their prices. This entry and the resulting price reduction improve the well-being of consumers.

Nevertheless, the health sector has developed many practices that effectively restrict competition. These practices include licensure requirements for providers, restrictions on provider advertising, and standards of ethical behavior that enjoin providers from competing with each other. We must explain the forces that generated such practices and understand their potential benefits, but we must also understand their anticompetitive impacts and measure the magnitudes of the higher costs they may impose on society.

Regulation to promote quality or to curb costs also reduces the freedom of choice of providers and may influence competition. There is often substantial interest in regulating the health care sector. The causes, as well as the impacts, of the regulations require considerable attention. The pharmaceutical industry, for example, contends that patent protection is crucial for its financial stability. Economists must consider how regulations are developed, as well as who gains and who loses from them.

Other types of government interventions, particularly antitrust action, can serve to promote competition. Economic theory suggests that monopolistic firms may reduce production, with a corresponding increase in price. Control of the monopolies or antitrust action to curb monopolistic practices may thus benefit the consumer.

## Role of Equity and Need

Poor health of another human being often evokes a feeling of concern that distinguishes health care from many other goods and services. Many advocates express this feeling by saying that people ought to get the health care they need regardless of whether they can afford it. In practice, “need” is difficult to define, and distributing care under certain definitions of need may cause more economic harm than good. Yet the word signals a set of legitimate concerns for analysis.

## Government Subsidies and Public Provision

In most countries, the government plays a major role in the provision or financing of health services. In the United States in 2009, Medicare and Medicaid alone accounted for 33 percent of national health care spending. However, there are many other government programs, both federal and state and local, including those for public health, military veterans, eligible children, and for mental health and substance abuse.

Federal government subsidies are prominent in the 2010 reform legislation. To expand insurance coverage by making it more affordable, those with incomes above the Medicaid thresholds but less than 400 percent of the poverty level will be eligible for subsidies on a sliding scale.

## CONCLUSIONS

From the mid-1960s, the U.S. health care system grew rapidly in its share of GDP as well as in absolute size. As we saw in Table 1-4, health expenditures rose from 5.2 percent of the GDP in 1960 to 9.2 percent in 1980, 12.5 percent in 1990, and 17.6 percent in 2009.<sup>2</sup> Buoyed by increased income, more generous insurance, and the availability of state-of-the-art technology and high-cost treatments, consumers and their insurers increased their health care spending dramatically.

Over the same period, extraordinary changes have occurred in the ways in which health care delivery is organized and financed. The role of government has greatly increased especially as insurer but in other ways as well; many of today’s reimbursement mechanisms would not have been recognized 40 years ago; private hospitals have flourished often as part of larger hospital networks; and information technology for both patients and providers has blossomed. Above all, a delivery system in which patients and their physicians once held a great deal of autonomy in decision making has been replaced by one dominated by “managed care” arrangements.

In light of the distinctive features of the health economy, health economists seek to analyze change meaningfully and to develop and evaluate alternatives that better meet objectives. To those who question its relevance to these tasks and argue that health care is unlike any other good, we will emphasize throughout the text that accounting for the features of the health sector falls well within the scope of economic inquiry. Uncertain events and information deficiencies affect much of life, including automobile safety, household security, and the fortunes of most businesses. Insurance programs cover many of these instances, and every financial decision entails an element of risk.

We commonly speak of needs for food and housing, and we have government programs to provide or subsidize the poor in these markets. Other professions are licensed in many states, and various industries are regulated. Nonprofit firms provide symphony music and college educations, as well as health care. Special tools of analysis are needed in all these cases, not just for health care.

Health and health care are undoubtedly unique in having this combination of distinctive features and perhaps in possessing the sheer number of them. It is unsurprising, therefore, that a distinctive field of economic inquiry with a distinctive set of questions has arisen.

---

<sup>2</sup> Projections by Keehan et al. (2011), prepared after passage of the 2010 Patient Protection and Affordable Care Act, indicate that the share of GDP will grow to 19.8 percent by 2020.

Health economics has enjoyed rapid growth and widespread acceptance since the mid-1960s, but its future undoubtedly rests on its ability to inform and influence those who make policy. Distinguished economist Victor Fuchs, a past president of the American Economic Association, is optimistic that health economics will meet these challenges and continue to flourish.

The greatest strengths of economics and economists are a framework of systematic theory, an array of concepts and questions that are particularly relevant to the choices facing policy makers, and skill in drawing inferences from imperfect data. Because health economists often take standard economic theory for granted (like being able to walk or talk), it is easy to underestimate the advantage this framework offers economics over the other social and behavioral sciences. When economists encounter a new problem, one with which they have had no previous experience, they immediately have a way to begin thinking about it long before data collection begins. Scholars in the other “policy sciences” do not. They typically require some detailed knowledge of the particular problem before they can begin to think productively about it. Economists’ framework of systematic theory facilitates the transfer of knowledge drawn from other fields of study to the health field.

Health economists have also inherited from economics a set of concepts and questions that have proven to be particularly relevant to the policy problems that have emerged in health during the past three decades. Scarcity, substitution, incentives, marginal analysis, and the like were “just what the doctor ordered,” although in many cases the “patient” found the medicine bitter and failed to follow the prescribed advice. (Fuchs, 2000, p. 148)

Professor Fuchs’s insights have become even more relevant following passage of the 2010 Patient Protection and Affordable Care Act. These reforms will bring unprecedented change, including a scheduled individual mandate for insurance coverage. We share Professor Fuchs’s optimism that the theoretical framework and tools used by economists will greatly improve our understanding of these changes and their potential effects.

---

## Summary

1. Health care spending has grown rapidly in absolute and relative terms. In 2009, it accounted for over one-sixth of U.S. GDP, and its share of GDP is projected to grow.
2. The growth in health care spending is attributable to more services, higher-quality services, and relative increases in the prices of health care services. Health economists seek to determine the underlying causes of these phenomena.
3. The size of the health economy is also reflected through other measures such as the number of jobs in health care professions and amount of capital.
4. Time spent obtaining and providing health care represents a key “unpriced” factor in the health economy.
5. The health economy is considerably larger in the United States than in other countries.
6. There are significant policy concerns not only with the growth of spending but also with access and quality.
7. Economists use models to explain economic behavior. The models are abstract simplifications of reality.
8. Health economists still disagree on some fundamental issues, such as the extent to which the competitive model applies to the health economy.
9. Health care services and the health economy possess a unique set of distinguishing features, such as the prevalence of uncertainty or insurance coverage. Health care is unique because of this entire set of features.
10. The health care system has changed dramatically over the past 50 years. The role of government,



reimbursement methods, and the dominance of managed care represent some of the major changes. The Patient Protection and Affordable Care Act of 2010 is the most important recent change.

11. An important consequence of many of these changes is the substantial drop in out-of-pocket costs for consumers, meaning that private insurance and public programs have correspondingly grown.
12. Technological change through improved procedures, and new drugs, provides potential improvements in health care, but also possibilities of increasing costs, in all countries.
13. Economics provides valuable theoretical tools and a systematic framework for understanding the health care system and evaluating alternative policy proposals.

## Discussion Questions

1. Suggest several reasons why health care spending is higher in the United States than in other countries. Is the fact that the U.S. population spends more per capita on health care than people in any other developed country evidence of a failure of the U.S. system? What issues do you think are involved in answering the question?
2. Describe several key issues facing policymakers with regard to health care spending.
3. If greater health care spending leads to more jobs, why is there such concern about the rapid growth rates of spending?
4. Do consumers take the net price (including insurance and time) they face into consideration when choosing health care? What evidence suggests that price matters? Suggest real life scenarios in which price may affect choices regarding health care.
5. Suppose that a woman works 40 hours per week with no opportunity for overtime. She also takes care of a sick parent. Can we say that her time has no value in providing this health care because she could not earn more at work?
6. What is meant by marginal analysis? Provide an example in which marginal analysis is useful in looking at policy questions.
7. Give three examples of quality of care in the provision of health services. Why might consumers be willing to pay more money to have each of them?
8. Describe the size of the health economy when measured by the quantities of capital and labor used to produce health care. What important inputs to the production of health are not being counted among these?

## Exercises

1. Health care spending ( $S$ ) can be summarized by the following equation:
 
$$S = \left( \begin{array}{c} \text{population} \\ \text{size} \end{array} \right) \times \left( \begin{array}{c} \text{health care} \\ \text{quantity} \\ \text{per person} \end{array} \right) \times \left( \begin{array}{c} \text{price per} \\ \text{unit of} \\ \text{health care} \end{array} \right)$$
  - (a) Identify three factors that might lead to the rapid growth of health care spending.
  - (b) Compare health care spending to housing expenditures and to food expenditures. How are the sectors similar? How do they differ?
2. Identify five distinctive features of the health economy. Examine each one separately, and describe other commodities or sectors that share those features. Do any other commodities or sectors have all the features you listed?
3. In Table 1-1, calculate which countries had the largest and smallest percentage increases in GDP share from 1960 to 1980. Compare these to similar calculations for the period 1980 to 2009. Discuss your results.
4. The United States, Canada, and the United Kingdom share the same language but have considerably different health care systems. Compare the three health shares of GDP from 1960 to 2009. What factors may explain the considerable differences among the three countries?
5. Table 1-4 provides indexes of the prices of health care inputs. Calculate the growth rates between 1980 and 2009 of the prices of hospital and physician services. Compare them to the growth rate of the overall consumer price index (CPI). Discuss your findings.
6. In Table 1-5, examine the private health insurance, Medicare, and Medicaid components. Which category grew the most between 1970 and 2009? Between 2000 and 2009? What factors might have led to the differences in the growth rates?

7. Several Web sites provide useful information on health care and health resources use. Use a Web browser to find sites of:
- Centers for Medicare & Medicaid Services (CMS)
  - National Institutes of Health (NIH)
  - Organization for Economic Cooperation and Development (OECD)
  - The Kaiser Family Foundation (KFF)
  - For students outside the United States, find governmental sites from your own country.
- Compare and contrast the data available from these sites.
8. The following chart shows health expenditures for the United States between 1965 and 2009. Using a spreadsheet program:
- (a) Calculate health expenditures per person for each year.
  - (b) Calculate percentage increases in health expenditures per person for each year.
  - (c) Can you find particular events in given years that might explain either small or large changes in the health expenditures per person or in the percentage changes?

Year	Population (in Millions)	Health Expenditures (\$ in Billions)
1965	200	42.2
1966	202	46.4
1967	204	52.0
1968	206	59.0
1969	208	66.3
1970	210	74.8
1971	213	83.2
1972	215	92.9
1973	217	103.0
1974	218	116.7
1975	220	133.0
1976	222	152.3
1977	224	172.8
1978	226	194.2
1979	228	219.9
1980	230	255.7
1981	233	293.6
1982	235	330.8
1983	237	364.7
1984	240	401.5
1985	242	439.3
1986	244	471.3
1987	247	513.0
1988	249	574.0
1989	251	638.8
1990	254	724.0
1991	258	781.6
1992	261	849.1
1993	265	912.5
1994	268	962.1
1995	271	1,016.6
1996	274	1,068.5
1997	274	1,125.1
1998	277	1,190.0
1999	280	1,265.2
2000	283	1,378.0
2001	285	1,469.2
2002	288	1,602.4
2003	291	1,735.2
2004	294	1,855.4
2005	296	2,021.0
2006	298	2,152.1
2007	301	2,283.5
2008	305	2,391.4
2009	307	2,486.3

# Microeconomic Tools for Health Economics

- Scarcity and the Production Possibilities Frontier
- Practice with Supply and Demand
- Functions and Curves
- Consumer Theory: Ideas Behind the Demand Curve
- Individual and Market Demands
- Elasticities
- Production and Market Supply
- The Firm Supply Curve Under Perfect Competition
- Monopoly and Other Market Structures
- Conclusions

This chapter provides an explanation of the microeconomic tools that are used in the text by reviewing material that may be encountered at the introductory and intermediate levels of learning. These tools are useful for a deeper understanding of ideas in health economics. This review does not substitute for an introductory course in the principles of economics, and a complete series of principles is necessary for a better understanding. Only by such study can one gain an understanding of the subtleties of the subject as well as the many large ideas that that we simplified for this review. Although we present the material in a way consistent with more detailed and rigorous treatments in a standard microeconomics course, we have eliminated the academic give-and-take of disputes over issues as well as much of the mathematical detail.

Note also that the “Basic Economics Tools” chapters deal mainly with microeconomic issues; we ignore the macroeconomic issues except in the sense that these subjects increasingly overlap. Microeconomic concerns involve individual decision makers such as households and providers, and specific industries such as the health insurance and hospital industries. The behavior of individual decision makers is the focus and subject of microeconomics. Health economics also addresses the problem of the efficient use of scarce resources, which too is a microeconomic issue.

This chapter starts with the concept of scarcity and reviews supply-and-demand analysis. After these introductory treatments of supply and demand, the chapter returns separately to demand first and then to supply, developing the underlying ideas behind these tools. We then combine them into models of market structures, emphasizing those featured in health economics.

## SCARCITY AND THE PRODUCTION POSSIBILITIES FRONTIER

A fundamental idea in economics is that there is no such thing as a free lunch. The fact that little if anything is free implies that to get something, one must usually give up something else, such as time or other resources. A helpful theoretical tool that illustrates this idea is the production possibilities frontier (PPF).

The PPF illustrates the trade-offs between two categories of goods. The curve shows how our choices are constrained by the fact that we cannot have all of everything we want. The idea that we face resource constraints and must make trade-offs is central to the PPF, but similar ideas also apply to the individual firm or the individual consumer.

We begin the production possibilities problem with a table illustrating a classic dilemma concerning society's trade-off between guns and butter. Table 2-1 Shows data on the amounts of guns or butter that a hypothetical society could produce with its resources. Guns and butter refer metaphorically to all goods and services with a military use versus those that have a domestic consumption use. The PPF could in principle also be drawn in many dimensions for many goods. What is essential is that the goods represented exhaustively account for all the goods in the economy.

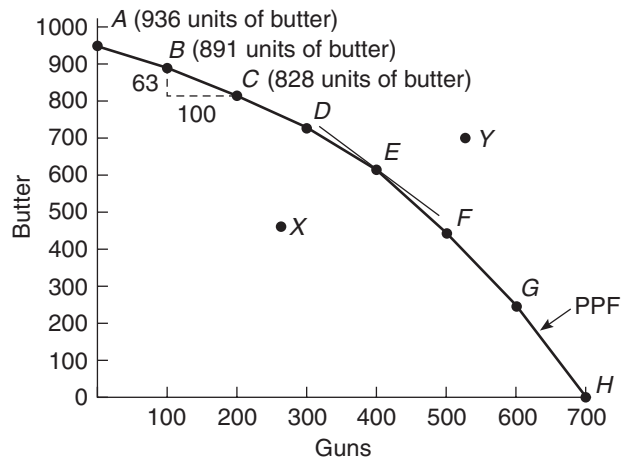
Table 2-1 contains many numbers but illustrates two central ideas. Note first that as the number of guns increases, the number for butter falls, indicating that to produce more of one good we must give up some of the other. The amount of butter given up in order to produce an extra 100 units of guns is called the *opportunity cost* of 100 units of guns. We can measure opportunity cost per 100 units of guns as here, but more commonly we look at the opportunity cost of the single next unit of guns, called the marginal unit. In either case, the opportunity cost represents what is given up.

The opportunity cost column reporting the costs of each 100 units of guns in terms of butter foregone illustrates a second idea—that of increasing opportunity costs. As the number of guns increases, the opportunity cost gets larger. If society is to increase its production of guns, say from 200 to 300 units, it must transfer the resources, labor, and capital previously used in butter production to gun production. The idea that this is a frontier means that we are representing society's best possible production. Thus, when we first shift butter resources toward gun production, we can arbitrarily choose to shift those resources relatively best suited to gun production first. By choosing laborers who are handier at gun-making than at butter-making, we will gain the most guns per unit of butter we give up. As we shift more resources toward guns, we will have to dig deeper into our relatively good butter-producing resources, and hence give up greater quantities of butter. Increasing opportunity costs also illustrate the specialization of society's resources of labor and capital.

We can transfer the data in Table 2-1 into the graph in Figure 2-1. Note that if this society devotes all its labor and capital to butter production, the most butter it can produce is 936 units. The two

**TABLE 2-1 Society's Trade-Off Between Guns and Butter**

Point	Butter	Guns	Opportunity Cost: Butter
			Given Up to Produce 100 Units of Guns
A	936	0	
B	891	100	45
C	828	200	63
D	732	300	96
E	609	400	123
F	444	500	165
G	244	600	200
H	0	700	244



**FIGURE 2-1** Society's Trade-Off Between Guns and Butter. Point X Is Inefficient; Point Y Is Infeasible

numbers—936 units of butter and zero units of guns—can be represented by point A in Figure 2-1. The other points are transferred in the same manner. We assume that the missing points between these data points will fit the same pattern, resulting in the PPF curve.

This graph illustrates the idea of no free lunch with the downward slope of the PPF curve. In this example, increased gun production means we must give up some butter production—hence, we get no free lunch. Second, the opportunity cost itself is illustrated in the slope of the curve. For example, the line between points B and C has a slope of 63 (the rise) over 100 (the run), and 63 units of butter per 100 units of guns is the opportunity cost we observed in the table.

The opportunity cost of one single unit of guns is the slope of the PPF at a single point, which equals the slope of a line tangent at that point. Therefore, at point E the opportunity cost is identical to the slope of the tangent line to the PPF at E. Finally, the idea of increasing opportunity cost is illustrated by the bowed-out shape of the PPF, showing its concavity to the origin. Recall that the slope is the opportunity cost. Thus, the slope becoming steeper means that the opportunity cost is increasing. Society could choose any point on the PPF, but society can be at only one point at a time. How society makes and achieves its choice are other matters to discuss, but at present we have merely illustrated the best possible practices of some hypothetical society. An interior point such as X means that the particular society is not doing the best it can; it is inefficient. A point such as Y is unattainable because of insufficient resources to produce the indicated amounts of both guns and butter.

## BOX 2-1

### There's Scarcity and Then There's Real Scarcity

Scarcity is clearly defined in economics; a good or service is “scarce” whenever it has a non-zero opportunity cost. So, goods and services generally are scarce, for the very rich even as much as they are for the very poor. The only question is whether one has so much as to have become satiated, at which point a good starts to become a “bad.”

However, the way ordinary people use the word *scarce* is very clear, too. If money is scarce, that clearly means tough times. Many American students, and perhaps Canadian and European students as well, may be surprised to discover just how vast the differences among countries are in terms of income, health, health expenditures, and the like. Germany and the United States are examples of the industrialized West; these populations experience three to five times the material income of the Russian Federation or Brazil, and six to eight times the income of either Albania or China. By far the most striking contrast is with the African countries of Ethiopia or Nigeria, who have little to spend on health care, and at the same time experience the largest health and health care deficit. Table 2-2 provides examples from a selection of countries across the globe.

**TABLE 2-2** There Is a Strikingly Diverse Experience in Income, Health, and Health Care Across the World's Cultures

Country	GDP per capita (in \$) 2009	Health expenditures per person (in \$) 2009	Life expectancy at birth (years) total 2009	Tuberculosis incidence per 100,000 people 2009	Infant Mortality rate (deaths per 1,000 live births) 2009
Afghanistan	486	51	44	189	134
Albania	3,808	265	77	15	14
Bangladesh	551	18	67	225	41
Brazil	8,230	734	73	45	17
China	4660	177	73	96	17
Ethiopia	344	15	56	359	67
Germany	40,670	4,629	80	5	4
India	1,192	45	64	168	50
Nigeria	1,118	69	48	295	86
Norway	79,089	7,662	81	6	3
Russian Fed.	8,684	475	69	106	11
United States	45,989	7,410	79	4	7

Note: We calculate GDP figures using the purchasing power parity method.

Source: <http://www.worldbank.org/data> accessed May 2011.

## PRACTICE WITH SUPPLY AND DEMAND

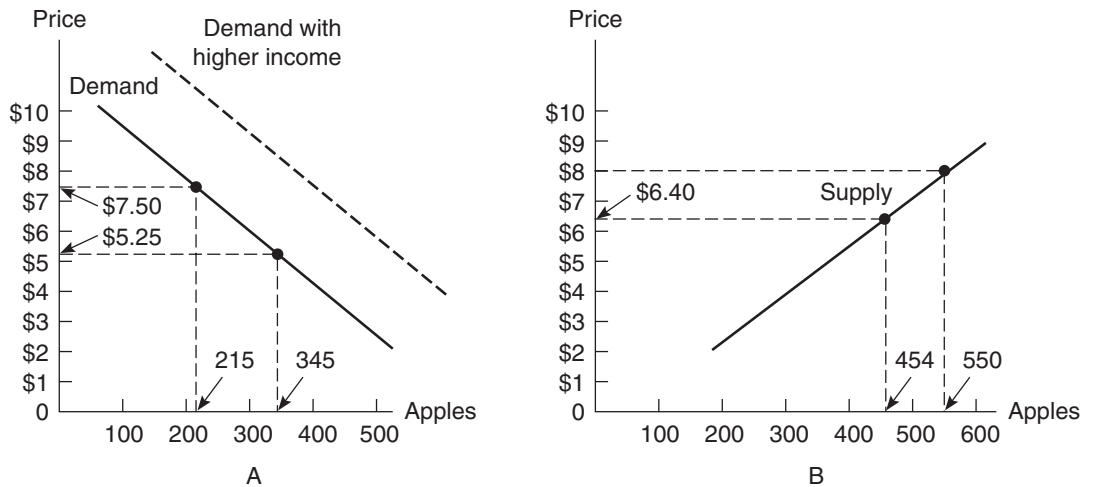
The most familiar ideas in economics are probably supply-and-demand curves. We will illustrate supply-and-demand analysis for a hypothetical market for apples.

### The Demand Curve and Demand Shifters

In Figure 2-2A, a demand curve illustrates the demand for apples in a hypothetical market for a given period, say one week. The demand curve drawn shows a quantity demanded in this market for this period for each possible price. For example, at a price of \$5.25 per bushel, consumers would wish to buy a total of 345 bushels that week. The theory of demand suggests that quantity demanded would be less at higher prices—for example, 215 bushels at a price of \$7.50. Some consumers may find that the price rise represents “the last straw” so that they buy none, while others may buy fewer apples than before, and yet others would not change their purchases. It is doubtful that anyone would use the occasion of a price rise as a reason to buy more.

This analysis is done *ceteris paribus*, meaning that we are assuming that all other things are held constant. The price of apples rises while people's tastes, perceptions, incomes, and so on stay the same. In life, it is common for two or more things to change at the same time. If, for example, the price of apples rises at the same time as tastes change, the result would be theoretically ambiguous, meaning that we cannot predict the direction of the change. In contrast, the demand curve depicts the behavior of consumers as price only changes.

As long as people buy less at higher prices, then the demand curve will be downward sloping. In statistical analysis, estimated demand curves are almost always downward sloping. The responsiveness of demand to price is measured by the elasticity. We will discuss elasticity in a later section. Other variables also will affect the demand for apples. For example, the amounts of various foods



**FIGURE 2-2** Changes in Demand and Supply

that people buy may depend on their incomes; richer people tend to buy more. Demand may be affected by the price of other substitutable goods. When any other variable affects demand, its effect will be shown as a shift in the curve. For convenience, we call such variables *demand shifters*. A list of demand shifters includes the following.

**Income** Generally, the quantity of a good that people wish to buy depends in part on their income. People with higher incomes tend to demand more of most goods. Such goods are called *normal* goods. But some goods, such as used clothing or generic brand goods, are purchased less often when people become richer. Such goods are called *inferior* goods. Let us assume that apples are normal goods. In Figure 2-2A, increased income in the community would tend to shift the demand curve outward.

**Other Prices** Prices of related goods also will affect demand. Related goods may be either substitutes or complements. If oranges are regarded as substitutes for apples, an increase in the price of oranges would cause the demand for apples to increase, shifting the demand curve to the right. In contrast, a complement is something that is used with apples, such as caramel. If the price of a complement rises, the demand for apples decreases or shifts left.

**Insurance** A variable that makes no sense for apples but that is essential on a list of demand shifters in health economics is insurance. Insurance reduces the price to the consumer at the point of service; given the lower price, a greater quantity of health care will be demanded. Although one can treat this as a movement down a given demand curve, we show in a later chapter that this is equivalent to a clockwise rotation in the original demand curve. Insurance plans have many complexities beyond changing the consumer's effective price, and these are also addressed in a later chapter.

**Tastes** Many other demand shifters may be grouped under the heading of tastes. Tastes can be literally what the word means, as when a new recipe increases interest in apples. The term can be less literal as well, as when we say that an older population has a greater demand for health care because it has a greater taste for health care.

### The Supply Curve and Supply Shifters

Supply can be approached in a similar way. Figure 2-2B shows an upward-sloping supply curve for apples. It illustrates, for example, that apple growers would be willing to offer 454 bushels of apples for sale if the price were \$6.40. At a higher price, say \$8.00, more would be offered. Apple growers

might be more willing to divert apples from cider production, to make greater efforts in harvest, or even to bring formerly unprofitable trees into production if the price were higher. Such reasons would suggest an upward-sloping supply curve such as the one shown.

We may likewise generate a list of supply shifters.

**Technological Change** As technology improves for producing a given product, the good becomes cheaper to produce. Certainly, technological changes that make products more costly without improving quality are ignored. As the product becomes cheaper to produce, suppliers are willing to offer more for sale at a given price. This increases supply, thus shifting the supply curve to the right.

**Input Prices** If the wages of apple pickers were to rise, this increase in an input cost would reduce suppliers' willingness to offer as much for sale at the original price. The supply would decrease, shifting the curve to the left.

**Prices of Production-Related Goods** The price of a good related in production, such as cider, also would be relevant. Because farmers can use raw apples for eating or for cider, generally a rise in cider prices will cause the supply of apples for eating to decrease, thus shifting the supply curve to the left.

**Size of the Industry** As more firms (in this case apple growers) enter a market, the supply of the product will be greater. Thus, entry of firms will cause supply to shift to the right.

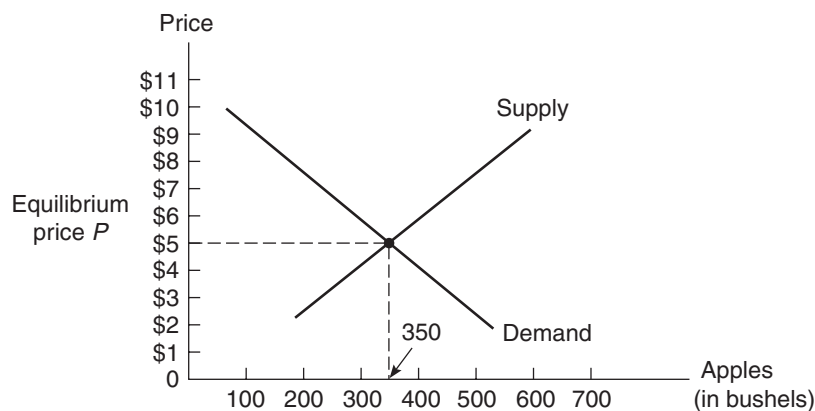
**Weather** For a number of products, acts of God such as the weather will tend to affect production. The direction of the effect is obvious: Good weather increases supply.

## Equilibrium

Under conditions of competition, the equilibrium in a market will occur at the point where the demand and supply curves intersect. This is the point at which demanders' and suppliers' plans agree. In Figure 2-3, the equilibrium occurs at the price of \$5.00 and at the quantity of 350 bushels. At higher prices, there will be excess supply, and suppliers who were unable to sell all their goods will be willing to bid prices down. At lower prices, there will be excess demand, and demanders who went undersupplied will be willing to pay more and will tend to bid prices up.

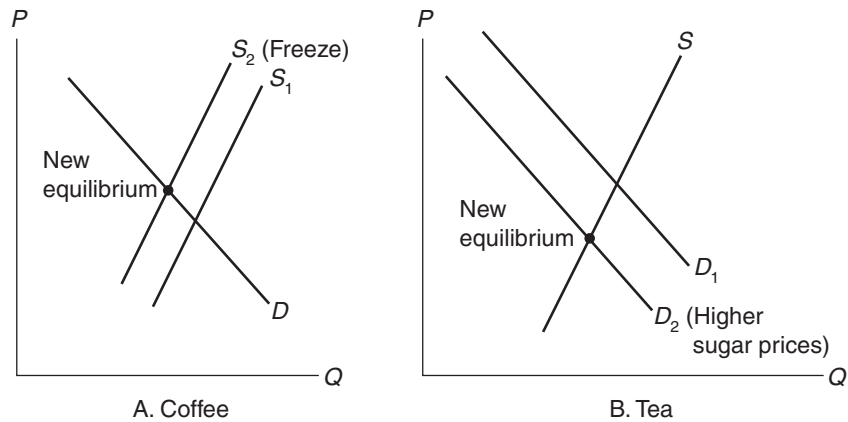
## Comparative Statics

An equilibrium as depicted in Figure 2-3 is a static equilibrium. It shows a picture of an unchanging equilibrium at a point in time. It is more interesting to assess how the equilibrium will change in response to some economic event. Figures 2-4A and B give two examples. Consider in Figure 2-4A



**FIGURE 2-3** Equilibrium Where Demand Equals Supply





**FIGURE 2-4** Market Effects of Supply-and-Demand Shifts

what would happen to the market for coffee if there were a freeze in Brazil. This worsening of the weather would tend to shift supply to the left as shown. At the new equilibrium, the price of coffee is higher, and the quantity consumed is lower.

Similarly, in Figure 2-4B, consider what happens to the market for tea when the price of sugar rises. Because sugar is a complement, this event causes a shift to the left in the demand for tea as shown. The new equilibrium will have a lower price and a lower quantity.

A few exercises help to generate experience with comparative statistics and to demonstrate the applicability of this analysis:

1. A national health insurance proposal is passed that provides comprehensive health insurance to millions more people than currently. How would this affect the markets for health care in the short run?

*Answer:* The competitive model is probably robust enough to give some guidance. Insurance coverage will probably increase on average, causing the demand for health care to increase, shifting the curve to the right. This will increase the equilibrium price of care, as well as the quantity consumed. The result will be an increase in the total money spent on health care. But recall that the analysis is conducted *ceteris paribus*: If an effective cost-control program were put in place at the same time this would reduce the pressure on costs, perhaps cancelling it out.

2. A new law requires that hospitals hire only nurses with baccalaureate degrees. How would this affect the market for hospital care?

*Answer:* Hospital markets are not perfectly competitive, but the predictions of the competitive model would probably give a good clue. Such a law would in effect increase an input price, shifting the supply of hospital care to the left. Under this interpretation, the equilibrium price of hospital care would tend to rise and quantity would fall.

3. Suppose that there is a big breakthrough in the technology for Lasik surgery, that is, surgery designed to correct nearsightedness. Suppose further that this cuts the price of Lasik to a tenth of its previous level with no loss in quality. How would this event affect the market for eyeglasses?

*Answer:* Lasik is a substitute for eyeglasses, and demand for eyeglasses would probably decline.

## FUNCTIONS AND CURVES

Most economic discussions consider how two or more economic variables are related to each other. For example, consider what we have theorized about the relationship of the price to the quantity demanded. We say that the quantity demanded is a function of the price. Mathematically, a function is an arrangement whereby we plug in the value of the independent variable, here the price, and the

function generates the value of the dependent variable, here the quantity demanded. Alternatively, we can say that quantity demanded depends on price.

### Linear Functions

Before considering the writing of supply and demand in functional notation, consider the linear function. A linear function is that of a straight line written as follows:

$$y = a + bx \quad (2.1)$$

where  $y$  is the dependent variable and  $x$  is the independent variable. A linear function, no matter what idea it represents, has characteristic features: an intercept and a slope, both of which are constants. The  $y$ -intercept is the value of the function evaluated when  $x$  equals zero. Here the intercept is  $a$ . The slope of a function is the increase in the vertical direction, or the “rise” divided by the increase in the horizontal direction, or the “run.”

To determine the slope of this particular function, examine this function drawn as a curve in Figure 2-5. We use the word *curve* for all functions, including the straight line. As noted, the  $y$ -intercept is  $a$ . Similarly, the function crosses the  $x$ -axis at a value of zero for  $y$ . Setting  $y$  to zero and solving for  $x$  yields a value of  $x = -a/b$ . Now, to find the slope, divide the change in  $y$ ,  $-a$ , by the change in  $x$ ,  $-a/b$ , thus generating a value for the slope of  $b$ , the slope of this linear function. The value of  $b$  in this case must be negative, as the slope is downward.

### Demand Functions

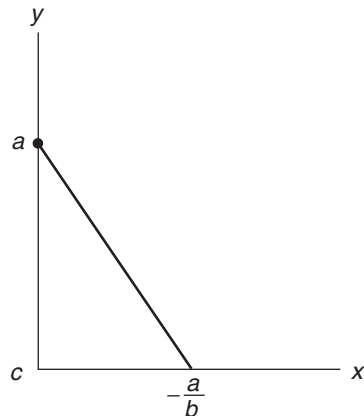
The demand functions up to this point have been linear. In general, though, linear demand is only one special case. Even when demand is linear, there is a minor complication. Consider the following linear demand function:

$$Q_d = a - bP \quad (2.2)$$

where  $Q_d$  is quantity demanded and  $P$  is price.

A complication arises because economists customarily draw demand and supply curves with the independent variable, price, on the  $y$ -axis and the dependent variable, quantity demanded, on the  $x$ -axis. In standard mathematics, the dependent variable is usually drawn on the  $y$ -axis.

Most commonly, we will be considering cases where the dependent variable, such as quantity demanded, is a function of not one but several variables. For example, the quantity of spaghetti demanded,  $Q_d$ , may depend not only on the price of spaghetti  $P_s$ , but also on the price of substitutes



**FIGURE 2-5** Graph of the Function  $y = a + bx$

for spaghetti (such as other pastas),  $P_o$ , the individual's income,  $Y$ , and a taste factor,  $Z$ . Mathematically, using the general notation, the demand function for spaghetti would be written as follows:

$$Q_d = f(P_s, P_o, Y, Z) \quad (2.3)$$

Here the notation  $Q_d = f(\dots)$  is read, "Quantity demanded is a function of  $P_s$ ,  $P_o$ ,  $Y$ , and  $Z$ ." If the function in (2.3) also happens to be linear, its more specific form would have a characteristic linear look to it. Statisticians frequently use this case, and it is useful to look at an example. A linear spaghetti demand function, for example, might look like this:

$$Q_d = 500 - 10P_s + 5P_o + 20Y + 40Z \quad (2.4)$$

Linear equations with several independent variables have some things in common with the simple linear equation in (2.1). An intercept constant is calculated by setting all the independent variables equal to zero; here the intercept is 500. The slope values in such linear cases will be the coefficients of the independent variables in question. For example, the slope value for the income variable  $Y$  is 20. The slope gives information regarding the contributions of changes in the independent variables to the value of the dependent variable. Again, it is worthwhile emphasizing that functions in economics need not be linear. For example, the true spaghetti demand function might instead look like this:

$$Q_d = 7P_s^{20.05} P_o^{0.002} Y^{0.8} Z^{0.01} \quad (2.5)$$

which is not linear. Our theory provides only a few strong conclusions about a demand function: It is downward sloping in its own price, shifting rightward (leftward) with higher prices of substitutes (complements), shifting rightward (leftward) with income increases for normal (inferior) goods, and shifting rightward with a positive shift in tastes. Beyond these features, the demand function mathematically could take on many different forms.

### Derived Demand

Demand by consumers for a final good or service may stimulate the provider of that service in turn to demand factors of production. There is no theoretical limit to how long this hierarchical chain can be, but in the health industry provides several good illustrations. Individual consumers may wish to improve their health or the health of family members. They may demand exercise equipment, healthful foods, and visits to a physician. These consumers generate a derived demand for factors. When considering just a clinic, the manager purchases electronics, technical equipment, drugs, and labor services, as well as the physician's time. The physician and other medically trained personnel, in anticipation of this, had a derived demand for medical education.

## CONSUMER THEORY: IDEAS BEHIND THE DEMAND CURVE

Consumer theory examines how rational individuals make consumption choices when faced with limited resources. The limited resources determine what options a consumer can afford. From among these options, the consumer attempts to pick the best one. The theory has two parts. One is a description of what the consumer prefers—what he or she thinks is best; for this description, we use the ideas of utility and of indifference curves. The second part is a description of what the consumer can afford; for this part, we use the idea of budget constraints. The use of indifference curves and budget constraints together constitutes indifference curve analysis.

## Utility

Consider a consumer, Kathy Richards. Let us suppose she is an experienced consumer who knows what her tastes are for kinds of houses to live in, cars to drive, food to eat, and books to read. She can't afford everything she would like to have, but she knows what she would prefer if she could afford everything.

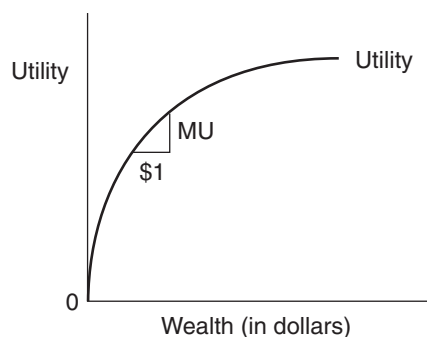
In summarizing this information about Kathy's preferences, we suppose that she has a utility function. Utility is a measure of her level of satisfaction with various combinations of consumer goods. It includes a market basket filled with a combination of housing, food, transportation, and so on, with perhaps many types of each. We assign a greater value of utility to bundles preferred over other bundles. Because more utility, thus defined, is always better, Kathy will logically seek to maximize her utility subject to the constraint of what is affordable to her.

Using functions, we say that Kathy's utility is a function of the goods and services she consumes. In practice, the level of detail we use will vary. On some occasions, we must specify most of the detailed consumption of Kathy's life. Then, we would describe her utility as a function of each good or service she buys, perhaps compiling hundreds of them. But in many cases, it is useful to abstract from this detail and describe Kathy's utility as a function of one or two goods of special interest, plus another general good representing all other goods she considers. At still other times, we will find it useful to describe Kathy's utility as a function of wealth.

We will develop two examples in the following discussion. Theories using the idea of utility may propose that utility is either cardinal or ordinal. Cardinal utility means a metric measure, like a measure of weight or volume. It is characteristic of cardinal measures that the difference as well as the ratio between two measurements has meaning. One can meaningfully say, for example, that five quarts is three more than two quarts, and, for that matter, it is two and one-half times two quarts. Under ordinal utility, to the contrary, only the ranking has meaning. Examples of ordinal numbers are first, second, third, and so on.

It is generally preferable to theorize that consumers' utility is ordinal. It is safer to assume that consumers can rank their preferences than to assume that they can both rank and scale them. It seems safer to suppose someone can say that he or she is happy to have gotten a raise, but it seems questionable to suppose that he or she is 1.07 times as happy. Most theories of demand assume only ordinality. In a few cases, such as the theory of behavior under risk and insurance, however, some degree of cardinality is assumed. To illustrate ideas about utility, we should begin with the simplest case to draw. Figure 2-6 depicts Kathy's utility as a function of her wealth. The curve illustrates two ideas. First, the upward slope indicates that utility increases with wealth. Second, the bowed shape, concave from below, indicates that her utility increases with wealth at a decreasing rate.

Marginal utility is defined as the extra utility achieved by consuming one more unit of a good. Here, the only good is wealth so marginal utility is the extra utility Kathy gets from one more dollar of wealth. An extra convenience of drawing the function graphically is that the marginal utility is the slope of the curve at a given point. Starting at any point on the curve and adding a single dollar to Kathy's wealth leads to extra utility for that dollar, which we have just defined as marginal utility (MU).



**FIGURE 2-6** The Utility of Wealth

If we understand marginal utility as the slope, the marginal utility of wealth for Kathy gets smaller as she gets wealthier. That is, the slope gets flatter. An extra dollar means more to Kathy when she is poorer than when she is richer. Does this notion apply to most people? That it might apply seems plausible to most students, but the notion also introduces an element of cardinal utility. This element of cardinal utility is essential when analyzing consumer decisions regarding risk and insurance, discussed in detail in Chapter 8.

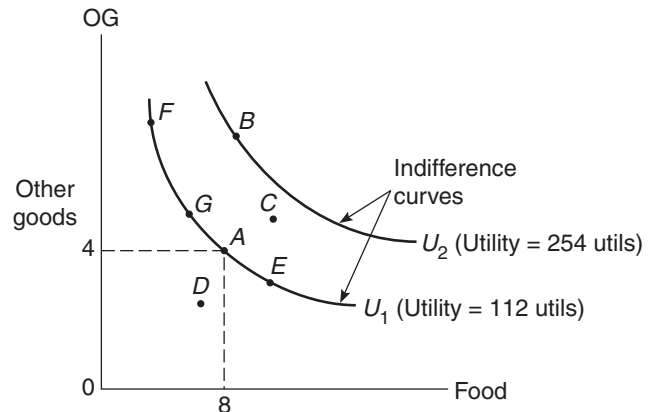
## Indifference Curves

Often, we wish to depict the consumer's preferences over two or more goods. The most convenient case to draw is when there are only two goods. To capture the sense of the real world in a two-good drawing, let one of the two goods represent all other goods generally, as if they were a conglomerate. Call this conglomerate OG, meaning Other Goods. Suppose that the good of special interest is Food. Figure 2-7 depicts a graph with these two goods, OG and Food, on the axes. Any point in the space, such as *A*, represents a consumer bundle. The bundle *A* includes the combination of eight units of Food and four units of Other Goods. Other bundles that are labeled include *B*, *C*, *D*, *E*, *F*, and *G*, but any other point in the space is also a bundle.

Suppose we focus on bundle *A* and that we hypothetically ask Kathy to identify all other bundles as well that for her are indifferent to *A* (that is, points that give her the same utility as *A*). The entire set of such points is labeled  $U_1$ ; as noted in the graph, any point along  $U_1$  affords Kathy 112 utils (we will call the units of utility *utils*).

This curve is downward sloping, as well as bowed toward the origin. Notice, for example, that Kathy did not choose point *C* as being indifferent to *A*. This seems plausible because *C* represents more of both goods, and as long as she is not satiated with these two goods, then she would prefer *C* to *A*. Likewise, she has not picked point *D* as indifferent to *A* because *D* has less of both goods. Instead, she has picked points such as *E*, which has more Food but less OG. Presumably, at *E* she has just balanced the loss of OG against the gain in Food. These considerations suggest that it is understandable that the indifference curve through *A* is downward sloping.

The curve is also convex to the origin. Consider that at point *F*, Kathy has relatively a lot of OG and little Food. As Food is relatively scarce for her, she is willing to give up a lot of OG to get more Food. We describe the rate at which she is willing to trade off the two goods by the slope of the indifference curve, which is steep at point *F*. In contrast, as we move down the indifference curve, Kathy gains relatively more Food, and the more she gets, the less ready she is to give up still further OG to gain yet more Food. Thus, the curve becomes flatter.



**FIGURE 2-7** Indifference Curves Between Food and Other Goods (OG)

Indifference curves for Kathy summarize and represent her preferences. Every possible combination of goods will lie on some indifference curve so that in principle there would be an infinite number of indifference curves in Figure 2-7, with higher curves representing greater satisfaction.

### Budget Constraints

Indifference curve analysis uses preference maps and budget constraints. The budget constraint indicates the set of bundles affordable with a given income. Suppose that Kathy must allocate \$30 of her family food budget per week between beef  $B$  and chicken  $C$ . If the price of beef is \$2.00 per pound, and the price of chicken is \$1.00 per pound, then she can afford any combination of  $B$  and  $C$  that costs less than or equal to \$30, and her budget constraint is:

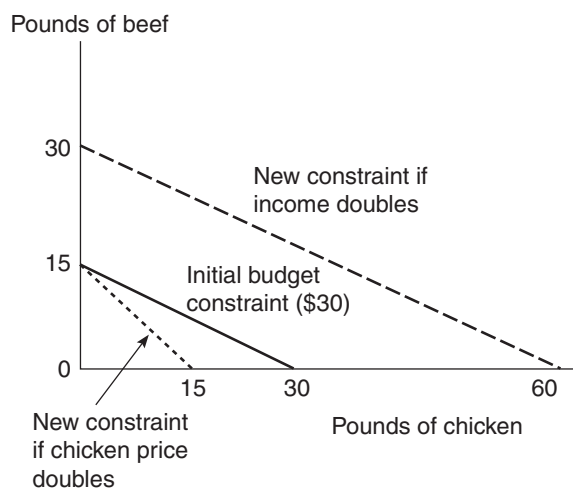
$$30.00 = 2.00B + 1.00C \quad (2.6)$$

If we draw this constraint in Figure 2-8 with beef  $B$  on the vertical axis, then the budget constraint will start at 15 pounds of beef and proceed downward sloping to 30 pounds of chicken as the horizontal intercept. It is convenient to demonstrate this pattern by examining the mathematical function. Equation (2.6) can be transformed using algebra so that  $B$  appears on the left-hand side and all other terms are on the right. Thus, the equivalent budget constraint is:

$$B = \frac{30}{2} - \frac{1}{2}C = 15 - 0.5C \quad (2.7)$$

a linear function with an intercept of 15 and a slope of  $-0.5$ . If, in Figure 2-8, the price of chicken rose, the amount that Kathy could buy, if she spent all \$30 on chicken, would be less than before. If the price doubled, the chicken axis intercept would shift inward, permitting her to buy only 15 pounds of chicken rather than 30. The beef intercept is not affected when the price of chicken rises.

Consider instead an increase in the portion of her budget allocated to beef and chicken. A doubling to \$60 would allow Kathy to increase the amount of beef from 15 to 30 pounds, or the amount of chicken from 30 to 60 pounds. As shown, the new budget constraint is parallel to the original budget constraint. Doubling the income in itself does not cause the prices to change. Because the slope of each budget constraint is the ratio of prices, the new constraint will be parallel. The intercepts will double.



**FIGURE 2-8** Changes in Budget Constraints Due to Changes in Price or Income (OG)

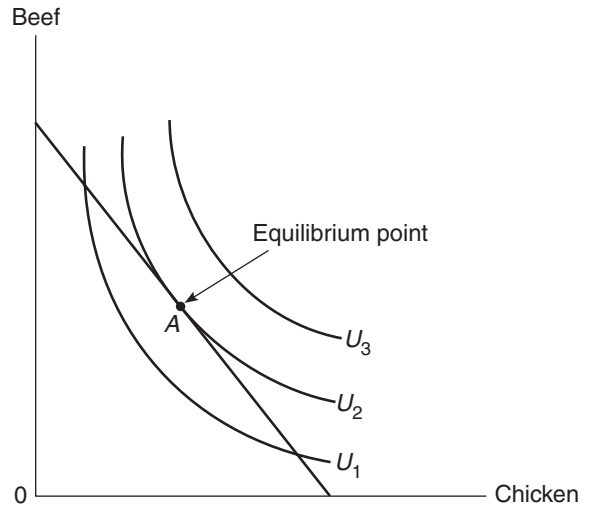


FIGURE 2-9 The Consumer's Equilibrium

### Consumer Equilibrium

To maximize satisfaction given a budget constraint, the consumer will seek the highest attainable indifference curve. In Figure 2-9, the indifference curve  $U_1$  is not the best possible, while the indifference curve  $U_3$  is unattainable. Rejecting such alternatives, the consumer will find that she maximizes utility at a point of tangency, shown as point  $A$  in the figure. At this point, the rate at which the consumer is willing to trade beef for chicken, the slope of the indifference curve equals the rate at which the consumer is able to trade the two goods at market prices—the slope of the budget constraint.

One can derive the equilibrium for different prices and/or for different values of income. For example, in Figure 2-10A, as the price of chicken rises consecutively, Kathy consecutively chooses points  $A$ ,  $B$ ,  $C$ , and  $D$ . The collection of such points is called the *price offer curve*. The information from these data points along the price offer curve can be used to generate her demand curve. For example, point  $A$  in Figure 2-10A represents her best choice for a given budget and set of prices. Suppose the price of chicken for budget constraint  $MN$  is  $P_0$ . Then, plotting  $P_0$  together with the quantity demanded at  $A$ , namely  $F_0$ , in the graph at right generates point  $A'$ . In the same manner,

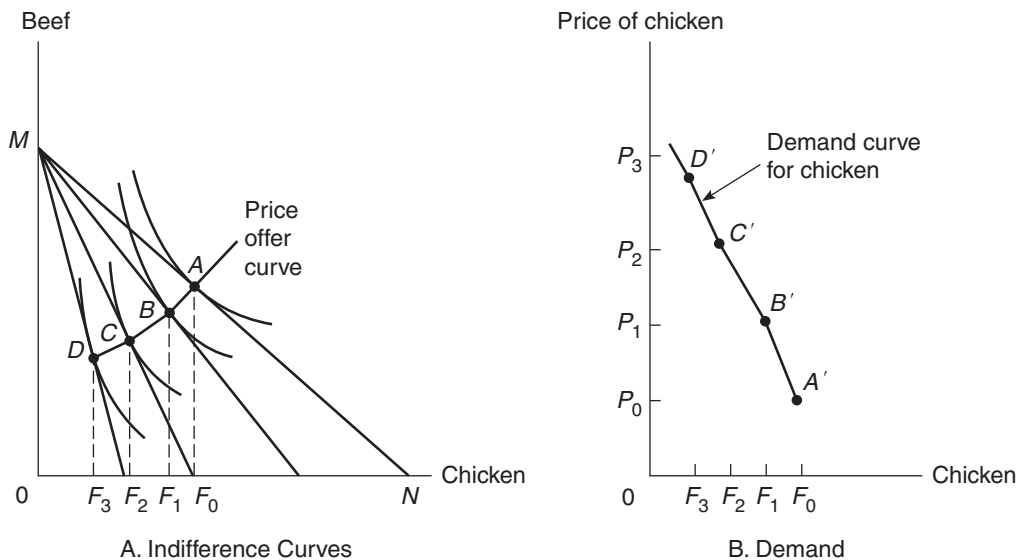
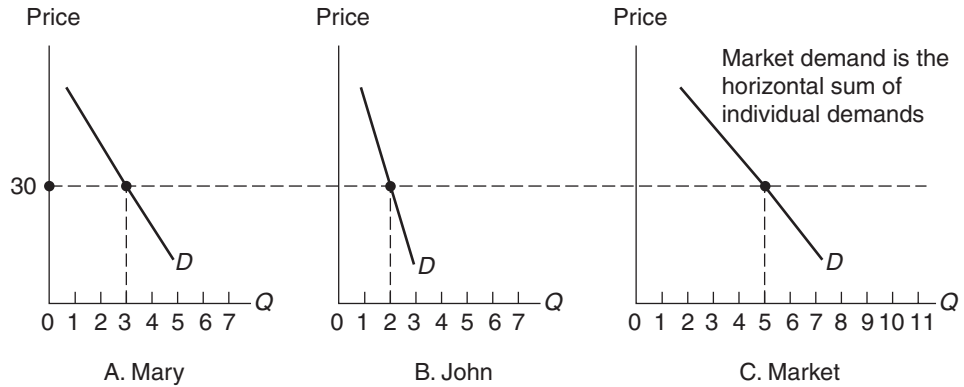


FIGURE 2-10 Derivation of the Consumer's Demand Curve



**FIGURE 2-11** Derivation of a Market Demand Curve

data on price  $P_1$  and the quantity demanded  $F_1$ , at point  $B$ , generate point  $B'$  in the graph at right. Repeating this process generates the demand curve.

## INDIVIDUAL AND MARKET DEMANDS

The theory of consumer behavior focuses on the demand relationships of individual consumers. Many applications of demand theory, however, consider market demand. The extension from individual to market demand is straightforward. In Figure 2-11, health care demand is shown for two people who constitute a market for health care. For example, Mary demands 3 units at the price  $p_H = 30$ , and John demands 2 units. Market demand here at  $p_H = 30$  is  $(3 + 2)$  or 5 units. The market demand in panel C is derived by adding the quantities demanded at every price. The process can be extended to all of the consumers in the market, and it yields a market demand curve. If everyone's demand curve conforms to the law of demand, the market demand curve must also be downward sloping.

Finally, note that as with individual demand functions, other variables such as income and the prices of related goods (the shift variable we discussed earlier) affect market demand. Thus, the market demand for some commodity  $X$  might be expressed in functional notation. Consider, for example

$$QD_x = f(P_x, Y, P_o, E) \quad (2.8)$$

where  $Y$  represents income,  $P_o$  represents the prices of other goods, and  $E$  represents a socioeconomic variable such as average educational attainment (in years of schooling).

## ELASTICITIES

We often seek to understand the responsiveness of the quantity demanded to other variables. For example, if the price of health care rises, will the quantity demanded fall by a large amount or a small amount? Economists use the term *elasticity* to describe the responsiveness of any term  $y$  (in this case, quantity demanded of health care) to changes in some other variable  $x$  (here, price of health care).

Elasticity is defined as the percent change in the dependent variable resulting from a one percent change in the independent variable under study.<sup>1</sup> In the case of the price elasticity of demand, it is as follows:

$$E_p = (\% \text{ change in quantity demanded}) \div (\% \text{ change in the price})$$

<sup>1</sup> Percentages allow us to "standardize" our measure. In particular, the use of percentages eliminates problems comparing various goods measured in pounds, liters, or meters. Here we measure them all in percent. For example, it becomes possible to compare demand elasticities for beef with automobiles, even though the price levels and quantities are different.



or

$$E_p = \left( \frac{\Delta Q/Q}{\Delta P/P} \right) = \frac{\Delta Q}{\Delta P} \left( \frac{P}{Q} \right) \quad (2.9)$$

where  $\Delta$  refers to change in the variable. The price elasticity is always algebraically negative because an increase in price leads to a decrease in quantity demanded. We derive other elasticities, such as the income elasticity of demand, similarly.

$$E_y = (\% \text{ change in quantity demanded}) \div (\% \text{ change in income})$$

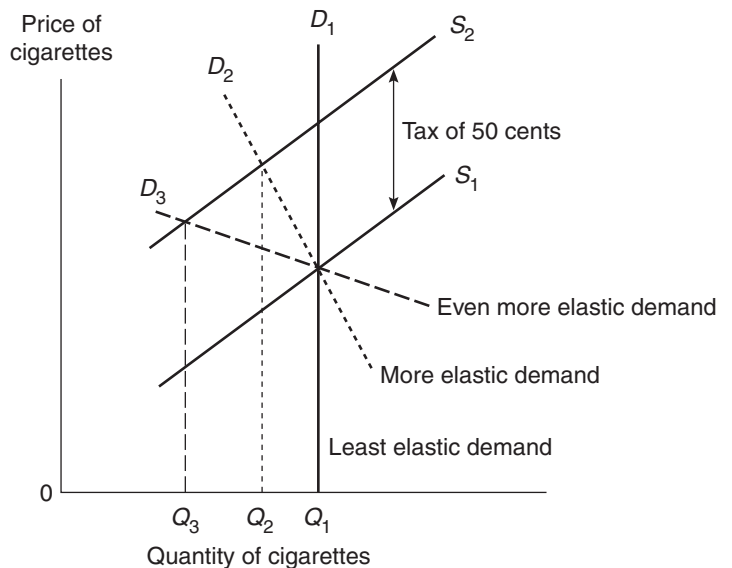
or

$$E_y = \left( \frac{\Delta Q/Q}{\Delta Y/Y} \right) = \frac{\Delta Q}{\Delta Y} \left( \frac{Y}{Q} \right) \quad (2.10)$$

Income elasticity may be positive (if a normal good) or negative (if an inferior good). If a variable elicits no response at all, then elasticity is zero.

Numerical values for price elasticities are often reported in absolute values, eliminating the minus sign. Absolute values for price elasticities indicate the responsiveness of demand to price in that the greater the elasticity, the greater the responsiveness. Absolute values greater than 1 are considered relatively responsive and are called elastic. Elasticities less than 1 in absolute value are called inelastic. In the limiting cases, zero price elasticity means that the demand curve is perfectly vertical, while infinite price elasticity means that the demand curve is perfectly flat.

The importance of price elasticity to policy questions can be illustrated with an example regarding cigarettes, which are clearly a health concern. Suppose that a state added 50 cents per pack to its tax on cigarettes. Together with supply-and-demand analysis, our elasticity concepts help us identify the main policy issues. Lawmakers hope that such a tax increase will curb smoking and bring in tax revenue, but these tend to be contradictory goals. The exact effects will be difficult to predict unless reliable estimates are available of the cigarette price elasticity. If one discovered that demand is perfectly inelastic ( $D_1$  in Figure 2-12), tax revenue would be at a maximum but with no effect on



**FIGURE 2-12** The Impact of a Cigarette Excise Tax

**TABLE 2-3 Demand Price Elasticities for a Variety of Goods**

Good or Service	Price Elasticity
Hospital Care <sup>1</sup>	−0.14 to −0.17
Physician Care <sup>2</sup>	−0.16 to −0.35
Apples (U.S.) <sup>3</sup>	−1.15
Bread (U.K.) <sup>3</sup>	−0.26
Gas, Short Run (Canada) <sup>3</sup>	−0.01 to −0.20
Cigarettes (U.S.) <sup>4</sup>	−0.30 to −0.50
Beer (U.S.) <sup>5</sup>	−0.20 to −0.40
Beef (U.K.) <sup>3</sup>	−1.45
Motion Picture Tickets <sup>6</sup>	−3.40
Foreign Travel <sup>6</sup>	−3.10

Notes: <sup>1</sup>Wedig (1988); <sup>2</sup>Manning (1987); <sup>3</sup>Mansfield et al. p. 103 (2002); <sup>4</sup>Keeler (1993); <sup>5</sup>Grossman (1998); <sup>6</sup>Ruffin & Gregory, p. 102 (1997).

smoking or health. Alternative scenarios of increasingly elastic demand ( $D_2$  and  $D_3$ ) create bigger reductions in smoking but at the cost of decreasing tax revenues. Thus, the more elastic the response, the greater the effectiveness of an excise tax in inducing people to reduce their levels of smoking. Lewit and Coate (1982) indicate that teenagers, for example, are more responsive to cigarette prices than are adults. In such cases, taxes on cigarettes will be relatively more effective with teenagers.

Market demand elasticities vary by industry and by product. Those goods and services that we call necessities tend to have elasticities less than one in absolute value, while luxuries are more elastic. Short-run elasticities are generally smaller in absolute value than long-run elasticities. Further, goods that cost only a tiny fraction of one's income motivate little or no "shopping around," making their demand elasticities very small in absolute value. Table 2-3 provides some common estimates of demand elasticities for a variety of products, and readers can compare these findings to their own personal experiences.

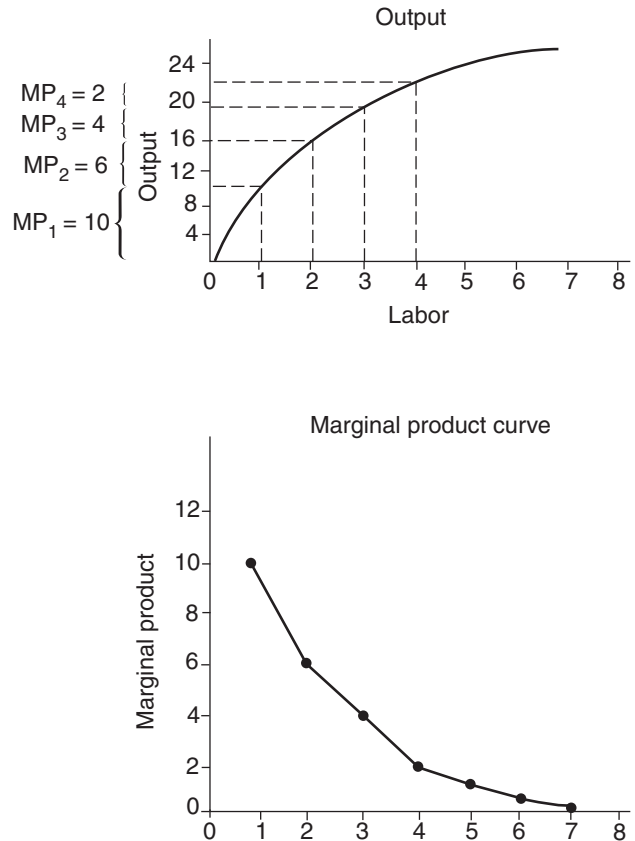
## PRODUCTION AND MARKET SUPPLY

If market demand is one "blade of the scissors" in determining the price of a good, market supply by individual firms is the other. A typical producer, or firm, faces decisions on production levels, prices, production methods to use, levels of advertising, and amounts of inputs to purchase. The theory of the firm, much like the theory of consumer behavior for buyers, develops a framework for understanding these choices.

The key assumption for most models of firm behavior is that the decision makers wish to maximize profits. It follows that the firm will try to minimize the costs of producing any given output and will undertake activities, such as advertising, only if they add to profits. Before examining such decisions, we will review production and cost relationships.

### The Production Function

The production function shows the maximum sustainable output that can be obtained from all of the various possible combinations of inputs such as labor, materials, and machinery, with existing technology and know-how. We begin our discussion of production functions with a simple case—one in which there is just one input and one output. Suppose that food, perhaps in a hunter-gatherer society, was produced solely with labor. We show the production function in the top panel of Figure 2-13. The fact that the production function is rising indicates the idea that labor is productive; more labor means that more food is produced. The bowed shape of the curve illustrates a second idea—the law of diminishing returns.



**FIGURE 2-13** Output and Marginal Product of Labor

It is understood in Figure 2-13 that the production takes place during a specified period of time. Thus, the output axis represents a flow of output per unit of time. Likewise, the labor input represents labor services applied during a specified period of time. The law of diminishing returns represents the idea that the marginal product (MP) of an input will eventually tend to fall as more is added. The MP (in this example, labor) is defined as the extra output that can be generated when one adds an additional hour of labor, holding all other inputs constant. In the figure, the food output increases from zero to ten units when the first hour of labor is added. Thus, ten units of food is the marginal product of the first unit of labor. When a second hour of labor is added, the output of food increases from 10 units to 16 units. The extra amount is six units, meaning that the marginal product of the second hour of labor is six units of food. The bottom panel of the graph illustrates the marginal product (MP) of consecutive hours of labor. The pattern of the MPs in this illustration is clear: They tend to get smaller as more labor is applied. This illustrates the law of diminishing returns. Notice that total output need never fall during diminishing returns. That is, the production curve itself never turns downward in this illustration, although it may in some applications.

**Production Functions**

In practice, production processes may involve several inputs, not just labor. It is convenient to express the production relationship for a firm, or a unit of the firm, as follows:

$$Q = f(X_1, X_2, \dots, X_n) \tag{2.11}$$

Here  $Q$  represents output;  $X_1, X_2,$  and so on are quantities of the various inputs. The ultimate output of the health industry is health, and its related production issues are termed the “production of

health,” a phrase that applies not only to larger populations but also to the consumer’s production of his or her individual health. When individuals seek to improve their family’s health, they often purchase health care from hospitals, clinics, nursing homes, home health agencies, and so on. One step removed logically from health production, these elements of health care must themselves be produced, and much economic research and concern are directed at this production stage, too.

Consider, for example, the output of hospital X-ray services, which require labor by technicians, nurses, and radiologists, and machinery such as X-ray machines, computers, and film. As with demand functions, production functions may take on many mathematical forms. The theory of production functions specifies only certain patterns for these functions.

One commonly applied functional form that fits the theoretical patterns for such functions is the Cobb-Douglas form. Historically, it was one of the earliest production functional forms to be studied and applied to firms. It derives its name from mathematician Charles Cobb and economist (and later U.S. Senator) Paul Douglas. Many other functional forms of production have since been investigated, but this form is still commonly used in the classroom to illustrate the mathematics of the production process.

If the production of X-ray services just discussed fits the Cobb-Douglas form, and if inputs of all kinds are grouped into the categories of capital,  $K$ , and labor,  $L$ , the production function actually estimated might look like this:

$$Q = L^{0.8}K^{0.2} \quad (2.12)$$

Here, as with any production function, there will be a unique level of maximum output for any combination of inputs. For example, Table 2-4 shows values of output corresponding to changes in  $L$ , holding  $K$  constant at five units. The change in output associated with a one-unit change in  $L$  is the marginal product of labor. More  $L$  (e.g., technicians) with the same amount of  $K$  will typically produce more  $Q$  (hence, a positive marginal product of labor), but as the machines become crowded or break down, the marginal product may fall as  $L$  increases. Hence, the marginal product in Table 2-4 diminishes. This decreasing marginal product again illustrates the law of diminishing returns. The average output or average product (AP) for each worker is  $Q/L$ , as shown in the last column.

Alternatively, we can derive various input combinations needed to produce a given output level. Table 2-5 illustrates several combinations that produce ten units of output for the production function represented by equation (2.12). This method closely parallels the indifference curve analyses introduced in an earlier section, except that we actually observe and measure the quantities produced (in contrast to the levels of utility that could only be ranked). Combinations of inputs producing equal output levels lie on an isoquant (literally, “the same quantity”). The isoquant in Table 2-5 is illustrated graphically in Figure 2-14. The isoquant map, representing all possible values of  $Q$ , would be the geometric representation of a production function.

**TABLE 2-4** Production Schedule for X-ray Services

K	L	Q	MP	AP
5	0	0.00	—	—
5	1	1.38	1.38	1.38
5	2	2.40	1.02	1.20
5	3	3.32	0.92	1.11
5	4	4.18	0.86	1.05
5	5	5.00	0.82	1.00
5	6	5.79	0.79	0.97
5	7	6.54	0.75	0.93

**TABLE 2-5 An Isoquant Schedule**

Q	L	K
10	1	100,000.00
10	5	160.00
10	7	41.60
10	8	24.40
10	10	10.00
10	11	6.80
10	13	3.50
10	20	0.63

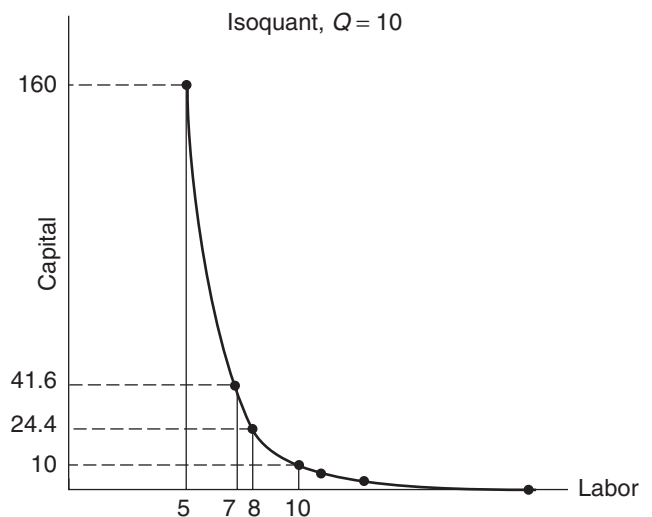
The negative slope to an isoquant indicates the possibility of substituting inputs in the production process and of the positive marginal product of the inputs. Consider, again, the example of X-ray services. The numerical value of the isoquant slope, indicating how much capital (X-ray machines, film, computers) must be given up per unit increase in labor (nurses, technicians, radiologists), is called the marginal rate of technical substitution of labor for capital ( $MRTS_{LK}$ ).

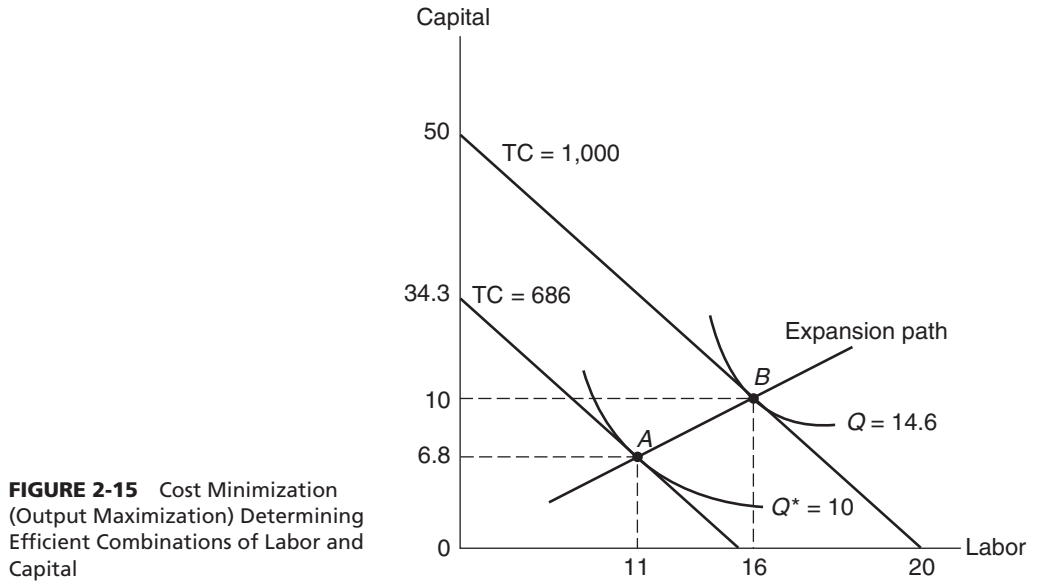
**IS SUBSTITUTION POSSIBLE IN PRACTICE?** Empirical estimates reveal substantial substitution possibilities between physicians' assistants and physicians. Other studies reveal substitution between nurses and residents and between hospital capital and hospital staff. How can physical capital substitute for a human medical practitioner? Later chapters discuss these issues.

### Isocost Curves

In order for the X-ray unit to maximize profits, the unit will want to minimize the cost of producing any given output. Letting  $TC$  represent total costs,  $w$  the price (wages, salaries, fringe benefits) of labor, and  $r$  the cost of buying or renting machines for the production period, the total cost is as follows:

$$TC = wL + rK \quad (2.13)$$

**FIGURE 2-14** A Production Isoquant



where  $L$  and  $K$  are the amounts of inputs used, labor and capital. For example, if  $w = 50$ , and  $r = 20$ , when the unit uses 30 hours of labor and 10 machines,  $TC = 1,700 = (50 \times 30) + (20 \times 10)$ . As with the consumer's budget problem, it is helpful to determine all of the combinations of  $L$  and  $K$  that cost a given amount, such as \$1,000. The equation for this isocost curve is

$$1,000 = 50L + 20K \quad (2.14)$$

Again, as with the consumer example, we can re-arrange the equation by placing capital (or labor) on the left-hand side, to yield

$$K = 50 - 2.5L \quad (2.15)$$

which is a linear equation as shown in Figure 2-15. The isocost curve for  $TC = 686$  is also shown. More generally, equation (2.13) can be written as

$$K = TC/r - (w/r)L \quad (2.16)$$

Equation (2.16) shows the impacts of changes in wages and/or rental rates on the costs of purchasing various amounts of labor and/or machinery.

### Cost Minimization or Output Maximization

In the example just given, the assumption that firms maximize profits requires that the X-ray unit select the least-cost method of producing its output. To produce  $Q^*$  units of output, the firm will wish to minimize the costs of that  $Q^*$  output, and hence be on the lowest possible isocost curve. The case where  $Q^*$  equals ten units is illustrated in Figure 2-15. This will occur at point A, where the isocost curve is tangent to the isoquant representing ten units of output. Higher isocost curves are unnecessarily wasteful; lower ones will not attain ten units of output.

Suppose instead that the firm has a budget of exactly \$686 and wishes to maximize its output. It is easy to see from Figure 2-15 that the same equilibrium condition will hold, and that the most

the firm can produce is ten units. Cost minimization and output maximization in the manner described lead to the same results.

**LOOKING AT COSTS FROM A DIFFERENT VIEWPOINT** Cost minimization can also be looked at from society's point of view. For example, hospitals may achieve cost minimization in applying hospital inputs, and home health care services may achieve cost minimization in applying home health care resources. But, for society as a whole to minimize its costs of care, we need to know which of these types of care is the most cost efficient for particular patients, home health care or inpatient hospital care, and we need to know if the quality of care is comparable.

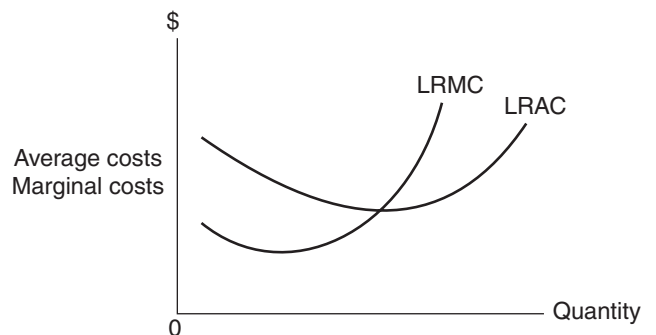
### Marginal and Average Cost Curves

By varying the production levels and finding the respective isocost curves, we can find the minimum cost of producing each output level. This is shown by the set of tangencies in Figure 2-15. The curve connecting these tangency points, *A* and *B*, is called the expansion path. Thus, the expansion path contains the information on the total cost and the average cost (cost/unit) of producing any output level.

If all inputs can be varied, then the long-run total cost and long-run average cost (LRAC) functions are generated. The LRAC curve is illustrated first in Figure 2-16. Total and average costs are related to the scale of the activity. If higher levels of production lead to improved ability to take advantage of specialization providing a better division of labor, it may be possible to reduce average costs; the case of decreasing long-run average costs is referred to as the case of economies of scale. If, on the other hand, the increased level of output leads to difficulties in managing and coordinating production activities, then long-run average costs may rise; this is referred to as the case of diseconomies of scale. Such issues are relevant for determining the optimal size for firms. For example, the socially optimal size and distribution of hospitals will depend on estimates of scale economies. As another example, it is clear that enough patient volume is needed to cover costs of such high-priced items as CAT scanners, a case of economies of scale. However, too many patients may lead to crowding of patients or to increased labor costs that could again increase costs, producing a case of diseconomies of scale.

In our specific Cobb-Douglas production function example, the LRAC will be a horizontal line reflecting constant average costs (about \$68.60 per unit). This occurs simply because of the specific production function chosen as an example. Figure 2-16 shows the classical U-shaped relationship, which starts with economies of scale and then yields to diseconomies of scale. The long-run marginal cost (LRMC) curve shows the cost of producing an incremental unit when all inputs (both machinery and labor) can be varied. It will go through the minimum point of the LRAC.

The short run corresponds to a period where at least one input (typically machinery or plant) cannot be changed. This form is particularly applicable to big-ticket machinery items in hospitals, for example. In such cases, there are some fixed costs (the machine costs or plant costs) that cannot be changed in the short run. The other costs are called variable costs (e.g., labor costs).



**FIGURE 2-16** Long-Run Cost Curves for a Firm

## THE FIRM SUPPLY CURVE UNDER PERFECT COMPETITION

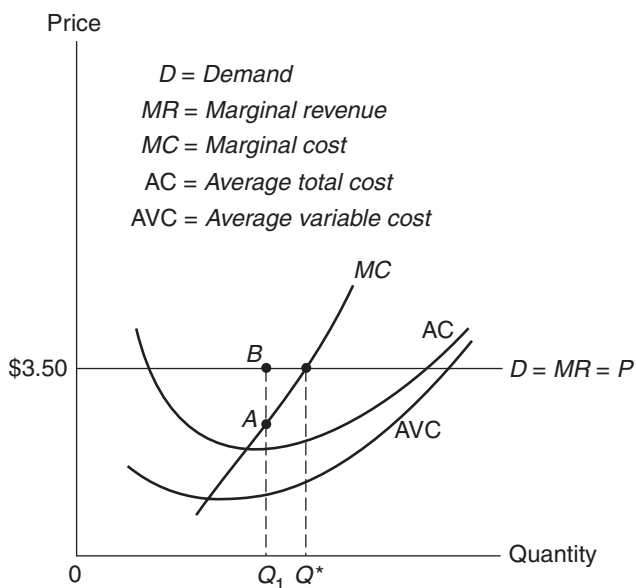
The cost curves we have reviewed can help to develop a theory of the supply curve for a firm, but to do so we must know something about the demand curve for the firm's product. In our earlier practice with supply and demand, the demand curve represented the market demand for the product. The demand for a single firm's product may be different. To gain an idea of what a typical firm's demand curve will look like, we must know what type of market structure we are talking about.

Several market structures provide insights to an idealized world or applicability to the real world. One defining principle that distinguishes the various market structures is the degree of control that individual firms have over the price that they get paid. Two cases define the extreme forms of market structure: the competitive and the monopoly cases. We will look first at the competitive model, then discuss market structure more generally, and finally follow that discussion with the monopoly case.

The competitive model is rarely seen in the world in its idealized form. It requires several assumptions that ensure perfect competition. The assumptions are as follows:

1. A sufficient number of buyers and sellers of the good exist so that no single buyer or seller has any power over the price.
2. The good is homogeneous; that is, all producers produce the exact same good so that the market cannot be segmented on the basis of difference of goods.
3. Information is perfect. All buyers and sellers have information on all relevant variables such as prices and qualities.
4. No barriers to entry or exit are present. A producer starts producing, buying necessary machinery, patents, or anything else on terms that are equivalent to those already in the industry.

These assumptions ensure that a short-run market equilibrium can be represented by the price and the quantity at which demand and supply curves intersect. Figure 2-17 illustrates the model. Under the assumptions of competition, the demand curve facing the firm will be flat, as shown by the curve  $D = MR = P$ . To understand this point, consider a mental experiment. Suppose that the market for wheat was competitive and that it had determined, by the actions of market demand and supply, some equilibrium price for wheat, say \$3.50 per unit, as shown. Suppose now that a single firm chose to raise its price above \$3.50. Would anyone buy its product? They would not because they



**FIGURE 2-17** The Competitive Firm's Supply Curve



know (perfect information) that they can buy an identical product (homogeneous product) elsewhere for \$3.50. In theory, at even a slightly higher price, the quantity demanded would slide horizontally to zero. On the other hand, suppose that the farmer wished to double the output. Would the farmer have to lower the price in order to sell it all? This would not be necessary because the farmer's output is small relative to the whole market (numerous buyers and sellers); hence, the farmer could sell as much as he or she wanted at the going price of \$3.50.

The demand curve for this firm is labeled  $D = MR = P$  indicating that it represents demand, as well as the marginal revenue for the firm and the going market price. Note that the market price is identical to the firm's marginal revenue. Marginal revenue is defined as the extra revenue obtained by selling one more unit of product. Because this firm can sell all it wants at the going market price, it can sell the marginal unit at that price, as well. Thus, marginal revenue equals price in the competitive model. It is only in cases where a firm has some monopoly power that marginal revenue will differ from price.

The profit-maximizing output for this competitive firm can now be deduced. The firm will maximize its profits at that output where marginal revenue (the price) equals marginal cost; this occurs in the figure at output  $Q^*$ . This output is called the firm's profit-maximizing output. The common sense of this seemingly technical proposition can be understood by examining a "wrong" output level—one that is not profit maximizing—for example,  $Q_1$ . Suppose a firm that is currently producing  $Q_1$  units were to produce one more unit. The cost of this one extra unit would be  $A$ , by definition the point on the marginal cost curve at  $Q = Q_1$ . The revenue from this one extra unit would be  $B$  (which equals the price), the point on the marginal revenue curve at  $Q = Q_1$ . The firm would increase its profits by producing that extra unit and would continue to increase its profits as long as the marginal revenue curve was above the marginal cost curve. Hence, maximum profits would occur only where  $MR = MC$ .

The supply curve for a firm shows the firm's profit-maximizing output at each possible price. The competitive firm is producing at the output where price equals marginal cost. If the market price were to rise in steps, the firm's adjustment steps would just be to follow the marginal cost curve on up.

The competitive firm's supply curve will be its marginal cost curve, as long as the price is sufficiently high to make it worthwhile to produce at all. Price must at least cover the firm's average variable cost (AVC).

The competitive market supply will be determined by the horizontal sum of the individual firm supply curves. This horizontal summing is done much in the manner in which we found market demand curves. The market supply curve in the competitive case, the sum of firm marginal cost curves, will also represent the industry marginal costs of production. In general, under competition, the supply curve is the industry marginal cost curve.

What then determines a good's market price? The answer is that combination of output and price at which market quantity demanded equals market quantity supplied, or the intersection of market demand and supply.

The assumption of free entry and exit, however, offers further insight into the workings of the competitive market. Suppose, for example, that the equilibrium price in the wheat market in the short run is high enough so that producers in the sector may earn attractive economic profits.<sup>2</sup> Any positive economic profit will be attractive to potential entrants. With perfect information and no barriers to entry, other suppliers will enter the market. This will increase market supply and drive down market prices. The entry process logically would continue in the long run until the prices have fallen enough to eliminate economic profits. In the long run, equilibrium profits will be zero, and price will be at the lowest point on each firm's long-run average cost curve.

However, if barriers to entry in the form of licensure or other restrictions exist, this adjustment process will be impeded. Sellers will be able to earn economic profits over long periods of

---

<sup>2</sup> Economic profits represent profits after considering all costs including opportunity costs. A "normal" level of profits is necessary to keep firms in the market, and is a factor payment to the entrepreneur, just like factor payments to workers (wages or salaries) or to owners of machines or materials.

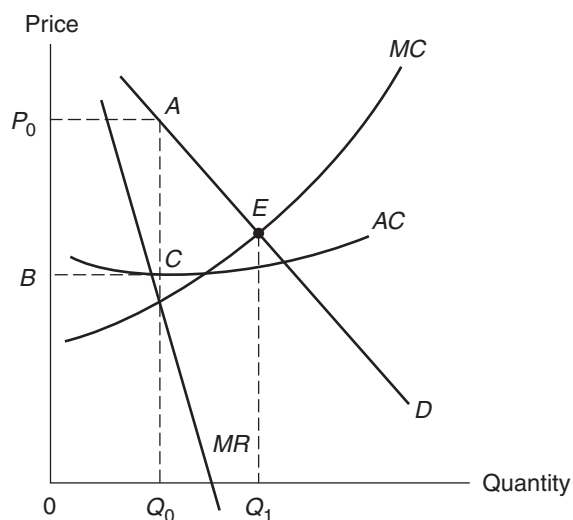
time, perhaps indefinitely. Such a situation is not perfectly competitive even though the forces of demand and supply determine price at any moment. It is important to evaluate the extent to which all of the four conditions for competition stated at the beginning of this section are satisfied.

## MONOPOLY AND OTHER MARKET STRUCTURES

Firms in other market structures, unlike the perfect competitor, have market power, which is the ability to affect market price. These market structures entail the pure monopoly (and the natural monopoly version), monopolistic competition, and the several forms of oligopoly. In nearly all instances, the market power concept determines the characteristic of choice of optimal price or quantity. This concept is well illustrated by the pure monopolist.

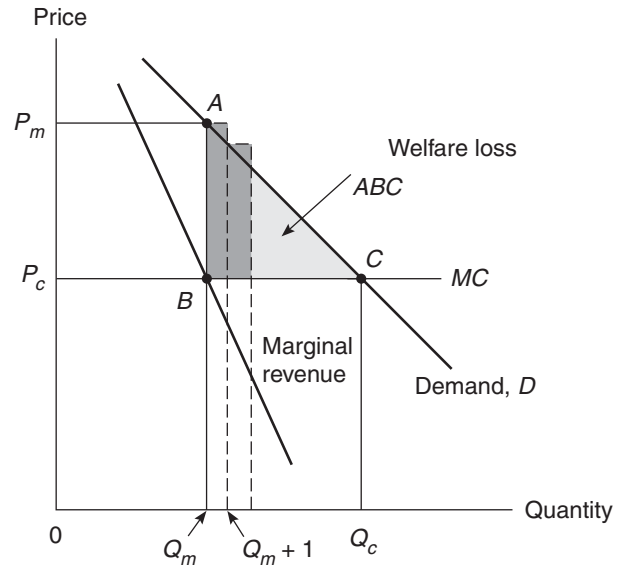
A pure monopoly is an industry with a single seller who has no close substitutes. As such, the monopolist faces the whole market demand curve, which is usually downward sloping. Downward sloping demand and market power are synonymous concepts; this is because the monopolist, unlike the perfect competitor, will not lose all its customers when raising its price.

In health sectors, pharmaceutical firms that control patents for certain drugs may be pure monopolists. Individual physician practices are not pure monopolies, but because the numerous competitors of each are differentiated by reputation, patient loyalty, and patient/practice distance, each physician probably has some market power. Many economists treat physician markets as monopolistically competitive. Finally, when a town has only a few hospitals, each hospital also has some market power. There being few firms in that market, the hospitals would be considered an oligopoly.<sup>3</sup> Equilibrium for the monopolist is illustrated in Figure 2-18. The demand curve facing the monopolistic firm is downward sloping because the monopolist faces the whole downward sloping market demand curve. With a downward sloping demand curve, the incremental, or marginal revenue (*MR*) is less than the demand price. Why is this the case? Suppose the monopolist were selling  $Q_0$  units at price  $P_0$ . Total revenue,  $TR_0$ , would be  $P_0Q_0$ . The monopolist would be selling to everyone who is willing to pay at least the price  $P_0$ . In order to sell one more unit of the good, the monopolist would have to induce more consumers to buy by lowering the price. It may be impossible to lower the price to extra consumers without also lowering to all previous consumers. In this case, because the monopolist must



**FIGURE 2-18** The Monopolist's Equilibrium

<sup>3</sup> Degree of monopoly is a recurring theme in the discussion of market power and the need for regulation. We will return to this subject repeatedly throughout the text.



**FIGURE 2-19** Welfare Loss of Monopoly

lower the price to everyone else, the marginal revenue will be the price of the extra unit of the good minus the loss of revenue from everyone else who now pays less.

To maximize profit, the monopolist produces where  $MC = MR$ , at  $Q_0$  in Figure 2-18. The corresponding price is  $P_0$  and total profit is the rectangle  $P_0ACB$ . If barriers to entry are persistent, the economic profits can be maintained and even increased through advertising, promotion, new product development, or other means. The fact that the monopolist is earning excess profits suggests that the monopolist has reduced the amount produced from the competitive amount. The monopolist in the graph chooses point  $A$  on the demand curve. If the monopolist had acted like a competitor by setting a price equal to marginal cost ( $MC$ ) it would have chosen quantity  $Q_1$ , point  $E$ , providing more output and charging a lower price. The induced scarcity caused by the monopolist necessarily raises the price to the consumer.

Economists seek to compare different economic situations or to examine the effects of different policies. In doing so, they often use the concept of allocative efficiency. One of the most widely used examples to illustrate the problem of inefficient resource allocation is found in the comparison of monopoly and competitive equilibria.

Consider once again the case of monopoly. Figure 2-19 shows the long-run marginal cost curve in an industry with constant marginal costs. With demand curve  $D$ , the competitive market equilibrium is at  $P_c$  and  $Q_c$ . Suppose, somehow, that the providers are able to form a monopoly. If so, it will be in their interest to raise prices by withholding services from the market. The resulting monopoly will produce at quantity  $Q_m$ , consistent with price  $P_m$ , where price is higher than marginal cost. The result reflects a loss to the consumer due to the monopoly. The total loss is indicated by the triangle  $ABC$ , called the welfare loss.

We can understand welfare loss better if we think in terms of marginals. Consider first that the demand curve measures the highest price that people are willing to pay for an extra unit of the good. The price they are willing to pay measures their marginal benefit. Now consider what the consumer and society as a whole have to give up when they face a monopoly. The monopolist will choose output level  $Q_m$ . If we somehow could have produced one more unit of the good, the  $(Q_m + 1)^{\text{th}}$  unit, we would have made a net gain for society. The benefit of that extra unit is  $1 \times A$  in the graph because by marginal benefit (here equal to the height of the demand curve) we mean the benefit of the extra unit. Similarly, the cost to the monopolist, and thus to society as well, of the extra unit is  $1 \times B$ . Because the marginal benefit exceeds the marginal cost, the extra unit yields society a net gain of the rectangular shaded area.

**BOX 2-2****Is Competition Better Than Monopoly?**

Virtually all economists greatly admire competition and competitive markets. But economists also understand that the benefits of competition are more likely to arise when competitive conditions are fully in place. Health economists are well aware that many health care markets don't fit the competitive conditions perfectly.

The question has been explored both theoretically and empirically. Brekke, Siciliani and Straume (2008) looked at the impact of hospital competition on hospital waiting times. First, they compared hospital markets with others having more firms in them. Second, they compared hospital markets with differing transportation costs, thus enabling consumers to reach more substitutes hospitals cheaply. They found mixed results on waiting times; lower transportation costs tended to increase waiting times, while increased hospital density tended to lower them.

Gaynor and Vogt (2003) approached the question empirically and focused on hospital prices. They exploited the fact that hospital mergers result in increased monopoly power. From estimates on California hospitals, they simulated the effect of mergers on price for two hospital chains. For San Luis Obispo County, where a merger would create a virtual monopoly, they projected a price increase of 57 percent.

Reasoning iteratively, another unit again yields another net gain to society, this time somewhat smaller than the first net gain. Net gains will continue to occur until we reach the output at which society's marginal benefit (demand) intersects its marginal costs, which occurs at *C*. The total net gain to society from increasing output up to the point at *C* equals the triangle labeled *ABC*.

**CONCLUSIONS**

The microeconomic tools developed in this chapter consist of the production possibility frontier, demand-and-supply analysis, utility and indifference curve analysis, production and cost curves of a typical firm, firm behavior under competition and monopoly, and the measure of welfare loss. The economic tools used later in the text apply or extend the tools developed here. By learning these ideas, you will gain an understanding of the terminology used in health economics, as well as an understanding of the type of reasoning used.

**Summary**

1. The concept of scarcity underlies much economic thinking. Scarcity necessitates that decision makers make trade-off decisions at the margin. The production possibilities frontier represents these trade-offs, and its slope represents the opportunity cost of one good in terms of the other.
2. Supply-and-demand analysis of competitive markets is a basic tool of economics and provides insights that extend beyond the theoretical, perfectly competitive markets. Supply reflects sellers' offers as a function of price, and demand reflects buyers' offers as a function of price. The intersection of demand and supply describes the market equilibrium.
3. Comparative static analysis of demand and supply finds the new equilibrium after economic events shift either curve. Demand-increasing (-decreasing) events tend to raise (lower) equilibrium price, while supply-increasing (-decreasing) events tend to lower (raise) equilibrium price.
4. A relationship between one or more independent variables yielding a unique value for the dependent variable is called a function. The linear demand function, showing demand as a straight line, is only one special case of the many possibilities.
5. The utility function summarizes a consumer's preferences. Higher utility numbers are assigned to

consumer bundles that provide higher levels of satisfaction, meaning that the consumer prefers these bundles.

6. Indifference curves are collections of points describing bundles that yield the same utility and hence the same level of satisfaction. Well-behaved indifference curves are downward sloping, continuous, and convex to the origin.
7. The budget constraint represents the combinations of goods that the consumer can afford given his or her budget. The budget constraint is downward sloping, and its slope is the negative of the ratio of prices.
8. In consumer theory, the consumer maximizes utility subject to a budget constraint. This means that the consumer picks the most preferred consumer bundle from among those he or she can afford. The equilibrium occurs at the tangency between the budget constraint and the highest attainable indifference curve.
9. Price elasticity depicts the responsiveness of demand to changes in price. It is defined as the ratio of the percentage change in quantity demanded to the percentage change in price. Each other elasticity also represents the ratio of a percentage change in a dependent variable to a percentage change in a given independent variable.
10. The production function describes the relationship of inputs to output. The marginal product of an input is the increase in output due to a one-unit increase in the input holding all others constant. That marginal product tends to decline as more input is added describes the law of diminishing marginal returns.
11. The average total cost curve of a firm shows the total cost per unit of output. The marginal cost curve shows the extra cost required to produce an additional unit of output.
12. The competitive firm in the short run produces that output where price equals marginal cost. The marginal cost curve is therefore the supply curve of the competitive firm.
13. In long-run equilibrium, entry by competing firms forces the typical competitive firm to produce an output level such that its price equals its minimum average cost. At this output, the competitive firm is producing the economically efficient output, and it is earning zero economic profits.
14. The pure monopolist faces the entire downward sloping market demand curve, and this implies that its marginal revenue lies below the demand curve. The monopolist restricts output, by comparison to the competitive case, and it charges a higher market price.
15. The pure monopoly case is one instance of a market in which a welfare loss occurs. A welfare loss, represented by an area under the demand curve and above the marginal cost curve, is an opportunity for mutual gains that is being foregone by the market.

---

## Discussion Questions

1. Explain the difference between cardinal and ordinal utility. Do you think that it is possible for researchers to find out which type of utility people actually have?
2. If a consumer always prefers more to less, can the indifference curves between the two goods be upward sloping? Explain. What if one of the “goods” is actually something unpleasant, like broccoli to a three-year-old?
3. The law of diminishing marginal returns states that eventually the marginal product of an input will tend to fall as more input is added. Describe real-life scenarios, explaining why this is likely to happen. For example, imagine a backyard garden of fixed size and all other inputs except labor also fixed; will adding a worker increase your output? Will adding another increase output by as much? Another?
4. If it makes sense that one type of labor can substitute for another in production, how can capital, a physical object, substitute for labor, a human being?
5. Describe the long-run equilibrium of the competitive firm. Conceptually remove a single assumption of perfect competition and analyze whether and how the process of long-run equilibrium would change. For example, if information were very imperfect, would the long-run equilibrium be achieved? If the firms’ products were not exactly alike? If there were barriers to the entry of new competitors?
6. When a welfare loss occurs because of monopoly, what exactly is lost? Who loses it?
7. Resolve the following: “The price of ice cream increased in the summer, yet quantity also increased. Therefore, the law of demand does not apply to ice cream.”
8. How is a production function affected by the invention of a new process related to it? Can this change result in lower prices to the consumer? What do you think? Do improvements in technological knowledge in the production of consumer goods necessarily reduce average family expenditures?

## Exercises

1. Draw a production possibilities curve for an economy that produces two goods, health and entertainment. Show how this PPF would change if the technology for improving peoples' health was to improve. Show the change if there were an increase in the underlying available quantities of capital and labor.
2. Draw a production possibilities curve between health and all other goods. Insert a point in the drawing that illustrates an economy with an inefficient health system. Insert two additional points that illustrate two efficient economies but two that contrast in their relative emphasis on health care versus all other goods. Is there a cost to society of policies that lead to increases in health care? Explain.
3. Using a supply-and-demand graph and assuming competitive markets, show and explain the effect on equilibrium price and quantity of the following:
  - (a) A technological change that reduces the cost of producing X-rays on the market for physician clinic services.
  - (b) Increased graduations of new doctors on the market for physician services.
  - (c) The virtual elimination of smoking in the population on the market for hospital services.
  - (d) A price ceiling placed on physician fees in the market for physician services.
4. Graph the following demand functions:
  - (a)  $Q_d = 110 - 3.3P$
  - (b)  $Q_d = 100P^{-1.3}$
5. In a graph with OG on the vertical axis and Food on the horizontal, what is the Food-axis intercept? How does the budget constraint shift if the consumer's income level and the two prices all double?
6. Calculate the price elasticity for a \$1 change in price at initial price level \$300 for the demand function  $Q_d = 1,500 - 1.5P$ .
7. What is the slope of the isoquant described by the data in Table 2-5 when evaluated from a labor input of 7 to 8?
8. Assume that a monopoly firm has a linear demand curve and a constant marginal cost curve. Graph this firm's optimal output choice before and after a per-unit excise tax is placed on the output. Does the equilibrium price rise by as much as the tax?
9. Using equation (2.4), what is the demand equation as a function of  $P_s$  if the price of other pastas ( $P_o$ ) is \$2, the individual's income ( $Y$ ) in thousands is \$25, and tastes ( $Z$ ) are represented by 20? What happens if the individual's income increases to \$30?

# Statistical Tools for Health Economics

- Hypothesis Testing
- Difference of Means
- Regression Analysis
- Multiple Regression Analysis
- Statistical Inference in the Sciences and Social Sciences
- Conclusions

The ideas from economic theory must be tested and measured according to the standards of real-world data. Statistical techniques applied to economics are collectively called econometrics. In Chapter 2, we discussed supply and demand, as well as the importance of price and income elasticities. Economic theory predicts that demand curves will slope downward, but it does not predict the degree of responsiveness of demand to price and other variables; it is the task of statistical analysis to estimate these magnitudes. When close substitutes are available for a good, theory predicts that demand will be more sensitive to price than if no close substitutes are available. Yet it is hard to know whether a 1 percent increase in price will decrease the quantity demanded of the good by 10 percent, 1 percent, or 1/10 of a percent, yielding elasticities of  $-10.0$ ,  $-1.0$ , or  $-0.1$ , respectively. Measurements of the economic behaviors of both people and firms may be crucial in analyzing whether drug companies raise drug prices, whether higher insurance copayments will lead people to use less treatment, or whether mandated levels of health care are economically efficient.

This chapter considers statistical methods that econometricians use to draw inferences from data that are collected. Many students with natural science backgrounds are familiar with laboratory experiments, where the environment is held as constant as possible and treatments are administered to experimental groups. The results are then compared to those of untreated control groups. One form of this design is called the dose-response model because the results or responses are generally related to the experimental treatment, or the dose. If statisticians determine the differences are *significant*, a term we will discuss in this chapter, then the dose is considered to be effective.

Social science analysis of human behavior is rarely so fortunate as to find an experimental group that can be matched with a convenient control group. Social scientists, economists among them, must usually collect information from people doing day-to-day activities. Using statistical methods, they try to control for the confounding differences among the people that they are analyzing. The more successful they are in controlling for such differences, the more reliable the analysis will be.

This chapter begins with discussions on how we form hypotheses. It then considers difference of means analysis as a way of introducing statistical inference. Most of the rest of the chapter concentrates on simple regression and multiple regression analyses that are most often used in economic and econometric analysis.

## HYPOTHESIS TESTING

Economists who study health care have been confronted on occasion by statements that, while plausible, demand some validation:

“Men and women don’t smoke the same numbers of cigarettes.”

“Rich people spend more on health care than do poor people.”

“The United States spends more on health care than does the United Kingdom.”

These are all statements that either logic or casual observation would suggest are true. It would be useful, however, to have a rigorous method of determining whether the assertions are correct. Statistical methods suggest formulating these statements as hypotheses and collecting data to determine whether they are correct.

Take, for example, the first assertion about smoking levels. We state clearly both the hypothesis we wish to disprove (the null hypothesis), as well as the hypothesis the theory suggests to be the case (the alternative hypothesis). The null hypothesis here,  $H_0$ , is that men’s levels ( $c_m$ ) equal women’s levels ( $c_w$ ), or

$$H_0 : c_m = c_w \quad (3.1)$$

The alternative hypothesis  $H_1$ , is that  $c_m$  does not equal  $c_w$ :

$$H_1 : c_m \neq c_w \quad (3.2)$$

It is necessary to show convincing evidence that  $c_m$  differs from  $c_w$ . Hypotheses that are designed to test for equality among two or more items are sometimes called *simple* hypotheses.

Consider the second hypothesis, which asserts that rich people spend more on health care than do poor people. If we define health care expenditures of the rich as  $E_r$  and the poor as  $E_p$ , then the null hypothesis is:

$$H_0 : E_r = E_p \quad (3.3)$$

The alternative is:

$$H_1 : E_r > E_p \quad (3.4)$$

In this analysis, it may not be enough just to show that  $E_r$  differs from  $E_p$ . Certainly, even convincing evidence that  $E_p$  is greater than  $E_r$  would not validate the hypothesis. Hypotheses used to test whether two or more items are greater (or less) than each other are sometimes called *composite* hypotheses. Having seen how one might construct the hypotheses in question, we now discuss how to test them.

## DIFFERENCE OF MEANS

Return to the hypothesis about men’s and women’s smoking. Smoking is the single most preventable health risk factor in all societies. People have been smoking for centuries. By the mid-1960s it was clear that smoking had numerous adverse public health impacts. Public health initiatives have sought to eliminate, or at least reduce, people’s smoking.

Smoking depends on many factors. Younger people often start in order to look more mature. People living in certain cultures (often where tobacco is raised) smoke more, and men have



traditionally smoked more than women (although women appear to be catching up). People with more education may better recognize the health impacts of smoking, and hence smoke less. Economists hypothesize that, like most goods, smoking is negatively related to cigarette price (as price increases, smoking decreases). The impact of income is unclear.

To compare men's and women's smoking rates, we can sample the population, yet we know that there are lots of different types of people, and we would like to avoid the confounding influences of age, education, or location. We could attempt to avoid this sort of distortion by drawing samples randomly from among all possible 20-year-olds, called the universe of data. Alternatively, we may try to choose samples of 20-year-old men and women from the same general income group. A sample of college sophomores, for example, from the same location and with similar socioeconomic status, may be a good group for holding many factors constant. Even this example shows how difficult things may be to control. People of the same age (at the same college) may come from different locations and different types of families. Some may have parents or other family members who smoke.

We need a test to determine the differences between two distributions of continuous data. Continuous data are natural measures that in principle could take on different values for each observation. Examples include height, weight, income, or price. Categorical data refer to arbitrary categories such as gender (male or female), race (black, white, or other), or location (urban or rural). In this chapter, unless we specify categorical data, our methods will refer exclusively to the analysis of continuous data.

Lots of people do not smoke, so let us concentrate only on smokers (looking at the decision whether or not to smoke is an important policy question, but is far beyond the scope of this example). The econometrician asks one woman and finds that she smokes 10 cigarettes per day ( $c_w = 10$ ). The first man asked smokes 15 cigarettes per day ( $c_m = 15$ ). This provides evidence that men smoke more than women, because  $c_m > c_w$ , or  $15 > 10$ . It is not very convincing evidence, however. The man, or woman, or either, may not be typical of the entire group. What if a different man and/or woman had been selected? Would the answer have been different?

It seems logical to test several men and to compute the mean or average level by summing the levels and dividing them by the total number of men tested. The National Institutes of Health (in 2001 and 2002) collected a database of over 43,000 individuals called the National Epidemiologic Survey on Alcohol and Related Conditions, or NESARC. They focused on potentially substance abusive activities including smoking, drinking of alcoholic beverages, and the taking of recreational (and harder) drugs. They asked a number of questions about smoking, and from the analysis of smokers the textbook authors found that

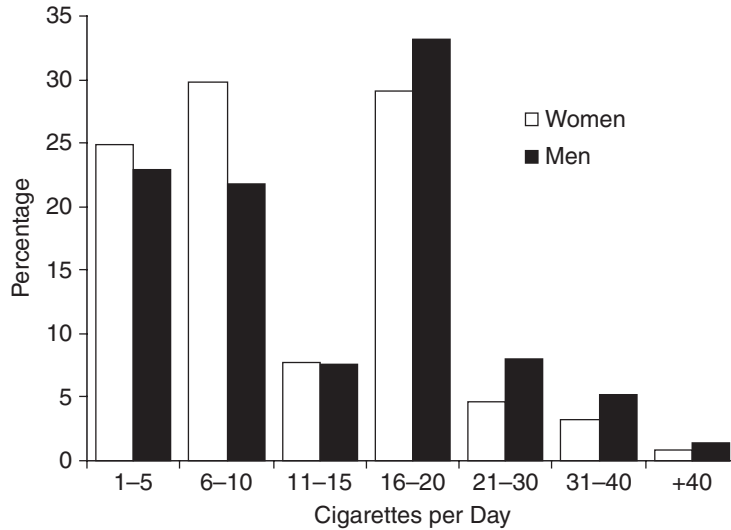
For 4,714 men, the mean, or average level,  $\bar{c}_m$ , was 15.60 cigarettes per day.

For 4,841 women, the average level,  $\bar{c}_w$ , was 13.47 cigarettes per day.

The difference,  $d = \bar{c}_m - \bar{c}_w$ , then, is 2.13 cigarettes per day.

### The Variance of a Distribution

Although a difference of the two means is improved evidence, the econometrician desires a more rigorous criterion. It could be that the true level for both men and women is 14 cigarettes per day, but our sample randomly drew a higher average level for men (15.60) than for women (13.47). Figure 3-1 plots the distributions in percentage terms. Almost 25 percent of the women smoke between 1 and 5 cigarettes per day, compared to a slightly smaller percentage of men; in contrast, while about 28 percent of the women smoke between 16 and 20 cigarettes per day, about 33 percent of the men smoke at this level. Although the mean levels differ (men are higher than women), some groups of women (those who smoke less) have higher percentages than some groups of men. Statisticians have found the variance of a distribution to be a useful way to summarize its dispersion. To calculate the variance of women's levels, we subtract each observation from the mean



**FIGURE 3-1** Cigarette Consumption by Gender, 2001–2002

(13.47), square that term, sum the total, and divide that total by the number of observations,  $N$ . Hence, variance,  $V_w$ , equals:

$$V_w = \frac{N_1x(1 - 13.47)^2 + N_2x(2 - 13.47)^2 + \dots + N_{80}x(80 - 13.47)^2}{4,841} \tag{3.5}$$

Here  $N_1$  is the number of women who smoke 1 cigarette per day,  $N_2$  is the number who smoke 2 cigarettes per day, and so on (and yes, there are some women who smoke 80 cigarettes per day!).

$V_w$  reflects the variance of any individual term in the distribution. If  $V$  is large, then the dispersion around the mean is wide and another woman tested might be far from our mean. If  $V$  is small, then the dispersion around the mean is narrow and another observation might be close to the mean.

### Standard Error of the Mean

The variance is often deflated by taking the square root to get the standard deviation,  $s$ , yielding:

$$s_w = \sqrt{\frac{N_1x(1 - 13.47)^2 + N_2x(2 - 13.47)^2 + \dots + N_{80}x(80 - 13.47)^2}{4,841}} \tag{3.6}$$

As with  $V$ , a large (small) value of  $s$  indicates a large (small) dispersion around the mean. Statisticians have shown that we can calculate the standard error of the mean itself by dividing  $s$  by the square root of the number of observations. In this sample, the standard deviation of the distribution for women equals 9.71. The standard error of the mean of the women’s distribution would then equal  $s_w$  divided by the square root of 4,841, or  $(9.71 \div 69.6)$ , which equals 0.14.<sup>1</sup>

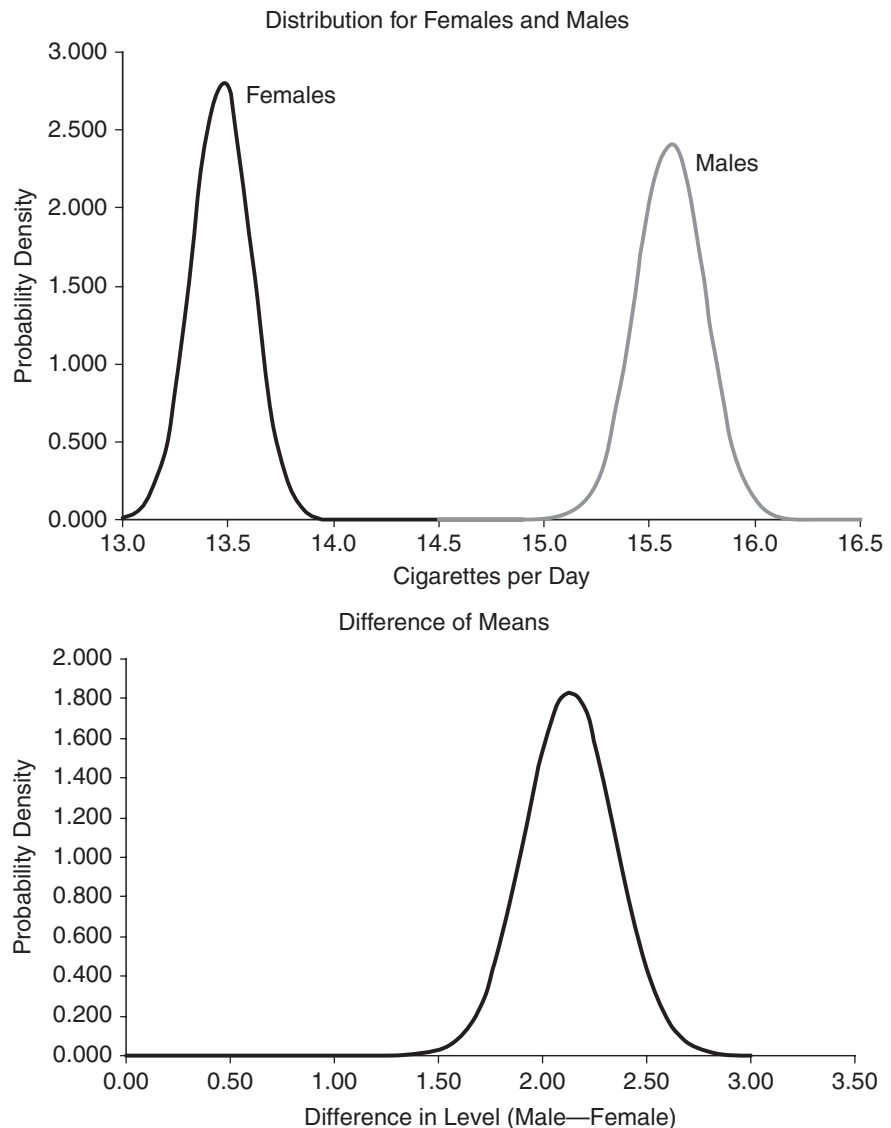
A powerful theorem in statistics, the Central Limit Theorem, states that no matter what the underlying distribution, the *means* of that distribution are distributed like a normal, or bell-shaped, curve. Hence, we can plot the normal distribution of means of women’s levels with a mean of 13.47 and a standard error of 0.14.

<sup>1</sup> Formally, in a sample (as opposed to the entire population), we calculate the standard error by dividing by  $n - 1$ . All calculations are rounded to the nearest hundredth.

Statisticians have also shown that a little more than 68 percent of the area under the curve would be within one standard error, or between levels of 13.33 (that is,  $13.47 - 0.14$ ) and 13.61 (i.e.,  $13.47 + 0.14$ ), and that 95.4 percent would be within two standard errors. This means that we could be about 95 percent sure that the true mean quantity of cigarettes smoked for women was between 13.19, [that is,  $13.47 - (2 \times 0.14)$ ], and 13.75, [i.e.,  $13.47 + (2 \times 0.14)$ ]. A similar calculation can be done for men, yielding a similar measurement. Intuitively, the further apart the means and the smaller the dispersions (standard errors), the more likely we are to determine that the average level for men is smaller than that for women. To test the hypothesis formally, we then construct a “difference of means” test. We wish to compare the measurement  $d = \bar{c}_m - \bar{c}_w$ , to zero, which was the original hypothesis.

Here  $d = 2.13$ . The variance of the difference is defined as the sum of the variances of the standard errors. If the standard error for women was 0.14 as we calculated it and the standard error for men given the sample in Figure 3-2A was 0.17, then the standard error of the difference would be:

$$s_d = \sqrt{0.14^2 + 0.17^2} = 0.216 \quad (3.7)$$



**FIGURE 3-2** Cigarettes per Day

The difference and its distribution also can be plotted.

The most probable value of the difference, as noted in Figure 3-2B, is 2.13. About 68 percent of the distribution lies between 1.91 (i.e.,  $2.13 - 1 \times 0.216$ ) and 2.35 (i.e.,  $2.13 + 1 \times 0.216$ ). About 95.4 percent of the distribution lies between 1.69 (i.e.,  $2.13 - 2 \times 0.216$ ) and 2.56 (i.e.,  $2.13 + 2 \times 0.216$ ).

This experiment would find very good evidence that among smokers, women smoke fewer cigarettes than men. The males have higher levels than the females, and the probability is well over 95 percent that this difference is statistically significant.

Alternatively, the  $t$ -statistic, comparing the numbers 2.13 and 0.0, equals  $2.13 \div 0.216$ , or approximately 9.86. Statisticians calculate tables of  $t$ -statistics, whose critical values are related to the size of the sample. With a sample of over 8,000, a  $t$ -statistic of nearly 10 is statistically significant at well over the 99 percent level. In other words, we can be 99+ percent certain that men smoke more cigarettes than women.

## Hypotheses and Inferences

This process illustrates the steps that are necessary to test hypotheses appropriately. The econometrician must:

1. State the hypotheses clearly

$$H_0: c_m = c_w, \text{ against}$$

$$H_1: c_m \neq c_w.$$

2. Choose a sample that is suitable to the task of testing.
3. Calculate the appropriate measures of central tendency and dispersion: the mean and the standard error of the mean for both men and women, leading to the difference of the two means.
4. Draw the appropriate inferences: men smoke more than women.

No matter how sophisticated the method used, good statistical analysis depends on the ability to address these four criteria and stands (or falls) on the success in fulfilling them. Box 3-1 provides a particularly good example of how analysts have examined the impacts of electromagnetic fields (EMFs) on children living near power lines.

There are, of course, measures of central tendency other than the mean (or average). Someone who smokes 4 packs per day (80 cigarettes) may unduly influence the mean. A different measure, the median, calculates a statistic such that half of the observations are greater than the median and half are less. Thus, a median smoking level of 15 cigarettes would imply that half of the people smoked more than 15, and half smoked fewer. The median is less sensitive to extreme values in the data (e.g., someone who smokes 60 cigarettes per day would have no more effect on a median of 15 than does someone who smokes 20). However, the median can present mathematical problems in hypothesis testing. Simple formulas for standard errors of medians have not been available, although popular numerical “bootstrapping” methods now provide intuitive and accurate standard errors. For a good discussion of bootstrapping, see Efron and Tibshirani (1993).

### BOX 3-1

#### No Link Between Childhood Cancer and Electromagnetic Fields

In 1979, Dr. Nancy Wertheimer and her assistant, Ed Leeper, reported that children who lived near power lines had twice the normal incidence of leukemia. The study was criticized because it was small and relied on indirect evidence rather than direct measurements of exposure to electromagnetic fields (EMFs). However, the report and its findings had major impacts. Parents of children with cancer sued power companies. Owners near power lines saw reductions in the values of their homes.

(continued)

Scientists began to examine the issue. While the studies tended to be small, and sometimes inferred exposure many years after the children had developed cancer, some did find associations between power lines and cancer.

In 1997, a collaborative study between scientists at the National Cancer Institute and specialists from the nation's leading medical centers reported that there was no evidence showing that electric power lines cause leukemia in children. The study compared 636 children with acute lymphoblastic leukemia, the most common childhood cancer, to 620 healthy children who were matched to the cancer patients by race, age, and residential neighborhood. The researchers asked, "Did the children who developed leukemia have a higher exposure to magnetic fields produced by power lines?"

To answer the question, the researchers assessed the children's exposure to magnetic fields produced by power lines in their current and past residences, as well as where the mothers lived when they were pregnant. They measured the fields in the children's yards, schools, and day care centers and also had the children walk around with measurement devices. Investigators did not know which children had had cancer.

The research team, headed by Dr. Martha S. Linet of the National Cancer Institute, found that the children living in houses with the highest EMFs were about 24 percent more likely to develop leukemia than the children in houses with the lowest EMFs. However, this difference was not statistically significant. Nonetheless, the fact that positive relationships exist over some range of the data is often misinterpreted as proof that definitive relationships exist.

Dr. Linet's group tested the hypothesis by relating the differences in means among groups, which were relatively small, to their variation, which was relatively large. This indicated that their measured effect was probably due to chance, rather than reflecting systematic differences. Their statistical test implied that they could not confidently reject the null hypothesis—that the true rates of leukemia were equal and unrelated to the level of EMFs.

*Source:* Martha S. Linet, Elizabeth E. Hatch, Ruth A. Kleinerman, et al., "Residential Magnetic Field Exposures and Childhood Acute Lymphoblastic Leukemia," *New England Journal of Medicine* 337 (July 3, 1997): 1–7.

## REGRESSION ANALYSIS

The difference of means analysis is extremely useful in treating continuous data that can be broken up by categories, such as gender, race, or location.<sup>2</sup> Yet many interesting economic variables occur naturally as continuous variables. Health care expenses, physician visits, firm profits, as well as prices and/or incomes could take large numbers of values naturally, and we group them into categories only with serious loss of information. If we have information on income, in terms of dollars per year, we can distinguish among households with incomes of \$10,000, \$20,000, \$30,000, and so on. If we were to define high income as greater than \$40,000, for example, separating all of the different incomes into two categories, we would then have no way of distinguishing between households with incomes of \$10,000 and \$20,000 (or, for that matter, between households with incomes of \$50,000 and \$100,000).

Regression analysis allows the econometrician to fit a straight line through a set of data points. How might this be used for policy analyses? Economists and health policy makers alike have long sought to determine how responsive cigarette consumption is to excise taxes. Recognizing that cigarette smoking is a dangerous habit, economists have reasoned that if the price is raised by taxes, consumption is likely to decline, but the question is "how much?"

Cigarettes are produced nationally and are subject to the same federal taxes but, as of 2010, excise tax rates varied from \$0.07 per pack in South Carolina to \$3.46 per pack in Rhode Island. The NESARC smokers, in addition to number of cigarettes smoked, were also asked a wide range of questions focusing on addictive behaviors, including smoking, drinking, legal and illegal drugs, as well as numerous questions about occupation, education, and income. Because the state of residence was known, an analyst could append readily available data on state level excise taxes.

<sup>2</sup> Although difference of means considers only two categories, *analysis of variance* methods allow the consideration of three or more categories. Newbold, Carlson, and Thorne (2007) present good discussions on this and other statistical topics.

Suppose that we wish to relate the amount of cigarettes smoked per day to the tax price of the cigarettes. Since cigarettes must be purchased in order to be consumed, we would like to know how responsive the quantity demanded is to the tax price (the tax price elasticity). Recall that the price elasticity relates the percent change in quantity to the percent change in price. It would be useful to draw a straight line that would summarize the relationship.

### Ordinary Least Squares (OLS) Regressions

Two rules are commonly used to determine this line. The first rule is that the deviations (unless the line fits perfectly) from the line must sum to zero. Positive deviations must be offset by negative deviations. We can show, however (see Figure 3-3), that many lines have this characteristic (for example, dashed lines  $R_1$  and  $R_2$ ). It is necessary to have a second criterion by which to distinguish among the large number of lines where the sum of the deviations equals zero.

The second criterion is to minimize the sum of the squared deviations of the actual data points from the line that is fitted. Even though the sum of the deviations equals zero, the sum of the squared deviations must be positive (any number multiplied by itself is either zero or positive). Hence, one can choose among the many lines that have sums of zero deviations by picking the one line with the minimum or least sum of the squared deviations. Such analyses are called ordinary least squares (OLS) analyses.

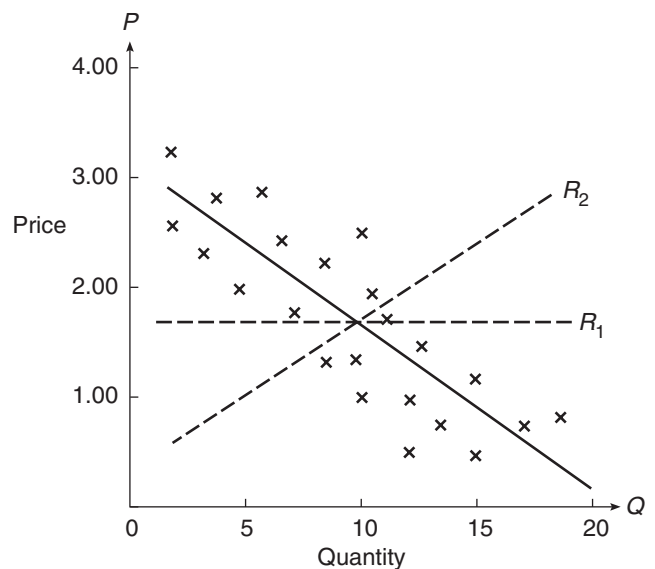
The resulting equation would have the following form:

$$Q = a + bP + \varepsilon \quad (3.8)$$

where  $P$  and  $Q$  refer to price and quantity, and  $a$  and  $b$  are the parameters to be estimated. Parameter  $a$  is sometimes referred to as the constant, or the intercept. It might refer to the demand for  $Q$  in the unlikely event that the tax price of cigarettes was zero.

Parameter  $b$  refers to the slope of the line and shows the direction and magnitude of the impact of a change in  $P$  on the quantity demanded. Because we expect a higher level of  $P$  to decrease the amount of cigarettes purchased (assuming that the subjects purchase the cigarettes that they smoke), we expect  $b$  to be negative.

The last parameter is the error term  $\varepsilon$ . No regression analysis will fit the data exactly. Errors are likely and may reflect several causes. We may have omitted a variable, such as age (older people may smoke more) or education (more educated people may be more aware of the dangers of smoking and smoke less). We may have measured one or more of the explanatory variables or the



**FIGURE 3-3** Plot of Quantity Purchased Against Price

dependent variable (the amount of cigarettes) inaccurately. All of these may stand in the way of our predicting the amount demanded exactly. In advanced econometric work, understanding  $\varepsilon$  is crucial for ensuring that the estimated parameters are accurate. Our exposition here will assume that  $\varepsilon$  obeys the rules to allow us to make appropriate inferences with OLS analyses. We will examine some exceptions later in the chapter.

### A Demand Regression

Table 3-1a shows the result of a simple regression of cigarette consumption against the tax price of cigarettes.

$$Q = 16.83 - 3.24 \times \text{tax per pack}, R^2 = 0.01 \quad (3.9)$$

(0.34)

This equation indicates that a \$1 increase in the tax per pack of cigarettes is correlated with a decrease in quantity demanded of 3.24 fewer cigarettes per day among those who smoke. The *standard error of estimate* for the coefficient of tax is 0.34. This term is similar to the standard error of the estimated mean in the example of cigarette smoking in men and women earlier in the chapter. As before, the smaller the standard error is relative to the estimated value of  $b$  (in this case,  $-3.24$ ), the better the estimate. In this regression, the standard error of 0.34 is relatively small compared to the coefficient,  $-3.24$ ; hence, the coefficient is significantly different from zero. The expression  $R^2$  is used to measure the fraction of the variation of the quantity of cigarettes explained by the price alone. An  $R^2$  of 0.01 implies that about 1 percent of the variation was explained.

It is useful to examine this simple regression in detail because it has many features that occur in more complex analyses. Consider the following hypothesis:

$H_0$ : Tax price doesn't matter; that is,  $b = 0$  against the alternative hypothesis, which is:

$H_1$ : Tax price is inversely related to quantity consumed; that is,  $b < 0$ . The test of the hypothesis is similar to a difference of means test. In particular, we are testing the difference between  $-3.24$  (estimated with standard error 0.34) and 0.

**TABLE 3-1 Excise Taxes and Cigarette Demand**

Variable	a - Simple			b - Multiple			c - Interactive		
	Coefficient	Std Error	t-stat	Coefficient	Std Error	t-stat	Coefficient	Std Error	t-stat
Intercept	16.83	0.19	86.78	17.22	0.63	27.34	17.15	0.63	27.19
Excise Tax	-3.24	0.34	-9.42	-2.28	0.33	-6.96	-2.29	0.33	-7.00
Income (\$ × 1,000)				-0.0020	0.0025	-0.80	-0.0021	0.0025	-0.85
Male				2.23	0.21	10.68	2.38	0.22	10.80
African American (AA)				-5.05	0.34	-15.04	-4.29	0.49	-8.77
Age				0.13	0.01	19.11	0.13	0.01	19.10
Educational Level				-0.67	0.05	-12.42	-0.67	0.05	-12.41
Hispanic				-6.50	0.37	-17.55	-6.51	0.37	-17.58
AA + Male							-1.43	0.67	-2.14
$R^2$	0.0092			0.1132			0.1136		
Elasticity	-0.0989			-0.0697			-0.0701		
$N$	9,555			9,555			9,555		

Source: Computations from National Epidemiologic Survey on Alcohol and Related Conditions (NESARC) by A. Goodman, 2008.

Remember that because demand is downward sloping, the coefficient will be negative. The  $t$ -statistic here is 9.5; that is, the value of the coefficient, 3.24, divided by the standard error of 0.34. The value of 9.5 suggests that we can be more than 99 percent sure that the tax price has an effect on quantity of cigarettes consumed. This term is statistically significant in its difference from zero.

If 1 percent of the variation of the quantity of cigarette demand is explained, then 99 percent is unexplained. In part, this occurs because the regression does not include some variables that are likely to be important. We have noted earlier that several other variables might help explain the consumption of cigarettes. If included in the analysis, we are likely to explain more of the variation in cigarette consumption. The inclusion of more variables in a multiple regression is explained later.

This example illustrates cross-sectional analysis, which provides snapshots of a slice of the population at one period in time. Because 2001–2002 was the first time that the NESARC was collected, it could not yet follow the people in the sample over time, and we could not be aware of continuing health problems, changes in wealth or income, or systematic differences in ability that cannot be measured, and that cross-sectional models treat as “random noise.” As a result, cross-sectional regressions often explain less variance than panel data, which follow households over time, or time-series data, which calculate aggregates over time.

### Estimating Elasticities

Regressions also are used to estimate elasticities. Recall that the definition of the price elasticity of demand ( $E_p$ ) is the percentage change in quantity demanded, elicited by a 1 percent change in price. This is written as

$$E_p = \frac{\% \text{ change in quantity}}{\% \text{ change in price}} = \frac{\left(\frac{\Delta Q}{Q}\right)}{\left(\frac{\Delta P}{P}\right)}, \text{ or } E_p = \left(\frac{\Delta Q}{\Delta P}\right)\left(\frac{P}{Q}\right)$$

The term  $\Delta P$  stands for a one-unit change in the price, while the term  $\Delta Q$  represents the resulting change in the quantity demanded. It follows that  $\Delta Q/Q$  is a measure of the percentage change in quantity, whereas  $\Delta P/P$  is a measure of the percentage change in price. In rearranging terms at the right, the term  $\Delta Q/\Delta P$  represents the ratio of changes and is the reciprocal of the slope of the demand curve. With the linear function here, this is  $-3.24$ .

In calculating an elasticity from the coefficients estimated in a regression, a different elasticity could be calculated for each different starting price that is assumed. Therefore, it is also necessary to have reference values for  $P$  and  $Q$ , and the mean (or average) values are often used. In our sample, the mean number of cigarettes smoked per day is 15.3 (about three-quarters of a pack), and the mean tax price is \$0.454 (about \$0.45 per pack).

Hence, calculated at the mean,

$$E_p = 3.24 \times (0.45 \div 15.3), \text{ or } -0.10 \quad (3.10)$$

This says that a 10 percent increase in the tax price of cigarettes would lead to a 1.0 percent decrease in quantity demanded. Does this make sense? Cigarettes, after all, are an addictive substance, and many people find it difficult to reduce their demand. Yet, there is clearly a negative effect, and one could also argue that doubling the tax (say from \$.50 per pack to \$1 per pack—a 100 percent increase) could reduce demand by 10 percent, a sizable amount.<sup>3</sup>

<sup>3</sup> This regression (and others in the chapter), and the estimated elasticities are provided primarily to illustrate how to read and use regression analyses. For a number of reasons they probably represent short-term responsiveness to price changes (we do not examine those who either start or stop smoking, and we do not see individuals over time, both of which could provide a long-run elasticity). For more development of cigarette issues, see Chapter 24, “The Economics of Bads.”



## MULTIPLE REGRESSION ANALYSIS

Real-world relationships are seldom two-dimensional, as useful as this situation would be in drawing graphs. As noted, demand for cigarettes may be related not only to the price, but to income,  $Y$ . Older people (variable  $A$  for age) may smoke more, having been addicted for longer, and more educated people,  $E$ , may recognize the dangers of smoking and smoke less. In addition, women have traditionally smoked less, and various groups, variable  $G$ , may have differing tastes toward smoking cigarettes. Indeed, the omission of important variables may lead to particular behavior in the error term,  $\varepsilon$ .

If each relationship could be graphed, assuming that nothing else was changing, then simple regression would work fine. Fortunately, the mathematics necessary to estimate the appropriate relationship can accommodate more than two dimensions. It is easy to write the following multiple regression:

$$Q = a + bP + cY + dA + eE + fG + \varepsilon \quad (3.11)$$

Although the example presented in Table 3-1 will summarize eight dimensions now rather than two dimensions, we use exactly the same least-squares criteria as before. The interpretation of the coefficients is similar to before but is done with more confidence. With the simple regression, relating  $Q$  only to  $P$ , the econometrician would not know whether income,  $Y$ , or age, or education, was varying as well. Including them in this regression allows us to “hold constant” these other variables and reduce the error. As a result, elasticities can now be calculated under the condition that “all else is equal.” The  $R^2$  measure of variation explained earlier also is available here.

$R^2$  will always rise with more variables. (If you add variables, you can’t explain less of the variation!) Several methods can be used to interpret  $R^2$ , and some statisticians wish to maximize  $R^2$ ; that is, to explain as much variation as possible.<sup>4</sup> While this may be desirable, most econometricians are at least as interested in the values of the parameters that are estimated.

### Interpreting Regression Coefficients

Table 3-1 shows both the original simple regression (a), and more complex multiple regressions (b) and (c) with standard errors of coefficients and  $t$ -statistics also included. Multiple regression (b) shows that a \$1 increase in the tax price of cigarettes,  $P$ , implies a decrease in quantity of tax demanded of 2.28 cigarettes per day. Income, measured in thousands of dollars per year, has a slightly negative effect, but that effect does not differ statistically from zero. Older people smoke slightly more, and more educated people smoke slightly less. Women, African-Americans, and Hispanics all smoke significantly fewer cigarettes than do white males.

As was done with the simple regression, predicted values of the dependent variable and values of the elasticities can be computed. As before, hypotheses can be tested. The addition of more variables leads to a fall in the price elasticity from about -0.10 to about -0.07, but it is still statistically significant from 0.

Most often, again, econometricians are interested in whether coefficients are positive or negative and whether they differ significantly from zero. In the cigarette example, excise tax increases clearly accompanied decreases in cigarette consumption. In a now famous example, Box 3-2 presents the surprising results of a detailed multivariate analysis looking at the impacts of hormone replacement therapy on postmenopausal women.

<sup>4</sup> Often,  $R^2$  is adjusted for the number of explanatory variables and the number of observations, giving:

$$1 - \bar{R}^2 = \frac{n - 1}{n - k - 1} (1 - R^2)$$

where  $n$  is the number of observations and  $k$  is the number of explanatory variables. Adding more variables,  $k$  will always raise  $R^2$  but it will not necessarily raise  $\bar{R}^2$ , termed the “adjusted  $R^2$ ”.

**BOX 3-2****Hormone Replacement Therapy—Rigorous Statistics Reveal Surprising Results**

As of July 2002, approximately 38 percent of postmenopausal women in the United States used hormone replacement therapy (HRT). While the U.S. Food and Drug Administration-approved indications for HRT included relief of menopausal symptoms (hot flashes, night sweats, and vaginal dryness) and prevention of osteoporosis, long-term use had been common to prevent a range of chronic conditions, especially heart disease. Advertisements by drug companies urged women to take HRT so they would stay “forever feminine.”

Although the drugs were widely used, many scientists had expressed concern that studies finding benefits of HRT were based on nonrandom samples of women who sought out the hormone therapy. The users of HRT were better educated and healthier than postmenopausal women who did not take HRT. Thus, some researchers felt that “selection bias” could account for the effectiveness of HRT because those women for whom it was not effective and those who found the side effects bothersome or harmful, as well as less educated and sicker women, were not included in the studies.

Between 1993 and 1998, a randomized clinical trial called the Women’s Health Initiative (WHI) studied 16,608 postmenopausal women aged 50 to 79. Roughly half of the participants were randomly assigned to the experimental group and received HRT, a daily tablet containing conjugated equine estrogen and medroxy progesterone acetate (progestin). The other half was randomly assigned to the control group and received a placebo (an inert pill with no medical properties). Study participants were contacted by telephone six weeks after randomization to assess symptoms and reinforce adherence. Follow-up for clinical events occurred every six months with annual in-clinic visits required.

A multitude of health outcomes related to cardiovascular disease, stroke, cancer, fractures, and death were measured. The statistical analysis was complex and compared the health outcomes for the experimental group who took the estrogen/progestin tablet to the control group who took the placebo. Formal monitoring began in the fall of 1997, with the expectation of final analysis in 2005 after an average of approximately 8.5 years of follow-up. An independent data and safety monitoring board (DSMB) examined interim results to determine whether the trial should be stopped early, in particular if the treatment proved either significantly beneficial or harmful to the experimental group, relative to the control group.

By May 2002, an average of 5.2 years into the analysis (recall that women had entered the study over a 5-year period), the DSMB determined that there were significantly higher risks of breast cancer, coronary heart disease, stroke, and pulmonary embolism in the experimental group, and that these increased probabilities outweighed some evidence of reduced risk of fractures and colon cancer. Therefore, the DSMB recommended an early stopping of the estrogen plus progestin component of the trial because it would be unethical to put more women at risk for adverse events by continuing HRT. The results were released in the *Journal of the American Medical Association*, and due to the sheer number of women being treated and the surprising findings, the results made front-page headlines and led off the nightly news for several weeks. A treatment that had been taken for granted as beneficial in reducing menopausal symptoms and preventing heart disease had been subjected to a carefully designed randomized clinical trial. The results sent shock waves that reverberate through the medical system even a decade later. Yet further studies have verified the original findings that whatever benefits hormone replacement may provide must be balanced against its demonstrable risks.

*Sources:* Writing Group for the Women’s Health Initiative Investigators “Risks and Benefits of Estrogen Plus Progestin in Healthy Postmenopausal Women,” *Journal of the American Medical Association* 288 (3) (2002): 321–333; Suzanne W. Fletcher and Graham A. Colditz, “Failure of Estrogen Plus Progestin Therapy for Prevention,” *Journal of the American Medical Association* 288 (3) (2002): 366–368.

**Dummy Variables**

In health care research, econometricians often are interested in whether particular groups of patients or subjects differ from others. For example, Table 3-1 denoted men, African Americans, and Hispanics by using 1 if the person was a member of such a group and zero otherwise. These groups were indicated by using binary, or *dummy*, variables. In other circumstances, econometricians may wish to indicate whether research subjects are white (white = 1), or not (white = 0), or whether the subject is a woman (female = 1) or not (female = 0).

**FIGURE 3-4** Interpreting the Effects of Dummy Variables

In Table 3-1 (b), coefficient  $b_a = -5.05$ , and coefficient  $b_m = +2.23$ . We see that African American women (“northeast box”) smoke 5.05 fewer cigarettes than white women. White men smoke 2.23 more cigarettes than white women. Finally, African American men smoke  $2.82 = (-5.05 + 2.23)$  fewer cigarettes per day than white women.

		African American, $b_a$	
		0 = No	1 = Yes
Male, $b_m$	0 = No	0	$b_a = -5.05$
	1 = Yes	$b_m = +2.23$	$b_a + b_m = -5.05 + 2.23 = -2.82$

Figure 3-4 shows how the estimated coefficients of race and gender could be treated, using the regression from Table 3-1 (b). The northwest box indicates an observation for which both male and African American equal 0. If male = 1, then the coefficient  $b_m$  for both lower boxes indicates whether the person smokes more or less (and whether this is significant). This is a row effect. If African American = 1, then the coefficient  $b_a$  indicates whether white households purchase more or fewer cigarettes (and whether that is significant). This is a column effect. If the household is both white and female-headed, then the combined effect is  $(b_m + b_a)$ .

In fact, if one felt that black males may have particular attitudes or preferences for smoking, one could estimate a coefficient that addresses the interaction of the two effects (1 if African American and male, 0 otherwise). This would be coefficient  $b_{am}$ . Here, the impact of being black and male would be  $(b_a + b_m + b_{am})$ . This is noted in Table 3-1 (c). Compared with white women (the “northwest” quadrant of the diagram), white men smoke 2.38 more cigarettes, African American women smoke 4.29 fewer cigarettes, and African American men smoke  $(2.38 - 4.29 - 1.43)$ , or 3.34 fewer cigarettes.

## STATISTICAL INFERENCE IN THE SCIENCES AND SOCIAL SCIENCES

Natural scientists attempt, not always successfully, to control experimentally for all of the other possible sorts of variation other than the relation being studied. By contrast, econometricians are seldom so fortunate. Occasionally, experimental economic studies are done, but such projects are expensive. One such study was the multimillion-dollar health insurance experiment conducted by the RAND Corporation in the late 1970s and early 1980s, funded by the federal government, and we discuss parts of that study in several later sections of the book. (There are excellent reviews by Manning and collaborators, 1987, and Newhouse and collaborators, 1993.) Even with the careful planning that went into the experimental design, this study could not avoid some major analytical issues.

Other fields have similar problems. A 1988 report from the Panel of the Institute of Mathematical Statistics referred to analytical problems in chemistry:

The data are frequently complex with a large number of dimensions, may sometimes have a time element, and can be further complicated because of missing values. In some instances, standard multivariate or time-series methods may suffice for analysis, but, more commonly, novel developments are required, for example, to handle the problem of multivariate calibration (Olkin and Sacks, 1988, p. II-1).

Econometricians must most often use natural experiments and must seek ways to account for the other variations. Because many policies, such as the provision of public health services or the regulation of the prescription drug industry, depend on accurate measurement of economic phenomena, it is essential that the measurements be accomplished carefully and scientifically.

## CONCLUSIONS

This chapter has provided a “taste” of the statistical methods necessary to address questions that occur in health economics and to clarify the analyses where statistical material is presented later in the text. To understand the text, it is important to be able to formulate questions in terms of hypotheses, read statistical test results to determine if the result is significant, understand statistical significance, and interpret reported regression results. The emphasis on problems to watch for in statistical analysis is not meant to generate undue skepticism over the statistical data to be reported. On the contrary, the discussion is meant to help distinguish the better studies where confidence can best be placed.

---

## Summary

1. Economists usually must collect information from people doing day-to-day activities and use statistical methods to control for the confounding differences among the people that they are analyzing. The more successful they are in controlling for such differences, the more reliable the analysis will be.
2. Statistical methods suggest formulating economic assertions as hypotheses, and collecting data to determine whether the hypotheses are correct.
3. Hypotheses that test for equality among two or more items are called *simple* hypotheses. Hypotheses that test whether two or more items are greater (or less) than each other are called *composite* hypotheses.
4. Several steps are necessary to test hypotheses appropriately. The econometrician must:
  - state the hypothesis clearly,
  - choose a sample that is suitable to the task of testing,
  - calculate the appropriate measures of central tendency and dispersion, and
  - draw the appropriate inferences.
5. Regression analysis allows the econometrician to fit a straight line through a set of data points. In ordinary least-squares regression, the sum of the squared deviations of the actual data points from the line is minimized.
6.  $R^2$  measures the proportion of the total variation explained by the regression model. While it may be desirable to maximize  $R^2$ , most econometricians are at least as interested in the values of the parameters that are estimated.
7. Important skills in statistical analysis include:
  - understanding statistical significance,
  - interpreting reported statistical results, and
  - detecting problems in reported statistical findings.

---

## Discussion Questions

1. List at least three ways in which natural experiments differ from laboratory experiments.
2. What is the difference between a simple hypothesis and a composite hypothesis? Why might economists choose one over another?
3. In considering the difference in smoking between men and women, what is the null hypothesis? What is the alternative hypothesis? Is the alternative hypothesis simple or composite?
4. Suppose that we wish to compare the health status of two groups of people. What variable might we use to measure the status? What variables might we wish to control in order to draw the appropriate inferences?
5. If someone reports that the mean weight for fourth-grade boys is 80 pounds and for fourth-grade girls is 78 pounds, what must you know to test hypotheses using the difference of means?
6. If we are trying to relate output to labor inputs and capital inputs using regression analysis, would we expect the coefficients of the regressions to be positive or negative? Why?
7. What are dummy variables? How are they useful in identifying differences among groups?
8. Suppose that you used regression methods to estimate the demand curve for physician visits and found a positive relationship; that is, you found that the higher the price is, the more visits are demanded. What problem has likely arisen? Explain the problem in words. Why might it make statistical inference difficult?
9. Rich people consume more health care services than poor people. Explain two ways one might test this hypothesis.

## Exercises

(For students with access to spreadsheet computer programs.) Consider the following data for a cross section of individuals in the population, in which

$Q$  = Quantity (in 100s) of aspirin purchased in a year

$P$  = Average price of aspirin in that year

$Y$  = Annual income

$A$  = Age of buyer

Observation	$Q$	$P$	$Y$	$A$
1	1	1.5	20	25
2	2	1.5	40	20
3	4	1	12	25
4	2	1	10	30
5	2	1	8	30
6	3	2	30	35
7	3.5	1.5	30	40
8	4	2	20	40
9	7	1	20	45
10	1	3	15	40
11	2	2	18	30
12	3	2	20	32
13	3.5	2	15	36
14	4	2	10	30
15	2	3	25	20

Now consider questions 1 to 4:

1. If we divide the population into two groups, up to age 35 and over age 35, which group purchases more aspirin?
2. Divide the population into three groups—up to age 30, over 30 and up to 45, and over 45. Do the purchases vary by age?
3. What is the relationship in a regression analysis between  $Q$  and  $P$ ? Between  $Q$  and  $Y$ ? Between  $Q$  and  $A$ ?

Observation	$Q$	$P$	$Y$	$A$
16	1	4	15	25
17	8	2	15	55
18	9	1	40	50
19	1	4	10	45
20	10	1.5	30	55
21	6	1.5	35	60
22	2	1	30	40
23	3	1	25	40
24	3	2	20	35
25	3	2	15	35
26	4	3	20	35
27	1	4	20	25
28	1	4	25	30
29	2	5	28	30
30	3	1	30	32

4. Calculate the multiple regression that relates  $Q$  with  $P$ ,  $Y$ , and  $A$ . Which variables are statistically significant? What is the elasticity of  $Q$  with respect to  $P$ , to  $Y$ , and/or to  $A$ ?
5. From Table 3-1, column b, suppose income is \$20,000, the excise tax on cigarettes is \$1, and the person is a 40-year-old white, non-Hispanic male who completed high school (education level = 9). Calculate the elasticities of demand for aspirin with respect to excise tax, income, and age.
6. Consider demand curves for aspirin, estimated for two different sets of consumers:
  - (a)  $Q = 20 - 5P + 0.2Y$
  - (b)  $Q = 30 - 5P + 0.2Y$

If  $Y = \$20$  and  $P = \$1$ , calculate the price and income elasticities for group (a) and group (b). Whose elasticities will be higher? Why?

7. Given the regression estimate of the demand equation of

$$Q_x = 1,000 - 3.3P_x + 0.001Y$$

where  $Y$  is income, what is the change in demand if price rises by \$1, holding income constant? What is the percentage change in demand if price rises by \$1 from an initial price of  $P_x = \$200$  given  $Y = 10,000$ ? What is the effect on demand of a \$1 increase in income, holding price constant?

8. Consider the estimate demand equation of

$$Q_x = 1,000 - 3.3P_x - 0.2P_z + 0.001Y \quad (3.5) \quad (2.1) \quad (0.5)$$

with  $t$  values in parentheses, where  $P_z$  is the price of another good  $Z$ , and  $Y$  is income. Is good  $Z$  a substitute or a complement? Can we say confidently whether good  $X$  is a normal good or an inferior good?

9. Look at Regression (b) in Table 3-1, and consider the following questions:
  - (a) Does cigarette consumption increase as income rises? Are cigarettes a “necessity” or a “luxury”?
  - (b) For the variable “Educational Level” a high school graduate is coded with level 9, and a college graduate is coded with level 12. What is the predicted difference in cigarette consumption between the two levels of education?
10. Table 22-1 shows GDP/capita and total health care spending per capita for 29 countries (the first two columns of numbers).
  - (a) Calculate the means of both variables.
  - (b) Calculate a regression relating health care spending to GDP/capita.
  - (c) Using the method discussed in equation (3.10) calculate the income elasticity of health expenditures.
  - (d) What does your answer to part c indicate about the “share” of GDP/capita going to health? Why?

# Economic Efficiency and Cost-Benefit Analysis

- Economic Efficiency
- Cost-Benefit Analysis: Background
- Cost-Benefit Analysis: Basic Principles
- Valuing Human Life
- Cost-Effectiveness Analysis
- Cost-Utility Analysis, QALYs, and DALYs
- QALYs Revisited: Praise and Criticism
- Conclusions
- Appendix—Discounting

In health economics, policy concerns require that we frequently and systematically evaluate alternatives. Just as rational individuals want to make the best choices given resource constraints, governments, too, face choices constrained by resource availability. For example, legislators and other policymakers must decide whether to spend more on preventive care versus giving more support to acute care facilities, or perhaps to medical research. When government regulates, its own expenditures may be relatively small. The economic consequences of its regulation can be very large, however, and corresponding care must be taken in evaluating the alternative scenarios. Economists base such decisions on the concept of efficiency. In developing the microeconomic tools for this textbook, we explained the “welfare loss” caused by a monopoly restricting quantity of production by charging too high a price. The nature of this welfare loss is that it describes inefficiency—society has foregone opportunities for mutual gain. Efficiency, however, applies to a broader range of phenomena than just monopoly, and we will begin here by developing the concept more fully.

## ECONOMIC EFFICIENCY

Economic efficiency exists when the economy has squeezed out every opportunity for net benefits possible through voluntary means. Consider a single market, such as a local market for apples. Consumers’ preferences for the apples can be measured by their willingness to pay for them; each person might have a different amount of money in mind. Likewise the opportunity costs to society of an apple can be measured by the marginal cost of production. In some cases, there will be extra or “external” costs or benefits involved, but for the moment, let’s assume that all the benefits and costs are the private willingness to pay and the marginal costs. An efficient

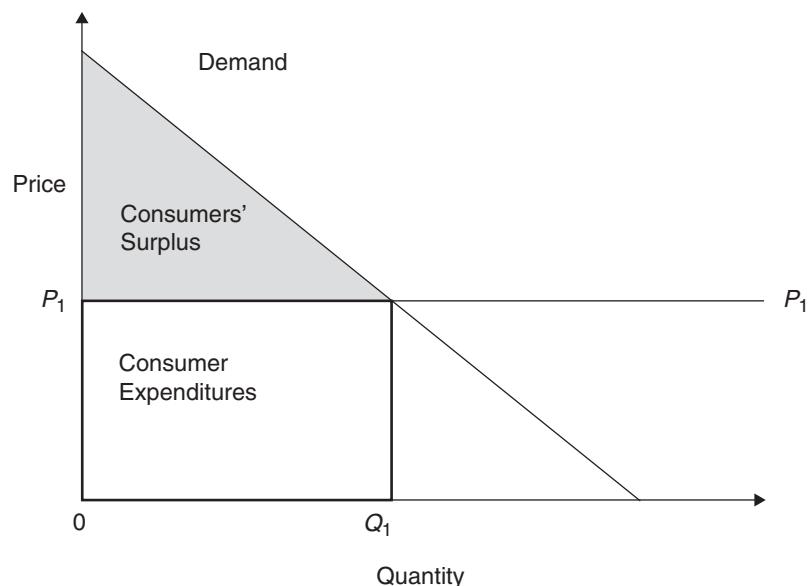
result for society will require that the wants of consumers, as reflected in their demand, be gauged against the costs to society of the required production.

Economists use the tools of demand and supply analysis to define the efficient allocation of resources in competitive markets, and this idealized sort of market is convenient to explain the concepts. The competitive market is a market form that functions “properly.” We will also see that for markets that do not function properly, or for cases in which markets do not exist at all, the underlying benefit and cost concepts often still apply.

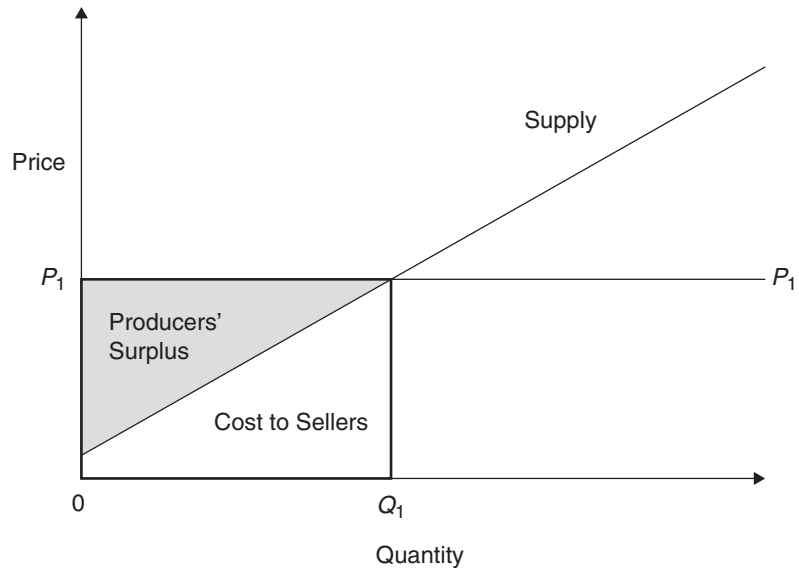
The demand curve for apples represents consumers’ willingness to pay for various amounts of apples. Marginal willingness to pay is another way of representing the demand function relating the quantity of apples demanded to the price of apples. Imagine, for example, lining up individual consumers from left to right in descending order on the willingness to pay for one apple. At small total quantities (along the market demand curve), the marginal willingness to pay is high, as only those who place considerable value on getting an apple are willing to pay for it. At larger quantities, the consumer’s marginal willingness to pay is lower; additional consumers would not buy unless the price was lower.

At price  $P_1$ , in Figure 4-1, consumers together spend amount  $P_1Q_1$ . This is part of the measure of what the apples are worth to the consumers, and this much is also what they pay out. The total value of the apples to the consumers, however, also includes the additional shaded area under the demand curve, referred to as the consumers’ surplus triangle. The consumers would have been willing to pay more than  $P_1$ , but did not have to do so. To see this, reconsider the example. Each consumer who buys an apple is willing to pay the price on the demand curve. That price reflects his or her benefit, but one must subtract the market price to get the net benefit. The shaded area then is simply the sum of the net benefits of each individual buyer. Alternatively, we say that the consumers’ surplus at  $Q_1$  equals the total value to consumers (the area under the demand curve up to  $Q_1$ ) minus the amount that they must spend,  $P_1Q_1$ .

Figure 4-2 presents a supply curve, which in the competitive market case measures the marginal costs for producers to bring apples to market—the higher the market price offered to them, the higher the marginal cost they are willing to bear. Suppose that the market price of apples were \$10 per bushel, but that some of the producers would supply apples even if the price were only \$5. The \$10 per bushel that they receive constitutes a surplus of \$5 per bushel in excess of the price that is necessary to induce them to transfer resources from other uses to producing apples. Total



**FIGURE 4-1** Consumers’ Surplus

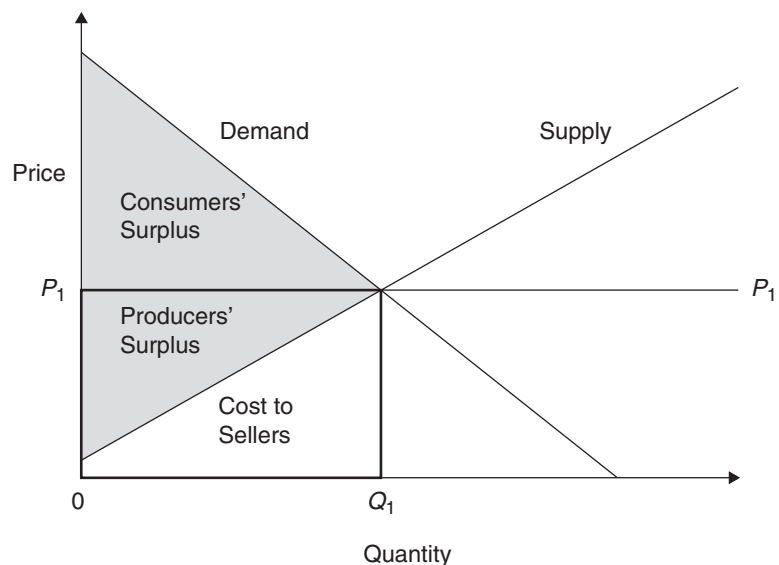


**FIGURE 4-2** Producers' Surplus

consumer expenditures  $P_1Q_1$  are apportioned into the resource cost to the sellers (the white area), plus the shaded area showing producers' surplus.

The economic criterion for maximizing well-being is to maximize the sum of the consumer and the producer surplus. Combining the supply and demand diagrams in Figure 4-3, we see the gains to consumers plus the gains to producers (the total net benefits to society) in the total shaded area. We calculate total "welfare" of the community in this market by using this equation: Total Welfare = Consumers' Surplus + Producers' Surplus.

Figure 4-3 shows that the sum of the surpluses is maximized at quantity  $Q_1$ , where demand equals supply. If quantity were less than  $Q_1$ , we could increase both consumers' and producers' sur-



**FIGURE 4-3** Efficient Quantity



plus (the shaded areas) by increasing  $Q$ . If quantity exceeded  $Q_1$ , we could increase the surpluses by reducing  $Q$  because the incremental quantity (beyond  $Q_1$ ) costs more (the supply curve) than it is worth to the consumers (the demand curve).

Understand that efficiency requires that the optimal quantity be brought to market. In several places in this textbook we find that monopolies are economically inefficient by bringing too few goods to market. In contrast, polluters generally produce too many goods whose by-products impose costs on society that exceed the goods' market prices.

In competitive markets, supply and demand provide the efficient quantities of goods to the market—prices ration supply and demand according to consumer preferences and producer costs. However, a wide range of goods exists for which such market signals are not readily available. These include bridges, parks, water purification systems, or mandated clean air. Decisions on whether to screen for certain types of cancers or whether to provide vaccines to the public, for example, must also be evaluated on criteria that do not easily lend themselves to market tests.

With absent or incomplete markets, various evaluation tools have been developed to measure and compare project costs and benefits. Characterized in general as cost-benefit analysis, these tools seek to determine the appropriate quantity by measuring incremental or total costs and incremental or total benefits. We address these methods in the following section.

## **COST-BENEFIT ANALYSIS: BACKGROUND**

Cost-benefit analysis (CBA) provides one good method for evaluating proposed public projects. Early forms of CBA appeared over a century ago when the U.S. Army Corps of Engineers sought to evaluate flood control and other water systems. With the large public works projects during the Depression of the 1930s, the need to justify expensive programs became readily apparent. In 1950, a congressional subcommittee recommended procedures for using the benefit/cost ratio to compare projects. The introduction of Medicare and Medicaid in 1965 led health economists to focus more on CBA. The federal government continued to improve and standardize methods, and since 1981 all new federal regulations must undergo CBA.

CBA measures the benefits and costs of projects in money terms. This often requires that we place dollar values on years of life or improvements in health and well-being. These challenges have led to the development of new ideas, and health analysts now use the general term *economic evaluation* to represent the entire collection of tools. Throughout the discussion, we will take the perspective of society as a whole rather than the narrower focus of the individual or firm. This is the appropriate perspective for public projects. We can also apply this logic to the investment decisions of a single hospital deciding whether to invest in a PET scanner or even to the decision problem of a young man or woman considering a career in medicine.

Cost-effectiveness analysis (CEA) and cost-utility analysis (CUA) have emerged as the principal alternatives to CBA. CEA applies to problems where the goal is accepted at the start and the problem is only to find the best, most efficient, means to achieve it. CUA is a special form of CEA that introduces measures of benefits that reflect individuals' preferences over the health consequences of alternative programs that affect them.<sup>1</sup>

## **COST-BENEFIT ANALYSIS: BASIC PRINCIPLES**

The basic idea of cost-benefit analysis mirrors the measurement problems that it addresses. For example, CBA evaluates public investment costs, including those that have no market. Likewise it investigates benefits that have no markets, to determine their prices. These may entail benefits or

---

<sup>1</sup> Excellent collections of articles providing comprehensive analyses of many of the issues raised in this chapter are found in Gold et al. (1996) and Drummond et al. (1997). C. Goodman (1998) provides a concise summary of the literature as well as a practical guide to the evaluation process.

costs over many periods. CBA also addresses public choices involving either benefits or costs that are external to the market they came from. Projects, such as putting a dam across a river or reconsidering the efficiency of a congressional program, are often controversial. Within health economics, controversies over the wisdom of immunization programs, patient screening, or heart transplants, for example, invoke many problems and criticisms that cost-benefit thinking must address.

CBA rests on the premise that a project or policy will improve social welfare if the benefits associated with it exceed the costs. The benefits and costs that are counted must include not only those directly attributed to the project but also any indirect benefits or costs through externalities or other third-party effects. Thus, where  $B$  represents all the benefits and  $C$  represents all of these costs, a project is deemed worthwhile if  $B - C > 0$ . We can also rank projects according to the benefit to cost ( $B/C$ ) ratio; thus, a higher  $B/C$  ratio generally indicates a project that will deliver greater social benefits for a given dollar of costs.

### Measuring Costs

As in standard economic theory, costs are measured as opportunity costs. The most common difference between public and private project evaluation is that public projects often have opportunity costs that have no market to serve as a guide for pricing. On the one hand, a dam project can destroy habitat for animal life, cover historical landmarks under water, and force whole towns to close down. On the other hand, it can provide people who fish with new territory, attract waterfowl, and provide campsites and swimming areas for recreation. None of these examples is either a cost or a benefit ordinarily bought or sold in markets. Much of the controversy of CBA derives from the imprecise task of placing dollar values on these difficult-to-evaluate costs and benefits. In some cases, the required values can be inferred. For example, we can measure the benefits of a newly created lake view by observing how much the market values of nearby homes increase. For many cases, however, there is no easy answer.

The previous example also mentions a second problem. Public investments may have side effects that impose additional measurement difficulties. Consider a public immunization program that would offer protection against the flu to a wide cross section of Americans and is especially concentrated on the elderly and the ill. The costs of materials and manpower can be measured directly, and the benefits to those immunized can be estimated. The program, however, also benefits people who never get immunized by reducing the number of infected carriers, thus reducing their exposure to the flu. These external benefits must be counted in the CBA, yet accurate estimation may prove difficult.

A pollution clean-up program raises similar questions of how to treat externalities. For example, factories that discharge contaminants into the air or water create external costs by damaging the environment and adversely affecting third parties. Conversely, pollution abatement creates external benefits to others (e.g., boaters and home owners) who are not directly involved in the firm's decisions. They typically do not pay for the benefits they receive.

### Risk Equity Versus Equality of Marginal Costs per Life Saved

Calamitous events often engage public sympathy to support rescue programs and emergency health care. These programs, in the abstract sense, seek to reduce health risks to the victims—for example, the risk of catching an infectious disease. Some argue that society ought to apply public resources so that health risks are shared equitably across the population. Perhaps equalizing life risks is impossible, but suppose that it were possible. Would it be the best choice for the use of society's resources? Viscusi (2000) explains why such a plan would cause society to fall short of its welfare potential. Contrast a plan that follows sound economic principles. Were we instead to distribute public investments so that the marginal cost of a life saved were equalized across publicly funded programs, we would maximize lives saved for a given overall budget. The idea is simply that each next dollar spent should be placed where it does the most good (marginal analysis in CBA is discussed in detail later in this chapter). Does U.S. spending follow this cost efficiency standard? In

**BOX 4-1****The Cost of Saving Lives**

The need to encourage rational social choice has been forcefully made by Harvard University's Center for Risk Analysis. After reviewing publicly available studies of life-saving interventions, the Center found enough data to compute the cost per life-year saved for 587 interventions representing an unprecedented range of both health and nonhealth programs. Estimates for selective interventions include:

Intervention	Cost per Life-Year (\$2010*)
<i>Medical</i>	
Childhood immunizations	less than 0
Prenatal care for pregnant women	less than 0
Influenza vaccine for all citizens	210
Mammography for women age 50	1,215
Annual mammography for women ages 40–49	142,500
Intensive care for ill patients with major vascular operations	1,275,000
<i>Nonmedical</i>	
Random motor vehicle inspections	2,250
Water chlorination	6,300
Pneumonia vaccination	18,000
Construction safety rules	57,000
Home radon control	211,500

Resource savings are subtracted from costs. Thus, some programs, such as childhood immunizations and prenatal care, have negative net costs because the costs of the programs are more than offset by the subsequent savings from lower health care costs.

\*Note: The costs were originally estimated in 1993 dollars. These have been converted to 2010 dollars the CPI increased 50 per cent by 2010.

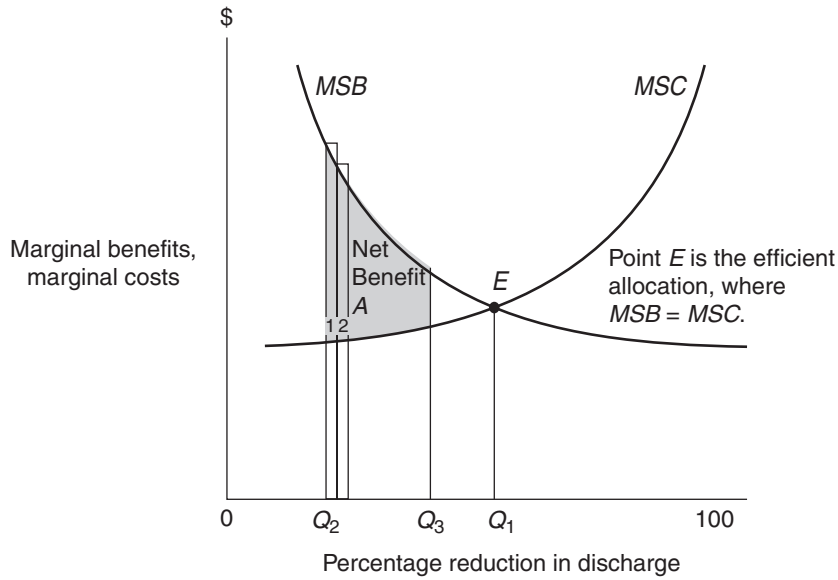
Source: Tengs, Tammy O., et al. "Five-Hundred Life-Saving Interventions and Their Cost-Effectiveness," *Risk Analysis 15* (1995): 369–384, with permission of the publisher.

Box 4-1, Tengs et al. reports on estimated costs of a life-year saved in several categories of programs. These cost figures are certainly far from equal!

**Marginal Analysis in CBA**

Figure 4-4 illustrates the marginal analysis principle applied to CBA. The marginal social benefits curve, representing the sum of all beneficial effects from increasing the abatement program by one unit, is shown as the downward-sloping *MSB* curve. The marginal social costs, representing at each point the sum of all costs of increasing the program by one unit, constitute the *MSC* curve. For many pollutants, the marginal social benefits will include the benefits to the public of improved health. As an example, the incidence of cancer and respiratory disease has been linked to various forms of air pollution.

Society's net benefit will be maximized where marginal social benefits equal marginal social costs. CBA represents an attempt to get the information with which to make the assessment. To illustrate the logic, suppose the current level of discharge at  $Q_2$  can be observed and that a project or legislation requiring  $Q_3$  is proposed. If *MSB* and *MSC* are properly measured, valid estimates of the project benefits are given by the area under the *MSB* curve between  $Q_2$  and  $Q_3$ , and similarly for costs by the area under the *MSC* curve. The net benefit is equal to area *A*.



**FIGURE 4-4** Efficient Use of Resources Where Marginal Benefits Equal Marginal Costs

Cost benefit principles may be helpful with such a minor decision as which lawnmower to buy. But the same principles can inform decisions that have an impact on the whole planet. The case at hand in Box 4-2 presents a modest invention that has intriguing possibilities beyond the village.

**BOX 4-2**

**Cookstoves, Global Warming, Health in Developing Countries and CB**

Much of the world burns biomass fuel for cooking and heating. Unfortunately this often results in smoky, unhealthy interiors, time lost in handling the material, and emissions of greenhouses gases. Cost benefit analysis proves useful in identifying the best, or at least, the most hopeful solutions. Of the several studies, a recent one by García-Frapolli et al.(2010) illustrates the method and some strong results. They focused on the Purépecha region of Mexico. The authors estimated the benefits and costs per year of replacing peoples’ current cookstoves with the more efficient Patsari design as follows:

Benefits/Year per cookstove	\$
Fuelwood savings	400.8
Job creation and income	19.1
Health impacts	208.6
Environmental impacts	103.2
Total benefits	731.7
Costs/Year per cookstove	
Costs of cookstove construction	83.3
Indirect costs (dissemination)	25.3
Total costs	108.6
Ratio of Benefits to Cost*	9 to 11

\*These ratios are calculated from an estimated lifetime of the stove and they depend on the discount factor assumed. See the next section for more details on discounting.

With a benefit cost ratio this high, they concluded that the project was well worthwhile.

## Discounting

For projects that last more than one period, the future costs and benefits must be discounted to put them on an equal basis with present values. This necessity arises from one of two basic reasons, both of which suggest that future dollars usually are not worth as much to people as present-day dollars. First, a dollar today has opportunities other than the project of study. That dollar could have been invested in an interest-bearing instrument, such as a bond, and then it would have been worth the dollar plus the interest in next period's dollars. Conversely, next period's dollar must be discounted at the market rate of interest to discover its present value.

Second, people have a tendency to prefer the present when allocating spending. Which would you prefer, a dollar right now or the same dollar 10 years from now? Most of us would choose the dollar now. One consequence of this time preference is that the equilibrium interest rate will be positive, rewarding those who wait and exacting a cost on those who can't wait or simply don't wish to. A simple introduction to discounting is provided in the appendix to this chapter.

Together, the time preference and the potential foregone interest from dollars spent on a project help explain why we discount future money values. The most commonly used method is relatively simple, even though the complexity of the underlying issues may be oversimplified by this approach (Frederick, Loewenstein, and O'Donoghue, 2002). Each period's costs and benefits are divided by a "discount factor" raised to the power  $t$ , the number of the time period, counting from the present as zero. The discount factor is  $(1 + d)$ , where  $d$  is the social discount rate, which for the present we assume is the market rate of interest. The present value equation is then

$$PV = \sum_{t=1}^{t=T} \frac{(B_t - C_t)}{(1 + d)^t} \quad (4.1)$$

where  $t$  is zero for the initial period and increases up to the period  $T$  in which the project ends. For example, in period one, the next period after the project starts, the discount factor is raised to the power one, and the net of benefit over cost in this future period is divided by  $(1 + d)$ . If a market interest rate of 6 percent were used for  $d$ , then the denominator would become  $(1.06)$ .

Notice also that the denominator in (4.1) becomes larger as  $t$  grows larger. This assumes that we discount the more distant future more heavily. This fact of discounting has raised difficult quandaries for both health care and environmental projects: Should those living in the present so disregard future generations? At first glance, time discounting may seem to guarantee that we will bequeath an unhealthful and polluted condition of life to the future generations. For reasons like these, some analysts propose that the social rate of discount,  $d$ , be set at a lower level than the current market interest rate or that we apply discount rate patterns that decline over time. It is clear in (4.1) that the lower the chosen value of  $d$  the greater is the emphasis placed on the future.

Two ideas help clarify why many economists prefer to use the market rate of interest in the present context. First, economists often temper the claims some people make on behalf of future generations with the knowledge that all human valuations must ultimately come from people living in the present. In a sense, we all speak for future generations through our actions in the capital markets. Second, using market interest rates also has the advantage of measuring what people actually do as opposed to their responses to more hypothetical questions. Yet many would disagree. In any case, it is ultimately a social decision and one that has surprisingly large consequences. For a practical illustration of these consequences, see Box 4-3, "Discounting and Global Warming."

A number of other troublesome issues arise, however, when proposing the market interest rate for discounting public projects. Some find the approach inequitable when high and low income groups differ in time preferences. As the feature suggests, the more keenly felt problem at present is the prospective loss of human life and the possibility of permanent damage to the environment. The

**BOX 4-3****Discounting and Global Warming**

On October 30, 2006, the British Treasury issued a disturbing report on the potentially cataclysmic consequences of greenhouse gases for our planet's future. Named after its leading author, this became known as the *Stern Review on the Economics of Climate Change*, the *Stern Review* (Stern, 2007) for short. It is widely credited with encouraging the world to understand the urgency of the problem, and it is worth reading in its entirety. Furthermore, the discounting issues it raised help illuminate health economic discounting issues in general.

The *Stern Review* chose a discount rate of nearly zero to treat future losses. William Nordhaus (2007) criticized this choice, arguing that our economies grow and that a zero rate ignores the fact that greater future wealth could make the estimated loss more affordable. Economists sometimes recommend a discount rate of 3 percent, which is closer to economic average growth rates. A few might recommend 5 percent, though most think this too high. Opponents counter that the population also grows, and the gains from avoiding calamity will be spread over more people. Economists reply that a zero discount rate fails to reflect how real people actually treat future values versus present values, though many allow that environmental rescue ought to get a special rate, say, 1 or 2 percent. Thus, should we discount the value of future lives saved at 0, 1, 2, 3, or 5 percent? People who think that this is a mere "academic" squabble will be the most surprised at the consequences.

To illustrate this, suppose we consider a calamitous climate event 100 years from now that would kill 1 million people and that we measure the value of each life in the future as \$6 million dollars. How much are we willing to invest today to save those future people? One hundred years from now the loss is \$6,000 billion (or \$6 trillion!), but what is the equivalent amount in today's dollars? The formula is Present Value = (Benefit at 100 years hence)  $\div (1 + d)^{100}$ , where  $d$  is the social rate of discount that is chosen. The accompanying table reports those calculations.

Benefit 100 years from now	The social rate of discount, $d$	Net present value
\$6,000 billion	0.00	\$6,000.0 billion
\$6,000 billion	0.01	\$2,218.3 billion
\$6,000 billion	0.02	\$ 828.2 billion
\$6,000 billion	0.03	\$ 312.2 billion
\$6,000 billion	0.05	\$ 45.6 billion

Clearly the chosen rate makes a difference. If one chooses a zero discount rate, then those future lives are valued at present at nearly one-half of the entire GDP of the United States; on the other extreme, if one chooses the 5 percent rate, then those 1 million lives are viewed as worth less today than what Americans spend annually on "paper and allied products." What rate  $d$  is correct in the environmental context? The economist generally argues that future values must be discounted. But at which rate? It is obvious that the consequences of a discount rate choice are huge. But a wider consensus has not been achieved. What is clear is that the *Stern Review* has forced the right issues into vigorous discussion.

accompanying feature illustrates the large consequences of choosing a discount rate. Finally, among health economists, there is also the proposal to discount ordinary costs at the market rate but benefits to life and environment at a separate and lower rate (Brouwer, 2005; Claxton et al., 2006).

**Risk Adjustment and CBA**

To apply the market interest rate, however, we must recognize that there are many market interest rates. Chief among the reasons that these differ is that projects that are relatively more risky tend to have relatively higher rates of interest. Often evaluators will adjust the social rate of discount to

reflect the riskiness of the public project. Even this is conceptually not an easy task, however. Private markets for capital projects have private bidders and sellers, and they naturally are capable of assessing perceived riskiness to themselves of adopting a prospective project. In contrast, public projects in principle represent the public at large, whose view of the project's risk might be difficult to discern.

Nobel laureate Joseph Stiglitz has recommended that a relatively lower social discount rate better reflects the public's role (Stiglitz, 1988). Stiglitz also proposes certainty equivalents to eliminate the biases that can result from including a risk-adjustment factor in the discount rate. Under the certainty equivalent method, the uncertain net benefit in any period, often represented by a probability distribution over project outcomes, is replaced by its equivalent (the value at which the decision-makers are indifferent between the risky set of outcomes and a value received with certainty). The more risky a project, the lower will be the certainty equivalent to someone who is risk averse. The cost-benefit criterion represented by equation (4.1) is then applied using the certainty equivalents for each period. The risk adjustment problem attracts the interest of health economists. Some researchers propose using a risk adjusted rate of return to compare programs with different risks (Sendi, Al, and Zimmerman, 2004).

### **Distributional Adjustments**

Although cost-benefit analysis is intended primarily to improve efficiency, as some of these examples have shown, changes in the income distribution often result from a project. In cases of narrowly focused projects, the tendency is to have a relatively small number of large gainers and perhaps many small losers. To the extent that society is concerned about equity, the distribution of the gainers and losers by income group should be a consideration. At the practical level, after projects are ranked according to their net benefits, decision-makers could invoke informal judgments as to the relative effects on the distribution of income and then adjust their rankings. More formally, Stiglitz proposes distributional weights through which the net benefits or losses give lower-income groups more weight than other groups. Of course, the method still will be subjective in that the weights themselves will necessarily reflect the judgments of the decision makers.

### **Inflation**

Another concern deals with the treatment of inflation. Conceptually, macroeconomic inflation is not a problem. Because estimates of the inflation rate often turn out to be incorrect, it is best to measure both benefits and cost in current or real terms and then discount at the real (inflation-free) discount rate. If an inflation factor is introduced, then the discount rate should be increased by that inflation rate to get the nominal rate. It is important, though, that the discount rate reflect that inflation factor and not some other rate.

## **VALUING HUMAN LIFE**

One of the most difficult but often unavoidable tasks in health care CBA is to place a value on human life. Several approaches are widely used. The first, known as the human capital approach, estimates the present value of an individual's future earnings. This approach has been especially favored in legal applications that require estimates of damages. It also measures the loss of national output from mortality and morbidity or the production gains from saving and extending life.

In other ways, however, the human capital approach is flawed as a welfare measure in that it does not directly measure an individual's willingness to pay to avoid risks of death, injury, or illness, nor does it measure what he or she is willing to accept as compensation for taking on such risks.

## Willingness to Pay and Willingness to Accept

*Willingness to pay* is exemplified in many everyday questions, such as “How much are you willing to pay for the reduction in risks provided by new locks on your door?” The flip side of this concept is the compensation you would require to accept an additional risk to life and limb. As in the previous case, the *willingness to accept* method has many everyday examples. In fact, its theoretical basis is found in the economic thought regarding the compensating differentials paid to laborers across various lines of work, a line of thought that dates back to the foundations of modern economics.

One of Adam Smith’s successes in *The Wealth of Nations* was to explain why the wages of workers differed across jobs—explanations that still resonate well with modern economic theory. Smith’s theory was the inspiration for modern-day economists to develop the theory of compensating differentials.

## Contingent Valuation

One can elicit an individual’s willingness to pay for a risk-reducing medical treatment, but not without difficulty. The method of contingent valuation poses sets of medical contingencies such as: “If you faced an X high risk of heart attack, how much would you be willing to pay for a medical procedure that would reduce your risk to Y?” The set of questions, visual images (if used), and researcher interactions is called the format. Contingent valuation has made it possible for practical studies to move forward. Yet it has proved vulnerable to challenge. One of the major problems is “hypothetical bias”, the idea that the subject will not or cannot answer hypothetical questions realistically. We find that changes in format can induce change in the valuations, a so-called framing bias (Whynes, Frew, and Walstenholm, 2005).

Potential resolution may come from studies that compare willingness to pay based on real-life behavior with measures of the same concept based on a question and answer format. A recent study that does this (Bryan and Jowett, 2010) concludes that the question-and-answer method does reasonably well for the given patient therapy they chose to study.

## How Valuable Is the Last Year of Life?

Health system analysts commonly express puzzlement or dismay over “Why do we Americans spend so much of our health care dollar on the last year of life?” Later in this chapter we describe “QALYs” and how this measure in contrast implies that a person has no claim on extraordinary health care when he or she gets very old. The elderly spend one quarter of their total health care expenditures on the last year of life. Even economists using standard utility analysis question why one’s utility of life would get to be so high when one has low natural prospects for living many years and when one’s motive to bequeath to one’s heirs is a ready alternative.

Becker, Murphy, and Philipson (2007) offer an interesting resolution to this puzzle. They begin by arguing against linearity in the calculation of the value of a life year. For example, the commonly used value of \$100,000 per life year is based on a linear extrapolation for peoples’ responses to marginal changes in their probability of death. It would seem irrational to us to save a life worth \$100,000 by incurring expenses worth \$500,000. But the \$100,000 would be incorrectly applied to a patient’s valuation of life when the risk change is life versus death right now. People facing a survive-or-die situation are not dealing with a marginal change; for example, a firefighter may increase his risk from 0.010 to 0.015, a marginal change. But the patient at life’s possible end may be considering a hoped-for reduction in the probability of death from 1.00, if not treated, to 0.60, for example, if treated.

Their analysis rests on four intuitive ideas. First, for many of the very old and sick, their resources have very low opportunity costs because they cannot enjoy their wealth once they have died. Second, they may rationally have “hope” for living, including the hope that more advanced



**BOX 4-4****The Price of Everything**

Oscar Wilde’s famous quip about cynics has evolved into a familiar version about economists: “An economist is someone who knows the price of everything and the value of nothing.” Does this phrase hit the nail on the head in the case of value-of-life measurement?

There are in fact both technical and philosophical issues that raise concern. One technical question has already been asked: Have “all other things” truly been held constant during and across the analyses, such other things as the psychic rewards one receives from work? Consider someone interested in becoming a fire-fighter, clearly a risky job. Will the extra risk be just offset by the extra salary (assuming the fire department pays better than alternative jobs) so as to give an accurate reading of his or her value of life? Not likely, especially when high-risk professionals like fire-fighters feel pride in the protection they provide to the community. The earned sense of honor is a “psychic” reward that may, in their own eyes, overcome a shortfall in their differential compensation in wages. A fire-fighter might even conceivably accept the job at less pay than is available elsewhere.

There remains, however, a bigger question than such technical issues because there is a larger sense to life than economists have intended to capture. The nature, meaning, and ultimate value of life are fundamental mysteries to economists, just as they are to humanity in general.

Yet on an everyday basis, humans must make choices about everyday risks, and most of us desire to trade and share these risks. We may do so to feel safer or simply to handle those consequences that can be planned for. When someone is buying insurance for a life risk or when a courtroom is deciding a just compensation to a widow, value-of-life assessments will be made. The economic approach we have described in this chapter helps in the making of such assessments, and, in this way, it helps people to share the risks of life.

health care will be developed within their extended lifetime. Third, their “social” value of life (the value of their life not only to themselves but to family, friends, and community) may be very high. Finally, these authors show that the value of an extended life year may be as high for frail patients as it is for those of higher quality health. Each of their hypotheses derives from rational behavior. While we may yet find that the high expenditures of the ill elderly result from irrational thinking, the analysis of these authors shows that they need not be.

The two methods we have described—willingness to accept, deriving from labor economic theory of compensating differentials, and willingness to pay, deriving from consumer purchasing behavior for risk-reducing devices—are nearly two sides of the same coin. A basic difference, however, is that purchasing behavior is limited by a person’s budget; we can afford to pay only so much to reduce our risks. In contrast, the willingness to accept has no theoretical limit, a fact most relevant for the world’s poor.

Researchers in health economics and other disciplines have applied both approaches. However, methods to elicit dollar values differ in practice. “Wage based” estimates of willingness to accept risk for extra pay, for example, observe real life behaviors. “Stated preference” methods ask subjects to state the risk/money tradeoffs they would prefer. Table 4-1 illustrates the wide range of estimates that investigators have found from various applications. Many have studied the value of a statistical life (VSL), and the table summarizes them using meta-analysis. These are estimates of VSL using simultaneously all the data from a given set of previous works. Given this variety of values, some health economists make a practical choice and pick what they believe is the best approximation. Many choose to describe the value of a human life as a bit over \$6 million dollars while \$100,000 describes the value of one life year (Cutler, 2007).

Nevertheless there are numerous complications. We are frequently discovering ways to refine the VSL estimates. A good example of this is a paper by Aldy and Viscusi (2008), which explores how and why VSL varies with age. On one hand, as you age you have fewer life years remaining,

**TABLE 4-1 How Much Is One Life Worth?**

<b>Meta analysis of wage based studies</b>	<b>Years covered by the studies</b>	<b>Value of Life in 2009 dollars</b>
Miller (2009)	1974–1990	5.2 million
Mrozek & Taylor 2002	1974–1995	2.0 to 3.3 million
Viscusi & Aldy (2003)	1974–2000	6.9 to 9.5 million
Kochi et al (2006)	1974–2002	11.1 million
<b>Meta analysis of stated preference studies</b>	<b>Years covered by the studies</b>	<b>Value of life in 2009 dollars</b>
Kochi et. al. (2006)	1988–2002	3.5 million
Dekker et al. (2011)	1983–2008	2.7 to 8.5 million
Lindhjem et al. (2010)	1973–2008	3.2 million

*Note:* These data are from Maureen Cropper, James K. Hammit, and Lisa A. Robinson, (2011), “Valuing Mortality Risk Reductions: Progress and Challenges,” Discussion Paper, Washington, DC: *Resources for the Future*.

suggesting that VSL might decline with age. On the other hand, income and wealth grow with age, and then decline in one’s later years. It may not be possible to transfer wealth easily between age levels. The authors’ estimates, all things considered, reveal that VSL increases in one’s earlier years and then declines. That is, VSL follows an inverted “U” shape over age.

### **Cost-Benefit Analyses of Heart Care Treatment**

Notwithstanding the difficulties of CBA, one finds excellent examples in the literature.<sup>2</sup> To illustrate, we examine a CBA published recently by David Cutler (2007) that compared treatments for patients at risk for recurrence of heart attack. The treatment in focus is “revascularization,” the use of bypass surgery and/or treatment with stents to improve blood flow to the heart. Cutler’s study not only examines econometric issues but also estimates the patient’s lifetime costs and benefits. To model the patient’s lifetime, he acquired data to permit the study of 17 years into the future following the treatment. Table 4-2 reports the change in risk of death at each of the future years; it compares improvements to survival for patients admitted to a revascularization-capable hospital and those admitted to a High Volume (assumed to be high quality) hospital but one not having revascularization capability.

By analyzing all 17 years we see some unexpected patterns. Note that at Year 1, the revascularization estimate shows that it actually worsens the chances of survival. Bear in mind here that Cutler’s columns labeled “Survival” are measured negatively as changes in the death probability, so a negative value indicates a reduction in death risk. It quickly changes to improve the survival rate for many years. The cost of the advanced treatment is much higher than costs via the control group, those hospitals that lack the revascularization capability and yet treat heart cases in high volume. Cutler calculated the increased life expectancy attributable to each of the two treatments. Revascularization increased life expectancy in this sample by 1.1 years (the sum of the revascularization survival rates) at a cost of approximately \$38,000, thus achieving its gains at a rate of \$33,246 for each life year. The High Volume hospitals increased life expectancy by only 0.06 years, and even though their costs were low, their costs per life year saved were \$175,719. Estimating the value of a human life year to be about \$100,000, Cutler concluded that the \$33,246 gain from revascularization easily proved cost beneficial.

<sup>2</sup> See also Weisbrod (1971) for an early, but classic, cost-benefit study involving research on poliomyelitis.

**TABLE 4-2 Costs and Benefits of Medical Technology for a Lifetime**

Time after MI	Revascularization		High Volume Hospital	
	Survival	Spending	Survival	Spending
1 Year	0.061	\$ 30,149	-0.009	\$ 4,065
2 Years	-0.029	\$ 27,339	-0.005	\$ 5,300
3 Years	-0.067	\$ 25,919	-0.004	\$ 5,993
4 Years	-0.043	\$ 26,820	-0.001	\$ 6,560
5 Years	-0.106	\$ 27,517	-0.005	\$ 7,296
6 Years	-0.119	\$ 29,662	-0.005	\$ 7,659
7 Years	-0.119	\$ 31,090	-0.005	\$ 7,953
8 Years	-0.108	\$ 32,919	-0.004	\$ 7,982
9 Years	-0.111	\$ 36,961	-0.006	\$ 8,087
10 Years	-0.119	\$ 38,028	-0.007	\$ 8,314
11 Years	-0.113	\$ 38,191	-0.006	\$ 8,532
12 Years	-0.120	\$ 40,804	-0.009	\$ 9,002
13 Years	-0.074	\$ 38,079	-0.006	\$ 9,161
14 Years	-0.064	\$ 38,708	-0.005	\$ 9,671
15 Years	-0.047	\$ 36,758	-0.005	\$ 9,524
16 Years	-0.041	\$ 37,200	-0.006	\$ 9,599
17 Years	-0.051	\$ 37,990	-0.007	\$ 9,770

*Note:* High Volume Hospitals are defined as hospitals that admit 75 or more heart attack patients in a year; these serve as comparisons and they are assumed to be high-quality hospitals in terms of heart care. High-quality hospitals do not necessarily have revascularization capability. “Survival” measures change in the risk of death compared to the hospitals that have neither high volume (quality) nor revascularization capability.

*Source:* Cutler, David, The lifetime costs and benefits of medical technology, *Journal of Health Economics*, 26 (2007), 1081–1100; data from his Table 5, p. 1094, with permission. Copyright © 2007 published by Elsevier B.V.

## COST-EFFECTIVENESS ANALYSIS

Given the difficulties of placing monetary values on life and health, as well as valuing other intangible benefits, cost-effectiveness analysis (CEA) sometimes provides a more practical approach to decision making than CBA (Garber and Phelps, 1997). CEA compares the costs of achieving a particular nonmonetary objective, such as lives saved. In cost-effectiveness analysis, one assumes that the objective is desirable even if the benefits have not been evaluated in monetary terms (strictly speaking, each of the projects might yield negative net benefits were it feasible to compute those net benefits). Though the analysis avoids valuation of benefits, the problems of determining costs remain.

The proper comparison of costs per output in CEA is based on the ratio of incremental costs to incremental output, as noted in equation (4.2). Let the change in social costs incurred due to a particular project be  $C_1 - C_0$ , and let the gain in health output be  $E_1 - E_0$ . Then the various projects are compared by the ratio:

$$\text{CEA Ratio} = \frac{C_1 - C_0}{E_1 - E_0} \quad (4.2)$$

The costs as usual are in dollars, while the outputs are the chosen health status measure. To compare projects with this method, the outputs must be measured in the same units across all projects considered.

### Advantages of CEA

The task under CEA is conceptually similar to a firm's production decision, which is to produce a chosen level of output from among alternative production methods at the lowest possible cost. Also, as in the firm-production decision, the objective must be quantifiable and measured in the same units across projects. Otherwise, a clear relationship between costs and output cannot be determined.

For example, cost-effectiveness analysis has been applied widely by the Department of Defense to determine the most cost-efficient means of achieving a particular level of military preparedness. Objectives can be quantified, in terms of ability to deploy forces, and the most efficient means of achieving the objectives is estimated. Note that CBA is not feasible in such cases because of the difficulty in evaluating benefits of the military capability in monetary terms. As another example, many communities are currently interested in recycling to the point where mandatory recycling is becoming more widespread. Assuming that a community has decided on the goal of reducing garbage mass, cost-effectiveness analysis can be used to compare recycling with incineration and other waste-management strategies.

Finally, CEA can be a useful first step toward undertaking a cost-benefit study. If the analysts run into significant problems in undertaking a CEA, it is unlikely that a CBA will be feasible. Conversely, good progress in developing a CEA can often determine whether it is possible to take the next step and extend the CEA into a cost-benefit study.

### COST-UTILITY ANALYSIS, QALYS, AND DALYS

The pressing and frequent need for cost-effectiveness analyses of health projects might account for the development of practical variations of the technique, which together can be described as cost-utility analysis. Principal among these is the quality-adjusted life-year (QALY); under this variation, each project is evaluated on the basis of its incremental costs per extra QALY delivered to the patients or other subjects (Garber and Phelps, 1997; Ried, 1998).

The QALY is a weighting system that assigns a value,  $q$ , ranging from 1 (perfect health) to 0 (death) to represent quality of life for each year. The details of QALY construction can be found elsewhere (Gold et al., 1996; Torrance, 1997). In its welfare economic version, the weights for QALYs are in principle derived by eliciting the individuals' preferences for different states of health. In practice, the QALYs are often summed over groups of people. In this case, the procedure departs from standard welfare economics by its comparison and in some cases imposing of interpersonal "utilities" across people summed in a cardinal manner, practices that welfare economists generally avoid. However, once the weights are agreed upon and the QALYs are acceptably calibrated, then the measure becomes:

$$\text{QALY} = \sum_{i=1}^{i=\max} \frac{F_i q_i}{(1+d)^i} \quad (4.3)$$

where  $F_i$  is the probability that the person is still alive at age  $i$ ;  $d$  is the time discount factor; and the value  $q_i$  is the quality weight, between 0 and 1, assigned to each year of the person's remaining life until a maximum value,  $\max$  (Garber, 2000).

Suppose, for example, that a patient has the opportunity for a treatment that will extend life by one year with a probability of 0.9 ( $F_1 = 0.9$ ) and by two years with a probability of 0.5 ( $F_2 = 0.5$ ). The patient will die with certainty after two years. Quality weight  $q_1$  is 0.8 in Year 1 and  $q_2$  is 0.6 in Year 2. The discount rate is 0.05 per year. Thus, using equation (4.3) the QALY computation is:

$$\text{QALY} = (0.9 \times 0.8 \div 1.05) + (0.5 \times 0.6 \div 1.05^2) = 0.96$$

indicating that the expected effectiveness of the treatment is 0.96 QALYs. This serves as the denominator for equation (4.2). Costs per QALY can then be used to compare alternative interventions.

### **An ACE Inhibitor Application of Cost-Effectiveness Analysis**

An ACE inhibitor is a medical treatment that can improve the life course of a patient with juvenile onset diabetes (Type 1). The issue for investigator Dong and his colleagues (2004) was whether there is an advantage for the patient to be treated with this drug at the conventional time, at the onset of a condition called microalbuminuria, or whether to treat right away when the diabetes is diagnosed. They chose as their standard of health outcome the quality-adjusted life-year (QALY). Their goal was to identify the cost per life-year saved by switching to the earlier treatment plan. They used cost-effectiveness analysis to do this.

These researchers investigated two other dimensions of analysis: patient age and blood category. These were studied for patients ages 20, 25, and 30 years, and for patient blood categories, HbA1c = 7% and HbA1c = 9%. These blood categories matter for the length of time for which treatment can be maintained, but we identify them here merely as two medically relevant patient categories. Costs included treatment and related exam costs as well as changes in costs to treat common disorders that affect diabetics during the course of their lives. All costs were discounted at a discount rate of 3 percent. ACE inhibitor treatment done immediately at the time of the diabetes diagnosis was termed as “Early” whereas treatment at the conventional time was termed as “Standard.” The QALY outcomes are the patient’s remaining quality-adjusted life-years, discounted at the 3 percent rate; this QALY figure should not be confused with life expectancy of the patient.

The “Difference,” that is Early less Standard, was positive in every case, showing that Early treatment costs more but achieves more life-years. Is the switch to early treatment cost-effective? The cost-effectiveness ratio, the one needed to determine which treatment method delivers the incremental QALY at least cost, requires data for the baseline QALY for each age/blood category. It is common, however, for such data to be unavailable, and it may even be unethical to generate data for untreated populations. The baseline QALY levels do matter, for example, for blood category HbA1c = 7%. Numerical experiments will show that cost-effective choice depends on the baseline QALY level.

Not having baseline QALY data, these authors focused on other questions: (1) at what age is early treatment most effective, and (2) for which blood group is early treatment most effective? To investigate these questions, the authors calculated the incremental cost between treatments in ratio to the incremental QALYs. Their data are very useful for this purpose, and they conclude that early treatment is most cost-effective for the youngest age groups, and that blood category HbA1c = 9% achieves the most cost-effective gains.

## **QALYS REVISITED: PRAISE AND CRITICISM**

### **Are QALYs Consistent with Standard Welfare Economics?**

As we have discussed, cost-benefit analysis is based on the economic standard of efficiency, the Pareto Principle, which states that if an option of society improves the well-being of some people while harming no one then that option enhances welfare. Put simply, CBA has developed to be the

standard of modern welfare economics. To say that CUA using QALYs is consistent with CBA would be to say that cost-utility analysis, too, chooses the efficient options for society; in the language of welfare economics, it would be “first best.” Recent work (Garber and Phelps, 1997; Bleichrodt and Quiggin, 1999) casts doubt; these findings conclude that very restrictive underlying conditions would be required in theory to cause cost-utility analysis to attain the welfare economic standard. Yet more recent is Blomqvist’s (2002) finding that CUA cannot, as typically applied, attain the first best result just described.

### **Extra-Welfarism**

Can a standard outside of standard welfare economics, one perhaps based on very different principles, give us a better world? Standards making this claim are called “extra-welfarist.” As you will see, extra-welfarist bases for decision-making, such as QALYs, often discard the economist’s idea of utility, and may reject the idea that people make their decisions rationally. Does it matter if we have two different approaches to health care project evaluation, one based within welfare economics and one outside of it?

Health economists often argue (e.g., Bala and Zarkin, 2000) that consistency with welfare economics is a valuable attribute to an evaluation of a public project. The fundamentals of welfare economics describe the net benefits to society of the project. Similarly, it is pointed out that the willingness-to-pay principal is appropriately broader than extra-welfarist approaches. Besides counting patients’ willingness to pay for extended life years, it includes their willingness to pay for side benefits to recreation and family life made possible by the treatment (Olsen and Smith, 2001).

The extra-welfarists, however, point to inadequacies in the standard economic welfare framework. The welfarist view, they complain, commonly counts up individual utilities without recognizing the interdependencies between people or their identification with the whole. Furthermore, the extra-welfarist approach using QALYs avoids a problem of willingness to pay; it avoids inequities that can be caused by an inequitable income distribution because ability to pay (due to higher income) is an important determinant of willingness to pay.

### **Sen’s Capability Approach and Qalys**

Amartya Sen (1985), a prominent critic of common conceptions of utility, proposes that each person is entitled to a life in which he or she can use a basic set of capabilities to achieve personal goals in life. Importantly, these capabilities would include basic health and functioning. Using this description, Cookson (2005) praises QALYs by showing that the quality index can be reinterpreted to represent a measure of Sen’s capabilities.

### **Linearity Versus What People Think**

A developing criticism of CUA with QALYs focuses on the method’s linear valuation of medical interventions as the simple sum of quality gains times life-years saved times the number of people treated. This is linear in a simple way, because increments to the quality index are merely added up. It doesn’t matter who you are or what your situation is. A given improvement in functioning is valued the same regardless. Dolan and colleagues (2005) tested these assumptions by reviewing the literature where people were sampled about QALYs. Here are some examples of many differences they found in ordinary people’s assumptions about QALYs.

- People are willing to sacrifice quality of life gains in order to give priority to the most severely ill.
- People do not like to discriminate by age, although they commonly weight the elderly somewhat lower.

- Health victims with dependent children are given more weight.
- People give much more weight to the health gains of people in the lowest social class.

### The Ageism Critique of Qalys

QALYs tend to place a reduced value on older people when evaluating a medical intervention. A successful treatment of an old patient saves fewer life-years; those years are already limited by nature. Many people ask, “Is this view of the elderly fair?”

The issue goes deeper with the philosophical assertion of Williams (2001), who asked a stronger question: Is an extra life year to be valued the same if an older person versus one younger receives it? His “fair innings” approach argues that the younger person merits more concern. We are each entitled, according to this view, to our “fair innings,” a normal human number of years (see also Daniels, 1985). Note that in this version, the old person’s shorter life expectancy is not the main issue; even an equal gain in QALYs is no longer assumed to have the same value between young and old.

An alternative approach, disability-adjusted life-years, (DALYs) points out that we humans tend to be dependent on the middle-age groups when we are very young or very old. To the adherents of this view, the greater social-related weights should be placed on people in the middle-age groups. The DALY theory provides a “hump-shaped” set of weights favoring the age groups in the middle.

Eric Nord (2001) argues that the value each of us attaches to our own lives depends on our point of view as situated in our time and place; that is, as we see it at our given age. Nord’s view is more consistent with standard economic methods of assessing value of life, which emphasize either the person’s willingness to pay for health or willingness to accept risks. Nord, in effect, raises the question of whether public valuation of an old person’s life should differ from the person’s own perception.

## CONCLUSIONS

The evaluation of prospective health projects has generated substantial interest among health economists in both the theory and the practical means to conduct evaluations so as to improve the society’s well-being. Cost-benefit analysis, which has a long history, requires the measuring of all benefits and costs attributable to the project both directly and indirectly. The need to identify external effects and to assign values to them in the absence of guidance from active markets poses both difficulty and controversy in practice. Future costs and benefits must also be adjusted to offset their differences from present values. Methods of discounting to accomplish this task are open to debate in that projects whose benefits are achieved only in the more distant future tend to benefit from lower rates of social discount.

The most difficult task of all for CBA is probably that of attaching dollar values to human life. Health economists who prefer an “extra-welfarist” approach view the willingness-to-pay principle, stemming from standard welfare economics as unsatisfactory. Cost-effectiveness offers a lesser but sometimes more practical evaluation result. By avoiding the assignment of dollar values to human life outcomes, the CEA focuses on providing useful guidance to the decision-maker. The problems of measuring costs still apply in CEA, but it avoids the most arbitrary and controversial steps. The CEA approach requires that the analyst describe the output of the various projects in common terms. Several methods of output measurement and various discounting techniques have led to variations on CEA as a group called cost-utility analyses.

Most prominent among cost-utility analyses is to measure health output as quality-adjusted life-years, QALYs. These methods, and the debate over health economic evaluation in general, are aimed at improving health policy for society. At its simplest level, economic evaluation recommends projects that achieve net positive benefits to society and prioritizes among these by their relative efficiency in doing so. Health and health care, however, provide some of the toughest applications and have helped to reveal the natural complexity of the seeming common sense behind the rules of thumb we apply to projects aimed at saving life and limb.

Despite its potential, CBA applications in health care still are less prevalent than one would expect. Difficulties in evaluating benefits, especially the value of life and improved quality of life, place limits on CBA and its usefulness to decision-makers. As a result, cost-effectiveness analysis using QALYs has emerged as an important tool for program evaluation.<sup>3</sup>

In 1993, the U.S. Public Health Service convened a panel of experts to help resolve controversies, standardize methods for cost-effectiveness studies, and develop consensus-based recommendations.<sup>4</sup> The panel recommended that these evaluative studies should take a societal perspective and that analysts should try to measure consequences through QALYs. Most importantly, the panel cautioned against decisions based solely on cost-effectiveness measures and noted that policy makers and analysts should examine equity and other considerations.

---

## Summary

1. Economic efficiency requires the maximization of total welfare, with the optimum quantity reflecting the sum of consumers' and producers' surpluses.
2. Unlike private decisions made in the marketplace, (social) cost-benefit analysis (CBA) involves evaluation of social benefits and social costs in public project analysis. Often, markets do not exist to evaluate the benefits and costs of such projects.
3. CBA rests on the principle that society's welfare will be improved whenever the benefits of a project exceed its costs.
4. CBA represents an example of marginal analysis. The social optimum is achieved when marginal social benefit is equal to marginal social cost.
5. Though in principle CBA appears simple, it can be difficult to apply. The difficulties include identifying all the relevant costs and benefits, including third-party effects, assigning monetary values, and making projections over many years for projects with long lives.
6. The monetary values of future net benefits must be discounted. The choice of the appropriate discount rate remains controversial.
7. Health care projects must often value human life. The human capital approach and the willingness-to-pay approach have been the most widely applied methods.
8. Cost-effectiveness analysis (CEA) can be used when it is difficult to place a monetary value on the benefits of a project. CEA is used to compare the costs under alternative projects of achieving some desired and quantifiable nonmonetary objective, such as the cost of detecting a case of cancer or the cost of a life-year saved.
9. Cost-utility analysis (CUA) is a special case of CEA in which the objective is measured in quality-adjusted life-years (QALYs) or some other indicator, such as disability-adjusted life-years (DALYs), that takes into account individuals' preferences for health.

---

## Discussion Questions

1. Would the concept of consumers' surplus be a sound welfare measure if the income distribution were deemed inequitable? If so, in what way?
2. In what ways is social cost-benefit analysis similar to a consumer's decision about allocating resources or to a firm's investment decision? In what ways is it different?
3. What external benefits or costs would you expect from a project designed to develop sanitary waste product disposal in a third-world village? Why do these need to be considered as part of a CBA of the project?
4. Why are jobs that are created as a result of a social project not normally considered as a benefit?

<sup>3</sup> Comprehensive bibliographies of CBA and CEA studies are found in Elixhauser (1993) and Elixhauser et al. (1998).

<sup>4</sup> Details are found in Gold et al. (1996) and a series of articles in the *Journal of the American Medical Association*.



5. Some suggest that a dollar value cannot be placed on life; that is, life is priceless. Explain how the dilemma to social decision created by this view might be resolved.
6. Does it matter whether a higher or lower discount rate is applied to the CBA of a social project? If so, why?
7. Discuss possible reasons why the estimates of the value of life presented in Table 4-1 differ so much.
8. How does the willingness-to-pay principle of welfare economics differ from the valuation of an extra life-year in applying QALYs?
9. Distinguish between cost-benefit analysis (CBA) and cost-effectiveness analysis (CEA). Can CEA replace CBA in all cases? If not, why not?
10. In the example of the ACE inhibitor, what decision issue is the marginal cost relevant to? The extra QALYs?
11. Consider the information provided in Box 4-1, “The Cost of Saving Lives.” If a society has a fixed budget that it can devote to all interventions, formulate a prioritizing rule that would save the greatest number of years of life for a given budget.

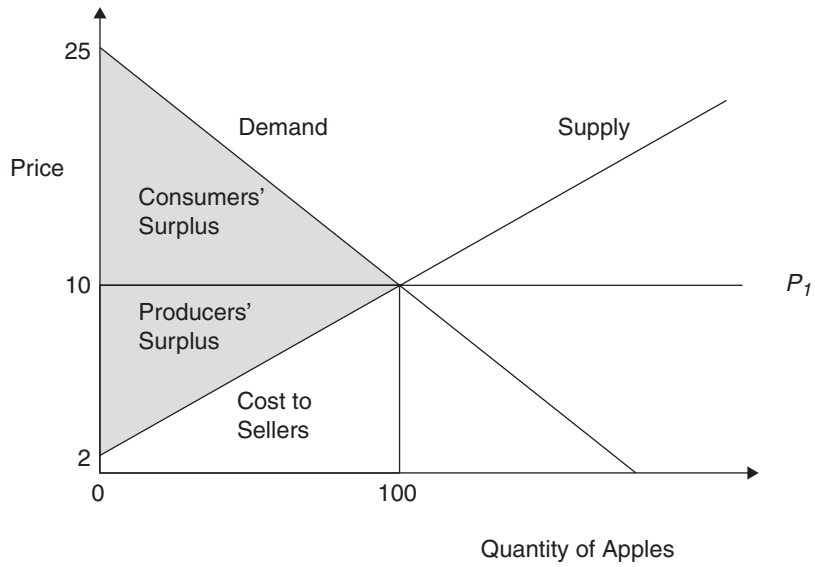
## Exercises

1. Using Figure 4-3, explain why a pollution abatement program that reduces discharge beyond  $Q_1$  is inefficient.
2. Consider the following two projects. Both have costs of \$5,000 in Year 1. Project 1 provides benefits of \$2,000 in each of the first four years only. The second provides benefits of \$2,000 for each of Years 6 to 10 only. Compute the net benefits using a discount rate of 6 percent. Repeat using a discount rate of 12 percent. What can you conclude from this exercise?
3. Consider the following table of costs and benefits from a governmental policy to clean the water in a local area.
  - (a) What level of abatement is most efficient by general economic criteria?
- (b) Would a 70-percent level of abatement pass a cost-benefit test? Is it efficient?
- (c) How would you respond to those who argue for 100-percent abatement?
4. Consider a project that costs \$10,000 today. It will provide benefits of \$4,000 at the end of Year 1, \$3,500 at the end of Year 2, and \$3,500 at the end of Year 3. If the discount rate is 6 percent, will this project be approved using cost-benefit analysis? Would your answer change if the discount rate is:
  - (a) 5 percent?
  - (b) 4 percent?
5. Consider a hypothetical three-stage screening test for a cancer with the following rates of detection and costs:

Level of Abatement	Total Costs	Total Benefits
0%	\$0	\$0
10	10	80
20	22	150
30	40	200
40	70	240
50	105	280
60	150	320
70	210	350
80	280	375
90	350	385
100	420	390

Stage	Number of Cases		Total Costs
	Detected		
1	100		\$200,000
2	105		260,000
3	106		300,000

- (a) Calculate the average cost per cancer detected in the three stages.
- (b) Calculate the marginal cost per cancer detected in the three stages.
- (c) Suppose that the marginal benefit per treated case is \$12,000 per person. What would be the optimal screening, given the costs?



6. Using the diagram, calculate
  - (a) Total consumer expenditures
  - (b) Total cost to sellers
  - (c) Total consumers' surplus
  - (d) Total producers' surplus
  - (e) The sum of the consumers' and the producers' surplus
  
7. Using the diagram, suppose that producers need to have licenses to sell apples, and that only 90 units of apples are licensed (i.e.,  $Q$  is limited to 90). Calculate
  - (a) the sum of the consumer surplus and producer surplus
  - (b) the reduction in consumer well-being because of the licensing

## APPENDIX—DISCOUNTING

Discounting is used in many applications involving comparisons of different streams of returns and/or costs over a number of periods. For example, an analyst might be asked to compare Investment A, which provides \$20 at the end of Year 1 and \$20 at the end of Year 2, with Investment B, which provides \$12 at the end of Year 1 and \$29 at the end of Year 2. Although Investment B returns \$41 over the 2 years compared to \$40 for Investment A, most of the return on Investment B comes later, at the end of Year 2. It is often necessary to compare the investments with a criterion that considers the timing of the returns. The analytical tool used for such comparisons is usually referred to as *discounting*.

Suppose George is offered the opportunity to buy a bond that will return \$1 one year from now. How much is he willing to pay now? George always has the option to keep his money and earn interest rate  $r$ . He will buy the bond if he can pay a price far enough below the \$1 return next year such that that price, multiplied by one plus the interest rate, equals \$1. Algebraically, if the rate of interest is  $r$  and the unknown amount is  $x_1$ , then:

$$x_1(1 + r) = 1$$

The value for  $x_1$ , then, is:

$$x_1 = \frac{1}{(1 + r)}$$

If the rate of interest is 5 percent (or 0.05), then  $x_1$ , the discounted value of \$1, one year hence, equals  $1/1.05$ , or 95.24 cents.

Similarly, the discounted value of \$1, two years hence, is:

$$x_2(1 + r)^2 = 1$$

The value for  $x_2$ , then, is:

$$x_2 = \frac{1}{(1 + r)^2}$$

If, again, the interest rate is 5 percent, then  $x_2$  equals  $1/1.05^2$ , or 90.70 cents.

Returning to the preceding example, we can calculate that the present value (the sum of  $x_1$  and  $x_2$ ), or *PV*, for Investment A will equal \$37.19, or  $(20/1.05 + 20/1.05^2)$ . The *PV* of Investment B is \$37.73. Thus, George will prefer Investment B.

In summation notation, the present value of a stream of returns  $R$  and costs  $C$ , over time, is:

$$PV = \sum_{t=1}^{t=T} \frac{(R_t - C_t)}{(1 + r)^t}$$

It is easy to demonstrate that the relative *PVs* of Investments A and B may depend on the interest rate (or in evaluation studies, sometimes referred to as discount rate  $d$ ) chosen for the analysis. If an interest rate of 15 percent were used instead of 5 percent in the previous example, the *PV* for Investment A would be \$32.51, compared to \$32.36 for Investment B. Intuitively the higher interest rate gives the larger but later return in Investment B less weight.

# Production of Health

- The Production Function of Health
- The Historical Role of Medicine and Health Care
- The Production Function of Health in the Modern Day
- How Does Health Care Affect Other Measures of Health?
- The Role of Schooling
- Conclusions

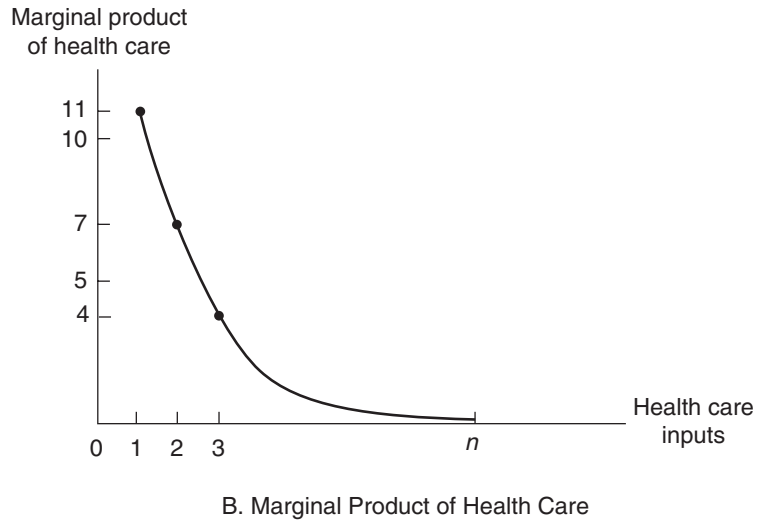
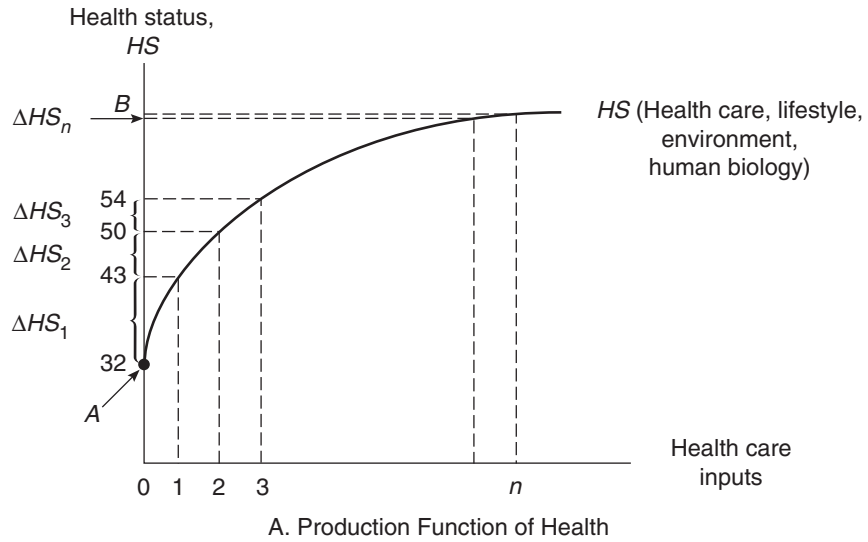
The production of health presents a central concern to the health economist and to public policy. Consider that the role of health care in society, including medical care provided by physicians, is ultimately a production question. We must learn about the determinants of health and about the contribution of health care. We can then better understand what decisions, both personal and public, will best produce health.

In medical terminology, this chapter addresses the efficacy and effectiveness of all those features of life, not only medical care, that plausibly contribute to our health. Unlike the typical doctor in practice, however, we look for evidence of the response of a “treatment” in the change in the health status of populations, as opposed to the treatment response of a medicine for the individual patient. We will see that the two approaches must remain in harmony and that both are fundamentally searches for causal relationships.

## THE PRODUCTION FUNCTION OF HEALTH

A production function summarizes the relationship between inputs and outputs. The study of the production of health function requires that we inquire about the relationship between health inputs and health. The answers that economists and medical historians offer to this question surprise many people. First, the contribution of practitioner-provided health care to the historical downward trends in population mortality rates was probably negligible at least until well into the twentieth century. Second, while the total contribution of health care is substantial in the modern day, its marginal contribution in some cases is small.

This distinction between total and marginal contributions is crucial to understanding these issues. To illustrate this distinction, consider Figure 5-1A, which exhibits a theoretical health status production function for the population. Set aside the difficulties of measuring health status in populations, and assume that we have defined an adequate health status (*HS*) measure. Health status here is an increasing function of health care. Also, to



**FIGURE 5-1** Production of Health

avoid a perspective that is too narrowly focused on health care, we specify further that health status depends at least upon the population’s biological endowment, environment, and lifestyle.<sup>1</sup> Thus,  $HS = HS$  (Health Care, Lifestyle, Environment, Human Biology). Improvements in any of these latter three factors will shift the curve upward.

A production function describes the relationship of flows of inputs and flows of outputs over a specified time period, so the inputs and output in Figure 5-1A are measured over an implied period, such as a year. In practice, we might use the number of healthy days experienced by the population per capita, mortality rates, or disability days, to indicate health status.

<sup>1</sup> This categorization of groups of inputs is not the only reasonable one, but it illustrates the main issues, and it has an excellent history. It was developed by the Canadian government for its pioneering work begun almost four decades ago (Lalonde, 1974).

To simplify the depiction, we have reduced all health care inputs into one scale called Health Care. In reality, health care consists of many health care inputs. Some of them include medical care provided by doctors of medicine or osteopathy, but other health care professionals also provide care. Conceptually, the health care measure,  $HC$ , may be thought of as an aggregate of all these types of health care, the aggregation being based on dollar values.

The marginal contribution of health care is its marginal product, meaning the increment to health caused by one extra unit of Health Care, holding all other inputs constant. Increasing Health Care from zero to one unit in Figure 5-1A improves health status by  $\Delta HS_1$ , the first unit's marginal product. Numerically, this first unit of Health Care has increased the health status index from 32 to 43;  $\Delta HS_1 = 11$  Health Status units. The next unit of medical care delivers a marginal product of  $\Delta HS_2 = 7$ , and so on.

These marginal products are diminishing in size, illustrating the law of diminishing marginal returns. If society employs a total of  $n$  units of Health Care, then the total contribution of Health Care is the sum of the marginal products of each of the  $n$  units. This total contribution as shown,  $AB$ , may be substantial. However, the marginal product of the  $n^{\text{th}}$  unit of medical care is  $\Delta HS_n$ , and it is small. In fact, we are nearly on the "flat of the curve." Marginal product is graphed on Figure 5-1B.

We have drawn the health production function as a rising curve that flattens out at higher levels of health care but never bends downward. Would the health production function eventually bend downward? Is it possible to get too much health care so that the health of the population is harmed? This is a logical possibility under at least two scenarios. *Iatrogenic* (meaning provider-caused) disease is an inevitable by-product of many medical interventions. For example, each surgery has its risks. Combinations of drugs may have unforeseen and adverse interactions. If the rate of iatrogenic disease does not fall while diminishing returns sets in, it is possible for the balance of help and harm from health care to be a net harm.

Medical scientists, such as Cochrane (1972), have pressed the case that much medical care as often practiced has only weak scientific basis, making iatrogenesis a real probability. Writing for the public audience, Dubos (1960) and Illich (1976) once warned of a medical "nemesis" taking away our abilities to face the natural hardships of life by "medicalizing" these problems. Illich argued that this medicalization would lead to less personal effort to preserve health and less personal determination to persevere; the result becomes a decline in the health of the population and thus a negative marginal product for medical care.<sup>2</sup>

Return to the distinction between total product and marginal product. Often, the marginals, rather than the totals, are relevant to policy propositions. For example, no one seriously recommends that society eliminate all health care spending. However, it is reasonable to ask whether society would be better off if it could reduce health care expenditures by \$1 billion and invest those funds in another productive use, such as housing, education, transportation, defense, or other consumption. We could even reasonably ask if health itself could be improved by transferring the marginal \$1 billion to environmental or lifestyle improvements.

Many of our government programs encourage health care use in certain population groups, such as the poor and elderly. Other programs, such as tax preferences for health insurance, provide benefits for those who are neither poor nor elderly and encourage their consumption of health care. The theoretical issues raised here suggest that we question the wisdom of each of our programs. The theoretical questions can be investigated with data of several kinds either directly or indirectly relevant to the production of health issue. We begin with the historical role of medicine, which indirectly bears on the issue of health production. After providing an overview of these efforts, largely the work of medical and economic historians, we then turn to econometric studies of the modern-day production function.

---

<sup>2</sup> The medicalization argument was reinforced by Thomas McKeown's (1976) research showing that medical care warranted little credit for the historical declines in mortality rates in England and Wales. The works by McKeown and by Cochrane remain the reference points for the continuing debate within medicine and the social sciences over the role of medicine (Alvarez-Dardet and Ruiz, 1993; Farmer and Nardell, 1998; Frenk, 1998).

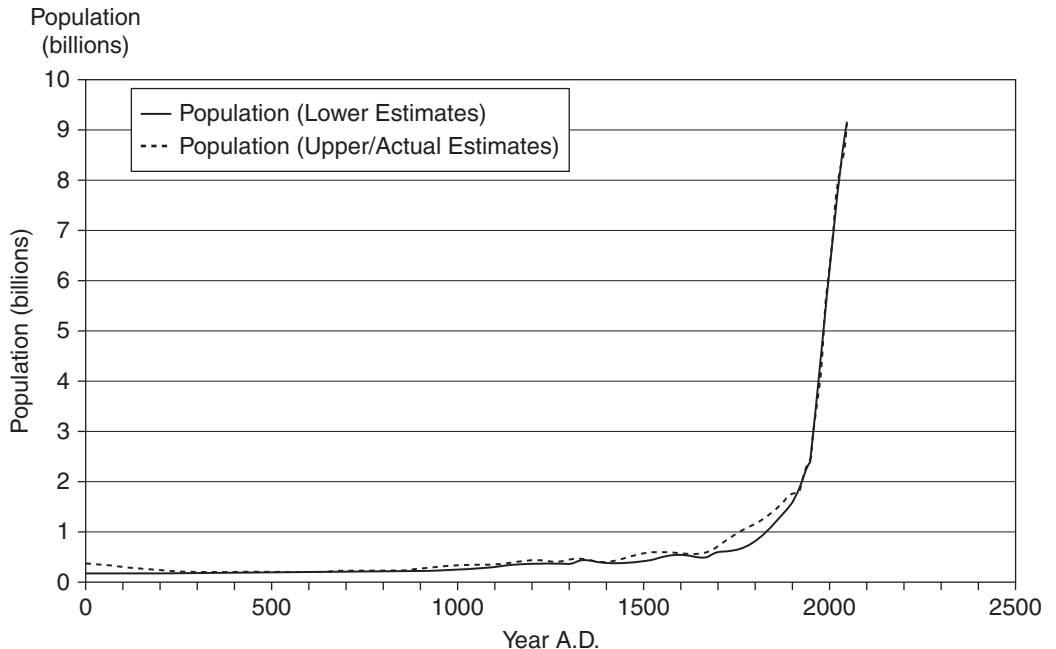
## THE HISTORICAL ROLE OF MEDICINE AND HEALTH CARE

Many medical historians agree that practitioner-provided medical interventions played only a small, perhaps negligible, role in the historical decline in population mortality rates. Effective medicine is a fairly recent phenomenon, and the delivery of effective medical interventions on a scale sufficient to affect population health indicators most likely appeared only well into the twentieth century. Though the magnitudes of other causes of mortality declines are still disputed, it is clear that a larger role, one of the most significant ones, might be attributed to public health measures and the spread of knowledge of the sources of disease. However, a number of scholars in this field attribute the largest share of the credit to improvements in environment, particularly to the greatly increased supply of foodstuffs that became available due to the agricultural and industrial revolutions.

### The Rising Population and the Role of Medicine

The notion that medicine played a relatively minor historical role is certainly not new, and it has been asserted by researchers of various ideologies. This point of view is associated with the work of Thomas McKeown (1976), who focused on the dramatic rise in population in England and Wales from 1750 to the modern day.

The pattern of world population growth, including population growth in England and Wales, has interested many scholars, including McKeown. World population is hard to estimate for the distant past, but research by the United Nations (1996) and others show that something extraordinary happened during the last 300 years. In the first century the population was roughly 300 million. For a thousand years thereafter, until the era of Viking ships, little or no change occurred. By the Age of Enlightenment, starting just before 1700, the population may have risen to 600 million. Then things began to change rapidly. Within a single century, the world population passed 1 billion people. The next 5 billion arrived within a mere 200 years. What had happened? Figure 5-2, based on United Nations data, reveals this startling pattern.



**FIGURE 5-2** World Population from 0 B.C.E. to the Modern Day *Source:* <http://www.ciese.org/curriculum/popgrowthproj/worldpop.html>. Data from U.S. Census Bureau.

Returning to the history of England and Wales, the large rise in their populations in the period following 1750 is to a large degree a story of the population's health. Population increase comes from increased birth rates, reduced mortality, or increased net in-migration. Migration was not an important source of population increase in England and Wales; when accurate birth rate and death rate data became available from 1841, these data alone proved able to account for the population change. Likewise, fertility probably did not account for the change because recorded birth rates have declined during the period since data have become available. Declines in birth rates are a common finding in countries undergoing industrialization and modernization. In contrast, recorded mortality rates did decline substantially.

McKeown began by investigating which diseases contributed to the decline in death rates. Mortality data are very limited prior to the mid-1800s, but the records revealed an emerging picture. Table 5-1 shows death rates by disease category for three time periods. The table shows that airborne infectious diseases account for the largest single portion of mortality reduction, and waterborne infectious diseases also make up a substantial portion of known causes. Regarding the airborne diseases, other data suggest that the main airborne diseases showing a decline in mortality include tuberculosis, bronchitis, pneumonia, and influenza.

### What Caused the Mortality Rate Declines? Was It Medicine?

Many presume that the declines in the mortality rates were due to improvements in medical science provided to the public through medical practice, but counterarguments to this proposition bring it into question. In most cases, an effective specific medical intervention was not available until late in the period, well after the greater part of the mortality decline had occurred.

The argument can be illustrated for the cases of respiratory tuberculosis and a group of three upper respiratory diseases—bronchitis, pneumonia, and influenza. Mortality rates for these diseases fell to relatively low levels prior to the availability of effective medical interventions, whose availability occurred respectively after 1930, and for some cases well into the 1950s and 1960s.

The argument can be illustrated for the cases of respiratory tuberculosis and a group of three upper respiratory diseases—bronchitis, pneumonia, and influenza. Mortality rates for these diseases fell

**TABLE 5-1** Death Rates\* (per million) in 1848–1854, 1901, and 1907

	1848–1854	1901	1907	Percentage of Reduction (1848–1854 to 1907) Attributable to Each Category	For Each Category, Percentage of Reduction (1848–1854 to 1907) That Occurred Before 1901
I. Conditions attributable to microorganisms:					
1. Airborne diseases	7,259	5,122	619	39	32
2. Water- and foodborne diseases	3,562	1,931	35	21	46
3. Other conditions	2,144	1,415	60	12	35
Total	12,965	8,468	714	72	37
II. Conditions not attributable to microorganisms	8,891	8,490	4,070	28	8
All diseases	21,856	16,958	5,384	100	29

\*Standardized to the age/sex distribution of the 1901 population.

Source: reprinted from McKeown, Thomas, *The Modern Rise of Population*. New York: Academic Press. 1976 (p. 54).

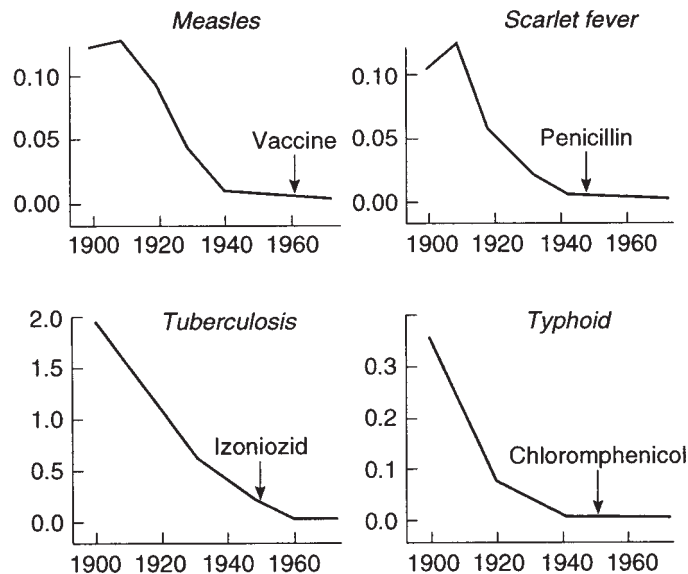


**BOX 5-1****Tuberculosis and *The Magic Mountain***

Thanks to the efforts of writers such as Nobelist Thomas Mann, the tuberculosis (TB) sanitarium of 100 years ago has found a permanent place in literature. Mann's novel *The Magic Mountain* describes with a mastery of medical detail, often admired by physicians, the characters' struggle with the deadly disease. Mann's incomparable development of character and dialog bring a reality to the society unique to this sanitarium, representing one he had himself visited. Hans Castorp, a young German engineer, is smitten with Clavida Chauchat, a young woman at the "good" Russian table. The Italian Settembrini's intellectual arguments with Naptha can result only in a duel. We grieve when Hans's cousin Joachim, a good, simple military man, succumbs to TB and his body is sledged down the mountain. Hans survives to return to the lands below.

Though effective chemical interventions were not available until after 1940, doctors did treat TB prior to 1940, notably with the widespread use of sanatoria such as the one in *The Magic Mountain*. Declines in TB mortality during the period studied by McKeown represented perhaps the most important example of declines in the mortality rate. Can we credit the sanatoria for this progress? This is unlikely, because it is unlikely that their capacity was ever large enough to affect the pattern of mortality in populations. For an account of the retreat of tuberculosis, see Smith (1988).

to relatively low levels prior to the availability of effective medical interventions, whose availability occurred respectively after 1930, and for some cases well into the 1950s and 1960s. The picture is shared by waterborne diseases. About 95 percent of the mortality declines in cholera, diarrhea, and dysentery occurred prior to the 1930s, when intravenous therapies became available. Likewise, typhoid and typhus mortality already had fallen to low levels by the beginning of the twentieth century. The pattern McKeown found for England and Wales also can be illustrated for the United States. McKinlay and McKinlay (1977) provided data for the United States from 1900 to 1973. Figure 5-3 shows these



**FIGURE 5-3** Fall in the Standardized Death Rate per 1,000 Population for Four Common Infectious Diseases in Relation to Specific Medical Measures for the United States Source: Reprinted from *Milbank Memorial Fund Quarterly/Health and Society*, John B. Mckinlay and Sonja M. Mckinlay, "The Questionable Contribution of Medical Measures to the Decline of Mortality in the United States in the Twentieth Century," *Milbank Memorial Fund Quarterly/Health and Society* 55 (1977): 405–428, with the permission of Blackwell Publishers.

patterns for several infectious diseases. In most cases, as is shown, the availability of the effective medical intervention occurs well after the majority of the mortality declines.

One of the most important changes in mortality in the twentieth century was the decline in infant mortality. Does this type of mortality follow the same pattern? A highly readable account of the modern historical pattern of infant mortality is offered in Victor Fuchs's *Who Shall Live?* (1975). Fuchs noted that infant mortality rates in New York City improved markedly from 1900 to 1930 and that this decline was due to declines in deaths from "pneumonia-diarrhea" complex. Fuchs concluded: "It is important to realize that medical care played almost no role in this decline. While we do not know the precise causes, it is believed that rising living standards, the spread of literacy and education, and a substantial fall in the birth rate all played a part" (p. 32).

Antimicrobial drugs were introduced in the 1930s. Between 1935 and 1950, the fall in infant death rates accelerated. Fuchs proposed that medical advances and rising living standards both contributed to the reduction in infant deaths during this period. Declines in infant deaths flattened somewhat beginning about 1950 but resumed a stronger decline about 1965. If specific effective curative medicines were not largely responsible for mortality declines, is it nevertheless possible that other tools in the physician's black bag were effective? Unfortunately, this too is unlikely. The problem is that there probably were few effective tools available until well into the twentieth century. Even a clear knowledge of what caused disease was not widespread until the 1900s.

**NUTRITION REDUCED MORTALITY** Two of the most respected students of the mortality decline, medical historian Thomas McKeown (1976) and economic historian Robert Fogel (2004), argued strongly that the main cause was improved nutrition. McKeown reasoned by process of elimination. As we have just seen, he showed the medicine interventions could not have been the cause, a claim that is still widely accepted. He considered other possibilities one by one. For example, some have suggested that perhaps the infectious organisms had spontaneously mutated and became harmless; he pointed out that the chances were remote that so many independent organisms had randomly mutated at about the same time.

McKeown also dismissed public health as a major cause, however, and this argument was to become controversial. If we re-examine his work in Table 5-1, we see that the largest portion of mortality decline from 1848 to 1971 was due to declines in mortality from airborne diseases. He argued that public health projects, which focused on improving water quality and the safety of food, could have little effect on airborne diseases. McKeown clearly understood that clean water and pasteurized milk were important to improved health, but he claimed that these benefits came late in the historical era of mortality declines. Supporting his claim about the timing of public health, consider that the role of germs was not understood until the mid-1800s, about the time that public health came into being, and pasteurization of milk did not start until around 1870 and its widespread commercial use did not come until well into the twentieth century. Having eliminated everything else, in his reasoning McKeown assumed that the great benefactor that transformed the developed countries from high mortality to low mortality must have been improved nutrition.

This argument for the primacy of nutrition, however, provided no direct evidence that nutrition improves health. Robert Fogel (2004) provided that needed evidence. He established that after the mid-eighteenth century, calorie intake of Europeans increased tremendously. At about the same time, their average height also increased substantially. The relationship of height to health is now well known; the Waaler Curve established that, for any given body mass, taller people (up to a point) have greater life expectancy (Fogel, 2004). We also now understand how better nutrition makes an individual better able to resist infectious disease. Fogel went on to study in great detail the heights and records of Civil War soldiers in the United States. His research led him to claim that nutrition played the major role in what the title of his recent book calls: *The Escape from Hunger and Premature Death, 1700–2100*.

**PUBLIC HEALTH REDUCED MORTALITY** Other historical analysts take issue with the proposition that nutrition was the main cause of the mortality reductions. The crux of the issue is when the era of mortality reductions began. Public health advocates claim, contrary to McKeown, that the major

declines did not start until around 1870, and if they began this late, then public health, which began about 1850, would have come in time to contribute. We know that the era from 1870 to about 1940 completed the “epidemiological transition.” This phrase describes the remarkable transition in developed countries from when infectious disease was the major cause of death to a time when it became of only minor importance to population health. It is instructive to examine what public health accomplished during this period.

By 1870, cities had grown rapidly without the planning and development we now consider essential to a healthy environment. During this era, urban centers eventually and painfully slowly overcame their status of having worse mortality rates and general health than the countryside, the “urban deficit.” Streets contained animal excrement, sewer systems were designed mainly for storm water, and water supplies were often delivered in lead pipes. The transition from water tainted with infectious organisms to clean water supplies was the most dramatic change in the health environments of city dwellers (see Box 5-2, “The Importance of Clean Water,” for this story).

To summarize, the period from 1750 to the present contains three strands of health-related phenomena: (1) growth in life expectancy; (2) improved nutrition; and (3) improved public health. The difficulty is how to sort out which relationships proved most important. We see the importance of nutrition to body mass and height, the keys to health in the Waaler Curve (Fogel, 2004). This is compelling evidence. However those who believe that public health was of key importance can point to the clean-up of cities, also compelling evidence. They point out, for example, that we know the modern Chinese people are not as tall as Americans, yet their life expectancies are much higher than would have been expected. Does not this point to the modern adoption of public health measures, which now can take place very rapidly? We will see shortly that these historical puzzles are not merely of “academic” interest but are critical for the growth in well-being of the lesser developed world.

## BOX 5-2

### The Importance of Clean Water

If transported by time machine back to the mid-nineteenth century, you would find it difficult to survive. This is because your modern body mass and height could barely be sustained by the small average quantities of available calories. But if you did survive and went to live in a city, you would find that in your weakened condition you would be very susceptible to infectious disease organisms permeating your environment, and especially in the water.

Even in 1900, waterborne infectious disease accounted for one-quarter of the deaths from infectious disease. Public health campaigns, which were painfully slow in gaining acceptance, cleaned up the water. They introduced the filtering of city water through sand. They fought to have sewage discharged at a safe distance from water intakes. In prior cases, cities had discharged waste directly into the same lakes or streams that provided drinking water. Water closets were introduced in about 1870, and these discharged human waste into a city sewer system that often could not handle it and overflowed even into the streets. Public health also introduced chlorination of the water supplies. If the earlier contaminating practices seem obvious and foolish to us, we need to remember that germ theory had only recently arrived, and pasteurization was discovered only in the late 1800s.

Cutler and Miller (2005) estimate that filtration of city water brought reductions in total mortality of 16 percent, and reduced infant mortality by 43 percent, in the 12 American cities studied. Applying cost-benefit principles, the researchers found that the ratio of benefits to costs in the filtration projects was about 23 to 1. This is history to us, but it is present-day reality to less developed countries, where over 1 billion people lack access to clean water (Cutler and Miller 2005). The United Nations has declared the 2005–2015 period the International Decade for Action on Water. More on the U.N. program can be found on the Web by searching for “millennium development goals.”

## What Lessons Are Learned from the Medical Historian?

We cannot conclude that medical research is unimportant in history or in the present day. Medical research contributes not only through improvements to medical practice, but also through its influence on health-enhancing practices. Typhoid provides a good example. Mortality from typhoid declined substantially well before the arrival by 1950 of chloramphenicol. Medical research, however, contributed to our understanding of the cause and transmission of typhoid and generated public health measures such as filtering public water supplies, chlorination of water supplies, and establishment of drinking-water standards. All of these factors occurred historically in time to have a major effect on mortality. Selma Mushkin (1979) estimates that medical research accounted for almost one-third of the cost savings to society from reduction in sickness or death rates in the United States from 1900 to 1975. The period 1900 to 1930 accounts for half the value of medical research effects, even though it came largely before the specific effective medical practice interventions.

Investments in medical research play a major role in our health and well-being. Murphy and Topel (2005) used people's willingness to pay for advances in medical knowledge as a measure of its value. They estimated that the contribution of medical research to mortality reductions from 1970 to 2000 added \$3.2 trillion to national wealth. Since 1970, reductions in heart disease mortality alone have been worth about \$1.5 trillion. The medical research share of these gains, even if this amounted to only 10 percent of the total, easily compensates for the \$36 billion we invest in it annually.

Second, perhaps the best result of this overview is a healthy skepticism toward the effectiveness of any given medical practice, and more importantly, to its significance and benefit to the population. It is in this spirit that the U.S. government has increasingly come to fund outcome studies. Outcome studies seek to address the effectiveness and appropriateness of specific medical practices on patient outcomes. The studies attempt to reduce the prevalent uncertainties in medical practice, and they offer important inquiries into the wisdom of using the marginal billion dollars on medical care delivery, particularly in terms of costs and benefits to the population as a whole. In the same spirit, "evidence-based medicine" aims to close the gap between outcomes research and physician practice.

Finally, and most importantly, these historical puzzles have relevance to the progress and public investment practices of lesser developed countries, who have scarce resources with which to invest in either industrial growth or to invest directly into health measures and public health improvements. If improved nutrition is the key to population health, then perhaps industrial growth will bring the best overall gains. If public health is the most direct and productive way to achieve health, it is also a route to develop a more productive workforce. These questions easily merit diligent research. We note also that there is also an immense quasi-natural experiment to observe. Fogel (2000) and others report that world poverty rates are dropping sharply; the number of people in extreme poverty, those living on only a \$1 a day, has dropped sharply in the last two decades. We will want to see population health status in the developing countries improving.

## THE PRODUCTION OF HEALTH IN THE MODERN DAY

The investigation of the modern health production function requires econometric techniques. An understanding of the strengths and limitations of these contributions requires attention to the underlying conceptual issues.

### Preliminary Issues

Two conceptual issues bear on our interpretation of the results. These two issues can be posed as questions faced by every researcher: (1) how to measure health, the dependent variable in these studies, and (2) how to eliminate biases in the estimates.

**HOW TO MEASURE HEALTH** Consider the measurement of health. We desire a measure of population health status that captures the aspects of health status that are meaningful and that we can measure with adequate precision. It is difficult to attain both of these goals. We are most confident in the accuracy of mortality rate data, but mortality rates do not adequately capture several meaningful aspects of health status, such as reduction in pain and suffering and other improvements in the quality of life. The approach of past research in this field is understandable. Researchers have emphasized mortality data because of their accuracy, as well as because of their importance in the public mind. However analysts have used other indicators of health status, such as morbidity (illness) rates and disability days. In discussing the empirical literature, we begin with the mortality studies and then consider studies of morbidity.

**ELIMINATING BIASES—REDUCED FORM VERSUS STRUCTURAL EQUATIONS** Consider that the statistical estimates often rely on *reduced form* equations, which are practical representations of the underlying true model of the phenomenon. The true model is based on what are called *structural* equations. Estimation based on the reduced form equations can lead to misinterpretations.<sup>3</sup>

### The Contribution of Health Care to Population Health: The Modern Era

Health economists inevitably use different study designs and data sources to estimate the marginal product of health care. The resulting numbers need to be converted to a common basis, and for this we use the elasticity of health with respect to expenditure on health care inputs:

$$E_{Health:Expenditure} = \frac{\% \text{ change in health}}{\% \text{ change in health care expenditures}}$$

Table 5-2 reports the elasticities from several studies of the production of health. Each study applies econometric methods to analyze survey data; these range from statewide data to data on county groups (Hadley, 1982, 1988) to data on individuals (Sickles and Yazbeck, 1998). Though not shown in the table, recent investigations (Cremieux, Oulette, and Pilon, 1999) confirm a significant contribution of health spending to reduce infant mortality (in Canada).

The several studies offer insights because of their differences. Some suggest that lifestyle and environment expenditures could provide more benefits per dollar of cost than health care; and efforts to improve schooling or reduce cigarette smoking offer appealing trade-offs with health care spending. The Hadley studies helped establish that health care spending makes a statistically significant contribution to health and argues that health care passes benefit-cost criteria at the margin. The latest study follows its subjects through time.

**TABLE 5-2 Measuring the Contribution of Health Care to Population Health**

Study Cited	Date of Study	Health Care Elasticity*	Significant?
Hadley	1982	0.12 to 0.17	Yes
Hadley	1988	0.20 to 1.00	Yes
Sickles and Yazbeck	1998	0.03 to 0.05	Yes

\*This is the elasticity of health with respect to health care expenditure.

<sup>3</sup> For further discussion of this issue and method, see Rosenzweig and Schultz (1983); Grossman and Joyce (1990); and Atkinson and Crocker (1992).

A study with both quality and novelty of design adds support to our finding that medical expenditures improve health (Almond, et al. 2010). Low birth-weight babies have a worse chance of surviving, and with “very low birth weight” defined as below 1500 grams. It will seem odd, but babies just below that cutoff had one percent *lower* mortality rates than babies just above it. The reason was that “very low birth weight” qualifies the baby for special medical treatment (with extra costs). The authors put these facts together finding that the marginal product of the extra care significantly contributed to health. They further concluded that the cost of saving a statistical life of a newborn with birth weight near 1500 grams is on the order of \$550,000 in 2006 dollars (Almond et al., 2010). Based on value of life estimates (see Chapter 4), these medical efforts are clearly worth it.

### Is Health Care Worth It?

Are we, then, literally on the “flat of the curve”, getting zero marginal product from the marginal investment in health care? The answer is clearly no. Murphy and Topel (2005), as we have seen, estimate that the American gains in life expectancy from 1970 to 2000, based on willingness-to-pay principles, are worth over \$3 trillion. We invest a small fraction of that in health technology research. Much of the gain is probably due to health care. Similarly Cutler (2004) ties 40 percent of the gains in life expectancy between 1950 and 2000 to medical improvements in just two categories: low-weight infant care and cardiovascular disease treatment. Again, based on willingness to pay, the life gains are worth the extra costs of medical care overall—not even counting the benefits from other forms of health care.

Look again at the marginal benefit of health care (technology held constant), which is small but not zero. We earlier suggested that the health production elasticity of health care on the margin is about 0.10. To illustrate the meaning of this elasticity, suppose that 0.10 were the true production elasticity of health care, and let Congress reallocate \$250 billion from other programs to health care expenditure. In 2011, the United States spent \$2.5 trillion on health care. The \$250 billion transfer would increase health care expenditure by 10 percent. We would extrapolate the improvement in health to be  $10 \times 0.10 = 1$  percent. If we define “health” by average life expectancy, a 1-percent gain would mean an increased average life expectancy of  $78 \times 0.01 = 0.78$  year, provided this increased expenditure continues indefinitely. Spread over the population (for a life span), however, the 0.78 year of life could lead to incremental benefits that exceed the incremental costs.

#### BOX 5-3

##### Sulfa: A Drug That Really Changed Things

At a time when cost-control planners seek to sort out medicine that is expensive but not very effective, sulfa is a good example to remind us that there are medicines and practices that work extraordinarily well. It was the miracle drug of the 1930s, made available in 1937 several years before penicillin appeared. Sulfa was a major player in the epidemiological transition from a time when infectious diseases were the fearsome killer to modern times when non-communicable diseases predominate: The following example provides an excellent way to measure sulfa’s effect.

Sulfa was not patented, appeared suddenly, and was distributed rapidly. These facts made the case ideal for study by health economists Jayachandran, Lleras-Muney, and Smith (2010). They tested population health effects of sulfa by comparing the time patterns of those mortality rates of diseases treatable by sulfa with those that weren’t. In most of these cases, the results were dramatic. Sulfa caused sharp reductions in mortality for treatable diseases such as MMR (measles, mumps and rubella), pneumonia, scarlet fever and maternal mortality. The drug lowered maternal mortality from 60 percent to 36 percent and caused similar or higher drops in the other sulfa-treatable diseases. The most telling statistic is that, by itself, sulfa raised U.S. life expectancy by 0.4 to 0.7 years.

## Prenatal Care

The importance of examining population subgroups comes from neonate mortality studies (for example, see Corman and Grossman, 1985; and Corman, Joyce, and Grossman, 1987). A neonate is an infant one month old or younger. Thus, the neonate mortality rate refers to deaths to neonates per 1,000 live births. How can we reduce these deaths? The early studies observed counties in the United States and identified several factors that seemed to cause higher neonate mortality rates. Table 5-3 displays the production of neonate mortality study by Corman, Joyce, and Grossman (1987).

The table presents their estimates of the contribution of each measured factor to the reported mortality rate decline. The WIC program is the shortened name for the governmental program designed to provide improved nutrition for women, infants, and children; it is a means-tested program, meaning that it is directed to the poor. The BCHS variable is the authors' measure combining various Bureau of Community Health Services projects, including maternal and infant care, as well as community health centers.

The data recorded in Table 5-3 indicate that of the total reduction in neonate mortality for whites during the period, 1.9 deaths per 1,000 live births, or 25.3 percent, can be explained by the observed factors, whereas for blacks a greater amount, 6.5 or 56.5 percent, can be explained. Blacks benefit more from health care on the margin, a finding that has been found in other studies. Interpret the table as follows: During the period studied, the WIC program resulted in a reduction of white neonate mortality rates of 0.425 deaths per 1,000 live births, while for blacks WIC reduced neonate mortality by 1.330 deaths per 1,000 live births.<sup>4</sup> Note too, that abortion tends to lower mortality, this result may reflect that many abortions were of fetuses that would not have survived infancy. Also note that prenatal care is effective in these data, especially for blacks; prenatal care shows up well in most studies.

A novel experiment (Evans and Lien, 2005) reinforces the value of prenatal care. When public transportation workers in Allegheny County, Pennsylvania, went on strike in 1992, prenatal visits

**TABLE 5-3 Contribution of Selected Factors to Reductions in Neonate Mortality Rates, 1964–1977**

Factor	Whites Total Effect	Blacks Total Effect
Organized family planning	0.084	0.526
WIC	0.425	1.330
BCHS	0.002	0.030
Neonatal intensive care	0.140	0.534
Abortion	0.824	2.109
Prenatal care	0.434	1.949
Total explained reduction	1.9	6.5
Total reduction	7.5	11.5
Percentage explained	25.3	56.5

*Note:* Figures record estimates of the reduction in deaths per 1,000 live births predicted to have been caused by various factors.

*Source:* Reprinted from Journal of Human Resources, Hope Corman, Theodore J. Joyce, and Michael Grossman, "Birth Outcome Production Function in the United States," *Journal of Human Resources* 22 (1987): 339–360, with permission from The University of Wisconsin Press.

<sup>4</sup> Currie and Gruber (1996) showed general Medicaid eligibility to improve birth outcomes in a study across states in the United States. (This is distinguished from the particular Medicaid spending directed to prenatal care.) These authors, however, raised issues about the cost-effectiveness of improvements in Medicaid eligibility.

among the poor were sharply curtailed. The results were significant reductions in the quality of birth outcomes, especially for those affected during the early stages of pregnancy.

Aizer, Currie, and Moretti (2007) observed the value of prenatal care indirectly. They measured the effects of a law transferring some California Medicaid patients to managed care, which in this situation offered distinctively lower quality prenatal care. Significantly greater numbers of low-weight births were observed. Another study, by Conway and Kutinova (2006), finds prenatal care to be effective in reducing the probability of low-weight births.

### **The World's Pharmacies**

Note also the contribution of pharmaceutical availability in determining population health. Recent research (Shaw, Horrace, and Vogel, 2005; Miller and Frech, 2004) finds that countries in the Organization for Economic Cooperation and Development (OECD) with higher drug consumption have greater life expectancies. The magnitude of the effect is on a par with the effects of reduced cigarette consumption and increases of fruit and vegetables in the diet.

## **HOW DOES HEALTH CARE AFFECT OTHER MEASURES OF HEALTH?**

Mortality rates are part of the story, but only a part, and for many, not even the most important part. Other measures include morbidity data, disability days, work loss days, and other indicators. Research available also measures the quality of life.

The RAND Health Insurance Experiment (RHIE) is one of the largest randomly controlled economic experiments ever conducted. It was designed to test the effect of alternative health insurance policies on the demand for health care and on the health status of a large and closely observed group of people from all walks of life.

RAND researchers discovered that the greater the portion of the health care bill that individuals are required to pay, the less health care they choose to purchase. While this should not have been surprising, what did surprise most health economists was how great the difference was; the fully insured purchased roughly 40 percent more health care than those who had to pay their own bills. This provides an opportunity to ask whether those with 40 percent more health care were also 40 percent more healthy. This was serendipitous from a research standpoint, but a perfectly valid way to test the real contribution of health care to people's health within the context of a scientifically controlled experiment.

Fortunately, RHIE analysts kept detailed records on each person, including a dozen or more measurements under each category of physical health, mental health, social health, and general health index. They also examined their subjects' dental health, persistence of symptoms, health habits, and disability days. The results are easy to summarize. For dozens of items, virtually no differences were found between the groups studied; health care and health insurance did not seem to matter.

A simple example from the RHIE illustrates the point. Table 5-4 provides detail on work-loss days per employed person per year—a measure of health status and morbidity that some economic researchers like to use because it ties directly to both health and productivity. This table separates the RAND subjects into four groups, which differ by type of health insurance policy. Some subjects pay nothing out of pocket for their health care/health insurance package; some pay 25 percent to 50 percent of their bill themselves; others pay all of their health care bills up to a certain amount, called a deductible. The subject's out-of-pocket cost ranges from zero (free) to about 95 percent of the bill. Newhouse et al. (1993) summarize: "Our results show that the 40 percent increase in services on the free-care plan had little or no effect on health status for the average adult."

The effects on children showed a somewhat similar pattern. Valdez et al. (1985) examined data for 1,844 children in the RAND study—children who differed primarily by the type of insurance plan



**TABLE 5-4 Work Loss Days per Employed Person per Year, by Plan**

Plan	Mean	Standard Error of Mean	95% Confidence Interval	Number of Persons
Free	5.47	0.42	4.65–6.29	1,136
Intermediate (25%, 50%)	4.82	0.37	4.09–5.55	983
Individual Deductible	4.54	0.36	3.83–5.25	787
Family Deductible (95%)	4.82	0.53	3.78–5.86	600

Source: Reprinted by permission of the publisher from *Free for All? Lessons from the RAND Health Insurance Experiment* by Joseph P. Newhouse et al., Cambridge, MA: Harvard University Press, 1993. Copyright © 1993 by the RAND Corporation.

their families obtained. Children under the cost-sharing plans consumed up to one-third less care. However, the reduction in care was not significantly related to health status measures.

It may seem from the RAND results that public provision of health insurance to both adults and children might not be justifiable on the basis of benefits to health. However, as Jonathan Gruber (2008) points out, this conclusion does not follow. No one in the RAND Experiment was “uninsured,” completely without insurance, as are close to 50 million Americans as of this writing (the Patient Protection and Affordable Care Act begins to address them in 2014). The least insured individuals studied by RAND had full coverage for health expenditures above a deductible, which was \$1,000. Studies of the truly uninsured began to appear showing significant health gains from the provision of public insurance (Currie and Gruber, 1996; Doyle, 2005; Hanratty, 1996). These studies report reductions in infant and neonate deaths of around 5 to 10 percent. The Institute of Medicine estimates suggest that even larger gains are possible; they claim that the uninsured face a 25 percent greater mortality risks.

Gruber further explains why these studies do not conflict with RAND. He proposes that the marginal effectiveness of medical expenditures is quite high for the first expenditures but then drops off precipitously, a plausible pattern given diminishing marginal returns.

Eventually, additional spending does no good and the effectiveness curve flattens out . . . This appears to be the case as we move from less to more generous coverage, as in the RAND Health Insurances Experiment (Gruber, 2008: 584).

### **On the Importance of Lifestyle and Environment**

Didn't we always know that much of our health depends on the wisdom of our own choices? The role of lifestyle was best illustrated by Victor Fuchs in his book *Who Shall Live?* (1995). He compared average death rates in Nevada and Utah for 1959 to 1961 and 1966 to 1968. These two states are contiguous, and they share “about the same levels of income and medical care and are alike in many other respects” (p. 52). Nevertheless, average death rates in Nevada were greater than those in Utah. Table 5-5 shows the results of Fuchs's work. Fuchs argued that the explanation for these substantial differences surely lies in lifestyle:

Utah was, and remains, inhabited primarily by Mormons, whose influence is strong throughout the state. Devout Mormons do not use tobacco or alcohol and in general lead stable, quiet lives. Nevada, on the other hand, is a state with high rates of cigarette and alcohol consumption and very high indexes of marital and geographical instability. (p. 53)

In 2009, Utah, with its low age-adjusted death rates, was still a national leader in health (this death rate equaled 507.8), while Hawaii (717.9), and Nevada (727.3) were much higher, but

**TABLE 5-5 Excess of Death Rates in Nevada Compared with Utah, Average for 1959–1961 and 1966–1968**

Age Group	Males (%)	Females (%)
Less than 1	42	35
1–19	16	26
20–39	44	42
40–49	54	69
50–59	38	28
60–69	26	17
70–79	20	6

*Source:* Reprinted from Victor R. Fuchs, *Who Shall Live? Health, Economics, and Social Choice*, Expanded Edition, Singapore: World Scientific Publishing Company, 1995, p. 52, with permission from the author and World Scientific Publishing.

significantly lower than the national average (793.7). Before concluding that a simple life and plenty of sun are the tickets to good health in and of themselves, consider that many of the top 10 healthful states, while they may be sunny, are known to be chilly: Minnesota (718.6), New Hampshire (761.6), Idaho (774.5), and Colorado (620.3). (Data source: CDC, Preliminary death rates, 2009.)

### **Cigarettes, Exercise, and a Good Night’s Sleep**

Many have chosen to quit smoking (or to avoid becoming addicted to cigarette smoking in the first place). Americans know that heart disease and cancer are the two leading killers, but most do not realize how substantial a part smoking plays. Using the category “malignant neoplasms of the respiratory system” (the category for lung cancer), we find that the 2002 death rate (51.5) is twice as high as that for any of the following: breast cancer (13.4), prostate cancer (9.2), pneumonia and influenza (17.5), diabetes mellitus (22.3), HIV (3.1), or motor vehicle–related injuries (11.8).<sup>5</sup> We already have seen the negative health production elasticity of cigarettes, which makes it clear that cigarette smoking affects the average health of the community and is statistically significant at that level.

However, economics searches for underlying causes, and human behavior can have many interwoven causes. For example, smoking and other lifestyle behaviors may themselves be determined by unobserved variables that affect health status. This common problem in economic empirical work has been addressed in recent research (Balía and Jones, 2008; Contoyannis and Jones, 2004). They address the problem by estimating both the determinants of lifestyle behaviors as well as the determinants of health status, giving a clearer picture of the importance of lifestyle. The authors showed that a good night’s sleep, avoiding smoking, and regular exercise each contribute importantly to self-reported health.

While smoking certainly causes ill health, it is pleasurable as well, and one’s degree of health can affect the decision to quit. For example, a healthy individual may be more likely to quit as a preventive measure; on the other hand, a critically ill individual may quit as a curative measure (Jones, 1996). Folland (2006) shows that greater life satisfaction means being less willing to risk death by smoking.

<sup>5</sup> These death rates are age-adjusted deaths per 100,000 resident population, National Vital Statistics System, 2009.

Granted that lifestyle is a major player in health comparisons between individuals, it is natural to ask whether it plays the same role when comparing countries. As we have seen earlier in the course, life expectancy in America is lower than in many developed countries. Commanor, Frech, and Miller (2006) investigated this question. They began by assessing U.S. efficiency in the production of health, finding it to be somewhat less efficient than other developed countries. What is most relevant to our present discussion is their finding that much of the U.S. deficit stems from the higher rates of obesity in the United States.

### **The Family as Producer of Health**

Women have long been warned to avoid cigarettes and alcohol while pregnant. Are such lifestyle factors important enough to be included as factors in the production of newborn health? The answer is yes. In the production of newborn birth weight (an important guide to infant health outcomes), maternal cigarette smoking has a significant negative effect (Rosenzweig and Schultz, 1983; Rosenzweig and Wolpin, 1995). Data on maternal smoking now show that taxing cigarettes leads to improved birth outcomes via its effect on smoking behaviors of expectant mothers (Evans and Ringel, 1999).

Maternal behavior also can have strong and tragic consequences in the case of drug use. Joyce, Racine, and Mocan (1992) found that the alarming increase in low-birth-weight births in New York City, particularly among blacks, was due in large part to an epidemic of illicit substance abuse by pregnant women. The explosion of cocaine use had horrible consequences for these babies.

Looking at this at a more abstract level, a study from Sweden (Bolin, Jacobsen, and Lindgren, 2002) develops the theory of how parents make health investments in themselves and their children. If parents individually make these health investment decisions strategically—that is, in response to the expected decisions of the others—the decisions, together, will not be optimal for the family. Even more significant health investment problems will occur, they warn, when parents split up in divorce, because the non-caregiver may lose some incentives to invest in the child's health. The parent's incentive to invest in the children's health is clearly a critical factor in child health.

### **Social Capital and Health**

Recent research has made it clear that family, friends, and community are associated with the health of the individual and the community. The networks of social contacts of an individual or the complex overlapping networks in a community have come to be called *social capital*. The effects, first described by political scientists, sociologists, medical researchers, and epidemiologists, suggest that social capital beneficially affects measures of health (see Islam et al, 2006, for a review).

Social capital may improve an individual's health in several ways: (1) it may relieve stress to have the support of more social contacts; (2) more contacts can provide additional information on healthful behaviors and health purchases; and (3) satisfying social relationships may provide reasons to re-evaluate risky health behaviors. This issue presents complex research obstacles; for example, not all social contacts are beneficial.

The bigger issue, however, is how to determine whether social capital in these studies causes better health or alternatively whether it is a result of some other factors. This is an important avenue by which economics and its econometric tools provide benefits to the ongoing research of other disciplines.

Health economists have taken interest in this area as a potential subject area in which to make a joint contribution with other disciplines. This is occurring both in theory, and in empirical work.<sup>6</sup> Findings generally support the hypothesis that social capital improvements lead to health improvements.

---

<sup>6</sup> Folland (2006, 2008), Folland, Kaarboe, and Islam (2011).

## Environmental Pollution

Pollution causes ill health and death in individuals, with the elderly and people with respiratory diseases more susceptible. The degree to which reductions in pollution will improve the health of populations is somewhat less clear. Pollution effects on health are sizable and statistically significant in both industrialized and lesser-developed countries (Cropper et al., 1997). Based on levels of total suspended particulates (TSP) in New Delhi between 1991 and 1994, the average pollution level was five times the limit recommended by the World Health Organization (WHO). Variations in deaths in New Delhi responded statistically to the variations in pollution; if these estimates prove true, then a reduction of pollution levels of about one-third would reduce deaths by more than 2 percent.

A similar study by Schwartz and Dockery (1992) in Philadelphia suggests that reducing the pollution level there by the same 100 micrograms per cubic meter would reduce deaths by more than 6 percent in the general population and nearly 10 percent for the elderly. This is because with our generally better health status in the United States, more people live long enough to become part of the population most sensitive to respiratory problems from pollution.

## Income and Health

While we know that good health during the years when an individual is forming a career can be a big boost to that person's income later in life (James Smith, 1998), we also know that being rich does not necessarily cause one to choose to live and eat wisely. Even programs designed to raise the income of poor families, such as (the late twentieth century) Aid to Families with Dependent Children in the United States, did not always correlate with good health habits among the recipients (Currie and Cole, 1993; Currie and Gruber, 1996).

Though earlier work had suggested that being richer in America was generally better for one's health, research by Deaton and Paxson (2001) brought that conclusion into question. Examining in detail both U.S. and British data over time, they find the relation of income and health to be complex and contradictory. There was a substantial decline in mortality after 1950, but rather than growing incomes as the cause, they conclude "a more plausible account of the data is that, over time, declines in mortality are driven by technological advances, or the emergence of new infectious diseases, such as AIDS" (p. 29).

Part of our problem thus far in researching the contribution of income to health in the industrialized world is that incomes do not vary greatly enough to detect the larger patterns. Pritchett and Summers (1996) leave little doubt that extremely low incomes have a strong effect on people's health. Though they treated the econometric challenges with great respect in establishing their conclusion, the most persuasive arguments may be those provided by simple graphs fitting various health statistics to per capita income data. These curves fit well and reveal that "modern" standards of good health are enjoyed solely by the industrialized countries with mortality experience turning sharply worse with lower income levels, conditions common in the underdeveloped world.

## THE ROLE OF SCHOOLING

What is the role of education? Since education includes both formal and informal training (such as experience or on-the-job training), some portion of education is impossible to measure accurately. Most often health economists focus on schooling as measured by years of schooling, or academic accomplishments such as diplomas or degrees. Health status correlates significantly with schooling as we have seen. If the marginal product of health care is truly small ("we are on the flat of the curve"), then perhaps we should reduce public health expenditures on health care at the margin and transfer the expenditures to education. However, the wisdom of such a policy depends on which of the two theories is correct.

## Two Theories About the Role of Schooling

The ideas and work of two health economists serve to develop and contrast the two theories. First, Michael Grossman's (1972a, 1972b) theory of demand entails a central role for education. Grossman contends that better-educated persons tend to be economically more efficient producers of health status.

In contrast, Victor Fuchs (1982) has suggested that people who seek out additional education tend to be those with lower discount rates. A decision-maker with a high discount rate will tend to prefer projects with immediate payoffs versus long-term projects. People with a lower discount rate tend to be those who value the long-term gains more. Now consider individuals facing a possible investment in education. Because education requires current costs to gain distant payoffs, individuals with relatively low discount rates will be more likely to invest in education and in health as well.

## Empirical Studies on the Role of Schooling in Health

Recent evidence supports the view that education makes one a more efficient producer of health (Lleras-Muney, 2005). Knowing that compulsory education laws came into being in various places at various times in the twentieth century, she reasoned that the related birth cohorts from that era would have experienced different levels of education but would have been similar in many other respects. This formed a natural experiment in which she could analyze the survival patterns of these people to detect a pure influence of education on health. Furthermore, the education laws could not have been directly manipulated by the study subjects, so they were good "instruments" for education. By this approach, she was able to conclude that education has a clear, causal, and positive effect on health. By 1960, the early century education experience appeared to have increased life years by 1.7 years, a substantial increase and one not due to time preferences of the subjects.

Lleras-Muney's study inspired new research of the effects of new laws extending the length of compulsory education in England and Ireland (Oreopoulos, 2006; Auld and Sidhu, 2005). These supported the earlier findings; an additional year of schooling caused an improvement in the affected student's health. To emphasize, the improved health was experienced by "likely dropouts," forced by the law to attend one more year of high school. Interestingly, Lindeboom and colleagues (2006) inquired through research as to whether the children born later on to these students *also* benefited from improved health, but the findings indicated that they did not.

In summary, research has supported the theory that education makes people more efficient producers of their own health. Cutler and Lleras-Muney (2006) add further support in their recent work by showing that education is associated statistically with better reasoned choices of health related behaviors. One finds as well that education plays a stronger role in health for cases where new medical knowledge is more important.

## CONCLUSIONS

In this chapter, we investigated many topics related both directly and indirectly to the production of health. The health production function exhibits the law of diminishing marginal returns. While the total contribution of health care is substantial, the marginal product is often small. Historically, we found that much of the decline in mortality rates occurred prior to the introduction of specific, effective medical interventions. Thus, historically the contributions of health care, at least as provided by the health practitioner, were probably small until well into the twentieth century. The small, modern-day marginal product of health care is statistically significant. Health care benefits people differentially and is generally more productive on the margin for women and blacks. Similarly, certain categories of health care have greater marginal effects on the population than others; prenatal care programs are examples of the more productive categories. Education has a strong association with health status. Whether this means that it causally improves health has long been an issue of contention. Recent research supports the view that education improves health.

## Summary

1. The production function for health exhibits diminishing marginal returns. In developed countries, the total product of health care is probably substantial at the same time that the marginal product is relatively small.
2. The historical declines in mortality rates in representative industrial countries were substantially responsible for the large growth of populations.
3. The historical declines in population mortality rates were not due to medical interventions because effective medical interventions became available to populations largely after the mortality had declined. Instead, public health, improved environment, and improved nutrition probably played substantial roles.
4. The marginal product of health care in the United States is small. Recent studies find elasticities in a narrow range around 0.10. Nevertheless, the total contribution of investment in health care technology over the past several decades is probably in the trillions of dollars; much of the improvement in life expectancy in this period can be attributed to health care improvements.
5. The RAND Health Insurance Experiment found that increased use of health care has little effect on the illness rates of the study population. However, studies of the totally uninsured now reveal gains in health due to publicly provided health insurance.
6. Lifestyle and environment are major and statistically significant determinants of population health status.
7. Health care contributes more substantially to health for subgroups of the population, including infants and also certain ethnic minority groups.
8. Social capital, produced by groups, is increasingly viewed as a substantive determinant of individuals' health.
9. Education, as measured by years of schooling, is positively related to population health. Recent research supports the view that the relation is causal, that increased education improves health.

## Discussion Questions

1. Assume that health production is subject to diminishing returns and that each unit of health care employed entails a constant rate of iatrogenic (medically caused) disease. Would the production of health function eventually bend downward? Explain.
2. What evidence is there to suggest that the United States is on the "flat of the curve" in health production? Is a typical developing country likely to be on the flat of its health production function? Discuss the differences.
3. Which of the following are important in explaining the modern rise in population in England and Wales: birthrates, death rates, and net migration rates? Describe the evidence.
4. "Medical interventions were not important in the historical declines in mortality rates, but that does not imply that medical research was unimportant." Explain this viewpoint.
5. What role did public health play in the historical decline in mortality rates?
6. Suppose you were hired as an adviser to a developing country and you were versed in the theory of production, the historical role of medicine, and the modern-day health production function studies. Their government seeks advice on the wisdom of a relative emphasis on health and health investment versus other forms of economic investment. What would be your advice?
7. Someone says the following: "Lifestyle may be the most important determinant of health status, but changing lifestyles may not be the least costly way to improve population health status." Explain the circumstances under which this opinion could be true. Is it likely to be true in reality? What does the evidence on lifestyle suggest about government policies to improve the public's overall health?
8. Summarize the two theories on how schooling is correlated with health status. Which of the two theories does the evidence support?
9. Research shows that the returns for prenatal health care are high, whereas it may cost hundreds of thousands of dollars to keep an acutely ill, elderly person alive. What does this suggest about the appropriate allocation of resources among members of society?
10. What is the total contribution of health care to health as estimated by researchers? What parts of health care are most effective in this regard?

---

## Exercises

1. Graph the production of health function  $HS = 10HC^{0.5}E^{0.3}LS^{0.4}HB^{0.2}$  in a graph with axes  $HS$  and  $HC$ , assuming  $E = 10$ ,  $LS = 5$ , and  $HB = 7$ . Graph the marginal product of health inputs. Is it increasing or decreasing? Show how the curve changes when  $E$  is increased to 15.
2. Which factors in Table 5-3 were important in explaining improvements in black neonate mortality rates? White neonate mortality rates? Speculate on why some of these factors may have been more important for blacks.
3. What are the differences between mortality and morbidity? Would you expect the two variables to be related to each other? If so, how?
4. We know that correlations never explain; it is our theories that provide explanations. Reexamine Table 5-3 and draft theories to explain why WIC has a larger contribution than Organized Family Planning. Why do the contributions differ between blacks and whites on WIC, abortion, and pre-natal care?
5. Pritchett and Summers argue that income per capita is strongly and positively related to health status when viewed across the world. From data in Table 2-2, in Chapter 2 of this text, plot a graph of GDP per capita against life expectancy for the countries shown. Does your plot confirm the Pritchett and Summers finding?

# The Production, Cost, and Technology of Health Care

- Production and the Possibilities for Substitution
- Costs in Theory and Practice
- Technical and Allocative Inefficiency
- Technological Changes and Costs
- Diffusion of New Health Care Technologies
- Conclusions

Recognizing that health is the ultimate output in the health sector, we understandably direct considerable policy interest toward the production and cost of the intermediate output, health care. Despite exceptions, both politicians and consumers seeking lower health costs do not seek cuts in “necessary” health care—only that which is “purely wasteful.” Yet, health care, even when needed, inevitably “costs too much.”

This chapter addresses production, cost, technology, and efficiency of health care. We investigate first the question of what degree of flexibility the manager or the public policymaker can find in the production process in order to make improvements. Must we always call physicians for certain tasks, or can we substitute nurses and other less expensive inputs? Can managers safely vary the mix of types of nurses employed? These questions reduce to the degree to which health care inputs substitute for one another.

Second, the study of cost functions can provide further clues to finding efficiency gains. Economists find that economies of scale and scope exist in many industries, and society would be better off if firms chose the size that minimizes average costs. Theory states that the perfectly competitive industry achieves this in the long run without outside interference. Health care firms are generally not perfectly competitive nor necessarily perfectly managed, and health economists inquire into the extent of any excess costs.

The next issue of the chapter is the technical and allocative efficiency of health care firms. These types of efficiency and the inefficiency they define articulate in economic terms the central research issue of the American consumer’s complaint: “I am paying an arm and a leg for my family’s health expenses. Am I really getting my money’s worth?”

We then examine how greatly the picture changes as technology, the major mover in the health care industry, changes. We will see that even when new health technology improves our lives, it can also make life more expensive. Insurance plays an important role as a shifter of demand, though probably its influence through its effect on technological innovations is as important.



Finally, we study the economics of how and why new health care technology diffuses through the health system and forms a “logistic” pattern. The diffusion can be rapid, although it may be slowed by regulatory or institutional realities.

### PRODUCTION AND THE POSSIBILITIES FOR SUBSTITUTION

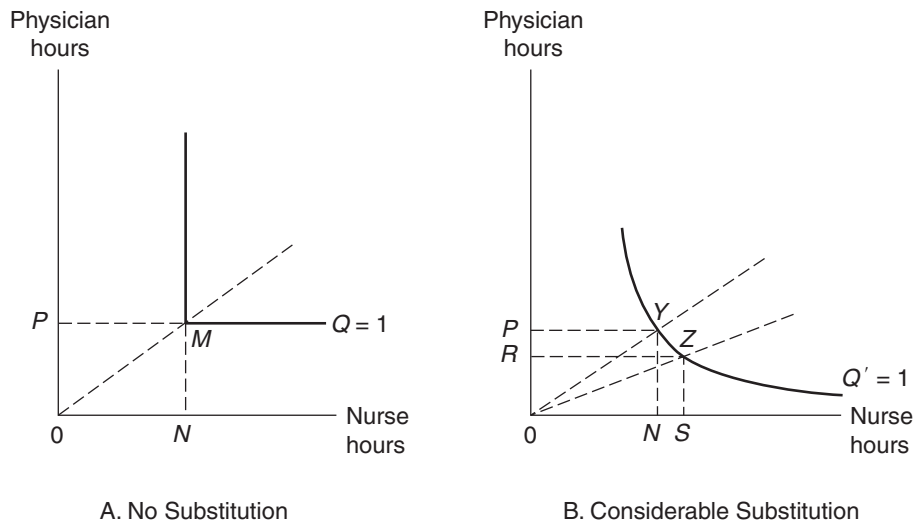
Economists often note that there is more than one way “to skin a cat,” that different techniques are available to produce the same product. A single technique is one recipe for production, meaning one specific combination of inputs. When multiple techniques are available, one can choose a relatively capital-intensive (labor-intensive) technique during times when capital (labor) is relatively cheap. It also means the ability to use cheaper forms of labor in substitution for more expensive forms.

While multiple techniques are common in many industries, health practitioners often recognize only one correct way of treating a given illness. The belief that only a single technique is possible or wise is what Victor Fuchs has called the “monotechnic view.” If such a view correctly described production processes, cost-saving substitutions would be difficult if not impossible without reducing either output or quality. A more flexible production process permits cost-saving improvements that may be beneficial to consumers. Economists investigate this question using the concept of substitution.

#### Substitution

Flexibility means the ability to substitute one input, such as capital, for another input, such as labor, while maintaining the level and quality of output. This does not mean that the two inputs are equivalent, but only that alternative combinations are possible. Figure 6-1 illustrates a case with no input substitution and a case in which an infinite number of techniques are available. In panel A, the isoquant shows the possible combinations of nurse hours and physician hours required to treat one patient case in a hospital; the isoquant is labeled  $Q = 1$ . Given this situation, only one sensible production technique combines the two inputs. Physicians and nurses must be combined in the ratio given by  $OP/ON$ , the ratio of inputs used at the corner point  $M$ . Notice that  $OP/ON$  is also the slope of line segment  $OM$ .

What does this mean? In panel A,  $OP$  physician hours are required to produce one case, and the addition of nursing hours beyond  $ON$  will not add to output unless physician hours also are



**FIGURE 6-1** Degree of Substitution Between Physicians and Nurses

increased. This applies to a production problem where patient care requires certain professional tasks that only a physician is trained and competent to perform.

The fact that the isoquant is flat when moving to the right from  $M$  means that adding nurses beyond the required combination produces no more output (i.e., they would be wasted). Likewise, the fact that the isoquant is vertical when moving upward from  $M$  means that additional physician hours beyond the required ratio combination are simply wasted resources. How would an isoquant look if substitution were possible? Panel B illustrates this. Again, a unit isoquant is depicted, but a smooth downward-sloping convex curve illustrates that many combinations of inputs could be chosen without being wasteful. For example, one case can be treated with the  $(OP, ON)$  combination of inputs or equivalently with the  $(OR, OS)$  combination. Each point on the isoquant represents a different technique. The slope of any isoquant, such as the isoquant labeled  $Q' = 1$ , is called the marginal rate of technical substitution (see Chapter 2), and it represents the rate at which nurse and physician hours can be exchanged while still maintaining output.

We emphasize that even though we can substitute in panel B, nurses and physicians are not equivalent. It is not even true that a fixed number of nurses always can replace a physician. Moving along the curve from point  $Y$  to point  $Z$ , the rate of substitution changes; that is, the slope becomes flatter, indicating a diminishing marginal rate of technical substitution. To replace one hour of physician time at point  $Y$  requires some number of nursing hours; however, at point  $Z$ , where we are using fewer physician hours, we require a much greater number of nursing hours. That is, as physician time grows scarcer relative to nursing hours, it becomes more difficult to replace. This retains the idea that while substitution is possible, it may be difficult, expensive, or unsafe to have nurses do certain physician tasks, and vice versa. The curve may even become flat at some point, indicating that we have reached a minimum of required physician time.

### What Degree of Substitution Is Possible?

The graphs in Figure 6-1 define terms, but they do not tell us which of the two cases is true of the health care world. From the 1970s on, economists addressed the question of whether physician extenders (as noted in Box 6-1) could substitute for physicians. Physician extenders refers to specially trained physician assistants or nurse practitioners who are utilized to perform certain tasks, including some that formerly were performed by the physician.

The estimates obtained are equivalent to the measurement of a few points along an isoquant, such as the ones just depicted. This work suggested that substitution could be substantial. Depending on the number of physician hours employed, one physician extender could substitute for 25 percent to more than 50 percent of a physician's services. Since physician extenders have a much lower training cost, this degree of substitutability could result in substantial savings. Reportedly, the observed physicians were not employing sufficient numbers of assistants to make the practice fully profitable to the physicians.

### Elasticity of Substitution

The hospital has provided a related focus of production studies. One study presents evidence of substitution among various categories of hospital inputs. It also illustrates a second and more convenient way to measure the possibilities for substitution between inputs, the elasticity of substitution ( $E_s$ ), which measures the responsiveness of a cost-minimizing firm to changes in relative input prices. It is defined as follows:

$$E_s = \frac{\text{Percentage change in factor input ratio}}{\text{Percentage change in factor price ratio}}$$

What does this elasticity mean? If a firm were a cost minimizer, then it would be responsive to changes in input prices, and it would tend to respond by shifting away from the now costlier input to

**BOX 6-1****Health Care Professionals: Expanding the Possibilities**

Most research on health professionals tends to focus on physicians, dentists, and managers. Yet each researcher is aware that a large share of “getting better” depends on the allied health professions. These professionals, besides often providing a human touch to care, can significantly improve the health production process.

First, the variety of specializations among allied health labor provides the benefits of Adam Smith’s “division of labor.” The theory explains how a focus by each expert on his or her specialized task yields greater output for the whole than were one to insist that each person be a generalist. Imagine a hospital where every professional was a generalist. This health professional then keeps the electronic records, prepares patients, sees patients, draws blood, takes X-rays, and so on. With present-day complex technologies, this is not even possible, let alone efficient. Office managers, registered nurses, medical technicians, X-ray technicians, physicians, and medical records specialists together solve this economic problem as a team.

Second, the availability of other health professionals enhances the possibilities for substitution in production. Nurse practitioners and physicians’ assistants can substitute for physician time and, if used wisely, can expand output by freeing physicians for tasks more directly suited to their training (Brown, 1988). Similar opportunities exist between physicians and nonphysician services in producing mental health services (Deb and Holmes, 1998). Expanded functions for dental assistants and dental hygienists have offered lower dental care prices without loss of quality (Liang and Ogur, 1987). Likewise, research suggests possibilities for substitution between registered pharmacists and pharmacy technicians (Okunade and Suraratdecha, 1998).

Nurse anesthetists are already generally accepted in practice, and while laws often limit what midwives can do, they have a long history of delivering babies. Finally, a stronger general emphasis on providing information to patients has made a wide variety of caregivers into information providers as well. The bottom line? Flexibility in the production of health care exists, in fact, and opportunities for substitution abound.

the now relatively cheaper input. Suppose that a hospital is currently at combination  $Y$  in panel B of Figure 6-1 using  $OP$  physician hours and  $ON$  nursing hours to treat one case. Suppose also that physicians are paid \$200,000 per year and nurses are paid \$40,000 per year. Suppose, finally, that the hospital employs 100 physicians and 100 nurses. If there is a 10 percent increase in the relative wage rate of physicians (from \$200,000 to \$220,000), the cost-minimizing hospital moves to combination  $Z$ , which substitutes  $NS$  nursing hours for  $PR$  physician hours. The relative physician input ratio decreases from the one shown by the slope of the line segment  $OY$  to the one shown by the slope of  $OZ$ .

Assume that the decrease in the slopes (i.e., the ratio of factor inputs) is 6 percent. The elasticity of substitution,  $E_s$ , has a value of 0.6, indicating that every 1 percent change in relative factor prices leads to a 0.6 percent change in the relative use of those factor inputs. Whereas the ratio of physicians to nurses was previously 1.00, a 10 percent increase in relative physician wages (from 5 to 5.5 times nurses’ wages) would change the input ratio to 0.94 (a 6 percent decrease). This would represent the replacement of one physician (\$220,000) with five nurses (\$200,000), hence lowering costs (due to the change in physician salary) by \$20,000.<sup>1</sup> We will report  $E_s$  in absolute value as positive numbers, even though we understand that the firm’s response is to decrease the relative use of a more expensive input.

The minimum value of  $E_s$  is zero, and a firm with isoquants represented by the one shown in panel A of Figure 6-1 will have an elasticity of substitution equal to zero because it always will use the same input combination to produce a given level of output regardless of relative factor prices. Higher values of  $E_s$  indicate a greater potential for substitutability.

<sup>1</sup> The new ratio is 99 physicians to 105 nurses, or 0.94.

**TABLE 6-1 Substitution Elasticities for Teaching and Nonteaching Hospitals Evaluated at the Mean**

Input Pair	Nonteaching Case-Mix Adjusted Admissions	Teaching Case-Mix Adjusted Admissions
1. Medical Staff with Nurses	0.547	0.159
2. Medical Staff with Beds	0.175	0.155
3. Nurses with Beds	0.124	0.211
4. Nurses with Residents	—	2.127
5. Medical Staff with Residents	—	0.292

Source: Gail A. Jensen and Michael A. Morrisey, "The Role of Physicians in Hospital Production," *The Review of Economics and Statistics* 68:3 (August 1986), pp. 432-442. Copyright © 1986 by the President and Fellows of Harvard College and the Massachusetts Institute of Technology, with permission.

A good example elasticity of substitution is used is provided by Acemoglu and Finkelstein (2006). These authors observed how health care firms responded to the introduction of the Medicare Prospective Payment System, which they found to cause labor to become more expensive. True to the theory of the firm with its responsiveness to the market, they found the firms' response was to increase the capital/labor ratio.

### Estimates for Hospital Care

Physicians are important to productivity of all hospital inputs. Pauly (1980) and Jensen and Morrisey (1986) incorporated this fact into their analysis of the elasticity of substitution between hospital inputs. They estimated a production function for hospital care and generated the isoquants and estimated the elasticities of substitution. Patterns of input use were observed in the process across a large number of hospitals.

Table 6-1 shows the estimates of elasticities of substitution between pairs of inputs. For example, a 1 percent increase in the price of medical staff relative to nurses would result in a 0.547 percent decrease in the ratio of medical staff to nurses.

How are we to judge the estimated degree of substitutability? Are these numbers large or small? The elasticities reported are at least sufficient to show that some substitutability exists between virtually all pairs of hospital inputs. In fact, the authors conclude, ". . . all inputs in both teaching and nonteaching hospitals are substitutes for each other." The smallest values for substitution reported here are between beds and categories of labor. It may seem nonsensical that one can substitute people for beds; but "beds" here represent a convenient measure of the various and often complex capital inputs used by a hospital. The data suggest that even in hospital care production, where labor is undoubtedly the critical input, capital can be substituted on the margin for labor.

Custer and Willke (1991) and Lehner and Burgess (1995) have shown these results to be sensitive to the particular definition of the physician's input. Even so, most health economists probably would agree that substitution possibilities exist among many health care inputs, though their range is still uncertain.

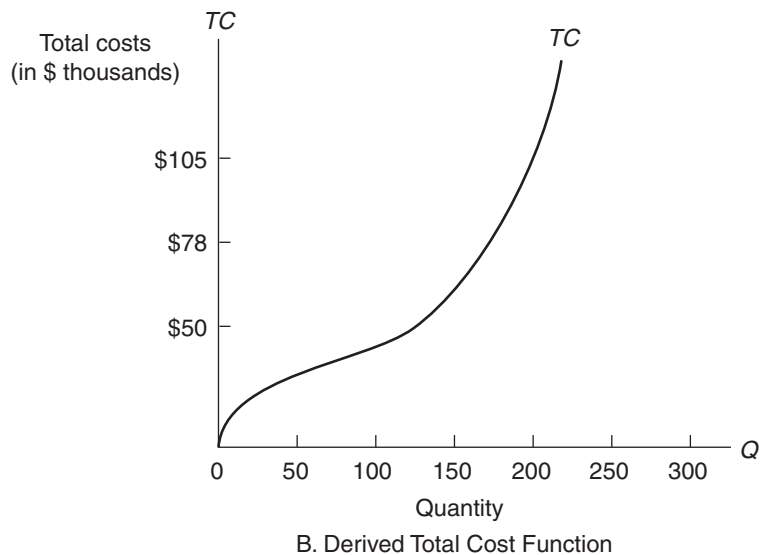
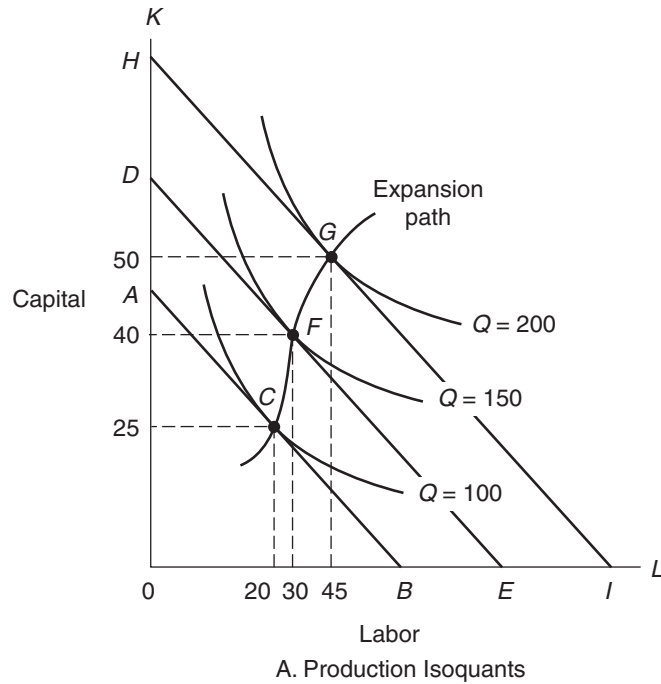
Recent studies have shown that home health care can substitute for hospital care. Is the "home hospital" as good? Is it less costly? Illife and Shepperd (2002) report that studies of randomized trials of home versus hospital care find little difference in mortality outcomes or in costs.

### COSTS IN THEORY AND PRACTICE

The production function, its isoquants, and the elasticities of substitution have consequences for costs. We will show the derivation of the cost function and explain the technical terms *economies of scale* and *economies of scope*.

### Deriving the Cost Function

As we noted, the production function describes the input/output relationships, and the cost function describes the cost/output relation. The two are closely related, and under the right conditions, the two functions can be derived one from another. We illustrate the closeness of this relationship in Figure 6-2.



**FIGURE 6-2** Production Function for a Hypothetical Physician Practice

Panel A, which depicts several isoquants, illustrates the production function, let's say for a physician practice. An infinite number of isoquants exist in principle, and we have chosen to show only three. The lowest isoquant shows that many points (or combinations) of capital and labor are capable of producing the 100 physician office visits. Higher isoquants, of course, produce more, 150 and 200 visits, respectively. Imagine that Figure 6-2, panel A, had no straight, slanted lines ( $AB$  and so on). The lines "removed" identify the cost consequences of the firm's input choices. Without them, the graph would characterize a firm that knows "only half of the story," only the output consequences of its input choices. Do firms exist that would simply ignore the cost inputs? Would they say: "We wish to help people get well regardless of what it costs and economics is not an appropriate consideration."

In reality, health care firms, like all others, must meet their creditors or close their doors—they must at least break even. A tougher question for economists, however, is whether the many nonprofit firms strive to minimize costs. We must further ask, how well do they achieve this goal? Economic theory derives the cost function only if the firm in question seeks to minimize its costs. Before showing how the derivation works, consider that Ellis (1993) makes a good case that we are "safe" in this regard. Ellis addresses this question: If a community hospital board discovered that it could treat 1,000 patients with proper care and still permit a waste of resources worth \$20,000, then some board member could sensibly suggest: "Eliminate the waste and we could treat even more patients." We will assume that health care firms reason in this way.

### Cost Minimization

While isoquants show the many combinations of inputs to produce a given output; they do not by themselves describe the cheapest combination of inputs to produce the given output. The firm minimizes costs of producing a given level of output with the aid of a second tool—the isocost curve, the downward-sloping straight lines in Figure 6-2.

An isocost curve is the collection of all combinations of capital and labor that together cost a given amount. Let the level of total cost being considered be  $TC$ ; this money can buy many combinations of capital and labor,  $K$  and  $L$ , in fact any combination whose costs add up to  $TC = rK + wL$ , where  $r$  is the rental price of capital and  $w$  is the wage rate of labor. The "rental price of capital" means the cost to the firm of using the capital for one period, regardless of whether it actually rents the capital or owns it. This particular isocost equation can be transformed algebraically so that  $K$  appears on the left-hand side and all other terms appear on the right, yielding the equivalent equation,  $K = TC/r - (w/r)L$ , which is an isocost line.

The firm wishing to produce a given output level, say 100 visits, will minimize its costs by choosing the lowest isocost curve that is tangent to the 100-visit isoquant. In the figure, least-cost production of 100 visits occurs at input combination  $L = 20$ ,  $K = 25$  at point  $C$  on isocost curve  $AB$ . Given knowledge of the input prices, we can calculate the output cost. For example, let  $r = \$1,200$  and  $w = \$1,000$ ; then, least-cost production of 100 visits will cost \$50,000. Alternatively, if the firm wishes to produce 150 visits, the least-cost production would occur at point  $F$ , which entails 30 units of labor and 40 units of capital for a total cost of \$78,000. In this fashion, the combination of the production function represented by the isoquants and the cost requirements represented by the isocost curves generates a set of outcomes, or points  $C$ ,  $F$ , and  $G$ . The set of all possible points of tangency, such as these, is called the expansion path.

The expansion path supplies the information that associates a given output with its minimum cost. When these cost and output data are recorded in a graph, as in panel B, the result is the firm's total cost function. The cost function has a lazy S-shape, a pattern thought to be typical of many firms in practice. It also goes through the origin, indicating that if this firm produces nothing, it will incur no costs, meaning that the firm has no fixed costs. Economists refer to a period long enough for the firm to alter or avoid any of its commitments to input suppliers as "the long run." Thus, the implication is that the cost function shown is depicting the firm in the long run. This cost function is a "frontier" in that it represents the minimum possible cost of producing a given output. Actual

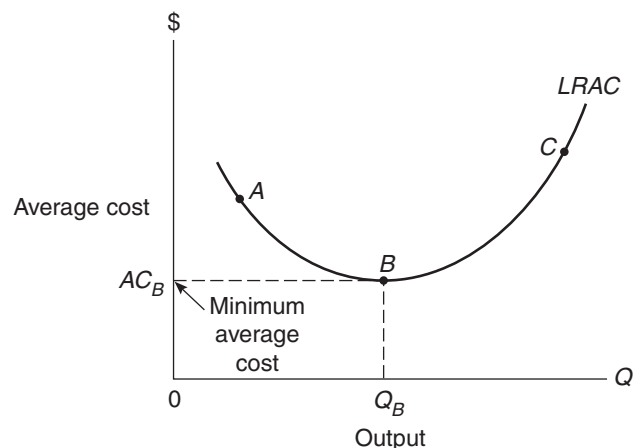
firms may operate somewhat inefficiently, and we will observe cost levels above the frontier. It is a contradiction to say that one could observe a cost level below the cost frontier. Analysts often wish to determine whether health care firms are operating on or above their cost frontier, and we will return to this efficiency issue later. At present, we address a different issue—whether a firm is operating at an economical point on the frontier from the point of view of society as a whole.

### Economies of Scale and Scope

To simplify the exposition, we separate the issues of economies of scale and scope, treating scale economies first. Consider a physician firm such as the one depicted in Figure 6-2. The long-run total cost function as shown in panel B can be transformed to express information about economies of scale. The average costs for this firm can be calculated by dividing the given cost level by the corresponding number of physician visits. The resulting long-run average cost (*LRAC*) function is in Figure 6-3. A firm experiences economies of scale when its long-run average cost is declining as output increases. Thus, the firm depicted exhibits economies of scale in region *AB*. Conversely, the firm experiences diseconomies of scale if and only if the long-run average cost is increasing as output increases, such as occurs in the region *BC*. What output level would a profit-maximizing firm choose in this case? It is tempting to suppose it would choose output  $Q_B$ , at which its average costs are lowest ( $AC_B$ ), but this is not necessarily the case. A firm is not in the business of minimizing its average costs and would do so only if, coincidentally, the output that minimized costs also maximized profits.

A related concept is economies of scope. By definition, economies of scope are possible only for a multiproduct firm and because many health care firms are multiproduct in nature, the concept is highly relevant. Economies of scope occur whenever it is possible to produce jointly two or more goods more cheaply than if we produce them separately. As an illustration, consider the provision of pediatric hospital care (for children) and geriatric hospital care (for the elderly). Suppose there were two hospitals in town—one that provided only pediatric care and one that provided only geriatric care. Would the total cost of pediatric plus geriatric care be lower if one single hospital provided both? It may be cheaper to combine the two hospitals and achieve scale economies, but that is not the point at present. However, it might still be cheaper to combine them when the inputs needed for both types of care interact well together. Perhaps things learned in pediatrics have applications in geriatrics and perhaps the two could support each other so that the result would be lower total costs. If so, by producing the two different outputs jointly, we achieve economies of scope.

Economies of scope are illustrated by equation (6.1). In mathematical notation two outputs are shown,  $Q_1$  and  $Q_2$ . For example,  $Q_1$  may represent pediatric care and  $Q_2$  may represent geriatric care. The example in equation (6.1) concerns the attempt to produce output levels  $Q_1 = 100$  and  $Q_2 = 150$



**FIGURE 6-3** The Long Run Average Cost Function

either jointly or separately. Economies of scope exist if the cost of producing the two outputs jointly, that is,

$$TC(Q_1 = 100, Q_2 = 150)$$

is less than the sum of the costs of producing each quantity separately, that is,

$$TC(Q_1 = 100, Q_2 = 0) + TC(Q_1 = 0, Q_2 = 150)$$

In summary, economies of scope occur in this example if the following inequality holds:

$$TC(Q_1 = 100, Q_2 = 150) < TC(Q_1 = 100, Q_2 = 0) + TC(Q_1 = 0, Q_2 = 150) \quad (6.1)$$

### Why Would Economies of Scale and Scope Be Important?

The concepts of economies of scale and scope are of considerable interest to both public policy and to managerial policy. We illustrate their significance using the idea of economies of scale, but we can use similar arguments to show the importance of economies of scope.

Recall that a profit-maximizing firm wishes to do just that, maximize profits. It has no intrinsic interest in producing at that level of output at which average costs are at a minimum unless that output also coincidentally maximizes profit, but this is not generally the case. Consumers would like firms to minimize average costs as long as the firms pass these cost savings on to the public. One of the reasons that economists promote the theory of perfect competition is that competition forces the firm in the long run to operate so that it minimizes average costs. The competitive firm is guided by competition, as if by an invisible hand, to serve society's interests in keeping costs low.

Most health care firms, such as hospitals and physician practices, do not operate in perfectly competitive markets. Therefore, competitive pressures will not necessarily force them to operate at the most efficient scale of operation. There may conceivably be too few or too many providers. If there are too many, existing hospitals may be smaller than is required for the efficient scale; small hospitals may be forgoing the profit-maximizing opportunities available through greater economies of scale.

Historically, area-wide health planning, promoted by various government programs since the 1940s, encouraged the reduction of "excess beds." These programs also promoted the "rationalization" of geographic patterns of critical and expensive diagnostic equipment, such as CT scanners. The programs expressly aimed at reducing the growth in health expenditures would make more sense to economic thinking if the health care were either unnecessary or if the rationalization were expected to take advantage of economies of scale.

The gain to society from exploiting economies of scale also illustrates the natural monopoly. Consider a simple example. Assume that Figure 6-3 shows a health care unit (perhaps a hospital) that provides the diagnostic services of an MRI scanner. If many such firms are in the market area, then perhaps no single MRI scanner is operating at an output level at which it achieves the lowest long-run average cost. Because of its competition, the hospital unit depicted operates at point *A*, where average costs are higher. If society had fewer MRI scanners, the remaining ones could operate at a higher capacity, say at point *B*, with lower average costs of  $AC_B$ . These arguments depend on finding empirical results showing that further advantages from economies of scale at the level of the firm are available to society as a whole. The average cost curve in Figure 6-3 records only the costs incurred by the hospital unit. From society's broader perspective, the costs incurred by the patients and their visiting families and friends also are relevant. To see the point more clearly, draw the MRI scan example to its extreme. Suppose the minimal average cost occurs when one scanner serves a rural region of 400 miles in radius. Would residents of this region be better off building only one centrally located scanner? Or would the necessarily large travel costs make such a plan foolish in the extreme?



## Empirical Cost-Function Studies

With the theoretical ideas understood, the fundamental questions become empirical. Are there, in fact, economies of scale and scope available to be exploited in real-world health care firms? At what level of output and for what combinations of outputs are these economies achieved? Two themes occur often and represent differences in approach. We will describe long-run versus short-run studies, and behavioral versus structural cost functions.

**LONG-RUN VERSUS SHORT-RUN STUDIES** We have seen the shape of the long-run average cost curve defines economies of scale. The difference between the long run and the short run is well-defined in economics. The long run is a period sufficiently long for the hospital to end any fixed commitments and to make any cost-saving adjustments that are possible. The short run is a period during which the hospital still has some fixed commitments, that is, some inputs that cannot be varied. An example is the number of beds set up for service. Research (Vita, 1990; Fournier and Mitchell, 1992) has demonstrated that results differ depending on the investigator's assumption of a long-run versus a short-run equilibrium.

Economists use short run cost estimates to test for the short run or long run. Econometricians can assess this question by simulating that the capital is variable in the estimated function. If simulated profits rise when capital is changed, they know that the firm is operating in the short run. Bilodeau and colleagues (2002) recently estimated a cost function which found the U.S. hospital system to be overcapitalized, which we interpret as short run phenomena.

**STRUCTURAL VERSUS BEHAVIORAL COST FUNCTIONS** Economic and health service analysts frequently distinguish between structural and behavioral cost functions. By structural cost function, we mean a cost function derived in a consistent manner from economic theory, just as we have derived it in the previous section. That is, we use the production isoquants and the isocost curves to derive the cost-minimizing level of costs for each possible level of output.

In contrast, behavioral cost functions (Evans, 1971) are derived from analyses of the patterns in costs in actual data across hospitals. Variables are included that distinguish real-world differences between hospitals. For example, teaching hospitals have higher costs due to the teaching and research services that they provide (Farsi and Philippini, 2008; Linna and Häkkinen, 2006). The variables matter for costs but often do not have a clear role in the theory of cost functions. Sometimes behavioral cost functions omit variables, like factor costs such as employee wage rates or equipment.

Of those who chose the theoretically most consistent "structural" approach, some (Conrad and Strauss, 1983) found economies of scale, some (Cowing and Holtmann, 1983) found constant returns to scale, yet others (Vita, 1990) reported diseconomies of scale. Researchers who applied the behavioral cost-function approach (Granneman, Brown, and Pauly, 1986) found economies of scale for the emergency department.

## Difficulties Faced by All Hospital Cost Studies

What do hospitals produce? The difficulty of measuring the heterogeneous hospital output trips up nearly all hospital cost studies. First, hospitals differ by type of cases they treat; this is the "case-mix problem." Medicare's Diagnosis Related Group (DRG) payment system identifies 745 groups of cases, so the hospital is a multiproduct firm to an unusual degree. Some studies have virtually neglected the problem; others have used multiproduct cost functions with adjustments for case mix and related variables. Although the multiproduct approach is superior, even multiproduct methods rarely incorporate more than four or five hospital output categories. Even the question of whether sufficiently refined hospital cost estimates are achievable in principle has not gained a consensus among health economists.<sup>2</sup>

<sup>2</sup> Work by Olesen and Petersen (2002) promises to provide ways to incorporate large numbers of hospital outputs into fewer output measures.

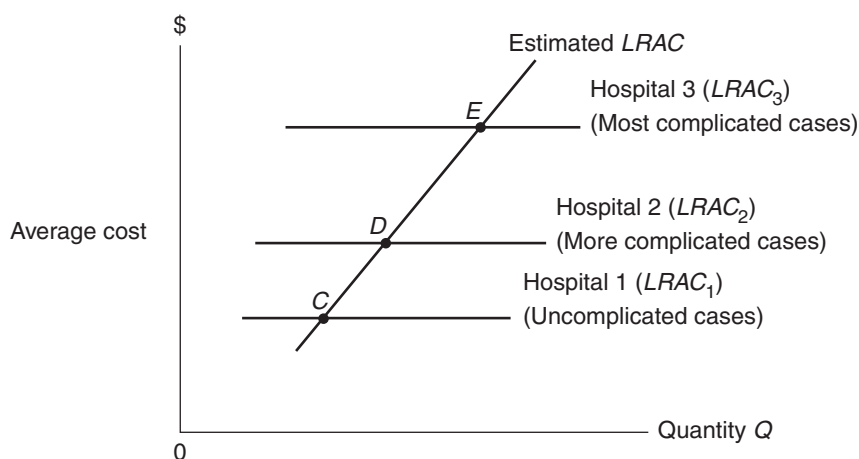
A related problem is how to treat quality. Unobserved or incorrectly measured variations in quality between hospitals may lead to errors in research examining economies or diseconomies of scale. Recent research established that hospital quality affects hospital costs, and quality is more expensive (Carey and Stefos, 2011).

The case mix issue is illustrated in Figure 6-4, and the quality issue, though not illustrated explicitly in the figure, is similar. The figure illustrates a situation where the true, long-run average cost function is flat, exhibiting neither economies nor diseconomies of scale. The three cost curves shown represent three hospitals, each with a distinct case mix. In order of ascending costs, these are Hospital 1, which treats uncomplicated medical cases; Hospital 2, which treats more complicated surgical cases; and Hospital 3, which treats the most highly complex level or tertiary cases.

Suppose that points *C*, *D*, and *E* represent the data observations available to the researcher for each hospital type for a given statistical study. The researcher may mistake the unobserved case-mix differences for diseconomies of scale, that is, mistakenly believe that the long-run average cost curve is rising as shown by the connecting line. This case is illustrative only because the underlying patterns of case mix, quality, and size of output could yield errors in either direction. Alternative approaches to treating the case-mix problem have been investigated extensively.<sup>3</sup> These studies show that case-mix differences between hospitals are materially important when estimating scale economies and cannot be overlooked.

Researchers may also lack reliable measures of hospital input prices. The cost of capital or the starting registered nurse's wage may seem simple data to collect (a matter of contacting an executive in the finance office), but they are often difficult to define adequately. Unfortunately, errors in measurement of hospital input prices have substantial consequences for the results.

Finally, hospital cost-function studies almost always omit physicians' input prices entirely. Physicians are not generally under hospital control, and this independence, as well as the accounting and the physician practices of billing the patient separately, become problems. Physician costs are properly part of production costs, and their omission results in biased econometric estimates of hospital costs. Early research experience with hospital costs (Pauly, 1980) established that physicians do matter. These studies establish that the difficulties in hospital costs center on the problems of measuring output in a multiproduct firm where quality matters a great deal.



**FIGURE 6-4** The Case-Mix Issue

<sup>3</sup> Hornbrook and Monheit (1985) study the importance of case mix. From data for 380 hospitals, they found that larger-scale hospitals in their sample tended to admit case mixes with relatively shorter lengths of stay.

## Modern Results

The most recent research supports claims that economies of scale exist in hospitals. Preya and Pink (2006) studied costs of Canadian hospitals prior to a massive consolidation, finding “large scale unexploited gains to strategic consolidation in the hospital sector” (p. 1049). Dranove and Lindrooth (2003) studied a large number of hospital consolidations, comparing them to matching hospitals that did not consolidate. They found “significant, robust, and persistent savings for mergers, 2, 3, and 4 years after consolidation” (p. 996).

## Summary: Empirical Cost Studies and Economies of Scale

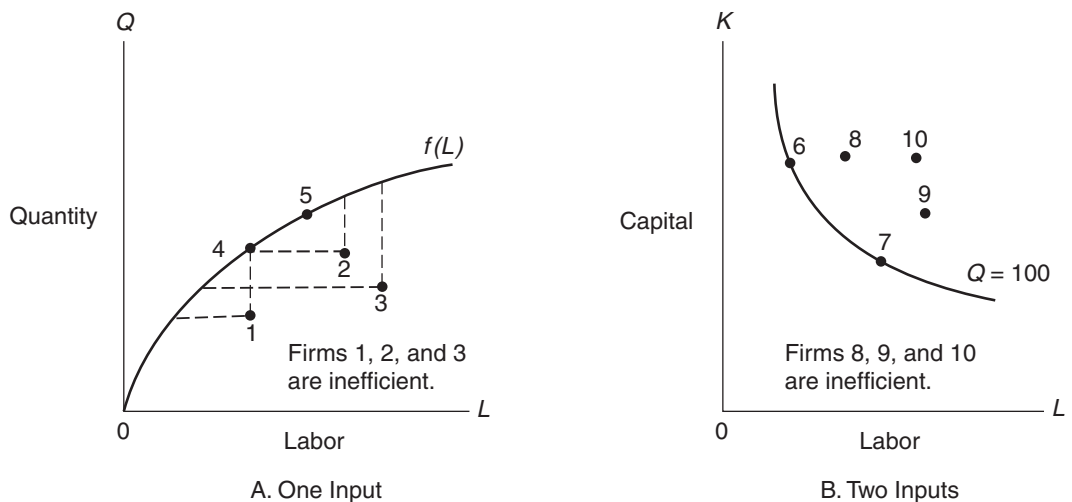
Early hospital cost studies led economists to believe that economies of scale existed, even claiming that the optimal hospital size was about 250 beds available for patient care. However, there followed a flurry of criticisms and corrections; the complex multiproduct hospitals, which also varied in quality, required more sophisticated methods. Studies that followed tended to dispel the earlier consensus with widely varied and sometimes contradictory results. Modern work (studies appearing since 2000) much more clearly reports that economies of scale exist in hospitals, a result that suggests that many hospital mergers might be justified on the basis of cost savings to society.

## TECHNICAL AND ALLOCATIVE INEFFICIENCY

In addition to issues of scale, efficiency can be measured in two other ways, each also of great concern to health care firms and policy makers. Economists refer to these as technical and allocative efficiency, or their lack, which is inefficiency.

### Technical Inefficiency

Technical inefficiency is illustrated in Figure 6-5, panels A and B. Panel A depicts a production process with one input, while panel B depicts a production process using two inputs, capital, and labor. Technical inefficiency implies that the producer is not achieving a maximum output from a given input combination. It is as if workers or machines were misused, not working at full capacity, or not cooperating well. In both panels, each firm’s actual experience is indicated by a firm number.



**FIGURE 6-5** Technical Efficiency and Inefficiency at the Firm Level

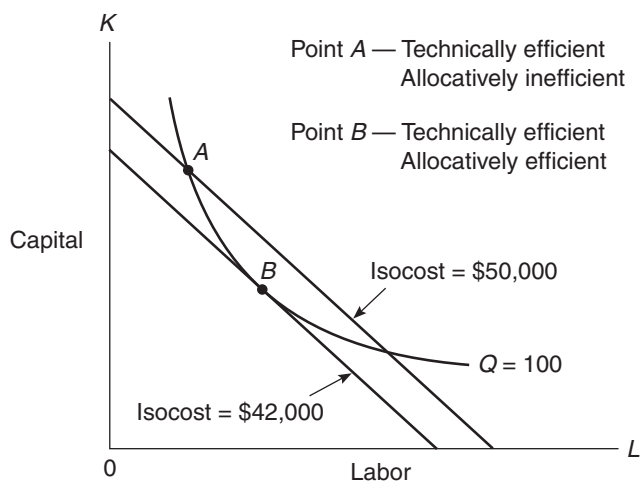
A technically inefficient firm falls off its frontier. In panel A, the production frontier is shown as the production function labeled  $f(L)$ . Firms 4 and 5 are on the frontier; that is, they are currently technically efficient. In contrast, Firms 1, 2, and 3 are off the frontier and thus are currently technically inefficient. In panel A, the firm inefficiencies are measured as relative distances from the frontier. The distance may be the output (vertical) distance—what output could have been achieved with these inputs—or it may be the input (horizontal) distance—how many fewer inputs could have achieved this output? Generally, these alternative approaches to inefficiency measurement will yield somewhat different results.

Panel B shows an isoquant representing frontier practice treating for 100 cases. Suppose that the actual current output of all firms depicted in panel B is known to be 100 cases. Firms 6 and 7 in the panel are on the isoquant for 100 cases and thus currently represent frontier practice and are technically efficient. Firms 8, 9, and 10 are off the isoquant, indicating that they have employed more input quantities than technically efficient production requires. As was the case in panel A, both output and input distance functions can be used as measures of inefficiency.

### Allocative Inefficiency

Technical efficiency applies conceptually to production within a given firm. By contrast, allocative efficiency requires the efficient allocation of inputs between firms and between outputs. Essentially, it requires that each type of capital and labor be put to its most rewarding use in society. Economic theorists have shown that allocative efficiency in production will result if each firm buys or hires inputs in competitive markets and if each firm minimizes production costs. Assuming competitive input markets—and thus fixed input prices common to all firms—permits us to describe allocative efficiency at the firm level. Here, allocative efficiency requires that each firm respond optimally to input prices; correspondingly, allocative inefficiency implies choosing an inappropriate combination of inputs in the sense that inputs and their prices have not been appropriately considered.

To illustrate, consider Figure 6-6. In this figure, the isoquant for 100 cases is illustrated as the curve labeled  $Q = 100$ . Assume that the firm being examined is currently producing its desired level of output, and that the desired level happens to be 100 cases. Two isocost curves also are depicted, with one indicating a cost level of \$50,000 and the other a cost level of \$42,000. Suppose that the firm in question was observed operating at a point A. Because this firm is treating 100 cases using an input combination on the 100 cases isoquant, we can say this firm is technically efficient. However, it is not allocatively efficient. At the current input prices, it uses too much capital, and not enough labor.



**FIGURE 6-6** Technical Allocative Efficiency

To confirm this idea, consider that the firm in question alternatively could have produced 100 cases at point *B*. As point *B* lies on a lower and thus less costly isocost curve, the firm at *B* would reduce costs from \$50,000 to \$42,000 by moving from point *A* to point *B*. Point *B* entails a tangency of the desired isoquant with the lowest feasible isocost curve. A tangency implies an equality of the ratio of input prices to the ratio of marginal products for the inputs (the equality of marginal output per dollar for each input). This is the firm's appropriate response to input prices, and its key condition for allocative efficiency.

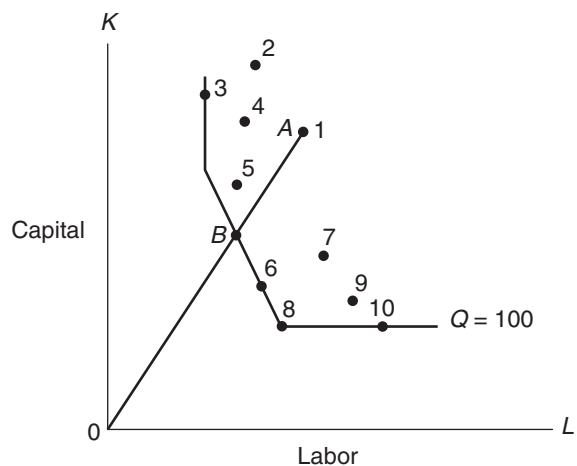
Though the conditions for efficiency of both types are well-defined, several different empirical techniques have appeared to address them. These techniques can be grouped into two categories: nonfrontier and frontier studies. In the nonfrontier studies, actual outputs or cost experiences for two or more groups of firms are compared while attempting to control for the effect of extraneous variables. In frontier studies, actual outputs or firm costs are compared to the best possible experience. We emphasize frontier studies because they are conceptually closer to the definitions of technical and allocative efficiency.

Two types of empirical frontier analysis have emerged. One, data envelopment analysis (DEA), was developed earlier. The other, the stochastic frontier analysis (SFA), was developed by 1977, with applications in health care proliferating in the late 1980s.

## Frontier Analysis

**THE DATA ENVELOPMENT ANALYSIS (DEA) APPROACH** DEA is illustrated in Figure 6-7. The frontier production, initially unknown to the researcher, is revealed as more firms are observed. A few such firms are represented by the points labeled 1 through 10. The data envelopment method finds the frontier isoquant for a selected level of output (for example,  $Q = 100$ ) by forming an envelope of the data. Researchers use linear programming to construct this efficient outer shell of the data points. Given the estimated frontier, technical inefficiency then is measured as a relative distance from the frontier. DEA is particularly useful for hospitals in that it easily handles multiple inputs and multiple outputs.<sup>4</sup>

DEA attracts researchers and research readers because it imposes no assumptions about the parameters of the underlying distribution of inefficiency. Analysts refer to this as "nonparametric." Its cousin SFA contrasts with DEA because researchers must guess the statistical distribution of the inefficiencies in advance. Those who prefer SFA argue that DEA assumes that all firms lying distant from the frontier are inefficient. Suppose that during a given year, the Huron Hospital nurses go on strike. Suppose further that the hospital is measured as experiencing substantial inefficiency; the hospital was operating at a substantial distance off the frontier for similar hospitals. Was the hospital



**FIGURE 6-7** The Data Envelopment Method

<sup>4</sup> See Fare and Lovell (1978) for economic applications of the DEA approach.

really inefficient or only apparently so? Should the researcher charge the entire shortfall in output to management error or should an adjustment be made (a “handicap” frontier) reflecting its special difficulties?

**THE STOCHASTIC FRONTIER ANALYSIS (SFA) APPROACH** SFA treats each firm uniquely by assuming it could be affected by a potential shock to its ability to produce care. Consider a hospital that is managed with perfect technical and allocative efficiency. Suppose this hospital discovers that its long-time major supplier has gone bankrupt. It takes months before comparable prices, qualities, and reliability are restored with new suppliers. Regardless of management’s response to these events, the hospital’s cost and output data for the year will differ from other apparently similar hospitals having standard experiences. This will be true even if the management is “perfect,” a model for other managers.

If each firm is randomly shocked during the period in ways that affect its production and cost performance, the firm’s best possible practice, its frontier, will be randomly shifted. When the frontier function is partly random, the result is a “stochastic process”—hence, the name stochastic frontier.

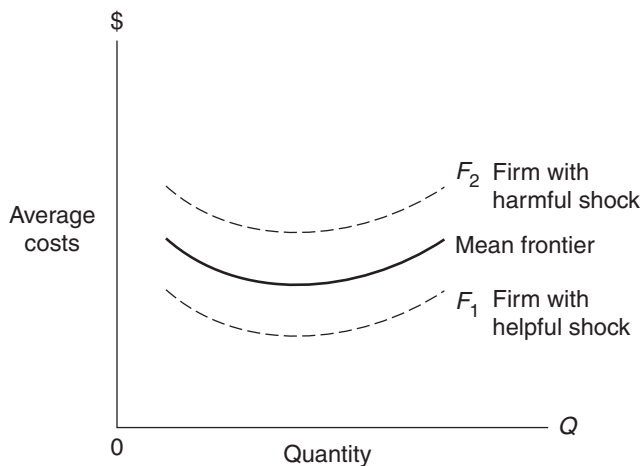
Techniques developed by Aigner, Lovell, and Schmidt (1977) make it possible to estimate both the individual firm’s expected frontier and the firm’s inefficiency. The stochastic frontier concept is illustrated in Figure 6-8, an example that focuses on average costs. The mean average cost frontier is the dark curve. The individual firm’s frontier is shifted by random shocks whose distribution must be assumed by the researcher from among known, parametric distributions. Thus, each hospital has a unique frontier and inefficiency distance from its frontier.

Many health economists recognize pluses and minuses to both approaches and treat them as complementary tools (Kooreman, 1994a, 1994b). Work with “panel data”—gathered as a sample of hospitals followed for many periods—suggests how the SFA method can be used without imposing the strong parameter assumptions (Lee and Schmidt, 1993).

### The Uses of Hospital Efficiency Studies

The estimates from the frontier analyses have stimulated investigations of substantial consequence. Consider these cases with background provided to explain the significance.

**TOTAL HOSPITAL EFFICIENCY** Critics argue that the U.S. hospital system has done an inadequate job of improving U.S. health status to justify its huge cost. We have seen that costs depend in part on whether one achieves economies of scale, and of course total costs also depend on the quantity demanded. But we have just seen that technical and allocative inefficiency also play a role. How efficient are U.S. hospitals?



**FIGURE 6-8** The Stochastic Frontier Method

Their reported efficiency levels have been quite high. The earliest DEA study (Valdmanis, 1990) reported technical efficiency levels of about 90 percent, while Magnusson's DEA (1996) study reached similarly high levels. SFA studies have tended toward similar levels; early SFA studies (Zuckerman, Hadley, and Iezzoni, 1994; and Folland and Hofler, 2001) found the sum of technical and allocative inefficiency to be only a little more than 10 percent. Since then studies have reported roughly similar levels of inefficiency. Frontier researchers are well aware of the many sensitivities of the method to variations across individual studies; but the various results support the hypothesis that the hospitals are highly efficient. One cannot blame hospital inefficiency for the high level of costs.

### **For-Profit Versus Nonprofit Hospitals**

Many people place greater trust in nonprofit hospitals than for-profits. Yet many economists and legal theorists are critical of nonprofit hospitals and demand to know why they deserve special benefits like freedom from taxes. To date, efficiency studies most often have not favored one organizational form over the other.

In many recent studies, nonprofit and for-profit hospitals appear approximately equal in efficiency. While the earliest studies (Valdmanis, 1990; and Ozcan et al., 1992) found differences between samples of public and for-profit hospitals, studies since then found no significant differences (Sloan et al., 2001). Burgess and Wilson (1998, p. 100) found "no evidence that differences in ownership affect technical efficiency after controlling for other factors." Looking at Italian hospitals, using DEA, Barbetta, Turati and Zago (2007) show a convergence of mean efficiency scores between not-for-profit and public hospitals. They believe that differences in economic performances between competing ownership forms result more from the institutional settings in which they operate, than the effect of the differing incentive structures.

If the efficiency data are neutral between organization types, this tends to favor the argument that nonprofits lack a justification for their tax breaks, at least on the grounds of technical and allocative efficiency.<sup>5</sup> Perhaps our impressions of relative efficiency of hospitals will change. Hollingsworth (2008) conducted a substantial literature review of over 300 frontier efficiency studies. Though cautious, he concluded that the public providers were somewhat more efficient.

### **Efficiency and Hospital Quality**

Mary Deily and Niccic McKay (2006) explain that hospital inefficiency may reduce the quality of care. Both care inputs in their study have been adjusted for quality. Quality of hospital output is measured by mortality rates. In principle, the process of combining inputs together may be affected by inefficiency. These authors test the proposition in a sample of about 140 Florida hospitals measured over three years. They found, using the stochastic frontier approach, that the inefficiency measure was a highly significant and positive contributor to a measure of hospital mortality rates.

Laine and colleagues (2005) attempted similar tests for long-term care. Although they detected no inefficiency effect on "clinical quality," they found inefficiency to contribute to the prevalence of pressure ulcers, "indicating poor quality of care was associated with technical inefficiency" (p. 245).

### **Are Hospital Frontier Efficiency Studies Reliable?**

This question was asked by both Folland and Hofler (2001) and by Street (2003). As an example, one of these papers estimated hospital efficiency values by three different versions of the stochastic frontier method. The versions were minor differences in the assumptions most investigators consider.

---

<sup>5</sup> Psychiatric hospitals (Mark, 1996), nursing homes (Kooreman, 1994a; Vitaliano and Toren, 1994); group homes (van Lear and Fowler, 1997), physician clinics (Defelice and Bradford, 1997; Gaynor and Pauly, 1990) and physicians working in hospitals (Chilingerian, 1995) also have been studied.

**BOX 6-2****Should We Close Inefficient Hospitals?**

Hospitals that are technically and/or allocatively inefficient will have higher costs than their more efficient peers. Should they be closed to save money? What about the utility loss of their former patients? Capps, Dranove and Lindrooth (2010) provide an interesting way to answer the questions. On the one hand, the cost savings from closing a particular hospital will be partly offset by increases at other hospitals who pick up these patients. On the other hand, recall that since the peer hospitals are more efficient, they will have higher occupancy rates after the change.

The authors' method allowed them to calculate the total travel time in the market that would be equivalent to the utility loss. Using industry estimates of the opportunity costs of driving, they find the dollar equivalent of the utility loss. The authors applied their approach to the cases of five recently closed hospitals, and they found the hospital closings provided a net welfare gain for the system.

Then they found the correlations between the versions. The correlations were rather poor, at 0.70 or lower. A rule of thumb here is that when the object is to test whether two series of numbers are valid equivalent measures of each other the correlation coefficient should be 0.70 or higher. The authors concluded that the frontier estimates seem adequate to discern mean differences between groups of hospitals, but they cannot be justified for the task of identifying inefficiencies by individual hospitals.

**Performance-Based Budgeting**

Yaisawarng and Burgess (2006) report success in devising an application of hospital efficiency data to the financial reimbursement methods of the Veterans' Administration hospital system. They have made a preliminary application to the hospitals, a method of payment to each hospital group in which the more efficient groups receive the highest payment. This provides healthier incentives, they argue, in contrast to the previous system where higher-cost hospitals received higher payment. If this works, it realizes a common dream for efficiency data, though it needs to overcome the skepticism generated by earlier research that found rankings of hospitals by efficiency scores to be sensitive to variations in estimation methods.

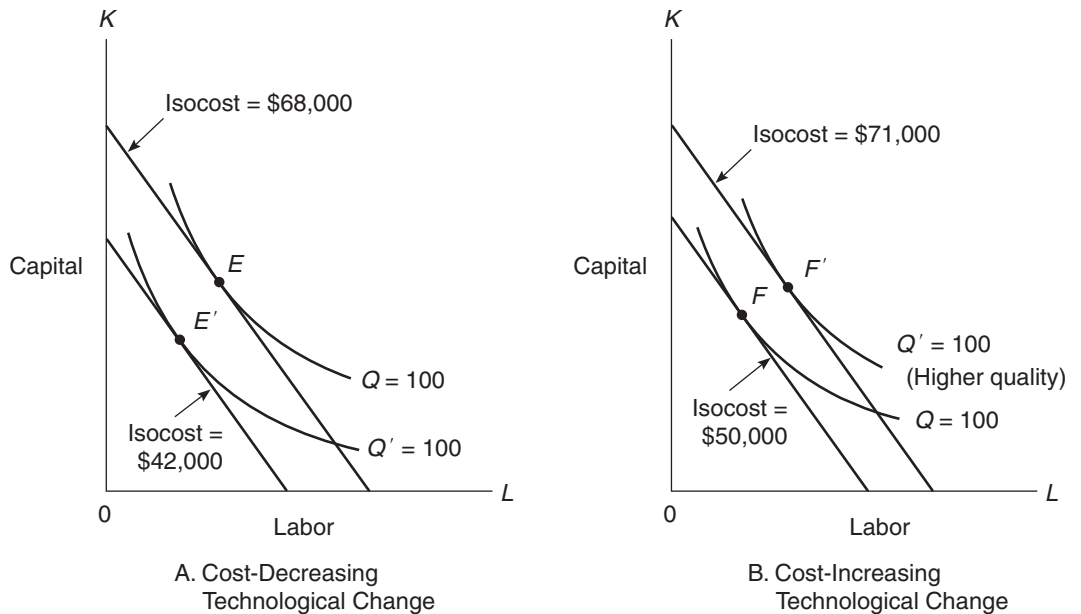
**TECHNOLOGICAL CHANGES AND COSTS**

The rapid pace of technological change in the health care industry raises economic questions about the effects these changes will have. Technological change may reduce costs when it improves the productivity of health care resources, or it may increase costs when it improves the quality of care or introduces new and costlier products. Because it often raises costs in the health sector, many researchers hypothesize technological change to be the major contributor to health sector inflation. Deep and widespread insurance coverage in the health sector may induce technological innovations of the type that increase costs. The effects on costs and the improvements to quality of care will depend on the diffusion of these new technologies to providers. Thus, patterns of diffusion have also become a critical subject of study. We investigate these issues in the remaining sections of this chapter.

**Technological Change: Cost Increasing or Decreasing?**

Technological change necessarily entails an improvement either by providing less costly production methods for standard "old" products, or alternatively by providing new or improved products. In either case, it will be less expensive to produce a given output, holding quality constant. However, the mix of products and services sold may change in directions that raise the average cost of a patient





**FIGURE 6-9** Cost-Reducing and Cost-Increasing Technological Change

day, a case treated, or a physician visit. Consequently, the total health care expenditure per capita may rise.

Figure 6-9 illustrates these concepts. In panel A, we hold quality of care constant and illustrate the isoquants representing 100 cases before and after a technological change. The technological improvement in panel A shifts the isoquant inward. The firm chooses an efficient combination of inputs at point  $E$  and after the change at point  $E'$ . This change results in the attainment of a lower isocost curve for treating the 100 cases.

Panel B illustrates the introduction of a new technology that makes it possible to treat 100 cases with better health status outcomes, thus providing a higher quality of care. Improvements entailing new products or, as shown in panel B, improved quality of care, are beneficial to the consumer, but they may be more costly. This is illustrated by a shift outward of the 100-case isoquant, resulting in production on a higher and costlier isocost curve. The typical patient will pay more for care. In some cases where patients are heavily insured, we may question whether the change is worth it to patients when they pay increased insurance premiums, or to society as a whole.

### Health Care Price Increases When Technological Change Occurs

How do we measure the cost of a treatment when the treatment changes radically over a mere one or two decades? For example, heart attack treatment (myocardial infarction) changed substantially from 1975 to 1995. Some new effective inputs proved extremely inexpensive (see the feature on aspirin in Box 6-3). Some materials did not exist in 1975, such as the intraortic balloon pump. Treatment practices changed; the average length of a hospital stay is now much shorter. Most important to the patient, the treatments are now more effective and have improved the length and quality of life for heart attack victims.

Treatment effectiveness improved, and in some cases less expensive inputs become available. Sometimes newly designed inputs were more costly than the ones they replaced. To see whether

**BOX 6-3****Aspirin, the Wonder Drug at a Bargain**

References to prototypes of aspirin date back to the works of the Greek physician Galen. These references mentioned salicylate-containing plants, such as willow bark and wintergreen. We today usually attribute aspirin to the Bayer Company in Germany in the latter 1800s. Throughout its history, professionals praised aspirin for its excellent powers to relieve pain and fever (Andermann, 1996).

We think of this humble product as a cheap, over-the-counter drug that is widely available at a few pennies per dose. While fairly safe, it can have serious side effects with overdosing. Physicians, for many decades, said that “if aspirin had been proposed now as a new drug it would probably require a prescription.” While it is doubtless that it will remain an over-the-counter drug, it now seems like a new product. Because of new discoveries of its benefits to heart patients, its influence continues to expand.

This is reflected in various treatment regimens on heart outcomes. David Cutler found that the basic three regimens—intensive technologies, non-acute pharmaceuticals, and behavioral change—have approximately equal contributions to improved outcomes. The non-acute drugs (pharmaceuticals) include those to control hypertension, reduce cholesterol, treat pain, dissolve clotting, and thin the blood. Aspirin is an effective blood thinner and providers now regularly prescribe it as a preventative measure as well as to reduce the recurrence of heart attacks (Cutler, 2001).

heart attack treatment in 1995 was more expensive per episode than in 1975, we must hold quality constant within the analysis.

These are the essential problems of any price index. Consumers are familiar with the Consumer Price Index (CPI) which is used to measure inflation. A subset of this index focuses on medical care. In this case, the multiproduct character of hospital care can be confusing.

Until recently, hospital prices comprised selected components of hospital care. Room charge, nursing, lab, and other service centers each had a price index and then the results were combined into a hospital price index. Though useful, this approach often distorted the cost picture. For many years the room charge was weighted heavily, and as room charges soared, the index tended to exaggerate hospital price inflation. Even more confounding, hospital average length of stay declined rapidly in the United States, and the old price index neglected this savings. Although patients paid much more per day, it was offset in part by shorter stays. Political debate centered on these price indexes, causing considerable misunderstanding. The DRG legislation that Medicare installed in 1983 reflected in part the wide public concern that health care inflation was out of control.

A good example of making the needed adjustment is a study done by Cutler and colleagues (1999). They developed two myocardial infarction treatment price indexes that measure patient gains in quality-adjusted life-years (QALYs). They evaluated these gains in dollar terms, and allowing for errors, particularly in valuing life years, these authors proposed a range of price inflation estimates. Their research shows clearly that

1. Technological change makes a difference in patient lives.
2. The fact that substantial quality improvement had previously been omitted meant that previous estimates of inflation in health care needed to be reduced.

How well are their ideas corroborated by the data? Table 6-2 indicates that technological change has improved the quality of heart attack treatment and that this quality adjustment can turn what at first looks like price inflation into one of price deflation. Their further research suggests that the true decline in the heart attack treatment price might be much larger than these figures.

**TABLE 6-2 Comparing the Unadjusted Approach with a Quality-Adjusted Measure of Price Increases of Treatment of Myocardial Infarction (1983–1994)**

Index	Avg. Annual Price Changes
<b>Unadjusted Indexes</b>	
Official medical care CPI	3.4%
• Hospital component	6.2
• Room	6.0
• Other inpatient services	5.7
Heart attack unadjusted episode approach <sup>a</sup>	2.8
<b>Quality-Adjusted Indexes</b>	
Quality (extra years of life)	–1.5
Quality (extra QALYs) <sup>b</sup>	–1.7

<sup>a</sup> Experts recognize that several alternatives are applied when selecting for analysis the market basket of goods and services whose inflation is to be measured. The table reports a fixed-basket method; the patient is assumed to purchase essentially the same combination of medical goods in each year studied. Chain indexes allow for the representative market basket to change over time, and therefore add realism. But, how frequently should the basket be recalibrated? When a 6-year calculation of the basket is used, the average annual percentage change becomes 2.1 percent; with an annual recalculation, it becomes 0.7 percent.

<sup>b</sup> QALYs are quality-adjusted life years. This quality of treatment is the same as the previous one except for the additional consideration of the degree to which the patient is able to lead a full, active life in the years after treatment.

Source: Adapted from Cutler et al. (2001). Reprinted with permission of University of Chicago Press.

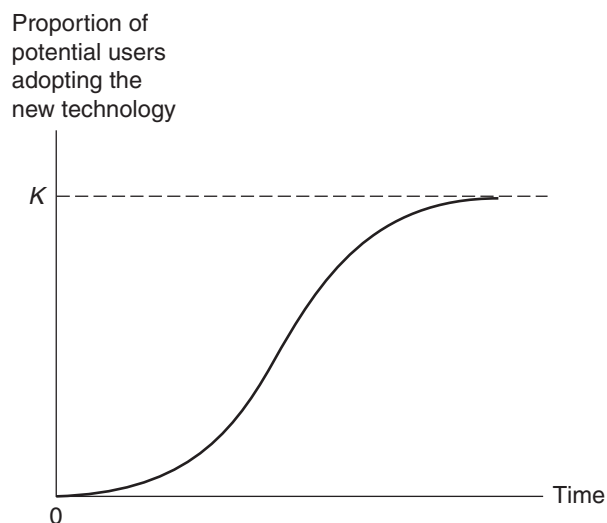
## DIFFUSION OF NEW HEALTH CARE TECHNOLOGIES

It takes time for a new product innovation to be widely adopted by providers. Some firms adopt rapidly, some slowly, and some not at all.

### Who Adopts and Why?

Those who study health technologies have found at least two basic principles that guide adopters: the profit principle and information channels. The first posits that physicians, for example, tend to adopt a new surgical technique if they expect to increase their revenues—this could happen through enhancing their prestige or by improving the well-being of their patients. The second is a compatible principle deriving originally from sociology, and it emphasizes the role of friends, colleagues, journals, and conferences in informing and encouraging the adoption decision.

Escarce (1996) emphasized the “information externalities” inherent in adoption by the first physician to adopt. An externality is the uncompensated, beneficial effect on a third person caused by the actions of a market, in this case, the actions of the first adopter. By adopting a technology, the physician communicates to friends and colleagues the expectation that the new product will benefit his or her patients and practice. The adopter paves the way for new infrastructure, new seminars, and library materials that reduce the cost of adoption for colleagues. The process tends to build on itself, perhaps at an increasing rate, until all the main body of potential adopters has acted, only then slowing the increase in total adopters until the community’s maximum potential is reached.



**FIGURE 6-10** The Diffusion of a New Technology

The data are consistent with this process; in many industries, adoption occurs slowly at first, then at an increasing rate that continues at a decreasing rate asymptotically approaching its limit. Compare this description in words with Figure 6-10, which illustrates the classic pattern of diffusion as a logistic curve. A new data set is tested empirically by estimating the logistic function where  $P_t$  is the proportion of individuals or firms adopting by time  $t$ , where the maximum potential proportion of adopters is  $K$ , with parameters  $a$  and  $b$  to be estimated:

$$P_t = \frac{K}{1 + e^{-(a+bt)}} \quad (6.2)$$

Escarce's data fit this time pattern quite well. He studied a new surgical procedure, laparoscopic cholecystectomy, introduced in 1989, which is a minimally invasive technique to remove diseased gall bladders. He found the diffusion curve to fit the logistic pattern common to diffusion studies. He then examined the differences between those who adopted and those who did not. Adopting surgeons were more likely to be younger, male, board-certified, U.S. medical school graduates, and urban-located. Younger individuals are more likely adopters. In fact, in one study older physicians proved less likely to adopt even though their expected gains in profit were much higher than for their younger counterparts (Rizzo and Zeckhauser, 1992). Others suggest that followers are more likely to emulate the "star" physicians, ones with the most impressive credentials (Burke et al., 2007).

### Other Factors That May Affect Adoption Rates

Economists believe that a firm will tend to adopt an innovation when the present value of future profits due to the innovation is positive. Waiting too long may provide competitors with an advantageous share of the market, which may be permanently sustained. However, waiting has benefits in that one may take advantage of future advances and learn from the experience of others. Waiting may reduce risks so more risk-averse firms may choose to wait somewhat longer. (For an introduction to the literature on hospital adoption of information systems, see McCullough, 2008.)

Sloan and colleagues (1986) found that mandatory rate-setting programs retard diffusion of technology in some cases, particularly in the instances of coronary bypass surgery, morbid obesity surgery, and intraocular lens implants. However, the degree of slowing tended to be small. Teplensky et al. (1995) conclude that restrictive rate-setting programs tend to retard significantly the adoption of new technology. Also, Caudill and colleagues (1995) report a slowing of the adoption of new blood dialysis technology when faced with restrictions on health care capital investment

imposed by Certificate-of-Need (CON) legislation. CON laws, which are generally applied at the state level, require hospitals to gain approval from planning agencies when they wish to expand or to make a major capital purchase (we discuss them in more detail in Chapter 20).

### Diffusion of Technology and Managed Care

Managed care arrived with hopes that it would control health care expenditure increases by removing the financial incentives for physicians to overprescribe, overtreat, and overhospitalize their patients. The same flattening of incentives—no extra money for extra treatment—potentially dampens the physician’s interest in cost-increasing technological change. As we have seen throughout this chapter, the reduction in incentives must be expected to slow innovation and the adoption of innovative technologies. Although we address the ultimate effect of managed care on health care inflation elsewhere in this text, it is appropriate here to ask: “Does a higher penetration of managed care into the health system tend to slow the growth in availability of new technologies?”

The answers seem to be “yes” for some technologies and “no” for others. Baker (2001) compared penetration of HMOs with adoptions of a new technology, magnetic resonance imaging (MRI); he found a slowing of adoptions associated with HMOs. Baker and Phibbs (2000) found that greater HMO penetration also retarded the adoption of neonate intensive care. Hill and Wolfe (1997) examine a managed care-like system in Wisconsin. These authors reported time trends of adoptions for selected technologies, and the data suggest a retardation of several of the technologies but continued growth of several others. Friedman and Steiner (1999) investigated the availability of intensive care units and found no difference in admission rates under managed care versus fee-for-service care.

## CONCLUSIONS

This chapter examined issues on the supply side of health care. Health care is fundamentally a production process, and it shares many characteristics with economic production generally. The production function, which summarizes the relationship of inputs and outputs, also embodies the technology. Technology that permits substitution between inputs provides better flexibility to the manager. The neoclassical cost function derives from the theory of production together with the theory of profit-maximizing behavior. Cost estimation describes the cost curves, which identify the economies of scale and scope. Health care firms may fail to achieve allocative or technical efficiency, or both. These analyses search for differences between for-profit, not-for-profit, and other kinds of firms.

Health firms may differ in technology because the adoption of new technologies differs among firms and is never instantaneous. Technology improvements in health care production may either increase or decrease costs depending on their effect on quality. Both market structure and regulation can affect the speed at which innovations are adopted.

---

## Summary

1. Health care goods and services can frequently be produced in different ways in the sense that they use different combinations of factor inputs.
2. The elasticity of substitution is used to measure substitution. It represents the percentage change in the ratio of factor inputs resulting from a 1 percent change in relative factor prices.
3. Economists have found some substitution not only among different kinds of medical staff but even between hospital beds and medical staff, as well as with the application of the large variety of allied health professionals.
4. The principles of cost minimization, as represented by the locus of tangencies between the firm’s isoquants

and isocost curves, are used to derive the cost curves (total and average). Economies of scale refer to a declining long-run average cost. Economies of scope represent situations where the cost of producing goods jointly is less than the sum of the costs of separate production.

5. Early empirical work on hospitals found evidence of economies of scale and an optimum size of about 250 beds. Several recent contributions find economies of scale that depend on the nature of the hospital.
6. Technical inefficiency occurs when a firm fails to achieve the maximum potential output from a given set of inputs. It can be measured as a relative distance from the frontier production function or correspondingly as a distance from the isoquants of the frontier production function. Allocative inefficiency arises in the case of competitive input markets when a firm fails to purchase inputs, given their prices, in a manner that minimizes costs.
7. We measure efficiency by frontier methods of two types. The data envelopment method estimates the frontier by statistically forming an envelope of data points representing production data. The stochastic frontier method estimates simultaneously a randomly shifting frontier and the firm's distance from that frontier.
8. Technological change in health care may reduce costs if it improves production technology of existing products or increase costs if it emphasizes new products and higher-quality versions of old products.
9. New health care technologies are adopted gradually in an industry, and the pattern of adoption fits the logistic curve. Adoption patterns are influenced by regulation, age of the adopter, profitability, and channels of communication.

---

## Discussion Questions

1. Explain whether there is any difference between goals in maximizing output for a given cost or minimizing the cost of producing a given level of output.
2. What are cross-sectional data? Why do economists find it so critical to control for case mix in studying health care cost functions? What are the analytical dangers if they do not?
3. How do legal restrictions on practice for nurses and physicians tend to affect the observed elasticities of substitution? Would the elasticities be higher if legal restrictions were removed? Would quality of care be affected?
4. Given the cost function and economies of scale and scope information reviewed in this chapter, does a policy encouraging large, centralized hospitals seem wise? Will market forces tend to reward centralization of hospital services?
5. Speculate on what types of services are more appropriate to large, regional hospitals, and what types of services are more appropriate to small, local hospitals.
6. Economists define the elasticity of substitution as the percentage change in the capital/labor ratio elicited by a 1 percent change in the factor price (wages/capital costs, for example) ratio. Would you expect the elasticity of substitution to be positive or negative? What would be the elasticity of substitution of a set of right-angled isoquants? Why?
7. Contrast technical and allocative efficiency. How can technical and allocative inefficiency in health care firms affect patient welfare?
8. What does "stochastic" mean in stochastic frontier efficiency estimation? Give several real life examples of events that could shift the production frontier.
9. Which of the following types of technological change in health care are likely to be cost increasing: (a) threats of malpractice suits that cause physicians to order more diagnostic tests on average for a given set of patient symptoms; (b) a new computer-assisted scanning device that enables physicians to take much more detailed pictures of the brain; (c) the introduction of penicillin earlier in this century; (d) greater emphasis on preventive care? Discuss.
10. As technologies diffuse, why do some firms adopt them before others? What types of technologies would you expect to be adopted most quickly? Most slowly? What factors can slow the rate of diffusion of new medical technologies?

---

## Exercises

1. Draw an isoquant that shows relatively little substitution between two factor inputs and one that shows relatively large substitution. Let the vertical axis represent capital and let the horizontal axis represent labor.
2. Draw isocost curves that are tangent to your isoquants in Exercise 1 and that each have the same slope. Mark the points of tangency and note the capital/labor ratio. Draw new, flatter isocost curves that are tangent, again each having the same new slope. Mark the points of tangency and note the capital/labor ratio. In which case is the change in the capital/labor ratio greater? Which will have a higher elasticity of substitution?

3. Determine the elasticity of substitution in the case of the isoquant in panel A of Figure 6-1.
4. Suppose a firm has the production technology shown below for Goods 1 and 2.
  - (a) Does Good 1 indicate economies of scale? Why?
  - (b) Does Good 2 indicate economies of scale? Why?
  - (c) Do the two goods indicate economies of scope? Why?

Good 1		Good 2		Both		
Q <sub>1</sub>	Cost	Q <sub>2</sub>	Cost	Q <sub>1</sub>	Q <sub>2</sub>	Cost
10	50	10	60	10	10	100
20	100	20	100	20	20	180
30	150	30	130	30	30	250

5. If any firm's price of labor and capital each double, what will happen to the expansion path (i.e., locus of tangencies between the isoquants and isocost curves)? What will happen to the firm's average cost curve?
6. Figure 6-5, panel A, illustrates technical inefficiency for firms with a one-input production function. It was explained that inefficiency could be measured by output distance or, alternatively, input reduction distance. Sketch and explain the comparable measures for the two-input production function.
7. In Escarce's account of diffusion, do improved "channels of information" matter regardless of the information content? Does all information increase the adoption rate? If not, what information does?
8. Calculate the average costs at points C, F, and G in Figure 6-2. Do they imply increasing or decreasing returns to scale? Why?

# Demand for Health Capital

- The Demand for Health
- Labor–Leisure Trade-Offs
- The Investment/Consumption Aspects of Health
- Investment over Time
- The Demand for Health Capital
- Changes in Equilibrium: Age, Wage, and Education
- Empirical Analyses Using Grossman’s Model
- Obesity—The Deterioration of Health Capital
- Conclusions

Chapter 5 considered the production of health in the aggregate by looking at the impacts of various factors on health for society as a whole. Here, we show how individuals allocate their resources to produce health. Economists’ understanding of this decision has been deepened by the important work of Michael Grossman. His model has enabled us to understand thoroughly the role of several variables, such as age, education, health status, and income in the production of health through the demand for health capital. In this edition, we introduce an analysis of obesity (being overweight) with its implications regarding human capital.

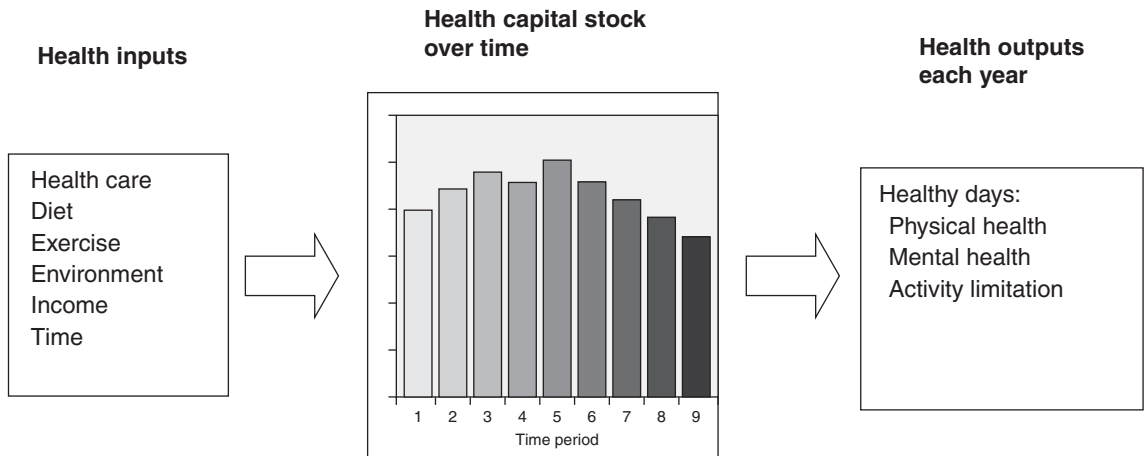
## THE DEMAND FOR HEALTH

### The Consumer as Health Producer

Grossman (1972a, 1972b) used the theory of human capital to explain the demand for health and health care. His research became a standard beginning point for much subsequent work. According to human capital theory, individuals invest in themselves through education, training, and health to increase their earnings. Grossman shows the way in which many important aspects of health demand differ from the traditional approach to demand:

1. It is not medical care as such that consumers want, but rather health. People want health; they demand medical care inputs to produce it.
2. Consumers do not merely purchase health passively from the market. Instead, they produce health, combining time devoted to health-improving efforts including diet and exercise with purchased medical inputs.
3. Health lasts for more than one period. It does not depreciate instantly, and it can be analyzed like a capital good.





**FIGURE 7-1** Investing in Health Capital

- Perhaps, most importantly, health can be treated both as a consumption good and an investment good. People desire health as a consumption good because it makes them feel better. As an investment good, health is desired because it increases the number of healthy days available to work and to earn income.

Figure 7-1 provides a simple diagram that explains the concept of health capital. Just as one thinks of a car or a refrigerator as a capital good (or “stock of capital”) that provides a stream of services over time, one can conceive of a person’s stock of health capital that provides the ultimate output of “healthy days.” One might measure the stream of output(s) as a single dimension of healthy days, or in several dimensions of physical health, mental health, and limited activity; for example, one can no longer play singles in tennis, but must play (less strenuous) doubles instead.

Consumers apply sets of health inputs, which might include not only market inputs of health care, but also diet, exercise, and time, to their physical makeup, thus making investments in health capital. These investments maintain or improve the consumers’ stocks of health, which in turn provide them with healthy days. Over time, the health stock may grow, remain constant, or decline (again, like a car or a refrigerator), either slowly with age, or more precipitously with illness or injury. As noted in Box 7-1, there may be many technologies available to produce health capital, using various amounts of time or market goods.

From Figure 7-1 and the accompanying discussion, we see how the end goal of “healthy days” guides consumer decisions as to how much time and money to invest in health stock. We will see that the prices of health care, the people’s wage rates, and their productivities in producing health, will dictate how resources are to be allocated among health capital, and other goods and services that people buy.

Consider a consumer, Ed Kramer, who buys market inputs (e.g., medical care, food, clothing), and combines them with his own time to produce a stock of health capital that produces services that increase his utility. Ed uses market inputs and personal time both to invest in his stock of health and to produce other things that he likes.

These other items include virtually all other things that Ed does. They include time spent watching television, reading, playing with and teaching his children, preparing meals, baking bread, or watching the sun set, a composite of other things people do with leisure time. We shall call this composite home good *B*.

### Time Spent Producing Health

An increment to capital stock, such as health, is called an investment. During each period, Ed produces an investment in health, *I*. Health investment *I* is produced by time spent improving health,

**BOX 7-1****Health Capital: A Substitute for Exercise?**

Some forms of health investment such as playing sports can be thought of as pleasurable, while others such as diet might not be. For all who have wondered if they could enjoy the benefits of exercise without the pain (or the time spent) of exertion, an article in the *New York Times* by Nicholas Wade notes that the answer may one day be yes: “Just take a pill that tricks the muscles into thinking they have been working out furiously.”

Researchers at the Salk Institute in San Diego reported that they had found two drugs that did wonders for the athletic endurance of couch potato *mice*. One drug, known as Aicar, increased the mice’s endurance on a treadmill by 44 percent after just four weeks of treatment. A second drug, GW1516, supercharged the mice to a 75 percent increase in endurance but had to be combined with exercise to have any effect. “It’s a little bit like a free lunch without the calories,” said Dr. Ronald M. Evans, leader of the Salk group.

The results, Dr. Evans said, seem reasonably likely to apply to people, who control muscle tone with the same underlying genes as do mice. If the drugs work and prove to be safe, they could be useful in a wide range of settings. They should help people who are too frail to exercise and those with health problems like diabetes that are improved with exercise, Dr. Evans said.

Experts not involved in the study agreed that the drugs held promise for treating disease. Dr. Richard N. Bergman, an expert on obesity and diabetes at the University of Southern California, said the drugs might prove to have serious side effects but, if safe, could become widely used. “It is possible that the couch potato segment of the population might find this to be a good regimen, and of course that is a large number of people.”

*Source:* Based on information from Nicholas Wade, *New York Times*, August 1, 2008. [www.nytimes.com/2008/08/01/science/01muscle.html?\\_r=1&ref=health&oref=slogin](http://www.nytimes.com/2008/08/01/science/01muscle.html?_r=1&ref=health&oref=slogin)

$T_H$ , and market health inputs (providers’ services, drugs),  $M$ . Likewise, home good  $B$  is produced with time,  $T_B$ , and market-purchased goods,  $X$ .

If, for example, we considered good  $B$  to be baking bread, the market goods might include flour, yeast, kitchen appliances, and gas, water, and/or electricity. Thus, Ed uses money to buy health care inputs,  $M$ , or home good inputs,  $X$ . He uses leisure time either for health care ( $T_H$ ) or for producing the home good ( $T_B$ ). Using functional notation:

$$I = I(M, T_H) \quad (7.1)$$

$$B = B(X, T_B) \quad (7.2)$$

These functions indicate that increased amounts of  $M$  and  $T_H$  lead to increased investment  $I$ , and that increased amounts of  $X$  and  $T_B$  lead to increased home good  $B$ .

In this model, Ed’s ultimate resource is his own time. Treat each period of analysis as being a year, and assume that Ed has 365 days available in the year. To buy market goods such as medical care,  $M$ , or other goods,  $X$ , he must trade some of this time for income; that is, he must work at a job. Call his time devoted to work  $T_W$ .

Because our focus is on the health aspects of living, we realize that some of his time during each year might involve ill health, or  $T_L$ . Thus, we account for his total time in the following manner:

$$\begin{aligned} \text{Total time} = T = 365 \text{ days} &= T_H(\text{improving health}) + T_B(\text{producing home goods}) \\ &+ T_L(\text{lost to illness}) + T_W(\text{working}) \end{aligned} \quad (7.3)$$

Recall that his leisure time is spent either improving his health or producing home goods.

**LABOR–LEISURE TRADE-OFFS**

The labor–leisure trade-off illustrates the potential uses of Ed’s time. Our variation on this analysis also helps illustrate the investment aspects of health demand.

### Trading Leisure for Wages

In Figure 7-2, the  $x$ -axis represents Ed's work and leisure time. Suppose that he considers his time spent creating health investment to be "health-improvement time" and that he calls  $T_B$  his leisure. In reality, he may do some health-improving activities at work, may obtain some enjoyment or satisfaction from healthful time, and so on, but assume here that these categories are exclusive. Assume further that the number of days lost to ill health and the number of days spent on health-enhancing activity are fixed (we relax this assumption later). Variables  $T_L$  and  $T_H$  refer to time lost and time spent on healthy activities, respectively. The maximum amount of time that he has available to use either for work,  $T_W$ , or leisure,  $T_B$ , is thus  $365 - T_H - T_L$ , so:

$$\text{Time Available for Work or Leisure} = 365 \text{ days} - T_H - T_L = T_B + T_W \quad (7.4)$$

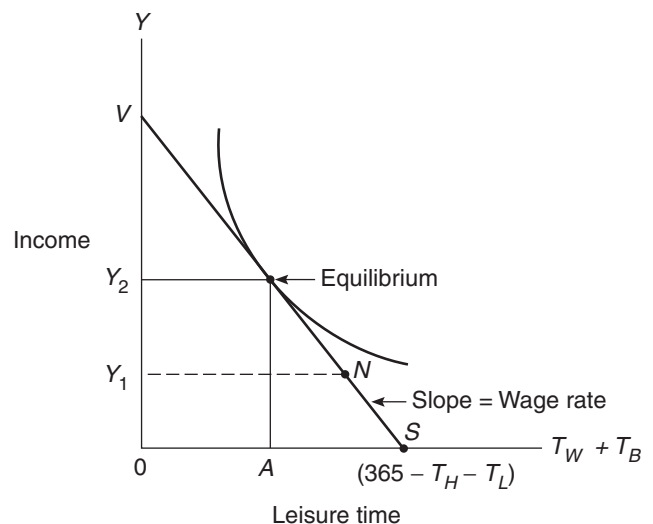
Leisure time,  $T_B$ , is measured toward the right while time spent at work,  $T_W$ , is measured toward the left. Figure 7-2 shows that if Ed chooses leisure time,  $OA$ , then he has simultaneously chosen the amount of time at work indicated by  $AS$ .

Recall that Ed's total amount of time available for either work or leisure is given by point  $S$ . If he were to choose point  $S$  for the period, he would be choosing to spend all this available time in leisure; that is, in the pursuit of the pleasures of life (albeit without the wage income to produce them). The  $y$ -axis represents income, obtained through work. This income will then purchase either market health goods or other market goods. Thus, if he chooses point  $S$ , he will not be able to purchase market goods because he has no wage income.<sup>1</sup>

If, beginning at  $S$ , Ed gives up one day of leisure by spending that day at work, to point  $N$ , he will generate income equal to  $OY_1$ , which represents his daily wage. In economic terms, this quantity represents income divided by days worked—that is, the daily wage. The slope of the line  $VS$  depicting the labor–leisure trade-off reflects the wage rate (if Ed pays Social Security and/or income taxes on his wage, then the slope reflects the after-tax wage rate).

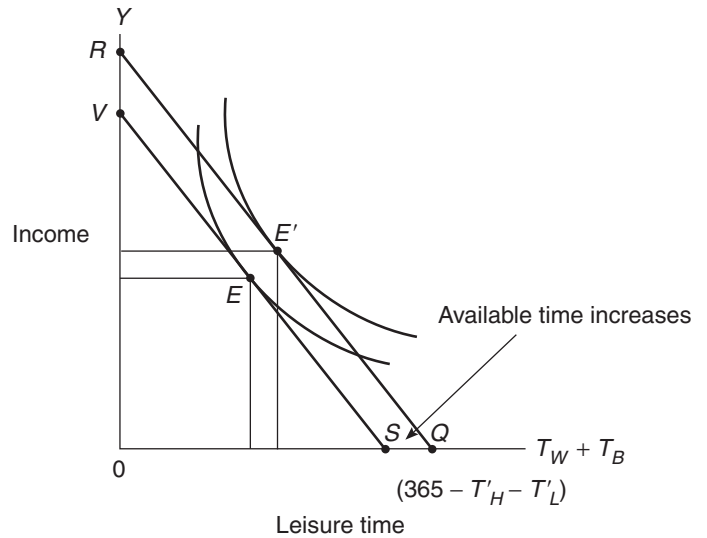
### Preferences between Leisure and Income

Consumers have preferences regarding income and leisure, just as they had among other goods in Chapter 2. As before, Ed would like more income and more leisure so the indifference curve map is



**FIGURE 7-2** Labor–Leisure Trade-Off

<sup>1</sup> We ignore here income from nonwork efforts—for example, through returns to financial investments, such as saving, stock, or bonds.



**FIGURE 7-3** Increased Amount of Healthy Time Due to Investment

shaped normally. In Figure 7-2, in equilibrium, Ed’s tradeoff of leisure and income is the same as the market trade-off, which is the wage rate. Here, he takes amount  $OA$  of leisure and trades amount  $AS$  of leisure for income,  $0Y_2$ .

In Figure 7-3, Ed has made a different choice with respect to time spent investing in health status. To illustrate, suppose that time spent on health-producing activities,  $T_H$ , is increased to  $T'_H$ . Correspondingly, suppose that the number of days lost to ill health has been reduced to  $T'_L$ . What effect will this change in time have on the horizontal intercept, which is the total time remaining for work or leisure? On the one hand, the time he spends producing health reduces his time available for other activities. Time spent on health investment increases health stock and, in turn, reduces time lost to illness.

If the net effect of  $T'_H + T'_L$  is a gain in available time, then this illustrates the pure investment aspect of health demand. The health investments “pay off” in terms that both add to potential leisure and also increase the potential income, shifting the income–leisure line outward from  $VS$  to  $RQ$ . The expenditure of time (and medical care, too) for health-producing activities may later improve Ed’s available hours (because he is sick less) of productive activity.

As a result of his investment, Ed can increase his utility, moving from point  $E$  to point  $E'$ . Not only does investment in health lead to his feeling better, but it also leads to more future income and may lead to more leisure, as well.

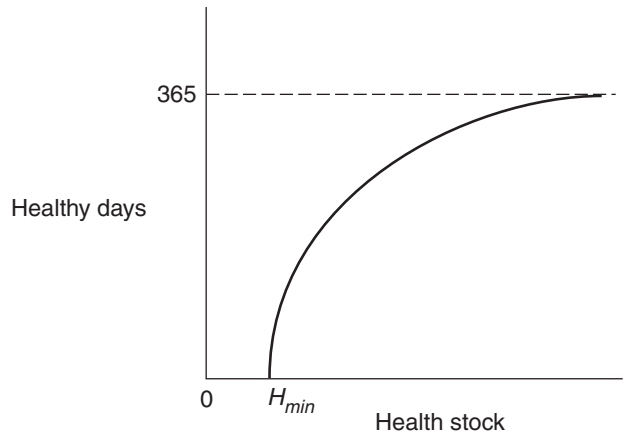
The improvement in health status also might increase Ed’s productivity at work, perhaps resulting in a higher wage and a steeper income–leisure line (why is it steeper?). In any case, the analysis shows that Ed might wish to engage in activities to improve his health, even if the only value of health is its effect on his ability to earn future income.

## THE INVESTMENT/CONSUMPTION ASPECTS OF HEALTH

The Grossman model describes how consumers simultaneously make choices over many periods or years. It can also be instructive to represent a whole life span as a single period. This can show the dual nature of health as both an investment good and a consumption good.

### Production of Healthy Days

For our purposes, we will view health as a productive good that produces a single output, healthy days, a production function relationship illustrated in Figure 7-4. The horizontal axis measures health stock in a given period. A larger stock of health leads to a larger number of healthy days, up



**FIGURE 7-4** Relationship of Healthy Days to Health Stock

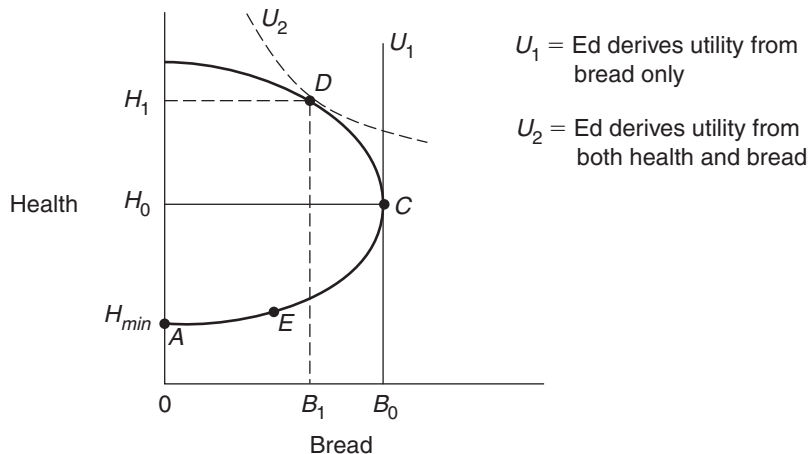
to a natural maximum of 365 days. The bowed shape of the curve illustrates the law of diminishing marginal returns (additional resources have decreasing marginal impacts on the output). Note also the concept of a health stock minimum shown as  $H_{min}$ . At this point, production of healthy days drops to zero, indicating death.

**Production of Health and Home Goods**

Consider the possibilities for producing health,  $H$ , and home good,  $B$ , given the total amount of time available. Figure 7-5 shows the production possibilities trade-off. The curve differs from the usual production possibilities curve in several respects. First, from point  $A$  to point  $C$ , health improvements increase the amounts of the home good,  $B$ , and health,  $H$ , attainable. It is necessary to increase health beyond  $H_{min}$  in order to obtain income and leisure time from which to produce  $B$ .

Moving along the production possibilities curve, Ed shifts his uses of available time and distributes his purchases of market goods. The move from  $E$  to  $C$  indicates that he has made more time available for health and that this move has reaped the side benefit (increased leisure time) of increasing the availability of market goods and time used to increase production of the home good, baking bread.

Suppose that Ed desires health solely for its effect on the ability to produce income and the leisure time to produce the home good bread,  $B$ . This would imply that his indifference curves between  $H$  and  $B$  are vertical lines. (Ed places no intrinsic value on  $H$ , so he would not trade any  $B$  to get additional health.) In such a case, he would maximize his utility by producing as much  $B$  as possible.



**FIGURE 7-5** Allocation of Production Between Health and Bread

The utility-maximizing choice would be at point *C*, a point of tangency between indifference curve  $U_1$  and the production possibilities curve. He produces amount  $B_0$  of the home good, and  $H_0$  of health.

Now assume instead that Ed achieves utility not only from producing  $B$ , but also directly from health itself (he likes feeling better). In this case, his (dashed) indifference curve,  $U_2$ , has the more familiar curved shape in Figure 7-5, sloping downward from left to right. It is more realistic to say that he values health both as a consumption good, as is shown in Figure 7-5, and as an investment in productive capacity. The consumption aspect suggests that he enjoys feeling healthy; the investment aspect, that feeling healthy makes him more productive, thus allowing him to earn more. In general, by including Ed's "feeling healthy" in this consumption feature of the model, he will choose a higher health stock than under the pure investment model. In Figure 7-5, health stock,  $H_1$ , exceeds  $H_0$ . The cost of this increase in  $H$  involves foregoing some of the home good  $B$ , such that  $B_1$  is less than  $B_0$ .<sup>2</sup>

## INVESTMENT OVER TIME

### The Cost of Capital

People make choices for the many periods over their life cycles, rather than just for one representative period. As a beginning point for each analysis, we feature the pure investment version of the model (point *C* in Figure 7-5). We then discuss the analytical changes when consumers, in addition, value health intrinsically (point *D* in Figure 7-5). We demand health capital because it helps us earn more and feel better. What does it cost? By analogy, a health clinic purchases thousands of dollars of X-ray equipment. The return to the X-ray equipment is in the future earnings that ownership of the equipment can provide.

Suppose that an X-ray machine costs \$200,000, and that its price does not change over time. Suppose that the annual income attributable to the use of the X-ray machine is \$40,000. Is purchasing the machine a good investment? Consider the alternative: Instead of purchasing the X-ray machine, the clinic could have put the \$200,000 in a savings account, at 5 percent interest, yielding the following:

$$200,000 \times 1.05 = 210,000 \text{ at the end of Year 1}$$

$$210,000 \times 1.05 = 220,500 \text{ at the end of Year 2}$$

$$220,500 \times 1.05 = 231,525 \text{ at the end of Year 3}$$

$$231,525 \times 1.05 = 243,101 \text{ at the end of Year 4}$$

$$243,101 \times 1.05 = 255,256 \text{ at the end of Year 5}$$

For the investment in an X-ray machine to be desirable by these criteria, it should provide at least \$55,256 in incremental revenue over the five years.

The problem is more complicated, however, because capital goods depreciate over time. Students will agree that a five-year-old personal computer is worth almost nothing. Even though it may do everything it ever did, new programs may not work on it, new equipment may not hook up to it properly, and it may be very slow compared to new machines. In economic terms, the machine has *depreciated*, and if parts wear out it may depreciate physically as well. If the computer originally cost \$2,000, and the students have not budgeted \$2,000 for replacement, they may find themselves without working computers!

Similarly, suppose that the clinic knows that the X-ray machine will wear out (or depreciate), so that it will be worth only half its original value in five years. The clinic must earn enough not

<sup>2</sup> Goodman, Stano, and Tilford (1999) provide a more detailed model addressing the production of health and home goods using both market goods, and people's time.

only to cover the opportunity cost from the bank, but also to maintain the value of the machine. For an investment in an X-ray machine to be worthwhile, then, it must not only earn the competitive 5 percent return each year, but it also must provide enough return to cover depreciation. This suggests that the cost of holding this capital good for any one year, as well as over time, will equal the opportunity cost of the capital (interest foregone) plus the depreciation (deterioration of value). In this example, the depreciation cost is \$100,000, or half of the \$200,000 original cost.

## THE DEMAND FOR HEALTH CAPITAL

Conventional economic analysis provides a powerful conceptual apparatus by which to analyze the demand for a capital good. The cost of capital, in terms of foregone resources (for health capital, both time and money), is a supply concept. The other needed tool is the concept of the marginal efficiency of investment (*MEI*) a demand concept that relates the return to investment to the amount of resources invested.

### Marginal Efficiency of Investment and Rate of Return

The *MEI* can be described in terms of the X-ray machine example. A busy clinic may wish to own more than one X-ray machine. But how many? The clinic management may logically consider them in sequence. The first machine purchased (if they bought only one) would yield a return as we have discussed. Suppose that return each year is \$40,000.

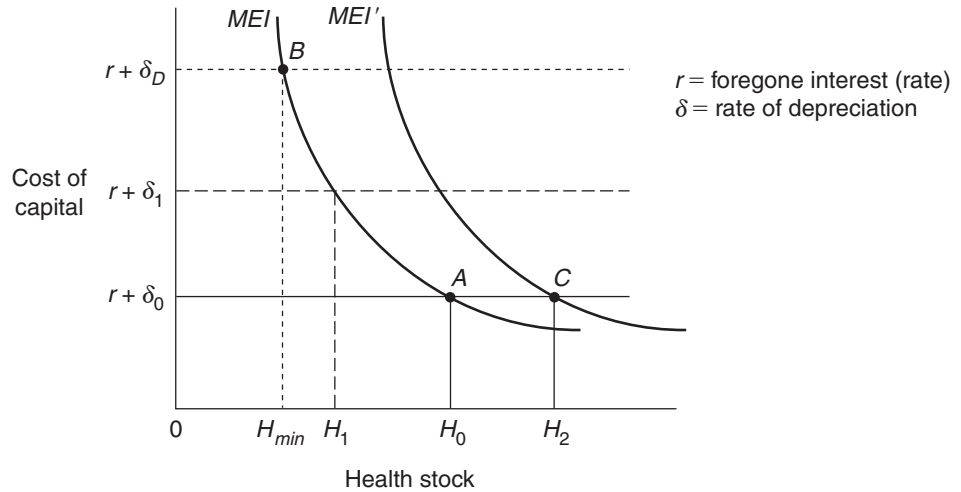
We also can calculate the annual *rate of return*, which would be  $\$40,000 \div \$200,000$ , or 20 percent per year. The management would buy this machine if the incremental revenue brought in covered its opportunity cost of capital and the depreciation. In terms of rates, management would choose to own the first X-ray machine as long as the rate of return, 20 percent, exceeded the interest rate (the opportunity cost of capital) plus the depreciation rate.

If management considered owning two machines, it would discover that the rate of return on the second X-ray machine would probably be less than the first. This is best understood by recognizing that a clinic buying only one X-ray machine would assign it to the highest-priority uses, those with the highest rate of return. If the clinic were to add a second X-ray machine, then logically it could be assigned only to lesser-priority uses (and might be idle on occasion). Thus, the second machine would have a lower rate of return than the first. The clinic would then purchase the second machine only if its rate of return were still higher than interest plus depreciation.

### The Decreasing *MEI*

Other machines probably could be added at successively lower rates of return. In Figure 7-6, the marginal efficiency of investment curve, *MEI*, describes the pattern of rates of return, declining as the amount of investment (measured on the horizontal axis) increases. The cost of capital (that is, the interest rate,  $r$ ; plus the depreciation rate,  $\delta_0$ ) is shown as the horizontal line labeled  $r + \delta_0$ . The optimum amount of capital demanded is thus  $H_0$ , which represents the amount of capital at which the marginal efficiency of investment just equals the cost of capital. This equilibrium occurs at point A.

Like the marginal efficiency of investment curve in this example, the *MEI* curve for investments in health also would be downward sloping. This occurs because the production function for healthy days (see Figure 7-4) exhibits diminishing marginal returns. The cost of capital for health would similarly reflect the interest rate plus the rate of depreciation in health. Understand that a person's health, like any capital good, also will depreciate over time. As we age, certain joints may wear out, certain organs may function less well than before, or we may become more forgetful. Thus, the optimal demand for health is likewise given at the intersection of the *MEI* curve and the cost of capital curve ( $r + \delta_0$ ).



**FIGURE 7-6** Optimal Health Stock

## CHANGES IN EQUILIBRIUM: AGE, WAGE, AND EDUCATION

Return to Ed Kramer and to Figure 7-1. Based on the analyses thus far, Ed has chosen an equilibrium level of health stock, by deciding how much to work, how much time to spend on health, what kind of diet, and how much exercise to do. He allocates his resources such that every year he maintains a constant level of health stock, and this provides him with an equilibrium level of healthy days per year. How does his investment in health change in response to changes in age, wage, and education? The model depicted in Figure 7-6 provides a useful way to investigate several important model implications. Consider age first.

### Age

How does Ed's optimal stock of health vary over a lifetime? In this model, death itself is *endogenous*, meaning that the age of death is determined as part of the model—it doesn't just happen! Ed chooses his optimal life span, a life span that is not infinite. By this model, all of us, at some time, will optimally allow our health stock to dissipate to  $h_{min}$ . This feature depends in a critical way on how the depreciation rate (a cost factor) varies with age, as well as how long the person expects to live (and enjoy the benefits of good health).

Looking first at costs, Ed's health stock may depreciate faster during some periods of life and more slowly during others. Eventually, as he ages, the depreciation rate is likely to increase. In other words, the health of older people is likely to deteriorate faster than the health of younger people.

Consider then the effect of aging on Ed's optimal health stock. Return to Figure 7-6. We assume that the wage and other features determining the *MEI* are not substantially altered by this aging. However, by hypothesis, the depreciation rate,  $\delta$ , increases with age from  $\delta_0$  to  $\delta_1$  and ultimately to  $\delta_D$ . These assumptions imply that the optimal health stock decreases with age.

This situation is shown in Figure 7-6 by the fact that the optimal health at the younger age,  $H_0$ , is greater than  $H_1$ , the optimal stock at the older age. Higher depreciation rates increase the cost of holding health capital stock. We adjust to this by holding more health capital in periods when health is less costly. In old age, health depreciation rates are extremely high,  $\delta_D$ , and optimal health stock falls to  $H_{min}$  at point *B*.



This conclusion is consistent with the observation that elderly people purchase a greater amount of medical care, even as their health deteriorates. Grossman explains the phenomenon:

Gross investment's life cycle profile would not, in general, simply mirror that of health capital. This follows because a rise in the rate of depreciation not only reduces the amount of health capital demanded by consumers but also reduces the amount of capital supplied to them by a given amount of gross investment. (p. 238)

Turning to the returns from investment, consider by way of analogy a consumer who has two cars, the same models, built by the same maker. One is two years old, with 30,000 miles of wear; the other is 12 years old, with 180,000 miles of wear. Each car has had the "bumps" and "bruises" that accompany driving in a major city. Suppose that the two-year-old car is damaged in the parking lot and will cost \$3,000 to repair. The decision is obvious. However, would the owner incur the same level of costs (more pointedly, would the car insurer reimburse these costs) to repair the 12-year-old car, which is most likely near the end of its useful life?

The analogy to individual health is immediate. A younger person may choose complicated surgery to replace knees or hips, to maintain earning capability or quality of life. An older person may choose not to do so. In Figure 7-6, this suggests that as the expected length of life decreases, the *MEI* curve shifts to the left, because the returns from an investment will last for a shorter period of time. This will reinforce the decrease in investment that occurs due to increased depreciation.

Other features of the model suggest that people will increase their gross investment (the amount of dollars spent) in health as they age. This suggests, in turn, that the elderly would demand more medical care than the young, as we frequently note to be the case.

Thus, the pure investment model generates the prediction that optimal health will decline as the person ages. Will this prediction change when we assume more realistically that an individual also will value health for consumption reasons (good health makes one feel better)? The issue relies on whether older persons get more or less direct utility from the enjoyment of healthy days. If people increase their valuation of healthy days as they age, this somewhat offsets the predicted health stock decline.

## Wage Rate

Figure 7-6 also illustrates the effect of a change in the wage rate on Ed's optimal level of investment. Increased wage rates increase the returns obtained from healthy days (8 hours' work will bring in \$160 rather than \$120 if the wage rate increases from \$15 to \$20 per hour). Thus, higher wages imply a higher *MEI* curve, or *MEI'*.

Assume now that the original *MEI* curve describes the lower-wage case and yields optimal health stock,  $H_0$ . The *MEI'* curve, above *MEI*, shows the marginal efficiency of investment for someone with higher wages. At new equilibrium point *C*, the higher wage will imply a higher optimal level of health stock,  $H_2$ , in this pure investment model. The rewards of being healthy are greater for higher-wage workers, so increased wages will generally tend to increase the optimal capital stock.<sup>3</sup>

The model illustrates one more implication of the wage factor. Consider that when Ed retires, his wage effectively drops to zero. The pure investment version implies that he would change his optimal health stock to  $H_{min}$  upon retirement. Once he retires, he would make no further investment in health, but instead would allow health to depreciate until death.

<sup>3</sup> This result *may* be ambiguous. Although the increased wage rate potentially increases the return to investment, it also represents an increased opportunity cost of time in producing health investment. Hence, the *MEI* curve could shift downward, and it is possible that the equilibrium demand for health investment will fall.

How would we amend this analysis by considering the consumption aspects of health—that good health makes people feel good? First, Ed would presumably continue to obtain utility directly from healthy days. Thus, optimal health stock would not necessarily drop to  $H_{min}$  directly upon Ed's retirement, but it would do so only when depreciation rates became sufficiently severe.

Second, if retirees and those who are still working obtain utility directly from healthy days, then the only significant change upon retirement would involve the pure investment aspects. Therefore, even when we include the consumption aspects of health, we would expect people to reduce their health stock upon retirement.

## Education

Education is especially interesting to those who study health demand. Those with higher education often have better health, and most economists believe that education may improve the efficiency with which people can produce investments to health and the home good. Examples of improved efficiency may include improved ability to follow instruction regarding medicines or better knowledge of harmful effects of smoking, drinking, or addictive drugs, to name just a few.

Figure 7-6 illustrates the effect of education. Here, the  $MEI$  curve illustrates the marginal efficiency of investment for the consumer with a low level of education (measured, for example, by years of schooling), while the  $MEI'$  curve illustrates the same person with a higher level of education. This model indicates that because education raises the marginal product of the direct inputs, it reduces the quantity of these inputs required to produce a given amount of gross investment.

It follows that given investments in health can be generated at less cost for educated people, and thus they experience higher rates of return to a given stock of health. The result, as shown, is that the more educated people will choose higher optimal health stocks,  $H_2$ , than the less educated, who will choose  $H_0$ .

This explains the widely observed correlation between health status and education. Educated people tend to be significantly healthier. However, this explains only the correlation of health status and education from the supply side in that it considers only the increased efficiency with which we produce health. One also might wish to explain education from the demand side.

Educated people most likely recognize the benefits of improved health. They may enjoy preparing and eating nutritious food or doing physical exercise. They may recognize the dangers of smoking and the long-term problems of overexposure to the sun. They may enjoy feeling and looking good. As such, all else equal, they would have a greater taste for health relative to other goods.

The demand for health due to education is difficult to separate from the supply effect of education, which implies more productivity in producing health. Clearly, however, both exist and both are important.

## EMPIRICAL ANALYSES USING GROSSMAN'S MODEL

In the last several years, many researchers have examined health production using Grossman's model. The resulting literature examines the separable impacts of age and education. It also looks at the various time components of health investment within families. The analytical methods and insights from the Grossman model permit researchers to examine the impacts of health status on the demand for health and health care. Box 7-2 shows how the model addresses addictive behavior.

Sickles and Yazbeck (1998) developed and estimated a structural model of health production that looks at the demand for leisure and the demand for consumption for elderly males. Measuring health production is a difficult problem. The authors use the Quality of Well-Being (QWB) scale, developed by Kaplan and Anderson (1988), based on mobility, physical activity, social activity, and physical symptoms and problems.

They find that both health care and leisure consumption tend to improve health. A 1 percent increase in health-related consumption increases health by 0.03 to 0.05 percent. A 1 percent increase in leisure increases health by from 0.25 to 0.65 percent.

**BOX 7-2****Rational Addiction**

Economists Gary Becker and Kevin Murphy (1988) describe conditions under which rational people choose to consume addictive goods. Their work challenges our common moral concepts and attitudes toward addiction. It also offers similarities and contrasts with Grossman's model of the demand for health.

Becker and Murphy argue that addictive behavior (as most current or former smokers would agree) must provide some pleasure or people would not pursue it. For a good to be addictive, in their model, past consumption (of the good) must increase the marginal utility of current consumption. Past consumption contributes to "consumption capital" of the good in question. Current smoking may entail a learning experience in the sense that future consumption of an additional cigarette becomes more enjoyable. This important feature also applies to many goods and services and does not necessarily separate those goods that society has made taboo or bad. For example, drinking coffee or listening to Mozart also may entail learning experiences. Mozart symphonies may become more enjoyable in the future as this learning experience continues. Coffee drinkers learn to enjoy the beverage but discover that stopping usage can lead to symptoms of withdrawal.

Many addictions, however, are harmful. Harmful, in the Becker and Murphy context, means that the capital good—the consumption capital—has harmful effects similar to a reduction in health status in the Grossman model. Smoking cigarettes may reduce healthy days and may reduce income. Such harmful effects are part of what Becker and Murphy call the "full price" of the addictive good.

A potential cigarette smoker might choose to start, even knowing that it is addictive, if smoking tends to increase future enjoyment from smoking more than it increases future harmful effects from smoking. This explains why a rational person may choose to ingest a harmful substance, knowing fully its harmful effects.

(See Chapter 24 for a more detailed discussion of rational addiction.)

Gerdtham and Johannesson (1999) estimate health demand with a Swedish sample of over 5,000 adults. They use a categorical measure of overall health status to measured health capital. They find that the health demand increases with income and education and decreases with age, urbanization, being overweight, and being single.

We can also study health investment by looking at the demand for preventive care. Kenkel (1994) estimates the determinants of women's purchases of medical screening services, designed for the early detection of breast and cervical cancer. Annual use of these screening tests decreases with age, a result consistent with women's rationally reducing care as the payoff period shortens over the life cycle and as the depreciation rate rises. Furthermore, increased schooling tends to increase the use of the screening services, implying more efficiency in producing health.

Mullahy and Sindelar (1993) examine the relationships among alcoholism, income, and working. Poor health may reduce income either by reducing productivity, which results in lower wages, or by reducing labor market participation (whether and/or how much one is working). Alcoholism is an illness that reduces people's health capital. Mullahy and Sindelar find the labor market participation effects are more powerful than the wage (productivity) effects in reducing earnings, and hence reducing the return to health. By these criteria, successful alcoholism treatment would appear to have significantly positive economic returns.

In a retrospective essay on research achievements and directions in the 30 years after Grossman's pioneering analysis, Leibowitz (2004) finds that increases in the parents' valuations of time will also affect the relative costs of alternative inputs to children's health. As a result, mothers who work outside the home may employ substitutes for their own time that are less effective in producing child health. For example, working mothers can substitute prepared foods for their own time in producing meals for their children. However, these prepared foods are often high-calorie and high-fat, perhaps leading to less nutritious diets for their children (Leibowitz, 2003; Anderson et al., 2003).

In yet another application, Borisova and Goodman (2004) examine the importance of time in the compliance of drug abusers with methadone treatment, as a substitute for heroin. Because most drug abuse treatment comes at only nominal or zero cost, the time spent traveling to, and receiving, the daily treatment becomes critical. The authors find, as expected, that increased travel and treatment time significantly reduce treatment compliance.

Finally, Bhattacharya and Lakdawalla (2006) examine the value to the U.S. labor force of improvements in survival and health over the years 1970 to 1999. They find that survival gains and reductions in the number of work-days missed due to poor health have added about 8 percent to the remaining labor force value of black males, and about the same to the value of 60-year-old white males. They note that these effects are almost as large as a full year of schooling. Gains for younger white males appear to have been approximately 5 percent, and gains for women are around 2 percent. Overall, they estimate that health improvements have added \$1.5 trillion to the value of labor market human capital over this period. Even a 5 percent return on this human capital would lead to an increase of \$75 billion per year in GDP, a substantial amount!

All of these examples indicate that Grossman's model has yielded considerable insight into the determinants of health and into the allocation of resources (both time and money) into health-creating activities. It has also crossed fields of economics, including labor, development, and growth economics, and has provided fruitful results in cross-disciplinary demographic and sociological research as well.

## OBESITY—THE DETERIORATION OF HEALTH CAPITAL

Obesity (excess weight) provides many insights in a model of health capital. Aside from aesthetic issues regarding appearance, obesity is a leading risk factor for heart disease, hypertension (high blood pressure), certain cancers, and type-2 diabetes.

According to reports from the Center for Disease Control in 2011, over one-third of U.S. adults (more than 72 million) people and 17% of U.S. children are obese. From 1980 through 2008, obesity rates for adults doubled and rates for children tripled. During the past several decades, obesity rates for all groups in society regardless of age, sex, race, ethnicity, socioeconomic status, education level, or geographic region have increased markedly.

Health analysts usually measure obesity in terms of Body Mass Index, or BMI, with the formula  $BMI = \frac{\text{Weight in kilograms}}{(\text{height in meters})^2}$ . BMI adjusts for the fact that tall people are generally heavier.

If Ed Kramer weighs 90 kg (almost 200 pounds) and is 1.75 meters (about 5 feet 9 inches) tall, he has a BMI of 29.4. Table 7-1 provides the generally accepted measures of weight relative to body size measured by height.

**TABLE 7-1 Weight Status Classified by Body Mass Index**

Category	BMI range
Severely underweight	less than 16
Underweight	16 to 18.5
Normal	18.5 to 25
Overweight	25 to 30
Obese Class I	30 to 35
Obese Class II	35 to 40
Obese Class III	40 and above

Source: World Health Organization. [http://www.who.int/bmi/index.jsp?introPage=intro\\_3.html](http://www.who.int/bmi/index.jsp?introPage=intro_3.html).

**TABLE 7-2 2009 US State Obesity Rates**

Alabama	31.0	Illinois	26.5	Montana	23.2	Rhode Island	24.6
Alaska	24.8	Indiana	29.5	Nebraska	27.2	South Carolina	29.4
Arizona	25.5	Iowa	27.9	Nevada	25.8	South Dakota	29.6
Arkansas	30.5	Kansas	28.1	New Hampshire	25.7	Tennessee	32.3
California	24.8	Kentucky	31.5	New Jersey	23.3	Texas	28.7
Colorado	18.6	Louisiana	33.0	New Mexico	25.1	Utah	23.5
Connecticut	20.6	Maine	25.8	New York	24.2	Vermont	22.8
Delaware	27.0	Maryland	26.2	North Carolina	29.3	Virginia	25.0
Washington DC	19.7	Massachusetts	21.4	North Dakota	27.9	Washington	26.4
Florida	25.2	Michigan	29.6	Ohio	28.8	West Virginia	31.1
Georgia	27.2	Minnesota	24.6	Oklahoma	31.4	Wisconsin	28.7
Hawaii	22.3	Mississippi	34.4	Oregon	23.0	Wyoming	24.6
Idaho	24.5	Missouri	30.0	Pennsylvania	27.4		

Source: <http://www.cdc.gov/obesity/data/trends.html#State>, accessed April 1, 2011.

As noted in Table 7-1, a BMI rating of over 25 is classified as overweight. By this criterion, Ed is overweight, and if he weighed 2 kilograms more, with a BMI slightly over 30, he would be considered (Class I) obese.

Table 7-2 indicates that 33 states in 2009 in the United States had obesity prevalence (BMI greater than 30) equal to or greater than 25% of their adult populations, and nine of these states (Alabama, Arkansas, Kentucky, Louisiana, Mississippi, Missouri, Oklahoma, Tennessee, and West Virginia) had a prevalence equal to or greater than 30%.

Obesity describes health capital, in that it may make the body less productive, more susceptible to disease, and possibly cause it to depreciate more quickly. We will therefore look to see what part of the health capital model may explain it. We then explain some of its economic effects and finish with economic explanations as to why it has increased.

### **An Economic Treatment of Obesity**

This discussion closely follows a model derived by Yaniv, Rosin, and Tobol (2009). They note that the human body needs energy to function, with food being the fuel that creates this energy. Potential energy exists in the form of calories burned in the process of daily functioning, and the body accumulates unburned energy in the form of fat tissues that increase bodyweight. People will gain or lose weight depending on the relationship of total calories consumed to total calories expended.

The body expends calories both in physical activity and at rest. The rest component, known as Basal Metabolic Rate (BMR), is the largest source of energy expenditure, reflecting blood circulation, respiration and daily maintenance of body temperature. While the BMR is determined by physical characteristics (such as gender, age, weight and height), calories expended through physical activity, as well as calorie intake through food consumption, are subject to choice. Differing BMRs among individuals indicate why one person can “eat like a horse” and gain little weight, while another may gain weight with far less intake of food.

This economic theory of obesity views weight gain as the outcome of rational choice that reflects a willingness to trade off some future health for the present pleasures of less restrained eating and lower physical activity. Although economic models of obesity usually focus on food consumption in general as the source of energy, foods vary in their calorie content. Junk food, for example, is relatively high in calories, while healthy food is lower in calories.

Suppose that overweight individuals can determine consumption of junk-food meals,  $F$ , and healthy meals,  $H$ . They may also choose their level of exercise,  $x$ . The model defines the weight gain during a period, or obesity,  $S$ , as:

$$S = \delta F + \varepsilon H - \mu x - BMR \quad (7.5)$$

where  $\delta$  and  $\varepsilon$  represent calorie intake per junk-food meal  $F$  and healthy meal  $H$ , respectively (with  $\delta > \varepsilon$ ), and  $\mu$  represents calorie expenditure per instant of physical activity. In plain terms, people who eat too much and do not exercise enough, will get fat.

Yaniv and colleagues note that people may eat even when they are not hungry, in social or stressful situations, and this type of eating may be composed of snacks, which are high in calories (i.e., part of  $F$ ). Using  $F^S$  and  $M$  to denote snacks and hunger-induced meals, suppose that people can satisfy their hunger through either junk-food  $F$  or healthy meals  $H$ . Hence,

$$\text{Meals} + \text{Snacks} = M + F^S = F + H: \quad (7.6)$$

Substituting Eq. (7.6) in Eq. (7.5), the obesity function becomes

$$S = \varepsilon M + (\delta - \varepsilon)F + \varepsilon F^S - \mu x - BMR \quad (7.7)$$

If people satisfy hunger with healthy meals and healthy snacks alone, then

$$S = \varepsilon(M + F^S) - \mu x - BMR \quad (7.7')$$

To the extent that they substitute a junk food meal for a healthy meal or a snack, the value of  $S$  increases by  $(\delta - \varepsilon)$ .

Here, healthy food  $H$  does not enter the obesity function explicitly, but it moderates the calorie contribution of junk-food meals that substitute in satisfying hunger and lead to increased weight. In this framework, taxes on junk food (reducing its consumption), or subsidies to healthy food (increasing its consumption) could have important impacts on the formation of health capital.

## Economic Effects

This model provides several useful inferences about obesity. Clearly, body weight is a measure of health capital, and most people have at least some say in what happens to their level of health capital. Healthy eating (consuming  $H$  rather than  $F$ ) exercising more (increasing exercise level  $x$ ), and appropriate diet practices will help maintain appropriate body weight.

From a health investment point of view, obesity is a bad investment, leading to both higher medical expenditures, and lower earnings. Finkelstein and colleagues (2009) report that across all payers, per capita medical spending for the obese in 2006 was \$1,429, or roughly 42 percent, higher per year, than for someone of normal weight. In aggregate, the annual medical burden of obesity increased from 6.5 percent (in 1998) to 9.1 percent of annual medical spending and could be as high as \$147 billion per year (in 2008 dollars). The authors also note that the 37 percent increase in obesity *prevalence*, and not per capita cost increases, was the main driver of the increase in obesity-attributable costs between 1998 and 2006.

Cawley (2004) addresses the measured impact of obesity on wages. He cites several previous studies that found negative correlations between body weight and wages among females, with three broad explanations for this finding. First, obesity reduces wages by lowering productivity or because employers discriminate against obese people, paying them less than others. A second explanation is that low wages cause obesity. This may occur if poorer people consume cheaper, more fattening, foods. The third explanation is that unobserved variables such as poor education cause both obesity and low wages.

In careful statistical analysis, he uses ordinary least squares (OLS) results to find that heavier white females, black females, Hispanic females, and Hispanic males tend to earn less, and heavier black males tend to earn more, than their less heavy counterparts. With more complex statistical models he shows that the effect is particularly strong for white females. A difference in weight of two standard deviations (roughly 64 pounds) is associated with a wage difference of 9 percent. The magnitude of this difference is equivalent in absolute value to the wage effect of roughly 1.5 years of education or three years of work experience.

### Why Has Obesity Increased?

It is difficult to isolate a single cause of increased obesity, but Cutler, Glaeser, and Shapiro (2003) show that there was increased caloric intake for both men and women from the late 1970s to the late 1990s. In Table 7-3 caloric intake increased for both men (almost 13 percent) and women (9.4 percent) over a twenty-year period.

While there are multiple causes, the authors attribute substantial explanatory power to changes in the time costs of food production. They note that technological innovations—including vacuum packing, improved preservatives, deep freezing, artificial flavors and microwaves—have enabled food manufacturers to cook food centrally and ship it to consumers for rapid consumption.

Table 7-4 shows that in 1965, married women who did not work outside the home spent over almost 138 minutes per day cooking and cleaning up from meals. In 1995, the same tasks took just about half the time. The switch from individual to mass preparation lowered the time price of food consumption and led to increased quantity and variety of foods consumed.

They argue further, that peeling and cutting french fries, for example, is a marginal time cost, while deep frying (in restaurants) is generally a fixed cost (up to the point where the fryer

**TABLE 7-3 Changes in Food Consumption, 1977–1978 to 1994–1996**

	Meal	Calories <sup>a</sup>		Change	Percentage of Total Change
		1977–1978	1994–1996		
Male	TOTAL	2080	2347	268	100%
	Breakfast	384	420	36	13
	Lunch	517	567	50	19
	Dinner	918	859	–59	–22
	Snacks	261	501	241	90
	Calories per meal	573	566	–7	
	Meals per day	3.92	4.53	.61	
Female	TOTAL	1515	1658	143	100%
	Breakfast	286	312	26	18
	Lunch	368	398	31	22
	Dinner	676	602	–74	–52
	Snacks	186	346	160	112
	Calories per meal	422	408	–14	
	Meals per day	3.86	4.44	.58	

Note: Data are from the Continuing Survey of Food Intake 1977–1978 and 1994–1996.

<sup>a</sup>Average calories except for the row reporting average meals per day.

Source: Permission AEA Publications

**TABLE 7-4 Time Costs by Demographic Group (minutes)**

	1965		1995	
	Meal Prep.	Meal Prep. + Cleanup	Meal Prep.	Meal Prep. + Cleanup
Adults				
Single male	13.6	18.1	15.5	17.3
Married male, nonworking spouse	6.5	9.4	13.2	14.4
Married male, working spouse	8.1	11.9	13.2	14.4
Single female	38.1	60.1	28.9	33.1
Married female, working	58.3	84.8	35.7	41.4
Married female, not working	94.2	137.7	57.7	68.8
Elderly				
Male	16.6	26.3	18.5	20.2
Female	65.9	104.4	50.1	60.3

*Source:* Authors' calculations from Americans' Use of Time Survey Archives, 1965 and 1995.  
Permission AEA Publications

is full). Mass preparation allows a restaurant to share the fixed time component over a wide range of consumers. In addition, mass preparation reduces the marginal cost of preparing food by substituting capital for labor. Finally, mass preparation exploits the division of labor. Food professionals instead of “home producers” now prepare food, reducing both fixed and marginal costs.

The authors calculate that reductions in the time required to prepare food reduced the per-calorie cost of food by 29 percent from 1965 to 1995. If the elasticity of caloric intake with respect to price is  $-0.7$ , this could explain the increase in caloric intake and the corresponding increases in obesity. If the calorie intake elasticity is a bit less responsive (say  $-0.5$ ), as is likely, then issues of individual self-control, food advertising, and perhaps lack of information on the true costs of obesity, may also explain the serious increase in obesity over the past 20 to 30 years.

## CONCLUSIONS

This chapter has addressed the demand for health and medical services from an individual perspective. It has treated health as a good produced like all others, using market inputs as well as an individual's time. People benefit from health in four important ways:

1. They feel better when well.
2. They lose less time to illness, and hence can work more.
3. They are more productive when they work and can earn more for each hour they work.
4. They may live longer.

By analyzing the demand for health in this way, we recognize that the demands for health care inputs—from physicians' services, to drugs, to therapy—are demands that derived from the demand for health itself. Consumers, jointly with providers, allocate resources among health care inputs based on the demand for health. We address the specific demand for health inputs in Chapter 9.



## Summary

1. It is not medical care as such that consumers want, but rather health itself. Medical care demand is a derived demand for an input that is used to produce health.
2. Consumers do not merely purchase health passively from the market, but instead produce it, spending time on health-improving efforts in addition to purchasing medical inputs.
3. Health lasts for more than one period. It does not depreciate instantly, and thus it can be treated as a capital good.
4. Demand for health has pure consumption aspects; health is desired because it makes people feel better.
5. Demand for health also has pure investment aspects; health is desired because it increases the number of healthy days available to work and thus earn income.
6. Consumers prefer more income and more leisure so indifference curves between income and leisure are negatively sloped. The slope of the line depicting the labor–leisure tradeoff is the wage rate.
7. Because health is a capital good, the cost of holding health for any one year, as well as over time, will equal the opportunity cost of the capital (interest foregone) plus the depreciation (deterioration of value).
8. The *MEI* curve for investments in health is downward sloping. This occurs because the production function for healthy days exhibits diminishing marginal returns. Thus, the optimal demand for health is likewise given at the intersection of the *MEI* curve and the cost of capital curve ( $r + \delta$ ).
9. The pure investment model generates the prediction that optimal health will decline as the person ages if the depreciation rate of health increases as a person ages.
10. The rewards of being healthy are generally greater for higher-wage workers so those with increasing wages will generally tend to increase their optimal health stock.
11. Health can be generated at less cost for more highly educated people, resulting in a higher optimal health stock than for less educated people.
12. Obesity is a prime example of health capital analysis. Many feel that decreased time and money costs of food preparation have had major impacts on caloric intake, and hence, obesity.

## Discussion Questions

1. Why do we treat leisure and earnings as ordinary utility-increasing goods?
2. Describe the aspects of health that make it a consumption good. Describe those that make it an investment good.
3. Give examples of how health is produced from time and market goods.
4. Why is the depreciation of a capital good a cost to society? In what ways does a person's health depreciate?
5. Why might older people's health care expenditures increase in the Grossman model even though their desired health stock may be lower?
6. List at least three factors that might increase an individual's marginal efficiency of investment in health capital.
7. Suppose that a young woman goes on to medical school and becomes a physician. Would you expect her expenditures on medical goods for her own health to be higher or lower than a nonphysician? Why?
8. From your experience, do you think the typical person becomes less healthy upon, or shortly after, retirement? What does the Grossman model predict?
9. People who earn a higher salary can afford more goods, including health care. However, according to Grossman, they will choose a higher desired health stock. Why is this so, according to the model?
10. Knowing the potential negative effects, would a "rational" person ever choose to become obese.

## Exercises

1. Draw an isoquant (see Chapters 2 and 6) for medical inputs and other inputs in the production of a given amount of health investment. What does the isoquant mean? How would the isoquant look if substitution was limited? If a high degree of substitution was possible?
2. Suppose that no amount of other goods can compensate for a loss in health. How would the individual's indifference curves look? Is this a reasonable assumption in terms of what we actually see taking place?
3. Suppose that John Smith gets promoted to a job that causes two changes to occur simultaneously: John earns a higher wage, and a safer environment causes his health to depreciate less rapidly. How would these two changes together affect John's desired health capital?

4. Suppose that John could work 365 days per year and could earn \$200 per day for each day he worked. Draw his budget line with respect to his labor–leisure choice.
5. Suppose that John chooses to work 200 days per year. Draw the appropriate indifference curve, and note his equilibrium wage income and labor–leisure choices.
6. Suppose, in Exercise 5, that John’s wage rises from \$200 to \$210 per day. Show how his equilibrium level of income and labor–leisure will change.
7. Suppose that John is ill ten days per year. Draw the impact of this illness on the equilibrium defined in Exercise 5. How will it change his equilibrium allocation of earnings and labor vs. leisure?
8. Answer the following.
  - (a) Depict Sara’s optimal stock of health capital at age 18, with a high school diploma and a wage of \$8 per hour.
  - (b) Suppose that she invests in a college education, expecting to get a better and higher-wage job. Show how her optimal stock of health capital changes by the age of 30 due to the increased wage. Then, show how her education would affect her optimal health stock if education also made her a more efficient producer of health.
  - (c) Suppose that after age 30 her wage stays the same. As she ages, show what happens to her optimal stock of

health capital, assuming that the depreciation rate of health increases with age.

9. Consider Fred’s investment in units of health capital with the following function:

$$I = 500 - 1,000 \times \text{cost of capital.}$$

- (a) Indicate some of the components of the cost of capital, and why they are costs.
  - (b) If the cost of capital is 10% each year, what is the equilibrium health investment in terms of units of capital? What is the equilibrium total investment expenditure? Explain both answers.
  - (c) If the cost per unit of health capital doubles to 20%, what will happen to the equilibrium level of health investment and to equilibrium health investment expenditures? Why?
10. Consider the obesity model where equation (7.7) refers to a daily obesity function. Suppose that Ed can either exercise 0 or 1 session per day, with  $\mu = 300$ . If Ed substitutes one junk food meal ( $\delta = 800$ ) for a healthy meal ( $\varepsilon = 600$ ) five days per week, how often per week will he have to exercise to avoid increasing  $S$ ? Why?

# Demand and Supply of Health Insurance

- What Is Insurance?
- Risk and Insurance
- The Demand for Insurance
- The Supply of Insurance
- The Case of Moral Hazard
- Health Insurance and the Efficient Allocation of Resources
- Income Transfer Effects of Insurance
- Conclusions

Health insurance underlies any discussion of the health economy. Most Americans, and indeed most citizens of other countries, do not pay directly for their health care. Rather, insurance companies or other programs indirectly pay for much of the care, with the consumer paying directly only a portion of the bill. The patient's portion is sometimes called coinsurance. Insurance coverage is provided through the payment of premiums (in privately financed systems) or taxes (when insurance is provided publicly). The premiums are often, although not always, purchased through the consumer's participation in the labor force.

Health care expenses are uncertain. Many illnesses occur rarely and seemingly at random. When they do, they may cost a great deal, and they can be financially troublesome if not ruinous to households. Costs could be so high that without financial help treatment might not be available.

Because insurance is so important to the demand and supply of health care, as well as the government's role in allocating health resources, we introduce several fundamental factors relating to the demand and supply of insurance. This provides a set of tools for addressing issues such as the demand for and supply of health care, the role of information in health care markets, and the variation of health care among various markets. We return to insurance issues in Chapter 11, which will look at the operation of insurance markets.

## WHAT IS INSURANCE?

Consider the demand for insurance without all of the detailed trappings (deductibles, premiums, coinsurance, etc.) that accompany modern insurance plans. Start with a club with 100 members. The members are about the same age, and they have about the same interests and lifestyles. About once a year one of the 100 members gets

sick and incurs health care costs of \$5,000. The incidence of illness seems to be random, not necessarily striking men, women, the old, or the young in any systematic fashion. Club members, worried about potential financial losses due to illness, decide to collect \$50 from each member and put the \$5,000 in the bank for safekeeping and to earn a little interest. If a member becomes ill, the fund pays for the treatment. This, in a nutshell, is insurance. The members have paid \$50 to avoid the risk or uncertainty, however small, of having to pay \$5,000. The “insurer” collects the money, tries to maintain and/or increase its value through investment, and pays claims when asked.

This example illustrates several desirable characteristics of an insurance arrangement.

1. The number of insured should be large, and they should be independently exposed to the potential loss.
2. The losses covered should be definite in time, place, and amount.
3. The chance of loss should be measurable.
4. The loss should be accidental from the viewpoint of the person who is insured.

Insurance generally reduces the variability of the incomes of those insured by pooling a large number of people and operating on the principle of the law of large numbers. That is, although outlays for a health event may be highly variable for any given person in the insurance pool, the average outlays for the group can be predicted fairly well. The law of large numbers shows that for a given probability of illness, the distribution of the average rate of illness in the group will collapse around the probability of illness as the group size increases.

This chapter considers the theory and practice of health care insurance. It shows the necessity of quantifying risk, as well as attitudes toward risk. With those ideas, we consider the structure of insurance policies and how markets evolve to provide them.

### Insurance Versus Social Insurance

We wish to distinguish between insurance as provided through the pooling of risk, and government programs (often referred to as social insurance), such as Social Security, Medicare, and Medicaid in the United States. Insurance is provided through markets in which buyers protect themselves against rare events with probabilities that can be estimated statistically. The government programs use the government as insurer and are distinguished by two features:

1. Premiums (the amounts paid by purchasers) are heavily and often completely (as in the case of Medicaid) subsidized.
2. Participation is constrained according to government-set eligibility rules.

In addition, government insurance programs often transfer income from one segment of society to another. Given the importance of such social insurance programs, we devote an entire chapter (Chapter 21) to them later in the text.

### Insurance Terminology

Consider some terms that are used to discuss insurance. Although much of the analysis uses the standard economic language of prices and quantities, the insurance industry has developed a particular set of definitions. These include:

**Premium, Coverage**—When people buy insurance policies, they typically pay a given premium for a given amount of coverage should the event occur.

**Coinsurance and Copayment**—Many insurance policies, particularly in the health insurance industry, require that when events occur, the insured person share the loss through copayments. This percentage paid by the insured person is the coinsurance rate. With a 20 percent coinsurance rate, an insured person, for example, would be liable (out of pocket) for a \$30 copayment

out of a \$150 charge. The insurance company would pay the \$120 remainder, or 80 percent. Thus, coinsurance refers to the percentage paid by the insured; copayment refers to the amount paid by the insured (such as a fixed payment for a prescription).

**Deductible**—With many policies, some amount of the health care cost is paid by the insured person in the form of a deductible, irrespective of coinsurance. In a sense, the insurance does not apply until the consumer pays the deductible. Deductibles may apply toward individual claims. Often in the case of health insurance they are applied only to a certain amount of total charges in any given year.

Coinsurance and deductibles often are used together. For example, an insurance policy may require that Elizabeth pay the first \$250 of her medical expenses out of pocket each year. It may then require that she pay 20 percent of each additional dollar in charges. This policy would be said to have a deductible of \$250 and a coinsurance rate of 20 percent.

Many feel that deductibles and coinsurance simply represent ways that insurance companies have found to separate consumers from their money. Economists, in contrast, have found that deductibles and coinsurance may lead to desirable economic consequences. The requirement of a copayment makes consumers more alert to differences in the true costs of the treatment they are purchasing. The charging of deductibles discourages frivolous claims or visits, and it also makes insured people more aware of the results of their actions. Both deductibles and coinsurance may serve to avoid claims and to reduce costs.

Finally, a few other terms describe various features of insurance:

**Exclusions**—Services or conditions not covered by the insurance policy, such as cosmetic or experimental treatments.

**Limitations**—Maximum coverages provided by insurance policies. For example, a policy may provide a maximum of \$3 million lifetime coverage.

**Pre-existing Conditions**—Medical problems not covered if the problems existed prior to issuance of insurance policy. Examples here might include pregnancy, cancer, or HIV/AIDS.

**Pure Premiums**—The actuarial losses associated with the events being insured.

**Loading Fees**—General costs associated with the insurance company doing business, such as sales, advertising, or profit.

With these in mind, we turn to a more formal analysis of risk and insurance.

## RISK AND INSURANCE

To this point we have assumed that all decisions occur under conditions of certainty; that is, consumers know what the prices, incomes, and tastes are and will be. Clearly, however, many decisions are made under conditions in which the outcome is risky or uncertain.

We begin by considering the insurance coverage of an event that occurs with the known probability,  $p$ , leading to a predictable loss and/or payment. This assumption will characterize people's choices under uncertainty. We will then extend the general characterization to health insurance, which differs because the payment may be affected by the insurance. We address this difference once the basic points regarding risk are developed.

### Expected Value

Suppose Elizabeth considers playing a game in which a coin will be flipped. If it comes up heads, Elizabeth will win \$1; if it comes up tails, she will win nothing. How much would Elizabeth be willing to pay in order to play this game? Analysts rely on the concept of expected value for the answer.

With an honest coin, the probability of heads is one-half (0.5), as is the probability of tails. The expected value, sometimes called the expected return, is:

$$E(\text{heads is called}) = (\text{probability of heads}) \times (\text{return if heads, i.e., \$1}) \\ + (\text{probability of tails}) \times (\text{return if tails, \$0}) \quad (8.1)$$

The expected value is \$0.50; that is  $(\frac{1}{2} \times 1) + (\frac{1}{2} \times 0)$ . If she uses the decision criterion that she will play the game if the expected return exceeds the expected cost, Elizabeth will play (pass) if it costs her less (more) than \$0.50. More generally, with  $n$  outcomes, expected value  $E$  is written as:

$$E = p_1R_1 + p_2R_2 + \dots + p_nR_n$$

where  $p_i$  is the probability of outcome  $i$ , (that is  $p_1$  or  $p_2$ , through  $p_n$ ) and  $R_i$  is the return if outcome  $i$  occurs. The sum of the probabilities  $p_i$  equals 1.

The special case where the price of the gamble is exactly \$0.50 and is equal to the expected return is analogous to a situation in insurance in which the expected benefits paid out by the insurance company are equal to the premiums taken in by the company. This equality of expected benefit payments and premiums is called an *actuarially fair* insurance policy. In reality, insurance companies must also cover additional administration and transaction costs to break even, but the definition of an actuarially fair policy provides a benchmark in talking about insurance.

### Marginal Utility of Wealth and Risk Aversion

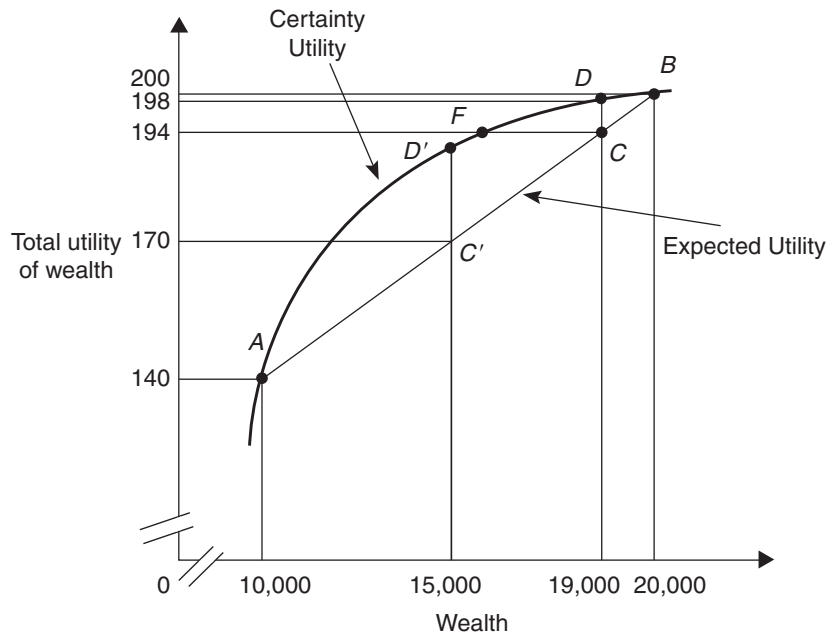
The foregoing example implies that Elizabeth is indifferent to risk. That is, for her the incremental pleasure of winning \$0.50 (the gain of \$1 less the \$0.50 she paid to play) is exactly balanced by the incremental displeasure of losing \$0.50 (the gain of zero less the \$0.50 paid to play). Suppose instead we increase the bets so that the coin flip now yields \$100, or nothing, and that Elizabeth is now asked to bet \$50 to play. Elizabeth may now think a little harder. She may now refuse an actuarially fair bet—\$50—on the grounds that she cannot afford to risk the \$50 loss if the coin lands tails. This suggests that the disutility of losing money may exceed the utility of winning a similar amount. This would occur if she felt that the utility of an extra dollar of wealth is worth more if she has less money than the utility of an extra dollar of wealth is worth when she has more. The utility from an extra dollar is called the marginal utility of wealth.<sup>1</sup>

It is important to incorporate the individual's utility of wealth function into the analysis. In Chapter 2, we assumed that consumers could only rank bundles but could not (and need not) compare magnitudes of satisfaction. To understand the utility model of risk behavior, however, we must further assume that consumers can rank alternatives and compare their magnitudes.

In Figure 8-1 suppose that Elizabeth's wealth is \$10,000. That wealth gives her a utility level of  $U_1 = 140$  and allows her to buy some basic necessities of life. This can be denoted as point *A*. Suppose her wealth rises to \$20,000. Will her utility double?

It is hard to know for certain, but it is plausible that the next \$10,000 will not bring her the incremental utility that the first \$10,000 brought. Thus,  $U_2$  will likely be less than twice  $U_1$ . Suppose, for example, that  $U_2 = 200$ . This is denoted as point *B*. Do all of the points on the utility function between  $U_1$  and  $U_2$  lie on a straight line? If they do, this is equivalent to saying that the utility from the 10,001st dollar is equal to the utility from the 19,999th dollar, and hence the marginal utility is constant. This also is unlikely. Because the marginal utility of earlier dollars is likely to be larger than that of later dollars, the utility curve is likely to be bowed out, or concave, to the  $x$ -axis.

<sup>1</sup> Wealth refers to the sum of the consumer's assets in money terms. It is related to income, which is the flow of funds in any given period. We may refer to one or the other for some discussions, but the substance of the insurance analysis refers to both.



**FIGURE 8-1** Total Utility of Wealth and the Impact of Insurance

The marginal utility of wealth refers to the amount by which utility increases when wealth goes up by \$1. This rise in utility, divided by the \$1 increase in wealth, is thus the slope of the utility function. The bowed shape of the utility function shows a slope that is getting smaller or flatter as wealth rises; the marginal utility of wealth is diminishing.

With wealth of \$20,000, Elizabeth understands that if she becomes ill, which may occur with probability 0.10, her expenses will cause her wealth to decline to \$10,000. If this occurs, she can calculate her expected wealth,  $E(W)$ ,

$$E(W) = (\text{prob. well} \times \text{wealth if well}) + (\text{prob. ill} \times \text{wealth if ill}) \quad (8.2a)$$

$$(0.90 \times \$20,000) + (0.10 \times \$10,000) = \$19,000$$

and expected utility,  $E(U)$ :

$$E(U) = (\text{prob. well} \times \text{utility if well}) + (\text{prob. ill} \times \text{utility if ill}) \quad (8.2b)$$

$$(0.90 \times \text{utility of } \$20,000) + (0.10 \times \text{utility of } \$10,000)$$

$$(0.90 \times 200) + (0.10 \times 140) = 194$$

Thus, the expected utility  $E(U)$  is 194 or point  $C$  because of the risk of illness. Geometrically, this is the line segment between points  $A$  and  $B$ , evaluated at wealth level  $E(W) = \$19,000$ . The expected utility due to risk must be compared to the utility of 198 (point  $D$ ), corresponding to the utility that she would receive if she could purchase insurance at an actuarially fair rate. As drawn, the risk leads to a loss of 4 units (198 to 194) of utility.

### Purchasing Insurance

Suppose that Elizabeth can buy an insurance policy costing \$1,000 per year that will maintain her wealth irrespective of her health. That is, if she stays well, her wealth will be \$20,000 less the \$1,000 premium. If she falls ill, she is provided \$10,000 in benefits, so that her wealth will be \$10,000 plus the \$10,000 in benefits, less the \$1,000 premium.

Is it a good buy? We see that at a net wealth of \$19,000, which equals her initial wealth minus the insurance premium, her certainty utility is 198. Elizabeth is better off at point *D* than at point *C*, as shown by the fact that point *D* gives the higher utility. If insuring to get a *certain* wealth rather than facing the risky prospect makes Elizabeth better off, she will insure.

We can, in fact, use Figure 8-1 to calculate the maximum amount that Elizabeth would be willing to pay for the insurance by moving southwest down the utility function to the level of  $U = 194$  and reading the level of wealth (off the  $x$ -axis) to which it corresponds, or point *F*. The distance *FC* reflects Elizabeth's aversion to risk. At point *F*, Elizabeth would be willing to pay up to \$4,000 (that is, initial wealth of \$20,000, less \$16,000 at point *F*) for insurance and still be as well off as if she had remained uninsured. If, for example, she were able to purchase the insurance for \$3,000, she would get \$1,000 in consumer surplus.

This analysis illustrates several facts:

1. Insurance can be sold only in circumstances where there is diminishing marginal utility of wealth or income—that is, when the consumer is risk-averse. By examining Figure 8-1, we can see that if marginal utility were constant, a requirement that Elizabeth pay an actuarially fair premium for insurance would leave her no better off than if she were uninsured. If Elizabeth was “risk-loving,” with increasing marginal utility of wealth or income, she would definitely refuse to buy insurance.
2. Expected utility is an average measure; Elizabeth either wins or loses the bet. If she is exposed to risk, Elizabeth will have wealth and hence utility of either \$20,000 (with utility of 200), or \$10,000 (with utility of 140), and a risky expected wealth of \$19,000. Insurance will guarantee her wealth to be \$19,000. If she does not fall ill, her wealth will be \$20,000 less the \$1,000 insurance premium; if she falls ill, her wealth will be \$10,000 plus the \$10,000 payment for the loss of health, minus the \$1,000 premium—again \$19,000.
3. If insurance companies charge more than the actuarially fair premium, people will have less expected wealth from insuring than from not insuring. Even though people will have less wealth as a result of their purchases of insurance, the increased well-being comes from the elimination of risk.
4. The willingness to buy insurance is related to the distance between the utility curve and the expected utility line. If Elizabeth is very unlikely to become ill, (near point *B*), then her expected utility will be almost identical to her certainty utility, and her gains from insurance will be small. If Elizabeth's probability of illness increases to 0.5 (point *C'*), her expected wealth will be \$15,000 and her expected utility will be 170. She will accrue significant gains by insuring as noted by the distance *C'D'*. However, if Elizabeth is almost certain to fall ill, (approaching point *A*), her gains from buying insurance decrease. Why? In this case, Elizabeth is better off “self-insuring,” by putting the (almost) \$10,000 away to pay for her almost certain illness rather than incurring the transactions costs of buying insurance and then filing claims.

## THE DEMAND FOR INSURANCE

If risk-averse people choose to purchase insurance, how much will they purchase? The next two sections present a classic model introduced by Mark Pauly in 1968 to consider the fundamental demand and supply decisions regarding insurance.

### How Much Insurance?

We have discussed why Elizabeth would choose to buy insurance so we now consider how much insurance she would choose. Recall that Elizabeth's expected utility involves her wealth when ill, with a probability of 0.10, or when healthy, with a probability of 0.90. If ill, her wealth will fall from \$20,000 to \$10,000.

We address Elizabeth's optimal purchase by using the concepts of marginal benefits and marginal costs. Consider first a policy that provides insurance covering losses up to \$500. Although it



might hardly seem worth buying a \$500 insurance policy when Elizabeth will lose \$10,000 if she falls ill, it is a useful place to start.

The goal of maximizing total net benefits provides the framework for understanding her health insurance choice. She benefits from health insurance only when she is ill and receives the insurance benefit payments. She still pays the insurance premiums when ill, but she gains financially net of those premiums. When well, only the premium applies to her, and is thus a net cost.

In turn, the marginal benefits, when she is well, will decline as additional insurance coverage is purchased; this is understood by applying the law of diminishing marginal utility of wealth. In contrast, again due to the diminishing marginal utility of wealth, the marginal costs when well will rise as additional insurance is purchased. Elizabeth's solution is to purchase so that the marginal benefits equal the marginal costs.

Suppose that she must pay a 20 percent premium (\$100) for her insurance, or \$2 for every \$10 of coverage that she purchases. The following worksheet describes her wealth if she gets sick.

Insurance Worksheet—\$500 coverage—Wealth If Ill		
	<b>Original</b> wealth	\$20,000
<i>less</i>	Loss	\$10,000
	Remainder	\$10,000
<i>plus</i>	Insurance	500
	Sum	\$10,500
<i>less</i>	Premium	100
<i>or</i>	<b>New</b> wealth	\$10,400

For the initial coverage, Elizabeth's wealth, if well, is \$20,000 less the \$100 premium, or \$19,900. Her marginal benefit from the \$500 from insurance is the expected marginal utility that the additional \$400 (\$500 minus the \$100 premium) brings. Her marginal cost is the expected marginal utility that the \$100 premium costs. We describe these benefits and costs in Figure 8-2. If

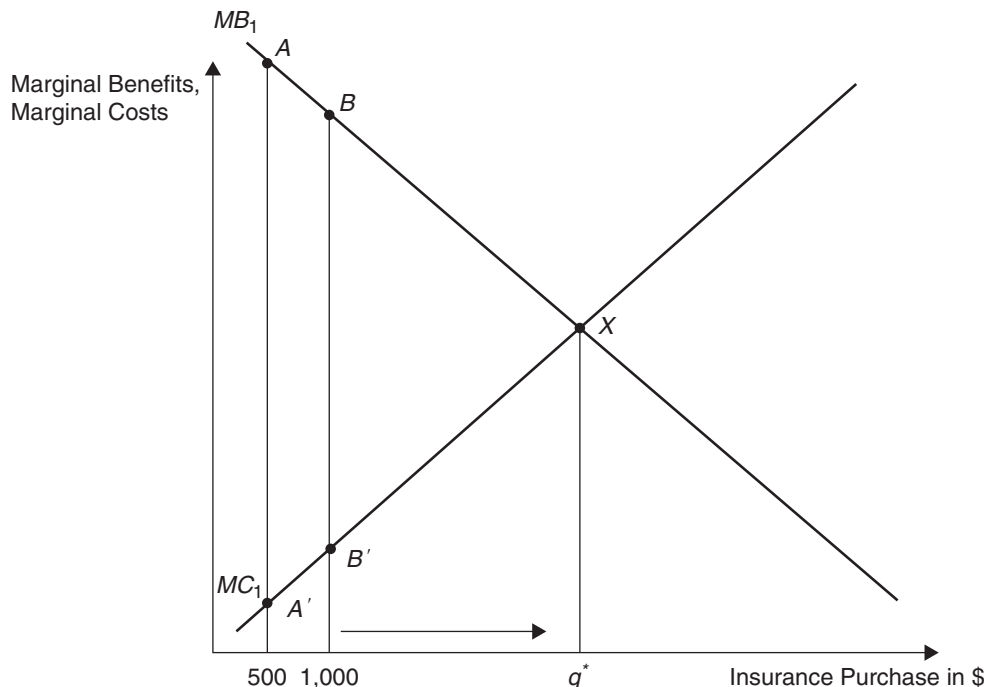


FIGURE 8-2 The Optimal Amount of Insurance

Elizabeth is averse to risk, the marginal benefit (point *A*) of this insurance policy exceeds its marginal cost (point *A'*).

Should Elizabeth increase her coverage from \$500 to \$1,000? She must compare the marginal benefits of this next \$500 increment to its marginal costs. Because Elizabeth is slightly wealthier than before, if ill (starting at \$10,400 rather than \$10,000) the marginal utility from an additional \$400 of wealth (calculated as before) will be slightly smaller than before. Hence, the marginal benefits from the second \$500 increment will be slightly smaller than for the first \$500 increment. Her marginal benefit curve,  $MB_1$ , is downward sloping, with her new marginal benefit at point *B*.

Similarly, because she is a little less wealthy than before if well, an additional \$100 in premiums will cost a little more in foregone (marginal) utility of wealth than the first increment at point *B'*. Thus, her marginal cost curve,  $MC_1$ , is upward sloping.

By reiterating this process, we can see that Elizabeth will adjust the amount of insurance,  $q$ , that she purchases to the point at which the marginal benefits equal the marginal cost. The quantity,  $q^*$ , at which they are equal (point *X*) is Elizabeth's optimum insurance purchase. The  $x$ -axis of Figure 8-2 is drawn to scale, and shows that  $q^*$  is approximately \$3,000.

### Changes in Premiums

How will her insurance decision change if premiums change? Consider first the impact of a higher premium, say 25 percent rather than the 20 percent used earlier. With the 25 percent premium (\$125), Elizabeth faces the following calculation regarding a policy that provides \$500 worth of coverage:

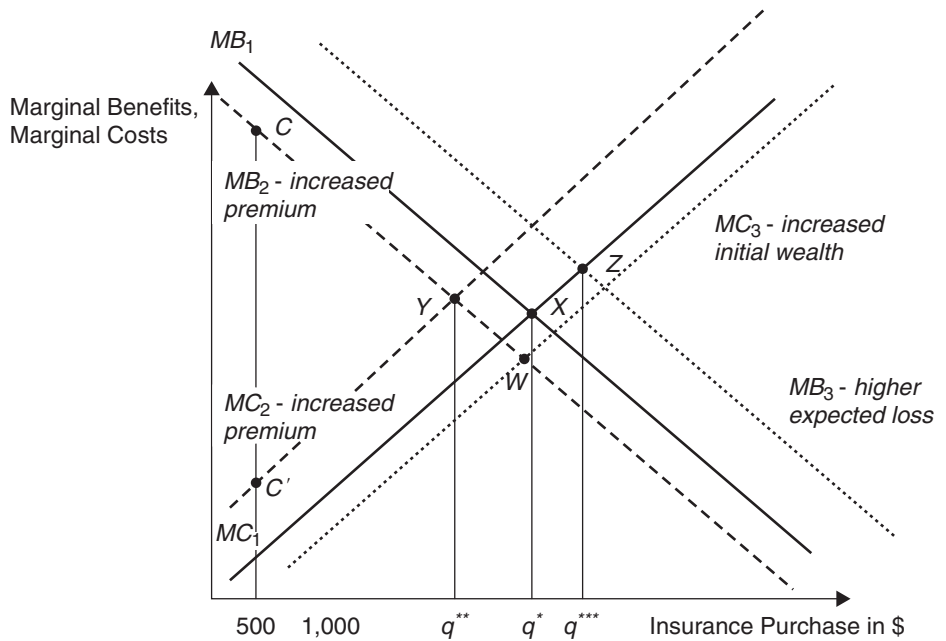
	<b>Original</b> wealth	\$20,000	
<i>less</i>	Loss	\$10,000	
	Remainder	\$10,000	
<i>plus</i>	Insurance	500	
	Sum	\$10,500	
<i>less</i>	<b>New premium</b>	125	←
<i>or</i>	<b>New</b> wealth	\$10,375	

If she stays well, her wealth is \$20,000 less the \$125 premium, or \$19,875. Look now at Figure 8-3. Elizabeth's marginal benefit from the \$500 from insurance is now \$375 rather than the previous value of \$400, so point *C* lies on curve  $MB_2$  below the previous marginal benefit curve,  $MB_1$ . We can fill in additional points on this curve, which reflects the higher premium.

Similarly, Elizabeth's marginal cost is the expected marginal utility that the (new) \$125 premium costs her. This exceeds the previous cost in terms of foregone utility, so point *C* lies on curve  $MC_2$  above the previous marginal cost curve,  $MC_1$ . Again, we can fill in additional points on this curve and find the intersection of  $MB_2$  and  $MC_2$  at point *Y*. The resulting analysis suggests that consumers react rationally to higher premiums by reducing their optimum coverage from  $q^*$  to  $q^{**}$ . In this example, the fall is from about \$3,000 to \$2,300.

### Changes in Expected Loss

How will the insurance decision be affected by changes in expected losses? Returning to the original example with a premium of 20 percent, suppose that instead of \$10,000, Elizabeth expected to lose \$15,000 if ill. Consider again the first \$500 of insurance coverage. Her wealth, if healthy, is



**FIGURE 8-3** Changes in the Optimal Amount of Insurance

\$19,900, so nothing changes with respect to marginal cost. Elizabeth remains on curve  $MC_1$ . The marginal benefit calculation, however, does change:

**Insurance Worksheet—Higher Expected Loss  
Wealth If Ill**

	<b>Original</b> wealth	\$20,000	
<i>less</i>	<b>New loss</b>	\$15,000	←
	Remainder	\$ 5,000	
<i>plus</i>	Insurance	500	
	Sum	\$ 5,500	
<i>less</i>	Premium	100	
<i>or</i>	<b>New</b> wealth	\$ 5,400	

As before, the insurance gives her a net benefit of \$400. However, this net benefit increments a wealth of \$5,000 rather than \$10,000. If we assume that an additional dollar gives more marginal benefit from a base of \$5,000 than from a base of \$10,000, then the marginal benefit curve shifts upward because of the increased expected loss. This provides equilibrium point Z on curve  $MB_3$  in Figure 8-3. It follows that in equilibrium, an increase in the expected loss will increase the amount of insurance purchased at point Z, or  $q^{***}$ . As drawn,  $q^{***}$  equals approximately \$3,500.

**Changes in Wealth**

Finally, consider the impact of changes in an individual’s initial wealth. Suppose Elizabeth was starting with a wealth of \$25,000 instead of \$20,000. Again, assume a premium rate of 20 percent.

**Insurance Worksheet—Increased Wealth  
Wealth If Ill**

	<b>Increased</b> wealth	\$25,000
<i>less</i>	Loss	\$10,000
	Remainder	\$15,000
<i>plus</i>	Insurance	500
	Sum	\$15,500
<i>less</i>	Premium	100
<i>or</i>	<b>New</b> wealth	\$15,400

At the higher level of wealth, the same insurance policy provides a smaller increment in utility, so the marginal benefit curve shifts down from  $MB_1$  to  $MB_2$ . However (for the same expected loss), the \$100 premium costs less in foregone marginal utility relative to the increased wealth, a downward shift of  $MC_1$  to  $MC_3$ . As a result of both downward shifts, the new equilibrium value of  $q$  at point  $W$  may be higher or lower than the original value of  $q^*$  (as drawn in Figure 8-3, it is slightly lower, about \$2,800). If, however, increased wealth is accompanied by increased losses, then the  $MB$  curve may shift down less. If it does shift down by less, the desired amount of insurance may increase because the increased expected losses would make a larger amount of insurance more desirable.

## THE SUPPLY OF INSURANCE

In the previous example, we assumed a 20 percent premium rate, but to determine the amount of coverage someone will buy, we must know how insurers determine the premium. We started with the club that insures its members against illness. The officers of the club do not know, nor necessarily care, who will file a claim.<sup>2</sup> All that is necessary for the club to function as an insurer is that revenues cover costs. In practice, insurance companies will also incur administrative and other expenses that also must be covered by premiums.

### Competition and Normal Profits

Let's return to Elizabeth's insurance problem from the previous section. Elizabeth is buying insurance in a competitive market, and under perfect competition, all firms earn zero excess profits. Recall that Elizabeth faced a potential illness with a probability of 0.10 (1 in 10). She sought to buy insurance in blocks of \$500, and at the outset, her insurer, Asteroid Insurance, was charging her \$100 for each block of coverage, or an insurance premium of 0.20 (\$100 as a fraction of \$500). Assume also that it costs Asteroid \$8 annually to process each insurance policy and (if necessary) write a check to cover a claim. Asteroid's profits per policy are:

$$\text{Profits} = \text{Total revenue} - \text{Total costs}$$

Revenues are \$100 per policy. What are Asteroid's costs? For 90 percent of the policies, the costs are \$8 because the insured does not get sick and does not collect insurance. The only costs are the \$8 processing costs per policy. The costs for the other 10 percent of the policies are \$508, consisting of the \$500 payment to those who are ill plus the processing costs of \$8.

<sup>2</sup> Insurers must recognize whether they are getting nonrepresentative slices of the risk distribution. Analysts refer to this as *adverse selection*, and it can lead to financial losses for the insurer.

Thus, the profits per policy for Asteroid are:

$$\begin{aligned}
 \text{Profits} &= \$100 - (\text{Probability of illness} \times \text{Costs if ill}) \\
 &\quad - (\text{Probability of no illness} \times \text{Costs if no illness}) \\
 \text{Profits} &= \$100 - (0.10 \times \$508) - (0.90 \times \$8) & \text{(8.3)} \\
 \text{Profits} &= \$100 - \$50.80 - \$7.20 \\
 \text{Profits (premium} &= 10\%) = \$100 \text{ (revenues)} - \$58 \text{ (costs)} = \$42
 \end{aligned}$$

These are positive profits, and they imply that a competing firm (also incurring costs of \$8 to process each policy) might enter the market and charge a lower premium, say, 15 percent, to attract clients. The cost side of the equation would remain the same, but the revenues, which equal the premium fraction multiplied by the amount of insurance, would fall to \$75. Hence, profits fall to:

$$\text{Profits (premium} = 15\%) = \$75 \text{ (revenues)} - \$58 \text{ (costs)} = \$17$$

which is still positive. We can see that entry will continue into this industry until the premium has fallen to a little less than \$0.12 per dollar of insurance (actually \$0.116), or about 12 percent, which would provide revenues of \$58, offset by the \$58 in costs, to give zero profits.

A little algebra can verify that the 12 percent premiums must be tied directly to the probability of the claim (0.10). Quite simply, for Asteroid, the revenue per policy is  $aq$ , where  $a$  is the premium, in fractional terms. The cost per policy in terms of payout is the probability of payout,  $p$ , multiplied by the amount of payout,  $q$ , plus a processing cost,  $t$ , which is unrelated to the size of the policy (assuming it costs no more to administer a \$10,000 policy than a \$500 policy). So:

$$\text{Profits} = \text{Revenue}(aq) - \text{Cost}(pq + t) = aq - (pq + t) = aq - pq - t \quad \text{(8.4)}$$

With perfect competition, profits must equal 0, so:

$$0 = aq - pq - t$$

We can solve for the competitive premium  $a$  as:

$$a = p + (t/q) \quad \text{(8.5)}$$

This expression shows that the competitive value of  $a$  equals the probability of illness,  $p$ , plus the processing (or loading) costs as a percentage of policy value,  $q$ , or  $t/q$ . If loading costs are 10 percent of the policy value,  $q$ , then  $(t/q) = 0.10$ . Hence, in equilibrium, if  $p$  equals 0.10, then  $a = p + (t/q) = (0.10 + 0.10) = 0.20$ . The premium for each dollar of insurance,  $q$ , is \$0.20. If insurers charge less, they will not have enough money to pay claims. If they charge more, firms like Asteroid will have excess profits, and other firms will bid down rates in perfectly competitive markets.

Previously, in the discussion on the bearing of risk, we considered insurance policies that would compensate the individual against the loss based solely on the probability of the event's occurring. Such rates are referred to as actuarially fair rates. The actuarially fair rates correspond to the rates in which the loading costs  $t$  as a percentage of insurance coverage,  $q$  (that is,  $t/q$ ), approach 0, hence:

$$a = p + (t/q) = p + 0 = p \quad \text{(8.6)}$$

Knowing that premium  $a$  equals  $p$  under perfect competition (with no loading costs), we now solve for the optimal coverage against any expected loss. To maximize utility, Elizabeth will add

coverage up to the point where her expected wealth will be the same whether she is ill or well. In the earliest example, the particular illness occurred with a probability of 0.10 and incurred a loss of \$10,000. In a competitive insurance market (ignoring loading costs), Elizabeth's wealth, if well, will be:

$$\begin{aligned}\text{Wealth (if well)} &= \$20,000 - \text{cost of insurance, or:} \\ &\quad \$20,000 - (\text{premium } a) \times (\text{coverage } q)\end{aligned}$$

Her wealth, if ill, will be:

$$\begin{aligned}\text{Wealth (if ill)} &= \$20,000 - \text{loss} + \text{insurance reimbursement} - \text{insurance premium or:} \\ &\quad \$20,000 - \$10,000 \text{ loss} + \text{coverage } q - (\text{premium } a) \times (\text{coverage } q)\end{aligned}$$

To maximize the expected utility we equate the wealth if well with the wealth if ill:

$$[\text{Wealth (if well)} = \$20,000 - aq] = [\$20,000 - \$10,000 + q - aq = \text{Wealth (if ill)}] \quad (8.7)$$

Subtracting  $\$20,000 - aq$  from both sides and rearranging terms yields:

$$q^{opt} = \$10,000$$

Elizabeth's optimal level of coverage for a loss of \$10,000 in the absence of transactions costs is \$10,000! It is more realistic of course to assume that transactions costs  $t$  will be positive; in fact, loading charges are often substantial. Elizabeth's best choice is to insure for less than the full health expense, we show in the examples above, where the transactions costs were positive and optimal insurance  $q^*$  was considerably smaller than \$10,000. This standard result from the mathematics of health insurance is in contrast with the propensity of consumers to seek full coverage.

## THE CASE OF MORAL HAZARD

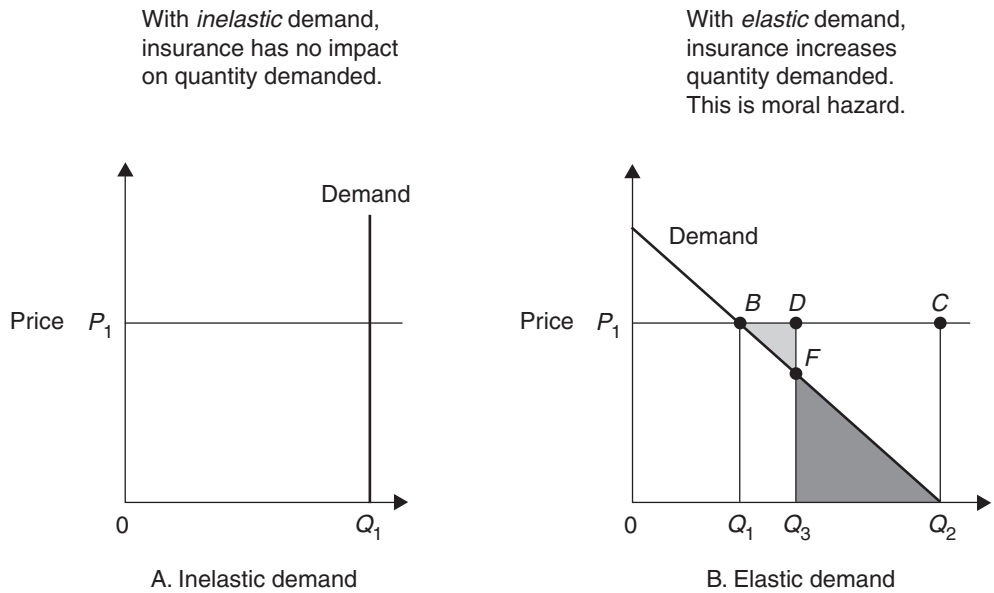
To this point, we have discussed the theory of risk, as well as the demand and supply of insurance when the events and the losses are random. The insurance policies discussed thus far represent indemnity policies, in which the insurer's liability is determined by a fixed, predetermined amount for a covered event. Indeed, the term *indemnity* comes from the Latin *indemnis*, meaning "unhurt." Insurance renders the insured party financially unhurt by the random event.

In the previous section, we showed that the optimal insurance policy covers the entire loss when there is no transaction cost, and less than the full loss in the more realistic case when transactions or loading costs are positive. We now address the effects of the price system on the provision of insurance.

Our discussions have assumed a fixed loss—that did not change merely because people bought insurance. However, in many cases, buying insurance lowers the price per unit of service to consumers at the time that they are purchasing services. If people purchase more service due to insurance, then we must modify many of the insurance propositions just presented.

### Demand for Care and Moral Hazard

Suppose Elizabeth faces the probability  $1 - p = 0.5$  that she will not be sick during a given time period and so will demand no medical care. She also faces probability,  $p$ , also equal to 0.5, that she will contract an illness that requires medical care. Elizabeth fears that she will contract Type 1 diabetes—if so, without insulin she will die. In Figure 8-4, panel A, we assume that her demand for insulin is perfectly inelastic, that is, unresponsive to its price. We saw earlier (ignoring the



**FIGURE 8-4** Demand for Care and Moral Hazard

transactions costs) that Elizabeth would be willing to pay insurance to cover expenditures  $P_1Q_1$ , her expenditures should she need care. An actuarially fair insurance policy would then charge Elizabeth  $\frac{1}{2}P_1Q_1$ , and she would purchase the policy because it insured her against the risk of diabetes.

Consider, instead, Elizabeth's demand for dermatological care (skin care for conditions such as acne or psoriasis). Elizabeth's demand curve for these elective treatments may very well respond to price; that is, the lower the price, the higher the quantity demanded. This is noted in Figure 8-4, panel B. If she purchases insurance that pays her entire loss, then this insurance makes treatment (ignoring time costs) free. Because the marginal price to Elizabeth is zero, she would demand  $Q_2$  units of care for a total cost of care of  $P_1Q_2$ , shown as rectangle  $OP_1CQ_2$ , which is obviously larger than rectangle  $OP_1BQ_1$ . Why only  $Q_2$  units when the care is "free"? Even free care entails time costs of visiting the provider or filling the prescriptions that keep the full price from equaling zero.

This leads to one of two possibilities that was not a problem either in the abstract or for a condition like diabetes:

1. If the insurance company charges the premium  $\frac{1}{2}P_1Q_1$  (where  $\frac{1}{2}$  refers to the probability of illness) for the insurance, the company will lose money. This occurs because the expected payments would be  $\frac{1}{2}P_1Q_2$ . Amount  $P_1Q_2$  exceeds  $P_1Q_1$  because the induced demand leads Elizabeth to consume more care ( $Q_2$ ) with insurance than she would have consumed ( $Q_1$ ) without insurance.
2. If the insurance company charges the appropriate premium,  $\frac{1}{2}P_1Q_2$ , for the insurance, Elizabeth may not buy insurance. This amount may exceed the medical expenses that she would have spent on average had she chosen to put away money on her own, or to "self-insure." While Elizabeth may be willing to pay more than  $\frac{1}{2}P_1Q_1$  to avoid the risk, she may not be willing to pay as much as  $\frac{1}{2}P_1Q_2$ .

The rational response to economic incentives brought about by the price elasticity of demand is termed *moral hazard*. It describes any change in consumer behavior occurring in response to a

contractual arrangement (here, the decision to insure). Here, there is an increased usage of services because the pooling of risks leads to decreased marginal costs to the consumer for the services. Failure to protect oneself from disease, because one has health insurance, would be another form of moral hazard (see Box 8-1 for an example). Our analysis gives a simple measure of the economic costs of moral hazard. Netting out the costs of servicing the insurance (which do not reflect increased use of services), moral hazard is the excess of premiums over Elizabeth's expected outlays had she not purchased insurance.

Elizabeth's insurance premium thus has two parts. The first is the premium for protection against risk, assuming that no moral hazard exists. The second is the extra resource cost due to moral hazard. As before, Elizabeth chooses insurance coverage  $q^*$  by weighing marginal costs against marginal returns, whereas before the returns were the utility gains when Elizabeth was ill. The twist here is that the costs now have two dimensions—the pure premium and the moral hazard. For some categories of care, the second may be important.

This analysis helps predict the types of insurance likely to be provided. It is clear that the optimal level of insurance will likely increase relative to the expected loss as the degree of moral hazard decreases. Suppose we use demand price elasticity as an indicator of the potential for moral hazard. Theory then suggests the following:

1. Deeper (more complete) coverage for services with more inelastic demand.
2. Development of insurance first for those services with the most inelastic demand, and only later for those with more elastic demand.

Data on current insurance coverage by area of service support the first hypothesis, and historical data support the second.

### BOX 8-1

#### Another Type of Moral Hazard—Health Insurance and Insecticide-Treated Bed Nets in Ghana

Although health insurance scholars worry about insurance price effects, health insurance may also influence disease prevention efforts. Zelalem Debebe and Luuk van Kempen (2011) examine the impact of the Ghanaian National Health Insurance scheme (NHIS) on households' efforts in preventing malaria. The National Health Insurance Act 650 was passed in August 2003 to improve access and quality of basic health care services through a National Health Insurance implemented at the district level. By the end of 2008, every district had enrolled and 61% of the total population was covered. The financing of the NHI includes premiums paid by the insured and the NHI fund that comes from taxes on goods, social security contributions, parliament budget allocation and returns from investment.

Sleeping under an insecticide-treated bed net (ITN) is a prominent malaria prevention strategy in sub-Saharan Africa and parts of Asia. Malaria obviously brings forward a utility loss, and possibly death, but people view sleeping under nets as inconvenient. One Ghanaian village chief was quoted as saying, "We have ITNs but we don't use them because the room is so hot and even hotter when you sleep under ITNs." Another insured person asked, "Why would you spend 8 Ghanaian Cedis [currency] on the bed net while you can take 2 Cedis to go to the hospital?"

In mixed statistical analyses the authors found that health insurance negatively impacted bed net ownership, number of members who slept under an ITN, and the number who slept under an ITN they got re-soaked (with insecticides) after they bought it. While the authors do not yet have firm evidence on whether the incidence of malaria had increased, they have shown that the insurance *for hospital care* reduced levels of user self-protection, unintended consequences from a contractual arrangement.



### Effects of Coinsurance and Deductibles

This analysis also provides insight into the impacts of deductible provisions and coinsurance in insurance policies. Returning to Figure 8-4, panel B, suppose that  $Q_1$  reflects \$500 of expenses (rectangle  $OP_1BQ_1$ ) and that  $Q_2$  is three times  $Q_1$  (rectangle  $OP_1CQ_2$ ), which reflects \$1,500 of expenses. If the insurance contains a deductible, Elizabeth will compare the position she would attain if she covered the deductible and received level  $Q_2$  free, with the position she would attain if she paid the market price for all the medical care she consumed.

Assume again that the probability of illness  $p$  equals 0.5. Consider first a policy containing a deductible, which requires Elizabeth to pay the risk premium plus the first \$500 of her medical care (expenses indicated by rectangle  $OP_1BQ_1$ ), after which all additional care is free. Elizabeth will buy this policy because it protects her from risk and allows her to purchase  $Q_2$  units of medical care for \$500. Her gain is the triangle under the demand curve,  $Q_1BQ_2$ .

Suppose now that the insurance company raises the deductible from \$500 to \$700. Will Elizabeth continue to buy the insurance? Recall that without insurance, Elizabeth would have purchased amount  $Q_1$  of health services; the \$700 deductible yields amount  $Q_3$ . When ill, Elizabeth is paying more for the amount ( $Q_3 - Q_1$ ) of incremental health care than she believes the value of incremental care to be. The incremental costs are rectangle  $Q_1BDQ_3$ ; the incremental benefits are the area under her demand curve (trapezoid  $Q_1BFQ_3$ ). The difference is triangle  $BDF$ , and this represents a welfare loss to Elizabeth. However, after paying the deductible, she can get as much additional health care as she wants at zero cost, and she will buy quantity  $Q_2$ . This yields welfare gain triangle  $Q_3FQ_2$  (incremental benefits less zero incremental costs). If  $Q_3FQ_2$  (her welfare gain) is larger than  $BDF$  (her welfare loss), she buys the insurance, even with the \$700 deductible. If  $BDF$  is larger than  $Q_3FQ_2$ , the loss exceeds the gain, and Elizabeth is better off self-insuring and spending  $P_1Q_1$  (in this example, \$500) with probability 0.5.

Hence, the deductible has two possible impacts. A relatively small deductible will have no effect on individual usage, here  $Q_2$ . A large deductible makes it more likely that individuals will self-insure and consume the amount of care they would have purchased with no insurance, here  $Q_1$ .

A wide range of coinsurance coverages have developed. Many analysts have considered how to formulate them to lead to more economically efficient outcomes. We turn to that analysis next.

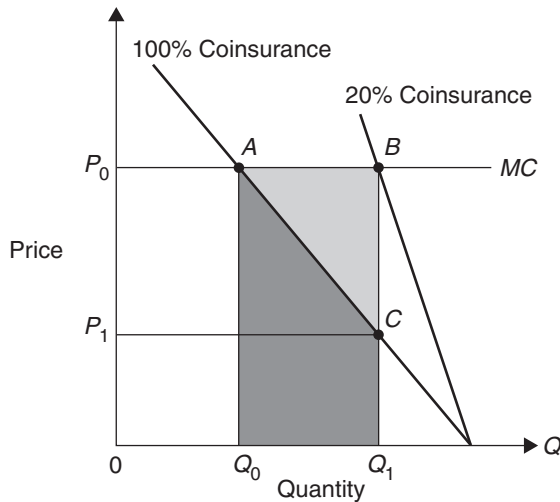
## HEALTH INSURANCE AND THE EFFICIENT ALLOCATION OF RESOURCES

This section examines the impact of health insurance on health care demand. Economists commonly examine the efficient allocation of resources, which occurs when the incremental cost of bringing the resources to market (marginal cost) equals the valuation in the market to those who buy the resources (marginal benefit). If the marginal benefit is greater (less) than the marginal cost, one could improve society's welfare by allocating more (fewer) resources to the sector or individual, and less (more) resources to other sectors.

Consider Figure 8-5, which shows the marginal cost of care at  $P_0$  and the demand for care by a consumer under alternative conditions of insurance. If this consumer is not insured, then the optimal choice of health care is  $Q_0$  units. The price (including travel time, parking, and the cost of bringing the service to market) reflects the cost to society of bringing the entire package to the market. Based on the consumer's (and the physician's) preferences, the marginal benefit, as described through the demand curve, equals the marginal cost. In economic terminology, this is an efficient allocation.

### The Impact of Coinsurance

What happens when Elizabeth pays only a small fraction of the bill, say, at a 20 percent coinsurance rate? If, for example,  $P_0$  was \$50 for an office visit, Elizabeth must now only pay  $P_1$ , or \$10, so her quantity demanded will increase. This is as if a new demand curve (labeled with 20 percent



**FIGURE 8-5** Health Care Demand with Insurance

coinsurance) were generated by rotating the original demand curve outward, and leading to a new equilibrium quantity demanded  $Q_1$ . The cost of bringing services to market has remained the same,  $P_0$ . Services valued at  $P_0Q_0$  are now being provided. The incremental amount spent (incremental cost) is  $P_0 \times (Q_1 - Q_0)$ , or the rectangle  $ABQ_1Q_0$ .

The incremental benefit (to Elizabeth) can be measured by the area under her original demand curve,  $ACQ_1Q_0$ . The remaining triangle  $ABC$  represents the loss in well-being that occurs because Elizabeth is purchasing more health care than is optimal. It is a loss in well-being because the incremental resource cost  $ABQ_1Q_0$  exceeds the incremental benefits  $ACQ_1Q_0$  by triangle  $ABC$ .

What exactly does this mean? It means that the insurance leads Elizabeth to act as if she was not aware of the true resource costs of the care she consumes. It also means that the insurance implicitly subsidizes insured types of care (organized health care settings, prescription drugs) relative to other types of health care (e.g., good nutrition, exercise, over-the-counter drugs, uninsured types of care) that may be just as, or even more, effective. It also subsidizes insured types of care relative to nonhealth goods. The degree of this distortion depends on the exact specification of the policy (that is, deductibles, maximum payments, rates of coinsurance), but it is apparent that insurance can distort the allocation of resources among health care and other goods.

Until recently, many insurance policies had flat rate copayments of as low as one or two dollars for all drugs, leading to circumstances under which it could cost more to drive to the pharmacy, than to pay for the drugs themselves. Then some insurers instituted two-tiered policies such as 5–10 policies, charging \$5 for generic drugs and \$10 for brand-name drugs. Box 8-2 examines recent changes in coinsurance rates for prescription drugs with four or five tiers. Tier 4 drugs, in this account, often come with coinsurance rates of 25 percent or higher.

The impact of moral hazard is intensified through interactions between primary and secondary insurance coverages. This type of interaction sometimes concerns “Medigap” plans, which provide additional coverage to the elderly above the amount paid by Medicare. Another example involves insured employees who have secondary coverage through their spouse’s insurance which may magnify moral hazard problems.

Elizabeth’s employer, General National, provides health insurance to all its workers, with policies that pay 60 percent of all medical expenditures. Many of General’s workers also receive coverage under their spouse’s insurance plans, but General’s plan is considered the primary insurer for these dually covered workers. The *secondary* policies cover 60 percent of the expenses left uncovered by General’s plan.

**BOX 8-2****Copayments Soar for Drugs with High Prices**

For many years, U.S. health insurance plans offered prescription drugs with very low copayment rates, suggesting the possibility of excess use due to moral hazard. By 2008 this had changed. In April 2008, Gina Kolata of the *New York Times* reported that insurers were adopting new pricing systems for very expensive drugs. With the new systems, many insurers had abandoned the traditional arrangement that has patients pay fixed amounts like \$10 (generic), \$20 (brand name), or \$30 (specialty) for a prescription, no matter the drug's actual cost. Instead, they were charging patients a percentage of the cost of certain high-priced drugs, usually 20 to 33 percent, which could amount to thousands of dollars a month.

Hundreds of drugs used to treat diseases that included multiple sclerosis, rheumatoid arthritis, hemophilia, hepatitis C, and some cancers were priced in this new way. Insurers say the new system reduces everyone's premiums at a time when some of the most innovative and promising new treatments for conditions like cancer, rheumatoid arthritis, or multiple sclerosis can cost \$100,000 and more per year. The system, often called Tier 4, began in earnest with Medicare (part D) drug plans and spread rapidly, incorporated into 86 percent of those plans. Some had even higher copayments for certain drugs, a Tier 5.

Observers also saw Tier 4 in insurance that people buy on their own or acquire through employers. Five years earlier Tier 4 was virtually nonexistent in private plans, but by 2008, at least 10 percent of them had Tier 4 drug categories. Private insurers began offering Tier 4 plans in response to employers who were looking for ways to keep costs down, said Karen Ignagni, president of America's Health Insurance Plans, which represents most of the nation's health insurers. When people who need Tier 4 drugs pay more for them, other subscribers in the plan pay less for their coverage.

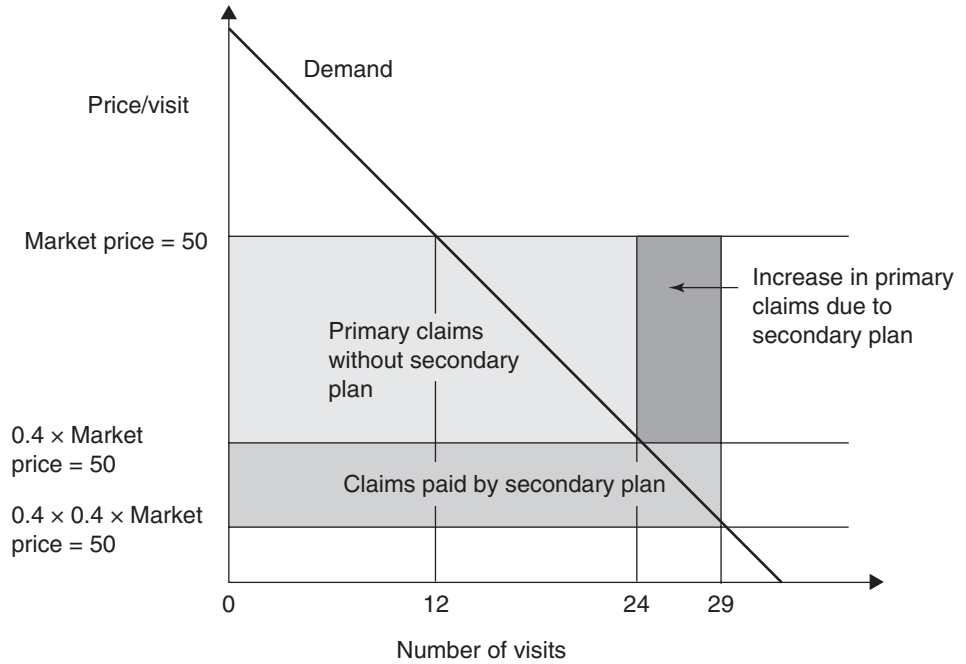
One patient, a 53-year-old woman, had been taking Copaxone since her multiple sclerosis was diagnosed in 2000, buying 30 days' supply under her husband's federal employee insurance. Even though the drug cost \$1,900 a month, insurer Kaiser Permanente required only a \$20 copayment under her husband's plan. Under a new policy, effective January 2008, however, she was to be billed 25 percent of the cost of the drug up to a maximum of \$325 per prescription. Her annual cost would jump to \$3,900, and unless her insurance changed or the drug dropped in price, it would go on for the rest of her life. In mid-April 2008 Kaiser had decided to suspend the change for the program involving federal employees in the mid-Atlantic region while it reviewed the new policy. Other insurance plans, however, continue to advertise plans with Tier 4 coinsurance rates of 25% and higher.

*Source:* Based on information from Kolata, Gina, "Co-Payments Soar for Drugs with High Prices," *New York Times*, April 14, 2008, [www.nytimes.com/2008/04/14/us/14drug.html?\\_r=1&ref=health&oref=slogin](http://www.nytimes.com/2008/04/14/us/14drug.html?_r=1&ref=health&oref=slogin), accessed April 15, 2008.

Figure 8-6 shows a demand curve for visits for the typical General National family if they had no insurance. The family would purchase 12 visits per year, at a price of \$50, spending \$600. If General National is the primary insurer, the out-of-pocket price to its insured will fall by 60 percent to \$20 per visit. As drawn, the lower out-of-pocket price to patients increases quantity demanded to 24 visits. General National will pay \$720, or 60 percent of the \$1,200 total cost; its employees will pay \$480.

Consider, however, the impact of secondary insurance. By paying 60 percent of the remainder, the secondary insurers reduce the out-of-pocket cost to the employees by another 60 percent, from \$20 per visit to \$8 per visit. Not surprisingly, the quantity of visits demanded increases again, this time from 24 to 29 visits. The secondary insurers pay \$12 per visit, or \$348 for the 29 visits. Moreover, the primary insurer, General National, faces increased claims due to demand induced by the coverage of the secondary insurers. General's liability increases from 60 percent of the original \$1,200 in expenditures to 60 percent of \$1,450 in expenditures—the higher level resulting from the secondary coverage.

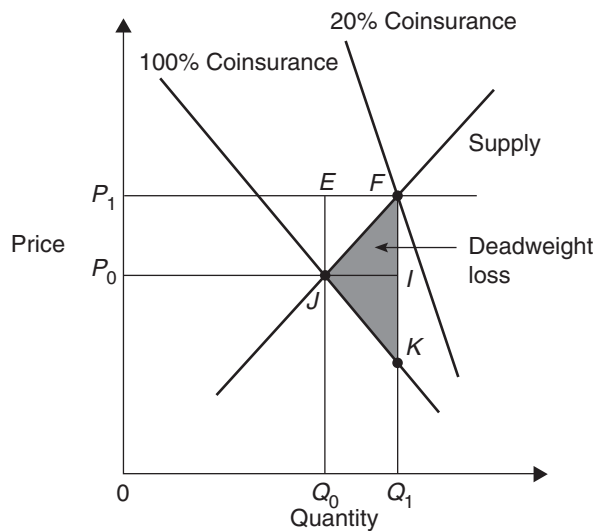
A combination of coverages, while providing additional employee benefits, exacerbates the moral hazard problem brought on in general by health insurance. The inefficiencies and welfare losses that occur when decisions of one firm increase the health care costs facing another pose a difficult problem for policy makers.



**FIGURE 8-6** Impact of Secondary Insurance on Primary Coverage and Utilization

Losses may be even more significant in the market context than in the individual context, as described by Figure 8-7. Clearly, as before, more services are used than are optimal. This comprises both a redistribution of resources (from consumers and insurers to providers) and a deadweight loss (referring to a loss that comes from the misallocation of resources between types of goods). At the original price,  $P_0$ , and quantity,  $Q_0$ , producers were getting the marginal cost to bring the products to market.

The deadweight loss comes from a misallocation of resources among goods (i.e., more health care is provided than should be, according to consumer preferences). Trapezoid  $Q_0JKQ_1$  indicates the incremental benefits induced by the establishment of a coinsurance regime (i.e., the area under the original demand curve).



**FIGURE 8-7** The Effect of Insurance Cost Sharing with Upward-Sloping Supply

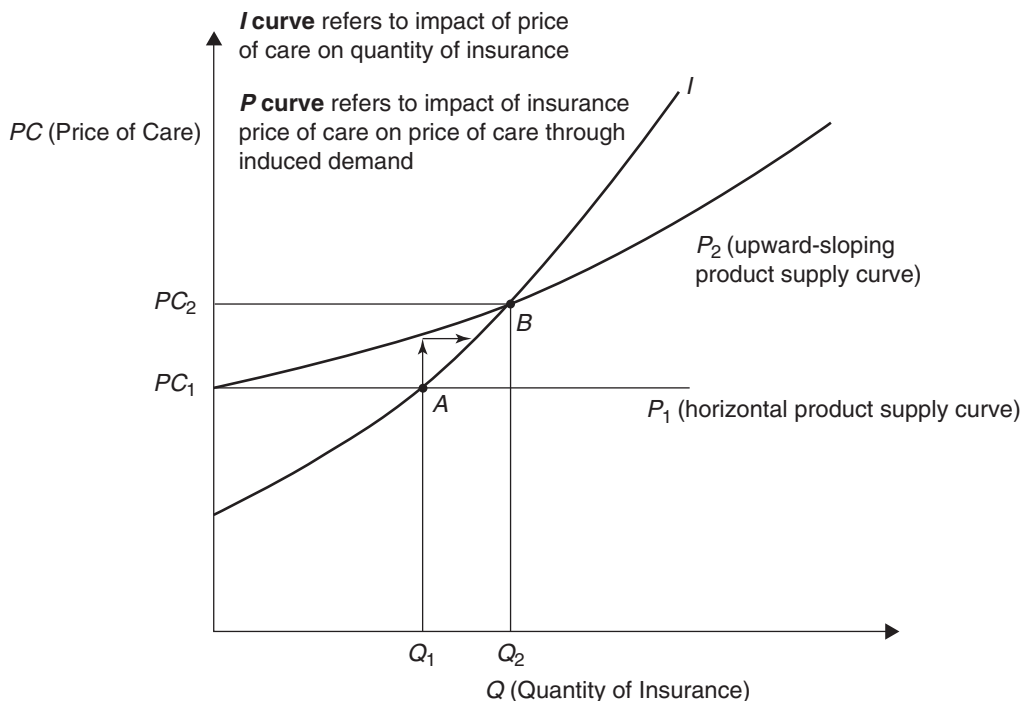
Similarly, the additional resource costs of bringing the treatment level ( $Q_1 - Q_0$ ) to society is trapezoid  $Q_0JFQ_1$ . The deadweight loss from the insurance-induced overproduction of health services is the difference in areas between the two trapezoids, or triangle  $FKJ$ .

### The Demand for Insurance and the Price of Care

Martin Feldstein (1973) was among the first to show that the demand for insurance and the moral hazard brought on by insurance may interact to increase health care prices even more than either one alone. Insurance is related to the expected loss; in health care, this is related in part to the price of care. Increased price of care is related to an increased demand for insurance, as noted in the upward-sloping  $I$  curve in Figure 8-8.

The second impact is that of insurance on the price of care. More generous insurance and the induced demand in the market due to moral hazard lead consumers to purchase more health care. Line  $P_1$  shows that if the supply curve for health care is horizontal, then increased insurance will not increase the price of care above  $PC_1$ . The equilibrium is at point  $A$ , with health care price  $PC_1$  and insurance quantity  $Q_1$ .

If, however, the product supply curve rises, more generous insurance causes market price to increase. We trace this impact as curve  $P_2$ . Start at point  $A$ . The increased product price (the vertical arrow) due to the moral hazard brought on by insurance leads to an increased demand for insurance (the horizontal arrow), which again feeds back on price of care and so on. The moral hazard together with the upward-sloping product supply curve leads to a new equilibrium,  $B$ , with higher price of care,  $PC_2$ , and higher quantity of insurance,  $Q_2$ . The combination of factors leads to a higher price of health care and a higher demand for insurance than would have occurred were there no insurance. Many feel that technology induced price increases along with improved insurance have further increased the price of care.



**FIGURE 8-8** The Interaction of Insurance and Price of Care

## The Welfare Loss of Excess Health Insurance

From the preceding discussion, one would ask why society would support policies that seem only to result in misallocations of resources. In fact, the foregoing analyses concentrate only on the costs. We emphasize that people willingly buy insurance, taking on additional costs to themselves, to protect against the risk of possibly substantial losses.<sup>3</sup> This protection provides major benefits through the protection against risk; the benefits from protection against risk offset the losses discussed here.

Feldstein (1973) was one of the earliest and most prominent of the researchers attempting to calculate the welfare losses of excess health insurance. He measured the cost of the excess insurance by measuring the demand for health care and the coinsurance rate, and calculating the size of the loss polygons in Figure 8-7.

Measuring the benefits also is straightforward conceptually. One can use Figure 8-1 to measure the horizontal difference between Elizabeth's expected utility and her actual utility. This represents the dollar amount that she would have been willing to pay for insurance over and above the amount that she was charged. Provision of insurance to Elizabeth at the actuarially fair rate provides a utility gain to Elizabeth through the reduction of uncertainty. Adding these gains across individuals provides a measure of net benefits to compare to the costs of the excess insurance.

In plain terms, insurance policies impose increased costs on society because they lead to increased health services expenditures in several ways:

- increased quantity of services purchased due to decreases in out-of-pocket costs for services that are already being purchased;
- increased prices for the services that are already being purchased;
- increased quantities and prices for services that would not be purchased unless they were covered by insurance; or
- increased quality in the services purchased, including expensive, technology-intensive services that might not be purchased unless covered by insurance.

Any procedures that raise the coinsurance rate will tend to reduce the costs of excess insurance but also will reduce the benefits from decreased risk bearing. Feldstein found that the average coinsurance rate was about one-third, or 0.33; that is, people paid \$0.33 of every \$1 of costs out of their own pockets. Raising the coinsurance to 0.50 or to 0.67 would cut the amount of insurance purchased, reducing the excess insurance, but also increasing the amount of risk borne by the clients.

The welfare gains from changed coinsurance, then, are:

$$\text{Welfare gains} = \text{Change in benefits} - \text{Change in costs} \quad (8.8)$$

Feldstein's analysis considers the welfare gains from increasing the average coinsurance rate from 0.33 to 0.67. He discovers that the costs fall much more than do the benefits as coinsurance rates rise. He estimates the welfare gains to be approximately \$27.8 billion per year (in 1984 dollars) under the "most likely" parameter values.

Feldman and Dowd (1991) updated Feldstein's 1960s estimates with 1980s parameters from the RAND Health Insurance Experiment regarding both price elasticity of the demand curve and attitudes toward risk. They calculate a lower bound for losses of approximately \$33 billion per year (in 1984 dollars) and an upper bound as high as \$109 billion. For perspective, the upper and lower bounds constituted between 8.9 and 29.1 percent of all 1984 health expenditures.

Manning and Marquis (1996) sought to calculate the coinsurance rate that balances the marginal gain from increased protection against risk against the marginal loss from increased moral

<sup>3</sup> Students might ask about people whose employers pay the entire insurance bill. Most economists believe that employees choose the insurance in lieu of a compensating take-home wage, thus paying for insurance themselves.

hazard, and found a coinsurance rate of about 45 percent to be optimal. Although the impacts of proposed changes depend crucially on the underlying econometric estimates (see Nyman, 1999, for further discussion), the fact that current coinsurance rates have remained far lower than 45 percent suggests an important role for restructuring insurance to reduce excess health care expenditures.

## THE INCOME TRANSFER EFFECTS OF INSURANCE

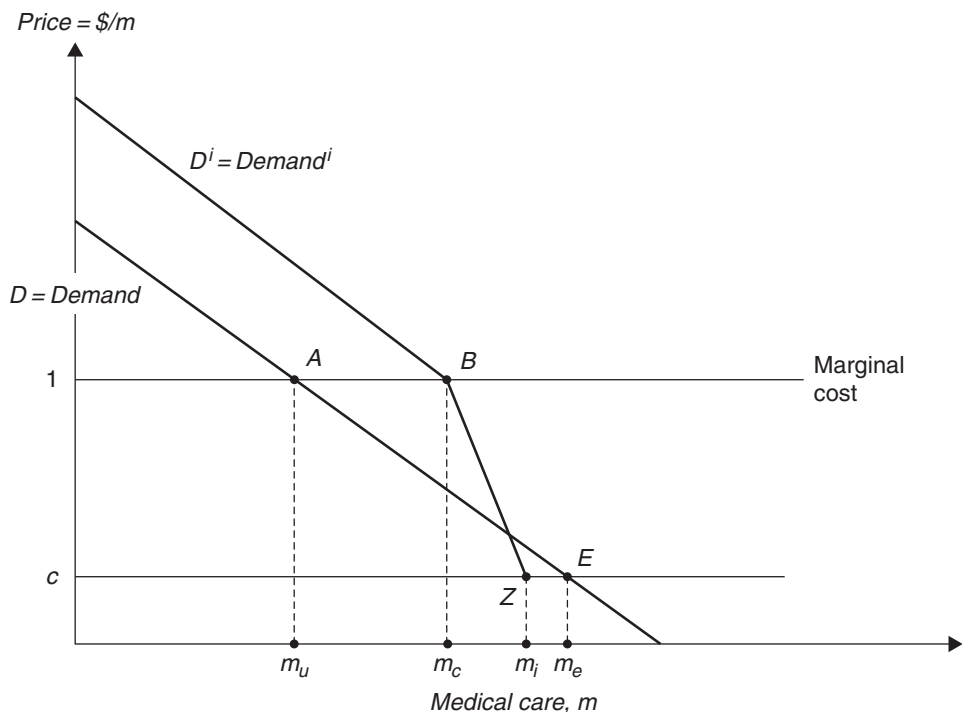
John Nyman (1999) argues that in contrast to the conventional insurance theory, we should view insurance payoffs as income transfers from those who remain healthy to those who become ill, and that these income transfers generate additional consumption of medical care and potential increases in economic well-being. Transfers occur because for most medical procedures, especially expensive procedures that treat serious illnesses, demand depends on the person's becoming ill.

The conventional analyses presented thus far in this chapter imply that:

1. there are no income transfer effects due to insurance, and
2. all moral hazard is due to pure price effects.

What healthy consumer, asks Nyman, would purchase a coronary bypass procedure (or bowel resection or organ transplant) just because insurance is available and the price has dropped to zero? Although the prices of such procedures may fall for all who purchase insurance, only those who are ill will respond to the reduction. Because only the ill respond, the price reduction is the vehicle by which income is transferred from the healthy to the ill.

Figure 8-9 presents the corresponding demand curve analysis with the original demand curve labeled  $D$ .<sup>4</sup> For the type of medical procedures in question,  $D$  represents the behavior of only those



**FIGURE 8-9** Nyman's Decomposition of Moral Hazard Using Demand Curves

<sup>4</sup> Nyman and Maude-Griffin (2001) provides the mathematical foundation for this analysis.

who become ill. If the price equals 1, quantity  $m_u$  is demanded, but if the price falls to coinsurance rate  $c$ , then quantity  $m_e$  will be consumed. The demand curve  $D^i$  illustrates the effect of the insurance contract on the behavior of the consumer who purchases insurance with a coinsurance rate of  $c$  and becomes ill, and consumes an amount of medical care equal to  $m_i$ .

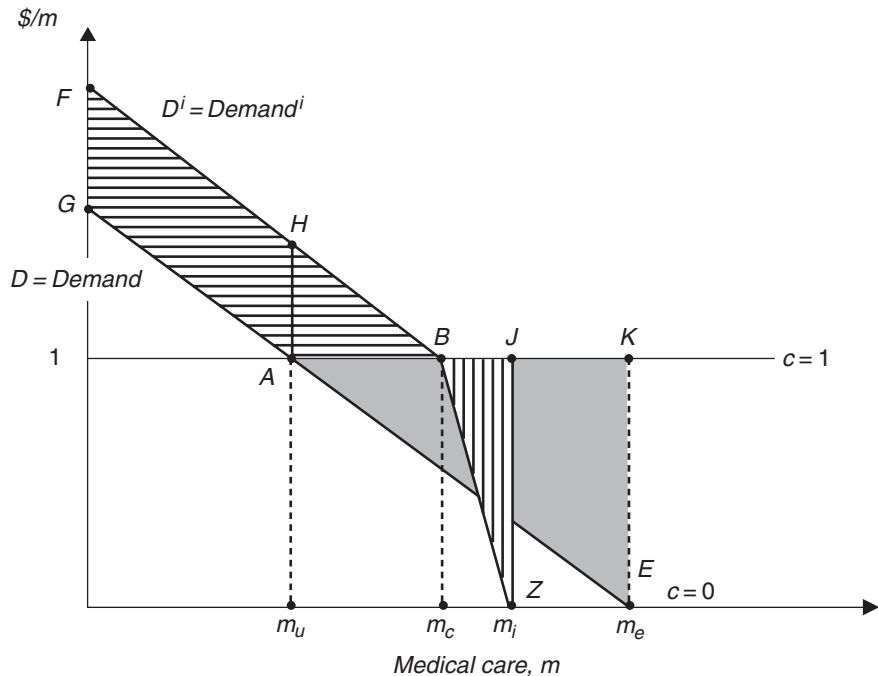
The insurance price decrease is the vehicle for transferring income to the consumer who becomes ill. In Figure 8-8, this income transfer is captured by the portion of  $D^i$  that lies above the original price because, if the income were simply transferred, the increase in willingness to pay would shift out demand at any of those prices. For any given probability of illness, the smaller the coinsurance rate that is purchased in the contract, the greater will be the income transfers and the shift in  $D^i$  compared to the original demand, representing the response to an exogenous price change.

The portion of  $D^i$  that lies below the original price of 1 reflects both behavior and the mechanics of the insurance contract. This portion of  $D^i$  is steeper than the original demand because in order to purchase an insurance contract with successively lower coinsurance rates, the consumer must pay successively greater premiums. That is, two changes occur simultaneously as successively lower coinsurance rates are purchased:

- First, the lower coinsurance rates generate a larger transfer of income to the ill consumer, causing the portion of  $D^i$  above the original price leading to shift horizontally and to the right, leading to point  $B$ .
- Second, the larger premiums associated with lower coinsurance rates generate an ever larger differential between demand curves  $D$  and  $D^i$ .

For example, purchasing a coinsurance rate of  $c < 1$  requires a premium payment that causes a demand differential equal to the horizontal difference between points  $E$  and  $Z$  in Figure 8-9 due to the assumed responsiveness to income. The purchase of a lower coinsurance rate would produce an even larger horizontal difference.

Figure 8-10 illustrates the gain from insurance for the ill consumer who purchases an insurance policy with coinsurance rate 0. The income transfer increases willingness to pay for medical



**FIGURE 8-10** The Net Welfare Gain of the Price Distortion Under Income Transfer Effects



care, shifting out the portion of demand that is above the existing market price of 1. This results in an increase in the consumer surplus of area *FBAG*. From this amount, however, we must subtract welfare loss *BJZ* generated by using a price reduction to transfer this income. Calculate the net welfare gain by subtracting area *BJZ* from area *FBAG*. In comparison, under conventional theory insurance only produces a welfare loss, which would be represented by area *AKE* (shaded in gray).

Here is a numerical example. Suppose that Elizabeth receives a diagnosis of breast cancer at an annual screen. Without insurance, she would purchase a mastectomy for \$20,000 to rid her body of the cancer. In this example Elizabeth has purchased insurance for \$4,000 that pays for all her care (zero coinsurance rate means that  $c = 0$ ) if ill. With insurance, Elizabeth purchases (and insurance pays for) the \$20,000 mastectomy, a \$20,000 breast reconstruction procedure to correct the disfigurement caused by the mastectomy, and an extra two days in the hospital to recover, which costs \$4,000. Total spending with insurance is \$44,000 and total spending without insurance is \$20,000, so it appears that the price distortion has caused \$24,000 in moral hazard spending.

Is this spending increase truly inefficient? We must determine what Elizabeth would have done if her insurer had instead paid off the contract with a cashier's check for \$44,000 upon diagnosis. After accounting for her \$4,000 premium payment, the payoff represents (\$44,000 – \$4,000) or \$40,000 in income transfers that she could spend on anything of her choosing. With her original resources plus the additional \$40,000, assume that Elizabeth would purchase the mastectomy and the breast reconstruction, but not the extra days in the hospital. In Figure 8-10, the mastectomy would be represented by  $m_w$ , the breast reconstruction by  $(m_c - m_w)$ , and the two extra days in the hospital by  $(m_i - m_c)$ . This implies that the \$20,000 spent on the breast reconstruction is efficient and welfare increasing, but the \$4,000 spent on the two extra hospital days (induced by the zero copayment) is inefficient and welfare-decreasing, consistent with the conventional theory.

Nyman's work provides an important extension to the theory of health insurance. The income effects that he identifies are justifiable additions to economic welfare, and we should net them out against potential excess costs brought on by moral hazard.

## CONCLUSIONS

This chapter has concentrated on the unique role of insurance in the health care economy. No other good in people's day-to-day budgets is so explicitly tied to the arrangements for insurance. Such arrangements affect not only expenditures for serious illnesses and injuries, but also plans for more routine expenditures, such as children's well-care visits (for infants and toddlers) and eye and dental care.

We have characterized risk and have shown why individuals will pay to insure against it. Under most insurance arrangements, the resulting coverage leads to the purchase of more or different services than might otherwise have been desired by consumers and/or their health care providers. Considerable health care policy debate focuses on how to structure insurance policies in order to reduce purchases and minimize insurance costs without compromising the health of the insured.

---

## Summary

1. Many illnesses occur rarely and seemingly at random, but when they do, they entail substantial costs.
2. Insurance reduces variability of people's assets by pooling a large number of units and operating on the principle of the law of large numbers. Although outlays for a health event may be highly variable for any given unit in the pool, average outlays for the group are fairly predictable. If they are predictable, they can be insured.
3. One should distinguish between insurance, as is provided through the pooling of risk, and government

programs, such as Social Security, Medicare, and Medicaid, which also redistribute wealth.

4. Insurance can be sold only in circumstances where there is diminishing marginal utility of wealth or income (i.e., when the consumer is risk-averse). With constant marginal utility, requiring that individuals pay actuarially fair premiums would leave them no better off than if they were uninsured.
5. Expected utility is an average measure; the individual either wins or loses the bet.
6. If insurance companies charge more than the actuarially fair premium, people will have less expected wealth through insuring than through not insuring. Even though people will be less wealthy by purchasing insurance, the increased well-being comes from the elimination of risk.
7. In theory, the optimal amount of insurance in the absence of loading costs leads to full insurance against the expected loss. With loading costs, the optimal coverage is less than the expected loss.
8. Moral hazard refers to the increased usage of services when the pooling of risks leads to decreased marginal price for the services. This suggests:
  - more complete coverage for price inelastic services, and
  - earlier development of insurance for services that are most inelastic.
9. Insurance policies lead to increased costs to society because they lead to increased expenditures on health services. They provide increased benefits through the reduction of risks.
10. Some analysts have calculated losses due to excess insurance as between 8.9 and 29.1 percent of all health expenditures. This suggests the importance of restructuring insurance to reduce excess health care expenditures.
11. Nyman shows that insurance payoffs represent income transfers from those who remain healthy to those who become ill, and that under many circumstances these income transfers generate welfare-increasing additional consumption of medical care.

---

## Discussion Questions

1. Discuss the difference between cardinal and ordinal utility. Why is cardinal utility necessary for the analysis of risk and insurance?
2. What does the term *moral hazard* mean? Give examples.
3. The deductible feature of an insurance policy can affect the impact of moral hazard. Explain this in the context either of probability of treatment and/or amount of treatment demanded.
4. Describe the benefits to society from purchasing insurance. Describe the costs. Define and discuss the welfare gains from changes in insurance coverage.
5. If only risk-averse people will buy health insurance, why do many people who buy health insurance also buy lottery tickets (an activity more consistent with risk taking)? Speculate on the differences and similarities.
6. Some brokers (called viatical brokers) offer cash settlements in advance to people with terminal diseases who have life insurance, paying them in advance of their death. Is this practice ethical? Is it ethical for the brokers to offer settlements to elderly people simply because they may die soon?
7. Because health insurance tends inevitably to cause moral hazard, will the population necessarily be overinsured (in the sense that a reduction in insurance would improve welfare)? Are there beneficial factors that balance against the costs of welfare loss?
8. By Nyman's arguments do all increased expenditures become welfare enhancing? Give examples of some that enhance welfare. Give examples of others that do not.
9. The game show *Deal or No Deal*, popular throughout the world, provides many elements of risk and expected value. Discuss the ways that the decision as to whether to "take the money" or to continue involves evaluation of risk and expected value.

---

## Exercises

1. Suppose that Nathan's employer provides a health insurance policy that pays 80 percent of \$1 over the first \$100 spent. If Nathan incurs \$1,000 in expenses, how much will he pay out of pocket? What percentage of his expenses will this be?
2. Suppose that rather than flipping a coin, one rolls a die. If the value is 1, 2, 3, or 4, the player wins \$1. If it is 5 or 6, the player loses \$1. Calculate the expected return.
3. A standard roulette wheel has an array of numbered compartments referred to as "pockets." The pockets are red,

black, or green. The numbers 1 through 36 are evenly split between red and black, while 0 and 00 are green pockets. For every \$1 that one bets on red, one wins \$1 if the roulette ball lands on red and loses if it lands otherwise. Similarly for black. What is the expected return to a red or a black bet? Why?

4. (a) Draw a utility of wealth curve similar to Figure 8-1 for consumers who are not risk-averse. How would its shape differ from Figure 8-1?
- (b) Draw a utility of wealth curve similar to Figure 8-1 for “risk-lovers.” How would its shape differ from Figure 8-1?
5. (a) Show the gains from insurance, if any, in Exercise 4a.
- (b) (Difficult) Show the cost of insurance in Exercise 4b.
6. We have discussed the role of utility functions in the purchase of insurance.
  - (a) Suppose Edward’s utility function can be written as:
 
$$U = 20Y$$
 where  $U$  is utility and  $Y$  is income per month.
 

What is his marginal utility if income is \$1,000 per month? \$2,000 per month? Is Edward likely to insure against loss of income? Why?
  - (b) Suppose instead that Edgar’s utility function can be written as  $U = 200Y^{0.5}$ . What is his marginal utility if income is \$1,000 per month? \$2,000 per month? Is Edgar likely to buy insurance against loss of income? Why?
  - (c) Suppose that Edmund’s utility function can be written as  $U = 0.5Y^2$ . What is his marginal utility if income is \$1,000 per month? 2,000 per month? Is Edmund likely to buy insurance against loss of income? Why?
7. Suppose, if ill, that Fred’s demand for health services is summarized by the demand curve  $Q = 50 - 2P$ , where  $P$  is the price of services. How many services does he buy at a

price of \$20? Suppose that Fred’s probability of illness is 0.25. What is the actuarially fair price of health insurance for Fred with a zero coinsurance rate?

8. In Exercise 7, if the insurance company pays Fred’s entire loss, what will Fred’s expenses be? How much will the company pay? Will it continue to offer him insurance at the actuarially fair rate? Why?
9. Suppose that the market demand for medical care is summarized by the demand function:

$$Q_d = 100 - 2p$$

and the market supply is summarized by the supply function:

$$Q_s = 20 + 2p$$

- (a) Calculate the equilibrium quantity and price, assuming no health insurance is available.
- (b) Suppose that health insurance is made available that provides for a 20 percent coinsurance rate. Calculate the new equilibrium price and quantity. (Hint: How does the demand curve shift?)
- (c) Calculate the deadweight loss due to this insurance.
10. Suppose, in Exercise 9, that the coinsurance rate was raised to 50 percent.
  - (a) Calculate the new equilibrium price and quantity. (Hint: How does the demand curve shift?)
  - (b) Calculate the deadweight loss due to this insurance.
  - (c) How does your answer compare to the deadweight loss in Exercise 9?
11. Consider the discussion in the text about Elizabeth’s breast cancer treatment. Using Figure 8-9, calculate the net welfare benefits if  $m_u = 20,000$ ,  $m_c = 40,000$ , and  $m_i = 44,000$ . To aid in the calculations, assume that point  $G$  has a value of 2 and point  $F$  has a value of 3.

# Consumer Choice and Demand

- Applying the Standard Budget Constraint Model
- Two Additional Demand Shifters—Time and Coinsurance
- Issues in Measuring Health Care Demand
- Empirical Measurements of Demand Elasticities
- Impacts of Insurance on Aggregate Expenditures
- Other Variables Affecting Demand
- Conclusions

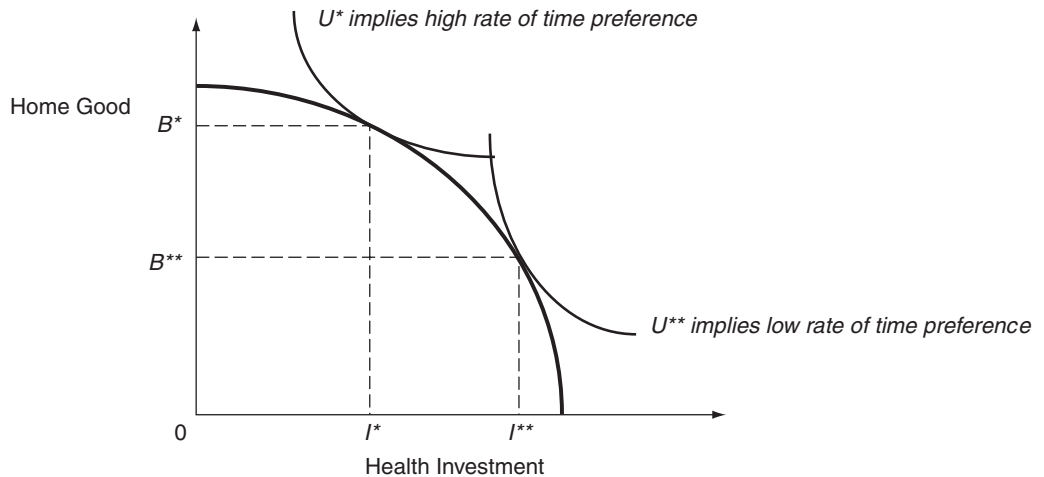
We have described health capital investment as a choice made by the consumer who values health but who also values the home good, which involves all the goods made possible by using income and time. These ultimate goods, health capital and the home good, are produced by the consumer by allocating a portion of time to each, as well as a portion of income to those marketed goods, such as medical care,  $M$ , and various other goods,  $OG$ , that are used in the production process.

Figure 9-1 helps illustrate this transformation from one model (see Chapter 7) to the other—the conventional analysis of choice over marketable goods. The production possibilities frontier in the figure illustrates the consumer's trade-off between health investment and the home good. Choosing these optimal quantities,  $I^*$  and  $B^*$ , the consumer also implicitly chooses an allocation of time available for these production tasks, to work that provides income, and to leisure.

Indifference curves  $U^*$  and  $U^{**}$  provide insights into how different people may choose between the present and the future through a concept known as the *rate of time preference*. Curve  $U^*$  represents an indifference curve for Tom. Tom has a high rate of time preference and places a large value on current consumption relative to *future* consumption. The latter is adversely affected by the higher mortality risks and other consequences (e.g., lower future earnings) of poor health. As such, he will choose high present consumption  $B^*$  and relatively low health investment level  $I^*$ .

His brother Jerry, in contrast, has a low rate of time preference, as noted by curve  $U^{**}$ . He places a low value on current consumption relative to future consumption and is more willing than Tom to invest in health. He chooses levels  $B^{**}$  and  $I^{**}$ .

We now wish to examine how income will be used to buy those marketable goods, such as medical care, that will in turn help the consumer produce health investment and the home good. The consumer faces a trade-off



**FIGURE 9-1** Demand for Health Capital Determines the Optimal Amounts of the Home Goods and Health Capital Investment

in the marketplace between the use of income to purchase medical care and the use of the money to purchase other goods. This trade-off is the budget constraint of standard consumer theory and is the focus of this chapter.

### APPLYING THE STANDARD BUDGET CONSTRAINT MODEL

As with the demand for health capital model, standard indifference curve analysis of consumer choice under a budget constraint describes the consumer with fairly strong assumptions. We assume that the consumer is rational and perfectly informed, that there is no uncertainty about the future, and that important decisions are made as if the future were known with certainty. Although we will later relax some of these assumptions, this model produces many reliable predictions on consumer behavior related to health.

Many might object to this approach at the start, believing that we have few choices when we need health care, at least for the urgently ill. If you are lying on the pavement and the ambulance arrives, do you ask for a list of prices and providers?

Nevertheless, a theory of rational choice over health care and other goods can be defended on several grounds. First, many health care options leave room for some thoughtful consideration or at least some planning. Second, the physician serves as an agent for patient-consumers and can make rational choices on their behalf even in urgent situations.<sup>1</sup> Finally, the ultimate test of any theory is whether it predicts well, and we will show empirically that people, as consumers of health care, do respond to economic incentives.

In economic theory, the logic of consumer choice is straightforward. It indicates that consumers can choose any affordable combination or bundle of goods, and from among these affordable bundles, they will choose the most preferred. The depiction of this choice requires two elements:

- The consumer's preferences—described by a set of indifference curves
- The consumer's budget constraint—described by the straight budget line

To make the graphical depiction possible, we abstract from the many goods available in the real world and assume instead that only two goods are available. The results for this two-good world

<sup>1</sup> Problems may occur in describing a relationship where the provider, acting as an agent for the patient, helps to determine the amount of care. This is known as supplier-induced demand, and it is explained in Chapter 15.

generally hold when the model is extended to many goods. Let one of these two goods represent a composite of other goods, and call this good Other Goods, or *OG*. Assume that the health care good is physician office visits consumed during a year, or VISITS. The consumer's name is Ellen Anderson.

### The Consumer's Equilibrium

Figure 9-2 depicts these elements of the choice problem. The indifference curves labeled  $U_1$ ,  $U_2$ , and  $U_3$  represent some of Ellen's indifference curves (not all are depicted), and together the indifference curves describe her preferences. The indifference curve  $U_1$ , for example, represents all points—that is, bundles of *OG* and VISITS—that provide her with utility level  $U_1$ . Utility is an index of preferences that can most easily be understood as a measure of satisfaction. Because *OG* and VISITS,  $V$ , are both goods to the consumer, it follows that indifference curve  $U_2$  is preferred to  $U_1$  and so on; that is, “higher” indifference curves are preferred.

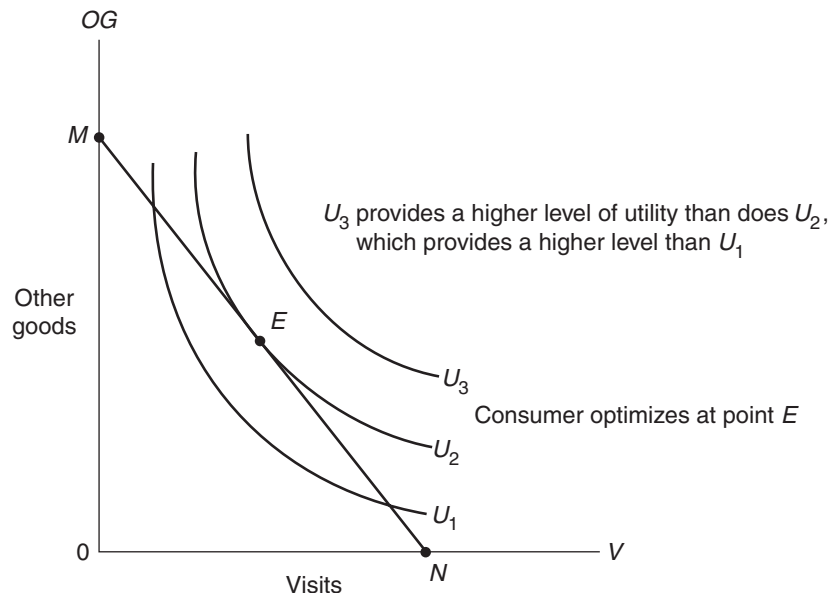
Let Ellen's budget be  $Y$  dollars for the period. If the price of other goods, *OG*, is  $P_{OG}$  and the price of VISITS is  $P_V$ , then the sum of her expenditures,  $P_{OG} \times OG$  plus  $P_V \times VISITS$ , cannot exceed her income,  $Y$ . To spend all her income means to be on the budget line, which is given by the equation:

$$Y = P_{OG} \times OG + P_V \times VISITS$$

Point  $M$  represents the amount of other goods consumed if no visits occur. Point  $N$  represents the amount of visits if no other goods are purchased. The budget line is shown as line  $MN$  in Figure 9-2, and its slope will be given by  $-P_V/P_{OG}$ , which is the negative of the ratio of prices.

The consumer equilibrium is shown as point  $E$  in Figure 9-2, a point of tangency between the highest indifference curve attainable,  $U_2$ , and the budget line. In contrast, all points on indifference curve  $U_3$  are unattainable, and points on  $U_1$  are not chosen because the consumer can afford points she prefers to these. The equilibrium point  $E$  is a point of tangency, meaning that the slope of the indifference curve equals the slope of the budget line at this point.

The slope of the indifference curve is called the marginal rate of substitution (MRS). It tells the rate at which Ellen is willing to trade other goods for physician visits. Recall that the slope of the



**FIGURE 9-2** Consumer Equilibrium Analysis

budget line is the negative of the ratio of prices. This is the rate at which she is able to trade other goods for physician visits at current market prices. An equilibrium is reached only if the rate at which she is willing to trade the two goods, the MRS, is equal to the rate at which she is able to trade the two goods,  $-P_V/P_{OG}$ . This will have the result that in equilibrium, a dollar's worth of  $OG$  will yield the same extra utility as a dollar's worth of VISITS.

### Demand Shifters

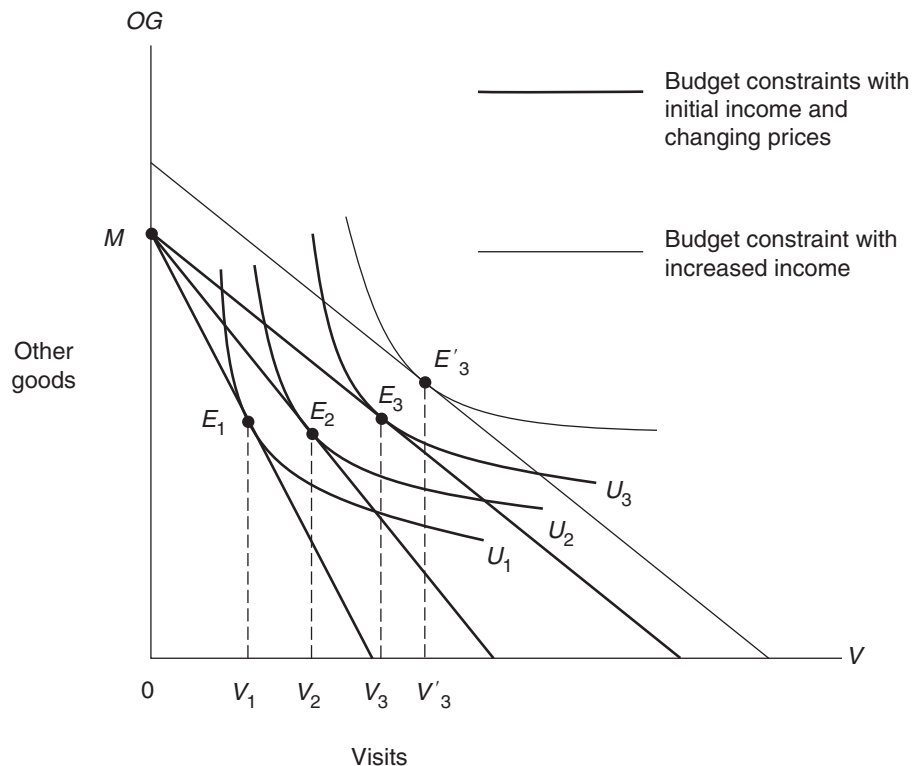
Ellen's response to price changes can be determined by examining the new equilibria that would occur as the price of  $V$  varies. Figure 9-3 shows the effects of changes in prices at initial income  $Y$ , dropping from the highest price,  $P_V^1$ , to a lower price,  $P_V^2$ , to the lowest price,  $P_V^3$ .

At the highest price,  $P_V^1$ , income  $Y$  buys  $V_1$  visits at equilibrium point  $E_1$ . At the lower price,  $P_V^2$ , Ellen chooses equilibrium point  $E_2$  (with  $V_2$  visits), and at the lowest price,  $P_V^3$ , Ellen chooses equilibrium point  $E_3$  (with  $V_3$  visits). The number of visits,  $V$ , increases because visits have become less expensive relative to other goods.

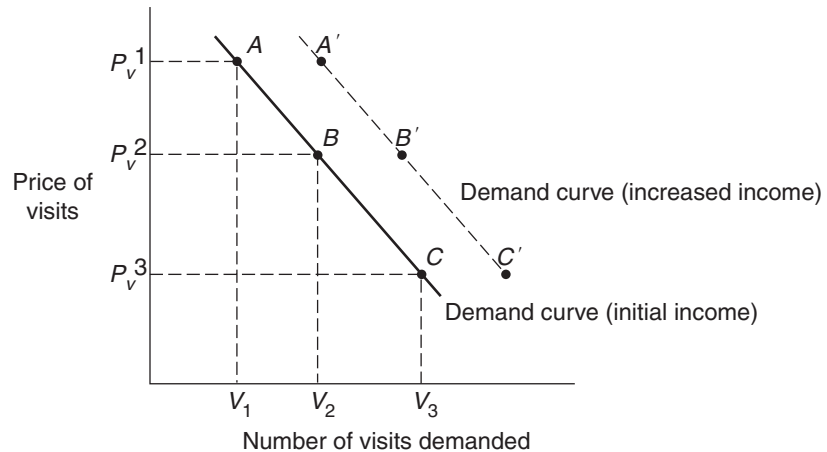
Figure 9-4 plots a demand curve relating the price of visits to equilibrium quantity demanded. The data come from Figure 9-3. Point  $E_1$  from Figure 9-3 corresponds to point  $A$  in Figure 9-4, and similarly points  $E_2$  and  $E_3$  correspond to points  $B$  and  $C$ . The demand curve summarizes Ellen's response to price changes, holding income and preferences constant.

We use price elasticity to measure the responsiveness of the consumer's demand to changes in price. Price elasticity,  $E_p$ , is the ratio of the percentage change in quantity demanded to the percentage change in price. Algebraically, it is:

$$\text{Elasticity} = E_p = \frac{(\Delta Q/Q)}{(\Delta P/P)} \tag{9.1}$$



**FIGURE 9-3** Change in Number of Visits as Visit Price and/or Income Changes



**FIGURE 9-4** Demand Curve Derived from Figure 9-3

Here, the numerator is the change in quantity divided by the initial quantity level; that is, the numerator is the percentage change in quantity. Because the demand curve slopes downward, the percentage change in quantity (the numerator) will always be negative in response to an increase in price. Likewise, the denominator is the percentage change in price. The higher the elasticity in absolute value (the farther away from 0), the more responsive the consumer is to price. Note that we can write equation (9.1) the *price elasticity of demand* as:

$$\text{Elasticity} = E_p = \left( \frac{\Delta Q}{\Delta P} \right) \left( \frac{P}{Q} \right)$$

A similar analysis develops the consumer's response to changes in income. Returning to Figure 9-3, recall that point  $E_3$  is determined by income, preferences, and price,  $P_V^3$ . Suppose that Ellen's income now increases. Because the relative prices are not changed by the income increase, the slope of the budget line is not changed, but Ellen can now buy more of both visits and other goods. Her new equilibrium point is  $E_3'$ . This is translated in Figure 9-4 to point  $C'$ . We can similarly draw points  $A'$  and  $B'$  on Figure 9-4 to indicate the impacts of an income change and prices,  $P_V^1$  (new point  $A'$ ) and  $P_V^2$  (new point  $B'$ ).

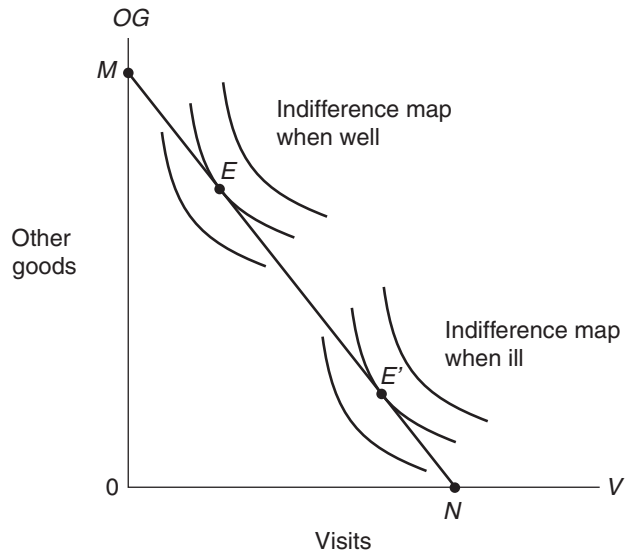
The responsiveness of demand to changes in income is measured by the income elasticity. Income elasticity,  $E_Y$ , is the percentage change in quantity demanded divided by the percentage change in income:

Income elasticity =  $E_Y = (\Delta Q / Q) / (\Delta Y / Y)$ , or

$$E_Y = \frac{\Delta Q}{\Delta Y} \left( \frac{Y}{Q} \right) \quad (9.2)$$

Finally, although two-dimensional indifference curves are not well suited to the handling of larger numbers of substitute and complement goods, the effects of changes in the prices of other goods can be analyzed. One would expect that increases in the prices of substitutes to physician visits (hospital outpatient services, visits to other providers) would increase the demand for office visits. In other words, an increase in the price of a substitute will shift the demand curve to the right in Figure 9-4. Increases in the prices of complements (diagnostic services) would reduce demand for office visits.





**FIGURE 9-5** Changed Preferences Due to Illness

### Health Status and Demand

Figure 9-5 illustrates how differences in patient health status can be handled. Suppose that Ellen is viewed in two different time periods in which her situation is the same in all economic respects except her health status. In Period 1 (equilibrium point  $E$ ), she is fairly healthy. In Period 2, her overall health status is lower because she is ill. The change in health status will affect Ellen's preferences (often referred to as tastes) over VISITS and  $OG$  as reflected by different sets of indifference curves and changed levels of physician care, here point  $E'$ .

Note that although Ellen consumes more visits in Figure 9-5 when she is ill, in both cases she has the same MRS of visits for other goods at the equilibrium. Indeed, as long as the prices of the two goods do not change, all consumers in equilibrium will adjust to the prices until all consumers have the same MRS.

The analysis thus far suggests that price, income level, tastes, health status, and other circumstances influence the consumption of physician services. However, other considerations, the roles of insurance and of time, cannot be overlooked.

### TWO ADDITIONAL DEMAND SHIFTERS—TIME AND COINSURANCE

Two demand-shifting variables “look” like changes in the price: time price and coinsurance. First, Ellen's time price, the value of the time she must give up for a physician visit, can represent a significant portion of her full price. Second, insurance causes Ellen's effective price, the price paid “out of pocket,” to be reduced below the market price.

#### The Role of Time

Recall from Chapter 7 that time is an important element in the demand for health. The consumption of health care services requires considerable time for some services and procedures. Economic observation suggests that people value their time. Many turn down additional work, even at increased wages, such as “time and a half” overtime. Still others decline to drive across town to save \$5 or \$10 on an item, even though the cost of driving is far less than the \$5 or \$10. These choices probably occurred because the additional time spent wasn't worth it to the consumer.

Given the opportunity cost of time, a focus on the money costs of health care ignores a substantial portion of the economic costs. The discrepancy between the total economic costs and the

money costs will be especially large for low-priced services, for services where patient copayments are small, and for patients with high opportunity costs of time.

As an example of time cost effects, suppose that Ellen must go to the doctor for a 10-minute visit. It will take her 15 minutes to travel each way (30 minutes in all), 20 minutes to wait in the office, and 10 minutes with the doctor. Suppose further that the money cost of the visit is \$25, and that she values her time at \$10 per hour. Traveling and parking cost \$5 total. The full price of each visit is then \$40:

- One hour of time valued at \$10
- One visit priced at \$25
- Travel and parking costs at \$5

Figure 9-6 illustrates that Ellen demands six visits when her full price is \$40. A money price increase of \$5 causes the new full price to be \$45, at which she demands five visits. Restating the price elasticity formula from equation (9.1), we find that the elasticity with respect to the full price is

$$E_p = \frac{(\Delta Q/Q)}{(\Delta P/P)} = \frac{(-1/5.5)}{(+5/42.5)} = -1.545$$

As appropriate, we use “arc elasticity” to evaluate the price at the midpoint (42.5) between the beginning (40) and the ending (45) price, and similarly for quantity.

Here,  $P$  represents the full price; that is,  $P = P_M + P_T$ . The full price is the sum of money price and time price. In contrast, the money price elasticity in this case is:

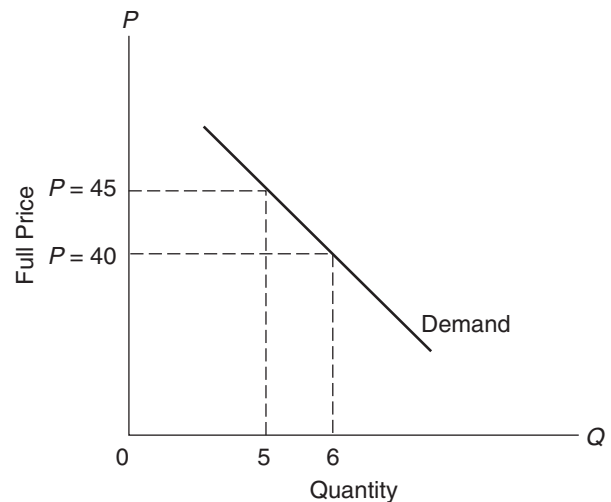
$$E_{pM} = \frac{(\Delta Q/Q)}{(\Delta P/P)} = \frac{(-1/5.5)}{(+5/27.5)} = -1.000$$

In general, the money price elasticity is smaller than the full price elasticity by the same proportion as the money price is smaller than the full price. To make sense of this, try comparing the ratio:

$$P_M \div (P_M + P_T) = 27.5 \div 42.5 = 0.647$$

to the ratio:

$$E_{pM} \div E_p = -1.000 \div (-1.545) = 0.647$$



**FIGURE 9-6** Demand and Time Price for Physician Visits

How might this apply? Assuming that the poor have a lower opportunity cost of time than the well-to-do, one would predict that they would more likely tolerate or endure long waiting times in clinics or physician offices. At the same time, even those poor whose physician fees are subsidized (e.g., by Medicaid) must pay their time price. Wishing to increase physician visits among the poor, we might choose to reduce the time price by building nearby clinics and expanding outreach programs, a strategy that has been developed in many localities.

In practice, does time price affect demand? In pioneering work on time price, Acton (1975, 1976) examines the effects of travel times, waiting times, and other variables on quantity demanded of outpatient visits and physician care. Table 9-1 reports his elasticity estimates. For example, outpatient care,  $T_{Out}$ , is the own-time price, and  $T_{Phys}$  refers to the other good (cross-time price). The own-time price elasticity for outpatient visits,  $-0.958$ , nearly reaches unity, and the own-time price elasticity for physician visits, though small, also shows the importance of time. The positive cross-elasticities suggest that outpatient and physician visits are substitutes.

Subsequent work usually supports an important role for time. Coffey (1983) finds time price also relevant to the decision to seek care initially, as well as the quantity consumed, though her estimates are small (with the exception of public provider care). Mueller and Monheit (1988) find time-price elasticities for dental care to significantly affect the quantity of dental care consumed. Likewise, the National Health Service in the United Kingdom, which has eliminated most price constraints on the use of health care resources, finds the remaining waiting time price to be an important rationing factor (Blundell and Windmeijer, 2000). Varkevisser and colleagues (2010) confirm the importance of time and its affect on demand. By studying non-emergency outpatient visits for neurosurgery in Dutch hospitals (where there are no patient copayments), the authors found that time elasticities across hospitals were consistently high though they varied widely ( $-1.4$  to  $-2.6$ ). Does time-price affect health care demand? Yes. It makes sense in theory, and it matters in practice.

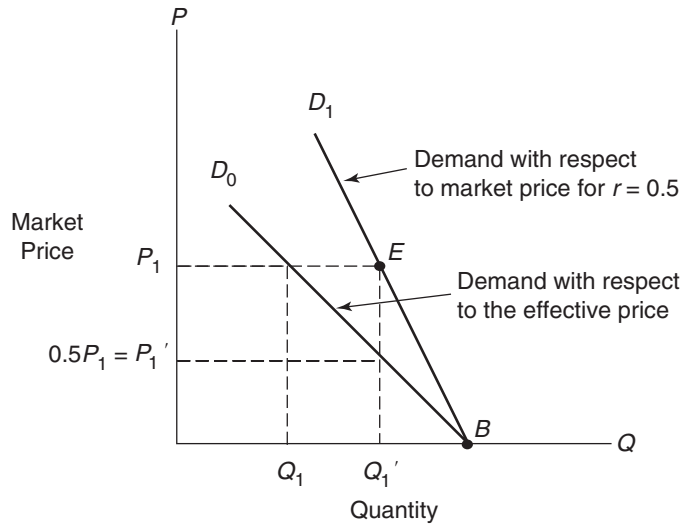
### The Role of Coinsurance

Building on our study of insurance in Chapter 8, we see coinsurance as a demand shifter that works by modifying the effective price. When a third party, such as an insurance company, pays a portion of the hospital bill, the remaining portion paid by the consumer is called the coinsurance rate  $r$ . Thus, more insurance means a lower  $r$ .

**EFFECTS OF REDUCED COINSURANCE ON THE INDIVIDUAL CONSUMER** Suppose Ellen has no health insurance and pays all her health care bills out of pocket. Figure 9-7 shows Ellen's demand for health care as  $D_0$ . Because she is uninsured, the market price also is always Ellen's effective (out-of-pocket) price. At a market price of  $P_1$ , her quantity demanded is  $Q_1$ . A simple thought experiment reveals the issue. Suppose Ellen is given a health insurance policy at no charge (that is, with no impact on the rest of her disposable income) that pays 50 percent of each of her bills, giving her a coinsurance rate of  $r = 0.5$ . The market price,  $P_1$ , is no longer the effective price; the effective price becomes  $0.5 \times P_1 = P_1'$ . Using her demand curve,  $D_0$ , as our guide, we see her now demanding  $Q_1'$ . This develops a first principle—her quantity demanded under coinsurance can be found along the out-of-pocket demand curve, provided we identify and apply the effective price.

**TABLE 9-1 Acton's Time Valuation Equations**

Dependent Variable	Outpatient Visits	Physician Visits
Elasticity with respect to $T_{Out}$	-0.958	0.640
Elasticity with respect to $T_{Phys}$	0.332	-0.252



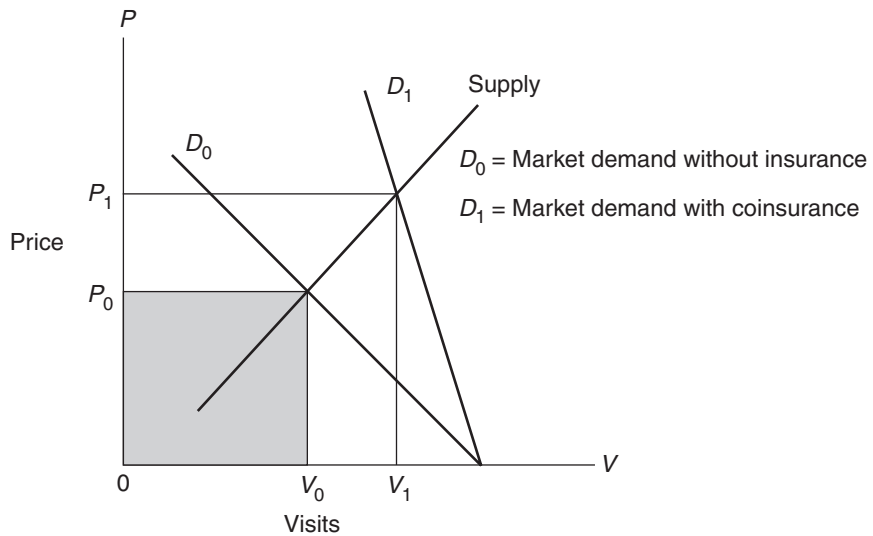
**FIGURE 9-7** The Effect of a Coinsurance Rate on Health Care Demand

It is more useful to identify her demand curve with respect to the market price. Ellen acted as though her health care demand had shifted, and this “rotating shift” can be shown to be equivalent to the previous analysis. In Figure 9-7, let us reverse the question and ask: If Ellen demands  $Q_1'$  when she has 50 percent insurance coverage and faces an effective price of  $P_1'$ , then what market price is she apparently willing to pay (part paid by her out-of-pocket and part paid by her through her insurance company)? The answer is  $P_1$ , and the resulting combination of quantity demanded and market price is Point  $E$  in Figure 9-7. By plotting such points (not forgetting the case where  $P = P' = 0$ ), we trace the demand curve with respect to the market price,  $D_1$ . Ellen, by receiving the insurance, will act just like an uninsured Ellen whose health care demand curve rotated to the right, hinged at Point  $B$ .

The exercise makes two theoretical facts clearer: Insurance will increase Ellen’s demand for health care, and insurance will make Ellen’s demand for health care less elastic. Suppose Ellen’s coinsurance rate were zero, meaning she pays nothing for health care. Would her demand be even less elastic? Most health economists would predict that her demand curve would become vertical, hence perfectly inelastic. Since she pays nothing, her demand is totally unresponsive to money price.

**MARKET EFFECTS** The effect of a reduced coinsurance alone, for Ellen, is an increase in the quantity demanded. Ellen does not demand enough care to influence market prices. Individual consumers are price takers; their individual actions have no effect on the price so they face essentially a horizontal supply curve. Suppose, however, that the coinsurance rate changes for many consumers in the market. For the market as a whole, the relevant supply curve slopes upward, indicating that higher prices might be required to motivate producers to offer greater market quantities.

Figure 9-8 shows an equilibrium of price and quantity with an upward-sloping supply curve. The original market equilibrium price is  $P_0$  and the equilibrium utilization is  $V_0$ . In this case, if coinsurance rates are generally reduced, the increased market demand will raise market quantity demanded to  $V_1$  and the market price to  $P_1$ . Total health care expenditures will rise from  $P_0V_0$  to  $P_1V_1$ . Many economists feel that such effects are major reasons for the increases in health care costs in the United States.



**FIGURE 9-8** Market Impact of Coinsurance

## ISSUES IN MEASURING HEALTH CARE DEMAND

With the current concern about health care expenditures, reliable estimates of demand elasticity become essential. Recall that elasticity measures the responsiveness of demand to a change in a related variable. Price elasticity helps determine the effects of health insurance practices and policies. The effect of public policies to improve the accessibility of health care will depend on the money and the time price elasticities.

Prices, incomes, time prices, and coinsurance rates do matter. Increased prices and coinsurance rates reduce demand for services. Raising income increases the demand for services. “Economic” factors are not consumers’ only considerations, but they matter.

In this section, we focus on variables of interest to science and public policy. We examine how health care demand responds to money price, insurance coverage, and time price. In addition, we examine the effects on market demand of income and other variables. Each study attempts to apply econometrics to estimate a demand function statistically; all variables relevant to demand are considered simultaneously. For ease in exposition, we consider the important variables separately.

Reviewing the difficulties faced by researchers and the differences between studies can be helpful to understanding the results. In effect, we ask the question: Why do the reported elasticities vary so often from one study to another? In most cases, the differences arise because of the different choices the researcher made in the face of problems common to research in this field. We identify five issues.

### Individual and Market Demand Functions

Our analysis so far has focused on the individual. It suggested the following type of demand function for physician visits, referred to as  $V$ :

$$V = f(P, r, t, P_0, Y, HS, AGE, ED, \dots)$$

where  $P$  is price per visit,  $r$  is the patient’s coinsurance rate,  $t$  is a time price,  $P_0$  is the price of other goods,  $Y$  is a measure of income, HS is the patient’s health status, and AGE and ED stand for

variables such as age and education to reflect other need and taste factors. This functional notation shows that certain variables are likely to affect  $V$ , but it does not specify the relation exactly.<sup>2</sup>

Often, however, economists are interested in market demand functions. Care is needed to move from individual to market demand. Even the measure of utilization poses a challenge. For example, most studies will use the number of visits per person (rather than the total quantity of visits) as the dependent variable. They then attempt to control for the size of the market by considering total population. This leads to serious problems in the interpretation of results.

### Measurement and Definitions

Unlike the carpenter's simpler problem of measuring the length of a wall, there are alternative definitions of health care quantities, as well as many alternative measuring tools. Investigators often measure the quantity of services in dollar expenditures. One problem is that expenditures reflect a complex combination of price of care, quantities of care, and qualities of care.

Alternative measures include quantity of visits, patient days, or cases treated, yet these do not necessarily measure the intensity of care. One person may spend five days in the hospital for observation; another may spend five days for brain surgery. Consequently, the literature contains a variety of measures and a variety of reported elasticities.

A related problem is to define the price of services. Because of the prevalence of health insurance, most patients do not pay the full price for their treatments. Moreover, the price they pay may be related to the size of the bill because of deductibles, coinsurance, or limits. A \$50 treatment, for example, may cost \$50 if it occurs before the deductible limit is reached, or \$10 (assuming 20 percent coinsurance) if it occurs after the deductible limit has been reached. The statistical problems in this case are fairly complicated, but it suffices here to note that the resulting price elasticities may vary.

### Differences in the Study Populations

Different researchers, naturally, use different samples or populations. Elasticities will differ between populations and even within populations at different points in time. For example, many health economists believe that income elasticities for health care have become smaller over the years in the United States, presumably because of the effects of programs like Medicare and Medicaid.

Furthermore, it is possible, in theory, for Californians to have a different price elasticity for physician services than Minnesotans. People in one state may be older (for example, Florida) or have better access to larger varieties of health providers. It is theoretically possible that people will exhibit different price elasticities for dental care than for pediatric care. Thus, some variation in reported elasticities is inevitable even when one uses the "same" measures, definitions, and techniques.

### Data Sources

Populations differ between studies, and the sources of data may differ in ways that result in different elasticity estimates. For example, a common source of health care data is the insurance claim. Claims data, however, are limited to services covered by insurance and used by the insured. Furthermore, claims data often lack detail on individuals' characteristics, such as education and income. In contrast, health interview survey data often incorporate personal data, but their accuracy depends on the recall ability of the people being interviewed.

<sup>2</sup> Econometricians often use the OLS method discussed in Chapter 3. In this case, the regression is:

$$V = b_0 + b_1P + b_2r + b_3t + b_4P_0 + b_5Y + b_6HS + b_7AGE + b_8ED + \varepsilon$$

with the variables defined as before, and  $\varepsilon$  is the error term.

## Experimental and Nonexperimental Data

Much of health care demand research used nonexperimental data, and thus the researcher could not control the environment or assure that other extraneous variables were held constant. These data typically represent samples across individuals or markets—that is, a slice of experience. If the necessary assumptions hold, then available statistical techniques can provide valuable analytical insights.

A useful alternative involving the natural experiment is sometimes possible. A natural experiment, for example, may occur when a given area changes its health insurance plan (e.g., Tilford and colleagues (1999) studied the response of previously uninsured school children in the Mississippi Delta region of Arkansas after a program provided them with health insurance). The change enables one to observe differences in health care utilization before and after. We presume that only the policy changes; all other factors are held constant. Unfortunately, other demand-related factors often change also.

In a controlled experiment, subjects are randomly assigned to treatment and control groups to measure responses directly to changes in the levels of demand-related variables. Such experiments, however, are costly to perform and are not without their own difficulties.

## EMPIRICAL MEASUREMENTS OF DEMAND ELASTICITIES

### Price Elasticities

Health care demand studies focus on price elasticity. Table 9-2 reports a selection of estimates by type of care. The dependent variable in each case is the quantity demanded. In some cases, it is a market aggregate, such as admissions per capita; in other cases, the unit of observation is the individual consumer. Most reported elasticities range between 0.0 and  $-1.0$ , indicating that consumers, while responsive to price, are not responsive to a substantial degree. Suppose that the price elasticity for physician services was between  $-0.08$  and  $-0.18$ , and physicians raised their prices by 10 percent. This would reduce consumption by 0.8 to 1.8 percent.

**TABLE 9-2 Price Elasticities from Selected Studies**

Study	Dependent Variable	Price Elasticity
All Expenditures: Manning et al. (1987)	All expenditures	$-0.17$ to $-0.22$
Physician Services:		
Newhouse and Phelps (1976)	Physician office visits	$-0.08$
Cromwell and Mitchell (1986)	Surgical services	$-0.14$ to $-0.18$
Wedig (1988)		
Health perceived excellent/good	Physician visits	$-0.35$
Health perceived fair/poor	Physician visits	$-0.16$
Hospital Services:		
Newhouse and Phelps (1976)	Hospital length of stay	$-0.06$
Manning et al. (1987)	Hospital admissions	$-0.14$ to $-0.17$
Nursing Homes:		
Chiswick (1976)	Nursing home residents per elderly population	$-0.69$ to $-2.40$
Lamberton et al. (1986)	Nursing home patient days per capita elderly	$-0.69$ to $-0.76$

**TABLE 9-3 Firm-Specific Price Elasticities**

Study	Dependent Variable	Price Elasticity
Physician Services:		
Lee and Hadley (1981)	Physician price	-2.8 to -5.1
McCarthy (1985)	Physician visits	-3.1 to -3.3
Hospital Services:		
Feldman and Dowd (1986)	Hospital patient days	-0.7 to -0.8
	Hospital admissions	-1.1
Gaynor and Vogt (2003)	Hospital discharges	-4.9
Nursing Homes:		
Mukamel and Spector (2002)	Case-mix adjusted days	-3.5 to -3.9

A further distinction among studies should be made. The price elasticities reported in Table 9-2 measure the individual consumer's or the market's response to price changes. That is, they do not relate to a particular seller, but instead represent the demand for the health care good or service in general.

The demand for physician care in general will be less elastic than the demand for the services of a particular physician. For example, suppose a medical symptom is worrisome, and the patient chooses to see a physician. The more worrisome the symptom, the less responsive he or she is likely to be to market price. Which physician to see is a completely different question. The consumer who knows the quality and price of each physician would choose the lowest-priced physician among those of equal quality. The point is that there are few substitutes for physician care, but there are many substitutes among individual physicians. Thus, firm (physician)-specific demand will be more price responsive than overall demand.

This is illustrated in Table 9-3, which reports studies of firm-specific demand elasticities. As we observe, these elasticities tend to be considerably higher in absolute value than most of the elasticities reported previously in Table 9-2.

The firm-specific elasticities have further significance. They indicate the degree of competition in the health services market. Under perfect competition, firm-specific elasticities will approach negative infinity because consumers will respond to a firm's price increase by instantly going to a competitor. The reported physician care elasticities may be large enough that competition is a reasonable approximation. In contrast, the smaller estimates for hospitals suggest considerable market power.

### Individual Income Elasticities

Economic theory suggests that increased income causes increased purchases for most goods. Most goods have positive income elasticities and are referred to as normal goods. Those with negative elasticities are referred to as inferior goods. Table 9-4 reports estimated income elasticities for a selection of studies by type of health care. In most cases, the magnitudes are small. This indicates that while health care is generally a normal good, the response is relatively small; that is, inelastic.

Income elasticities also help define when goods are necessities or luxuries. We call goods "necessities" when the income elasticity is between 0 and +1. When income elasticities exceed +1, goods are called "luxuries." From the properties of elasticities, a 1 percent rise in income increases the budget share devoted to a luxury and decreases the budget share devoted to a necessity. From Table 9-4, the results are not surprising; people commonly perceive health care to be a necessity.<sup>3</sup>

<sup>3</sup> Issues also arise regarding the appropriate definition of income, particularly with respect to short-term versus long-term (or permanent) income. Those interested should examine Goodman (1989).



**TABLE 9-4** Income Elasticities from Selected Studies

Study	Dependent Variable	Income Elasticity
All Expenditures:		
Rosett and Huang (1973)	Expenditures	0.25 to 0.45
Hospital Services:		
Newhouse and Phelps (1976)	Admissions	0.02 to 0.04
Physician Services:		
Newhouse and Phelps (1976)	Visits	0.01 to 0.04
Nursing Homes:		
Chiswick (1976)	Residents per elderly population	0.60 to 0.90

### Income Elasticities Across Countries

Given these findings, it may be surprising that studies of aggregate health expenditures across countries report substantially higher income elasticities. Often the magnitudes exceed unity. An early cross-national study was published by Newhouse (1977) who regressed per capita medical expenditures for 13 developed countries on a variable representing per capita income. From the estimated coefficient of this equation, he then calculated the implied income elasticity for various levels of income. The elasticity estimates ranged from 1.15 to 1.31. He concluded that despite within-country results showing health care to be a necessity, health care in fact is a luxury good.

Parkin and colleagues (1987) pointed out several potential weaknesses in most existing cross-national studies, but despite their objections, offered improved results that tended to support the finding of cross-national income elasticities greater than 1.0.<sup>4</sup> Gerdtam et al. (1992) and Getzen and Poullier (1992) also lend support to the result.

Is it inconsistent that within-country health care income elasticities are small, while cross-national estimates exceed 1.0? Can health care be a necessity at the individual and market levels but a luxury at the country level? A hypothetical example illustrates that income elasticity results at the national aggregate level do not necessarily apply to individual or market level. Suppose that two countries, one rich and one poor, each provided free health care to their citizens irrespective of income. Then within-country income elasticities might be small if not zero. Yet the richer country might provide greater quantities, higher technology, and better qualities of health care to each of its citizens. Thus, the cross-country income elasticities could be high.

These and related ideas are more fully developed by Getzen (2000), who shows that the individual's response to more income is different than the nation's response to more income. He also notes that symptoms of illness and pain are often more important reasons we as individuals seek out the doctor, while the available health care resources and technologies at the national level often reflect the nation's economic well-being.

The results of this small but well-established line of research have been challenged by two Canadian researchers, Blomquist and Carter (1997). By studying a large set of countries over time, observing time patterns and country-specific effects, they tentatively concluded that health spending grows about 2 percent faster than income in a manner suggesting the role of technological

<sup>4</sup> Three of these arguments are especially noteworthy. First, highly aggregated data such as national income or national health care expenditures do not necessarily imply individual behavior. Second, most studies use exchange rates to convert foreign country values to U.S. dollars, but exchange rates may not accurately reflect the purchasing power of the currency. Finally, the results are sensitive to the functional form (econometric method) used.

progress. But what about the original research goal; are the income elasticities for health spending greater than 1.0? Of 18 countries studied, they find that 11 income elasticities were either less than 1.0 or so close to 1.0 that the null hypothesis (that the elasticities equaled 1.0) could not be ruled out. Whatever direction this line of research work takes in the future, researchers are gaining increasingly sophisticated understanding of the methods and challenges of performing cross-national studies.

### Insurance Elasticities

Consumer responses to changes in insurance are important because insurance coverage has grown dramatically in the past 30 to 40 years and because we frequently must consider possible changes in social insurance. While the issue of insurance effects must be treated separately from price effects, they are closely intertwined with the issue of price elasticities.

Consider a health insurance policy where the consumer pays a fixed percentage of the bill—that is, a fixed coinsurance rate,  $r$ . In such a case, the net price that the consumer pays would be a simple multiple of the market price,  $P$ :

$$\text{Net price} = rP$$

When the market price increases by 1 percent, so does the net price; that is:

$$1.01 \times \text{Net price} = r(1.01P)$$

Under such an insurance plan, the coinsurance elasticity would be the same as the price elasticity.

However, most health insurance plans are not so simple. In practice, they include deductibles and maximum dollar expenditure (MDE) limits in addition to the coinsurance rate. The result is that the effective coinsurance rate depends in part on the size of the bill. In practice, price and coinsurance elasticities will differ somewhat.

A further difficulty arises because most studies have been performed on nonexperimental data. One of the major concerns with nonexperimental data is that the groups compared are not always randomly selected. For example, suppose that a company allows its employees to enroll in either a high-coverage plan or a low-coverage plan. Some people may choose to work for the company because it offers the high-coverage insurance plan. Others who expect to use large (low) amounts of services naturally enroll in the high- (low-) coverage plan. Still others, if sufficient insurance is not available from the company, will purchase more generous insurance privately. If the demand analysis proceeds by comparing these groups, the results may misstate the true effect of coinsurance. This is because the major decision was made in deciding which group to join. The behavior is known in economics as *adverse selection*.

The RAND Corporation, funded by the United States Public Health Service, mounted an experiment beginning in 1974. Households at six sites across the nation were randomly assigned to groups that had different levels of cost sharing, ranging from free care to care with 95 percent coinsurance and including a maximum dollar expenditure limit. The families were paid a lump-sum payment to be sure that no family was made worse off by the experiment. Because the assignment was random, adverse selection could be minimized, and the random assignment of coinsurance also allowed researchers to investigate the effects of coinsurance on expenditures.

They observed family health care use and expense experience over a period that varied from three to five years for various experimental groups. This intensive and expensive experiment improved our understanding of the response of health care consumers to economic incentives. Table 9-5 summarizes some of the key results of the RAND Health Insurance Experiment.

We note that coinsurance has a considerable effect on the level of average medical expenditures. From an extreme of a 95 percent coinsurance to the opposite extreme of free care, or zero coinsurance,

**TABLE 9-5 A Summary of the Effects of Coinsurance on Mean Annual Use of Medical Services in the RAND Health Insurance Experiment**

Plan	Likelihood of Any Use (%)	One or More Admissions (%)	Total Expenses (\$ 1991)
Free	86.8 (0.8)	10.3 (0.5)	982 (50.7)
Family Pay			
25 Percent	78.8 (1.4)	8.4 (0.6)	831 (69.2)
50 Percent	77.2 (2.3)	7.2 (0.8)	884 (189.1)
95 Percent	67.7 (1.8)	7.9 (0.6)	679 (58.7)
Individual Deductible	72.3 (1.5)	9.6 (0.6)	797 (60.3)

Source: Reprinted by permission of the publisher from *Free for All? Lessons from the RAND Health Insurance Experiment* by Joseph P. Newhouse, Cambridge, MA: Harvard University Press, Copyright © 1993 by the RAND Corporation.

the average family's medical expenses increase by nearly 50 percent, from \$679 to \$982. Even hospitalization rates are responsive, increasing from 7.9 percent of those in the 95 percent coinsurance group to 10.3 percent in the free care group, representing an increase of about 30 percent. We must conclude from the RAND experimental data and analysis that both price and insurance do matter considerably. Newhouse and colleagues (1993) conclude:

All types of service—physician visits, hospital admission, prescriptions, dental visits, and mental health service use—fell with cost sharing. There were no striking differences among these services in how their use responded to plan. Another partial exception was demand for mental health services—which, the results indicate, would have been more responsive than other services to cost sharing had there been no cap on out-of-pocket expenditure. (pp. 338–339)

The RAND study, which has proved very valuable, was not designed to track the effects of insurance on the elderly. Other researchers, however, find that the elderly consume more health care when they are more fully insured. The elderly, who are automatically eligible for Medicare hospital insurance, may opt for additional coverage. A confounding factor, one which the randomized experimental approach of the RAND study eliminates, is the possibility that those who buy the extra insurance might be those who expect to be more ill, another form of adverse selection. Hurd and McGarry (1997) separated out this confounding issue, and they conclude that the insurance effect among the elderly is due primarily to the way in which insurance changes the economic incentives that accompany illness rather than adverse selection.

Finally, although the RAND study is often considered the methodological gold standard, it was conducted over 30 years ago. Much has changed in the health economy since then, especially the growth of managed care. Meyerhoeffer and Zuvekas (2010) use comprehensive annual surveys of the U.S. civilian population over 1996–2003 to estimate more recent price elasticities for physical and mental health care. Elasticities for both services were low but, surprisingly, the price elasticity of demand for mental health visits (−0.05) was even lower than ambulatory visits for physical health problems (−0.12).

## IMPACTS OF INSURANCE ON AGGREGATE EXPENDITURES

RAND researchers, estimating coinsurance and income elasticities to be approximately 0.2, sought to calculate the demand-related portion of the post–World War II real increase in U.S. health expenditures due to the spread of health insurance. The answer was “not much”—only 10 percent of the increase.

Using the RAND income elasticity of 0.2, the post-war income increase accounted for about another 10%. Therefore, according to the RAND investigators, coinsurance and income accounted for about one-fifth of the total increase in real health expenditures. Subsequent research (Peden and Freeland, 1998) determined that about half of the expenditure increase was due to induced technological innovation. Those authors also attributed a higher impact (20%) to increased income.

## OTHER VARIABLES AFFECTING DEMAND

The studies we have reviewed often incorporate many other variables of interest in the demand function estimates, and considerable information relevant to policy issues has been obtained.

### Ethnicity and Gender

Many studies of demand examine the influence of race, and find that blacks tend to consume less medical care than the other large, self-identified ethnic groups when other factors are held constant.<sup>5</sup> Because the disparities in utilization across racial and ethnic groups have been so large and persistent over time, in 1999 Congress mandated the Agency for Healthcare Research and Quality to publish an annual disparities report. (See Box 9-1 for further discussion.) Although the majority of studies of ethnic differences in health care have focused on the experience of blacks and Hispanics, other ethnic

#### BOX 9-1

#### Disparities in Health Care: A National Priority

Disparities across racial, ethnic, and socioeconomic groups in health outcomes and health care utilization are well-documented. The *Healthy People 2010* (U.S. Department of Health and Human Services, 2000) initiative placed the elimination of disparities on the national agenda. Yet, the most recent report from the Agency for Healthcare Research and Quality (2010b) indicates that disparities remain common. The Agency uses core measures of quality (e.g., pregnant women receiving prenatal care in the first trimester) and of access (e.g., people who have a specific source of ongoing care). Overall, blacks received worse care than whites in 10 out of 20 core measures; better care in 3 core measures. Low-income people received worse care than high-income people in 15 cores measures; better care in 1 core measure. There are many other examples including those for specific and serious conditions. More disturbing, the Agency found that many disparities have not been decreasing despite the national attention and policy priority given to this problem.

Why? Is there discrimination in health care delivery against certain population groups as some have suggested? There are no easy answers but the Institute of Medicine's report to Congress on the extent and sources of the disparities (Smedley, Stith, and Nelson, 2002) greatly raised awareness of the complexity of the underlying issues. The report recognized that differences in access to care are major contributors to disparities in utilization and health outcomes, but also that there are many other confounding factors including discrimination and differences in preferences and propensities to seek care across groups.

Economists have sought to develop methods that distinguish among the various sources that account for disparities. The growing literature that addresses these issues includes Balsa, Cau, and McGuire (2007) and David and Harrington (2010).

<sup>5</sup> Extensive literature reviews are found in Mayberry, Mili, and Ofili (2000) and Weinick, Zuvekas, and Cohen (2000).

differences also have been noted, often when a given disease, for example Tay-Sachs disease, appears predominantly within one group—in this case, Jews of Eastern European origin.

Part of the differences may be of social origin; for example, blacks and other ethnic groups may be reluctant to seek care or may be treated differently by white physicians. However, ethnic differences, more broadly understood, may help explain geographic patterns across the United States. For example, westerners tend to rely less on hospitals and physicians, while residents of the more snowbound north central states rely more heavily on nursing home care for the elderly than other areas on a per capita basis.

Females differ from males most clearly in their time pattern of medical care usage. During childbearing years, women are relatively heavy users of health care, but women are healthier in the long run and they predominate in the numbers of the elderly. Thus, they are predominant among physicians' older patients. Though a great deal of public attention and concern in the past decade has turned to the science of treating diseases prominent among women, death rates for cancers are often as high or higher among men as among women. For example, mortality rates from prostate do not differ much from mortality rates from breast cancer. Myocardial infarction (heart attack) is a notorious killer of men, though women's rates have been increasing.

Thus, researchers will continue to study the differences in medical demand between the sexes, and the many differences among ethnic and cultural groups. These include not just differences among skin color groups but among the many subcultures and local cultures within these larger groups. These differences may offer the explanation for demand variations not accounted for by the usual demand variables and may yet help identify and explain many health demand questions that have remained unanswered.

Variations in sexual behavior have proved tremendously important in explaining variations in the pattern of infections with HIV and mortality due to AIDS. Homosexuals in the United States were long the major group at risk and most prominent among deaths with intravenous drug users second in numbers. Though much public effort was addressed to the prevention of a feared epidemic among heterosexuals, especially youth, the pattern has remained steady. In contrast, the world's attention is turned more toward Africa, where transmission of the disease is primarily heterosexual, and the size of the epidemic forms a crisis for world public health efforts.

### **Urban Versus Rural**

Studies sometimes find differences in health care usage due to rural status. If rural residents use less care, the reasons why are not necessarily clear. Rural dwellers may differ culturally, and some analysts argue that this factor is more important to one's perception of life than ethnicity is. Whether born to rural life or to have adopted it, it may become linked to tastes, health status, and relative reluctance to seek out a physician. The lesser health demand by westerners, already identified previously, could be understood in this view as an artifact of the predominance of rural areas in the region.

The contrasting argument is made that it is the greater travel distances required to obtain health care in rural areas rather than rural culture or tastes that account for the demand patterns. Thus, studies of geographical patterns of health care demand must take special care to measure the full price of physician or hospital care, that is, to include the travel time price.

### **Education**

Education is strongly associated with better health. If you are a college student, the odds are very good that you are healthier than your noncollege counterparts. As in the demand for health capital model, this may be because you are a more efficient producer of health, you are less likely to smoke, and you are more likely to eat a healthful diet. Or it may be that you are the sort of person with a long-term goal, and to meet that goal you have identified the need to take proper care of yourself. For the researcher sorting out such questions, the complexity of the issue is multiplied by confounding factors, especially income. Educated people tend to earn more, a fact not lost on most college students. We then must determine whether education improves one's health, or whether the income it brings affords a healthier life.

## Age, Health Status, and Uncertainty

Older people consume three to four times more health care than the younger population. Though the relationship is no surprise, Grossman's theory of this pattern makes the issue more intriguing to the research community. If we invest in our health each period and yet our health depreciates somewhat during each period, why do we necessarily buy greater quantities of health care as we get older? Very plausibly, as Grossman assumed, the depreciation rate gets larger as we age, thus greater inputs of our own time and health care are required to restore our health. This makes the correlation of health demand and age appear logical. Perhaps more interesting is the relationship of health status to the price elasticity of health care demand. Theory is not clear on this point, but it is plausible that sicker people will tend to be less sensitive to price.

Wedig (1988) finds that the price elasticity of the decision to seek health care tends to be lower in absolute value for those with poorer health status, regardless of which measure is used to record health status. However, no clear pattern over health status can be determined with respect to level of care—that is, the amount of health care consumed given that the consumer has chosen to seek health care.

Finally, uncertainty will affect health care demand. When a consumer, worried about a future health risk, seeks advice or preventive treatment, we call this a precautionary demand (Picone, Uribe, and Wilson, 1998). Elderly patients, for example, may smooth their utility over time by spending now to avoid sharp drops in well-being and mobility in the future. Some empirical evidence suggests that older people have somewhat less tolerance for risk and that one's degree of risk tolerance influences one's decision whether to buy health insurance.

## CONCLUSIONS

Demand theory is crucial to our understanding of health care markets. The substantial increases in out-of-pocket costs for prescription products experienced by many patients have affected utilization of drugs in the expected negative direction<sup>6</sup> (see Chapter 17 for specifics). Hospitals and other providers continue to compete for patients as well as for contracts with managed care organizations. The more recent estimates of price elasticities for hospitals and nursing homes shown in Table 9-3 indicate that the demand facing both types of providers is even more sensitive to price than prior studies have shown. (See Box 9-2 for an example of consumer sensitivity to the cost of nursing home care.) Time and distance can also be important as theory suggests. In a dramatic demonstration, Currie and Reagan (2005) found that each additional mile to the nearest hospital reduces the probability that central-city black children have a check-up by 3 percentage points, regardless of whether the children are privately or publicly insured.

An analysis of the demand for physician care in 12 European Union countries illustrates the universal relevance of demand theory. Jiménez-Martin and colleagues (2004) show that one-third to one-half the variability in demand across countries is explained by differences in age, income, and the physician's role in the health care system, for example, whether the general practitioner (GP) acts as a gatekeeper and whether physicians are capitated, salaried, or paid on a fee-for-service basis. In fact, the frequency of GP visits increases and the probability of contacting specialists as well as the number of visits to specialists decrease in countries where GPs are gatekeepers. Such results can help policymakers design reforms that better meet their efficiency and cost targets.

Finally, a good understanding of demand theory serves as the rationale for market-based, consumer-driven approaches to health system reform. Under health reform legislation in the United States, the future of market-based strategies remains unclear. Nevertheless, in 2009, there were nearly 6 million enrollees in consumer-directed health plans (CDHPs), with another 17 million enrolled in

---

<sup>6</sup> As Chernew and Newhouse (2008) discuss, it may be important to consider that such reductions can have undesired effects including significant reductions in use of important medications to treat chronic conditions and the health inequities created by greater price sensitivity among lower-income patients.

**BOX 9-2****Seniors Head to Mexican Nursing Homes**

For those who are not poor enough to qualify for Medicaid or do not have private insurance for long-term care, the costs of nursing homes can be overwhelming. The national average cost for a private room in a nursing home in 2006 was nearly \$78,000 per year, and nearly \$36,000 for a resident of an assisted living community. As with cosmetic surgery and other expensive procedures that are not covered by insurance, it should not be surprising that some American patients are taking advantage of much lower costs for nursing home care in other countries. An article in *USA Today* described the experiences of several who had chosen nursing facilities in Mexico. One patient paid about \$1,300 per month in 2007 for a studio apartment, all her meals, and 24-hour care. This represented about 25 percent of the average nursing home cost in her native Oregon. Another patient, who moved to Mexico from Nevada, paid about \$550 per month for similar services, and just \$140 per year for full medical coverage from the Mexican government. Although such low costs can be very enticing, prospective patients need to be aware that Mexican nursing and assisted living facilities are very loosely regulated, so standards and quality may vary widely. These facilities also typically operate out of private homes and can go out of business or change the terms at which they provide care on very short notice.

Sources: MetLife Mature Market Institute, *The MetLife Survey of Nursing Homes & Assisted Living Costs*, October 2007 ([www.metlife.com/FileAssets/MMI/MMIStudies2007NHAL.pdf](http://www.metlife.com/FileAssets/MMI/MMIStudies2007NHAL.pdf)), accessed July 14, 2008; Chris Hawley, "Seniors Head South to Mexican Nursing Homes," *USA Today*, May 16, 2007, p. 1A ([www.usatoday.com/news/nation/2007-08-15-mexnursinghome\\_N.htm](http://www.usatoday.com/news/nation/2007-08-15-mexnursinghome_N.htm)), accessed July 14, 2008.

high-deductible health plans (HDHPs). Described more fully in Chapter 23, CDHPs typically involve high deductible catastrophic insurance and other features, including a Health Savings Account, that enable consumers to take greater control over their spending decisions. By "empowering" consumers, supporters of this strategy envision a more competitive health care system in which cost-conscious patients restrain the power of providers to reduce questionable services and to keep fees low. It is still too early to be able to evaluate fully the impact of CDHPs, but their growth has created opportunities for applications of demand theory.

**Summary**

1. The theory of rational choice over health care and other goods helps explain our decisions because many health care options are not urgent, leaving room for thoughtful consideration or at least some planning.
2. In addition, physicians serve as the patient-consumer's agents and can make rational choices even in urgent situations.
3. Depicting the consumer's choice requires knowing preferences, as described by a set of indifference curves, and resource constraints, described by the budget line indicating income and market prices.
4. Consumer equilibrium occurs only if the rate at which they are willing to trade two goods, or MRS, equals the price ratio at which they are able to trade the two goods. In equilibrium, a dollar buys the same marginal utility from all goods.
5. Price elasticity,  $E_p$ , is the ratio of the percent change in quantity demanded to the percent change in price. Income elasticity,  $E_Y$ , is the percent change in quantity demanded divided by the percent change in income.
6. The time spent acquiring services constitutes a substantial portion of the economic costs. The discrepancy between the total economic prices (including time) and the money prices will be especially large for low-priced services, services with small patient copayments, and for patients with high time costs.
7. Insurance plays a major role in health services demand. Many health care purchases are at least partially covered by health insurance so that a portion is paid for by someone other than the consumer.
8. The impact of coinsurance depends critically on the price elasticity of demand for health care. If consumers do not respond to price changes in the absence of insurance, changes in coinsurance will have no impact on quantity of services demanded.

9. Coinsurance makes the demand curve for health services less responsive (less elastic) with respect to the price.
10. Quantity of services is often measured by dollar expenditures. One problem is that expenditures reflect a combination of price of care, quantity of care, and quality of care. Alternatively, quantity may be measured in numbers of visits, patient days, or cases treated.
11. It is often difficult to define prices of services since insured patients usually do not pay the full price. Moreover, the net price paid by consumers is influenced by deductibles, coinsurance, or other limits.
12. Most reported price elasticities indicate that consumers respond to price changes. However, these elasticities (between 0.0 and  $-1.0$ ) are not large compared to many other goods and services.
13. In most cases, income elasticities are low. While health care is a normal good, since its demand increases with income, the response is relatively small. However, at aggregate levels, across countries income elasticities often exceed  $+1.0$ .
14. Coinsurance has a considerable effect on the level of average medical expenditures. Both price and insurance matter.
15. Income and insurance changes since World War II may explain approximately one-fifth of the increase in U.S. health expenditures, through increased demand for services. Researchers attribute much of the remainder to increased costs brought on by technological change.

---

## Discussion Questions

1. Discuss how time costs affect health care demand, and speculate on this and possible other reasons for the lower observed per capita demand for health care in the western United States.
2. Define *price elasticity of demand*. How does an increase in the coinsurance rate affect the consumer's price elasticity?
3. Why are firm-specific demand price elasticities higher than elasticities for demand in general? Why does a high elasticity indicate a very competitive market?
4. For the following pairs of services, which of the two services would you expect to be more income elastic? More price elastic?
  - (a) Surgical services versus allergist services
  - (b) Heart surgery versus cosmetic surgery
5. It has been discovered that countries with higher per capita incomes spend more than proportionally as much on health care. What does this imply about the cross-national income elasticities? Why might this occur, even though individual income elasticities seem to be quite low?
6. The frequencies of health care visits are often used to measure service demand. Many, however, criticize the use of this variable. What are some pros and cons of the use of visits?
7. We often speak of how price rations goods. What are other rationing measures in clinics in which free care is provided?
8. Explain or show why the impact of changes in coinsurance rates on demand depends on the elasticity of demand. What sorts of health care goods or services will be responsive to changes in coinsurance rates? What sorts will tend to be relatively less responsive?
9. A profit-maximizing firm finding that its demand is inelastic will necessarily find it profitable to increase its price; therefore, its equilibrium price elasticity will necessarily be greater than 1.0 in absolute value. Are the market- and firm-specific elasticity data reported here consistent with this theory?
10. The consumer's indifference curves in Figure 9-2 indicate substitutability between visits and other goods. What will the indifference curves look like if the consumer perceives no substitutability? What will happen to the elasticity of demand in this case?
11. Some argue that wide disparities in utilization rates across racial and ethnic groups are indicative of discrimination (see Box 9-1). Use indifference curve analysis to explain why it may be difficult to distinguish between discrimination and differences in socioeconomic factors such as incomes and preferences.

---

## Exercises

1. Suppose that Martha's income is \$40,000 per year. She can spend it on health care visits, which cost \$80 per visit, or on groceries (standing for all other goods), which cost \$100 per bag of groceries. Draw Martha's budget constraint. Using indifference curves, show Martha's optimum if she buys 300 bags of groceries per year.
2. Suppose that Martha's income rises to \$42,000 per year, and that she increases her consumption of health care visits by five visits. Using the graphs for Exercise 1, draw the new equilibrium. What is her income elasticity of demand for health care visits?
3. Consider the following information on Alfred's demand for visits per year to his health clinic, if his health



insurance does not cover (100 percent coinsurance) clinic visits.

- (a) Alfred has been paying \$30 per visit. How many visits does he make per year? Draw his demand curve.
- (b) What happens to his demand curve if the insurance company institutes a 40 percent coinsurance feature (Alfred pays 40 percent of the price of each visit)? What is his new equilibrium quantity?

<i>P</i>	<i>Q</i>
5	9
10	9
15	9
20	8
25	7
30	6
35	5
40	4

4. Suppose that a consumer makes  $V_0$  physician visits each year at a price of  $P_0$ . If the price elasticity is  $-0.4$ , what will happen to the number of visits if the price increases by 10 percent? What will happen to total physician expenditures? Why?
5. If the price elasticity of demand is  $-0.5$  and the income elasticity is  $+0.3$ , then what will be the effect of a simultaneous 10 percent increase in price and a 10 percent increase in income on health expenditures?
6. Draw a diagram for hospital care that reflects the income-elasticity estimates found empirically. As income increases, what happens to the proportion of income spent on hospital care?
7. Would the opportunity cost of waiting time be higher for higher-income people or lower-income people? Given your answer, for which income group would money price tend to be a smaller portion of the full price?
8. Explain how the demand for health insurance is related to the demand for health care. Would the demand for health care then depend also on whether the person paid for the insurance or alternatively was provided the insurance at a subsidized cost?

# Asymmetric Information and Agency

- Overview of Information Issues
- Asymmetric Information
- Application of the Lemons Principle:  
Health Insurance
- The Agency Relationship
- Consumer Information, Prices, and Quality
- Conclusions

The traditional theory of demand, as we have seen, begins with the assumption that individuals are fully informed about prices, quantities, and the relationships of medical care and other inputs to their levels of health. We examined decision making within a model that assumed perfect information. Depending upon the purpose of the model, such an assumption may be justified even if it is not realistic. However, a more complete understanding of the health economy requires particular insight into the effects of various informational problems in health care markets.<sup>1</sup>

Though imperfect information has long been regarded as a feature of the health economy, insight into its specific effects was not well understood until the economics of information emerged as a distinct specialty. The development of tools to study asymmetric information and agency relationships has greatly enhanced the field of health economics. Asymmetric information encompasses situations where buyers and sellers have different levels of information; agency concerns situations where, for lack of information, buyers or sellers rely on other parties to help make decisions.

---

<sup>1</sup> The emergence of health economics as a distinct field is often traced to Kenneth J. Arrow's (1963) seminal article, "Uncertainty and the Welfare Economics of Medical Care." Arrow emphasized the role of imperfect information and uncertainty, especially the features of health care markets due to the "imperfect marketability of information." McGuire (2000) provides a comprehensive review of the relevant literature. See also a special issue of the *Journal of Health Politics, Policy and Law* (October 2001) that examines Arrow's contribution within a contemporary context.

## OVERVIEW OF INFORMATION ISSUES

The markets for many health care services and for insurance in particular are marked by significant degrees of asymmetric information and agency relationships. For example, adverse selection, a phenomenon in which insurance attracts patients who are likely to use services at a higher than average rate, results from asymmetric information. Most analysts agree that potential beneficiaries have better information than the insurer about their health status and their expected demand for health care. As a result, premiums for higher-risk patients will be underpriced, encouraging such patients to overinsure, whereas the opposite holds true for lower-risk patients. Adverse selection reduces the efficiency of health insurance markets while redistributing income from the healthy to poorer risks.

Information and agency problems account for many other important characteristics of health care markets. The possible preference for health care delivery by nonprofit hospitals and nursing homes (Chapter 13) has been attributed to patients' lack of information and inability to discern quality. For some patients, a nonprofit status might be taken as reassurance of higher quality because decisions are made independent of a profit motive. Lack of quality information also is an important motive for licensure and other regulatory measures.

The present chapter has three goals. First, we introduce information asymmetry, describe its relative prevalence, and determine its consequences, especially for insurance markets. It will quickly become clear that adverse selection in insurance is only one consequence of asymmetric information.

Asymmetric information, as when a patient is less well informed about appropriate treatments than the attending physician, typically leads to an agency relationship between the patient and provider. The second goal of this chapter is to describe the agency relationship and examine some of the problems arising in health care markets from imperfect agency. The special and controversial case of supplier-induced demand (SID) is revealed as an asymmetric information/agency problem in Chapter 15.

Finally, we seek to examine the effects of imperfect consumer information on the price and quality of health care services. Despite consumers' informational disadvantages, they often influence markets in predictable ways. Here and throughout the chapter, we identify arrangements that commonly evolve to reduce the disadvantages for the less well-informed parties. Thus, the ultimate consequences of asymmetric information and imperfect agency on the efficient functioning of markets are often not as severe as one might initially assume.

## ASYMMETRIC INFORMATION

Basic microeconomic theory usually includes an assumption that the market being analyzed exhibits perfect information. Under conditions of perfect information, all consumers and producers have complete information on all prices, as well as the quality of any good or service available in the market. Consumers also will be as well informed about the product as the seller.

Although information is never perfect in the real world, perfect information serves as a useful starting point because the properties and predictions of the standard models relying on this assumption are so well understood. Also, as students of economics are repeatedly taught, the relevant issue is whether the predictions derived from standard models apply to real-world markets. In many cases, the predictions derived from these models hold up reasonably well.

Cases do arise, however, where imperfect information does seem to matter. During the past four decades, economists have developed new insights into the effects of imperfect and asymmetric information. This section examines some of that work, including contributions by health economists to the specific problems of the health sector.

### On the Extent of Information Problems in the Health Sector

Before investigating several contributions to the economic theory of information, we begin by asking how prevalent information problems are in the health sector. It is obvious not only that information is

imperfect in health care markets, but also that information is asymmetric. Levels of information will differ among participants, such as between physicians and patients in many transactions. Often the patient is poorly informed compared to the provider about his or her condition, the treatment available, expected outcomes, and prices charged by other providers. Furthermore, we presume that the information problems that exist in the health sector are prominent enough to require the special analysis of the economics of information.

Although we can agree that information problems arise in health care markets, we must avoid the temptation to overemphasize this point. To say that information problems exist in the health sector does not mean that these problems are necessarily worse than in any other market. Markets for insurance, other professional services, automobile and appliance repairs, and many other goods and services also exhibit asymmetries. We should not necessarily conclude that information asymmetries in health care markets make it impossible for corrective institutions, practices, or products to evolve; nor do they necessarily preclude the possibility of competition.

Why? Pauly (1978) noted that at least half of physician visits customarily are made for services, such as general checkups or chronic care, for which the patient has some if not considerable experience. From data on the portion of medical expenditures attributable to ambulatory physician care, we can estimate that if half of this care is reasonably well informed, then about 8 percent of total medical care is informed. Reasoning in this manner about all sorts of medical care and products, Pauly concluded that plausibly “one-fourth or more of total personal health-care expenditures might be regarded as ‘reasonably informed’” (p. 16). By adding nursing home services and chronic conditions, Pauly (1988a) subsequently argued that this ratio is about one-third.

Further, for several medical care issues, the provider shares in the information gap with the patient. As we will emphasize when describing the “small area variations” literature in Chapter 15, the provider often is uncertain if not uninformed about the outcomes of many medical procedures. In such cases, information asymmetry does not necessarily arise even though it may be correct that the patient is ill-informed.

Finally, economic analysts of information asymmetry problems have been able to show that markets may perform well in the face of some degree of information asymmetry provided a sufficient portion of the consumers are reasonably well informed. Perhaps a majority of consumers who use personal computers are relatively poorly informed about their technical aspects and relative qualities and prices. However, a significant minority of consumers tends to be highly informed. In markets like this, the informed minority may be sufficient to provide the economic discipline it takes to make the market perform well so that the rest of us will find that the higher-priced computers also tend to be of higher quality.

We conclude this section by summarizing its main point. Certainly information gaps and asymmetries exist in the health sector. They are perhaps more serious for health care than for other goods that are important in household budgets. This makes it useful for the student of health economics to investigate the theory of information asymmetries and its application to health care. However, one should not assert that information gaps preclude the possibility of having a high degree of competition. In particular, mechanisms to deal with information gaps should not be overlooked. These mechanisms include licensure, certification, accreditation, threat of malpractice suits, the physician-patient relationship, ethical constraints, and the presence of informed consumers.

Will a state of relative consumer ignorance preclude high levels of competition? Will health care markets be characterized by a high degree of price dispersion and the provision of unnecessary care or care that is not in the patient’s best interests? Can some of the characteristics of health care markets and the evolution of their institutional arrangements be related to asymmetric information? The following sections address these and other questions by beginning with the pioneering work on asymmetric information.

### Asymmetric Information in the Used-Car Market: The Lemons Principle

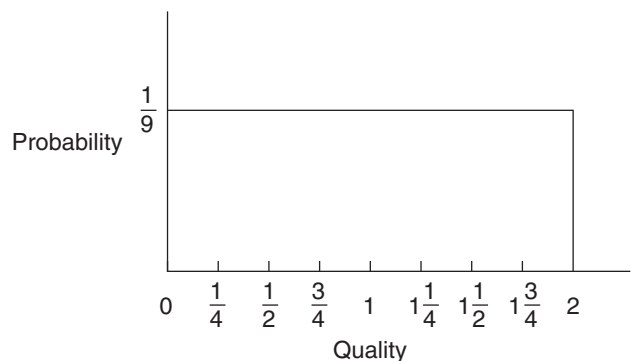
Nobel Laureate George Akerlof (1970) is often credited with introducing the idea of asymmetric information through an analysis of the used-car market. Though seemingly unrelated to health care, his classic article is worth studying for two reasons. First, it tells us much about adverse selection and the potential unraveling of health insurance markets. Adverse selection provides a key to our understanding of some major contemporary issues, such as the reasons that some may remain uninsured, or the performance of health maintenance organizations or other delivery systems. Second, Akerlof's example leads right into the issue of agency.

In Akerlof's example, used cars available for sale vary in quality, from those cars that are still in mint condition to some that are complete lemons. Information asymmetry arises if, as is plausible, the sellers know better the true quality of their cars than do the potential buyers. Akerlof devised an example that showed that cases may arise where such information asymmetry causes the market for used cars to perform poorly or even to disappear entirely.

To illustrate the nature of the used-car example, consider a somewhat simpler example, but one that retains the essential features. Suppose that nine used cars are to be sold (potentially) that vary in quality from 0, meaning a lemon, to a high of 2, meaning a mint-condition used car, a "cream puff." In fact, suppose that the nine cars have respectively quality levels ( $Q$ ) given by the cardinally measured index values of  $0, \frac{1}{4}, \frac{1}{2}, \frac{3}{4}, 1, 1\frac{1}{4}, 1\frac{1}{2}, 1\frac{3}{4},$  and 2. Under a cardinal index, a car with a value of 1 has twice the quality of a car with an index of  $\frac{1}{2}$ . The distribution of these cars is shown in Figure 10-1, where the horizontal axis shows the quality level and the vertical axis shows the uniform probability, in this case, of randomly picking a car of each given quality.

Suppose further that a car owner knows its quality level exactly but that the potential buyers know only the distribution of quality. It is known that the owners have a reserve value on their cars, so that reserve value to the seller =  $\$5,000 \times Q$ . That is, the owners would sell their cars only if they could get at least \$5,000 for every unit of quality that the car has. On the other hand, the nonowners are more eager for used cars and value them so that the value to nonowners =  $\$7,500 \times Q$ . To make this experiment a complete market, suppose that an auctioneer is hired to call out market prices; sales take place when the auctioneer finds a price that successfully equates quantity demanded with quantity supplied.

**DOES A MARKET EXIST?** Consider what would happen under asymmetric information. Suppose the auctioneer calls out an initial price of \$10,000 per car. At this price, all owners know it is worthwhile to sell their cars, so all nine cars will be supplied. However, nonowners, knowing only the distribution of quality but not the quality of each individual car, will make a best guess that a given car is of average quality; that, is  $Q = 1$ . They would not buy any cars at a price of \$10,000 because they are willing to pay only \$7,500 per unit of quality. They guess that all cars have a quality of 1, for a product of  $\$7,500 \times 1 = \$7,500$ , which is less than the \$10,000 asked. They would be willing to buy cars only if the price were less than or equal to \$7,500.



**FIGURE 10-1** The Availability of Products of Different Quality (Uniform Probability of Picking Each Car)

So the auctioneer, perhaps trying to accommodate the potential buyers, tries a lower price, say \$7,500. Unfortunately, at this price, the owners of the two best cars will withdraw from the market. Why? The owner of the car with two units of quality is receiving only  $\$7,500 \div 2$ , or \$3,750 per unit of quality; the owner of the car with  $1\frac{3}{4}$  units will act the same way. The withdrawal of the two best cars causes the average quality of the seven remaining cars to fall. With nine cars, the average of the distribution was  $Q = 1$ . Now at a price of \$7,500 per car, the best car offered will have a quality level of  $\frac{3}{2}$ , and the average quality will be  $\frac{3}{4}$ . Potential buyers would now be willing to pay only \$7,500 per unit of quality  $\times \frac{3}{4}$  unit of quality per average car, or \$5,625 for any car. Just as the previous price of \$10,000 per car was too high, the new price of \$7,500 is too high for buyers.

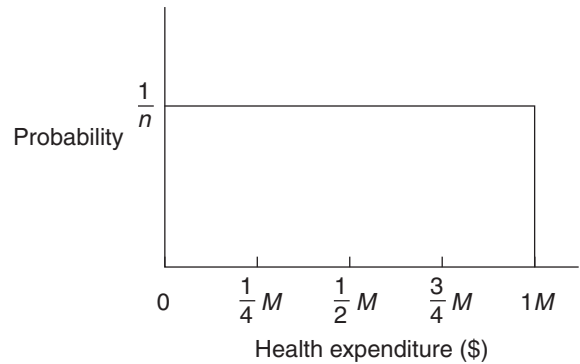
Will an equilibrium price ever be found? Surprisingly, in this example, no equilibrium that satisfies both buyers and sellers will be found. The reader can discern this by trying several successively lower prices. In the end, the cars will not be sold even though nonowners value the cars considerably more than their current owners. Akerlof saw the problem this way: When potential buyers know only the average quality of used cars, then market prices will tend to be lower than the true value of the top-quality cars. Owners of the top-quality cars will tend to withhold their cars from sale. In a sense, the good cars are driven out of the market by the lemons. Under what has become known as the Lemons Principle, the bad drives out the good until, as in some cases such as this one, no market is left.

**IMPERFECT VERSUS ASYMMETRIC INFORMATION** To see that the problem is asymmetric rather than imperfect information for both buyers and sellers, consider what would have occurred if information had been symmetric. Suppose that both owners and nonowners were uncertain of the quality, that they knew only the average quality of used cars on the market. Again, let the auctioneer start with a price of \$10,000 per car. All owners, at their best guess, may presume that their car is of average quality, and that the average will again be  $Q = 1$ . Thus, at a price of \$10,000 per car, all nine cars would be offered for sale. However, the nonowners would be willing to pay, at most, \$7,500 based on their guess that a given car is of average quality (that is, \$7,500 per unit of quality, multiplied by expected quality of  $Q = 1$ ). Again, suppose the auctioneer tries to accommodate the potential buyers by offering a lower trial price, say \$7,500. If the owners have imperfect information rather than better information, they will guess that their cars are of average quality, and thus worth (to them) \$5,000 per unit of quality, multiplied by the average quality of 1. So the owners are willing to supply nine cars at a market price of \$7,500, and the buyers are willing to purchase them at that price. The market thus exists, and clears (supply equals demand) if the information is symmetric—in this case, equally bad on both sides.

This example is extreme in several respects. The assumption of an auctioneer, the assumption that there is only one price for the used cars, the implicit assumption that the parties are not influenced by risk, and even the assumption that the quality of the cars is exogenously given could each be modified to add more realism. Since the lemons example was published, several analysts have worked on models that modify these assumptions. In some cases, this changes the result significantly. However, Akerlof's main point remains illuminating.

## APPLICATION OF THE LEMONS PRINCIPLE: HEALTH INSURANCE

Adverse selection applies to markets involving health insurance and to analyses of the relative merits of alternative health care provider arrangements. The application of the Lemons Principle to health insurance can be seen directly with the help of the previous example, a mirror image of the insurance problem. In Figure 10-2, let the horizontal axis measure the expected health expenditure levels of a population of  $n$  potentially insured people, instead of measuring the quality of used cars. Assume that they have the same demographic characteristics and that their expected health expenditure levels for the insured period range from a low of \$0 up to an expenditure level of \$ $M$ . The vertical axis represents



**FIGURE 10-2** Uniform Probability of Expenditure (Expected Health Expenditure Levels)

the probability with a uniform distribution (so that the probability of any level of spending is  $1/n$ ). The insurer must at least break even, which means that the premium (or price) received from each insured must cover the insured population's average expenditure and other expenses (including marketing and overhead).

Information asymmetry will likely occur because the potential insureds know more about their expected health expenditures in the coming period than does the insurance company. To illustrate, assume a potential insured knows his or her future expenditure exactly but that the insurance company knows only the distribution of expenditures for all insured persons.

Again, use the device of the auctioneer to illustrate the point. Suppose this time that the auctioneer attempts a first trial price of \$0! All potential beneficiaries would certainly demand coverage at this price. Just as certainly, the insurance company, expecting an average expenditure of  $\frac{1}{2}M$ , would require a premium of at least  $\frac{1}{2}M$ .

Following Akerlof's analysis, suppose the auctioneer tries a higher price, say  $\frac{1}{2}M$ , hoping that this will clear the market. In this case, all potential beneficiaries who expect an expenditure level below  $\frac{1}{2}M$  will choose to self-insure, that is, leave the insurance market altogether, because this premium is higher than their privately known levels of health expenditure. When these healthier people leave the market, the average expected expenditure level of the remaining insured persons, those with expected expenditures from  $\frac{1}{2}M$  to  $M$ , rises to  $\frac{3}{4}M$ . Thus, the higher health risks tend to drive out the lower health risk people, and a functioning market may even fail to appear at all for some otherwise-insurable health care risks.

Observe again that it is the asymmetry of information rather than the problem of incomplete information that leads to this result. If patients were no better at predicting their health expenditures in our example than the insurer, adverse selection would not take place. That is, all potential beneficiaries would have expected expenditures of  $\frac{1}{2}M$  and would be willing to purchase insurance at the premium of  $\frac{1}{2}M$ .

### Inefficiencies of Adverse Selection

This example illustrates the effects of adverse selection. Health insurance industry analysts recognize that even in its less extreme forms, adverse selection will appear. Even if functioning health insurance markets do evolve in the presence of information asymmetry of this kind, the resulting adverse selection leads to economic inefficiencies.

What are the inefficiencies? Unlike the example, few people can know exactly their future level of expenditures. Risk is the main reason for insurance. However, if the lower risks are grouped with higher risks and all pay the same premium, the lower risks face an unfavorable rate and will tend to underinsure. They sustain a welfare loss by not being able to purchase insurance at rates appropriate to their risk. Conversely, the higher risks will face a favorable premium and therefore overinsure; that is, they will insure against risks that they would not otherwise insure against. This,

too, is inefficient. In addition to inefficiency, income will be redistributed from consumers who are lower risks to those who are higher risks.

Is adverse selection merely a theoretical prediction or a serious problem? Evidence of adverse selection has been found in markets for supplemental Medicare insurance (Wolfe and Goddeeris, 1991) and individual (nongroup) insurance (Browne and Doerpinghaus, 1993). Elsewhere, Cardon and Hendel (2001) found that those who were insured spent about 50 percent more on health care than the uninsured. Although this gap appears to support the existence of adverse selection, it could also be due to two other reasons. First, by lowering the price of health care to the patient, the consumption of additional health care is encouraged, that is, moral hazard occurs. Second, the insured may have different observable characteristics, such as age, that are associated with higher spending. However, insurers could incorporate observable characteristics in setting premiums. It is the unobservable characteristics that are the source of asymmetric information, and hence adverse selection. Cardon and Hendel found only a small and statistically insignificant effect of unobservables in explaining the spending gap.

As this study suggests, if information asymmetry threatens to lead to inefficiency and even to the elimination of functioning markets in some cases, we would expect consumers, providers, and insurers to resort to other economic devices and institutions to help mitigate the problem. To illustrate, while the lemons problem in used-car markets is real, a buyer may hire a mechanic to examine the car of interest, the seller may offer a warranty, and agencies or consumer unions may arise to provide quality information. In health insurance markets, beneficiaries have often not been covered for pre-existing conditions, although this will likely change under the 2010 PPACA. Premiums for individual policies may be based on other information that insurers use to predict expenditures. They may consider factors such as age, employment status, and occupation.

## Experience Rating and Adverse Selection

Group insurance can often be a more useful mechanism to reduce adverse selection. Most employees and their families in the United States are insured through employer group plans rather than through individual policies. Group plans enable insurers to implement experience rating, a practice where premiums are based on the past experience of the group, or other risk-rating systems to project expenditures. Because employees usually have limited choices both within and among plans, they cannot fully capitalize on their information advantage.

Although experience rating can reduce adverse selection, it has come under increasing attack with the rapid growth of managed care plans such as health maintenance organizations (HMOs). HMOs receive predetermined premiums to provide the contracted health care for their enrollees. In contrast to traditional insurance where providers are independent of insurers, HMOs integrate insurance with the provision of health care.

HMOs and other managed care organizations have been promoted heavily on the belief that they have a powerful self-interest in eliminating the inappropriate care that might be recommended by providers in traditional fee-for-service systems. However, the intense competition to enroll healthy populations in managed care plans, known as “cream skimming” or “cherry picking,” has led to concerns that insurers are more interested in finding favorable groups than providing quality care. Experience rating also redistributes income toward healthy populations, and the perceived inequity of such redistributions has become a public issue.

To deal with these concerns, several states have introduced some degree of mandated community rating—a practice in which an insurer charges all groups within an area the same premium. In a less rigid form, upper and lower limits on premiums are established through rate bands. The effects of these changes have not yet been determined, but Goldman et al. (1997) predict some serious redistributive consequences of the community rating schemes considered for California. In particular, because health care spending in wealthy, urban areas is relatively high, their model predicts large regional transfers of income from poorer, rural communities to those areas.



## THE AGENCY RELATIONSHIP

An agency relationship is formed whenever a *principal* (for example, a patient) delegates decision-making authority to another party, the *agent*. In the physician-patient relationship, the patient (principal) delegates authority to the physician (agent), who in many cases also will be the provider of the recommended services. The motive behind this delegation of authority is that the principals recognize that they are relatively uninformed about the most appropriate decisions to be made and that the deficiency is best resolved by having an informed agent. Thus, asymmetric information and agency are closely related phenomena.

### Agency and Health Care

What would the perfect agent do? The perfect agent physician is one who chooses as the patients themselves would choose if only the patients possessed the information that the physician does. (See Box 10-1 for a study in which physicians and other “experts” are the patients.) This is in line with the medical code of ethics to the extent that the patient’s own interest focuses on his or her health. When any conflict arises, the perfect agent focuses on the patient’s preferences, not his or her own.

The problem for the principal is to determine and ensure that the agent is acting in the principal’s best interests. Unfortunately, the interests may diverge, and it may be difficult to introduce arrangements or contracts that eliminate conflicts of interest.

As an example, Dranove and White (1987) ask why we do not reimburse physicians on the basis of improvements in patient health. More simply, why are they not reimbursed only if the patient is cured? It would appear that such a contract would arise naturally to merge the interests of

#### BOX 10-1

#### What Happens When the Patient Is a Medical Expert?

In the United States and some other countries, observers have raised concerns over the high and increasing rates of caesarean sections (C-sections). Some have attributed this phenomenon in part to imperfect agency where the financial self-interests of providers lead to them to recommend and perform more C-sections than would otherwise take place. They have suggested that the higher rates of reimbursement for privately insured than publicly insured patients in the United States act as a principal determinant of the substantially higher C-section rates for the former group (though Grant (2009) recently estimated that other factors account for much of the gap).

What happens, however, when the expectant mothers are themselves experts: physicians, obstetricians or midwives? Grytten and colleagues (2011) examined childbirths in Norway where obstetricians receive fixed salaries and where hospitals are under tight budget controls so that unnecessary C-sections would be discouraged. The authors sought to understand the role of education, hypothesizing that less educated mothers would be the easiest to persuade to have a normal vaginal delivery when complications arise, and that expert mothers would be the most difficult to persuade.

Agency theory predicts that expert mothers would have higher C-section rates, which is the case in Norway, but this prediction is also consistent with an alternative known as “statistical discrimination.” Under the statistical discrimination model, the preferences of expert patients are more closely met because they are better communicators. Grytten’s research methodology distinguishes between these competing hypotheses by considering new technologies that help detect fetal distress and consequently reduce clinical uncertainties regarding the appropriate delivery. Under the statistical discrimination hypothesis, these new technologies should reduce disparities in C-section rates between expert and non-expert mothers. The disparities would not decrease under the agency model because the new technologies enable providers to retain their information advantages and influence over non-expert mothers. Empirical estimates of differences between expert and non-expert mothers over 1967–2005 lend greater support for statistical discrimination theory. In other words, the new techniques did reduce disparities.

both principal and patient. The authors suggest that such contracts do not exist because of the problem of asymmetric information, although in this case it is the physician who may lack information about the patient's well-being.<sup>2</sup>

To illustrate, consider patients with low back pain. Regardless of their improvement, the patients have financial incentives to understate the extent of their improvement. The provider also has an incentive to overstate the difficulty in treating the patients and in improving their health in order to increase the payment (which, let us assume, is based on the difficulty of the case). Further, it is these information problems and not other special characteristics of health care delivery that preclude payment based on the degree of improvement of the patients' conditions.

Dranove and White further apply agency theory to explain other features in the organization of health care delivery. Patients often establish a long-term relationship with a physician and pay that physician on a fee-for-service basis. As discussed earlier, such an arrangement would appear prone to lead to conflict between patient and provider. It is thus natural to ask the following questions: Why does this particular physician-patient arrangement arise and why is it so common?

Dranove and White argue that a continuous relationship between patient and physician provides the patient with increasing information with which to monitor the physician. This information places constraints on the extent to which the provider is able to deviate from an agency responsibility. Monitoring also encourages the physician to make appropriate referrals to other providers when he or she is unable to provide the services alone.

We can add to their argument by pointing out that a continuous relationship reduces the cost of transferring information about medical history, circumstances, and preferences from patient to provider. These advantages of the usual physician-patient relationship would be eroded if patients and providers were to switch to limited-period contracts under which providers are reimbursed on a different basis.

## CONSUMER INFORMATION, PRICES, AND QUALITY

We next examine the effects of imperfect information on the price and quality of medical services. Would relatively poor consumer information reduce the competitiveness of markets? Does increasing physician availability increase competition and lower prices as traditional economics suggests? What happens to quality? How do consumers obtain and use information? Several studies provide helpful insight.

### Consumer Information and Prices

Satterthwaite (1979) and Pauly and Satterthwaite (1981) introduced one of the most novel approaches to handle issues involving consumer information and competition. The authors identify primary medical care as a reputation good—a good for which consumers rely on the information provided by friends, neighbors, and others to select from the various services available in the market. Physicians are not identical and do not offer identical services. Because of this product differentiation, the market can be characterized as monopolistically competitive.

**REPUTATION GOODS** Under these conditions, the authors show that an increase in the number of providers can increase prices. The reasoning behind this surprising prediction is logical. Recall that a typical consumer relies on other consumers for information regarding their experiences with physicians. Thus, when physicians become numerous, the average number of friends who see any provider diminishes; this, in turn, diminishes the average level of information available. The consumer's responsiveness to prices and other practice characteristics depends on his or her knowledge

---

<sup>2</sup>There is actually a serious literature on "warranties" regarding maximum total charges for certain surgical procedures. De Brantes, D'Andrea, and Rosenthal (2009) describe a payment model with warranties that has improved quality while also increasing the profits of providers.

of—that is, information about—the available alternatives. Thus, this reduced information reduces the price responsiveness (i.e., the elasticity) of the firm demand curves, causing the equilibrium prices to rise. The economic idea is that reduced information tends to give each firm some additional monopoly power.

That reduced information enhances monopoly power and reduces the elasticity of the firm demand curves is consistent with standard theory. That such a situation may arise as the number of sellers increases is an unconventional and counterintuitive idea. The authors have, however, provided empirical support for their theory, and the interested reader is referred to their work for further study.

**THE ROLE OF INFORMED BUYERS** The degree to which imperfect price information contributes to monopoly power should not be overemphasized. Recall that it is not necessary for every buyer of a commodity to have perfect price information to elicit relatively competitive pricing conditions. Realistically, most consumers lack complete price information about many of the goods and services they buy (they don't know what alternative sellers are charging). Yet, despite variations in the prices of individual items among, for example, grocery stores, the average charges for a set of items across similar types of stores are likely to be similar and close to competitive pricing. A growing body of literature shows that it is sufficient to have enough buyers who are sensitive to price differentials to exert discipline over the marketplace. This will likely hold especially where the damaging threat exists of having any systematic differentials publicized by consumer organizations or the low-priced merchants themselves.

These arguments suggest that while imperfect price information will likely produce higher prices, this phenomenon may be substantially limited. In health care markets, where many services are fully or partially covered by insurance, there are added considerations. While a patient may become less sensitive to price levels and price differentials in the choice of providers, third-party payers, such as insurers, have assumed a monitoring function. Through selective contracting and other fee agreements, the actual reimbursement is often lower than the provider's charges.

**PRICE DISPERSION** The distinction between the effective transaction price and a provider's charge also obscures evidence of dispersion of fees as distinct from the average level of fees. Under conditions of imperfect consumer information, Nobel Laureate George Stigler (1961) argued that variation in prices will increase.

Building on Stigler's insight, Gaynor and Polachek (1994) developed measures of the degree of both buyer and provider ignorance by using frontier regression methods. These authors separated price dispersion into measures of incomplete buyer information, incomplete seller information, and random noise. They found that both patients and physicians exhibited incomplete information with the measure of ignorance being one and one-half times larger for patients than for physicians.

### Consumer Information and Quality

Many reports have documented high rates of medical errors and inappropriate care. For example, McGlynn and colleagues (2003) evaluated the medical records over a two-year period for a random sample of adults in 12 metropolitan areas. The study participants received only 55 percent of the recommended care overall with about the same proportions for recommended preventive care (55 percent), recommended acute care (54 percent), and care recommended for chronic conditions (56 percent).

Because consumers cannot easily monitor quality, the search for information regarding quality can be costly. At the same time, the consequences of poor-quality care can be severe or even fatal. Thus, as the previous discussion suggested, despite asymmetric information, patients rely on a variety of countervailing arrangements that are intended to reduce their search costs. These include licensure and certification, the threat of malpractice suits, codes of ethics, and various quality-assurance

**BOX 10-2****Innovative Research on Patient Perceptions of Quality**

The study of patient perceptions of health care quality is of great interest, and several recent contributions provide valuable insights. In a simple yet effective randomized experiment showing that manipulation of price can affect expectations of health benefits, Waber et al. (2008) recruited 82 paid volunteers to rate the pain caused by electric shocks applied to their wrists before and after taking a pill. Each participant was informed that the pill (actually a placebo) was a new FDA-approved analgesic that works similarly but more rapidly than codeine. Half of the participants were told that the pill had a regular price of \$2.50, and the other half that it had been discounted to \$0.10. Although the pill was reported to reduce pain for a majority of participants in both groups, the effect was much higher for the regular-price group (85 percent) compared with the discount-price group (61 percent). Thirty-nine percent of the latter group reported a pain increase compared to only 15 percent of the regular-price group.

Patients' abilities to make rational choices including those that involve quality attributes are not easily observed. These choices are especially important to strategies that involve trade-offs between regulatory and market-based reforms. If the Waber study raises concerns about the ease with which consumer quality perceptions can be manipulated, two "real world" economic analyses give comfort to the proposition that consumers respond to improved quality in expected ways.

Leonard (2008) examined whether patients in Tanzania (presumably less sophisticated than those in United States and other developed countries) can detect changes in the quality of care provided by physicians, as measured by physicians' adherence to treatment protocols. The study design takes advantage of the Hawthorne effect, which describes a temporary change to behavior in response to a change in the environmental conditions—in this case, significantly improved provider adherence through observation by a research team. Adherence slowly returned to usual levels even though the physicians continued to be observed. Leonard examined patient responses to the temporary increases in quality and found that every 1 percent increase in adherence to protocols increased the probability that a patient will be very satisfied with the doctor's quality by 0.4 percent. From the responses to the equivalent of an experimental change in quality, he concluded that patients "recognize and value quality care."

Howard (2005) examined registrations of candidates who were suitable for kidney transplants. These transplants are typically performed at major medical centers, and many factors in addition to perceived quality can influence patient decisions, for example, provider recommendation, distance to the medical center, insurance restrictions, and quality. Using graft failure as a measure of quality, Howard found that a one standard deviation increase in the one-year graft failure rate, a clear indicator of poor quality, was associated with a 6 percent reduction in patient registrations at a center.

schemes that are either mandated or voluntary. The Internet is also becoming a major source of information despite concerns about the accuracy of online information. Box 10-2 provides some recent examples of how consumer perceptions of quality are formed.

At this point, we pursue the consumer's direct role through the Dranove and White argument that the physician–patient relationship enables patients to monitor providers and encourages physicians to make appropriate referrals. To the extent that many specialists rely on referred patients, these specialists would seem to have incentives to maintain quality. Are they also rewarded with higher prices for higher-quality services? Theory suggests that if consumers have the ability to distinguish between quality levels, then the demand for higher-quality providers and thus price should be greater than for lower-quality providers. Haas-Wilson (1990) examined this proposition using data from the psychotherapy services market. She investigated whether the prevalence of referrals from informed sources affects the price of social workers' psychotherapy services. Informed sources include other health providers and other professionals such as school counselors and clergy.

Regression analysis of a sample of social workers' fees indicated that fees are positively and significantly affected by the percentage of clients who were obtained through informed referrals. The evidence shows that patients rely on informed sources (agents) for information and that higher quality, as measured by informed referrals, is rewarded by higher fees.

Other mechanisms can also help reduce the problems created by asymmetric information. As previously discussed, the lack of a clear profit motive may make nonprofit organizations more attractive to patients when they cannot easily observe or determine quality. Although this argument seems plausible, empirical support for it had been lacking until Chou (2002) developed a novel application to quality of care in nursing homes (where nearly two-thirds are for-profit). Because many nursing home patients are too cognitively or physically incapacitated to be able to monitor and evaluate their care, Chou used the absence of visits by a spouse or children (the patient's representatives) within one month of admission as an indicator of information asymmetry. There were no significant differences between for-profits and nonprofits when asymmetric information was not present. With asymmetric information, a very different picture emerged. For-profits had higher mortality rates as well as higher rates of decubitus ulcers, dehydration, and urinary tract infection. Chou concluded that for-profits "have more incentive to compromise on those aspects of quality of care which are hard to monitor."

### Other Quality Indicators

With the increasing dominance of HMOs and other managed care organizations, the availability of information to help consumers select among plans, and to monitor how consumers respond to the information, have become major issues. The National Committee for Quality Assurance (NCQA), a private accreditation body for HMOs, issues report cards based on about 50 standardized measures of a plan's performance (such as childhood immunization rates, breast cancer screening, and asthma inpatient admission rates). *Newsweek*, *U.S. News & World Report*, and various consumer groups also regularly rate HMOs. A key assumption behind these efforts is that information about quality will, like price information, help discipline providers through patient choices. Low-quality HMOs will presumably not survive, or at least they will not be able to charge high-quality prices.

Initial evidence on the intended effects of plan performance ratings brings the report card strategy into question. Tumlinson et al. (1997) found that independent plan ratings are relatively unimportant to consumer choices. Only 17 percent of survey respondents indicated that such ratings are essential, compared with 72 percent for specific plan benefits and 62 percent for out-of-pocket costs. Chernew and Scanlon (1998), employing multivariate statistical methods on consumer choice of plans, confirm that "employees do not appear to respond strongly to plan performance measures, even when the labeling and dissemination were intended to facilitate their use" (p. 19).

In subsequent work, Scanlon and colleagues (2002) examined a flexible benefits system introduced by General Motors in 1996 and 1997 under which employees and retirees received a fixed amount of dollars that could be spent across a variety of fringe benefits categories. Excesses in spending over the allotted "flex dollars" are paid out of pocket. GM developed ratings (e.g., superior performance, below expected performance) for six performance categories (e.g., preventive care, access, patient satisfaction) to help those who wanted to select an HMO choose among the available plans (typically two to six depending on the employees' geographic area). As expected, the study found that higher out-of-pocket prices imply lower enrollments. Also as expected, employees tended to avoid plans with many below average ratings. However, plans with many above-average ratings were not much more successful in attracting enrollees relative to plans with many average ratings.

Together with other results on plan switching (Beaulieu, 2002), we can conclude that the provision of quality information does influence consumers, particularly when the quality ratings are negative. Nevertheless, it is clear that much more work is needed regarding quality indicators and patients' perceptions of quality. Box 10-3 provides examples of some recent and innovative work on this subject.

**BOX 10-3****Do Report Cards Matter? More Recent Evidence**

Two recent contributions for very different medical interventions provide new evidence on report cards and how they affect patient decisions. Wang and colleagues (2011) examined coronary artery bypass graft surgery (CABG) in Pennsylvania, a state that has published report cards on CABG providers since 1992. Beginning in 1998, the report card information for both hospitals and individual surgeons (e.g., number of cases, mortality and readmission rates) were made available on the Pennsylvania Health Care Cost Containment Web page (<http://www.phc4.org/cabg/>). The Wang study considers both demand side effects (e.g., decrease in demand for poor performance surgeons) and supply-side effects (e.g., “dumping” or avoiding high risk patients) of report card information. At the surgeon level, volume decreased for poor or unrated surgeons indicating a patient avoidance effect. Surprisingly, though there was no increase for high performing surgeons, and also no effects on hospital volume of any rating.

Unlike most CABG procedures, fertility treatment involves completely elective procedures that are not often covered by insurance. Does information on fertility clinics’ success rates affect clinic choice? You bet it does! Bundorf et al (2009) compare the three-year lagged birth rates of fertility clinics before and after report card information became publicly available in 1995. The authors found that the differential effect of the pre- and post-reporting for clinics with higher birth rates increased, thus supporting the hypothesis that consumers respond to quality information.

**CONCLUSIONS**

There is little doubt that information gaps, asymmetric information, and agency problems are prevalent in provider–patient transactions. However, for some health care services, the problems are not necessarily greater or larger than those for other goods. Patients are likely to be relatively poorly informed about treatment for conditions that they have not previously experienced and about care involving newer technologies. The informational asymmetries and reliance on provider-agents are likely to be most pronounced in these situations.

Although there is a potential lack of competition, even wide information gaps do not necessarily lead to market failure. Leaving aside the role of licensure and regulation, arrangements have evolved to help patients or their insurers to monitor the quality and prices of providers. Furthermore, higher-quality producers are generally rewarded by greater demand and higher prices. The use of referrals, accreditation, and other arrangements reduce the provider’s ability to raise prices above those charged by others and to sell low-quality services at high-quality prices. Nevertheless, improved quality remains an elusive national goal. Despite efforts to provide quality information and many private and government initiatives to improve quality, a series of influential reports released over a decade ago by the Institute of Medicine (1999, 2001) suggested that as many as 98,000 deaths and a million excess injuries annually in the United States can be attributed to problems with quality and safety. By calling for a fundamental overhaul of the U.S. health care system, the Institute raised troubling questions about existing safeguards as well as patients’ perceptions of quality and their ability to monitor it.

In the decade since those publications, the Agency for Healthcare Research and Quality (AHRQ), which evaluates quality through an annual report covering 200 measures of quality, safety and effectiveness, found that “quality is improving, but the pace is slow, especially for preventive care and chronic disease management” (AHRQ 2010a, p.2). Its conclusions echo an earlier assessment by the Commonwealth Fund (2008) which warned that the “U.S. health system is on the wrong track.” From a National Scorecard of 37 indicators of health system performance based on comparisons with top performing states, regions, health plans, and other nations, the Fund concluded (p. 13) that the United States “is losing ground in providing access to care and has uneven health care quality. Average U.S. performance would have to improve by 50 percent across multiple indicators to reach benchmark levels of performance.” The challenge is to develop and implement system-wide reforms that can narrow the gap between processes that work and what is actually done.

## Summary

1. Health care markets tend to be characterized by both imperfect information and asymmetric information. Asymmetric information describes a situation in which those on one side of a transaction have better information than those on the other side.
2. Often, providers are relatively well informed (e.g., about the patient's illness and possible treatments). In other cases, buyers are relatively well informed (e.g., the purchaser of insurance knows more about his or her health status and pertinent habits than the insurer does).
3. The extent of consumer information problems should not be exaggerated. Consumers are reasonably well informed on about one-fourth to one-third of their health care spending.
4. One possible consequence of asymmetric information is that a market will not exist. Even if it exists, a general reduction may occur in the quality of goods available (the "Lemons Principle").
5. The Lemons Principle appears as the problem of adverse selection in health insurance and other health care markets.
6. Adverse selection results from asymmetric information, not equally imperfect information. Adverse selection in insurance results in inefficiencies through higher-risk consumers overinsuring, relative to the amounts they would purchase at actuarially fair rates, and lower risks correspondingly underinsuring.
7. An agency relationship tends to be formed when a party (principal) delegates decision making to another party (agent). The problem for the principal is to develop a contract or relationship to ensure that the agent is acting in the principal's best interests.
8. Various agency relationships have evolved to mitigate the problems associated with asymmetric information between patient and provider. These include the continuous physician-patient relationship and the health maintenance organization.
9. Other constraints, such as licensure and accreditation, codes of ethics, and the threat of litigation, limit the ability of providers to deviate from their agency responsibilities.
10. Many health care services are reputation goods. In markets for reputation goods, an increase in the number of providers can lead to an increase in monopoly power and higher prices.
11. The existence of informed buyers helps exert discipline over the market by limiting price increases and price differentials among sellers.
12. Though challenged by some, the proposition that higher quality tends to be rewarded with higher price is supported in economic studies. Patients also respond to quality indicators in selecting a hospital. However, they rely only modestly on objective plan ratings in their selection of HMOs. Negative ratings have a greater impact on consumer decisions than positive ratings.
13. Improving quality has become a national priority. Many studies have found high levels of inappropriate care and medical errors.

## Discussion Questions

1. The market for higher education is another example where a high degree of information asymmetry is likely. What mechanisms have evolved to help students in their choice of schools and classes within schools? Do you have confidence that higher-priced institutions provide higher-quality education?
2. The situation in which an individual is interviewing for a job also exhibits information asymmetry. Explain why. How does the relatively poorly informed party deal with this?
3. The use of professional and independent buyer-agents to help individuals purchase automobiles or houses is becoming a more common phenomenon. Given the conflict of interest facing the physician-agent, why do we not see greater use of a buyer-agent who is retained by the patient?
4. The used-car market has publications that provide information on the quality and prices of used cars. Are similar avenues of information available to health consumers? What kind of information do they provide? Is it more or less effective than the information available on used cars? How would you, as a patient, find information about a provider's quality or prices? How would you assess the confidence you have in that information?
5. What is a reputation good? What are examples of reputation goods outside the health care sector? Show what Pauly and Satterthwaite predict will happen to the demand curve for health services as a result of an increase in the number of providers.
6. Stigler argued that the variation in fees increases as buyer information decreases. Suppose you observe that each seller

- in a market is charging the identical price. What potentially conflicting inferences can you draw?
7. Why don't physicians guarantee their work as do many auto repair shops?
  8. Various commentators have suggested that only 15 to 20 percent of all health care services have been subject to rigorous, controlled investigation, that is, care based on what is commonly called "evidence-based medicine." Assume that this statement is correct. What are some implications for efficiency of health care delivery?
  9. Is it possible to have a situation where higher costs, as measured by the resources used to provide care, do not produce higher quality?
  10. According to clinical research, nearly one-half the care provided in the United States falls short of recommended treatment protocols. Discuss how imperfect and asymmetric information contribute to this phenomenon. How can health plans or markets be reformed to reduce the quality gap?

---

## Exercises

1. Suppose that in the Akerlof example, there are only eight cars ranging in quality from  $\frac{1}{4}$  to 2 (i.e., there is no complete lemon). Hence, the mean quality level is 1.125. Determine whether the market disappears completely, and, if not, how many cars will be sold.
2. Consider the agency relationship in malpractice cases under a contingency fee system. The plaintiff (party that sues) typically pays his or her attorney about one-third of any monetary damages that are awarded (and nothing if the case is lost). Supporters of this system claim that client and attorney share a common goal of maximizing the award. Is there, however, an inherent conflict between attorney and client in the amount of attorney time and other resources that are devoted to the case?
3. Give three examples of asymmetric information in which the health consumer has information that is unavailable to the health provider. Give three concrete examples in which the health provider has information that is unavailable to the health consumer.
4. In the Akerlof example, the individuals are treated as indifferent to risk. What would you expect to see in these markets if individuals wanted to avoid risk? What if there were some "risk lovers"?



# The Organization of Health Insurance Markets

- Loading Costs and the Behavior of Insurance Firms
- Employer Provision of Health Insurance: Who Pays?
- Employer-Based Health Insurance and Labor Supply
- The Market for Insurance
- The Uninsured: An Analytical Framework
- Technological Change, Higher Costs, and Inflation
- Conclusions

Chapter 8 introduced the concept of insurance, an arrangement that allows risk-averse people to reduce or eliminate the risks they face, with a primary focus on health insurance. Consumers buy insurance to replace the uncertainty of a large loss or major expenditure with the more certain prospect of regular premiums. In most countries, profit-seeking firms supply various types of insurance, although the provision of health insurance varies from country to country. In an idealized market, the insurance premium (as a percentage of the potential loss) will approach the probability of the event occurring.

In previous chapters, we concentrated on the impact of insurance on individuals. In this chapter, we focus on the insurance market and the behaviors of firms within that market. Within the context of the employer-provided health insurance common in the United States, we establish who pays for health insurance. We continue with an examination of employer-provided insurance and job mobility. We then look at the traditional community-rated health insurance (where individuals or groups all pay the same premium) and show how that market has changed, and we follow with an analysis of the uninsured. We finish the chapter by examining linkages among insurance, technological change, and health care cost inflation.

## LOADING COSTS AND THE BEHAVIOR OF INSURANCE FIRMS

Consumers can improve their well-being by sacrificing a (relatively) small but certain premium to insure against the probability of a considerably larger loss. It is important now to demonstrate how within competitive markets the policies will be offered to specific groups and why, in fact, some groups will find it difficult to get insurance at all.

We have referred to the model of a competitive industry, in which the firms will compete to where economic profits become zero, or normal. With higher (lower) profits, firms will enter (leave) the market. Only

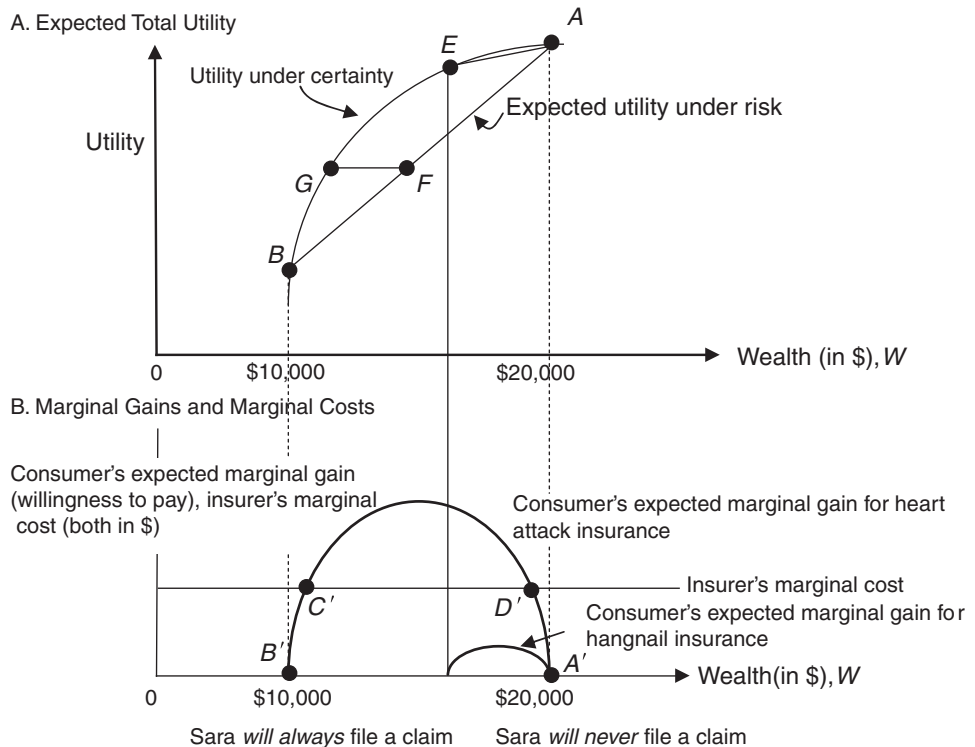
when profits are zero, or normal, will entry and exit cease. In this model, the insurance carriers collect money during the year and pay some of it out. In good years, carriers pay out less than collected; in bad years, they pay out more. Economic analysis suggests that the good (and bad) years will be random. Systematically good (bad) years suggest excess profits (losses), and the probability of entry into (exit from) the industry by other firms.

We also have previously shown how moral hazard can lead firms to offer certain types of coverage and not others. In particular, firms have often shown themselves to be reluctant to cover conditions accompanied by price-elastic demands for services.

### Impacts of Loading Costs

Insurance firms incur costs of doing business that are added to the claims payouts. These loading costs are largely related to the numbers and types of customers and claims processed. Even in perfect competition, these costs must be passed on to consumers, or else the insurers will not be able to cover all costs and will be forced to leave the market. The incidence of these costs suggests that insurers will shy away from covering events that are almost certain to occur, or those that seldom occur.

Consider consumers who behave as though they have a utility of wealth ( $W$ ) function exhibiting diminishing marginal utility of wealth. Figure 11-1A relates total utility to total wealth and Figure 11-1B looks at corresponding marginal gains and marginal costs related to various actions. In Figure 11-1A, Sara has \$20,000 in wealth yielding utility at point  $A$ , with various possibilities of losses up to \$10,000, or point  $B$ . The amount Sara would be willing to pay over the actuarially fair amount (also interpreted as Sara's consumer surplus) is shown by the horizontal distance between the expected utility line and the (curved) utility function, measured in dollars. For example, at point  $F$ , this horizontal distance is  $FG$ . On inspection we note that the horizontal distance between the



**FIGURE 11-1** Impacts of Loading Costs on Availability of Insurance

expected utility line and the utility function is zero if the event never occurs (i.e., if we are at point *A*). It increases up to some point as we move in a southwest direction (with increased probability of illness) and then decreases to zero, as the illness becomes more certain, toward point *B*.

Because insurance is taken against risk, as the probability of the uncertain event approaches either 0 or 1, insurance becomes less desirable. Near point *A*, the expected loss—that is, the probability of the event—multiplied by the loss if the event occurs, is not large enough for Sara to bother to insure. This is noted as point *A* in Figure 11-1B, the lower diagram, where dollars replace units of utility on the vertical axis. Going back to Figure 11-1A, at point *B*—because the event is almost certain—Sara might as well set the money aside (self-insure) and avoid the trouble of dealing with the insurer. The corresponding point on Figure 11-1B is *B'*.

### Insurance for Heart Attacks and Hangnails

In comparing types of losses, for any probability of illness, the larger the expected loss, the larger the gain from the insurance. We see this in Figure 11-1A by comparing the distances between the expected utility line and the utility curve for a small loss (line segment *EA*) and for a large loss (line segment *BA*). Segment *EA* shows a small distance; segment *BA*, a larger one. Hence, if Sara has equal probabilities of a hangnail (small loss) and a heart attack (large loss), the expected gain to her from insurance for heart attack coverage will exceed the gain for hangnail coverage.

Consider now the insurers' decisions in providing insurance. If the event is almost certain, the insurers' costs of administering the policy may exceed the benefits to the consumers.

In Figure 11-1B from *B'* to *C'*, it will not pay to insure claims because for the firms to earn profits they must charge marginal costs, which here exceed the expected consumers' marginal benefits. Between points *C'* and *D'* expected marginal benefits exceed marginal costs. To the right of point *D'*, again the marginal costs exceed the expected marginal benefits, and no insurance will be provided. As the diagram is drawn, no firm could afford to offer hangnail coverage.

### Loading Costs and the Uninsured

The analysis of loading costs provides one avenue for addressing the problem of those who cannot get insurance. Health insurance in the United States has been largely available through participation in the labor market. Those who do not participate in the labor market, and many of those who are employed by small businesses, self-employed, or sporadically employed have found it difficult to get insurance.

Many explanations have been proposed, but it is apparent that the per-person costs of processing information and claims of those individuals who are outside larger organizations (either companies or unions) are higher. This results in an increase in the firms' marginal costs relative to the consumer's marginal benefits and can reduce or eliminate the range of services that may be offered.

The analysis also helps address the impacts of entry and exit in the insurance market. More efficient processing and information handling presumably will lower the premiums that must be paid by customers in the market. If we look again at Figure 11-1B, we recognize that improved information handling and processing would not only lead to lower marginal costs and hence lower prices, but also would permit firms to offer services (based on probability of occurrence) that had not previously been offered.

Consider points *C'* or *D'*, where the expected marginal benefit was previously just equal to (or possibly just below) the marginal cost. An insurer who lowers costs can offer coverage for types of events that previously were uncovered. Conversely, increased costs, due either to market forces or to mandated coverage, would force firms to cut back coverage on events for which they could not (due to limited consumer surplus) pass along the increased costs on to the customers.

### EMPLOYER PROVISION OF HEALTH INSURANCE: WHO PAYS?

The largest segment of the American population acquires health insurance through the workplace, and this began almost by accident in the 1940s. During World War II a booming economy coupled with wartime shortages left consumer goods in short supply, so the federal government imposed

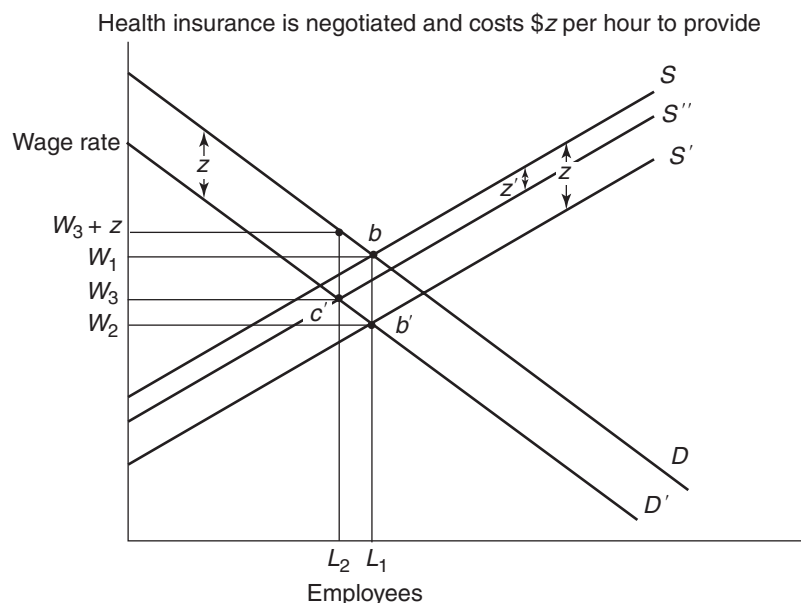
wage and price controls as anti-inflationary devices. Predictably, employers had to devise new ways to attract workers because wage controls in a full-employment economy prevented companies from raiding workers from one another. Fringe benefits were not legally considered as part of the wage package, but they could provide flexibility in worker compensation, and to improve the allocation of workers among sectors of the economy. One of these fringe benefits was health insurance.

Economists start their analyses by looking at the labor market. We assume that a lower market money wage rate leads an employer to hire more workers for two reasons: (1) the employer can substitute labor for more expensive equipment or resources; and (2) the employer can sell more products at lower prices, hence requiring more workers. Assume at the outset there is no health insurance benefit, and that the market wage is \$20 per hour. Employers will hire workers as long as the incremental (marginal) revenue from the goods those workers produce exceeds the \$20 per hour wage. To begin, assume that the employer employs 1,000 workers, at an equilibrium money wage of \$20 per hour.

Suppose that workers negotiate a health insurance benefit worth \$1 per hour to them, and costing exactly \$1 for the employer to provide. The employer, who was previously willing to pay a wage of \$20, will now be willing to pay \$20 less the \$1 cost. Other points on the employer's demand schedule, showing the number of workers it would hire at different wages, will also change by the \$1 cost of the benefit.

The workers of course would prefer the \$1 benefit on top of the previous \$20 wage, but if they were previously willing to accept a wage of \$20, they will now be willing to supply their labor for \$1 less, because they value the benefit at \$1. In the resulting equilibrium, the net wage or total compensation (money wage + the value of the benefit) remains unchanged at \$20, but the equilibrium money wage falls to \$19, or by exactly the amount of the benefit. Workers accept lower money wages, and the same 1,000 workers are employed at the same net wage, \$19 in money wages plus the \$1 benefit. The workers are no worse off at a wage of \$19 with the health insurance than at \$20 without the health insurance because the insurance is worth the \$1 that it cost in reduced wages. The employer earns no less profit for providing the health benefit.

For a more detailed analysis, consider (following Lee, 1996) a labor market with a typically downward-sloping demand for labor,  $D$ , and a typically upward-sloping supply of labor,  $S$ , as noted in Figure 11-2. The demand for labor is related to the marginal productivity of workers. The supply



**FIGURE 11-2** Interaction of Health Insurance and Employee Wage

of workers reflects the wage in this industry relative to the wage in other industries. Workers will choose to work in this industry as long as the wage they can earn exceeds their opportunities in other jobs. In Figure 11-2, at equilibrium point  $b$ , the equilibrium wage is  $W_1$  and the equilibrium quantity of labor demanded and supplied is  $L_1$ . Suppose that workers in the market negotiate a health insurance benefit worth  $\$z$ /hour at that margin, and it costs employers exactly  $\$z$ /hour to provide. What happens? Employers who were previously willing to pay  $W_1$  per hour for workers will now pay  $W_1$  less  $\$z$ . Other points on the demand curve will shift downward in a similar manner so the demand curve will shift downward by exactly  $\$z$  to  $D$ . What will happen to the supply curve? Because the workers were willing to supply various amounts of labor at various wage rates according to the supply curve before, now that they are receiving a benefit worth  $\$z$  they will offer their labor for  $\$z$  less. Hence, the supply curve will shift downward by exactly  $\$z$  to  $S$ .

What is the result? Similarly to our earlier discussion, the net wage remains the same at  $W_1$ , but the money wage falls by  $\$z$ . The equilibrium wage has fallen to  $W_2$  at point  $b$  or by exactly the amount of the benefit. Workers have taken their benefits in lower money wages, and the same number of workers,  $L_1$ , is employed at the same net wage. For a real-world example of who pays, see Box 11-1.

There are several reasons that the marginal benefits of the insurance to the employees may fall short of the employers' marginal costs. Some contracts negotiate subsidized coverage for prescription drugs, at a cost to the employer. However, some employees are healthy and do not use prescription drugs. This benefit has no value to them.

In addition, recall from Chapter 8 that for many types of health care, fractional coinsurance lowers the consumers' marginal costs of treatment and leads them to buy more insured care than otherwise. As a result the benefits on average may be worth less to the workers than what they cost the employers to provide. Without moral hazard, prescription drug coverage would simply reduce the cost of drugs to the workers. However, the drug benefit might induce workers to purchase prescription shampoo or prescription cold medicine rather than less expensive over-the-counter brands.

If the average benefit is worth  $\$z$ /hour, or less to the workers than the  $\$z$ /hour that it costs to provide, then the new supply of labor curve,  $S''$ , will have fallen by less than the demand for labor

### BOX 11-1

#### "It Never Goes Down"

Figure 11-2 discusses the interaction of wage rates and benefit packages in abstract terms, but the "real world" shows these effects to be valid. In September 2010, reporter Phil Galewitz wrote of J.W. Cheatham LLC, a construction company in West Palm Beach, Florida. Until 2008 it paid full cost of employees' health insurance premiums. In 2008, the firm began charging employees about \$26 per week, or roughly \$1,300 per year. Due to the extra expense, 10 of the company's 96 employees chose to drop the coverage and were now uninsured.

In 2010, the company initially faced a 9 percent increase in health premium costs, but it negotiated that down to 4 percent after agreeing to bundle the purchase of health and life insurance with its insurance carrier. In addition, the company eliminated paid vacation and holidays for employees. Those cost-cutting measures helped Cheatham maintain the premium costs for workers and employees at \$26.80 per week. Although figures were not given, eliminating ten days of paid vacation, and the six major federal holidays, when evaluated at even the minimum wage of \$7.25 per hour, cost the workers \$928, or approximately 6 percent. This is almost certainly an underestimate of the reduction in compensation.

The co-owner and personnel director at Cheatham was already worrying about how much higher insurance costs will go to next year. "It never goes down."

Source: Galewitz, Phil, "Rising Health Care Costs: Workers Pay Bigger Share," *The Fiscal Times*, September 2, 2010, <http://scauth.thefiscaltimes.com/Issues/Health-Care/2010/09/02/Workers-Pay-More-for-Health-Care.aspx>, accessed May 9, 2011.

(still  $D$ , reflecting what it costs to provide the benefit). Equilibrium will be at  $c$ , rather than  $b$ , the money wage will be  $W_3$ , and the total wage will be  $(W_3 + z)$  rather than  $W_1$ . Employers will react to the higher gross wages  $(W_3 + z)$  by reducing employment, here from  $L_1$  to  $L_2$ .

### Spousal Coverage: Who Pays?

Working members of the same family often have coverage from several sources. Using the logic from the previous section, an analysis derived from Mark Pauly (1997) helps to examine the issue of spousal insurance coverage. What happens if the husband has family insurance coverage where he works, and the wife chooses not to take coverage where she works? Who pays in this case? The subtleties of the analysis that occur through the labor market may surprise some readers.

To simplify, suppose that employees in a town can work in either the Alpha or Beta sector. Alpha employers employ only married men; half of the spouses do not work, and half of the spouses work in the Beta sector. Half of the Beta employees are spouses of the men in the Alpha sector and half are single. At the outset, assume that no health insurance is provided and the “pure” wage ( $W_1$  from Figure 11-2) for each employee in both the Alpha and Beta sectors is \$80,000 per year. Assume that whatever health insurance is implemented, all employees of each firm are given the same take-home pay regardless of insurance cost.

Suppose now that the Alpha firms (employing only married men) offer to buy family coverage for their employees worth \$8,000 per year. The Beta firms (half spouses; half single) offer to pay \$4,000 per year per person for their employees, as long as those who elect coverage pay an additional \$20 per month, or \$240 per year. In this situation, all Alpha workers will choose family coverage; it is a better buy because the employee is covering himself, his spouse, and his children for \$8,000. As a result of market processes similar to those in Figure 11-2, money wages for Alpha workers fall by \$8,000 to \$72,000. In the Beta sector, wages per worker will fall by \$2,000 to \$78,000 (because half of the workers use coverage that costs \$4,000 per person). Who pays?

- *Two-worker* families covered through Alpha firms pay \$10,000 for \$8,000 in coverage. This occurs because wages have fallen by \$8,000 in the Alpha sector (where the men work) and by \$2,000 where their spouses work.
- *Single-worker* families covered through Alpha firms pay \$8,000 for \$8,000 in coverage, again because wages have fallen by \$8,000 in the Alpha sector.
- Single-worker families covered through Beta firms pay \$2,240 (reduced wages of \$2,000 + \$240 from the monthly payments) for \$4,240 (\$4,000 + \$240 from the monthly payments) in coverage.

Are the Beta firms better off because they don’t pay for the health care for half of their employees? No, because they are still paying a net wage of \$80,000 for the labor services that they use. Are the Alpha firms worse off because they are paying benefits for people who don’t work for them? No, they too are still paying a net wage of \$80,000 for the labor services that they use. Single workers at the Beta firms benefit because they receive \$2,000 more in coverage than they are paying for—a transfer from married households, which receive \$2,000 less in coverage than they are paying for!

### How the Tax System Influences Health Insurance Demand

One of the most important factors in the increased demand for health insurance since World War II has been the tax treatment of health insurance. Suppose Sara earns \$1,000 per week and would like to buy health insurance. Ignoring state and local taxes, assume that she is in the 25 percent marginal tax bracket, so her take-home pay is \$750 per week. Suppose further that health insurance would cost her \$60 per week. Her net take-home pay would then be the take-home pay of \$750 less the health insurance of \$60, or \$690 per week.

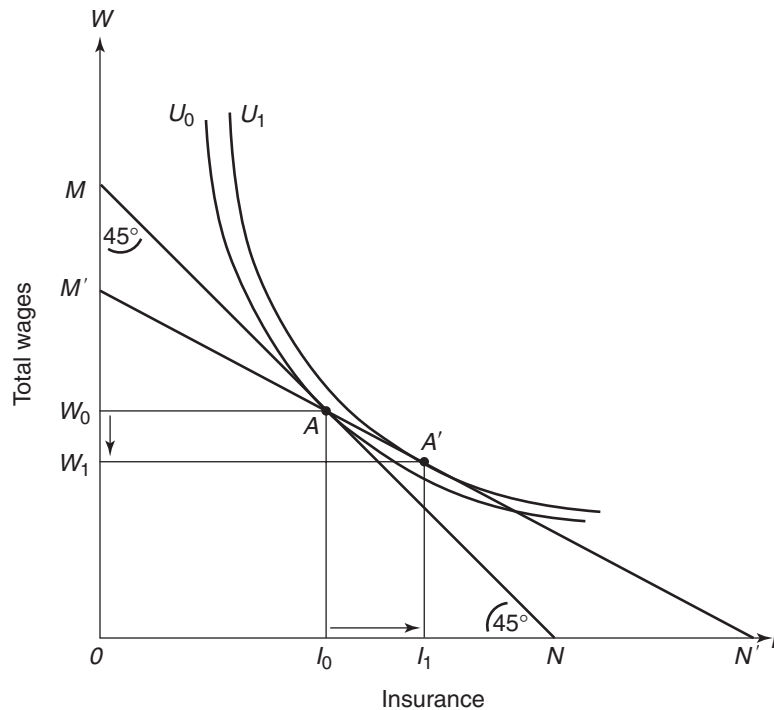
Suppose instead that Sara’s employer purchases insurance for her, again at a price of \$60 per week. This fringe benefit is exempt from income taxation, as it has been since World War II.

Although Sara’s total compensation is still \$1,000 per week, she is taxed only on the wage portion, or \$940. Her take-home pay will now be 75 percent of \$940, or \$705 per week. Her \$15 improvement in well-being occurs because she does not pay \$15 in tax on the \$60 insurance benefit. The \$705 in net compensation with insurance is clearly superior to the \$690 net take-home pay without insurance.

If marginal tax rates increase, consumers have incentives to increase employer health expenditures. Employers also benefit from this arrangement because their levels of Social Security taxes will fall. In 2011 employees pay 4.20 percent for the Social Security portion, and employers pay 6.20 percent.<sup>1</sup> Both groups pay 1.45 percent for the Medicare portion. Because insurance is an expense to the employer rather than a wage (on which Social Security and Medicare taxes must be paid), it is exempt from Social Security and Medicare taxes. In the 1950s, federal marginal income tax rates went as high as 91 percent, and even today many people pay marginal (federal plus state) tax rates of 40 percent or more.

The allocative problem within the economy occurs because health expenditures have been singled out for special treatment. Consider Figure 11-3. This figure shows an entire wage package consisting of the sum of total wages,  $W$ , and total insurance,  $I$ . Intercept  $M$  on the  $y$ -axis shows the amount of wages if no insurance is in the package. Similarly, intercept  $N$  on the  $x$ -axis shows the amount of insurance in the unlikely case that Sara received her entire compensation as insurance benefits. If no special tax treatment is given, then line  $MN$  has a 45-degree relationship to the  $x$ - and  $y$ -axes, a slope of  $-1.0$ . In other words, \$1 of insurance trades for \$1 in wages, and the initial allocation is at point  $A$ , with wages  $W_0$  and insurance  $I_0$ .

The subsidy of health insurance through the government policies lowers the price of \$1 of insurance relative to \$1 of wage remuneration. Suppose that the employees even recognize the



**FIGURE 11-3** Impacts of Preferential Treatment of Employee Insurance

<sup>1</sup> Without further changes in the law, the employee tax rates will return to 6.2 percent in 2012.

subsidy and are prepared to give up some wages for an insurance subsidy. The “give back” rotates the  $x$ -intercept down to  $M'$ , but the subsidy causes the budget constraint line to rotate to point  $N'$  on the  $x$ -axis. Hence, \$1 of wages actually buys (trades off for) more than \$1 of insurance. Figure 11-3 shows that without special tax treatment, Sara consumes package  $A$  of  $I_0$  and  $W_0$ . Even if the “give back” left Sara unchanged at point  $A$ , the changed relative prices will now make it more attractive to move to a more insurance-rich package. Thus, the tax system leads Sara to choose combination  $A'$ , with more insurance at the expense of lower wages, and increased utility at  $U_1$ . Not only will Sara buy more insurance, but the tax subsidy may encourage her to insure for the kinds of low- or high-probability events (e.g., routine dental care) that might otherwise be left uninsured.

### Who Pays the Compensating Differentials?—Empirical Tests

The compensating differentials, with respect to wages and insurance, merit serious empirical investigation. Morrissey (2001a) notes that many empirical studies have associated health insurance with higher, rather than lower, wages. Because compensation (wages plus insurance) is based on productivity, employers spend considerable effort identifying workers who are better motivated, more dependable, more highly skilled, and better able to interact with clients and customers. Many researchers have had only age or schooling measures to capture productivity and other attributes of more or less productive workers have been unobserved. Because more productive workers get both higher wages and more health insurance, the substitution between wages and insurance may be swamped by the productivity effect. Despite these problems, several researchers have developed creative tests to identify the wage–insurance trade-off.

Gruber and Krueger (1992) examine workers' compensation insurance, and Gruber (1994) looks at mandated maternity benefits coverage. Both studies confirm the existence of “group specific” average wage adjustments. That is, those groups that were paid more in benefits received lower wages.

Jensen and Morrissey (2001) use 1994 and 1998 data from the Health and Retirement Survey (HRS) to examine the wage-coverage tradeoff for workers born between 1931 and 1941. The HRS provides information on wages, health insurance coverage, demographic characteristics, and health status measures for workers and their spouses and the authors used the health status of spouses as an indicator of the demand for health insurance.

They find evidence of compensating differentials for older workers. Other things equal, those workers with health insurance had annual wages about \$6,300 lower than those without. Since annual family health insurance coverage provided through an employer often costs \$6,000 to \$7,000 per year, the evidence suggests that workers do pay for their health insurance through lower wages.

Bhattacharya and Bundorf (2005) look for differentials in the context of obesity. They find that the incremental health care costs associated with obesity are passed on to obese workers with employer-sponsored health insurance in the form of lower cash wages. In their study obese workers in firms *with* employer-sponsored insurance received lower wages, while those *without* employer-sponsored coverage, who had individual coverage, or no coverage, did not.

Adams (2007) examines the impacts of the 1993 New York imposition of pure community rating on firms in the small group market. Community rating prevents carriers from charging different premiums based on age. If smaller firms had been cutting wages of older workers prior to pure community rating to offset their higher health care costs, then one would have expected the reform to lead to higher relative wages for older workers at these firms. The reform did increase the relative wages for older workers, both in relation to older workers in other states and in relation to older workers at large firms within the state.

Emanuel and Fuchs (2008) sum up the tradeoff between wages and premiums as “not a point merely of economic theory but of historical fact.” Since the late 1970s insurance premiums increased by 300 percent (a factor of 4) after adjustment for inflation. Corporate profits per employee flourished, with inflation-adjusted increases of 150 percent before taxes and 200 percent after taxes. In contrast, average hourly earnings of workers in private nonagricultural industries were *stagnant*,



actually decreasing by 4 percent after adjustment for inflation. Rather than coming out of corporate profits, the increasing cost of health care resulted in relatively flat real wages for 30 years.

### Other Impacts of Employer Provision of Health Insurance

Employer provision of health insurance has other impacts as well. Because the employer is a large, single buyer of coverage, the purchase of insurance through the employer provides scale economies of dealing with insurance providers that single purchasers could never enjoy. This tends to lower the effective price of coverage to the employees.

In addition, group purchase by employers addresses the problem of adverse selection in the provision of insurance. Recall that in Chapter 8 we considered a club whose members participated in an insurance arrangement. The arrangement worked well because the contract provided a necessary service to the members. In particular, the probability of a claim was a random event that could be calculated, and that was independent of the actions of the members.

Central to this result is the proposition that the probability of usage is independent of the insurance plan. Suppose that Karen smokes cigarettes and knows that her probability of a claim is not the 5 percent assumed by the insurance company, but rather 10 percent. If able to convince an insurer that she indeed belonged to the less risky (5 percent) category, Karen would be able to buy insurance much cheaper than the actuarially fair premium. Karen would get a bargain; the insurer would lose money. The inability to identify probabilities, and hence their impacts on the insurance market, is often referred to as *adverse selection*.

As an example, consider an insurance plan that offers major hospitalization coverage. Consider also that many heavy smokers may recognize their higher probabilities of lip, throat, or lung cancer and heart disease. If they can prevent their insurers from finding out about their smoking, then they can purchase much cheaper insurance than the appropriate premium, given their prior conditions.

It can be argued that the purchase of insurance by employers minimizes adverse selection by providing a more appropriate pool for the fixing of insurance rates. Box 11-2 shows the advantages of group insurance, relative to individual policies. These advantages accrue because most groups contain a broad mix of risks, by virtue of having been formed for some purpose *other* than insurance.

#### BOX 11-2

#### After Caesareans, Some Women Can't Get Insurance—"Man, Is That a Scary Thing!"

Group coverage pools insurance risks, especially when those insured joined the group for reasons other than getting insurance. We have also seen how the insurers' experiences with the groups allow them to charge lower rates. In the individual insurance market, however, the absence of explicit pooling, and the lack of group-related experience, may prevent some people from getting insurance at all.

In June 2008, Denise Grady of the *New York Times* reported the story of Peggy Robertson, a 39-year-old woman from Centennial, Colorado. Ms. Robertson had been shopping around for individual health insurance, the kind that people buy on their own. She already had insurance but was looking for a better rate and was turned down because she had given birth by caesarean section. Having the operation once increases the odds that it will be performed again, and if she became pregnant and needed another caesarean, Golden Rule did not want to pay for it. A letter from the company explained that if she had been sterilized after the caesarean, or if she were over 40 and had given birth two or more years before applying, she might have qualified. After being rejected by Golden Rule, she kept her existing coverage.

Insurers' rules on prior caesareans vary by company and also by state, since the states regulate insurers. Some companies ignore the surgery, but others treat it like a pre-existing condition. "Sometimes the coverage will come with a rider saying that coverage for a caesarean delivery is excluded for a period of time," said Susan Pisano of America's Health Insurance Plans, a trade group.

Golden Rule, which sold individual policies in 30 states, wrote a letter saying it would insure a woman who had had a caesarean only if it could exclude paying for another one for three years. However, Colorado law considered such exclusions discriminatory and forbade them, so Golden Rule was *simply rejecting* women who had had the surgery, unless they had been sterilized or met the company's age requirements. Colorado residents who were denied individual health insurance could obtain it through a state program, but the premiums were high, at 140 percent of standard rates. A spokeswoman added that some women had enrolled specifically because prior caesareans had disqualified them from private insurance.

Ms. Robertson said that had she known a caesarean was grounds for rejection, she would not have even applied to Golden Rule. Insurers routinely ask applicants if they have ever been denied, and deny anyone who says yes. "My understanding is that once you're denied it's hard to get other insurance," Ms. Robertson said. "Man, is that a scary thing."

*Source:* Based on information from Grady, Denise, "After Caesareans, Some See Higher Insurance Cost," *New York Times*, June 1, 2008, p. 1. [www.nytimes.com/2008/06/01/health/01insure.html?\\_r=1&hp&oref=slogin](http://www.nytimes.com/2008/06/01/health/01insure.html?_r=1&hp&oref=slogin), accessed June 1, 2008.

## EMPLOYER-BASED HEALTH INSURANCE AND LABOR SUPPLY

Because most private health insurance is obtained through employment and is typically not portable to different employers, researchers have sought to determine the extent to which health insurance may affect labor supply. The two major impacts relate to retirement age and job mobility.

### Health Insurance and Retirement

Aging workers face a dilemma. Gruber and Madrian (2002) show that compared with those age 35 to 44, those age 55 to 64 are:

- twice as likely to report themselves in fair health and four times as likely to report themselves in poor health,
- seven times as likely to have had a heart attack and five times as likely to have heart disease, and
- 40 percent more likely to have a prescribed medicine (with twice as many medicines if receiving a prescription).

As a result, their medical spending is almost twice as large and twice as variable as the younger group.

While declining health makes retirement more attractive, it also makes employer-provided insurance more attractive, especially for those younger than 65 years of age, at which time Medicare will provide insurance. Thus, individuals face an incentive to postpone retirement until they are eligible for Medicare at age 65.

Researchers have generally focused on the impact of retiree health insurance on retirement behavior. Gruber and Madrian summarize 16 studies and report that the availability of retiree health insurance raises the odds of retirement by between 30 and 80 percent.

### Health Insurance and Mobility

Health insurance may also affect worker mobility between jobs. Employees may fear losing coverage for pre-existing conditions, which are generally defined as any medical problem that has been treated or diagnosed within the past six months to two years. This job lock may have several economic effects:

- Less productive workers may stay at jobs for insurance reasons only, leading to decreased economic output because they would not be replaced by more productive workers.

- Even if all workers are equally productive, some workers may stay in jobs for fear of losing the health insurance benefits to the exclusion of those who would otherwise fill the jobs.
- Those who do change jobs may be denied coverage, face higher premiums, or only obtain insurance subject to a waiver that excludes coverage of their health condition.

Both Cooper and Monheit (1993) and Madrian (1994) address the issue. We look more closely at Madrian's presentation.

Madrian creates a simple matrix of the probability of job mobility to consider the impact of job lock. Because job lock is caused by the potential loss of health insurance coverage with changing jobs, one would not expect those with coverage through both their own employment and an outside job to face job lock.

<b>The Probability of Changing Jobs</b>		
	<b>Employer-Provided Health Insurance</b>	
	<b>No</b>	<b>Yes</b>
No other health insurance	<i>a</i>	<i>b</i>
Other health insurance	<i>c</i>	<i>d</i>

She tests for the magnitude of job lock by examining whether those workers with employer-provided health insurance and other coverage are more likely to change jobs than those without alternative coverage, or:

$$(\text{cell } d \text{ probability}) - (\text{cell } b \text{ probability}) > 0$$

However, if a man is in cell *d*, it may be due to the insurance provided by his wife, who may be providing income as well; all else being equal, the additional income could lead to increased mobility. Hence, Madrian derives a second test: Whether having other health insurance increases mobility more for those who have employment-based insurance than for those who do not, or:

$$(d - b) - (c - a) > 0$$

This test is referred to as “difference-in-difference,” that is, the difference between  $(d - b)$  and  $(c - a)$ .

Her most general model looks at the probability of turnover of married men. Inserting predicted job turnover probabilities into Madrian's matrix, the raw estimate indicates that the mobility rate under job lock (cell *b*) is 0.085 or 26 percent lower than cell *d*, which shows a mobility rate of 0.115. The difference-in-difference estimates, attempting to account for any independent effect of other health insurance on mobility, give an alternative estimate of 31.1 percent.

Subsequent research has tended to support Madrian's results. Sanz de Galdeano (2006) uses data from the Survey of Income and Program Participation (SIPP) 1996–2000 panel. Her study suggests that employer-provided health insurance adversely affects job mobility for all population subgroups by about 31 to 58 percent. She also evaluates the impact of the Health Insurance Portability and Accountability Act of 1996 on job mobility, and finds evidence that is contrary to the intended objectives of the 1996 HIPAA. That is, the 1996 HIPAA failed to remedy the insurance-induced job lock in labor markets.

Rashad and Sarpong (2008) provide a good review of the literature. They find that individuals with employer-provided health insurance stay on the job 16% longer and are 60% less likely to voluntarily leave their jobs than those with insurance that is not provided by their employers.

Job lock thus appears as an unintended consequence of employment-related coverage in the United States. It could be addressed through changes that are broadly consistent with prudent insurance practices. These include elimination of pre-existing condition clauses and the development of

health insurance pooling mechanisms in local labor markets that might promote continuity of coverage across employers. Job lock also must be considered as a consequence of mandated approaches to health insurance benefits, both at state and national levels.

## THE MARKET FOR INSURANCE

Having discussed the provision of insurance in theory, we now consider some institutional features of the health insurance market. Such a discussion must address the roles of the Blue Cross insurers, which were originally nonprofit firms, and the commercial insurers, which were typically for-profit firms. With increased competition in the health care sector, many distinctions have blurred. Nonetheless, to understand the current insurance market, it is helpful to consider how it has developed over time.

### The Market for Private Insurance

The number of those privately insured in the United States burgeoned after World War II. From a base of 12 million insured in 1940, it increased by a factor of more than six by 1950 (to 76.6 million), and doubled again (to 158.8 million) by 1970. By 1980, over 187 million US residents had private health insurance.

Well into the 1970s, most of the coverage was provided either by insurance companies (usually in group settings) or by Blue Cross and Blue Shield plans. Since the late 1970s, use of other plans has increased, with declines in both the shares and the absolute numbers of those covered by both the insurance companies and the “Blues.” This reflects the movement toward self-insurance by large firms, as well as toward various arrangements through health maintenance organizations (HMOs), preferred provider organizations (PPOs), point-of-service (POS) providers, and other forms of managed care.

Table 11-1 provides more recent health insurance coverage status, as well as type of coverage. The largest portion of the population has private coverage, and the largest portion of that comes through the workplace. In 2009, 194.5 million people had private coverage, with over 87 percent

**TABLE 11-1 Health Insurance Coverage Status and Type of Coverage, 1990–2009 (in thousands)**

Year	Total People	Total Insured Private or Gov't	Private			Government				Not Covered
			Total	Employer Based	Direct Purchase	Total Insured	Medicaid	Medicare	Military	
1990	248,886	214,167	182,135	150,215		60,965	24,261	32,260	9,922	34,719
1995	264,314	223,733	185,881	161,453		69,776	31,877	34,655	9,375	40,582
2000	279,517	241,091	202,794	179,436	26,799	69,037	29,533	37,740	9,099	38,426
2001	282,082	242,322	201,695	178,261	26,309	71,295	31,601	38,043	9,552	39,760
2002	285,933	243,914	200,891	177,095	26,846	73,624	33,246	38,448	10,063	42,019
2003	288,280	244,876	199,871	175,844	26,783	76,755	35,647	39,456	9,979	43,404
2004	291,166	247,669	200,924	176,247	27,551	79,486	37,955	39,703	10,789	43,498
2005	293,834	249,020	201,167	176,924	27,055	80,213	38,104	40,177	11,166	44,815
2006	296,824	249,829	201,690	177,152	27,066	80,270	38,281	40,343	10,547	46,995
2007	299,106	253,449	201,991	177,446	26,673	83,031	39,554	41,375	10,955	45,657
2008	301,483	255,143	200,992	176,332	26,777	87,411	42,641	43,029	11,560	46,340
2009	304,280	253,606	194,545	169,689	27,219	93,167	47,758	43,440	12,412	50,674

Source: 1990 and 1995 data from *Current Population Survey*, March 1988–2005; 1999 data forward from <http://www.census.gov/hhes/www/hlthins/data/historical/files/hihist1.xls>

(169.7 million) getting the coverage from the workplace. Over 93 million had health insurance through the government—some (particularly those with Medicare) were dually covered by both the private market and the government. Based on these government estimates, the number of uninsured increased from 38.4 million in 2000 to over 50 million in 2009.

The period between 2007 to 2009 represents a major departure from longer term trends. During the “Great Recession,” the number of Americans with employer-provided insurance fell by 7.8 million, and the number with government-provided insurance rose by over 10 million, with an over 8.2 million increase in those receiving Medicaid. The number of uninsured jumped from 45.7 million to 50.7 million. Almost certainly, the passage of the 2010 Patient Protection and Affordable Care Act stemmed in large part from perceived problems in insurance coverage, as well as concerns about health costs. We discuss this in considerable detail in Chapter 23.

### Insurance Practices

At least two organizational and practice issues characterize the health insurance industry and link it to the emerging issues of managed health care. The first issue is a conflict between insurers and the insured (most often represented by the health care providers) regarding the amounts of claims, and indeed whether the claims should be paid at all. From the earliest instances of health insurance, providers, most particularly physicians, have argued that their judgments must not be questioned on cost grounds. Insurers, in contrast, have felt that they could increase their profits and reduce premiums by judiciously questioning treatments and costs.

Why would providers consent to having someone second-guess their decisions? Goldberg and Greenberg (1977) traced the growth of health insurance in Oregon in the 1930s. At that time, physicians shared in the Great Depression with the larger population, and they saw acceptance of health insurance, even with its accompanying oversight, as a way of increasing earnings. Although insurance plans were attractive to physicians when times were difficult, groups such as the Oregon State Medical Society threatened to expel physicians who participated in the plans and to establish their own plans. The plans established by the physicians tended to be less strict in their cost reviews.

Blue Cross and Blue Shield plans were started by medical providers, with Blue Cross providing hospital payment and Blue Shield providing physician payment. The Blues typically offered more complete and more comprehensive coverage than other insurers, and they paid participating providers directly. They were also usually generous in the payment of hospital care. The development and success of large hospitals in the post–World War II United States was at least in part supported by the generosity of Blue Cross and Blue Shield reimbursement of hospital stays on per-diem bases. The fiscal distress facing many large hospitals in the last 25 years may be traced to changes in financing that have accompanied the reduced power of Blue Cross and Blue Shield leadership in the provision of health insurance, in part related to the strictures placed by Diagnosis Related Group (DRG) controls. Medicare introduced DRGs to provide prospective fixed payments for specific diagnoses. These payments induced cost-containment measures that generally resulted in reduced hospital stays.

The second organizational pattern that has characterized the industry is the change in “rating” clients. We have noted that insurers pool their clienteles to determine risk premiums based on their experiences with the groups. Blue Cross plans began with a method called *community rating* in which all subscribers in a given location, irrespective of age or health experience, were charged the same premium. This contrasts with experience, or risk, rating, in which the insurer charges group premiums (to a company or a fraternal or service organization) based on its experience with the group.

Community rating provides a fundamental information problem. Low-risk clients are overcharged, and their premiums are transferred to higher-risk clients in the same pool. Advocates of community rating argue that this allows high-risk and low-income clients to buy insurance that would otherwise have been unavailable.

However, insurers that can identify groups with low risk may offer members of these groups lower-cost insurance. Many large firms choose alternative carriers who, again, will charge lower premiums than

the community-rated Blue plans. The large firms may in fact choose to self-insure. This “cream skimming” practice may leave the plans that continue to community-rate their clienteles with risk pools that are so risky that they require almost prohibitively high premiums. Clearly, the extent to which community rating can survive is related to the amount of competition in the health insurance market.

### The Past 30 Years

The shift toward managed health care through HMOs and PPOs from about 1980 onward induced a change in philosophy among Blue Cross and Blue Shield insurance plans. The typical Blue Cross and Blue Shield system faced stiff competition from smaller organizations that were peeling off supposedly more desirable clients from the community-rated system leading to higher costs for those who remained in the system.

The major explicit response to the changing marketplace was the 1994 approval of a change in organizational status. Blue Cross plans are now either for-profit firms or establish for-profit subsidiaries.

Why change? One fundamental reason involved the ability to raise capital. Nonprofit organizations must generate funds through their revenues, whereas for-profit firms may sell stock. With more competition in the marketplace and renewed emphases on lowering premiums and costs of care, the option to sell stock became more attractive.

In a wide-ranging interview (Iglehart, 1995), Blue Cross of California chief executive Leonard Schaeffer offered three reasons for the changes in the Blue Cross system:

1. Most Blues plans were big and had not reacted quickly to marketplace changes.
2. The Blues had been successful for a long period of time, and traditional operations got “embedded.” This led to a resistance to change.
3. The national association consisted of an “unwieldy” national association with 69 independent plans.

At the time, Schaeffer envisioned that there would be “30, perhaps even 20” associations, organized on a regional basis, but this estimate proved pessimistic, as the Blues have maintained a substantive presence in the majority of the states.

According to Robinson (2006), from 1980 through 2000, the industry experienced extensive creation of new health plans, sponsored by medical groups, hospital systems, employers, labor unions, consumer cooperatives, and other entities and that the traditional Blue Cross and indemnity insurers lost substantial market share to the new entrants. Over time, however, the “best-managed firms,” some of which were upstart health maintenance organizations (HMOs) and others of which were renovated Blue Cross and commercial indemnity carriers, demonstrated their ability to grow market share and to absorb weaker competitors. By 2006 the industry comprised four national plans (United, WellPoint, Aetna, and CIGNA); state-specific Blue Cross and Blue Shield plans; a few regional for-profit plans (such as Humana, HealthNet, and Coventry); and, in some markets, regional nonprofit plans (such as Kaiser Permanente, Tufts Health Plan, and HealthPartners).

A 2010 American Medical Association (AMA) report indicates that as of 2008, Blue Cross Blue Shield (BCBS) plans still had the largest market share in 24 states, with over 50 percent of the market in 16 states. Individual states had even higher shares with Alabama showing 93% BCBS market control, Hawaii having 75% and Illinois 72%.

Metropolitan Statistical Areas (MSAs) correspond to the large cities and their environs. At the MSA level, the market shares of health insurers were even higher than at the state level. According to the AMA report, in 2008

- In 96 percent (344 of the 359 MSAs) at least one insurer had a combined HMO+PPO market share of 30% or greater.
- In 48 percent, at least one insurer had a combined HMO+PPO share of 50%.
- In 18 percent (63 of the MSAs), one insurer had a combined HMO+PPO market share of 70% or higher.

These measures indicate considerable monopoly power (to set prices and to earn potentially excess economic profits) in large portions of the US market.

Austin and Hungerford (2010) note that Blue Cross plans were originally designed to avoid competition by requiring exclusive territories and barring plans linked to specific hospitals. They surmise that those requirements may have been aimed at supporting community rating policies and broadly based risk pools, benefitting many consumers. As commercial insurers and managed care strategies rose in prominence, market forces along with merger and acquisition strategies have helped reshape the health insurance market. Some insurers may have engineered mergers and acquisitions to enhance market power and the success of that strategy depends on underlying factors that determine the structure of the market. We address these competitive and potential anti-trust issues in Chapter 20.

## THE UNINSURED: AN ANALYTICAL FRAMEWORK

With the high costs of health care and the inevitability to most people of incurring at least some expenses, it has become crucial for individuals to have access to health insurance. In the United States, health insurance availability has been linked to the workplace. Yet various surveys have shown that over 50 million Americans have no health insurance at any moment in time (see Box 11-3), that the number has soared recently, and that a large fraction of these people are employed. 2011 numbers from the U.S. Census (2011), using the Current Population Survey, showed that 16.6 percent of Americans had no public or private health insurance in 2009. A total of 50.7 million were uninsured, up from 45.7 million in 2006.

Table 11-2 shows that the workplace does not provide insurance for all families, and that coverages vary widely by economic circumstance. Among the many statistical breakdowns available in this table, in part A we see 28.8 percent of those aged 25–34 were uninsured in 2009, compared to 23.8 percent in 2002. In the 35-to-44 range, 21.8 percent were uninsured in 2009, compared to 17.8 percent in 2002.

### BOX 11-3

#### Counting the Uninsured

Many might be surprised to learn that estimates of the uninsured come from surveys of the population, rather than by censuses, in which all are counted. Most estimates of the uninsured come from the National Health Interview Survey (NHIS), the March supplement to the Current Population Survey used by the Census Bureau (CPS), the Medical Expenditure Panel Survey (MEPS), and the Survey of Income and Program Participation (SIPP).

The CPS identifies individuals as uninsured if they have lacked coverage for the entire previous calendar year (although many analysts believe that respondents incorrectly provide information about their current insurance status). The SIPP identifies individuals who are uninsured for each month of a four-month reference period. The NHIS identifies individuals as uninsured if they lacked coverage in the month prior to the survey. The MEPS data count as uninsured those without coverage for the entire interview round (an average of three to five months).

Insurance definitions vary. MEPS defines private insurance as coverage for hospital and physician services, thereby eliminating serious and dread disease, workers compensation, accident, and disability policies from counting as coverage. As well, MEPS does not include single service plans (such as dental plans) as private insurance except for single service hospital coverage. CPS and SIPP instruct interviewers not to count single service plans as private insurance, but researchers believe that some single service coverage may get misreported as comprehensive coverage.

In short, estimates of the uninsured may differ according to the source. Estimates from different sources generally move up and down together, but they are not likely to be identical.

*Source:* U.S. Department of Health and Human Services, Understanding Estimates of the Uninsured: Putting the Differences in Context, <http://aspe.hhs.gov/health/reports/05/uninsured-understanding-ib/index.htm>, accessed May 4, 2011.

**TABLE 11-2 Health Insurance Coverage of the Nonelderly, 2009**

	Nonelderly (millions)	Percent Distribution by Coverage Type				
		Private		Public		Uninsured
		Employer	Individual	Medicaid	Other	
<b>Total - Nonelderly</b>	<b>264.7</b>	<b>56.5%</b>	<b>5.2%</b>	<b>16.7%</b>	<b>2.7%</b>	<b>18.9%</b>
<b>A. Age</b>						
<b>Children—Total</b>	<b>79.3</b>	<b>50.9%</b>	<b>4.0%</b>	<b>33.2%</b>	<b>1.4%</b>	<b>10.4%</b>
<b>Adults—Total</b>	<b>185.4</b>	<b>59.0%</b>	<b>5.7%</b>	<b>9.6%</b>	<b>3.2%</b>	<b>22.5%</b>
Adults 19–25	29.1	41.0%	10.8%	13.5%	1.6%	33.0%
Adults 26–34	36.4	54.7%	4.5%	10.4%	1.6%	28.8%
Adults 35–44	40.2	63.3%	4.1%	8.9%	1.9%	21.8%
Adults 45–54	44.3	66.1%	4.7%	8.0%	3.3%	17.8%
Adults 55–64	35.4	64.2%	5.8%	8.5%	7.6%	13.9%
<b>B. Annual Family Income</b>						
<\$20,000	66.2	14.7%	6.2%	37.2%	4.1%	37.8%
\$20,000–\$39,999	52.1	43.4%	5.1%	21.7%	3.4%	26.3%
\$40,000+	146.5	80.1%	4.8%	5.6%	1.8%	7.7%
<b>C. Family Poverty Level<sup>c</sup></b>						
≤138%	<b>74.1</b>	<b>15.3%</b>	<b>5.5%</b>	<b>40.8%</b>	<b>3.5%</b>	<b>34.9%</b>
<100%	55.6	12.3%	5.5%	43.4%	3.1%	35.7%
100–138%	18.5	24.2%	5.3%	33.0%	4.8%	32.8%
<b>139–399%</b>	<b>101.9</b>	<b>60.8%</b>	<b>5.5%</b>	<b>11.6%</b>	<b>3.1%</b>	<b>19.0%</b>
139–250%	51.0	47.9%	6.0%	17.1%	3.7%	25.3%
251–399%	50.9	73.8%	5.0%	6.1%	2.4%	12.7%
<b>400%+</b>	<b>88.7</b>	<b>86.1%</b>	<b>4.6%</b>	<b>2.4%</b>	<b>1.6%</b>	<b>5.4%</b>
<b>D. Household Type</b>						
Single Adults Living Alone	19.5	55.4%	7.8%	10.6%	5.2%	20.9%
Single Adults Living Together	33.5	40.2%	8.5%	10.7%	3.2%	37.3%
Married Adults	54.5	68.2%	5.4%	5.7%	4.3%	16.3%
1 Parent with children	33.4	33.1%	4.1%	41.5%	1.4%	19.9%
2 Parents with children	108.6	66.4%	4.3%	15.3%	1.6%	12.4%
Multigenerational/Other with children	15.2	33.3%	2.9%	31.8%	2.9%	29.0%
<b>E. Family Work Status</b>						
2 Full-time	68.2	81.9%	3.3%	5.9%	1.2%	7.7%
1 Full-time	135.4	60.9%	5.1%	13.5%	1.8%	18.8%
Only Part-time	24.1	25.8%	10.7%	27.5%	2.7%	33.4%
Non-Workers	37.1	14.0%	5.6%	41.2%	8.7%	30.5%
<b>F. Race/Ethnicity</b>						
White only (non-Hispanic)	166.0	65.3%	6.4%	11.5%	2.8%	14.1%
Black only (non-Hispanic)	33.5	43.1%	2.8%	27.8%	3.8%	22.6%
Hispanic	46.0	35.2%	2.5%	26.8%	1.6%	33.9%
Asian/S. Pacific Islander only	12.9	59.4%	7.0%	13.3%	2.0%	18.3%

(continued)



**TABLE 11-2 (continued)**

	Nonelderly (millions)	Percent Distribution by Coverage Type				
		Private		Public		Uninsured
		Employer	Individual	Medicaid	Other	
<b>Total - Nonelderly</b>	<b>264.7</b>	<b>56.5%</b>	<b>5.2%</b>	<b>16.7%</b>	<b>2.7%</b>	<b>18.9%</b>
Am. Indian/Alaska Native	1.8	38.4%	—	29.1%	2.5%	28.4%
Two or More Races	4.4	51.0%	4.5%	28.2%	4.0%	12.3%
<b>G. Citizenship</b>						
U.S. citizen - native	231.8	58.5%	5.3%	17.2%	2.8%	16.1%
U.S. citizen - naturalized	12.8	58.4%	5.5%	10.6%	2.5%	23.0%
Non-U.S. citizen, resident for < 5 years	4.1	30.8%	4.6%	18.8%	1.3%	44.5%
Non-U.S. citizen, resident for 5+ years	16.1	33.1%	3.2%	13.8%	1.0%	48.9%
<b>H. Health Status</b>						
Excellent/Very Good	178.0	62.2%	5.9%	14.2%	1.6%	16.2%
Good	63.3	49.3%	4.2%	18.7%	2.8%	24.9%
Fair/Poor	23.5	33.1%	3.0%	29.9%	11.1%	23.0%

( ) = Estimate has a large 95% confidence interval of +/- 5.0 - 7.9 percentage points. Estimates with larger margins of error or with standard errors greater than 30% are not provided

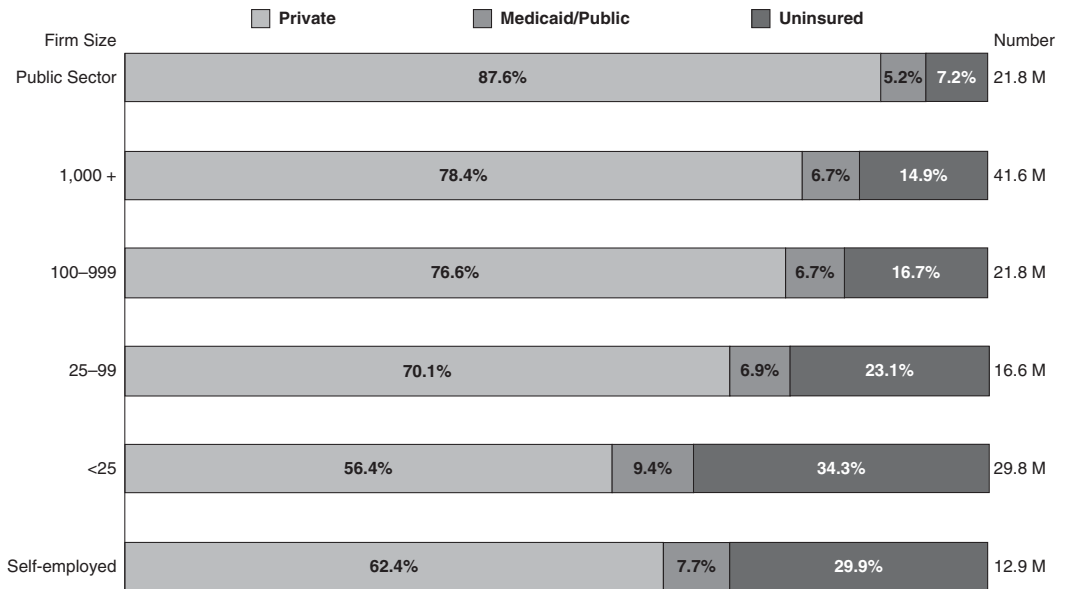
Source: [http://facts.kff.org/upload/relatedtable/Health\\_Coverage\\_of\\_Nonelderly\\_2009.xls](http://facts.kff.org/upload/relatedtable/Health_Coverage_of_Nonelderly_2009.xls), with permission, accessed May 4, 2011.

Part B shows that problems obtaining insurance were severe for those with incomes less than \$20,000 (37.8 percent, or 3 of every 8), and families with only part-time, or no workers. Of 135.4 million nonelderly working Americans (almost all of the elderly are covered through Medicare) in single-worker households (part E), 25.5 million, or 18.8 percent, were uninsured.

Problems in getting insurance were particularly severe for individuals and families getting coverage through small employers (Figure 11-4). Of the 29.8 million people working in firms with 25 or fewer employees, about 10.2 million people (or 34.3 percent) were uninsured. Contrast this to workers in firms with 1,000 or more employees. A total of 41.6 million workers are in this category, with 6.2 million (or 14.9 percent) uninsured. Even among the largest firms the number of uninsured workers had increased by almost 1 million, and from 13.0 percent in 2006, to 14.9 percent only three years later.

Repeated surveys of the uninsured (Kaiser Foundation, 2005) have exploded common myths about the numbers and motivations of the uninsured. Although it is plausible that some people rationally “choose” not to have insurance, only 7 percent asserted that they do not think they needed it. The majority said the main reason was that was is too expensive. Some members of the public, and some politicians, argue that the uninsured can get the care they need, including emergency room settings, when they really need it. However surveys indicated that over one-third of the uninsured report needing care in the previous year but not getting it, and nearly half of the uninsured report postponing care—rates at least three times higher than those with insurance. Box 11-4 updates the discussion to the most recent 2010 health care debate.

Under any circumstance, it is important to examine the reasons that some individuals lack insurance. Insurers must be able to lower the loading factors, which are the costs of determining probabilities of claims and processing claims. The ability to insure through the workplace gives the



**FIGURE 11-4** Health Insurance Status of Workers by Firm Size, 2009

### BOX 11-4

#### Why Being Insured Matters

In the debate about the Patient Protection and Affordable Care Act of 2010, there was a heated discussion about the number of uninsured, whether they were uninsured by choice, and what the impact was on their health. A 2010 report by the Center for Disease Control addressed these issues:

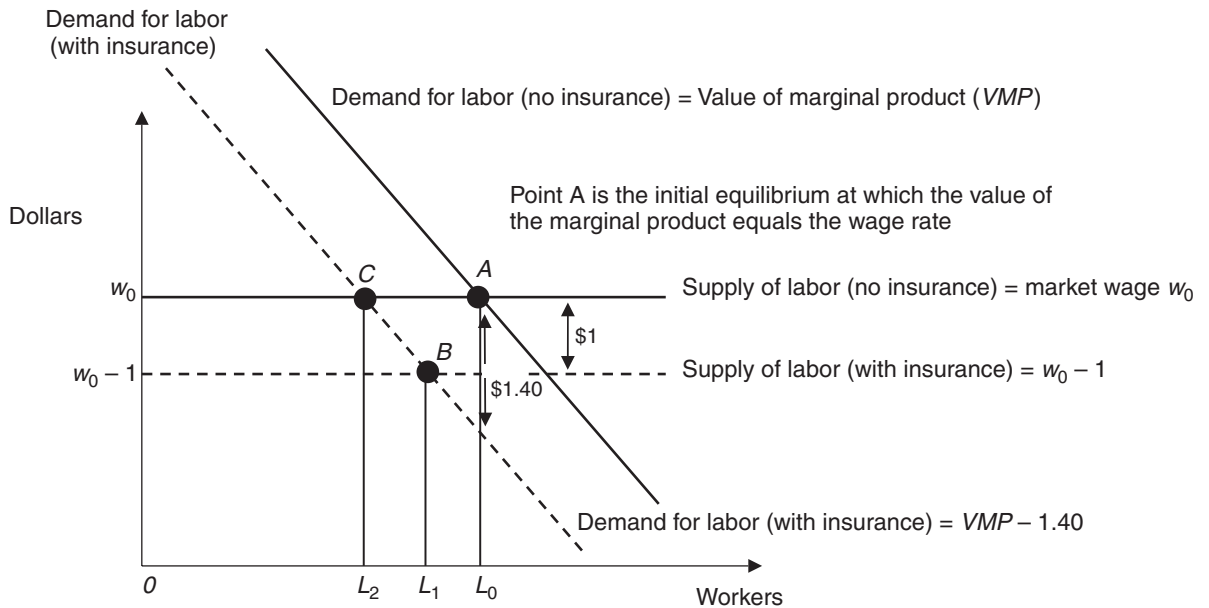
1. More than 1 in 4 adults 18–64 years old—about 50 million people—had no health insurance for at least part of the previous year. Over the previous several years, the number of adults 18–64 years old without health insurance for at least part of the year had increased by an average of 1.1 million people each year.
2. Middle-income people accounted for half of that increase.
3. Not having insurance has a greater impact among those adults who need health care the most. Delays in receiving health care can lead to poorer health and higher medical costs over time, especially for those individuals who already have health issues, including the approximately 40% of the U.S. population with one or more chronic diseases. About 60% of adults ages 18–64 with a disability who had recent gaps in the past 12 months in their insurance skipped or delayed care as well.

*Source:* Centers for Disease Control and Prevention, “Health Care: See Why Being Insured Matters,” <http://www.cdc.gov/Features/VitalSigns/HealthcareAccess/>, November 2010, accessed May 24, 2011.

opportunity to improve the experience rating. In principle, private insurers can insure those outside the workplace, and many do, particularly those in affinity groups, such as organizations and clubs, or the elderly. If high costs lead to onerous payments, then the problem may be poverty rather than high prices. This would suggest the need for governmental subsidies in a social insurance scheme.

#### The Working Uninsured

Consider now the working uninsured and start in Figure 11-5 with a firm that is a price taker (facing a horizontal supply curve) in the labor market, pays wage  $w_0$ , and offers no insurance. At initial equilibrium point  $A$ , the firm hires  $L_0$  workers. Suppose that the workers negotiate an



**FIGURE 11-5** Insurance and Employment

insurance contract worth \$1 per hour to them. As in Figure 11-2 earlier in this chapter, if the workers were willing to accept wage  $w_0$  before, they will now be willing to accept  $w_0 - 1$  dollars per hour in wages plus the insurance, and their supply curve shifts downward to reflect this willingness.

The literature suggests, however, that many businesses, especially small businesses, may have limited experience ratings; as a result, their loading costs may be 40 percent or more higher than those paid by large firms. If the loading costs are high, then it may cost much more than one dollar to provide a dollar's worth of insurance. In the Figure 11-4 example, the demand for labor curve (with \$1 per hour of insurance) shifts down by \$1.40.

At employment level  $L_0$ , the marginal cost to the firm of the workers,  $w_0 - 1$ , exceeds the marginal product net of the health insurance,  $w_0 - 1.40$ . To continue employing  $L_0$  workers, the firm must reduce the amount of insurance to less than \$1 per hour so that the sum of the value of marginal product plus the insurance equals  $w_0$ . If it cannot reduce the amount of insurance, then the firm will have to reduce its employment to  $L_1$ , at point *B*.

McLaughlin and Zellers (1994) summarize three major types of barriers to small business provision of health benefits. The first, affordability, is a price-related barrier. Particularly in low-wage industries, a combination of low profit margins, low wages, and high premiums may make it difficult to provide health benefits.

Another barrier involves insurance redlining or pre-existing-condition clauses, which may exclude specific individuals, or companies that employ them, from insurance coverage. The authors note that redlined industries typically employ older workers (over age 55), have high employee turnover, seasonal workforces, or workforces paid by commission or on the basis of other contractual terms. Lawyers' offices may be redlined because of fears of litigation, and physicians' offices because of fears of high utilization of health care services.

A third barrier to the provision of coverage may be termed attitudes. Many firms are uninterested in offering insurance. Sometimes a large number of their employees can piggyback on health plans of spouses' or partners' employers. In such cases, employees prefer to be compensated in higher wages rather than in benefits.

Simon (2008) provides important insights on arrangements to provide the pooling structures necessary to insure small employers. Proposed policy changes seek to reduce adverse selection by using state high-risk pools (mechanisms to insure the sickest workers outside of the regular commercial insurance market), and making small firms look like large firms through pooled purchasing arrangements.

An efficient pool should be as large as possible and have binding features, such as long-term contractual requirements. Simon notes that in other industries, such as in the concrete industry, small firms band together to purchase supplies. However, if it were simply a matter of size, the private market would have solved this problem using pooling arrangements. This has not happened, possibly because of adverse selection.

The particular type of adverse selection faced by a small firm with healthy employees involves being the first to join a pool that contains one or more small firms with sicker employees. The “healthy” firm must be assured that the pool will be large and stable enough and that the other small firms will not drop out once their workers become healthy again. Simon notes that whether because of legal reasons or unenforceable contracts, these arrangements have not become popular.

Literature on small employer coverage has been sparse and often anecdotal. Feldman and colleagues (1998) analyze a 1993 survey of Minnesota small employers. A problem with this type of research is that the costs of premiums are available for those firms that do buy insurance, but they are not for those firms that do not buy insurance. (Consider asking people who did not buy cars how much they didn’t spend!) The researchers carefully assigned insurance premiums to those who did not buy, and found that a \$1 decrease in the price of individual insurance increased the probability of coverage from 0.610 to 0.627. This seemingly small change implies that a 1 percent decrease in insurance premium leads to a substantial 3.91 percentage point increase in small firm demand for insurance (an elasticity of  $-3.91$ ).

In contrast, Kronick and colleagues (2008) find much smaller responses in a trial that randomly assigned small businesses to experimental treatments which varied by how much employers and their employees were required to pay for coverage. At \$20 per month, an estimated 40 percent of eligible businesses purchased insurance. At \$100 per month, 13 percent purchased insurance. Small businesses that were not currently offering insurance to employees did not respond much to large reductions in the price of coverage. The authors conclude that programs to subsidize insurance for small businesses and their employees are unlikely to reduce substantially the number of uninsured persons.

### **The Impacts of Mandated Coverage**

Many have advocated that the federal government or individual states mandate some, more, or particular types of coverage. Returning to Figure 11-5, if such coverage is provided through employee insurance, the mandates would raise the amount of insurance as well as its costs, thus lowering the (dashed) demand for labor curve net of insurance. Responses of employers to the increased marginal costs brought on by mandates may result in two adverse impacts. First, the company may stop offering insurance entirely because it is too expensive. Thus, rather than having modest health coverage with the benefits of whatever experience rating may exist within the workplace, there may be no coverage at all. (Alternatively, “play-or-pay” proposals might require firms to provide a minimal level of insurance or to pay into an insurance pool that would insure the uninsured.) The employer then may have to raise the wage to keep employees who would have to buy their own insurance. This may allow for employee choice, but it also denies the employee workplace-related experience rating.

The second adverse impact may also be understood by examining Figure 11-5. The equilibrium value of marginal product and net wage at point *A* represents a labor force of the appropriate size

**TABLE 11-3 20 Most Common Conventional Mandates—2010**

Mandates Regarding:	Number of States
Breast Reconstruction	50
Mammography Screening	50
Maternity Minimum Stay	50
Mental Health Parity	48
Diabetic Supplies	47
Alcoholism/Substance Abuse	46
Emergency Room Service	45
Mental Health General	42
Diabetic Self-Management	38
Off-Label Drug Use	36
Prostate Cancer Screening	36
Colorectal Cancer Screening	34
Drug Abuse Treatment	34
PKU/Metabolic Disorder	33
Well Child Care	33
Cervical Cancer / HPV Screening	31
Dental Anesthesia	31

Source: Victoria Craig Bunce, JP Wieske *Health Insurance Mandates in the States 2010*, with permission. [http://www.cahi.org/cahi\\_contents/resources/pdf/MandatesintheStates2010.pdf](http://www.cahi.org/cahi_contents/resources/pdf/MandatesintheStates2010.pdf), accessed May 24, 2011.

to maximize profits for the producer. Suppose, for example, that  $w_0$  represents a binding minimum wage, either by federal or by local “fair wage” laws. Even if the workers value \$1 in mandated benefits by the full dollar and would be willing to reduce their supply price, the money wage *cannot* fall lower than  $w_0$ . By this analysis, the new equilibrium is at point *C*, and the mandated coverage is likely to result in additional unemployment by the amount  $(L_0 - L_2)$ , which of course means sharply curtailed insurance benefits. How much employment falls is related to the elasticity (or responsiveness) of the labor demand curve to the increased gross wage.

Mandated insurance benefits have become most popular at the state level. In 2010 the 50 states plus the District of Columbia had imposed a total of 2,156 mandates related to benefits, treatments, and covered persons. Table 11-3 shows that the most popular benefits included breast reconstruction (50 states, of the 50 + the District of Columbia), maternity stays (50) and mammograms (50). While most of the categories reflect relatively conventional categories, other mandated categories include Lyme disease (3 states), in vitro fertilization (15 states), and attention deficit disorders (2 states).

Jensen and Morrisey (1999) found that smaller firms were affected disproportionately by mandates in part because they were less likely than larger firms to avoid the costs of mandates by self-insuring. Because health insurance would be more expensive for smaller firms due to the requirement that they include the mandated benefit, they would be less likely to offer coverage to employees.

## TECHNOLOGICAL CHANGE, HIGHER COSTS, AND INFLATION

We have seen that increased insurance coverage increases health care demand, and we know that increased demand raises price levels. However, health insurance also can induce innovations, which may be either cost decreasing or cost increasing. Furthermore, as Weisbrod (1991) has explained, a

two-way relationship can exist between insurance and innovation as new health products also may stimulate the demand for insurance.

### The Cost-Increasing Bias Hypothesis

Because increased health insurance usually reduces the out-of-pocket portion of the bill, patients may be willing to spend considerably more when insured. Consumers and their providers may realize that premiums will tend to rise if all patients act likewise, but they wish to make use of the best, new, expensive technology to the maximum extent affordable, often beyond the point where the marginal benefits would equal the full marginal costs. Consequently the payoff to owning a patent to new and costly health care technology that is covered by insurance, increases.

Process innovations (improvements to existing treatment processes that make them less costly) and product innovations (improvements through the introduction of new products, which could make the treatment more costly even when lowering the cost of improvements to the patient's well-being) may be affected differently. Insurance can encourage new products and influence innovations toward those that increase costs.

### Goddeeris's Model—Innovative Change over Time

By its very definition, technology involves innovative change over time. Economists call this a *dynamic* process. The model we use, formulated by John Goddeeris (1984), looks at *changes* in health expenditure  $\Delta m$  on the *x*-axis and *changes* in health care  $\Delta h$  on the *y*-axis. Consider providers who currently offer levels of health care  $h_0$ , which is paid for by medical expenditures  $m_0$ . Ignoring inflation, if there are no changes in health technology,  $\Delta h$  will equal zero (no change in health), as will be  $\Delta m$  (no change in expenditure). We denote that point as the origin of the graph, point 0. At the original equilibrium, then, both  $\Delta h$  and  $\Delta m$  equal 0.

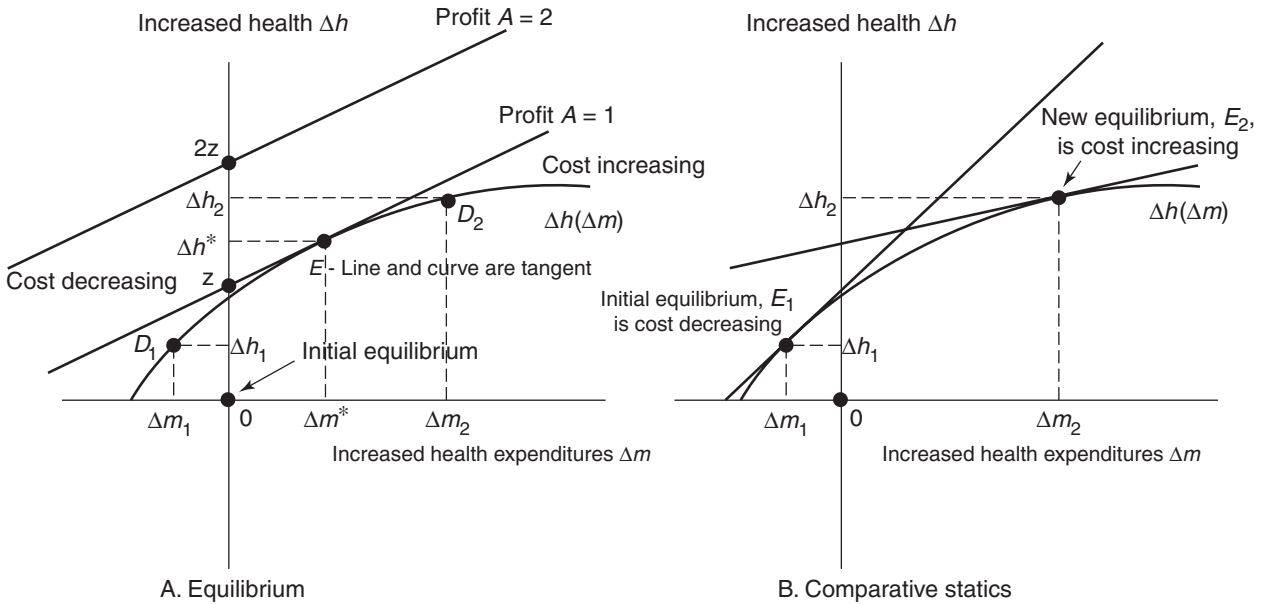
Providers may choose innovations to improve health care if the innovations lead to increased profits. Cost-reducing innovations give negative values of  $\Delta m$ . Other, cost-increasing, innovations may provide better health, but at higher cost levels entailing positive values of  $\Delta m$ . The curve in Figure 11-6  $\Delta h(\Delta m)$ , "change in health as a function of change in expenditures," shows the theoretical impacts of increased medical expenditures  $\Delta m$  on improvements in health  $\Delta h$ . Goddeeris hypothesizes that cost increasing innovations (higher values of  $\Delta m$ ) also yield greater improvements in health (higher values of  $\Delta h$ ). Under what circumstances will innovations be chosen, and how big will they be?

We have seen that health insurance policies generally require consumers to pay only a fraction,  $z$ , of the provider's bill out of pocket. Potential innovations have consequences for the equilibrium health of the consumers and for the costs of treatment. Look at the function  $\Delta h(\Delta m)$ . This means that an innovation may raise health by  $\Delta h_1$  and *reduce* treatment costs by  $\Delta m_1$ , resulting in point  $D_1$  in panel A. An alternative innovation may *increase* treatment costs, say by  $\Delta m_2$ , but substantially increase health, say by  $\Delta h_2$ , the outcome of which is shown by point  $D_2$  in panel A. Which is chosen?

By investing in a given innovation, providers seek to increase profits. Their extra profit obtained through the innovation equals the extra revenue minus the extra costs. If patients value each additional unit of health at a constant rate of \$1, then they are willing to pay at most a dollar for \$1 improvement in health status. However, the providers can charge a full price per unit of  $\Delta h/z$  for the innovation. Suppose that consumers have a 10 percent coinsurance rate. Providers could charge up to \$9.99 for an improvement worth \$1 to the consumer, because the consumer would be paying less than \$1 ( $0.1 \times \$9.99$ ) out-of-pocket for the improvements.

Providers evaluate innovations by comparing extra revenues to extra costs. In Figure 11-6A, we represent profits by the isoprofit lines, defined as a collection of points that yields a constant profit level. The isoprofit equation for profit level  $A$  is thus:

$$\text{Extra profits (A)} = \text{Change in Revenues} - \text{Change in Costs} = (\Delta h/z) - \Delta m \quad (11.1)$$



**FIGURE 11-6** The Impact of Increased Innovation Expenditures on Health

If we algebraically rearrange equation (11.1), the isoprofit equation relates change in health,  $\Delta h$ , to terms  $A$ ,  $z$  and  $\Delta m$ . That is, the isoprofit equation is:

$$\Delta h = z(A + \Delta m) = zA + z\Delta m \tag{11.2}$$

This is a linear (straight line) equation with intercept  $zA$  and slope  $z$ . Thus, the lower the coinsurance rate (the more comprehensive the insurance), the smaller the slope of the line.

We show two isoprofit lines in panel A. The isoprofit line for extra profit level \$2 ( $A = 2$ ) is so labeled in the graph. Its vertical intercept is found by setting  $m$  equal to zero in the function  $\Delta h = 2z + z\Delta m$ ; thus, the vertical intercept (where  $\Delta m$  equals 0) is  $2z$ . Similarly, the \$1 isoprofit line has an intercept of  $z$ . Higher isoprofit lines yield higher extra profit levels. Although entrepreneurs would like high profits, they are constrained by the technology. In panel A, the technology cannot provide a profit of 2; isoprofit line 2 is unattainable given the function  $\Delta h(\Delta m)$  shown. In contrast, isoprofit line 1 is just attainable, and the equilibrium choice will be a point of tangency at point  $E$ , with change in medical expenditure  $\Delta m^*$ , and corresponding change in health  $\Delta h^*$ , which is slightly higher than vertical intercept  $z$ .

Compare, for example, the profit-maximizing equilibria for two values of  $z$ , a 20 percent coinsurance rate and a 10 percent coinsurance rate. From equation (11.2) the larger the value of  $z$ , the steeper is the isoprofit line. In panel B, innovators find their optimal choices to have changed when the isoprofit curves became flatter. Assume that if the initial coinsurance rate is 20 percent (or 0.2), the optimal choice initially is point  $E_1$ , a point of tangency. If the insurance coverage becomes more generous, coinsurance rate  $z$  falls to 0.1, implying that the highest extra profit will be attained at a new equilibrium, in this case at point  $E_2$  in panel B. We can conclude that the more generous insurance coverage induces a higher level of  $\Delta m$  than for the initial equilibrium. The innovator has switched from an emphasis on *cost-reducing* (because equilibrium  $\Delta m_1$  is negative) innovations, at point  $E_1$ , to *cost-increasing* innovations (because equilibrium  $\Delta m_2$  is positive) at point  $E_2$ .

## Evidence on Technological Change and Inflation

Newhouse (1978b, 1988) set out the basic empirical models designed to test the underlying hypotheses and measure the contribution of health insurance to health care inflation. First, the conventional view emphasized the role of insurance as a demand shifter; this is measured by the variable “change in insurance.” Second, the level of health insurance (indicated by the average coinsurance rate) was hypothesized to induce innovations, thus causing inflation indirectly through cost-increasing technological change. By 1988, Newhouse had found, though somewhat tentatively, evidence that supported the induced-innovation hypothesis, leaving uncertain whether any role at all existed for the conventional demand-shifting avenue.

Investigating a larger data series and providing more testing of underlying econometric issues, Peden and Freeland (1998) addressed further questions. They studied the coinsurance level, the level of noncommercial research spending, and the annual change in average coinsurance from 1963 to 1993, during which time real per capita health spending grew by 373 percent. About half of this expenditure increase was due to the level of coinsurance, giving support to the induced-innovation hypothesis. In contrast, their data offered little support for the demand-shifting effect of reduced coinsurance. Personal income growth over the period contributed more than 20 percent of the inflation, which was more than other researchers had previously thought. A 2000 review of the literature by Smith, Heffler, and Freeland brackets the impact technology-related changes in medical practice on real per capita health care spending at between 38 and 62 percent of the overall change.

## CONCLUSIONS

Chapter 8 introduced readers to insurance and to the specific issue of health insurance for individuals. In the current chapter, we have focused on the insurance market and the behaviors of firms within that market. We established that in a market setting, insurance constitutes an important part of the wage package, and to the extent that it is valuable to the workforce, higher insurance is reflected in lower money wages. This market result occurs irrespective of who contractually pays for the insurance.

We have also shown how many of the trappings of the U.S. health care system are related to the employer base of the health insurance. The system of tax deductibility pushes employees to ask for higher proportions of untaxed insurance relative to taxed wage benefits. The linkage of insurance to the workplace also tends to lock employees into certain jobs, constraining mobility.

We continued with a brief discussion of the evolution of the health insurance industry. This industry was formerly defined by the Blues (Blue Cross and Blue Shield) with their nonprofit status, community rating, and predilection for hospital care. Although they are still large, the decline in the primacy of the Blues has led to profound changes in the provision of health insurance and the delivery of care.

The chapter also included several implications about the uninsured. Some are not employed and hence ineligible for health insurance. There are others, however, whose health, employment, or lifestyles may not permit commercial insurers to provide insurance profitably. Government mandates to employers that they insure everyone if they insure anyone, may cause employers to drop insurance plans entirely, thus leaving larger numbers at risk. A safety net for those who are difficult to insure requires a social contract that only the government can provide. While this safety net has been established in most advanced countries for some time, in the United States, the Patient Protection and Affordable Care Act of 2010 represents a major, although hardly complete, step in this direction.

We finished the chapter by looking at the impact of insurance on cost-increasing technological change. We noted that a major portion of health care cost inflation has been attributed to technological change. Having looked at the structure of the insurance industry, we turn next to issues of managed care.



## Summary

1. Individual health insurance, in theory, trades off a guaranteed reduction in wealth (the insurance premium) for a reduction in uncertainty due to ill health through the pooling of risk. The organization and cost conditions in some health insurance markets, however, suggest that some contingencies may not be insurable.
2. The economies of scale in processing information suggest that smaller firms or unions may see high marginal insurance costs relative to marginal benefits levels. These higher costs may reduce or eliminate the range of services that they offer.
3. If workers in an industry value health insurance, then competitive pressures lead to reduced money wages' offsetting increased health benefits.
4. Subsidizing health insurance through government tax policies lowers the price of \$1 of insurance premium relative to \$1 of wage remuneration. This leads employees to purchase more health insurance relative to wages than would otherwise occur.
5. Health insurance is a key determinant in the decision to retire. Studies suggest that the availability of retiree health insurance raises the odds of retirement by between 30 and 80 percent.
6. Employer-provided health insurance may inhibit worker mobility between firms, thus locking employees into jobs. Researchers find this job lock to be responsible for substantial reductions in employee mobility.
7. Much of the American health insurance environment has been defined by Blue Cross and Blue Shield plans, which as nonprofit firms were typically exempted from profits taxes, property taxes, and federal and state corporate taxes. Analysts once felt that such tax treatments gave the Blues considerable competitive advantages, but over the past two decades the Blues have faced considerable competition.
8. Many groups have advocated that the federal government or individual states mandate either more coverage or various types of coverage. In reaction:
  - Companies may stop offering insurance entirely because it is too expensive.
  - Marginal workers may now cost more than they are worth, and some of them will be let go.
9. Goddeeris shows how insurance coverage can influence technological change. More generous insurance coverage increases the relative profitability of cost-increasing technological innovations versus cost-reducing ones.

## Discussion Questions

1. Suppose each person's health expenditures can be predicted with certainty by both the insured and the insurer. What are the implications for insurance markets? Explain the prevalence of insurance for highly predictable events, such as routine dental services.
2. In 1986, the U.S. federal income tax system changed marginal tax rates so that the top federal marginal rate fell from 50 to 33 percent. More recently many marginal rates have fallen further. From what you know about how fringe benefits are negotiated, what would you expect to happen to the demand for health insurance as a fringe benefit? Why?
3. Suppose your health insurance allows you, a worker, to buy whatever prescription drugs you wish for \$5 per prescription. In contract negotiations it is proposed to change this benefit to "10–20," that is, you pay \$10 for generic drugs and \$20 for brand name drugs. What would your reaction be? What would economic analysis predict?
4. Suppose that a company pays its workers \$20 per hour and provides an additional \$2 per hour worth of fringe benefits, including a basic health insurance policy. Discuss the firm's reaction to a state mandate that requires it to expand the items covered in the health care policy. What is likely to happen to the number of people employed?
5. Blue Cross plans typically have practiced community rating. If other insurance firms are seeking healthier patients at reduced rates, what impact will this have on Blue Cross net revenues? Why?
6. According to the 2010 Patient Protection and Affordable Care Act, adults with pre-existing conditions became eligible to join a temporary high-risk pool, which will be superseded by the health care exchange in 2014. To qualify for coverage, applicants must have a pre-existing health condition and have been uninsured for at least the past six months. Analyze the impacts of such regulations on insurance markets.
7. What is job lock? Would you expect job lock to increase or decrease if employer-based health insurance were to be replaced by government-provided health insurance?
8. Suppose a household does not carry health insurance. Can we conclude that this reflects failure of insurance markets? Why or why not?

## Exercises

1. Using Figure 11-1, illustrate the probability that someone will obtain insurance for treatment for
  - (a) A hangnail
  - (b) A broken arm
  - (c) A “bad hair” day
  - (d) Viral meningitis
2. Using Figure 11-2:
  - (a) Calculate an initial labor market equilibrium (wages and employment) determined by the demand and supply of labor.
  - (b) Indicate the wage and employment impacts of a health insurance policy that costs \$2 per hour to employers and is worth \$1.50 per hour to the workers.
  - (c) Indicate the wage and employment impacts of a health insurance policy that costs \$1.50 per hour to employers and is worth \$2 per hour to the workers.
3. Using Figure 11-2, consider an insurance policy that provides free “purple aspirin” to all workers. This benefit provides no conceivable advantage (workers don’t care whether their aspirin is purple or white) but comes with cost  $z$ . Show the new labor market equilibrium indicating the wage and employment impacts.
4. Using Figure 11-2, indicate the wage and employment impacts of a health insurance policy that costs \$2 per hour to the workers and is worth exactly \$3 per hour to the workers. Why do your answers to exercises 2, 3, and 4 differ?
5. Consider the market labor demand  $L_D$  and labor supply  $L_S$ , where  $W$  is the market wage.
 
$$\text{Demand: } L_D = 1,000 - 20W$$

$$\text{Supply: } L_S = -200 + 400W$$
  - (a) What is the equilibrium market wage? What is the equilibrium employment level?
  - (b) Calculate the equilibrium market wage and employment level if the workers negotiate a benefit worth \$1 that costs the employers \$2.
  - (c) Calculate the equilibrium market wage and employment level if the workers negotiate a benefit worth \$2 that costs the employers \$1.
6. Consider two workers, Ralph and Steve. Both of them work for the same employer, and each earns \$15 per hour. Steve is taxed at the 15 percent marginal rate. However, Ralph is married, and due to his wife’s income, he is taxed at the 28 percent marginal rate. Using Figure 11-3, indicate which one would be expected to seek more health insurance and why.
7. Suppose that Charlie’s Pizzeria in Kalamazoo, Michigan, employs 10 employees at a wage level of \$8 per person. All other costs (ovens, rent, advertising, return to capital) total \$40 per hour, and the pizzeria sells 12 pizzas per hour at a cost of \$10 per pizza. Suppose the state of Michigan mandates health coverage that can only be covered at a cost of \$1 per hour, if it is offered at all. Charlie finds that if he offers insurance, he could maintain production by letting one worker go and running his pizza ovens a little hotter, leading to costs of \$45 per hour.
  - (a) What are Charlie’s original profits?
  - (b) What is Charlie’s elasticity of demand for labor? How is this calculated?
  - (c) What will happen to Charlie’s profits in the short run if he chooses to pay for mandated insurance?
  - (d) What will Charlie’s long-run decision be? Why?
8. In the Goddeeris model in Figure 11-5, how would the equilibrium change in expenditure,  $\Delta m$ , be affected if there were a vertical parallel shift in the  $\Delta h(\Delta m)$  curve, say caused by increased support of medical technological research? Show graphically the alternative possibility if support for research tended to emphasize cost-reducing innovations.

# Managed Care

- What Is the Organizational Structure?
- What Are the Economic Characteristics?
- The Emergence of Managed Care Plans
- Development and Growth of Managed Care—Why Did It Take So Long?
- Modeling Managed Care
- How Does Managed Care Differ?—Empirical Results
- Growth in Spending
- Competitive Effects
- The Managed Care Backlash
- Conclusions

The previous chapter described how conventional health insurance will generally increase consumers' health care utilization. Those who are insured consider the out-of-pocket cost of care rather than the true full cost at the point of service. In the absence of restrictions, they will purchase services beyond the point at which the marginal benefit of the care equals its marginal cost. For insurance to provide a net benefit to society, the costs of this increased health care consumption must be overcome by the benefits of the reduced financial risk to patients. Even with the benefits of risk reduction, improved insurance coverage leads to increased costs to society.

A simple analogy may help address the dilemma. Suppose that rather than health care insurance, employers provided food and clothing (F&C) insurance for their workers. A “fee-for-service” F&C plan would allow the consumers to purchase their food and clothing at any merchant they choose and would reimburse the consumers subject to coinsurance rates and deductibles. A consumer facing a 20 percent coinsurance rate could purchase filet mignon for \$20 per pound and have the insurer pay \$16 per pound, or 80 percent of the price. Another consumer could purchase designer athletic shoes for \$200 and have the insurer pay \$160, again 80 percent of the price. Consumers would likely buy more (or more expensive) filet mignon or athletic shoes than if they had to pay the full amount themselves. The market effect of such plans would likely cause consumers and their insurers to worry about F&C cost and expenditure inflation.

Suppose that in response to the perceived high costs of food and clothing, and the consequent high cost of the F&C insurance plan, a group of consumers and their employers organized and offered a “managed F&C” plan. In this plan, members (consumers and the employers) would pay a fixed amount per person per month for

food and clothing, presumably less than they were paying (together) under the fee-for-service plan. In return for this reduced cost plan, the consumers would be limited to shopping at a single shopping center with which plan managers had negotiated lower prices for food and clothing. Moreover, the plan managers could limit the types of goods purchased (no filet mignon) and might attempt also to curb total consumer expenditures in other ways. It is likely that at least some consumers would find such a plan attractive due to its lower costs.

Analysis of this “managed care” F&C arrangement raises a host of questions. For individual consumers, one might ask:

- Are they getting the same quality of goods as before?
- Are they being denied goods that they “should” be getting?
- Are their expenditures reduced?
- Are they less well-fed, less healthy, or less well-dressed than before?

At the market level, one might ask:

- Do aggregate F&C expenditures decrease or does their growth rate decrease?
- Do the managed F&C plans meet consumer preferences?
- Can the merchants earn sufficient returns to stay in business?
- Is there competition in the managed care F&C market, and do the managed care F&C plans influence the fee-for-service F&C plans in terms of the prices or selection of goods?

With this in mind, we turn to the issue of managed health care. One might argue that physician practice must be managed in order to address high health care costs. This suggests that networks of providers, including HMOs (health maintenance organizations), PPOs (preferred provider organizations), and individual practice associations (IPAs), are widely seen as means to restore competition to the health care sector and as means to control expanding health care costs. We devote this chapter to managed care with particular attention to the distinctive combination of insurance and care exemplified by HMOs and similar organizations. Unless distinguishing the individual types of institutions, we will refer to them as managed care organizations (MCOs). The HMO receives special attention in this chapter for its pioneering role and for the fact that much of the scholarly and policy research has focused on HMOs.

HMOs appear to overcome the information problems inherent in fee-for-service (FFS) health care markets that ordinary insurance coverage may exacerbate. Under FFS, the provider provides health care and advises the consumer on how much to get. At first glance, it appears that the consumer’s imperfect information about health care, when combined with FFS remuneration, may create the incentives for substantial overconsumption. The HMO organizational form appears to eliminate the overconsumption incentives and replace them with cost-control incentives and even possibly incentives toward underconsumption.

We begin this chapter by describing managed care and its cost-cutting potential. We then turn to HMOs as a form of health care organization that combines the functions of insurance and the provision of care. We describe the HMO and its organizational relatives, and we assess the theory and evidence on their effects. We then turn to the market effects of managed care on providers, insurers, and the adoption of new health care technology.

## **WHAT IS THE ORGANIZATIONAL STRUCTURE?**

It is instructive to provide a general description of MCOs, leading to a more specific discussion of HMOs, while recognizing that the concept of managed care is undergoing constant changes. Analysts speak of an organized delivery system as a network of organizations (e.g., hospitals, physicians, clinics, and hospices) that provides or arranges to provide a coordinated continuum (from well care to emergency surgery) of services to a defined population. This system is held clinically and fiscally accountable for the outcomes and the health status of the population served. It is tied

together by its clinical (treatment) and fiscal (financial) accountability for the defined population. Often the organized delivery system is defined by its association with an insurance product.

In principle, managed care creates incentives for keeping people well by emphasizing prevention and health promotion practices, and by treating those who become ill at the most cost-effective location in the continuum of care. Through a more centralized management of services, the goal is to provide additional quality-enhancing features for a given price, or to provide a given set of quality attributes or outcomes for a lower price. A primary provider typically serves as the patient's gatekeeper to help ensure appropriate care and limit overutilization.

Two features characterize the contemporary MCO. The first is the extensive reliance on health care information systems. Initially, these systems were developed mainly to replace clerical functions such as billing and record keeping. Indeed, the "embryonic" development of clinical information systems constituted a fundamental barrier to the success of managed care organizations. In the 1990s, large health centers budgeted tens of millions of dollars per year to integrate systems that often were developed separately and almost never "talked to each other" (Shortell et al., 1994). A major challenge facing managed care is the design of information networks that provide direct clinical support to improve the process of care.

A second feature of MCOs is their de-emphasis of the acute care hospital model. Hospitals provide expensive care, and moving toward cost-effective systems necessarily moves away from hospital care. As noted earlier, primary care physicians are often the gatekeepers of managed care systems, directing patients to appropriate (i.e., cost-effective) treatment settings. If they "feed" patients into the hospital instead, this leads to increased costs. Managed care seeks a vertical integration of what had previously been a generally unintegrated system of health care treatment. Through coordination of care and improved information, such integration has the potential to address the health care costs in a manner that would appear to address criteria of economic efficiency. Yet the integration is costly, and the quality of the resulting care may not match all consumer preferences. Some also claim that managed care systems have incentives to underprovide services, which may be harmful to patients.

The HMO represents a prime example of managed care on which there has been considerable research. We begin by describing HMOs and we continue with their history and with the rationale for a government policy that has promoted their development.

## WHAT ARE THE ECONOMIC CHARACTERISTICS?

Managed care features a health care delivery structure involving the integration of insurers, payment mechanisms, and a host of providers, including physicians and hospitals. What distinguishes managed care from the fee-for-service care that also might plausibly attempt to integrate the various health care system parts?

There are four related mechanisms by which health insurance plans seek to contain costs and/or improve quality of care:

1. *Selective contracting*, in which payers negotiate prices and contract selectively with local providers such as physicians and hospitals. There may be price differences across providers and other contract features such as volume limits and discounts based on volume.
2. *Steering* of enrollees to the selected (in-network) providers. If patients select non-network providers, they may have to pay substantially higher out-of-pocket costs and, in some plans, pay the entire costs of these services.
3. *Quality assurance* through meeting voluntary accreditation standards. Practice guidelines, "best practices," and disease management programs are often incorporated into quality improvement activities.
4. *Utilization review* of the appropriateness of provider practices. The utilization review process may be prospective (in advance), concurrent (at the same time), or retrospective (looking back).

Of the four, most analysts find selective contracting to be most important. Dranove, Simon, and White (1998) point out that contracts may be awarded on the basis of the providers' willingness

to accept particular payment plans and monitoring of treatment styles and quality of care. Morrisey (2001b) notes that under selective contracting (unlike FFS care), some providers get contracts and some do not. Service price becomes important in managed care system negotiations with providers.

The selective contracting and the steering distinguish managed care from the more standard FFS care. Managed care is also distinct through its quality assurance emphasis. Most plans seek NCQA accreditation. The NCQA, a private nonprofit organization, establishes performance measures through HEDIS (Healthcare Effectiveness Data and Information Set), and it also issues report cards. Despite these efforts, as discussed later in this chapter, many analysts are disappointed at the progress managed care has made in improving quality. As for utilization review, Morrisey argues that traditional indemnity plans that include pre-hospital admission certification would be classified as managed care plans under this definition. Almost everyone now reviews utilization, so utilization review in itself is not helpful in discussing managed care.

What, then, is a good analytical way to conceptualize between MCOs and FFS? Cutler, McClellan, and Newhouse (2000) provide a useful conceptual model that asks how much a patient would have to be compensated to move from FFS to MCO coverage. The compensation presumably would be related to the patient's difference in utility (satisfaction) between FFS and MCO coverage. If an MCO and an FFS plan were identical, the compensation would be zero; if the MCO leads to less (more) utility, compensation must be positive (negative) to make the client indifferent.

Three differences between MCOs and FFS might affect compensation:

1. ***Difference in health.*** If the MCO provides reduced health (relative to FFS) due to reduced treatment, then the compensation must be positive for those who choose the MCO. This positive compensation might be offset if the MCO is better at managing the overall care process or at providing “well care.”
2. ***Cost savings.*** If, holding health constant, the MCO provides savings due either to less treatment or cheaper treatment, the compensation must be negative, because the MCO is saving money for its clients.
3. ***Financial risk from different out-of-pocket payments.*** Clients may prefer an MCO if it ensures them from having to make large out-of-pocket payments. If so, the compensation will be negative because payment variability is reduced. The size of the compensation would depend on the MCO's cost-sharing provisions, as well as reimbursement for out-of-plan use.

This framework suggests that one must measure the differences between managed care and fee-for-service along several dimensions: health, price of care, and quality of care. In fact, patients who value health less (or other things more) may choose less health and/or health care by choosing an MCO, or possibly even no insurance, rather than FFS care.

It does not necessarily tell us which mechanism provides the appropriate level of care at which marginal benefits equal marginal costs. Recall that under FFS, with fractional coinsurance, clients may overuse services. Under managed care, they may use fewer services and possibly not enough of them, but it is not clear whether they will use the efficient amount.

With this framework established, we can look at the emergence of managed care plans and what the market for managed care will look like. We also can look at the differences in health, price, or quality of care, recognizing that consumers and employers, acting on their behalf, will evaluate all of these dimensions in spending their health insurance dollars.

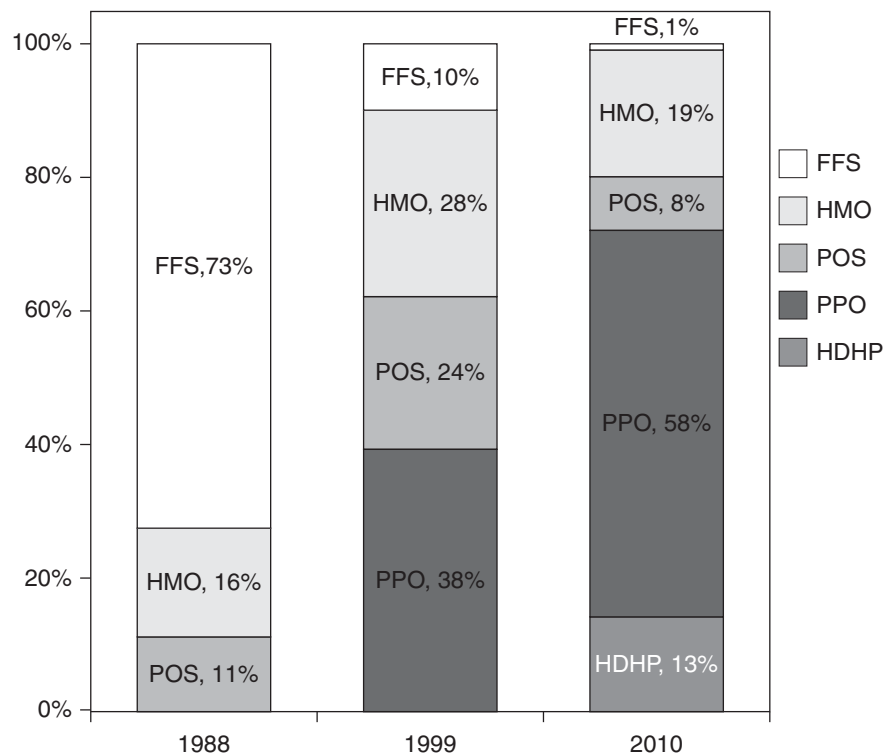
## THE EMERGENCE OF MANAGED CARE PLANS

Managed care describes a variety of arrangements with the following common features. First, much, if not all, of the patient's care is provided through a specific network of hospitals, physicians, and other health care providers. Second, considerable centralized oversight of resource use, often referred to as utilization review, occurs within the network. Here, we provide a brief overview of the types of managed care plans and the extent to which physicians and hospitals now contract with insurers under capitation arrangements.

## Employer-Sponsored Managed Care

Employer-sponsored insurance dominates the private health insurance market. Of the 167 million non-elderly persons with private health insurance in 2009, 150 million obtained coverage through the workplace. The Kaiser Family Foundation (KFF), through surveys and other sources, provides a wealth of information on the health care marketplace and employer-sponsored health plans.<sup>1</sup> The KFF data in Figure 12-1 document the historic and dramatic shift to the three main types of employer-sponsored managed care plans. Traditional indemnity (FFS) insurance accounted for just 1 percent of enrollments in 2010 compared to 73 percent in 1988. Figure 12-1 also shows the recent emergence of high deductible-health plans (HDHPs). In 2010, these plans represented 13 percent of covered workers, up from 4 percent in 2006.<sup>2</sup>

Many employer-sponsored plans are self-funded, i.e., self-insured. In contrast to fully-insured plans, where the entire risk is borne by the insurance company, an employer bears the financial risks under a self-funded plan, although it may purchase various insurance protections against unexpectedly large claims. Self-funded plans are governed by the Employee Retirement Income Security Act



**FIGURE 12-1** Health Plan Enrollment for Covered Workers, by Plan Type, Selected Years

Source: Kaiser Family Foundation/Health Research and Educational Trust, *Employer Health Benefits, 2010 Annual Survey*, Exhibit 5.1. Available at [www.kff.org/insurance](http://www.kff.org/insurance) accessed May 23, 2011. This information was reprinted with permission from the Henry J. Kaiser Family Foundation. The Kaiser Family Foundation is a non-profit private operating foundation, based in Menlo Park, California, dedicated to producing and communicating the best possible information, research, and analysis on health issues.

<sup>1</sup> Available at [kff.org](http://kff.org). The KFF sources used in this chapter include: *Employer Health Benefits 2010 Annual Survey*, *Trends and Indicators in the Changing Health Care Marketplace*, and *Kaiser Slides*. Other sources are: U.S. Department of Health and Human Services, *Health United States, 2010*, and the *2011 Statistical Abstract*.

<sup>2</sup> Employer-sponsored HDHPs with health reimbursement arrangements (HRAs) or health savings accounts (HSAs) were part of the 2003 Medicare Modernization Act. Although these plans often include managed care features, they are usually considered as a distinct category. We elaborate more on these plans in Chapter 23.

(ERISA) of 1974 which offers employers considerable flexibility by exempting them from state insurance laws including mandated benefits. According to the KFF, 59 percent of covered workers in 2010, and 83 percent of workers in firms with 200 or more employees, belonged to ERISA plans.

There are wide variations across health plans and provider organizations in terms of management of utilization and other features. With the emergence of many hybrid forms, the taxonomy of managed care is continuously evolving. For simplicity, we describe the three basic types of employer sponsored managed care plans.

*Health maintenance organizations (HMOs)* provide relatively comprehensive health care, entail few out-of-pocket expenses, but generally require that all care be delivered through the plan's network and that the primary care physician authorize any services provided. Each subscriber is assigned a primary care physician ("gatekeeper") upon joining the HMO. If health care services are provided without gatekeeper authorization, then the HMO usually does not cover the services. The subscriber is personally liable for payment of the nonauthorized services. HMOs that directly employ physicians in their network are called staff model plans. In its simplest characterization, these physicians are paid salaries by the HMO, although some HMOs do base payments on factors such as patient load. Alternatively, plans that set up their network by contracting with physicians in geographically spread out, independent solo or small group practices are called independent practice associations (IPAs). Both types assign primary care physicians as gatekeepers for covered services. IPAs are more common than staff model HMOs.

*Preferred provider organizations (PPOs)* give subscribers two distinct tiers of coverage. When subscribers use the PPO's preferred provider network, the required cost sharing with deductibles or coinsurance is lower than when they use nonnetwork providers. Although a network is formed, PPOs have no physician gatekeepers. Rather, patients simply must pay more out of pocket if they choose to go outside the network. In this way, PPOs create financial incentives for subscribers to use network providers rather than go outside the network for care.

PPO contracts with physicians and hospitals generally address the prices providers will charge the PPO. In return for promising to charge a lower price, selected providers become part of the PPO's preferred network. No guarantee is given that the provider will see patients under the plan, but if the network is not too large and the PPO's cost-sharing provisions for subscribers are network-favorable, then the provider may enjoy a large increase in patient care business by joining the network. Prompt payment for services is another advantage.

Providers often agree to submit themselves to some form of utilization review under the contract. Most PPOs require pre-admission certification for a hospital stay and concurrent utilization review for such stays. About half require a mandatory second opinion for a recommendation of surgery.

*Point-of-service (POS)* plans are a hybrid of HMOs and PPOs. Like PPOs, POS plans offer two tiers of insurance benefits. Coverage is greater (out-of-pocket costs are lower) when members use network providers and less generous (out-of-pocket costs are higher) when they use non-network providers. Like HMOs, however, POS plans assign each member a physician gatekeeper, who must authorize in-network care in order for the care to be covered on in-network terms. Most POS plans do not require authorization for a member to use out-of-network services, but such care is covered on less-generous terms.

Table 12-1 categorizes the organizational structures. The matrix rows indicate whether an organized provider network is formed. The columns indicate whether a gatekeeper is part of the

**TABLE 12-1 Different Health System Organizational Structures**

		Gatekeeper	
		No	Yes
Provider	<b>No</b>	Fee-for-service (FFS)	Point-of-service (POS)
Network	<b>Yes</b>	Preferred provider organization (PPO)	Health maintenance organization (HMO)



arrangement. The gatekeeper and the provider network represent two particular forms of health system control. HMOs provide both, while FFS plans provide neither.

### **Medicaid and Medicare Managed Care Plans**

In the last few years, many states have adopted managed care models for the Medicaid coverage they provide to families with dependent children and pregnant women who meet their low income criteria for Medicaid eligibility. They believe that managed care may help contain program costs, which are major parts of most states' budgets. As of 2009, 36.2 million Medicaid beneficiaries nationwide (72 percent of Medicaid recipients) were enrolled in some form of managed care. This represents a sharp increase from the 2.7 million as recently as 1991.

As with employer plans, Medicaid managed care plans vary considerably. In some areas, states have contracted directly with HMOs that already exist in local markets. In others, states have created their own loosely structured provider networks, which contract with selected providers for discounted services and use physician gatekeeping to control utilization. Some Medicaid programs combine the two approaches.

Unlike Medicaid and private insurance, traditional fee-for-service coverage dominates Medicare, the federal program for the elderly. Of the nearly 46 million Medicare enrollees in 2009, only 10.9 million selected a Medicare Advantage plan (also known as a Part C plan). These are private plans that receive a fixed monthly amount per enrollee from Medicare. Medicare Advantage includes fee-for-service plans but HMOs and PPOs account for nearly 80 percent of the enrollments.

There are significant ongoing developments that are relevant to both Medicaid and Medicare managed care, and the major ones are further discussed in Chapters 21 (Social Insurance) and 23 (Health System Reform). At this time (2011), some states, such as Florida, have shifted or are planning to shift their entire Medicaid populations into managed care. Medicaid managed care programs will also get a boost from the Patient Protection and Affordable Care Act (PPACA). In a dramatic expansion, Medicaid eligibility will include all individuals under 65 with incomes below 133 percent of the federal poverty level by 2014. This contrasts with Medicaid's historic emphasis on coverage for pregnant women, children, and the disabled.

Medicare managed care will also be impacted, though adversely in this case, by the PPACA legislation. President George W. Bush and many political conservatives favored the expansion of Medicare Advantage as a way of increasing competition and ultimately holding down the growth of Medicare costs. Government payments to Medicare Advantage plans exceeded the cost per regular Medicare enrollee. Political liberals viewed this as a threat to traditional Medicare and, under PPACA, cuts in payments to Medicare Advantage plans will begin in 2011 until the disparity is eliminated.

### **Managed Care Contracts with Physicians**

Managed care contracts with physicians vary considerably. Most HMO and POS plans pay their network physicians on a capitation basis. Under capitation, the plan pays the physician's practice a fixed fee, generally an actuarial per-member-per-month (PMPM) dollar amount, in return for the treatments provided to members of the insurance plan. Physicians also may be responsible for the costs of referrals, laboratory tests, and hospital services. Thus, HMO and POS plans shift the costs of care, as well as the risk associated with those costs, directly onto physician practices. In so doing, these contracts put physician earnings at risk. If care provided under such arrangements turns out to cost less than the fixed-dollar plan payment, the practice makes a profit. If instead care costs more than the payment, the practice must take a loss.

In contrast, PPO contracts with physicians rarely involve capitation. Instead, they specify the discounted fees for various services that the plan will pay in exchange for the privilege of being in that plan's network. If a physician joins the PPO's network and happens to provide services to one

of that plan's subscribers, the practice must accept the pre-negotiated fees as payment in full. "Balance billing" of the patient (for the remainder of a higher bill) is not allowed.

Managed care contracts, whether they are HMO, PPO, or POS plans, commonly contain utilization review procedures for physicians. Most managed care contracts also require a certain degree of physician record-keeping on their enrollees (e.g., plan-specific patient encounter forms may have to be filed with the insurer each time care is provided).

Medicaid managed care contracts with physicians parallel those of private managed care plans, although specific service packages are determined heavily by the state's policies. In states that have set up their own Medicaid provider networks, the state contracts directly with individual gatekeeper physicians, agreeing to pay them a small fixed fee (e.g., \$3 per month) for each Medicaid enrollee under their jurisdiction. In return for this payment, the physician serves as the gatekeeper for Medicaid-covered services.

The KFF reported that 89 percent of physicians had at least one managed care contract in 2004–2005, accounting for 44 percent of practice revenues. Some of these contracts are capitated contracts where physicians receive a flat monthly fee for each patient they cover. In 2001, the latest year for which such KFF data are available, 32 percent of all physicians had capitated contracts. Primary care physicians exhibited higher prevalence of capitation contracts than did specialists because the primary care physicians serve more often as gatekeepers in HMO and POS plans. To reduce the risk associated with capitation, some physicians purchase reinsurance against large losses or negotiate stop-loss provisions in their contracts to limit the liability per enrollee.

### Managed Care Contracts with Hospitals

HMO and PPO plans contract with only a subset of the providers (physicians and hospitals) in the areas that they serve. This key feature of the managed care sector allows plans to promote price competition among hospitals that might otherwise lose plan business.

As recently as the early 1980s, fewer than 20 percent of the insured population was enrolled in managed care plans, with most in Kaiser-model HMOs (named after Kaiser-Permanente HMO system) where the HMO owned the hospitals that its members used. In the 1980s, many states passed "selective contracting" laws, which provided insurers with greater flexibility to develop alternative health plans and to test different design features. These laws led to growth in PPOs and allowed more flexibility than the "closed-system" HMOs such as Kaiser.

By 2005, the KFF estimated that the proportion of hospitals reporting revenue from capitated contracts had increased to 38 percent (from 30 percent in 1998). Hospitals in urban areas, and particularly inner-city facilities, are more likely to report capitation revenues than are rural hospitals.

The probability and characteristics of contracts between individual managed care organizations and hospitals appear to depend on three sets of factors:

1. *Plan characteristics*, including whether it was a PPO or an HMO (and possibly what type of HMO), plan size, whether the plan serves several localities, and how old the plan is
2. *Hospital characteristics*, including size, ownership (including for-profit versus nonprofit status), location (city versus suburb), teaching status, and cost structure (reflecting prices)
3. *Market characteristics*, generally measured at the metropolitan area level, including the penetration and rate of growth of managed care plans

Research has found equivocal results, most often on the important issue of hospital costs, which are used to reflect prices to the plans. Early studies found that before managed care plans became popular, more competitive markets had higher hospital costs. This occurred because under cost-based FFS reimbursement, hospitals could (and did) compete on the basis of services and quality rather than price. More recent research has suggested that competition in hospital markets can lead to lower costs when the insurance market includes sufficient managed care penetration.

Zwanziger and Meiorowitz (1998) examine the determinants of plan contracts with hospitals in a study that looks at the three categories. For HMOs and PPOs in 13 large, metropolitan statistical areas (MSAs), they report:

- Managed care plans prefer to contract with nonprofit hospitals, preferring even public hospitals to for-profit ones.
- Plans will more likely contract with large hospitals compared with medium-sized hospitals, and with medium-sized hospitals compared with small ones.
- Hospital cost factors (which reflect hospital prices) do not significantly affect contracts.

It is useful to elaborate on some of the findings. The authors suggest that for-profit hospitals may be less willing to agree to managed care plan concessions than are nonprofit hospitals. Alternatively, plan choice may reflect consumer demand, and consumers might prefer nonprofit hospitals, using the nonprofit status to indicate a higher quality of care.

The insignificant role of costs, reflecting prices, indicates a need for further analysis. The authors believe that the price of services ought to matter in the decision to affiliate, particularly when one has controlled for other factors. They conclude that their cost terms may not reflect the prices, as they believed they would. It is also plausible that managed care plans use costs as quality proxies, or that the public correlates low costs with perceived lower quality.

## **DEVELOPMENT AND GROWTH OF MANAGED CARE— WHY DID IT TAKE SO LONG?**

Why did it take managed care so long to become a force in the marketplace? A variety of institutional, economic, and political forces has influenced its development and growth. Above all, the story of managed care in the United States requires an account of the strenuous historical opposition from organized medicine. Early on, organized medicine fiercely opposed alternatives to free patient choice and particularly alternatives to FFS reimbursement. These alternatives posed a threat to a physician's ability to earn excess profits (known as economic profits or rents). Kessel (1958) described the historic political actions of organized medicine and hypothesized that FFS enabled physicians to charge some patients a higher fee than others for essentially the same service, a pattern known as price discrimination. Such price discrimination is difficult or impossible under the contracts that characterize prepayment-based organizations. First, providers will find it difficult to determine how much individual consumers value the services. Second, the prepayment-based organization may be able to shop among providers, thus limiting the providers' monopoly power.

Organized medicine also created other barriers to managed care development. For example, it opposed physicians' participation in plans that were not controlled by physicians and/or that were not offering a free choice of physician. The Federal Trade Commission successfully challenged such restrictions in the 1970s.

### **Federal Policy and the Growth of Managed Care**

The HMO Act of 1973 represented a turning point in federal policy in promoting the development of alternative delivery systems as a cornerstone of a cost-containment strategy. The act enabled HMOs to become federally qualified if they provided enrollees with comprehensive benefits and met various other requirements. Loan guarantees and grants for startup costs were made available, but the main advantage accruing to a federally qualified HMO was that it could require firms in its area with 25 or more employees to offer the HMO as an option. Other regulatory barriers subsequently were reduced.

Despite these changes, the number of HMOs and HMO enrollees did not accelerate until the 1980s. When incentives to enroll Medicare and Medicaid recipients improved, the entry of for-profit HMOs led to growth from 235 in 1980 to 623 in 1986. Growth slowed in the late 1980s and

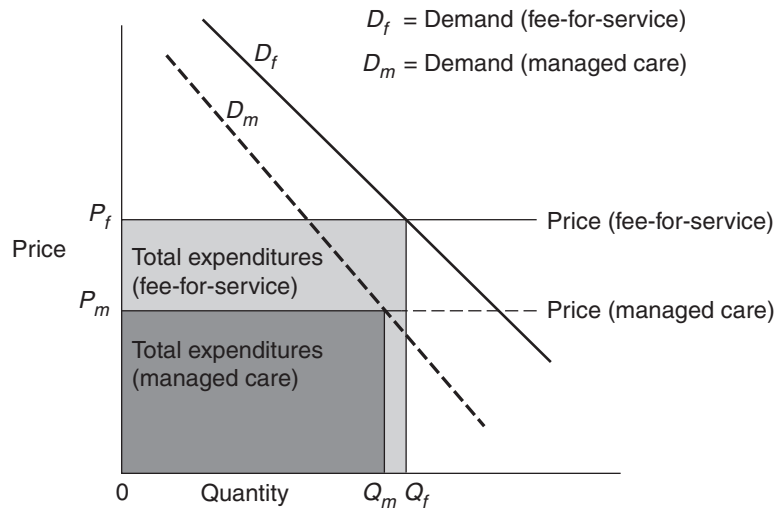
some consolidation occurred in the number of HMOs, but the upward trend in enrollment resumed in the 1990s. While the number of plans declined slightly from 1990 to 2000, total enrollment increased from 33 million to 81 million persons. Since then, HMO enrollments and the number of plans have decreased, but the gap has been filled by the growth of PPOs, which offer more flexibility in choosing providers. For example, HMO and PPO memberships in employer-sponsored plans were roughly equal in 1996 at 31 percent and 28 percent, respectively, of covered workers. As seen in Figure 12-1, in 2010, 58 percent chose a PPO, compared to just 19 percent choosing a HMO.

Total HMO enrollment stood at 75 million in 2008. Although percentage enrollments are highest in the Northeast and the West (California alone accounts for 22 percent of the national HMO total), HMOs and PPOs have increased their market penetration in other geographic areas. KFF data regarding the composition of HMOs indicate that mixed and IPA models dominate, accounting for nearly 60 percent of HMO membership.

### The Economics of Managed Care

The erosion of barriers to prepaid plans, coupled with some provider markets that characterized by economic profits, gives us the key elements to understand the rationale for managed care. Figure 12-2 starts with consumer demand  $D_f$  under fee-for-service insurance. Assuming that the price is constant at  $P_f$ , total expenditures,  $P_f Q_f$ , are represented by the larger rectangle. These total expenditures also would represent the expected insurance component of the individual's wage package.

A managed care option by constraining choice of provider as well as various coverages is viable only if it reduces expenditures. By exerting market power over suppliers, HMO managers may lower prices from  $P_f$  to  $P_m$ . A price reduction is possible when providers have been earning economic profits.<sup>3</sup> In effect, managed care reduces or eliminates those economic profits.



**FIGURE 12-2** Treatment and Expenditures Under Managed Care

<sup>3</sup> From a provider's perspective, such as a physician or hospital treating both FFS and MCO patients, its pricing decision is very similar to one involving price discrimination. Chapter 17 formally covers price discrimination within the context of pharmaceutical products. Here, the provider's demand from the MCO market is likely to be far more elastic than its demand from the FFS market because the MCO can contract with other doctors or hospitals if the provider tries to raise rates for the MCO's enrollees. As shown in Chapter 17, price will be lower in the market with a more elastic demand.

Demand also may be reduced from  $D_f$  to  $D_m$  by reducing inpatient care, by limiting length of stay, minimizing supplier-induced demand, and, in general, by encouraging more cost-effective care through the use of information technology and financial incentives to providers. The resulting expenditures are shown by the smaller rectangle representing total expenditures,  $P_m Q_m$ .

It follows that managed care trades some constraint of consumer choices for lower per unit prices for care. As drawn in Figure 12-2, both decreased price per unit of care and decreased quantity of care contribute to decreased expenditures. Note, however, that a natural response to decreased prices is to increase quantity demanded. Total consumer expenditures will fall as long as the price decreases are not fully offset by increased quantity demanded.

## MODELING MANAGED CARE

The changed incentives from managed care have caused many to predict success in constraining utilization and cost of care. The underlying logic behind this proposition is intuitive. When providers agree to handle all of their patients' health care needs for fixed, pre-arranged fees, the providers accept and bear a substantial part of the financial risk. By bearing the risk of delivering services at a fixed premium, managed care plans have strong incentives to reduce excessive care. The fixed premiums would also seem to provide incentives to offer forms of preventive care that are cost-effective from the care managers' perspectives. Such preventive care could decrease the need for more subsequent (expensive) curative care.

Will managed care plans provide enough care or the right types of care? The following analyses, developed by Goodman and Stano (2000), treat managed care plans as "staff model" HMOs where physicians are HMO employees.<sup>4</sup>

### Modeling Individual HMOs

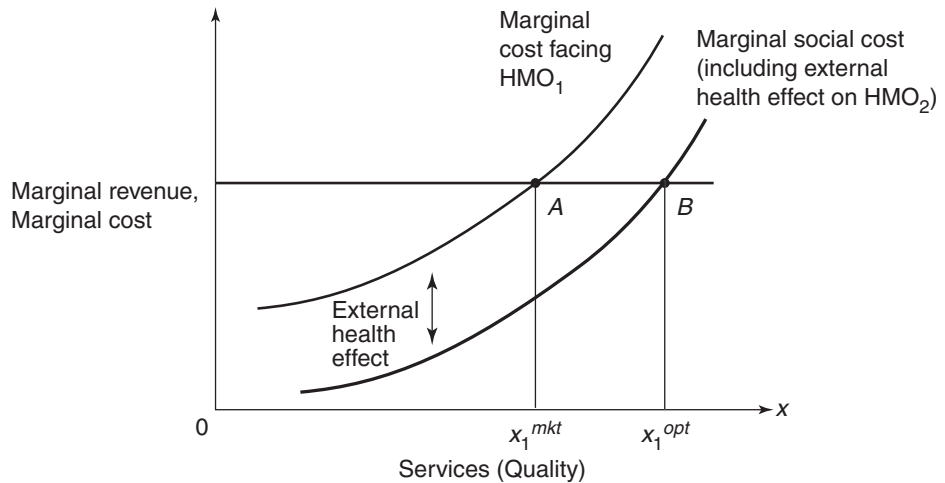
Individual HMOs need to determine the number of consumers to serve, or quantity, and the level of service to provide, or quality. One might compare them to urban suburbs, which set the qualities of public schools and the sizes of the police forces, for example, by the preferences of their residents and by the costs of providing them. Just as suburbs provide menus of public services, HMOs provide menus of care in inpatient and outpatient settings. To keep things simple, we will assume that HMOs provide only one type of service (visits), and that the HMOs are differentiated in quality by how many visits each offers.

We assume that HMO treatment costs are related to member health status, which is a function of care received at the HMO and elsewhere, as well as other factors including diet, environment, and lifestyle. People's long-term health relates in part to short-term decisions about how much care to provide. This follows directly from the long-term nature of human capital, most particularly health capital. Thus, because patients live for many years, treatment decisions at one HMO may affect treatment costs at other HMOs. Moreover, HMOs have the dual incentives of keeping people healthy and attracting healthy people.

An HMO's total annual costs are higher if it provides more services per enrollee (quality) or if it has more members (quantity). Having healthier members lowers HMO costs. Assume that health care works—that health is related positively to the level of services by all providers. Because at any time in the future these individuals may be members of a given HMO, treatment of the whole population by all providers affects the average health status of the population. This will influence the costs for these providers in that better health will lower costs.

---

<sup>4</sup> Economists have yet to develop a distinct generalized theoretical model of managed care. This section describes some simple approaches that provide useful insights. See Olivella and Vera-Hernández (2007) and Brekke, et al (2010) for more sophisticated contributions that attempt to deal with quality and differentiation across managed care plans.



**FIGURE 12-3** Externality Model of HMO

### How Much Care?

The previous section indicated that profit-maximizing HMOs might not recognize the system-wide impact of health services on consumers' health. Although an individual HMO seeks to maintain the health of its own clientele, it does not recognize that its decisions may affect the costs of other HMOs. We call this an externality because it is a benefit that affects others but is not considered by any individual HMO.

Suppose instead that an HMO has clinics at a number of locations, referred to as HMO<sub>1</sub>, HMO<sub>2</sub>, and so on. Although HMO<sub>1</sub>, for example, may not account for this possibility, the HMO entrepreneur, to maximize profits, must consider the impacts of health services on the health of others elsewhere.

We see this in Figure 12-3. Without the externality, HMO<sub>1</sub> optimizes at point A, providing quality level  $x_1^{mkt}$ . Level  $x_1^{mkt}$  is economically inefficient, however, because it does not account for the fact that improved treatment at HMO<sub>1</sub> lowers costs throughout the system.

The optimal quality level of  $x_1$  is at point B, or  $x_1^{opt}$ . This occurs by recognizing the downward shift in the marginal cost curve by a factor that reflects the effect of the health externality on the costs facing HMO<sub>2</sub>, HMO<sub>3</sub>, and so on, as well as HMO<sub>1</sub>. Level  $x_1^{mkt}$  indicates an inefficiently small level of HMO services, and by implication a substitution of non-HMO and/or non-health care inputs (such as the patient's own time) for the HMO care.

Although the potential loss of patients may influence treatment decisions of FFS providers as well as HMOs, the capitation method of payment to HMOs makes the problem of potential disenrollment particularly important. Fee-for-service providers are paid for each unit of care. Aside from uncollectible bills, they do not risk losses on services provided currently or in the future.

In contrast, by integrating insurance with the provision of health care, the HMO receives a fixed payment per enrollee to cover costs in the current period, and over time, for those who remain enrolled. Unlike FFS care, where payment in every period is likely to cover costs, the HMO must consider the timing of expenditures and the financial losses of overspending on patients who may disenroll. One way for an HMO to protect against long term losses attributable to disenrollment is to economize on care for those currently enrolled.

### What Types of Care?

In this section, we consider the long-term consequences of potential movement among HMOs on decisions among treatment practices. If an HMO receives constant revenue per patient each period,

its goal is to minimize costs. Patients stay in the HMO for two periods, which might be considered as early and late in their lives. The HMO may offer:

- High-tech, possibly capital-intensive procedures leading to high Period 1 costs and zero Period 2 costs
- Low-tech, less capital-intensive procedures, leading to low costs in both periods

Because the revenues are the same with either high-tech or low-tech care, the HMO's problem is to minimize costs, using the cheaper (over time) of the two procedures.

Suppose that increased competition through increased choice raises disenrollment rates. With higher disenrollment rates, low-tech continuing care becomes the more financially viable option, even if the present discounted values are equal and even if high-tech treatment is more economically efficient in producing health. Here, HMOs protect themselves against future disenrollment by reducing current costs through (low-cost) continuing care rather than high-tech treatment.

In evaluating how much, and what types of, care HMOs offer, we see that the HMO faces an economic externality because it cannot capture fully the gains of its treatment over time. As a result, it will offer less care and lower-tech care than FFS plans. As noted, if the HMO were to merge with others forming a larger network, the larger firm might internalize this positive externality because the network owners could expect that clients who leave one HMO might join another HMO within the network. The receiving HMO would then take advantage of the now-healthier clients who had benefited from their earlier treatment.

### Framework for Prediction

Clearly, this model simplifies the situation because many forms of managed care exist providing a myriad of services. It provides a framework, however, for addressing possible HMO cost savings relative to FFS plans. FFS plans encourage overutilization to the point where marginal private benefits can be far less than marginal costs. HMOs are widely believed to discourage this deadweight loss and other forms of overutilization, such as supplier-induced demand. To evaluate utilization, however, one must control for the health of the client population, which might be impacted by HMO "cream-skimming" of the healthier clientele.

This simple model provides predictions that are consistent with the rising conflict between HMOs and their members. The popular press reports stories about inadequate levels of services provided by HMOs, and the unavailability of expensive, high-tech treatment options.

### Where Managed Care Differs from FFS—Dumping, Creaming, and Skimping

We recognize that one of the key differences between HMOs and FFS plans involves the form of payment. Hospitals and other FFS providers are paid for each treatment in order to cover costs. HMOs are paid fixed rates per person irrespective of the amount of treatment used. These differences have led to discussions within the medical and policy communities regarding three purported practices.

1. **Dumping.** Refusing to treat less healthy patients who might use services in excess of their premiums.
2. **Creaming.** Seeking to attract more healthy patients who will use services costing less than their premiums.
3. **Skimping.** Providing less than the optimal quantity of services for any given condition in a given time period.

Because the HMO's costs will depend on the average health of its clientele, practices such as dumping (unhealthy patients) and creaming (healthy patients) may occur. If the HMO can identify and dump patients who are sicker, it can improve the average health of its clientele and hence lower its costs. As noted in Ellis (1998), the patient and the public at large may not notice this decision.

The HMO may honestly say that it is not equipped for this kind of case and that the patient would be better served elsewhere. This behavior contrasts with cost-based FFS care. Inasmuch as every penny spent on even the most severe case theoretically is paid back to the hospital, the cost-based hospital will not dump anyone.

Creaming is the practice of seeking out or emphasizing low-severity patients. The HMO benefits from creaming because all patients of sufficiently low-case severity require few hospital services, so that premiums for these patients exceed their costs. Some patients believe that their malady is so mild that the hospital care is hardly worth the trip; these patients correspond in practice to the young, vigorous, and healthy people that HMOs prefer to have in their service populations. The FFS provider also creams because its reimbursement covers costs for each of its patients.

Skimping involves cutting back on services to the point that patients' welfare is reduced. FFS providers will not skimp because each nursing hour, electronic instrument, and surgery that they employ will be reimbursed. In contrast, skimping provides the opportunity for the HMO to increase profits. Because the HMO has received the premium in advance, reducing the amount of services will lower costs and may raise profits. The uncertainty regarding profits occurs because reducing the benefits for a given severity of illness may cause the HMO to lose some patients to FFS plans or to other HMOs.

### Equilibrium and Adverse Selection in a Market with HMOs

Providers may have incentives for seeking patients, but patients also may select themselves into particular types of care. Cutler and Reber (1998) demonstrate the potential adverse selection of sicker consumers toward FFS care and healthier patients toward HMOs. This work also explains how HMO penetration can respond strongly to small changes in relative pricing.

Suppose that Jeff has just graduated from college and landed a job with Santa Fe Futons. Santa Fe offers its employees a choice of membership in one of two health plans. The HMO charges a flat prearranged price  $P_{HMO}$ ; the deluxe care FFS plan features deductible  $D$  and coinsurance rate  $r$ . Letting the severity of the illness,  $s$ , be measured by the expenditures required to treat at the FFS level, Jeff calculates the extra cost  $E$  of FFS to be:

$$\begin{aligned} E &= \text{FFS cost} - \text{HMO cost} = (\text{Deductibles} + \text{FFS copay}) - \text{HMO price} \\ E &= (D + rs) - P_{HMO} \end{aligned} \quad (12.1)$$

Is the extra cost,  $E$ , to use FFS, worth it, or should Jeff use HMO care instead?

To evaluate Jeff's decision, we calculate market values for the terms in Equation (12.1) and then compare term  $E$  to the value received from FFS care. If this extra cost of using FFS exceeds the value to Jeff of FFS, he will choose HMO care. If not, he will choose FFS. Under perfect competition, both FFS and HMO providers earn zero profits in the long run. FFS provider profit  $\pi_F$  for the average patient is written as:

$$\pi_F = D + \bar{r}\bar{s} - \bar{s}_F = 0$$

indicating that the firm collects patient deductibles  $D$  and patient copayments,  $\bar{r}\bar{s}$ , and spends  $\bar{s}_F$  for treatment (where the "bars" indicate FFS and HMO averages). Re-arranging terms, this means that the market deductible is:

$$D = (1 - \bar{r})\bar{s}_F$$

For the HMO parameters, we again assume that in the long run, competition will drive profits to zero. Cutler and Reber assume that HMOs achieve efficiencies in providing care, as well as restricting the amount of hospital care provided, reducing the cost of care. They summarize the efficiencies



using parameter  $\alpha$ . With no cost reductions,  $\alpha$  would equal 1; Cutler and Reber assert that the real-world value of  $\alpha$  is about 0.9. As a result, premium  $P_{HMO}$  must cover average payments for condition  $\bar{s}_H$ , or  $\alpha\bar{s}_H$ , so that:

$$\pi_H = P_{HMO} - \alpha\bar{s}_H = 0$$

Substituting these terms into (12.1) provides the following expression for Jeff:

$$E = [(1 - \bar{r})\bar{s}_F - \alpha\bar{s}_H] + rs \tag{12.1'}$$

Thus,  $E$  consists of a market-determined constant (in brackets) plus an increment of expenditures,  $rs$ , depending on the severity of the condition. It is reasonable to assume that the market-determined term is positive, so that  $E$  is the upward-sloping line in Figure 12-4.

To this point, we have looked at extra costs of FFS care. We also recognize that Jeff, like many, may see an additional benefit in FFS, the benefit of being able to choose one's own doctor. Cutler and Reber believe that this additional benefit would increase with the seriousness of the illness, in that people with serious illnesses would prefer to choose their own physicians. If so, the additional benefit of FFS starts at the origin (if  $s = 0$ , there is no benefit) and rises as  $s$  increases. We label this curve  $V$  in Figure 12-4.

Start at expected level of care  $s = 0$  where Jeff is young, healthy, and does not expect to use much service. If he were to join the FFS plan, he would have to pay the average FFS client expenditure, which exceeds the average HMO fee (the term in brackets in equation 12.1'), plus the share  $r$  of the services  $s$  that he uses. Because the extra FFS costs are higher than the value  $V$  that he puts on them (the  $E$  curve is above the  $V$  curve), Jeff chooses to join the HMO. If Jeff has a chronic condition requiring ongoing treatment, he recognizes that each unit of ongoing treatment is subsidized by the FFS plan at the rate  $(1 - r)$ . With increasing severity of treatment,  $s$ , FFS becomes a more attractive option as  $V$  exceeds  $E$ .

At severity level  $s'$ , Jeff would be indifferent between the HMO and the FFS plans because the extra value of FFS just equals the extra cost. If all consumers were similar to Jeff in everything

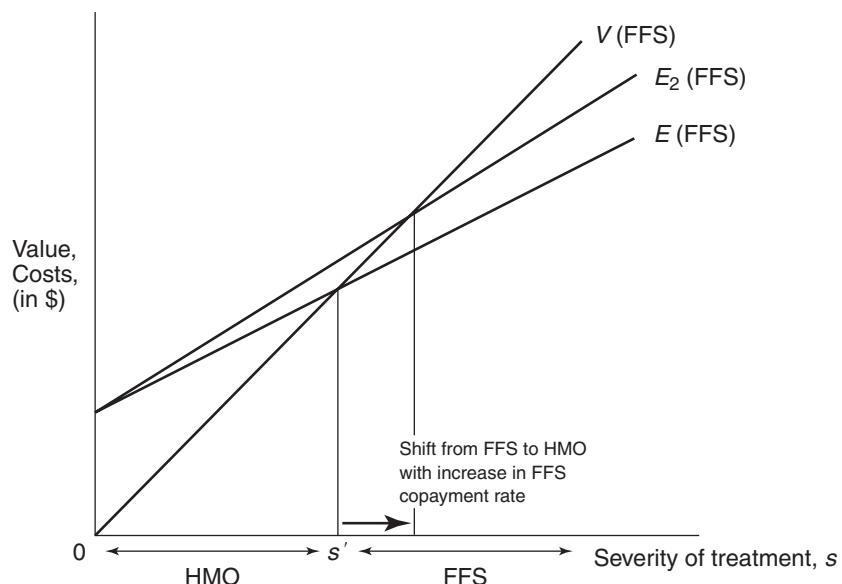


FIGURE 12-4 Selection into HMO and FFS Settings

except their health, consumers with expected severity less than  $s'$  would join the HMO, while those expecting severity greater than  $s'$  would choose FFS.

Suppose the FFS plan increased its coinsurance rate  $r$ . Then the FFS plan becomes less attractive; this can be seen by rotating the  $E$  line counterclockwise to  $E_2$ . Consumers who were previously indifferent between HMO and FFS plans will shift to the HMOs. The younger and healthier HMO patrons are now joined by some of the (not as young and not as healthy) former FFS patrons. As a result, the healthiest among the former FFS patrons become the sickest HMO members and the average severity of illness in both plans increases.

## HOW DOES MANAGED CARE DIFFER?—EMPIRICAL RESULTS

Economic and organizational theories have suggested that managed care will differ from more traditional fee-for-service plans. One might predict that managed care organizations will spend less per member, reducing health care costs. Theory would also predict, however, that if fewer resources are used, quality of care may also suffer. Policymakers have considerable interest in whether this theoretical proposition is true. Researchers early on reported that total costs—that is, the sum of premiums and out-of-pocket expenses—were from 10 to 40 percent lower for HMOs. They attributed the cost differences largely to lower hospitalization rates, not to lower ambulatory care-use rates.

### Methodological Issues—Selection Bias and Quality of Care

Conceptually, it would seem fairly simple to compare health care costs in managed care and fee-for-service plans. Researchers would collect data on cost of care across a wide spectrum of the population. Controlling for items such as patient age (older people have higher costs) and existing health status (sicker people have higher costs), one could use multiple regression statistical methods to compare costs.

Two major issues complicate the comparison—selection bias and quality of care. The methods previously discussed work only if patients are randomly assigned to either HMO or FFS treatment. Analysts worry that this random assignment does not exist in the real world. On the one hand, HMOs offer comprehensive benefits and so they may attract and retain sicker members. If we do not address this feature, studies may make HMOs look more expensive than they really are. On the other hand, HMOs may attract disproportionately younger members and families who tend to be healthier, and for whom the costs of care tend to be relatively lower. Studies that ignore this problem may make HMOs look less expensive.

Does managed care offer quality of care that is comparable to care under FFS? Managed care provides incentives to reduce the costs of care. Does it also provide incentives to cut corners by reducing the quality of care? Although some consumers may choose to pay less for lower quality care (just as some buy cheaper tires or cheaper cuts of meat), it is essential both to measure quality and to control for quality differentials in evaluating differences in health care costs.

The definition of quality is by no means obvious. Cutting health care quality would likely lower costs in the short term, but it might increase the longer-term costs if patients required additional services later. Furthermore, if information about quality were available to consumers, lowering quality would tend to erode demand. Despite the importance of quality-of-care issues to health care analysis in general and to analysis of managed care in particular, quality issues are difficult to resolve.

Quality may range from consumer perceptions of the provider-patient relationship to the outcome effects of health care on health status. Donabedian (1980) provides three general descriptors:

1. **Structure.** The quality and appropriateness of the available inputs and their organization.
2. **Process.** The quality of the delivery of care.
3. **Outcome.** The ultimate quality of care but the most difficult to measure scientifically.

### Comparative Utilization and Costs

Due to the evolving nature of both managed care and fee-for-service provision, some of the previous studies are more interesting for their historical perspective than for their current applicability. Luft (1978, 1981) found that HMO enrollees, especially prepaid group practice members, had lower hospitalization rates. No clear evidence showed that these lower rates were attributable to reductions in the less important, discretionary procedures. Furthermore, the evidence at hand could not dismiss the possibilities that biased self-selection of HMO membership or underutilization in HMOs was responsible for the observed differences.

Arnould and colleagues (1984) confirmed Luft's conclusion that length of stay was not significantly different between the HMO and the FFS patients. They also found that the use of surgeon visits, as well as lab charges, per patient were lower for the HMO users (significantly lower for hysterectomy and appendectomy), although total hospital charges were significantly lower for the HMO patients only in the case of appendectomies. Thus, although differences occurred in costs of elements of hospital care, no strong case could be made to conclude that HMOs produce hospital care more cheaply overall.

### The RAND Study—A Randomized Experiment

In the RAND Health Insurance Experiment, patients were assigned randomly to different plans in a controlled experiment, thus apparently eliminating selection bias. Would HMO costs still be lower under such circumstances?

The RAND study (Manning et al., 1984) compared HMO and FFS patients in the Puget Sound area, where 1,580 individuals were assigned randomly to either an FFS physician of their choice or to the Group Health Cooperative (GHC) of Puget Sound—an HMO in Seattle, Washington. The 431 FFS individuals were in one of four groups:

1. Free care
2. 25 percent of expenses up to a maximum out-of-pocket liability of \$1,000 per family
3. 95 percent of expenses up to a maximum out-of-pocket liability of \$1,000 per family
4. 95 percent coinsurance on outpatient services, up to a limit of \$150 per person (\$450 per family)

In addition to the experimental GHC group of 1,149 persons, a control group consisted of a random sample of 733 GHC members who had been enrolled for at least one year.

Total expenditures per person were \$439 for the experimental group, including out-of-plan use (which may be substantial), compared to \$609 for the free care FFS group (group 1). As seen from Table 12-2, ambulatory utilization, as indicated in the columns labeled "Face-to-Face Visits" and "Preventive Visits," was about the same. Thus, the 39 percent increased spending for FFS members (or 28 percent reduction for GHC) was due largely to a much higher admission rate and increased hospital days per person. The study could not pinpoint the reasons for GHC's lower hospital use.

To put the potential cost savings into better perspective, note that the use rates for the experimental HMO patients did not differ materially from the last two categories of FFS, the categories of "95 percent" coinsurance (group 3) and "Individual deductible" (group 4). Thus, for some population groups, a shift to HMOs would not lead to savings, although the cost savings for other population groups might conceivably be important.

### Recent Evidence

In a series of studies, Miller and Luft (1994, 1997, 2002) summarized findings regarding quality of care, utilization, and customer satisfaction. In the late 1980s and early 1990s, managed care (compared to FFS) plan enrollees received more preventive tests, procedures, and examinations (such as cancer screening; pelvic, rectal, and general physical examinations). Outcomes on a wide range of

**TABLE 12-2 Annual Rates of Admission and Face-to-Face Visits**

Plan	Admission Rate	Hospital Days	Face-to-Face Visits	Preventive Visits
	Per 100 Persons		Per Person	
GHC experimental	8.4 (0.67)	49 (9.6)	4.3 (0.14)	0.55 (0.02)
GHC control	8.3 (1.01)	38 (9.0)	4.7 (0.17)	0.60 (0.02)
Fee for service				
Free	13.8 (1.51)	83 (26)	4.2 (0.25)	0.41 (0.03)
25%	10.0 (1.43)	87 (28)	3.5 (0.35)	0.32 (0.03)
95%	10.5 (1.68)	46 (9.9)	2.9 (0.34)	0.29 (0.04)
Individual deductible	8.8 (1.20)	28 (5.1)	3.3 (0.33)	0.27 (0.03)

(Standard errors)

Source: Manning et al. (1984). Copyright © 1984, Massachusetts Medical Society. All rights reserved.

conditions (including congestive heart failure, colorectal cancer, diabetes, and hypertension) were better or equivalent to those using FFS plans. HMO enrollees were less satisfied with quality of care and physician-patient interactions but more satisfied with costs.

Their 1997 article found that HMO plans and providers cut hospitalization and use of more costly tests and procedures, often with little visible effect on quality of care “given the high prices of the indemnity insurance/fee-for-service system.” However, simply carrying out the same clinical processes but with fewer resources can negatively affect quality of care in some cases, such as Medicare home health care.

In their 2002 review, which covered the period 1997–2001, Miller and Luft included HMOs and some mixed models, but excluded studies purely on PPOs. Table 12-3 summarizes their analysis along six dimensions: (1) quality of care, (2) access to care, (3) satisfaction, (4) prevention, (5) length of stay, and (6) use of expensive resources. In the first line, for example, 14 studies showed better (HMO) quality, 15 showed worse quality, and 18 were in the middle. The overall results summarized in Table 12-3 are similar to those in the two previous articles. Compared with

**TABLE 12-3 HMO Plan Performance Update: An Analysis of Published Literature: 1997–2001**

Indicator	Favorable to HMOs	Mixed	Unfavorable to HMOs	Total
Quality of care	14	18	15	47
Access to care	2	4	4	10
Satisfaction	0	3	8	11
Prevention	7	3	0	10
Length of stay	5	5	0	10
Expensive resources	8	7	0	15

Source: Miller and Luft (2002).

**BOX 12-1****What Do HMOs Actually Do?**

There seems little doubt that HMO expenditures per member are substantially lower—sometimes 30 to 40 percent lower—than expenditures in traditional indemnity plans. Are these reductions achieved mainly by managing access and utilization, and lowering payments to providers that reduce their economic profits? What is the role of risk selection and quality? Several contributions provide considerable insight into these challenging questions.

Using methods described earlier in this chapter to decompose differences between FFS and HMOs, Cutler, McClellan, and Newhouse (2000) focused on Massachusetts patients with newly diagnosed heart disease, both those with heart attacks (acute myocardial infarction), which are relatively expensive to treat, and those with less severe forms of ischemic heart disease. By selecting one condition, the authors avoid some of the problems associated with aggregation across conditions. By studying heart disease, risk selection is minimized because even if a patient chose a plan based on some expectation of heart disease, the choice would not likely be based on expectations regarding the severity of the disease.

The study results are very clear and powerful: “Essentially all of the differences in reimbursement . . . [result from] differences in the prices paid for particular services, rather than differences in quantity or quality of services received” (p. 327). The authors caution about generalizing findings based on a life-threatening condition, such as a heart attack, where insurance status may have little effect on treatment. They also suggest that cardiac care is well known for providing sizable economic rents to both hospitals and cardiologists. Thus, the price effect may not be as large for other kinds of services.

Polsky and Nicholson (2004) also decompose the differences between HMOs and non-HMOs into differences in risk selection, utilization, and prices. They use a national sample for overall expenditures which were \$188 (9.3 percent) lower for HMO members. Consistent with the Cutler study, lower prices paid by HMOs were the main determinant of expenditure differences. Prices were actually \$269 less for the HMOs, or more than the expenditure difference because their utilization was \$81 higher. Risk selection accounted for only \$35 of this amount.

With their consistent results, the two studies provide a clearer picture of the role of HMOs. HMOs are quite successful in using their leverage to negotiate lower fees, and they do this without obvious reductions in quantity or quality. Risk selection also is not a major factor. However, it must come as a disappointment to many proponents of managed care that there is no indication that HMOs have been able to fulfill their promise and potential of applying information technology and better management to improve the process of health care delivery.\*

\* We do not mean to imply that MCOs have not engaged in major effort to improve quality, but, until recently (Landon et al., 2008), this effort has not been systematically described. Recent work by Wu (2009) also provides new information about hospital price discounts. Large plans and those with a greater ability to channel patients to alternative hospitals are able to extract higher discounts.

non-HMOs, HMOs had similar quality of care, more prevention activities, less use of hospital days and other expensive resources, and lower access and satisfaction ratings.

Recent work confirms some of these findings (see also Box 12-1). Rizzo (2005) concludes that HMO patients get substantially more preventive care than FFS patients and that this is not due to a selection effect that patients/physicians with preferences for preventive care are more likely to choose HMOs. After accounting for self-selection, Deb and colleagues (2006) estimate that an individual in a managed care plan would receive about two more physician visits and 0.1 emergency room visits per year than had the same person enrolled in a nonmanaged care plan. Because of these and other research developments, we now have a much better understanding of the performance and role of HMOs.<sup>5</sup>

<sup>5</sup> Glied (2000) summarizes many articles that examine impacts of managed care on costs and on utilization. Her tables discuss the populations, comparison groups, controls for differences in patient characteristics, and impacts on treatment aspects, such as charges and length of stay.

## GROWTH IN SPENDING

Analysts believe that managed care reduces utilization, especially of hospital care. A different but related question is whether managed care organizations also have lower growth rates in spending. If they do, a continued shift toward managed care will result not only in reductions in spending levels, but also in the long-term rate of increase.

It is important to provide a framework for discussing the relationship among FFS, MCOs, and total costs, particularly because terminology can be confusing. Suppose we are concerned about costs per person for treating a particular illness over three periods. Let us assume that people use either FFS or MCO and that the population is fixed. We can calculate the total treatment costs as:

$$\begin{aligned} \text{Total treatment costs} &= (\text{Number in FFS}) \times (\text{FFS costs/FFS enrollee}) \\ &+ (\text{Number in MCO}) \times (\text{MCO costs/MCO enrollee}) \end{aligned} \quad (12.2)$$

Dividing both sides by the total population, we get:

$$\begin{aligned} \text{Treatment costs/Person} &= (\% \text{ of population in FFS}) \times (\text{FFS costs/FFS enrollee}) \\ &+ (\% \text{ of population in MCO}) \times (\text{MCO costs/MCO enrollee}) \end{aligned} \quad (12.3)$$

Suppose, in Period 1, that FFS treatment costs \$2,000, MCO treatment costs \$1,000, and that 60 percent use FFS and that 40 percent use MCO care. The treatment costs per person will be:

$$\text{Treatment costs/Person} = (0.6 \times \$2,000) + (0.4 \times \$1,000) = \$1,200 + \$400 = \$1,600$$

This is in column 1 of Table 12-4 as Period 1.

If FFS and MCO costs were to stay constant and patients were to switch from the more expensive FFS to the less expensive MCO, costs per person would fall. As noted in Table 12-4, if a 10 percentage-point movement occurs from FFS to MCO, treatment costs per person would fall because 1 in 10 people would be substituting a (MCO) treatment that is half the price of the other (FFS treatment). As calculated in Table 12-4, the total costs per person would fall by \$100, or 6.25 percent.

What is less obvious is that unless the population shift from FFS to managed care continues, cost reduction and cost containment may be difficult. Suppose that in Period 3 the percentage in FFS remains at 50 percent, but the costs of both FFS and MCO increase by 10 percent. The total cost per person accordingly increases by 10 percent from \$1,500 to \$1,650. If total costs per enrollee in each sector were to continue to increase by 10 percent and no change occurred in the MCO market share, the total costs per person on aggregate also would increase by 10 percent. This occurs

**TABLE 12-4 Managed Care and Cost Containment—An Example**

	Period 1	Period 2	Percent Increase	Period 3	Percent Increase
Fraction of population MCO	0.4	0.5		0.5	
FFS costs per enrollee	2,000	2,000	0.00	2,200	10.00
MCO costs per enrollee	1,000	1,000	0.00	1,100	10.00
Total costs per person	1,600	1,500	-6.25	1,650	10.00

irrespective of the fact that a larger proportion of the population (Period 3) is now being treated in managed care settings than was being treated in Period 1.

This analysis provides important insights into the impacts of managed care plans on health care costs and health care cost increases. To the extent that large shifts of insureds into managed care have led to lower-cost treatments, there may have been one-time cost decreases relative to what they would have been. However, if managed care and FFS plans face the same cost inflation for the services that they offer apart from the one-time decreases, overall cost inflation is unlikely to abate.

Early studies by Luft (1981) and by Newhouse and colleagues (1985) found the growth rate of HMO spending to be roughly the same as the growth rate under FFS, and recent studies have not contradicted those findings. In the 1990s, policymakers relied on shifts into managed care to reduce health care costs and their growth rates. With most persons outside the Medicare program now covered by a managed care plan, such shifts can no longer continue. In the absence of other innovations, costs will continue to inflate, albeit from lower bases.

## COMPETITIVE EFFECTS

Up to this point, we have concentrated on the direct effects of managed care and managed care organizations. We have asked what managed care organizations look like, whether they provide less costly care, and whether they provide different quality care. In this section, we address the indirect effects of managed care through the market mechanism. By indirect, we mean that existing health providers must respond to competition from the managed care sector.

After discussing some of the theoretical concerns, we will evaluate the impact of competition in three ways. We will look at the impact of managed care on hospital markets, on insurance markets, and finally on the adoption of technological change.<sup>6</sup>

### Theoretical Issues

The spread of alternative delivery systems can elicit substantially greater competition in other sectors only if competition is absent at the start. Otherwise, both providers and insurers would be operating at, or close to, their minimum costs of production. Although there is bound to be disagreement on the extent of the degree of market imperfections, most would agree that the insurance and provider markets are less than highly competitive.

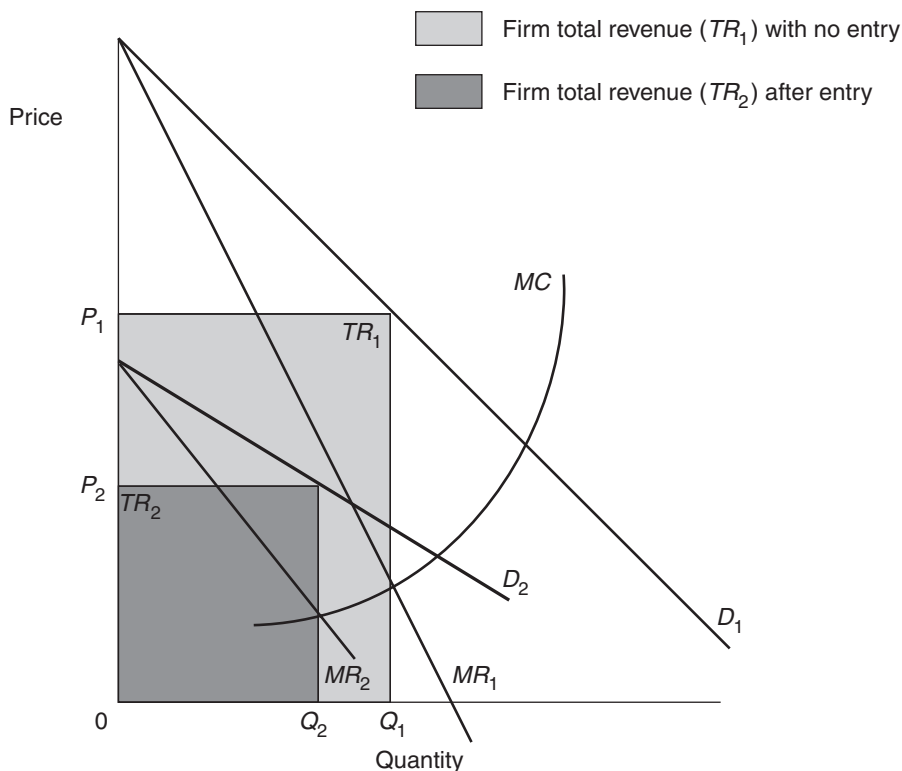
Consider the consequences of having a larger number of products and competitors to an existing monopolistic seller. In Figure 12-5,  $D_1$  is the current market demand and  $P_1$  is the monopoly profit-maximizing price (where marginal revenue  $MR_1$  equals marginal cost  $MC$ ) for each firm (average cost is omitted for clarity). The entry of other firms will have the following effects on each individual firm:

- Shift the demand curve to the left to  $D_2$
- Shift the marginal revenue curve to the left to  $MR_2$
- Increase the elasticity of demand at any price because there are now more competitors

With the same costs facing each firm, the new profit-maximizing price (where marginal revenue  $MR_2$  equals marginal cost) for each falls to  $P_2$ . If the decrease in firm demand is sufficiently large, it is possible that an individual firm will no longer be able to earn a competitive return at  $Q_2$ . This would occur if the demand curve shifts (due to the entry of competitors) so that it is everywhere below the firm's average cost curve.

---

<sup>6</sup> Competition can also affect quality and other dimensions of care. Scanlon and colleagues (2005) found that lower levels of HMO competition, as measured by the Herfindahl-Hirschman Index, actually produce better results on several quality dimensions. Higher HMO penetration was associated with higher quality. The study also found that plans that publicly report their data have higher quality than those that do not.



**FIGURE 12-5** Impact of Entry of Alternate Providers into a Monopolistic Market

The existing firm also may respond in other ways. It may attempt to reduce its administrative costs. More importantly, it may try to court customers by attempting to market plans that limit utilization of services, and hence the costs of the services, through various devices. These include utilization review and the adoption of health care plans with increased cost sharing. Of course, it could establish its own HMOs and PPOs, further increasing competition. It also could improve the quality of its service. Further, it may embrace forms of emerging nonprice competition, such as advertising and marketing.

Is this theoretical scenario applicable? Many items are likely to be affected by the spread of HMOs and PPOs. These include the provider and health insurance markets, the phenomenon of biased selection, the roles of employers, and the rate of innovation and diffusion of technology. To focus our discussion, we will limit it to three areas: (1) the impact of HMOs/PPOs on hospital markets, (2) their impact on insurance markets, and (3) their impact on the adoption of technological change.

### Managed Care Competition in Hospital Markets

A first step in looking at the impacts of managed care penetration is to examine the determinants of the penetration itself. What is it about particular markets that lend themselves to high levels of HMO activity? Dranove, Simon, and White (1998) use a demand-supply framework to address this question.

What are the demand characteristics? The authors view employers as the primary health insurance shoppers. Serving as agents for their employees, they seek the best deals.

What are the supply characteristics? The authors seek to identify market features that would reduce costs of payers to contract with providers. They expect that the supply of managed care will



be higher in markets where the MCOs can negotiate with and monitor the providers more cheaply. They also expect that excess hospital capacity may make hospitals more amenable to negotiation.

Regarding demand, higher managed care penetration accompanies more educated and more urbanized populations. Regarding supply, higher managed care penetration is related to lower percentages of physicians in solo practice, hospital market concentrations, and hospital occupancy. In short, managed care comes in where large numbers of half-filled hospitals are found. The authors express concern that the low rate of managed care penetration in more concentrated markets may imply anticompetitive behaviors, meriting antitrust considerations.

Early work viewed the hospital as competing for patients, physicians, or both, arguing that widespread health insurance allowed individual consumers and their physician-agents to be much less concerned about the price of care. Hospitals would compete on services and amenities rather than price. McLaughlin (1988, p. 207) argues that the “providers are responding not with classical cost-containing price competition but, instead, with cost-increasing rivalry, characterized by increased expenditures to promote actual or perceived product differentiation.”

Feldman and colleagues (1990) examined the Minneapolis/St. Paul area, addressing the degree of competitive bidding for HMO contracts, whether HMOs can get discounts, and whether HMOs tend to rely on low-price hospitals. They found that HMOs generally did not extract major discounts. In fact, price did not seem to be the major HMO consideration in the selection of hospitals with whom to affiliate. Rather, it was hospital location and the range of services that the hospital offered. There was no indication as to whether HMOs tend to affiliate with lower-priced hospitals.

Melnick and colleagues (1992) obtained hospital transaction prices negotiated by a large California PPO in 1987. Their regression analysis controlled for hospital characteristics such as ownership, teaching, Medicare and Medicaid demand, and market structure. Four key findings emerged from the analysis:

1. Controlling for other factors, the PPO paid a higher price to hospitals located in less competitive markets.
2. If the PPO had a larger share of the hospital’s business, it was able to negotiate a lower price.
3. The more dependent the PPO was on a hospital, the higher price the PPO paid.
4. Hospitals with high occupancy located in markets with high average occupancy charged the PPO higher prices.

Morrisey (2001b) terms the Melnick research “a watershed,” demonstrating that traditional concepts of competition could apply to hospital markets and potentially to other health care markets as well.

### **Managed Care Competition in Insurance Markets**

Commensurate with the spread of HMOs, PPOs, and various forms of managed care, the health insurance industry is changing rapidly. Many have noted that health insurers were previously lethargic in introducing innovative insurance products and in their cost-containment efforts. In addition, we consider the historic opposition by organized medicine, tax advantages and provider control of the Blues, and the constraints imposed by state regulation and antitrust laws, all of which previously inhibited change in the insurance market.

Frech and Ginsburg (1988) identified the dramatic changes that occurred after 1977 when the insurance market was divided about equally between the Blues and commercial insurers. The growth of HMOs and PPOs was accompanied by substantial increases in patient cost sharing, increased utilization review, and self-insurance (or self-funding as described earlier in this chapter) by many large firms.

In a self-funded plan, a Blue Cross and Blue Shield or another organization will act only as a third party in processing claims and providing other administrative services such as utilization review. More competition is introduced because self-insured firms have more control over their health plans and more direct interest in cost-containment measures.

Baker and Corts (1995, 1996) identify two conflicting effects of increased HMO activity on conventional insurance premiums:

1. **Market discipline.** HMO competition may limit insurers' ability to exercise market power, thus driving down prices, a standard competitive argument.
2. **Market segmentation.** HMOs may skim the healthiest patients from the pool, thus driving insurers' costs and prices up.

Their model suggests that if increased HMO penetration does lower the premium levels, the market discipline effect becomes relatively less important, so at higher levels of penetration, the market segmentation effect may raise premiums. Their empirical work verifies this hypothesis: Across metropolitan areas, market penetration up to 14 percent decreases premiums, but market penetration beyond that point raises them. Wickizer and Feldstein (1995) find comparable market discipline effects, although they do not investigate market segmentation effects.

Joesch, Wickizer, and Feldstein (1998) investigated nonprice impacts of HMO market competition. They found that increased HMO penetration reduced insurers' likelihood of increasing insurance deductibles, or "stop-loss" levels (the levels limiting losses to those insured). Moreover, groups located in markets with higher HMO enrollments were more likely to adopt utilization management or PPO options.

### **Managed Care and Technological Change**

The impact of managed care on technological change is also potentially significant in controlling health care costs. Increased incomes and changed coinsurance rates are insufficient to explain the increases in health care costs over the past 50 years. Most analysts attribute a major role to the advances in high-cost technologies fed by payment mechanisms that were at best indifferent to controlling costs. To the extent that managed care plans explicitly seek cost containment, one would expect careful monitoring of attention directed toward high-cost technologies.

Baker and Spetz (2000) compiled an index using 18 technologies available in 1983, including cardiac catheterization and neonatal intensive care units. They then aggregated hospitals within metropolitan areas and compared metropolitan areas on the basis of degree of HMO market shares. Fundamentally, they found that HMO market shares did not matter. Although they detected modest variations, no substantive differences were seen in technology at given points in time or in the dispersion of technologies over time.

Managed care plans, through their emphases on cost containment, would seem to be important vehicles for reining in the usage of high-cost-high-technology procedures and facilities, but research has not (yet) proved this. Medical facilities and medical practice styles change slowly, and it may be too early to see changes induced by the recent market penetration of managed care plans. Or it may be that the public demands high (and costly) technology, irrespective of who provides it.

### **THE MANAGED CARE BACKLASH**

In the first half of the 1990s, many managed care plans placed increasingly severe restrictions on patient choices, including prior approval for access to specialists and certain high-cost procedures. The results of a 1997 survey conducted by Blendon and colleagues (1998) documented the public's anxiety about the direction of managed care at that time. Only 34 percent of American adults who were surveyed thought that MCOs were doing a "good job," 51 percent believed that MCOs had decreased the quality of care, and 52 percent favored government regulation even if it would raise costs.

Concerned about timely access to care, California voters passed a law in 2002 intended to ensure that HMO members do not face undue delays in receiving medical attention. The rules were not approved until 2010 by the California Department of Managed Care but their specificity is unprecedented.

HMO members will face maximum waiting periods for nonemergency care, e.g., 48 hours for urgent care with no prior authorization, and 15 business days for nonurgent specialty care. While it is too early to assess the effects of the California measures on cost, quality and compliance, the “drive-through delivery” provides an example that has received extensive media, legislative, and scholarly attention.

“Drive-through delivery,” which refers to managed care’s movement in the early 1990s toward one-night hospital stays for mothers expecting a normal (or non-caesarean delivery) childbirth, became fodder for late-night talk-show jokes and ridicule of HMOs. Fueled by stories of instances in which children died shortly after the mother’s release, there was an enormous public outcry to require managed care plans to provide at least a second night of hospital care. The economics behind this example, however, are serious, and Jensen and Goodman (1999) provide us with an overview.

As recently as 1980, nearly 70 percent of mothers experiencing vaginal delivery had hospital stays of three days or more (Gillum, Graves, and Wood, 1998, Table R). Inpatient care is very costly. Yet, as long as hospitals received reimbursement for what they charged, they had no incentive to send the new mother home earlier. Almost certainly, the marginal benefits to the woman of being in the hospital for a third day did not measure up to the costs of keeping her there.

This cost inefficiency became apparent to HMO managers. Pressure to reduce the stay was considerable, and by 1995, the average length of stay for a mother with a vaginal delivery was 1.7 days, with 46.8 percent of all mothers staying one day or less. For the vast majority, home care (starting the second day) along with appropriate outpatient follow-up has become the alternative. Home care carries its own costs, requiring the woman’s time and usually assistance from family, friends, or hired caregivers. Nonetheless, it generally provides an appropriate level of care at a far lower cost than the inpatient care that it replaces.

The key phrase is “normal” childbirth. Medical care is not exact and mistakes are made. Managed care opponents have seized on cases in which a baby sent home the second day after birth developed an ailment and needed to return to the hospital or, worse still, died. Keeping the baby an extra day, they argued, would prevent these problems.

Many states passed legislation requiring insurers to cover at least two nights of hospital stay to all mothers with normal deliveries. Maryland’s 1996 Early Discharge of Mothers and Babies Bill guaranteed that mothers and babies have coverage in the hospital for 48 hours for a normal vaginal delivery and 96 hours for a normal caesarean delivery.

Liu, Dow, and Norton (2004) analyze the state length of stay mandates in 32 states, comparing the costs of the increased length of stay to the estimated health benefit, relying on infant mortality estimates that one infant life could be saved for each 1,400 normal newborns moved from early discharge (less than 30 hours) to longer lengths of stay. They find that for normal vaginal deliveries the average state law decreased early discharge (less than two-night stays) by 16 percentage points and increased average hospital charges, implying a \$1,281 cost per early discharge averted by the law. This could be converted to a crude estimate of \$1.79 million per life saved. They characterize their estimate as neither “highly cost-effective [n]or hugely cost-ineffective relative to estimates of the value of a life (often in the range of US\$ 1–10 million).”

Economists have also been interested in another issue closely related to the backlash. Did consumers respond to the restrictions by “voting with their feet”? Marquis and colleagues (2004/2005) examined HMO market penetration in two periods, 1994–1998 and 1998–2001, with the former representing the pre-backlash period. There was little evidence of substantial consumer exit and plan switching even in markets where consumers had more options.

Cooper et al. (2006) provide a more extensive set of results through their analysis of enrollments in HMOs and other types of plans over the period 1997–2003. They analyze both aggregate enrollments as well as enrollments by firm size. Aggregate enrollments remained stable until 2002, well after the most intense backlash years, because enrollment declines in large firms were offset by increases in smaller firms. Beginning in 2002, HMOs experienced sustained aggregate decreases in enrollment.

For various reasons including the moderation in the growth rate of health care costs in the late 1990s, many MCOs eased some of their restrictions. Rather than trying to curtail patient choices

and utilization, MCOs have engaged in other strategies to deal with quality and cost containment. One of the most visible is *pay-for-performance* (P4P). P4P describes incentive plans in which physicians and sometimes hospitals are given cash bonuses to meet guidelines for quality care and to implement evidence-based medicine. As a largely untested strategy, the value of P4P is still under debate. Box 12-2 provides evidence of its potential impacts.

We close this section with a revealing report by Fang and Rizzo (2010). There are widespread perceptions that managed care plans have become less restrictive but very little empirical evidence to support this proposition. Fang and Rizzo take advantage of surveys for 2000–2001 and 2004–2005 of large, nationally representative samples of physicians. The proportion of a physician’s practice revenues obtained from MCOs was used to measure managed care. The purpose of the research was to compare any changes over the two time periods of managed care and non-managed care in constraining the provision of physician care. Although the effects of managed care in limiting services in 2004–2005 declined somewhat relative to the earlier period,<sup>7</sup> the bigger story is that it declined relative to non-managed care. The authors conclude (p. 100) that “non-managed care plans appear to be getting more restrictive, similar to their managed care counterparts.”

These conclusions should not be surprising. We have already seen that, with the exception of Medicare, managed care dominates other health insurance sectors. It is likely that the remaining

## BOX 12-2

### Pay-for-Performance

Rosenthal et al. (2006) documented the extent to which HMOs adopted P4P by 2005. Their survey of a large number of HMOs indicates that more than half (52 percent) representing 81 percent of enrollees used P4P in their provider contracts, although much more so with physicians than with hospitals. Adoption of P4P was positively associated with HMOs that are nonprofits, those that use primary care physicians as gatekeepers, and those that use capitation to pay them.

P4P is still in an early stage of development, and there are wide variations in program design (Trude et al., 2006). Preliminary results from one of the largest efforts, known as the *Rewarding Results* program, indicated that financial incentives can motivate change if the rewards are substantial. Other evaluations are less encouraging. Rosenthal and colleagues (2005) compared a broad set of quality performance measures for a large health plan that introduced P4P for its California medical groups in 2003, with its plans in Oregon and Washington that did not subject medical groups to P4P. The results for three measures of clinical quality (cervical cancer screening, mammography, and hemoglobin A<sub>1c</sub> testing for diabetes) showed that the rates for each increased in California after P4P was introduced. However, only the cervical cancer screening rate increased faster than the rate for the Pacific Northwest medical groups.

In a more extensive follow-up to this research, Mullen, Frank, and Rosenthal (2010, p.85) concluded that P4P “may not necessarily have the dramatic or even predictable effects touted by its enthusiasts.” For example, appropriate medication rates for asthma even declined after P4P was introduced in California. The effects of P4P on health outcomes were also mixed.

Nevertheless, many analysts remain optimistic about the potential for P4P. The Centers for Medicare and Medicaid Services (CMS) have funded several multiyear P4P demonstrations, and some have shown cost savings and improved patient outcomes. A common criticism of U.S. health care is the lack of financial incentives for quality care. Although this criticism can be debated, the Tax Relief and Healthcare Act of 2006 mandated a P4P program for Medicare. The program, known as the Physician Quality Reporting System, is still voluntary, but the CMS paid an average bonus of nearly \$19,000 per participating professional practice in 2009. CMS reported an average increase of 10.6 percent over 2008 across 99 measures of performance.

<sup>7</sup> In another report using the same survey, Fang, Liu, and Rizzo (2009) did not find a change over the two periods in the assignment by HMOs of primary care physicians as gatekeepers.

non-managed care plans feel the same pressures faced by MCOs, as well as competition from MCOs. Thus most traditional FFS plans have incorporated some managed care features such as prior authorization, general utilization review, or specialty utilization review for mental health and other specialized services.

## CONCLUSIONS

This chapter has considered HMOs and other managed-care delivery systems that combine the functions of insuring patients and providing their care. We begin by describing managed care and providing the rationale for a government policy that promotes it.

Our discussion has emphasized that HMOs and other care managers have incentives to curtail costs because they serve as both insurers and providers. Thus the incentives for additional, less essential procedures are reduced, and we can show that systemwide even less care may be given than would be economically efficient. One key finding is that managed care organizations tend to reduce hospitalization—one of the most expensive components of health care costs. While other findings are mixed, little evidence suggests that the quality of the care provided in HMOs is inferior to FFS care. Another key finding is that MCOs have been able to reduce fees paid to providers.

In the early 1990s, cost pressures moved more users into managed care. By 2010, traditional fee-for-service (FFS) health care enrollment for covered workers had fallen to 1 percent, from 73 percent two decades earlier. Yet customers also rebelled against the more stringent cost controls of HMO plans, preferring what some analysts refer to as “managed care light”—as exemplified by PPO or POS plans. Less stringent controls were palatable with the deceleration of health care costs in the late 1990s. However, costs were increasing again at double-digit rates in the first years of the twenty-first century. The passage of PPACA in 2010 is also creating uncertainties in both private and public health insurance markets. To complicate matters further, there is growing awareness of widespread deficiencies and inconsistencies in health care quality. The enormous task facing health economists, policy analysts, and plan managers is to innovate and evaluate delivery models that promote a better balance between costs and quality.

---

## Summary

1. Managed care seeks to integrate what previously had been a non-integrated system of health care treatment. Such integration has the potential to reduce health care costs, but the integration is costly and may limit choice of provider and treatment options.
2. In HMOs, PPOs, and other MCOs, the functions of insurance and the provision of care are combined. In return for a prepaid premium, MCOs agree to provide enrollees with comprehensive health care over a given period.
3. By agreeing to handle all of a patient's health care needs for a fixed, prearranged fee, a provider is bearing a substantial part of the financial risk. By bearing such a risk, the managed care organization has a strong incentive to develop strategies for reducing excessive care and minimizing other inefficiencies.
4. HMOs control utilization and costs by imposing physician gatekeepers and requiring treatment within defined provider networks. PPO plans maintain the provider networks but do not require physician gatekeepers. Point-of-service (POS) plans maintain the physician gatekeeper role but do not require treatment within defined provider networks.
5. There has been a dramatic shift from FFS to managed care in employer sponsored plans. By 2010, only 1 percent of these workers had traditional FFS insurance, 86 percent were in managed care plans, and 13 percent had a high-deductible health plan.
6. A theoretical model of the HMO shows that, due to the impact of potential disenrollment that does not face providers in the fee-for-service sector, one might expect inefficiently low levels of care in the managed care sector.

7. If providers can charge different consumers different amounts, they can earn additional profits. Such price discrimination by providers is difficult under the contracts characterizing prepayment-based organizations because:
  - Providers will find it difficult to determine how much individual consumers value the services.
  - Prepayment-based organizations may be able to shop among providers, thus limiting the providers' monopoly power.
8. Managed care organizations typically provide comprehensive ambulatory and inpatient care, including routine office visits and preventive care, generally with low coinsurance or deductibles. Paperwork for patients is reduced, and uncertainty over their coverage is minimal. These features make membership attractive to consumers, especially to those who are concerned about out-of-pocket costs.
9. In contrast to FFS arrangements, the managed care enrollee's choices of providers and access to hospitals (aside from emergency care) are limited. Also, direct access to specialists may require referral from the patient's gatekeeper—the primary care physician.
10. There is a strong consensus that managed care reduces utilization, especially of hospital care. There is little evidence that the quality of care is inferior to the quality found in a FFS system.
11. The lower expenditures per enrollee under managed care are associated largely with the lower fees they are able to negotiate with some providers. Risk selection is not a major factor.
12. In addition to individual cost-related impacts of managed care organizations, there are also competitive impacts. Although higher market penetration of managed care does not always lead to lower hospital prices, it does appear to reduce insurance premiums.
13. MCOs, through their emphasis on cost containment, would seem to provide important opportunities for reining in the usage of high-cost-high-technology procedures and facilities. Research findings to date have not supported this hypothesis.
14. MCOs adopted more stringent restrictions on utilization in the early 1990s. There was a strong media backlash. Some evidence indicates that consumers did not react by “voting with their feet.”
15. Pay-for-performance (P4P) describes incentive programs in which providers, most often physicians, are given cash bonuses to meet quality performance targets. Many MCOs have adopted P4P but the preliminary evidence of its effects is not yet clear.
16. Traditional insurers have increasingly adopted managed care features and MCOs have become less restrictive. As a result, differences between managed care and traditional insurance have narrowed.

---

## Discussion Questions

1. What are the key elements that distinguish managed care from FFS plans?
2. What are the principal differences among HMO, PPO, and POS plans?
3. How do the economic profits (rents) that may be earned by some groups of providers enable MCOs to limit expenditures? What role does the price elasticity of demand play in this process?
4. Why is selection bias such an important issue in measuring HMO performance?
5. Discuss ways that managed care organizations may be able to reduce costs of care to their clientele.
6. Why do some critics argue that managed care organizations provide lower-quality care than FFS plans? Evaluate this possibility from a societal perspective.
7. After a large increase in membership, HMO enrollments flattened in the late 1980s and many HMOs suffered financial difficulties. How might this be explained according to what is known about the supply and demand for HMOs?
8. If everyone chose to join an HMO, would average HMO expenditures per case tend to rise or fall? Would national health expenditures tend to rise or fall?
9. What features of managed care organizations tend to inhibit or discourage people from joining? What features tend to attract people? Discuss the advantages and disadvantages of managed care enrollment.
10. Why is the growth of managed care plans a relatively recent phenomenon? Describe governmental policies and practices that have encouraged managed care organizations and inhibited them.
11. If traditional FFS leads to demand inducement, what constrains the HMO from underproviding care?
12. Explain how the availability of alternative delivery systems is expected to produce competitive effects throughout the health economy.

13. Discuss the ways that managed care organizations can influence the adoption of new technologies.
14. Some critics argue that providers do not have sufficient financial incentives to provide quality care. Describe some of the existing safeguards. Evaluate the potential role of P4P.
15. Differences between managed care and traditional insurance plans appear to be narrowing. Describe how competition and other market forces can account for this phenomenon.

---

## Exercises

1. Consider an HMO with a demand curve of the following form:  $Q = 100 - 2P$ . Suppose that its marginal and average costs were \$20. If the firm maximizes profits, determine its price, output, and profits.
2. In Exercise 1, if the firm must act as a perfect competitor, in the long run what will happen to equilibrium price and equilibrium output?
3. Consistent with Figure 12-2, assume that the FFS price was \$100 per visit and the average patient made eight visits per year. A competing managed care organization came in and charged \$80 per visit, providing seven visits per year.
  - (a) Calculate the change in total expenditures.
  - (b) Graph the FFS and the managed care market equilibria as was done in Figure 12-2. What do our findings suggest about demand for managed care compared to demand for FFS care?
4. Consider the discussion on adverse selection into HMOs and FFS care, as noted through equation (12.1') and Figure 12-4. Suppose that, on average, FFS clients bought \$2,000 in services ( $\bar{s}_F = 2,000$ ) and HMO clients bought \$1,500 in services ( $\bar{s}_H = 1,500$ ), with an efficiency factor of 0.9. The FFS plan charges a 10 percent coinsurance rate.
  - (a) Set up this problem graphically, labeling the  $E$  and  $V$  curves.
  - (b) If a client expects to spend \$250 on care, will he or she choose an HMO or an FFS plan? Why?
  - (c) At which value of  $s$  would the client expect to be indifferent between an HMO and an FFS plan? Why?
  - (d) How would your answer to parts (b) and (c) change if the HMO adopted a 20 percent coinsurance rate?
5. Assume that in Figure 12-5, so many providers entered the health care market that individual demand curves fell below the average cost curves. Draw the new equilibrium. What would happen to short-run profits in the health care market?
6. Exercise 5 discusses a short-run equilibrium in the health care market. With entry and exit into and from the market, graph and discuss the long-run equilibrium.
7. In Table 12-4, the market penetration for managed care rises from 40 percent to 50 percent between Periods 1 and 2 and stays at 50 percent. Suppose, instead, that it fell from 40 percent to 30 percent and stayed at 30 percent.
  - (a) What would happen to total costs and to rates of cost increase?
  - (b) How do your results compare to the discussion regarding Table 12-4? Why?

# Nonprofit Firms

- An Introduction to Nonprofits
- Why Nonprofits Exist and Why They Are Prevalent in Health Care
- Models of Nonprofit Hospital Behavior
- The Relative Efficiency of Nonprofits Versus For-Profits
- Conclusions

Nonprofit firms account for only 5 percent of GDP, but they make up a significant portion of the health care sector. The 60 percent of community hospitals that are nonprofit provide nearly 70 percent of the beds and treat a similar proportion of the nation's hospital patients. Nonprofit firms also offer 30 percent of nursing home care and half of the inpatient specialty mental health and substance abuse treatment. We also see nonprofits providing education, collecting and providing blood, operating symphony orchestras and opera companies, and facilitating the conservation of land.

Inasmuch as nonprofits are prominent in health care, especially in the important hospital industry, they pose questions of considerable interest. Will the economic behavior of nonprofits differ distinctly from the more common for-profit firm or from government enterprise? A growing body of literature addresses such questions and adds to our understanding of nonprofit firms in general. Studies within health economics contribute to both the general theory of nonprofit firms and the understanding of vital issues of the health economy.

## AN INTRODUCTION TO NONPROFITS

What is the nonprofit firm? People commonly assume that nonprofits are firms organized to provide charitable goods or community services and that they obtain their revenues largely from donations. While many fit this category, the definition does not serve well conceptually, nor does it make an important economic distinction. On one hand, numerous profit-making firms provide important goods and services to the community, and many provide basic goods, such as food and housing, to the poor. On the other hand, nonprofit firms often serve the well-to-do, and they often compete with for-profits. Many obtain the lion's share of their revenues from the sale of goods and services at prices similar to their for-profit competitors. Health care nonprofits obtain more than 90 percent of their revenues from "sales and receipts."



In economic language, the important distinction of the nonprofit is the nondistribution constraint. This means that no one has a legal claim on the nonprofit's residual, the difference between the revenues and its costs, or what an ordinary firm would call its profits. Because there is no residual claimant, the nonprofit's objectives may differ from profit-making. Two secondary distinctions between nonprofits and for-profits are also consequential. First, nonprofits are exempt from corporate income taxes and often from property and sales taxes. Second, donations to nonprofits receive favorable tax treatment. These distinctions give them an advantage and make nonprofits a potentially different, but are they?

## WHY NONPROFITS EXIST AND WHY THEY ARE PREVALENT IN HEALTH CARE

In any economy like that of the United States, there are three types of firms: private profit-making, government, and voluntary nonprofit enterprises. Why do each of these types of firms exist? Societies do not create economic institutions at random. They must reflect some basic economic incentives and unmet needs that the firms and the laws establishing them were created to satisfy.

### Nonprofits as Providers of Unmet Demands for Public Goods

Burton Weisbrod's analysis (1975) guides the following exposition. In the United States and other Western economies, economists view a perfectly competitive industry under certain circumstances as economically efficient, and empirical data suggest that competition often fosters growth. Under this account, we need government enterprises only in cases where competitive markets fail. As this theory unfolds, we need nonprofits, in turn, when government enterprise also fails. We begin by reviewing the standard explanation of the role of government enterprise, and we then introduce Weisbrod's explanation in which the nonprofit firm satisfies demands for public goods unmet either by private markets or the government.<sup>1</sup>

Under the standard economic explanation, government enterprise might possibly—though not necessarily—have a role in improving market efficiency in cases where competitive markets tend to fail. The most prominent cases of market failure involve externalities and public goods. Two examples will clarify these issues.

**MARKET FAILURE: EXTERNALITIES** Most consumer goods provide private benefits and little or no externalities. An externality is an uncompensated direct effect of the production or consumption of a good on persons other than the producers or consumers. Consider goods without no externalities. When one enjoys a hamburger at a local restaurant, the pleasure is primarily private, and the benefit goes to the one who consumes. There is no effect on parties external to the market, other than the producer or consumer. However, consider the case where one purchases a vaccination for influenza. This good entails a private benefit: The purchaser will less likely suffer from influenza. In addition, there is an external benefit to others because the purchaser will be less likely to infect others.

Free markets tend to underproduce goods for which there exist significant external benefits. For example, the purchasers of vaccinations will tend to consider primarily the private benefits and will ignore the external benefits to the community. The benefit to society is the sum of the private benefits and the external benefits to the community. Because demand will represent only the private benefits, it will understate society's benefits and give a false or inadequate signal to the market. The market then produces less than the amount that would maximize net social benefits. This is economically inefficient, and we call this situation a case of market failure.

---

<sup>1</sup> For further discussion on the nature of nonprofit firms, see Weisbrod (1988).

The existence of a large externality raises the possibility of a role for government, but it is only a possibility? If we recognize that markets may fail, we must recognize that governments too may fail to act efficiently.

**MARKET FAILURE: PUBLIC GOODS** The vaccination is a private good with an external benefit, but it has public good aspects. Consider first the case of a pure public good, a good that is both *nonexcludable* and *nonrival*. Nonexcludable means that people cannot be economically excluded from consuming the good even if they refuse to pay for it. Nonrival means that one person can consume the good without depleting it for others. Our hamburger is an example of a private good in that McDonald's can easily refuse the hamburger to someone who refuses to pay. Likewise, the hamburger is rival because when one person consumes the burger it is then depleted, eaten up.

In contrast, consider a pure public good like national defense. Imagine an antimissile system that puts a defensive umbrella over a country. This defense system would be nonexcludable because those living in the country would benefit whether they paid or not. It would be impossible, let alone costly, to allow the nonpayer alone to be subject to an enemy missile strike. Likewise, the defense system is nonrival because the protection of one more individual does not diminish the defense enjoyed by others.

The government often provides public goods like this. If private enterprise tried to attempt to provide defense, it would find many citizens choosing to be free riders. A free rider is a person who consumes the public good but refuses to pay. Only government has the power to force consumers to pay.

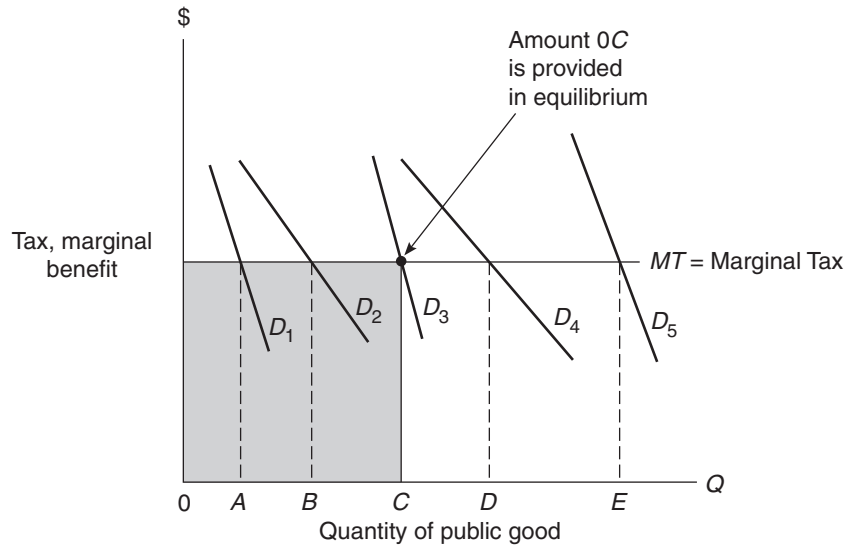
### The Public Good–Private Good Aspect of Donations

Now consider charitable donations. A donation to the health of others has the characteristics of both a public good and a private good. The need for both concepts becomes clear in Richard Steinberg's (1986, 1987) work, which argues that donations to public goods motivate the donor as both private and public goods. If you donate toward the health of a poor person, you may get a "warm glow" (an increase in utility from the act of giving; Arrow, 1975; Andreoni, 1990). The warm glow may come from the act of donating or simply from the pleasure in knowing that a suffering person's health improved. However, other charitably minded persons will also have this pleasure, whether or not they have donated. They are free riders who receive an external benefit free. The charity market then provides too little charity to be efficient.

Now consider a case where the government provides a public good. This case explains Weisbrod's rationale for the existence of nonprofits. In Figure 13-1, let the curves  $D_1$  through  $D_5$  represent the demand curves of five different voting individuals for a public good that the government will provide. Let the demand curves represent the external benefits to these different groups of taxpayers. These demand curves represent the marginal benefits to the taxpayer donors.

To pay for this public good, assume that these five different taxpayers will be equally taxed at a per-unit tax rate of  $MT$ , the marginal tax. Because the five individuals will pay equal taxes, the marginal tax for each will be exactly one-fifth of the marginal cost to society. Thus, if the government were to provide output  $OC$ , then each of the taxpayers would be charged  $OC \times MT$ , and the total tax receipts would exactly pay for the project. If the output were  $OB$ , then  $OB \times MT$  would be collected from each individual, and so on.

The government must choose a single level of output. What level would it choose leaving the outcome to the democratic political process? Economists find it convenient to imagine an elected manager whose continued term in office is determined by the ability to provide the "right" level. If level  $OA$  is proposed, four of the five voters would prefer more (why?) and would vote against it. If level  $OB$  is proposed, three of the five would prefer more and would vote against that level. In contrast, level  $OD$  would provide too much public good for three of the five voters, and similarly for level  $OE$ . Only level  $OC$  will gain a majority vote. This majority consists of Voter 3, who is exactly



**FIGURE 13-1** The Median Voter Model of Public Goods

satisfied because her marginal benefits equal the marginal tax, and Voters 4 and 5, who would like to see more but are satisfied with level  $OC$ . Suppose the manager proposed just a little more than  $OC$ . This is too much for Voter 3, who now votes against the manager in favor of someone else who would reduce the public good level back to  $OC$ .

This example illustrates Weisbrod's point. The level chosen entails dissatisfied voters on the margin, whose demands are not exactly satisfied. The marginal tax rate perfectly matches the preferences only of Voter 3, the median person. It is theoretically possible to design a tax system so that for each person the marginal tax equals his or her marginal benefit, but it is difficult to accomplish this in practice. Without such a tax system, some voters will prefer less of the public good and some will prefer more of it. The government enterprise will have failed to satisfy the demands of those who would prefer more. Such unsatisfied voters will have the incentive to form a nonprofit enterprise and provide the good themselves.

It follows from Weisbrod's theory that nonprofits arise because both private markets and government will tend to underproduce goods or services entailing beneficial externalities. The underserved citizens are those who feel the external benefit most keenly. Such will found and support nonprofit corporations.

### Relevance to Health Care Markets

The Weisbrod analysis will apply in principle to services that provide external benefits to the community at large. Nonprofit health care enterprises may arise where a sufficient minority of voters are dissatisfied with the quantity or quality of such services provided by the for-profit sector or government. The theory fits the historical rise of nonprofit hospitals. Hospitals in the United States and in many countries often began as charitable institutions, providing care primarily to the poor and relying heavily on donations.<sup>2</sup>

However, the purely charitable aspects of hospital care have become less important (see Gruber and Hungerman, 2007). Improvements in medical technology transformed hospitals into workshops for doctors—places for treating all patients, both rich and poor. By 1996, donations accounted for less

<sup>2</sup> The history of hospitals and the relative importance of nonprofit versus for-profit status are further explored in Bays (1983), Frech (1990), and Temin (1988).

than 3 percent of average hospital revenues, and only about 3 percent of patients relied on donations or welfare as their principal source of payment. Hospital finance departments wrote off unpaid patient bills, but these accounted for less than 1 percent of billings. Experts have attributed the demise of hospital donations to the growth of hospital insurance coverage, especially since the advent of Medicare and Medicaid. The decline in donations is consistent with the Weisbrod theory. Individual demands have stayed stable while increases in public sector provision or financing have occurred.

With the decline in cash donations, other forms of donations as well as other forms of nonprofit hospital advantages have become relatively more important. These include donations of time and energy for board members and others.

To be sure, cash donations still play an important role in certain capital fund drives, but in the present day, the financial advantages enjoyed by nonprofit hospitals stem more importantly from the subsidization of nonprofits by the government through tax exemptions. Generally, nonprofits do not pay corporate, property, or sales taxes although for-profits do. The only tax advantage enjoyed by for-profits is the ability to write off losses during bad years.

### **Nonprofits as a Response to Contract Failure**

A related theory with early origins complements Weisbrod's ideas. Arrow (1963) suggested that the prevalence of nonprofits is due to the uncertainty of identifying quality of care. Hansmann (1980) expanded this idea and theorized that the nonprofit sector helps to repair the problems of contract failure that occur when the quantity or quality of output is difficult to observe. Thus, asymmetry of information between the firm and the buyer of services becomes important (as noted in Chapter 10) in explaining the nonprofit role.<sup>3</sup>

A classic example of the contract failure illustrates the problem. Suppose you wish to contribute food and clothing to suffering people in Haiti. You can find a firm to deliver the care. However, it would be prohibitively costly to verify that the firm actually is delivering the desired goods to the designated population. You would hardly fly to Haiti just to check on this. Under these circumstances, you may prefer to employ a nonprofit firm.

People will perceive a for-profit firm to have a conflict of interest as a deliverer of the aid packages. Such a firm could increase its profits by renegeing on its promise. The nonprofit cannot distribute its residual so it would have less incentive to renege. In this case, the nonprofit, by better serving the donor's interests, also serves the market more efficiently.

### **Applications of Contract Failure to Health Care**

Contract failure does not occur solely in the cases of donated goods and services, as it can exist even when the purchaser is nearby. Contrast the management of hotels and nursing homes. Hotels are profit-seeking enterprises that provide rooms and suites along with housekeeping services, dining, and recreation to travelers and residents. Nursing homes also provide rooms and suites, dining, and recreation along with housekeeping services and of course special nursing services largely for an older and often infirm population. In fact, many nursing homes began as hotels.

We assume that hotel patrons are utility-maximizing consumers who compare benefits and costs in deciding whether to stay at the hotel. Nursing home patients, in contrast, may not be able to assess the quality of the facility and the care they receive accurately, perhaps because of their health impairments. Relatives or friends may obtain only limited impressions upon visiting and may not be sophisticated assessors of the quality of long-term care. A for-profit home, in contrast to a for-profit hotel, may appear to have a conflict of interest in the eyes of some demanders. That nonprofit ownership is viewed as a signal of higher quality is supported by both theory and data, and Christensen and Arnould (2005) provide an example of this. But do for-profit nursing homes in practice provide lower

---

<sup>3</sup> The essential idea was developed further in a model by Easley and O'Hara (1983).

quality? It is difficult to compare them. For example, does a nursing home with more nurses per patient provide higher quality or is it simply a measure of inefficiency? Outcome measures, such as changes in patient health status, are best in principle, but reliable outcome measures are difficult to obtain.

Much of the early evidence suggested an apparent for-profit advantage in cost. However, other reports that took greater effort to account for quality disagreed with the earlier view. In practice, nursing homes with more personnel per patient tend to have better patient outcomes, but it is difficult to sort out efficiency and quality. Gertler (1989) addressed these problems and found a higher quality level in nonprofit homes.<sup>4</sup>

Hirth (1999) and Santerre and Vernon (2007) demonstrated that under plausible conditions, an influx of nonprofit homes will drive up the average quality in the market, making the nonprofit a productive agent for change whether or not it exhibits a higher quality itself. Related research further supports this theory (Grabowski and Hirth, 2003).

Contract failure theory does not appear appropriate for hospitals, a point noted by Hansmann (1980). Hospital patients are under the close supervision of physicians acting as the patients' agents. Physicians have little incentive to misinform patients by overstating the quality of care.

Sloan (1988) further suggests that those who favor the contract failure theory often apply a double standard. The field of physician services, which are nearly all provided on a for-profit basis in the United States, seems ripe for the application of contract theory. Consumers would find it just as hard to assess the quality of physician care as they would hospital care, so why do nonprofit firms not take over the physician care sector?

### **Financial Matters and the Nonprofit**

Lacking the ability to distribute net revenues, the nonprofits do not issue equity stock and lack this avenue for raising capital. Although this is a disadvantage when it needs to respond to changing market conditions, the nonprofit also has some financial advantages. It is exempt from corporate, property, and sales taxes, and its bonds are generally tax-exempt as well. It may also have market advantages, such as the responsiveness and loyalty generated by the consumer's response to nonprofits under the contract failure situation. Finally, it is more likely to attract donations than is the for-profit.

What is the bottom line when all advantages and disadvantages of the nonprofit are considered? Which ownership form can respond to a rapidly expanding demand the quickest? Research generally has indicated that rapid demand change favored the for-profit as measured by changes in their market share. Hansmann, Kessler, and McClellan (2002) determined that over the previous 20 years, the for-profit hospital form was the most responsive to demand change in an era of declining hospital demand. Chakvarty et al. (2005) support this assessment by finding for-profit hospitals to be quicker to either enter or exit a market as conditions change.

### **Summary of the Reasons for the Prevalence of Nonprofits**

Weisbrod accounts for nonprofit firms that arise to provide for unmet demands for public goods, when there are significant external benefits from the good or service. Hansmann's view complements this account, emphasizing the role for nonprofit firms in cases of contract failure. Under either of these analyses, the tax preferences for nonprofit firms make economic sense.

## **MODELS OF NONPROFIT HOSPITAL BEHAVIOR**

Health economics offers many descriptions of hospital behavior. We begin with a nonprofit model that applies in principle to an entire class of nonprofit firms.

---

<sup>4</sup> For further study, see Cohen and Spector (1996) and Davis (1991).

## The Quality–Quantity Nonprofit Theory

When economists model a nonprofit hospital, they begin by positing an objective of the hospital decision makers. Most frequently, they choose either a utility-maximization or a profit-maximization type of model. The utility maximizing model, most clearly approximating the altruistic firm, was proposed by Joseph Newhouse (1970). For Newhouse, the hospital's objective was to maximize the utility of the decision makers. Utility of the firm is an index of the decision maker's preferences, a measure of satisfaction, similar to the utility of a consumer.

It is especially difficult to describe the complex elements and interactions of the heterogeneous set of hospital decision makers. Nonprofit hospitals in the United States tend to have three parties with considerable decision-making authority. The trustees are nominally in charge, but boards of trustees may include people with widely varying backgrounds, knowledge of health care, and management expertise. The trustees' decision-making agent is the hospital administrator or CEO. This manager may have varying degrees of actual power and authority. Finally, the arbiters of medical decision-making are the physician staff. The physicians also may exercise considerable decision-making authority. We assume that this complex decision-making apparatus resolves into a single utility function and describes a set of well-behaved indifference curves.

**THE UTILITY FUNCTION** In Newhouse's model, the hospital's preferences are defined over quantity and quality of output. Quantity of output could be measured in several ways, but assume we measure it by the number of cases treated. We further assume that there is only one type of case to treat, though there could be hundreds. Output quality can entail many different characteristics of the care provided. Some top decision makers may value the quality or beauty of the hospital structure, and others may emphasize expertise of the physician or nursing staff. Still others may emphasize prestige in the medical community, and yet others may stress the quality of the tender loving care provided. Graphically, we shall suppose that just one index of quality exists.

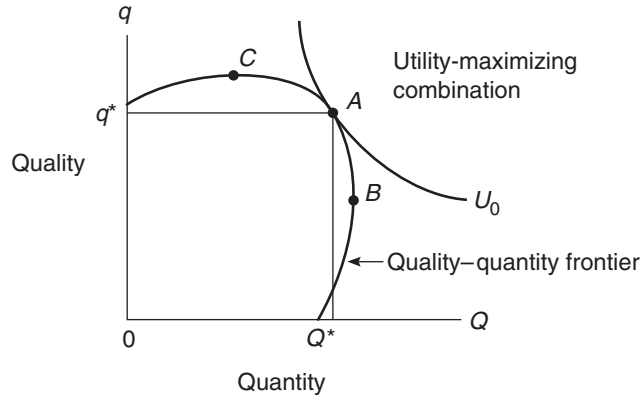
This conception of the nonprofit hospital is consistent with the external-benefits account of the role of nonprofit firms. The utility derived from producing quantity and quality might arise because care to these patients entails an external benefit to the community at large. Consider this model as a description of the hospital decision-makers' having altruistically internalized the community benefit in providing quantity of care.

**THE QUANTITY–QUALITY FRONTIER** The hospital selects a combination of quantity and quality that maximizes utility. It faces a budget constraint, as it must pay its bills and cannot run negative net revenue. Furthermore, the nondistribution constraint, which applies to all nonprofits, implies that this hospital has no incentive to maximize net revenues as a general rule. Thus, by the budget constraint the sum of patient-generated revenues plus donations equal the hospital's costs.

Figure 13-2 shows its possible choices as the Quantity–Quality Frontier. This frontier comes from demand and cost analysis, but we can understand it intuitively as follows (see Spence, 1973, for a detailed explanation). At zero quality, point  $Q^*$ , this hospital can achieve both a higher quality and a higher quantity by choosing a point to the northeast. This can occur if the higher quality attracts more customers but costs do not grow as rapidly. The frontier eventually bends backward, indicating that quality improvements no longer strongly attract customers but quality is increasingly costly.

**MAXIMIZATION OF UTILITY** Given the possibility frontier, the hospital decision makers choose the point that maximizes utility. The constrained utility maximization point occurs at a point of tangency between the frontier and the highest indifference curve attainable. In Figure 13-2, utility maximization occurs at point A.

On the one hand, let the hospital preferences place sole weight on quantity. Then, the hospital would behave like a constrained quantity maximizer and produce at point B. On the other hand, let the



**FIGURE 13-2** Hospital Quality-Quantity Trade-Offs

hospital primarily value quality, this quality emphasis by top decision makers who may be altruists or who may alternatively be primarily interested in the prestige of the hospital as compared to its peers. It would produce at point  $C$ .

### The Profit-Deviating Nonprofit Hospital

We have just seen a model where the nonprofit focuses solely on the quality or quantity of hospital care. Although some have suggested that such behavior might be a form of altruism, Lakdawalla and Philipson (2006) see the nonprofit differently, as a mix of altruism and profit motives. Their model makes clearer the entry and exit responses of nonprofits to changes in market conditions and government regulation.

Let the hospitals maximize utility  $U = U(q, \pi)$  over the quantity of hospital service,  $q$ , and define the nonprofit's "profit"  $\pi^N$ , as the sum of profit from sales,  $\pi_S$ , and receipts from donations,  $D$ , and require that it can pay its bills,  $\pi^N = \pi_S + D > 0$ . This model is particularly useful to contrast the behavior of nonprofits and for-profits. Thus let the for-profit goal be to maximize its profits  $\pi^F$ .

Notice that the profit-deviating model described this way includes the purely altruistic model (a version where the decision maker has no concerns for personal profit) and the pure profit maximizing model (where decision makers have no concern for the health of the community separate from profit) as special cases.

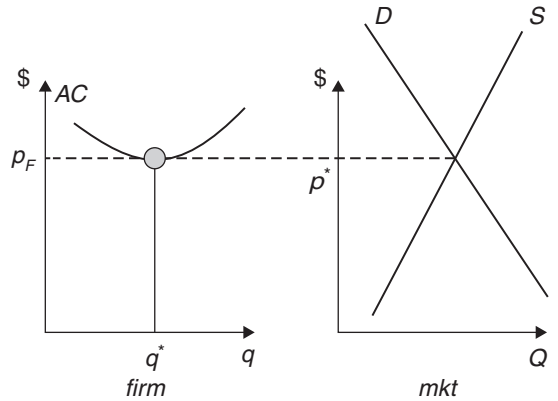
Consider entry and exit of the profit-deviating hospital. To enter a market, it must cover its opportunity costs elsewhere, that is, it must be able to attain the utility level that it could achieve in other markets (or by simply not producing in this market). We call its minimum required utility in this market the *Reservation Utility*,  $U^* = U(0, 0)$ . As in all such models, the hospital of both types must pay its bills; here the profit-deviating hospital has an advantage over the pure for-profit because it receives donations. Thus the operating constraint of the for-profit hospital is that its profits are nonnegative,  $\pi^F > 0$ , while the operating constraint of the profit-deviating hospital is  $\pi^N = \pi_S + D > 0$ .

Figure 13-3 illustrates their differences. The first graph shows the for-profit's Entry Conditions Curve, which is the same as its Long Run Average Cost Curve (LRAC). It records all price quantity combinations that yield a zero profit; it must attain at least this much revenue to stay in the market. This much is as in conventional microeconomic theory, and in the long run equilibrium, the market price would equal  $P_F$ .

Alternatively, consider the nonprofit. It too has a break-even curve, though this does not determine its entry conditions. This break-even curve is found by subtracting the average donations from

**FIGURE 13-3** Entry Conditions into the Market for Nonprofit and For-Profit Firms

- Look first at the “traditional” market model.
- In the LR, the “marginal” firm will produce output  $q^*$ .



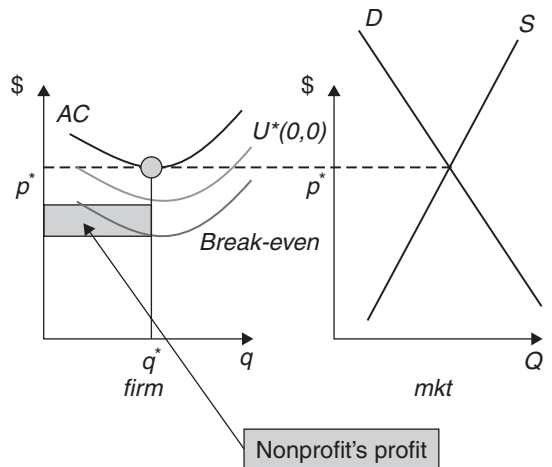
the LRAC curve shown, that is, nonprofit must make at least  $LRAC - D/q$  to be able to pay its bills. It will do better. Consider the curve labeled  $U(\pi, q) = U^*(0, 0)$ , in Figure 13-4. This is an indifference curve of a sort, with all points on the curve yielding the same utility. At the utility level shown, the firm is just indifferent between producing in this market or not.

We finish up by explaining the implications of the model. First suppose some more nonprofits besides the one shown enter this market. This would shift the market supply curve to the right and cause the market price to fall. If enough nonprofits entered this could drive the market price below  $P_F$ , which would drive all of the for profits firms out of the market. The nonprofit donations are a very powerful advantage. Why do we have any markets at all with a mix of for profits and nonprofits?

Keep in mind that these nonprofits will differ in their level of Reservation Utility (much like consumers will differ in their appreciation of a consumption good). This will mean that the number of nonprofits that choose to enter cannot be determined a priori; though, in principle, there could be so many as to drive all for-profits out. Most states in the United States do not have for-profit hospitals, though this must derive in part from legal or other institutional restrictions.

If there are enough nonprofit hospitals waiting in the wings, their entry would drive out all the for-profits. However, suppose profit-deviating entry stops before that so that there temporarily is unmet demand for hospital care. In contrast to the nature of the nonprofit entry, standard

- Here, the for- $\pi$  firm faces constraint,  $\pi \geq 0$
- Non- $\pi$  firm has a reservation utility curve below for- $\pi$  curve.
- Non- $\pi$  firm faces constraint  $\pi + D \geq 0$ .  $D$  = donations.



**FIGURE 13-4** Co-existence of non- $\pi$  and for- $\pi$  firms



for-profit theory suggests that there will always be for-profit firms to take advantage of profit opportunities. However, in order for for-profit hospitals to enter the available revenue must be high enough for them to cover their costs. Given unmet demand, the market price will tend to rise until it is profitable for them to enter (in the process some new nonprofits may also enter). The equilibrium is determined by the for-profits' greater response to market signals. In the graph, this price,  $P_F$ .

Finally, if prices went up further, then more for-profits would enter. Their entry, as in the standard theory, would drive the price back down to equilibrium. By this process, in the long run, the for-profits are the marginal firms and their cost structure determines the market equilibrium price. Because they are the marginal firms in the industry, they are the ones that determine a new market price after a new regulation changes hospital cost structures. The authors conclude that in markets where the two firm types coexist, the for-profits, as the marginal firms will determine the effects of the new regulation.

### The Hospital as a Physicians' Cooperative

A different account of the nonprofit hospital comes from theories that believe that hospitals maximize the pecuniary gain to the decisive set of decision makers. Mark Pauly and Michael Redisch (1973) describe the nonprofit hospital as a "physicians' cooperative," assuming that the hospital is controlled by a physician staff who operate the hospital to maximize their net incomes.

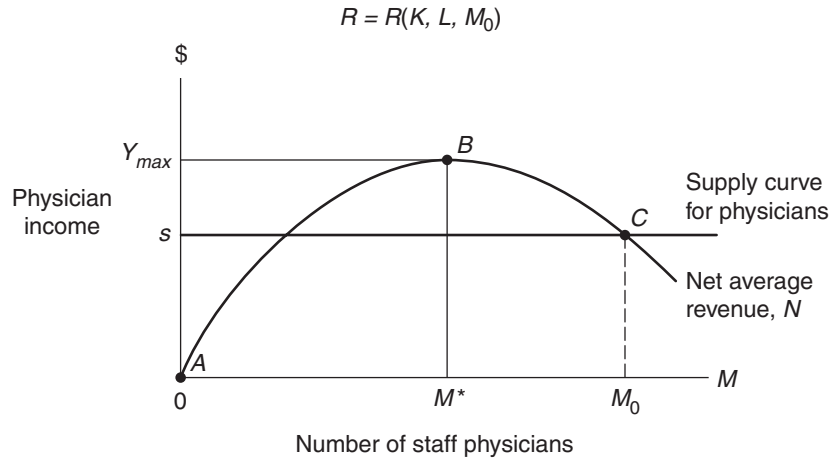
This view of the hospital focuses on the "full price" of the hospital care, meaning the total charges to the patient by both the hospital and the physician. Assume that the patient pays a single bill representing the full price of the care and let donations be zero. The full price of care depends on consumer demand and on the total quantity of care offered by the physicians' cooperative. The amount of care produced and offered to patients depends, in turn, on the quantity of inputs chosen by the top decision makers, here the physicians. Summarize hospital inputs as capital,  $K$ , and labor,  $L$ . Let the physician input be  $M$ , a fixed input if the hospital selects a "closed staff."

### Maximizing Net Revenue per Physician

In this model the co-op runs the hospital to maximize the net revenue ( $NR$ ) per physician ( $M$ ), or  $NR/M$ . The net revenue is the sum of all of the revenue less factor payments to nonphysician labor and payments to capital. The net revenue per physician divides that revenue over the total number of physicians,  $M$ . An increase in the number of physicians,  $M$ , initially increases revenues per physician. Eventually, revenues per physician must fall because (due to the fixed levels of nonphysician labor and capital) the percent increase in revenues (in the numerator) will be smaller than the percent change in number of physicians (in the denominator).

Figure 13-5 shows the optimal staff size if the physicians can limit the size of the staff, or "close" the staff. Here dollar values are on the vertical axis and the number of physicians on staff,  $M$ , is on the horizontal axis. The curve  $N$  denotes average physician income. The  $N$  curve starts at the origin point  $A$  (no revenue), rises to a maximum at point  $B$ , and then falls. Curve  $s$  depicts the supply curve of physicians, which is infinitely elastic, plausibly representing a case in an urban, physician-dense environment. For physicians who are on the staff, the optimal staff size would be  $M^*$ , where curve  $N$  reaches its maximum.

In contrast, if the hospital has an open staff, physicians are free to enter as long as their resulting average income,  $N$ , equals or exceeds their supply price,  $s$ . The open-staff equilibrium occurs at point  $C$ , where net revenue (the demand for physicians) equals supply,  $s_0$ , with  $M$  physicians hired. Regardless of the number of physicians on staff (either a closed-staff or an open-staff equilibrium), the hospital inputs are chosen to maximize residual income for the medical staff. If we view the Newhouse model as resulting from the maximization of external benefit perceived to accrue to the community, then the Pauly-Redisch objective is the complete opposite.



**FIGURE 13-5** Maximizing Net Average Physician Revenue in Hospital

### A Comparison of the Quantity–Quality and the Physicians’ Cooperative Theories

It is useful to contrast the two models on the extreme ends of the spectrum by comparing them on the same graph. To do so, we represent combined (physician and hospital) revenues as a single function:

$$R = R(K, L, M_0)$$

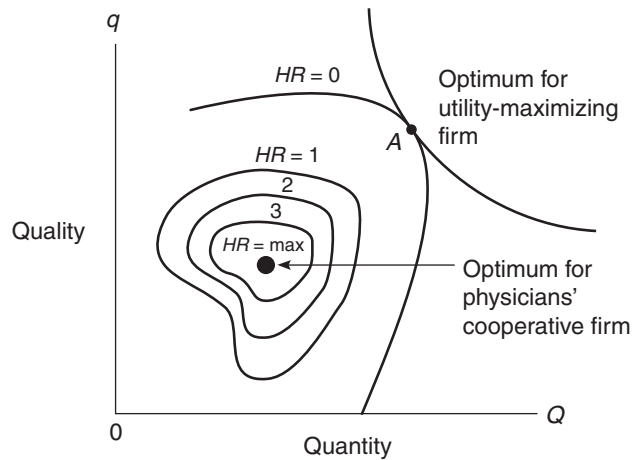
The combined revenues depend on the quantity and quality of care produced, which in turn depend on the amounts of the inputs used. Like Newhouse, assume that the hospitals/physicians produce care efficiently so that a higher quality of care requires necessarily a higher level and hence cost of inputs. The hospital may receive additional revenues in the form of donations,  $D_0$ , and government subsidies,  $G_0$ . Let physician supply be perfectly elastic at a constant supply price,  $s$ . Constant input prices,  $r$  for capital and  $w$  for labor, complete the description. Finally, define the hospital residual,  $HR$ , revenues as the following equation:

$$HR = R(K, L, M_0) - wL - rK - sM_0 + D_0 + G_0 \quad (13.1)$$

Under the Pauly-Redisch model, the physicians on the staff usurp the hospital residual,  $HR$ . For a given level of physicians on the staff,  $M_0$ , the physicians will maximize their average incomes by maximizing  $HR$  itself.

In contrast, the Newhouse hospital will maximize utility of quantity,  $Q$ , and quality,  $q$ , subject to the constraint that the hospital residual is zero; that is, the hospital breaks even. To simplify the graphical representation, let the hospital residual function,  $HR$ , form a rounded hill, following Spence (1973). The contours of that hill in Figure 13-6 graph quantity and quality of care on the axes. The contours are analogous to contour lines on a topographical map. For example, the contour line labeled  $HR = 1$  represents the collection of all combinations of quality and quantity of care that yield a hospital residual of \$1 million. Contours farther away from the maximum residual point,  $HR = HR_{max}$ , yield successively lower levels of hospital residual. The contour curve labeled  $HR = 0$  indicates the combinations of quality and quantity that yield a zero residual.

**QUANTITY–QUALITY CONTOURS AND HOSPITAL RESIDUAL** Begin with the Pauly-Redisch analysis of a closed hospital staff. Let the hospital physician staff be fixed at some level,  $M_0$ . The Pauly-Redisch hospital chooses the quantity-quality combination that maximizes the hospital



**FIGURE 13-6** Hospital Quantity–Quality Contours and Hospital Residual

residual, point  $HR = HR_{max}$ . The hospital residual then goes to the staff physicians, and these physicians will thus have maximized their average incomes. In contrast, the Newhouse hospital maximizes utility at point A, the point of tangency between the  $HR = 0$  contour, representing the Newhouse budget constraint, and the highest indifference curve attainable. Thus, in the closed-staff case, the models yield very different results. As depicted, the Newhouse hospital tends to produce more quantity and quality of care. Under the Pauly-Redisch behavior, however, the physicians indirectly usurp the hospital care residual, and this includes the donations and government subsidy as well. It is as if the nonprofit hospital is a for-profit firm in disguise. If we believed that the nonprofit hospitals behaved like this, we would likely call for an end to government tax exemptions for this nonprofit status (Clark, 1980).

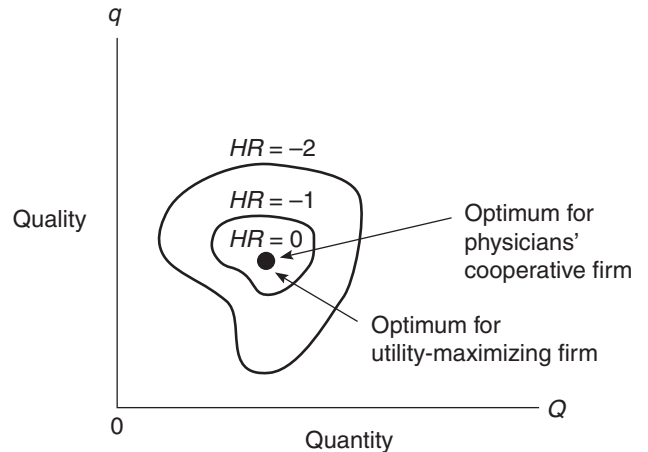
**EFFECTS OF INCREASED COMPETITION** Examine an effect of increased competition in the hospital sector. If entry were free, then all potential firms that may want to compete for hospital care patients are free to do so. Potential competitors could include alternative delivery systems as well as other hospitals. As more competitors enter the market and compete for business, the demand for care at any existing hospital will tend to fall. For our purposes, this means that competition will tend to shrink the hospital residual hill in size.

Figure 13-7 depicts a case where competition has continued until the maximum hospital residual attainable is zero. As we move away from the top of the hospital residual hill in any direction, the contours reflect negative and successively more negative residuals. The result is that the Newhouse and the Pauly-Redisch nonprofit hospital in this result will converge in their choices of quantity and quality. The only difference remaining in the long run will be simply the result of the nonprofit's cost advantage.

The hospital industry has experienced increasing competition in the last 30 years. Especially noteworthy are the competition from alternative delivery systems and the competitive incentives introduced under Medicare's prospective payment system. A tendency for nonprofit hospitals to converge toward for-profit-like behavior has long been noted in theory (Newhouse, 1970), and more recently it has been proposed as a probable result of hospital competition.

### The Evidence: Do Nonprofit Hospitals Differ from For-Profit Hospitals?

Since the physicians' cooperative would choose nonphysician inputs and outputs much like a pure profit-oriented firm, we look at empirical comparisons of nonprofit and for-profit hospitals. Despite what appears to be a big difference in orientation, only a few behavioral differences are clear in theory. Sloan and colleagues (1998), in fact, conclude that there is not a "dime's worth of difference," basing this conclusion on studies of quality, cost, and efficiency of hospitals by ownership type.



**FIGURE 13-7** Convergence of Hospital Models in Long-Run Equilibrium Under Free Entry

Theoretically the nonprofit will produce at a higher quality, a higher quantity, or both. The Newhouse nonprofit could be very quantity oriented; it may strive to provide large quantities of basic hospital care to a deserving, perhaps destitute, population. Careful observation, however, finds that the public hospitals are first to serve areas of poverty (Ballou, 2008). Nonprofits may instead strive single-mindedly for the highest quality of care possible (Newhouse argued that there would be a bias toward quality). These (potentially) different types of nonprofits make comparisons of average quality of all nonprofits versus for-profits difficult to interpret. Thus, lack of quality differences on average is consistent with theory.

Are there differences empirically in quality? An early study by McClellan and Staiger (1999) found higher mortality rates for the elderly in for-profit hospitals overall, but the small difference on average masked substantial variation with a number of markets showing quality superiority in the for-profit hospitals. Shen (2002) and Lien et al. (2008) also found quality advantages with the nonprofits, though Eggleston et al. (2008b) warn that these kinds of results depend on data sources, time periods, and regions studied. There is mixed evidence on the effect of competition on quality (Gaynor, 2006), but at least one study finds that competition from nonprofits tends to provide spillover effects so as to improve the quality of the for-profits (Grabowski and Hirth, 2003).

Who provides the most charity care? Norton and Staiger (1994) measured differences in care provided to the uninsured by nonprofit and for-profit hospitals. They found that hospitals in the same market area tend to serve the same number of uninsured. Private charity care declined historically, probably due to crowding out by government charity that started with President Franklin Roosevelt's 1933 New Deal (Gruber and Hungerman, 2007). It may be possible, nevertheless, to encourage nonprofit hospital markets to increase their charity care. This could occur if the financial incentives of tax-exempt status encourage development of nonprofit hospitals, which, according to Hassan and colleagues (2000), "are forced to provide higher levels of charity as a condition for gaining access to the tax exempt [bond] market."

Several studies investigate potential differences in managerial practices. Ballou and Weisbrod (2003) find substantial differences among religious, secular nonprofit, and government hospitals in patterns of CEO compensation. However, Brickley and van Horn (2002) find for a large sample of nonprofit hospitals that compensation incentives for CEOs are significantly related to financial performance. They also find little evidence that nonprofit hospitals provide "explicit incentives for their CEOs to focus on altruistic activities."

Several technical issues have shown the differences between for-profit and nonprofit hospitals more clearly. As we discussed, Hansmann et al. (2002) found the for-profit to be quicker in adjusting to market demand changes, presumably due to better access to the capital markets.

Chakvarty et al. (2005) support this finding that the for-profits to be more “nimble” in adjusting to new economic conditions. Hirth et al. (2000) studied the responses of firms providing dialysis care during a period of decreasing dialysis payments. The nonprofits tended to maintain quality and offset the dollar losses by cutting back on amenities, while the for-profits took the reverse route.

Finally, Silverman and Skinner (2001) found different patterns of DRG “creep” in which hospitals may “upcode” or recategorize patients into more profitable Diagnosis Related Groups (DRGs) used to determine Medicare payments. The authors compared the DRGs of “pneumonia” and “respiratory illness”—where the line between the two DRG categories is fuzzy for many patients. The authors found generally that hospitals facing tougher financial conditions tended to upcode patients into the more profitable DRGs. The upcoding response, however, was greater for the for-profit than for the nonprofit hospitals, and greatest for hospitals that had just recently converted to for-profit status.

**REALISM IN HOSPITAL MODELING** The three analytical hospital models that we have presented can be understood as contrasts between the roles of the two principal power groups in the modern hospital, the physician staff group and the trustee-administrator group.

It is tempting to view the utility function as representing the top management as constrained by the realities of physician behavior. The Pauly-Redisch model posits the de facto control by the physicians. In either case, there is presumably a resolution of power that occurs, perhaps even harmoniously. The hospital theory proposed by Jeffrey Harris (1977) represents an entirely different approach. The hospital, under Harris’s account, is the scene of continual conflict within an organization inherently split into two parts—what Harris describes as a noncooperative oligopoly game.

The Harris model is noteworthy for two additional reasons. Harris is a physician with a Ph.D. in economics and he devised his account after experience as a physician in an urban hospital. As an economist, he understands complex economic organizational theories, but he presents his work in realistic words and images describing experience within a hospital.

**THE HOSPITAL AS TWO FIRMS** Harris proposes that the hospital’s internal organization is really two separate firms interacting in a complex way. The hospital’s structure has two main parts: the trustee-administrator group that serves as the supplier of inputs, and the physician staff that serves as the demanders. Like an automobile parts-and-service shop, the hospital has technical experts providing the actual service and requesting supplies from the organization. Unlike the auto shop, the physician’s loyalties lie almost entirely with the patient. The physician is the patient’s “hired gun.”

Rather than having a predictable technological and inventory problem, the administrator has the problem of providing inputs to the physician-agent during a complicated, uncertain sequence of events. Consider Harris’s account of a hypothetical case history:

Mr. X comes to Dr. A with a fever and a cough. A chest X-ray reveals a density. He is hospitalized. Penicillin is administered. Although the fever subsides with this treatment, a repeat X-ray shows that the density has not disappeared. A sputum cytological examination is performed and lung cancer is diagnosed. Further studies suggest that the cancer can be removed surgically. An operation is performed. Unfortunately, massive postoperative bleeding occurs. Matched whole blood is administered. Despite this treatment, a cardiac arrest ensues and an emergency resuscitation (code call) is announced. Mr. X is transferred to Intensive Care with chest tubes and a respirator. A special contrast study (angiogram) reveals the site of bleeding. A repeat operation is performed. (p. 469)

Here, rather than describing a predictable assembly-line procedure, the hospital service is unique and customized for each patient and it plays out in an unpredictable way. The physician hurries to various supply centers, demanding services on behalf of his patient. The administrator supplies these services. Under the circumstances, market-like negotiations between the supplier and demander are impossible, as are lengthy determinations of an exact price for the patient prior

to treatment. The problem is not merely that the patient is usually heavily insured, but instead, the unusual agency relationship that requires the physician to make noneconomic decisions on the patient's behalf. As Harris states,

There is a special negative externality in an arrangement in which one makes repeated marginal decisions about life and death. This externality is so important that the physician's participation in the "market" for angiograms and code calls is explicitly foreclosed. Whether or not it is justified, this notion has an important influence on the way the hospital is organized. (p. 473)

The hospital, according to Harris, solves the rationing problem with a variety of nonprice decision rules. "There are loosely enforced standards, rules of thumb, side bargains, cajoling, negotiations, special contingency plans, and in some cases literally shouting and screaming" (p. 478).

**IMPLICATIONS** Harris's model provides three implications. First, given the role of physicians, we can expect that the hospital's preferences for new technology will be driven by the preferences of the physician demanders. The physician-agent will tend to prefer technologies that are complementary to existing capacity and that are thus quality enhancing. Second, hospital regulation aimed primarily at the trustee-administrator group may have little effect. That is, regulation to limit hospital costs must establish incentives for and constraints on the physician-agent as well as the administrator. Third, reorganization of the hospital along product lines might make it a more effective organization. For example, reorganizing departments by cardiology, neurology, and so on may better help to integrate the physician into the decision-making apparatus.

### Summary of Models of Hospital Behavior

While many theories of nonprofit hospital behavior exist, we have described some main types. Most commonly, researchers depict hospitals as either utility maximizers or profit maximizers. Of the utility-maximizing type, the Newhouse model is a prominent example, and it depicts the top hospital decision makers as choosing the best combination of quantity and quality of care. The middle ground is occupied by the Lakdawalla-Philipson theory, which proposes that nonprofit preferences include altruism and profit maximization. Of the profit-maximizing models, the Pauly-Redisch physicians' cooperative version is most prominent. This approach depicts the nonprofit hospital as effectively making choices that serve to maximize the pecuniary gain to physicians—the decisive set of decision makers. The Harris model forms a distinctive alternative, describing the hospital as two separate firms representing opposing interests.

The continued attempts to distinguish the theoretical models will remain a serious interest for policy in debating the nonprofit's favored status. For example, would we continue to extend tax-exempt status to the Pauly-Redisch type of nonprofit? Ironically, if we become successful in identifying nonprofit hospitals by behavioral type, we may find that hospitals of all types coexist.

### What Causes Conversion of Nonprofits into For-Profits?

About 7 percent of nonprofit hospitals converted to for-profit status between 1970 and 1995, and the conversion rate has increased in recent years. Usually, conversion means the sale of the hospital assets to a profit-oriented corporation and the use of the proceeds to fund a nonprofit foundation. Several ideas recur in this new literature. As we have seen, Lakdawalla and Philipson (1998, 2006) explain that if nonprofit hospital decision makers value profit goals and output goals, they will convert to for-profit status when opportunities for making and enjoying profits provide greater utility than any other combination of quantity and profits. Several Blue Cross organizations have converted in this way.

Thus increases in consumer demand that improve profitability will increase the chance that some nonprofits will convert. In such cases, shedding the nonprofit constraint has become more attractive. Others make essentially the same observation in noting that the secular decline in the

importance of donations to nonprofit hospitals makes continued nonprofit status less important to many decision makers. A reduction of tax benefits can also erode loyalty to the nonprofit status; the data show that tax rate changes have the corresponding effect on nonprofit market share.<sup>5</sup>

More idiosyncratic, individual motives may be just as important. Financial distress frequently motivates conversion, and the new management often succeeds in relieving the distress by providing new cash and by streamlining operations. However, case studies reveal that personal financial benefits sometimes accrue to the nonprofit decision makers who agree to the conversion. In cases when the nonprofit management goals stay in place after conversion, the conversion may be motivated by the desire to gain greater access to capital; raising equity capital is possible for for-profits but violates the nondistribution constraint for nonprofits.<sup>6</sup>

The question of whether society gains a net benefit from these increasing conversions has come to interest health economists. There are at least two possible efficiency gains. Dynamic efficiency involves better access to capital enabling the rapid development of needed facilities, and efficiency of operation. Efficiency of operation may occur if new management and control eliminate entrenched practices and streamline operations. In addition, this type of efficiency difference should be observable in cost-efficiency studies that compare nonprofit hospitals with for-profits, a subject to which we return shortly.

The experience with conversion provides another opportunity to test the implications of ownership status. Picone, Chou, and Sloan (2002), Shen (2002), and Farsi (2004) all find that a short-term decline in quality measured by mortality rates usually follows conversion from nonprofit to for-profit status. Santerre and Vernon (2005) warn that conversion, by altering the nonprofit/for-profit mix in a market area, has implications for market-level economic efficiency. They suggest that the typical U.S. hospital market has an inefficiently large number of nonprofits.

## THE RELATIVE EFFICIENCY OF NONPROFITS VERSUS FOR-PROFITS

Economic efficiency takes on several meanings. The “social efficiency” of a market structure requires that society maximize total benefits net of opportunity costs. The early criticism of nonprofits pointed to a tendency to overproduce. However, the gains from the provision of underprovided public goods can overcome this drawback.

Hospitals may be efficient in production but even so, they cannot control the other inputs into the community’s health production, such as exercise and diets. Ellis (1993) contends that the combination of nonprofit hospital care and the community’s other health production inputs will most probably produce community health inefficiently.

### Are Nonprofit Health Care Firms Less Technically or Allocatively Efficient?—Hospital and Nursing Home Studies

Frontier studies examine hospital efficiency by attempting to identify the best possible practices. A firm is observed to be inefficient when it falls short of the best possible production practice frontier or above the cost frontier. Zuckerman et al. (1994) as well as Folland and Hofler (2001), found little or no difference between for-profit and nonprofit hospitals. A meta-analysis by Hollingsworth (2008), examining over 300 studies, led him to conclude cautiously that the public hospitals were somewhat more efficient than the other two organizational forms.

Nursing homes provide another test for nonprofit efficiency. Nonprofit homes sometimes appear to be less efficient merely because they are offering a higher quality of care (see Box 13-1 for a

<sup>5</sup> Gulley and Santerre (1993) and Hansmann (1987) examine the effects of tax rates on nonprofit market shares.

<sup>6</sup> For further investigation of conversions, see also Mark (1999), Cutler and Horwitz (1998), and Goddeeris and Weisbrod (1998).

**BOX 13-1****Why Are RNs' Wages Higher in Nonprofit Nursing Homes?**

Holtmann and Idson (1993) observed, as have others, the differentially higher wages that RNs receive in the nonprofit nursing home sector. They proposed that nonprofits pay nurses higher wages to get higher quality. Using econometric analyses, the authors discovered that the differential wages reflect quality-enhancing characteristics of the nurses, such as years of experience and length of tenure on the current job.

discussion). Nevertheless, Garavaglia et al (2011), when adjusting for quality, found the for-profit homes to be more efficient. Santerre and Vernon (2007) contrast this by looking at industry level efficiency. The two authors find that nursing home industry efficiency is higher when the mix of ownership types includes a larger share of nonprofit nursing homes.

**CONCLUSIONS**

This chapter has examined the nonprofit firm in the health care sector. We began with a description of the nonprofit firm, noting the importance of the nondistribution constraint. We then asked why nonprofit firms exist. We found two principal accounts. Weisbrod explained how nonprofit firms might arise to provide public goods that are neglected by the private markets and the government. Hansmann explained how nonprofit firms might reduce or eliminate a contract failure that arises because consumers may not trust the profit-motivated firm to perform faithfully certain functions, often charitable ones. Under these theories, the nonprofits can improve the well-being of the community, overcoming the for-profit firm's tendency to underproduce in the presence of beneficial externalities.

We then investigated three analytical models of nonprofit hospital behavior. The Newhouse hospital model examines the desire to provide service to the community, with the quality of care often being as important as the quantity. The Lakdawalla-Philipson model exploits a middle ground to explain the entry and exit behaviors of nonprofits. In contrast, the Pauly-Redisch hospital model involves physicians' control, used to maximize the average physician's income. We contrasted these two hospitals, which show the range of behaviors from purely altruistic concerns for the community to pure profit-maximizing interests. We showed the irony that under pressure of competition, the differently motivated firms may behave nearly the same. The Harris hospital model casts a different and illuminating picture of what a nonprofit hospital looks like on the inside, to a physician/economist at work in the hospital.

Finally, the nonprofit and for-profit health care firms were contrasted regarding efficiency. The data from efficiency studies show relatively little difference between the ownership types.

The first half of the twentieth century saw radical changes in the structure of the health care industry, particularly in the prevalence of nonprofit firms. Since then we have gained a better understanding of the economic factors that make nonprofits attractive, and we are beginning to understand what can make them increase or decrease in number.

**Summary**

1. The defining characteristic of a nonprofit firm is the nondistribution constraint. Furthermore, nonprofits are typically tax exempt, and donations to nonprofit firms receive favorable tax treatment.
2. The Weisbrod theory for the existence of nonprofits proposes that nonprofit firms arise to fulfill unmet demands for public goods.
3. The contract failure theory for the existence of nonprofits proposes that nonprofit firms are advantageous under circumstances where it is difficult or impossible for the purchaser of the good to verify the delivery and the quality of the good.
4. Altruistic conceptions of the nonprofit are exemplified by the Newhouse model, in which the hospital



decision makers choose preferred combinations of quality and quantity of care subject to a break-even constraint.

5. A model where the nonprofit values both profit and altruistic service to the community illuminates entry and exit behavior.
6. The physicians' cooperative model depicts the hospital under de facto physician control exercised to maximize average physician income.
7. The Harris hospital model is distinct in that it adds more realism to the depiction of the hospital. Harris depicts the nonprofit hospital as the outcome of two noncooperative interacting firms represented by administrators on one side and physicians on the other.
8. Evidence suggests that nonprofit hospitals in the United States are not very different in economic efficiency from for-profit hospitals.

---

## Discussion Questions

1. What is the nondistribution constraint? In what way is the nondistribution constraint circumvented in the Pauly-Redisch model? What implication does this have for the efficiency of the Pauly-Redisch nonprofit hospital?
2. If an agency could cheaply, and accurately, monitor the delivery and quality of care by health care firms, would there be any contract failure in health care remaining? Would there be any need for nonprofits in health care? Would any arise?
3. In the Lakdawalla-Philipson model, why do the authors claim that the minimum average cost of the for-profit will determine the industry price?
4. Suppose that population growth expands the quality-quantity frontier of a Newhouse utility-maximizing nonprofit hospital. How would its choice of quantity and quality change? In your view, is a for-profit hospital likely to respond more quickly to population growth? Discuss.
5. Under which of the models of hospital behavior described in this chapter does the tax-exempt status of nonprofit hospitals make the most sense? Under which does it make the least sense?
6. Can we say which are the most efficient hospitals—nonprofits or for-profits? Which are the most efficient nursing homes? What qualifications apply to our present knowledge in each case? What is your view?
7. Why might information problems lead to consumer preferences for nonprofit provision of some goods and services? Reconcile your answer with the observation that most physician care, drug products, and many other services are provided by for-profit firms.
8. In what sense do nonprofits earn “profits” and need to earn “profits” to survive?
9. What are some cost advantages that nonprofits have over for-profits? Are there any disadvantages?
10. Explain the logic behind the argument that donations have characteristics of a public good.
11. Weisbrod and Hansmann present different theories on the existence of nonprofit organizations. Compare and contrast them in regard to the types of firms and the ways they are financed.

---

## Exercises

1. In Figure 13-1, if two additional voters had demand curves equal to  $D_0$ , what amount of the public good would tend to be provided by the democratic government? Which voters would be unlikely to promote a nonprofit? Which would be the most likely?
2. In Figure 13-1, suppose that Voter 5 comes to value the public good even more than before. Will there be an increase in the amount provided through the median voter model of the voting process? Why or why not? Suppose that Voter 5 can bribe one of the other voters to change his or her preferences. Which one will Voter 5 approach?
3. Under the physicians' cooperative model, if the supply price of physicians were to rise, how would this affect the equilibrium staff size in the open-staff case? How would it affect the optimal staff size in the closed-staff case?
4. Of the Newhouse and the physicians' cooperative models, which nonprofit hospital is likely to produce more quantity and quality in equilibrium with barriers to entry? In long-run equilibrium, with free entry and exit?
5. Consider Figure 13-3. Are these long run average cost curves ( $LRAC$ )? What accounts for the nonprofit advantage?

# Hospitals and Long-Term Care

- Background and Overview of Hospitals
- Hospital Utilization and Costs
- Closures, Mergers, and Restructuring
- Quality of Care
- Nursing Homes
- Hospice, Home Health, and Informal Care
- Conclusions

In 2008, there were 5,815 hospitals with close to 1 million beds in the United States (Table 14-1). Of the 15.5 million persons employed at all health service sites in 2009, 6.3 million were employed at hospitals. Hospital care is the most visible component of total health care spending, and the hospital remains at the center of an evolving health economy. At the same time, the aging of the population has created a major challenge for the adequate provision of long-term care. This chapter provides an overview of the hospital and long-term care sectors. It also examines several controversies. For hospitals, these include the “medical arms race,” cost shifting, hospital quality, and the effects of managed care and consolidation. For nursing homes and long-term care, we examine various quality, demand, and cost issues, as well as the possible substitution of informal care for nursing home care.

## BACKGROUND AND OVERVIEW OF HOSPITALS

We distinguish among the many types of hospitals by using four criteria: length of stay, type, ownership, and size. Hospitals are categorized as short stay (usually less than 30 days) or long term (usually more than 30 days). The community hospital is the type with which the general public is most familiar. It consists of all nonfederal general hospitals that provide acute, short-term care.<sup>1</sup>

Many community hospitals are also teaching hospitals, with residency programs approved by the Accreditation Council for Graduate Medical Education. Other hospital types are mental, including those treating alcoholism and other chemical dependencies; tuberculosis and other respiratory diseases; and other specialties (e.g., maternity, orthopedic, and rehabilitation).

<sup>1</sup> More extensive discussions of many of the topics in this section are found in Starr (1982), Temin (1988), and Raffel, Raffel, and Barsukiewicz (2002).

**TABLE 14-1 Hospital Data**

	1980	1990	2000	2008
All hospitals	6,965 (1,365) <sup>a</sup>	6,649 (1,213)	5,810 (984)	5,815 (951)
Federal hospitals	359 (117)	337 (98)	245 (53)	213 (46)
Nonfederal hospitals	6,606 (1,248)	6,312 (1,113)	5,565 (931)	5,602 (905)
Community	5,830 (988)	5,384 (927)	4,915 (824)	5,010 (808)
Nonprofit	3,322 (692)	3,191 (657)	3,003 (583)	2,923 (557)
For-profit	730 (87)	749 (102)	749 (110)	982 (121)
State-local government	1,778 (209)	1,578 (169)	1,163 (131)	1,105 (131)
Psychiatric and other long-term	702 (256)	892 (183)	631 (105)	576 (95)
Community hospitals				
6–199 beds	4,120 (341)	3,730 (314)	3,489 (290)	3,605 (269)
200–499 beds	1,393 (430)	1,369 (417)	1,179 (358)	1,135 (348)
500 or more beds	317 (218)	285 (196)	247 (176)	270 (191)
Occupancy rate <sup>b</sup>	75	67	64	68
Admissions (per 1,000 pop.)	159	125	117	118
Average length of stay (days)	7.6	7.2	5.8	5.5
Outpatient visits (per 1,000 pop.)	890	1,207	1,852	2,053

<sup>a</sup>Numbers in parentheses are beds in thousands.

<sup>b</sup>Percent of beds occupied.

Sources: U.S. Department of Commerce, *Statistical Abstract of the United States* (2011 and earlier issues); and U.S. Department of Health and Human Services, *Health, United States* (2010 and earlier issues).

Hospital ownership can be private or public (federal, state, county, or local). The former category consists of either nonprofit or proprietary (for-profit) hospitals. Table 14-1 indicates that there were 1,105 public short-stay hospitals in 2008, mainly state or local. Nonprofits dominate the remainder.

Hospital size is generally measured by number of beds. This does not mean that one doubles hospital size by putting a second bed in each room, but rather that the support services, types of equipment, and to some extent administrative staff, are related to the number of people that the hospital can house, and hence the number of beds.

Most short-stay hospitals are relatively small with fewer than 200 beds. However, the categories consisting of hospitals that have more than 200 beds account for 67 percent of all beds. The largest hospitals usually are affiliated with university medical schools and provide tertiary care in addition to primary and secondary care. Tertiary care consists of the most complex procedures such as organ transplant surgery and open-heart procedures. The typical community hospital is limited to secondary care, consisting of the more common surgical and medical procedures. Primary care consists of the kinds of preventive and curative care received by patients who are not hospitalized.

## History

Hospitals date back to ancient Egypt and Greece. Since then, places of healing in many countries were organized by religious establishments. Illness was closely associated with a lack of faith or superstition, and priests often administered care. Even today, it is not unusual for the afflicted to believe that they are being punished or cursed. and in some parts of the world, shamans and other “medicine men” are called upon to exorcise evil spirits.

Early hospitals in the United States were associated with the poor or with mental and infectious diseases, and medicine was practiced mainly at the home. This picture changed as more effective surgery became possible following scientific and technological advances in the last half of the nineteenth century. The modern U.S. hospital emerged at the turn of the twentieth century. In particular, important advances in antisepsis to help fight off infections greatly increased the probability of surgical success. Major advances in anesthesia, anatomy, and physiology and the invention of the X-ray also contributed.

Two nonscientific factors helped accelerate the process. One was the rapid pace of urbanization resulting from industrialization. Rural areas could not support sophisticated hospitals because of transportation problems and low population densities. Urbanization also created health problems, such as outbreaks of infectious disease that were much less common in rural areas and that required hospitalization.

The second factor was a financial one. Early hospitals relied on philanthropic contributions or state and local government funds. These alone would have been inadequate to support the growing numbers and costs associated with the modern hospital. Urbanization created wealth, and the rise of an urban middle class led to a greater ability to pay, as well as third-party payment through private insurance and workers' compensation, which originated shortly in the early 1900s.

The opening of The Johns Hopkins Hospital in Baltimore in 1885 was a significant milestone. Though few other hospitals would ultimately be able to emulate or compete with its international reputation, The Johns Hopkins Hospital introduced the latest advances in medical technology and hospital design. Through its affiliation with The Johns Hopkins University, it became a model of the teaching and research hospital.

## Organization

The typical nonprofit community hospital is governed by a board of trustees that selects the president and approves most major decisions. Traditionally, board members were often leading members of the community known for their ability to contribute or raise funds. In recent years, the increasing financial complexity of decisions facing the modern hospital has made a strong business background an extremely important qualification.

The hospital's decision-making power rests with the medical staff rather than the administrators or the board. To appreciate the significance of this phenomenon, consider that the medical staff in most hospitals is composed of admitting physicians, who are not hospital employees. Physicians apply for staff privileges to admit patients and perform certain procedures. Because physicians often have admitting privileges at several hospitals and bring patients to the hospital, admitting physicians have considerable influence on hospital decision making by serving on many committees relating to hospital governance and patient care. As such, the hospital has been referred to as the physician's "rent-free workshop," where the physician can direct substantial resources for patient care but is not held directly accountable for those resources. Physicians also bill separately for their services.

To deal with the conflicts and cost pressures created by the traditional system, more hospitals now rely on permanent physician-employees who are paid salaries or combinations of salaries and bonuses, the latter driven by various incentives. As these staff physicians are no longer the driving force to admit patients, hospitals with permanent staff physicians must have other means to attract patients. The source of patients for such hospitals is often affiliation with or ownership of HMOs. Hospitals also advertise directly through the broadcast or print media or purchase physician practices to gain new patients.

The hospital industry has undergone major change. Due to declining inpatient utilization, many smaller hospitals have closed while others have merged or reorganized. Hospitals face considerable pressure to join networks of providers in order to participate in managed care plans and to become diversified health care centers with expanded primary care facilities. Many hospitals have concentrated resources on freestanding outpatient surgery units and other outpatient programs such as cardiac rehabilitation. Table 14-1 reveals the extent of these changes. Inpatient admissions per

capita dropped 26 percent from 1980 to 2008, and the occupancy rate dropped from 75 to 68 percent. Over the same period, the number of outpatient visits per capita more than doubled.

Despite this dramatic shift to outpatient care, hospitals will continue to face challenges, especially as Medicare and Medicaid budgets tighten. Governments provide 57 percent of hospital revenues, although this figure can exceed 75 percent for many urban and small hospitals that rely heavily on poor and elderly patients. Proposals to reform Medicare and Medicaid also tend to reduce hospital spending and reimbursement rates. Thus, hospitals that disproportionately depend on Medicare and Medicaid are especially vulnerable. Nevertheless, in an increasingly competitive, cost-conscious environment, all hospitals are under considerable pressure to respond quickly to new incentives and opportunities.

### **Regulation and Accreditation**

Hospitals are subject to a wide variety of state and federal regulations over quality, costs, and reimbursement. Hospitals are licensed at the state level, although licensure is often focused on the adequacy of the hospital's physical plant and other inputs. Hospitals have their own quality assurance programs but federal legislation established professional standards review organizations (PSROs) in 1971 to monitor quality while limiting utilization. After considerable controversy regarding their effectiveness, PSROs were replaced in 1984 by peer review organizations (PROs) that performed case-by-case peer review and monitored Medicare utilization in hospitals and other facilities. PROs were often dominated by physicians and hospitals and their impact was questioned. In the mid-1990s, the Centers for Medicare and Medicaid Services (CMS) adopted the position that case-by-case inspection brings only marginal improvements in quality. In 2002 PROs were replaced by quality improvement organizations (QIOs), intended to monitor and improve care.

Hospitals also are subject to numerous other regulations and requirements. Many of these relate to reimbursement, such as Medicare's prospective payments system (PPS) and various forms of state rate regulation. Certificate-of-Need (CON) laws limit capital spending, and hospitals are subject to antitrust laws intended to promote competition.

In addition to meeting licensure and regulatory requirements, most hospitals and many other health care facilities seek accreditation from the Joint Commission on the Accreditation of Healthcare Organizations (JCAHO). The JCAHO is a private, nonprofit organization that was founded in 1952 and has a board dominated by representatives from physician and hospital associations. It sets standards for patient safety and quality of care. Hospitals seeking accreditation are evaluated by a visitation team, which examines hospital compliance with JCAHO standards. To maintain accreditation, the hospital must undergo an on-site review every three years.

Many third-party payers reimburse only for care provided in accredited hospitals. Although hospitals can be evaluated by federal inspectors to qualify for Medicare and Medicaid reimbursement, JCAHO accreditation also satisfies the federal requirement. The JCAHO is clearly a powerful organization, and the potential for conflict between professional self-interests and public interests is evident. The influence of the JCAHO can be used to limit hospital competition and to protect physicians against other groups of providers, such as chiropractors and doctors of osteopathy, by denying them access to hospitals or influence within hospitals.

### **HOSPITAL UTILIZATION AND COSTS**

The relentless growth of hospital costs has served as the impetus for many forms of regulation and other policy initiatives. Table 14-2 shows the increases in total hospital costs, as well as costs per day and costs per admission. Hospital costs account for 30.7 percent of national health expenditures and although they have decreased over the past 2 decades as a percentage of all health care costs, they have nonetheless increased at an annual rate of 9.5 percent since 1960. Table 14-2 also shows

**TABLE 14-2 Hospital and Nursing Home Costs**

	1960	1970	1980	1990	2000	2008
National health care spending (\$ billions) <sup>a</sup>	27	73	246	696	1,353	2,339
Hospital care (\$ billions)	9.2	28	102	254	417	718
Percent of total health spending	34.1	38.4	39.2	36.5	30.8	30.7
Percent of hospital care paid by						
Out of pocket	20.8	9.1	5.2	4.4	3.3	3.2
Private health insurance and other private	37.0	36.0	40.5	42.4	39.8	39.9
Government	42.2	55.0	54.3	53.2	56.9	56.9
Medicare	—	19.4	26.0	26.7	29.3	29.4
Medicaid	—	9.6	10.4	10.9	17.3	17.1
Other government <sup>b</sup>	42.2	26.0	17.9	15.6	10.3	10.4
Average hospital cost (\$)						
Per day	—	—	245	687	1,149	1,782
Per stay	—	—	1,851	4,947	6,649	9,788
Nursing home care (\$ billions)	0.8	4.2	18	53	95	138
Percent of total health spending	3.2	5.8	7.1	7.6	7.0	5.9
Percent of nursing home paid by						
Out of pocket	77.9	53.6	40.0	37.5	30.1	26.7
Private health insurance and other private	6.3	5.1	5.7	13.3	13.1	11.1
Government	15.7	41.2	54.2	49.2	56.9	62.2
Medicare	—	3.4	1.7	3.2	10.6	18.6
Medicaid	—	22.3	50.2	43.9	44.1	40.6
Other government	15.7	15.5	2.3	2.1	2.2	3.0

<sup>a</sup>All monetary values are in nominal dollars.

<sup>b</sup>Includes the Department of Veterans Affairs, CHIP, and state and local spending.

Sources: U.S. Department of Health and Human Services, *Health, United States* (2010 and earlier issues); and U.S. Department of Commerce, *Statistical Abstract of the United States* (2011 and earlier issues).

that a small and dwindling share has been paid out of pocket, especially after Medicare and Medicaid were introduced in 1965. Governments now pay for 57 percent of all hospital costs compared to 38 percent in 1965. Despite a growing and aging population, the cost increases do not result from more inpatient admissions. The number of hospital beds has been declining for many years (Table 14-1), and occupancy rates for community hospitals have rebounded only slightly from the 62 percent bottom reached in 1997. Rather, the cost per day and cost per admission and the shift to outpatient services, leaving hospitals with high fixed costs, are the main driving forces. The influence of these determinants of hospital costs is intertwined closely with numerous features of health care markets including changes in technology and reimbursement methods. We develop these issues and many others relating to hospital costs and health care system reform in other chapters. In this section, we address two of the many other concerns relevant to the growth of hospital costs. The first concerns the effect of competition on costs, and the second deals with cost shifting.

### Competition and Costs

Consumers generally welcome increased competition as a driver of lower prices, greater availability, and improved innovation and quality. Whether increased competition among hospitals provides similar benefits is of considerable policy and academic interest. Such concerns are interesting and

unique. With the exception of natural monopolies resulting from economies of scale, economists usually endorse competition as being in the best interest of consumers. Evidence is substantial that higher levels of seller concentration in most markets lead to higher prices and reduced choices. Indeed, this is the premise behind federal and state antitrust laws.

Some analysts suggest that the hospital market is an exception to the standard paradigm. They argue that hospital competition has encouraged an unproductive and costly medical arms race (MAR), as described in Box 14-1, with unnecessary duplication of expensive capital equipment as well as unnecessary expenditures on advertising in order to attract patients. Competition also may create pressure to fill beds through questionable admissions.

Why should hospital care be different? The answer lies with the reimbursement mechanisms traditionally used by insurers, which paid hospitals on a retrospective cost basis. Higher costs generally meant higher payments to hospitals. Unlike other industries, where sellers must compete on the basis of price for customers, retrospective reimbursement meant that hospitals were largely immune from the discipline exerted by the competitive process.

### BOX 14-1

#### Game Theory and the Medical Arms Race (MAR)

The 2002 Academy Award–winning movie *A Beautiful Mind* brought considerable public attention to John Nash and his contributions to game theory. Game theory is a powerful analytical tool used increasingly in economics and many other disciplines. It can be used, for example, to show why it may be in the best interests for each hospital to engage in a MAR even when hospitals as a whole are negatively affected. Game theory begins with a payoff matrix of the type shown in Figure 14-1. Suppose there are two large hospitals, A and B, in a market, each facing the decision of whether to add an expensive heart transplant unit without knowing what its rival will do. The payoff matrix shows the total profit for each hospital (with values for A's profit shown first) resulting from the four combinations of strategies. For example, if both adopt (the “northwest” cell), each hospital will have a total annual profit of \$100 (million). If A alone adopts (the “northeast” cell), assume that it will have a significant advantage resulting in a profit of \$200 (million), while B loses \$50 (million).

Game theory tries to predict a solution, that is, the strategy chosen by each participant. It is clear that both hospitals with a combined profit of \$300 (million) will be better off if neither introduces the unit. However, if the hospitals cannot agree (e.g., they may not trust each other or they may believe that antitrust laws preclude cooperation), game theory predicts a solution in which each hospital will adopt the unit and combined profits will be \$200 (million). Why? Given the payoff matrix, each hospital has a dominant strategy. That is, regardless of what Hospital B does, A will always have a higher profit by adopting rather than not adopting, that is, \$100 (million) versus –\$50 (million) if B adopts and \$200 (million) versus \$150 (million) if B does not adopt. Similarly B's dominant strategy is to adopt and, hence, a scenario results consistent with the MAR hypothesis.

Students of game theory will recognize this as an example of the prisoner's dilemma and the solution as a Nash equilibrium. McKay (1994) and Calem and Rizzo (1995) provide other applications of game theory to other decisions including hospital quality and specialty mix. In addition to decisions involving the acquisition of technology and introduction of new services, game theory can provide insight into hospital advertising and other forms of nonprice competition.

		Hospital B	
		Adopt	Do Not Adopt
Hospital A	Adopt	100, 100	200, –50
	Do Not Adopt	–50, 200	150, 150

**FIGURE 14-1** Payoff Matrix (millions of dollars)

This situation has greatly changed over the past two decades. Hospitals as well as insurance companies must compete for their managed care business through price and quality. Hospitals also are now reimbursed by many major third-party payers on a prospective basis at rates that are independent of their actual costs. It would thus appear that hospitals have a strong financial stake in being efficient and in avoiding capital investments that are not profitable.

Kessler and McClellan (2000) examined the effects of hospital competition on the costs and outcomes for Medicare beneficiaries who incurred heart attacks. They found that, prior to 1991, competition improved outcomes in some cases, but also raised costs. After 1990, there were substantial decreases in costs and substantial improvements in outcomes. A part of the welfare improvement resulted from the higher HMO enrollments over this period.

Elsewhere, Zwanziger and Mooney (2005) studied HMOs in New York State which until 1996 regulated the rates (determined largely by historical costs) private insurers were required to pay for inpatient care. After the 1996 reforms, HMOs were able to negotiate lower prices with hospitals that were located in more competitive markets.

Finally, we observe other effects of managed care, including its role in the large decline in the number of hospitals and beds since 1980. Dranove and colleagues (2002) describe the financial pressures created by managed care for hospital consolidation and improved efficiency. Their work demonstrates the substantial impact associated with the growth of managed care. For the average market, the consolidation between 1981 and 1994 attributable to managed care represented the equivalent of a decrease to 6.5 equal-sized hospitals from 10.4 such hospitals.

A second study by Dranove and colleagues (2008) examined whether the “managed care backlash” and consequent easing of restrictions on patient choices beginning in the mid-1990s affected the relationship between hospital prices and hospital concentration. MCOs should be able to extract greater price concessions in more competitive hospital markets especially when their members are more willing to be directed to selective hospitals. In the limiting case of a single hospital in a market, the MCO will have no bargaining ability. The research results show that the price-concentration relationship grew stronger between 1995 and 2001, but that it peaked in 2001, and possibly even reversed after 2001. The authors conclude (p. 374) that despite growing concentration, there has not been a “collapse in the price-concentration relationship. MCOs still appear to be continue to be playing competitive hospitals off against each other to secure discounts, though with possibly less effectiveness than in the peak year of 2001.”

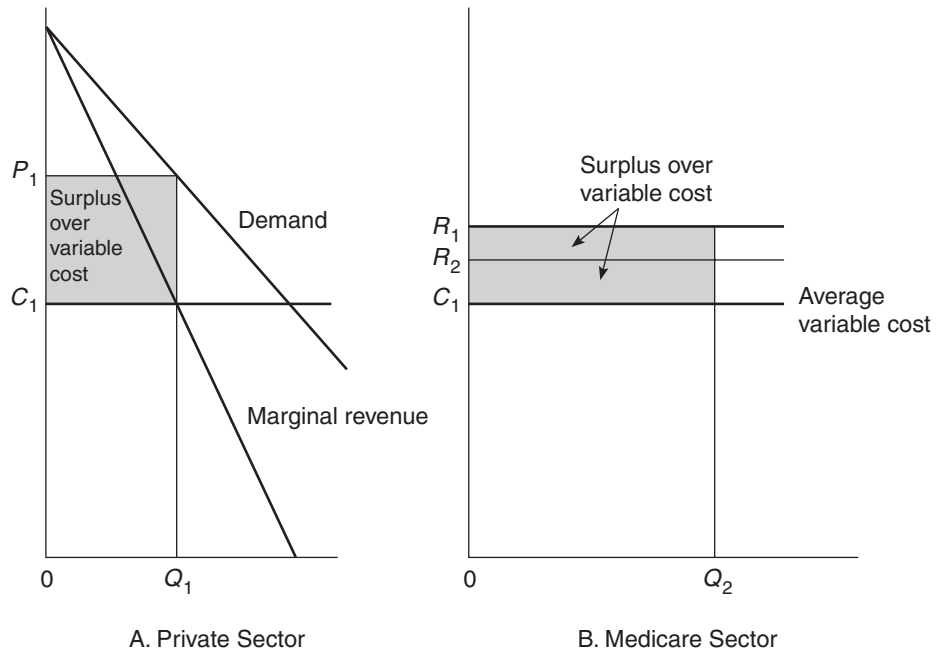
**HOSPITAL COST SHIFTING** For various legal and ethical reasons, hospitals provide substantial amounts of uncompensated care. Most of this care is provided to uninsured indigents, but uncollectibles from incompletely insured patients are also considerable. In addition, many third-party payers place stringent limits on reimbursement rates, and proposals to reduce Medicare and Medicaid expenditures typically call for further reductions. After an initial period of generous payments under PPS, by 1993 the payment-to-cost ratios for both Medicare paid Medicaid were only 90 percent, compared to 130 percent for private patients. The ratios for Medicare and Medicaid subsequently increased but dropped back to the 1993 levels by 2004.<sup>2</sup>

Are the costs of uncompensated care and “discounts” to some third-party payers passed on by hospitals to other patients as is often claimed? If Medicare and Medicaid cuts are passed on to others, there would be no savings to society but merely a shifting of the hospital cost burden. Similar shifting would occur if the number of uninsured or poorly insured increases because of an increase in part-time employment in the services sector and cutbacks in fringe benefits by some employers.

Intuition suggests that these costs are shifted. After all, services must be paid for and it stands to reason that the burden for nonpayers must be picked up by others. However, the issue may not be as simple as it first appears. To see why, we develop a model of hospital fee determination.

<sup>2</sup> American Hospital Association and The Lewin Group, *Trendwatch Chartbook 2006*, March 2006, Table 4-4.





**FIGURE 14-2** Analysis of Changes in the Medicare Hospital Reimbursement Rate

We examined a variety of hospital behavior models in Chapter 13, but for simplicity assume that hospitals maximize profits (or the undistributed residual in the case of nonprofits). Suppose also for simplicity that there are just two groups of patients: private (insured or self-pay) and Medicare. The downward-sloping demand curve for the private sector and the constant Medicare hospital reimbursement rate ( $R_1$ ) per patient are shown in Figure 14-2. The private demand curve (panel A) is negatively sloped because at least some patients economize or substitute other services as their out-of-pocket obligations increase, and hospitals that raise fees lose patients to other hospitals. Assume further that  $R_1$  covers the average variable cost ( $C_1$ ) for a fixed number of Medicare patients seeking admission (i.e.,  $Q_2$  Medicare patients in panel B) but that the rate does not necessarily cover all costs. Finally, assume that the hospital is operating below capacity, as is the case for many hospitals, and that  $C_1$  is constant over the relevant range and equal to marginal cost.

If the hospital is unable to price discriminate in the private sector, it will accept  $Q_1$  private patients (the quantity where marginal revenue equals marginal cost) and charge the price,  $P_1$ . It also will accept all Medicare patients ( $Q_2$ ) at the Medicare rate<sup>3</sup> so that the hospital treats a total of ( $Q_1 + Q_2$ ) patients. Total revenues of ( $P_1Q_1 + R_1Q_2$ ) produce a surplus over variable costs equal to  $(P_1 - C_1)Q_1 + (R_1 - C_1)Q_2$ .

Suppose that the Medicare reimbursement rate is lowered to  $R_2$ , but that the rate still covers the average variable and marginal costs so that the hospital continues to accept  $Q_2$  Medicare patients. Contrary to intuition, it makes no sense to increase prices in the private sector. Hospital surpluses diminish at prices above  $P_1$  because the hospital will lose private-sector patients whose marginal revenue exceeds marginal cost. The optimal private rate remains at  $P_1$  for the  $Q_1$  private patients, and no cost shifting occurs.<sup>4</sup>

<sup>3</sup> With capacity limitations, it will first raise the price to eliminate those private patients whose marginal revenue is below the Medicare rate.

<sup>4</sup> Cost shifting could arise if the hospital was not previously maximizing profit and was accepting “unprofitable” patients whose marginal revenue fell short of marginal cost. The lower Medicare rate would encourage the hospital to reduce the number of these patients by raising the private rate. Santerre (2005) describes the welfare loss that would arise. His estimates for 1992 indicate a maximum welfare loss of just 0.84 percent of private hospital expenditures.

This is not the end of the story because the hospital's revenues are reduced. In the long run, revenues must cover all costs, or else the hospital cannot survive. If revenues exceed costs for the hospital in Figure 14-2 after the lower Medicare rate, then the private rate may be unaffected in the long run as well. Suppose, however, that the hospital's revenues now fall short of total costs. If it cannot reduce costs, the hospital and others in the same situation may be forced out of business or forced to merge. As this happens, demand will increase for the remaining hospitals and, as a result, the private rate also could rise.

Determining the burden of lower public reimbursement is complex. The burden ultimately can be borne by many groups through reductions in the number of hospitals, lower compensation to hospital employees as the demand for their services diminishes, reduced access to care for those with public insurance or those receiving uncompensated care, and higher fees to the private paying groups.

A review of the evidence on cost shifting by Morrisey (1995) indicates that cost shifting through higher prices has taken place but that it is far from complete. One study included in his review shows that California hospitals reduced the amount of uncompensated care by 53 cents for every \$1 decrease in their discounts to third parties. This would have been unnecessary if the hospitals could have shifted the costs to others.

More recently, following reductions in Medicare payments to hospitals, Wu (2010) found relatively little cost shifting overall but large variations across hospitals. Those where Medicare reimbursement was small relative to private insurance were able to shift nearly 40 percent of the Medicare cuts. Hospitals that relied more heavily on Medicare patients were much more limited in shifting costs.

## **CLOSURES, MERGERS, AND RESTRUCTURING**

We have already referred to some of the dramatic changes affecting hospitals. The growth of managed care and the introduction of reimbursement methods that discourage inpatient care and long lengths of stay have contributed to declining inpatient utilization. In response, capacity has been reduced through the sharp drop since 1980 in the number of hospitals and beds (Table 14-1). Even so, with occupancy running at just 68 percent in 2008, excess capacity remains one of the most visible and significant characteristics of the hospital industry.

Although hospital closures can be painful to a community, the restructuring of the hospital industry should be viewed as a market response to cost-containment efforts. Nevertheless, it remains important to determine just how well the market works for this sector, and, in particular, whether inefficient hospitals are more likely to close.

Cleverly (1993) examined 160 community hospitals that closed between 1989 and 1991. Most were small, located in rural areas, and had sustained progressively larger losses for several years before closing. High costs and high prices, low utilization, and little investment in new technology were common features. From the characteristics of failed hospitals, Cleverly describes the road to failure. High prices and lack of investment in technology drive patients away. With lower utilization, costs per patient increase and cash flows become negative. The deteriorating liquidity ultimately leads to closure.

The relatively large number of small, rural hospital closings has challenged policymakers to maintain access for rural populations. To prepare for unexpected influxes of patients, small hospitals have higher rates of excess capacity and, hence, lower occupancy rates than larger hospitals. Various federal programs provide subsidies to these hospitals. Nevertheless, rural hospitals can increase their chances of survival by practicing good management and responding to competitive pressures. Succi and colleagues (1997) found that rural hospitals gain an advantage and reduce the threat of competition by differentiating their services. Those that offer more basic services and high-tech services compared to the market average are less likely to close.

With the continued growth of managed care in the 1990s and hospitals' increased reliance on managed care patients, the pace of hospital restructuring accelerated. Nearly every hospital was facing serious financial and competitive challenges. Hospitals responded by merging, participating in multihospital systems, and by forming various strategic alliances with other hospitals and physician

practices. There are two intended effects of these activities. First, by downsizing administrative units, eliminating duplication, and taking advantage of economies of scale through integration, hospitals seek to become more efficient and, therefore, more successful in competing for managed care contracts. Jantzen and Loubeau (2000) found that price and hospital participation in networks are very important to managed care organizations in awarding contracts. Second, hospitals and hospital systems, through their size and partnerships, seek to counter the pricing pressure and other demands that have been placed on them by managed care organizations.

Health economists continue to seek a better understanding of the two principal effects of the restructuring, whether hospital care is being produced at lower costs than would otherwise have been observed, and whether prices have risen as a result of less competition among hospitals. For example, Harrison's (2007) examination of closures and mergers using comprehensive national data covering the period 1981–1998 found that increased market power, rather than improved efficiency, is the principal driving force for consolidations. Consistent with these findings, Melnick and Keeler (2007) showed that hospitals that were members of multihospital systems increased their prices between 1999 and 2003 at much higher rates than nonmembers. Because these effects are undesirable from society's perspective, regulators and policymakers must be up to the challenge of assessing the restructuring efforts especially in light of evidence showing that there are cases where hospital closures on balance increase economic welfare (Capps et al., 2010).

We close our discussion of hospital consolidation with a contribution containing a rich set of results. Ciliberto and Lindrooth (2007) developed a revenue premium concept to measure the revenue each hospital in a market receives relative to competitors. Not surprisingly, hospitals are more likely to exit if they are relatively less efficient and have a lower revenue premium. The authors also found that not-for-profit hospitals, after controlling for efficiency and the revenue premium, are less likely to exit, presumably because their capital flows are not as stringently subject to market forces

## BOX 14-2

### Hospitals and Airlines: What Are the Lessons?

A provocative article, "Could U.S. Hospitals Go the Way of U.S. Airlines?," raises important and troubling questions about the potential adverse consequences of downsizing in the hospital sector. Altman and colleagues (2006) draw interesting parallels between the hospital and airlines industries, most notably through their historical lack of price transparency, limited competition, and cross-subsidies. With increased price transparency and competition from specialized low-cost airlines, the legacy airlines downsized, merged, cut unprofitable routes and capacity, and reined in wages and other costs. Despite these efforts, their financial state remains precarious.

What might happen if the hospital industry faces similar pressures? There are already strong efforts to increase price transparency through posting of prices on the Internet and other mechanisms. (See Reinhardt [2006] for his description of hospital pricing as "chaos behind a veil of secrecy.") Specialized clinics, ambulatory surgery centers, and other freestanding outpatient facilities are increasingly competing with the general hospital. The Altman article suggests that hospitals might be forced to follow the airlines' example by paring back capacity, services, staff, and the quality of care. Unprofitable patients and hospital units could be the first to go. High-cost communities that depend heavily on Medicaid and Medicare patients would be especially vulnerable if cross-subsidies were to be reduced.

The tight government budgets and cutbacks in employer-provided insurance that are likely to continue in coming years would appear to make prospects for the hospital sector even more dire. However, actual developments in the health care industries often surprise analysts. For example, Courtemanche and Plotzke (2010) show that the growth of ambulatory surgical centers, a seeming threat to hospitals, has had little effect on their outpatient surgical volume and no effect on their inpatient volume. Similarly, Bates and Santerre (2008) found that managed care organizations, another potential threat to hospitals, have not usually taken advantage of their monopsony power.

as those governing for profits. Finally, they showed that hospitals that have diversified to high tech services and outpatient care are more likely to survive.

## QUALITY OF CARE

In discussing asymmetric information in Chapter 10, we have already described how quality of care has emerged as a national priority. Hospital quality is often understood in two ways. Chapter 13 introduced theoretical objective functions for hospitals that include quality, where quality is represented in a broad sense through characteristics such as beauty of facilities, and the expertise and prestige of staff. The availability of high tech units and services also falls into this category. However, quality can also be understood in terms of hospital mortality and error rates, readmission rates, and the rates at which a hospital meets established treatment processes and protocols. Much of the current concern with quality, and efforts to improve it, revolves around the latter sorts of performance indicators. Through a variety of voluntary and mandated mechanisms, hospitals are constantly being evaluated on the premise that closer scrutiny and publicly available report cards will encourage quality improvements. Patients can now compare hospital performance measures for specific conditions and procedures under the Centers for Medicare and Medicaid Services 'Hospital Compare' initiative ([www.hospitalcompare.hhs.gov](http://www.hospitalcompare.hhs.gov)). If embarrassment about poor performance is not a sufficient motivator, hospitals have a major financial stake in raising quality to secure managed care contracts and to attract the growing numbers of patients that are choosing consumer-directed health plans.

Two reports published in the July 21, 2005, *New England Journal of Medicine* provide considerable insight into the quality of hospital care and the impact of public reporting. One (Williams et al., 2005) evaluated an initiative implemented by JCAHO in 2002 that required most accredited hospitals to report 18 standardized performance indicators for several common conditions. Seventeen indicators assessed processes of care; one was mortality. Analysis of quarterly data for heart attacks, congestive heart failure, and pneumonia indicated substantial gains in 15 of the 18 measures over the two-year study period. The gains were greatest for hospitals that had been the worst performers at the start of the evaluation period.

A second report (Jha et al., 2005) examined 10 quality indicators for patients discharged at 3,558 hospitals in 2004, also for heart attacks, congestive heart failure, and pneumonia. Mean performance scores (representing proportions of patients who satisfied the criteria) were 89 percent for heart attacks, 81 percent for congestive heart failure, and 71 percent for pneumonia. Performance varied substantially among hospitals, and many failed to provide patients with broadly accepted treatments at surprisingly high rates. It is possible that the scores reflect, in part, record keeping omissions rather than failure to provide treatment, but such omissions themselves would indicate a quality problem.

Regional variations are also startling. Among the 40 largest hospital regions, Boston led the way in treating heart attack, with a performance score of 95 percent. San Bernadino was at the bottom with 83 percent. Overall, hospitals in the Midwest and Northeast outperformed those in the South and West.

Hospital quality has not been ignored in economic analyses and Chapter 10 describes several quality-related issues. Here we close our discussion with a revealing study by Tay (2003) that examined the impact of quality and distance on (non-HMO) Medicare patients admitted to a hospital for heart attacks. Medicare patients who are not enrolled in managed care plans have a wide range of hospital choices, and price is not an issue to them. Consistent with economic theory, Tay found that distance has a significant negative impact on demand, while quality, measured by various input and health outcomes indicators, has a significant positive impact. Remarkably, some patients are willing to travel much further for higher quality, even for conditions where a small delay to treatment can have a big impact on outcomes.

The bottom line is that quality matters a great deal to patients. The challenge for the hospital is to understand the production function (and associated costs) for quality improvements and the increased demand from patients and managed care organizations that results from it.

## NURSING HOMES

The rapid growth of the elderly population in many countries has led to considerable interest in the problems associated with long-term care. Long-term care encompasses a wide variety of services and arrangements used to care for the elderly and others with serious functional impairments.

In this section, we concentrate on the nursing home. Numerous economic issues have been investigated in the nursing home literature (Norton, 2000). Following some background information, we will focus on those involving quality, cost shifting, and financing the care.

### Background and Costs

Traditionally, the elderly were cared for until death by their families so the nursing home is a relatively recent phenomenon. Raffel and colleagues (2002) trace its origins. The first “nursing homes” in the United States were the county poorhouses of the eighteenth and nineteenth centuries, established for invalids and those without families. Most of the patients were elderly, and the conditions were dreadful. Other state and local facilities evolved and some, usually with higher standards, were sponsored by religious and fraternal groups. The Social Security Act of 1935 provided funds for patients in private nursing homes, but the major increase in funding and nursing industry growth came after the 1965 passage of Medicare and Medicaid.

In 1960, nursing home care amounted to less than \$1 billion, with 78 percent coming out of pocket (Table 14-2). Twenty years later, in 1980, spending reached \$18 billion, with Medicaid’s share at 50 percent and only 40 percent coming out of pocket. By 2008, spending had grown to \$138 billion, with Medicaid contributing 41 percent (another 19 percent came from Medicare) and patients or their families paying just 27 percent. Between 1963 and 2000, the nursing home population over age 65 grew from 446,000 to nearly 1.5 million. Since 2000, it has leveled off and currently stands at 1.4 million. About two-thirds are in for-profit facilities.

The burgeoning nursing home population and the growth of costs are connected closely to Medicare and Medicaid. Medicare typically covers beneficiaries who are discharged from hospitals and require skilled nursing care to help recover from an acute illness. A skilled nursing facility provides round-the-clock nursing care and other medical supervision. Historically, Medicare reimbursed on a cost basis, but the Balanced Budget Act of 1997 mandated a shift to prospective reimbursement with fixed per diem payments determined by the category in which a patient is placed. The goal of this change was to shift the financial risk to nursing homes.

Medicaid, in contrast, pays for the long-term care of the poor including the nonelderly as well as elderly. It covers both skilled nursing care and intermediate (custodial) care. Because Medicaid is administered by the states subject to federal requirements, eligibility requirements and payment methods can vary widely. In the 1980s, many states phased in prospective payment systems that distinguished only between patients in skilled and intermediate care facilities. The rates were often set too low for heavy-care patients, discouraging nursing homes from admitting them. To provide better incentives, states increasingly are adopting case-mix reimbursement systems in which payment depends on a more extensive classification of patient types.

A 1987 legislative change created new standards that drove costs higher. All nursing homes participating in federal programs, including Medicaid-only facilities, must now meet the same standards as Medicare’s skilled nursing facilities. Also, nursing homes participating in federal programs must evaluate each resident’s needs and “provide services and activities to attain or maintain the ‘highest practicable level’ of function (physical, mental, and psychosocial well-being).” These requirements led to large increases in staffing and the use of rehabilitation services.

## Quality of Care

Despite increasingly stringent state and federal requirements, the public cares most about quality. To many, nursing homes evoke powerful images of neglected and poorly treated patients. It is thus natural that health economists and other scholars have concentrated their efforts on nursing home quality.

As we have seen from those chapters that deal with quality issues, researchers must find ways to define and measure quality. Regulatory standards for nursing homes tend to focus on structure measured by the quantity and categories of inputs used to provide care, rather than evaluating how the care is delivered (process) or monitoring outcome indicators such as patients' satisfaction or their quality of life.

It follows that the most comprehensive and readily available information on nursing homes and other health care institutions comes from input data—their types and costs. Scholars have used these data to examine the relationship among quality and nursing home size, ownership, expenditures, and source of payment.

One would expect a positive association between size and quality as a result of economies of scale and scope. If such economies occur over some range of output, larger institutions would be able to provide the same quality of care at a lower cost, or a higher quality of care for a given level of spending per patient. Davis's (1991) review of a large number of studies, including those that used process and structure measures of quality, suggests that no clear relationship exists.

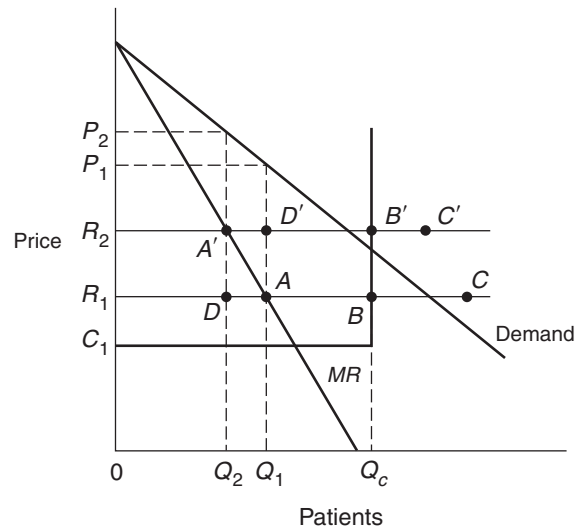
A more important issue involves quality and cost or expenditure per patient. Do “you get what you pay for” in nursing care? Using structure measures of quality, not surprisingly, most analysts find a positive relationship, but these results are not meaningful. If inputs represent quality, one expects a close relationship between various measures of input and quality. Of the 18 process and outcome studies included in Davis's literature review, only 6 indicate a positive relationship between quality and cost or inputs, while the results in 11 are insignificant (one is negative).

Subsequent research reached similar conclusions. In one of the largest studies of long-term outcomes, Porell and colleagues (1998) found little relationship between quality and facility characteristics, such as size and ownership type. They even determined that higher RN staffing does not improve outcomes. Except for nursing homes at the highest staffing levels, where the top 10 percent were significantly better, Schnelle et al. (2004) also found few quality differences.

The failure to find positive relationships consistently is troubling. It indicates that improvements are needed in measuring quality as well as in formulating the statistical models used to estimate the relationships.

A third area of interest is in the relationship between type of ownership and quality. The previous chapter discussed the concept of contract failure within the context of nursing homes. Put simply, contract failure arises when quality is not easily observable. In the case of nursing homes, patients may believe that nonprofit organizations are more likely to serve their interests than ones motivated by profits. Is this view justified by the evidence? Nursing home costs per patient are higher for nonprofits (which have higher proportions of private-pay patients) so that structure measures are clear on this point. However, because analysts have not been able to detect an unambiguous positive relationship between quality and costs, it follows that they would have great difficulty in detecting any relationship between type of ownership and process or outcome measures of quality. Clearly, the relationship between quality and ownership is a complicated one, where profit status possibly plays a secondary role relative to other factors (Decker 2008).

Finally, Davis also reviews the literature on quality and the proportion of public pay (Medicaid) patients. Many believe that nursing homes dominated by Medicaid patients are inferior. Expenditures per resident are lower in homes with higher proportions of Medicaid patients so that structure measures unequivocally support a negative relationship between quality and the proportion of Medicaid residents in a nursing home. Troyer (2004) found that Medicaid resident mortality rates were 4.2 percent and 7.8 percent higher than those for private-pay residents after one and two years,



**FIGURE 14-3** Analysis of Changes in the Medicaid Nursing Home Reimbursement Rate

respectively. However, these differences declined when the analysis included controls for resident, market, and facility characteristics. It appears that, while Medicaid patients may be concentrated in lower quality nursing homes, the care given to patients does not vary by payment source.

### Excess Demand

For many years nursing home observers have characterized the nursing home industry as having excess demand, and have argued that excess demand is one of the reasons for the allegedly inferior quality of care provided to public pay patients. Economists are naturally intrigued by, and at the same time skeptical of, claims of persistent shortages of any commodity. Put simply, they believe that prices, and ultimately supply, will increase to eliminate the excess demand.

To examine the possibility of excess demand for nursing home care, Figure 14-3 introduces a variant of an approach originally developed by Scanlon (1980). It shows the demand and cost conditions for a representative nursing home. The demand curve reflects only the private demand (self-pay or insured), while  $R_1$  represents the Medicaid reimbursement rate. The segment  $AC$  along  $R_1$  shows the number of Medicaid patients seeking admission. For simplicity, assume a constant (horizontal) marginal and average variable cost ( $C_1$ ) up to the capacity level ( $Q_c$  patients) where it becomes vertical (no more patients could be served at any cost).

Under the conditions represented here, the profit-maximizing nursing home will first select all private patients whose marginal revenue exceeds  $R_1$  and then fill the remainder of beds with Medicaid patients. The nursing home admits  $Q_1$  private patients paying a price,  $P_1$ , and  $(Q_c - Q_1)$  Medicaid patients, leaving an excess demand of  $BC$  Medicaid patients.

The shortage can be reduced or even eliminated by raising the Medicaid rate. At  $R_2$ , nursing homes will raise the private fee to  $P_2$  and substitute  $A'D'$  Medicaid for  $AD$  private patients. The excess demand is reduced to  $B'C'$  from  $BC$ .<sup>5</sup> Conversely, reductions in the Medicaid rate will lower Medicaid admissions and the private fee but increase the excess demand. Similarly, an increase in private demand resulting from higher incomes or more prevalent private long-term insurance will reduce Medicaid admissions and increase the excess demand.

<sup>5</sup> The rate that would eliminate the Medicaid excess demand is found by sliding the segment  $AC'$  further up the marginal revenue curve until the quantities such as  $B'C'$  are eliminated. Clearly, the rate must be above  $R_2$ .

One additional aspect should be considered. If the nursing home industry is profitable, and evidence suggests that it is, one would expect entry of additional nursing homes to reduce the excess demand. However, nursing homes also are subject to Certificate-of-Need (CON) regulations, and it has been argued that legislators have intentionally used CON and other restrictions to limit nursing home entry as a way of limiting Medicaid spending.<sup>6</sup> There is a similar theme of legislative aversion to policies that would raise the Medicaid rate which, as shown previously, increases public spending while squeezing out private patients.

Various tests have been proposed for the shortage hypothesis. For example, in our model, changes in reimbursement rates do not affect total utilization, just the composition between private and Medicaid patients. Any new beds will more likely be filled by Medicaid than by private patients because that is the population for which there is excess demand. Using national data for 1969 and 1973, Scanlon's empirical tests indicate considerable excess demand for Medicaid patients. However, states may vary widely in their willingness to fund public patients, and changes may have taken place in more recent years to reduce the shortages. For example, Nyman's (1993) estimates for 1988 for Minnesota, Oregon, and Wisconsin generally do not support the shortage hypothesis.<sup>7</sup>

### Financing Long-Term Care

The continuous growth of the population that will need long-term care (LTC); the requirement that patients must meet income, asset, and home equity tests to qualify for nursing home benefits under Medicaid; and the budgetary problems created by the growth of Medicaid spending have led to many proposals to reform Medicaid. The need to deplete one's assets is especially burdensome to the middle class. The Deficit Reduction Act of 2005 even extended the "look-back" period, where Medicaid examines the recipient's financial records, from three to five years. As a result, a variety of proposals have been introduced to help resolve this contentious issue. These range from proposals allowing individuals to have higher incomes and retain a higher proportion of their wealth to qualify for public assistance, to those that would cover everyone who meets certain medical requirements.

Federal policy to contain public spending centers on two strategies: (1) encourage home care and other potentially less costly substitutes for nursing home care, and (2) encourage more private coverage for LTC. The market for private LTC insurance is growing, and legislation passed in 1996 allows employers 100 percent deductions on their contributions to group plans. The legislation also provides some tax relief to individuals who itemize their returns and purchase tax-qualified policies. Nevertheless, only 10 percent of the elderly have LTC coverage. Lapse rates are substantial (Konetzka and Luo (2010) provide a detailed analysis), and 2009 marked the first year in which the total number of existing LTC policies did not increase.

What are the barriers to growth of private LTC insurance? They include confusion over the benefits provided by Medicare and private health plans, perceptions that such insurance is unaffordable, and perceptions that governments will somehow meet LTC needs. The low probability (about 1 percent) that employed workers will need long-term care before they retire, coupled with uncertainty over future premiums and medical technology, increase the difficulty of selling LTC insurance to the working population. To overcome these barriers, some analysts propose additional tax incentives by allowing all taxpayers a 100 percent tax deduction for premiums or allowing tax-deductible LTC accounts, similar to retirement accounts. The Medicare drug legislation, effective in 2004, allows individuals to create health savings accounts that could be used to pay for qualified LTC premiums.

---

<sup>6</sup> CON regulations require health care providers to obtain approval from state planning agencies for capital expenditures exceeding a threshold level such as \$500,000. CON was federally mandated until 1987 and then left to the states.

<sup>7</sup> Recent estimates of the effects of changes in public programs are also very revealing. Grabowki and Gruber (2007) analyze national data over a 20-year period covering wide variations in Medicaid policies and in policies within and across states. Their work suggests that the demand for nursing home care is inelastic with respect to public policies.



The more visible strategy to contain public spending is the federal government's emphasis on alternatives to nursing home care. The proportion of the elderly found in nursing homes has been decreasing. The elderly population grew nearly 18 percent between 1985 and 1995 and the number of nursing home residents also increased, but there was a striking 8 percent drop in their use rate. Bishop (1999) estimated that had the 1995 nursing home share remained at the 1985 level, a quarter of a million additional elders would have been placed in nursing homes. Where have all these patients gone? The high rates of growth of home health care and other LTC arrangements may provide an answer.

## **HOSPICE, HOME HEALTH, AND INFORMAL CARE**

The budgetary pressure of caring for the growing elderly population in hospitals and nursing homes has promoted interest in other less costly arrangements. Hospice and home health programs are perceived to be cost effective and are heavily funded at the federal level.

Hospice care is intended for the terminally ill. Most hospice patients receive care in their own homes, but the use of special facilities is becoming more prevalent. In hospice, an interdisciplinary team of health professionals provides individualized care that emphasizes patients' physical and emotional comfort (i.e., palliative as opposed to curative care), as well as support for family members. Hospices strive for improved quality of life in a patient's final days and death with dignity.

Medicare introduced hospice benefits in 1983, but higher reimbursement rates in 1989 accelerated growth in the number of hospices. Higher reimbursement rates will increase significantly the number of Medicare-certified providers, improving access for Medicare beneficiaries. Hamilton (1993) estimated that every dollar increase in the daily rate raises the probability of certification by 1.7 percent. There were about 106,000 hospice patients at any time during 2000. With evidence that hospice programs offer savings, many private insurers also have added coverage for hospice care.

Home health care, the larger and more important program, provides care to patients with acute and long-term needs, including those with disabilities, those recuperating from a hospital stay, and even the terminally ill. The home care benefit was included in the initial Medicare legislation of 1965 that was extended in 1973 to certain disabled persons under 65. By 1996, about 2.4 million home health care patients were on the rolls of agencies at any time, and their numbers were growing rapidly. With a doubling of patients in just four years between 1992 and 1996, home health care became one of the fastest-growing components of total spending. Between 1992 and 1996, Medicare spending for home health care grew from \$7.7 to \$18.1 billion. Medicare spending, however, fell to \$7.6 billion by 1999 following the Balanced Budget Act of 1997. As with nursing home care, the act mandated a shift from a cost-based to a prospective-based system of reimbursement, and installed an interim system of payment limits over the intervening period. By 2000, there were only 1.4 million home health care patients at any time and national spending on home health care stabilized at \$31 billion. Since then, national spending again climbed rapidly, reaching \$65 billion, with governments accounting for 79 percent, by 2008.

The rationale for public funding for home health care rests on the premise that it is much less expensive than either hospital or nursing home care. Even though a home health visit is unquestionably far less costly than a day spent in an institution, the effect on total health spending is not entirely clear. The principal issue is the extent to which home health substitutes for "unpaid" care by family members and other caregivers, or for institutional care.<sup>8</sup> Policymakers are concerned that more generous public funding for home care will substitute for previously "unpaid" care without significantly increasing the overall care for patients. Yoo and colleagues (2004) have shown that formal and informal LTC are close substitutes. From data for 15 OECD countries, they estimated that the availability

---

<sup>8</sup> An extensive literature is available on the impact of home care on hospital utilization. Hughes's (1997) meta-analysis found that home care unambiguously reduces hospital days for the terminally ill. For others, the effect on hospital days is negative though not overwhelmingly large. See also Forder (2009) and Bonsang (2010) for recent analyses of various substitution effects in European countries.

of a spouse caregiver, measured by the male-female ratio, reduces LTC spending by \$29,000 (in \$1995), a figure that exceeded the annual Medicaid payment to an intermediate care facility.

The substitutability of informal care with LTC is confirmed by Van Houtven and Norton (2004), who introduce a utility-maximizing framework to analyze the informal care decisions by children and their elderly parents. Children select the optimal amounts of consumption, leisure and informal care, subject to their budget constraints. The utility function includes the parent's health status. The utility functions maximized by parents include consumption, formal medical care, and the informal care available from their children. The simultaneous maximization process predicts that informal care could be either a substitute or complement to formal care, and that the substitute or complement effects may vary with the type of formal care consumed (e.g., substitute for LTC and complement to inpatient/outpatient hospital care). The results of their empirical estimation indicate that informal care provided by children is a net substitute for both LTC and hospital care and physician visits. It is a complement to outpatient surgery.

Does it matter whether the caregiver is a son, daughter, or someone else; whether the adult child is married; or whether the recipient is married? Subsequent work by the same authors (Van Houtven and Norton, 2008) focused on these and other differential effects. Informal care by adult children is a net substitute for skilled nursing home and inpatient care. Children are less effective for married than for single recipients and children of single elderly are more effective than other caregivers. However, the gender of the adult child caring for a parent does not matter.

The policy and cost implications of the significant substitutability of informal care for LTC are substantial. Social changes that reduce children's abilities or commitments to care for their parents would further strain private and public budgets. Conversely, well-designed programs that provide incentives for children or other family members, especially those who are not in the labor force, to care for parents could lead to substantial government savings.

## CONCLUSIONS

Following a discussion of the history and organization of hospitals and hospital costs, we examined two economic and policy issues—the medical arms race (MAR) hypothesis and cost shifting. Careful analysis indicates that both issues are far more complex than they first appear. Common perceptions of a wasteful MAR and complete cost shifting do not accurately represent how hospital markets function. We also examined the determinants of consolidation and exit in the hospital industry, and some of the consequences of this restructuring. There are significant deficiencies in the quality of hospital care, and improving quality is one of this nation's most important goals.

In the long-term care sector, we focused on nursing homes, emphasizing three issues: quality, especially for Medicaid patients; shortages; and financing nursing home care. We found that economic theory and empirical evidence can provide useful and sometimes surprising results. For example, no clear relationship exists between costs and quality. It is also possible to have a persistent shortage of nursing home beds without any mechanism, such as price, that would alleviate the shortage.

Finally we examined potentially less costly alternatives to nursing homes and the role of informal care. Home health care represents one of the fastest-growing components of health care spending. Informal care and LTC are close substitutes, so policies that encourage informal care could substantially reduce public spending for LTC.

---

## Summary

1. The modern hospital evolved at the turn of the twentieth century following the invention of the X-ray and significant advances in antisepsis, anesthesia, and the biological sciences.
2. Hospital spending has grown rapidly in recent decades as a result of the growth of private and public insurance and other factors. It accounts for about 31 percent of health expenditures.

3. The hospital industry has experienced rapid change, including reductions in the number of hospital beds and inpatient utilization and significant growth of outpatient services. Hospitals are facing competitive pressure to restructure through mergers, participation in hospital networks, and other partnerships.
4. Hospitals are licensed and subject to a wide range of state and federal regulation.
5. Many analysts believe that the hospital industry is in a medical arms race resulting in unnecessary duplication of expensive technology. The limited empirical evidence does not support this view.
6. Intuitive reasoning suggests that the costs of discounts or uncompensated care to some patient groups must be passed on to other paying groups. More formal analysis leads to a richer set of results including situations where costs cannot be shifted. The empirical literature indicates that cost shifting is far from complete and the welfare loss associated with any shifting is relatively small.
7. Increased market power, rather than improved efficiency, is the principal driving force for consolidations. Less efficient and less profitable hospitals are more likely to exit.
8. Improvements in the quality of hospital care, and the role of publicly available performance indicators, have emerged as major policy and research themes.
9. Patients are concerned about and react to hospital quality differences.
10. The nursing home population has grown dramatically since the introduction of Medicare and Medicaid. Governments pay 62 percent of all nursing home costs.
11. Nursing home quality has been examined through structure, process, and outcome indicators. Surprisingly, no conclusive evidence relates cost to quality or supports the view that nursing homes with higher proportions of Medicaid patients produce lower-quality care.
12. A familiar theme in the nursing home literature is one of persistent excess demand by Medicaid patients. A model of chronic excess demand is plausible, although recent evidence indicates that excess demand is not a universal phenomenon.
13. The nursing home population has leveled off, but financing nursing home and long-term care remains a great social challenge. Medicaid has tightened its eligibility thresholds for nursing home care.
14. Home health care and other alternatives are growing rapidly in number of patients and costs. They can be cost-effective alternatives to hospital and nursing home care.
15. Informal care and long-term care are close substitutes. By encouraging or discouraging informal care, policies have significant budgetary implications for Medicare and Medicaid.

---

## Discussion Questions

1. Explain why it is often claimed that hospitals compete for doctors rather than patients. What are some of the implications of this phenomenon, assuming that it is true?
2. Even nonprofit hospitals must earn a “profit.” Evaluate this statement.
3. What is the medical arms race (MAR) hypothesis? What features of hospital markets make the presence of an unproductive MAR possible?
4. Suppose that the Medicare rate of hospital reimbursement is reduced. Explain why the costs may not be shifted to other patients in the short run.
5. Explain why only about 5 percent of adults buy long-term care coverage.
6. Hospital costs have grown following the growth of private and public insurance. Describe other factors that could account for some of the growth.
7. The headline of an August 21, 2005 article in the *New York Times* was “It’s the Simple Things, but Some Hospitals Don’t Do Them.” Use the “Quality of Care” section to discuss and explain why hospitals may fail to provide some simple and effective life-saving procedures.
8. The article in Box 14-2 describes similarities between the hospital and airlines industries. What are some significant differences that may prevent or minimize some of the outcomes for hospitals that are suggested by that article?
9. Explain how excess demand for nursing home beds may persist over long periods. How can the hypothesis be tested?
10. Nonprofits are dominant in the hospital industry, while for-profits dominate the nursing home industry. Develop some possible explanations for this difference.
11. Informal care provided by children and other family members are good substitutes for LTC for parents. Describe some potential social and demographic changes that may reduce the availability of such informal care. Develop policies that may help take advantage of the substitutability to delay entry of the elderly into LTC facilities.

---

## Exercises

1. What is a dominant strategy in game theory? Using Figure 14-1, replace the entry in the first row and second column with (125, -50). Does A have a dominant strategy? Does B? What is the solution to this game?
2. Assume that there are three groups of hospital patients (instead of two as shown in Figure 14-2): private, Medicare, and Medicaid (which has a lower fixed rate of reimbursement than Medicare). Explain how a hospital would select patients in order to maximize profits.
3. Suppose that Medicaid's hospital reimbursement rates do not cover the variable costs of patient care. Will a profit-maximizing hospital accept Medicaid patients? If not, under what circumstances will the hospital accept such patients?
4. Assume that a patient's health can be improved by home care or institutional care. Use isoquants and isocost curves to determine the condition for efficient utilization of these two forms of care. Under what circumstances will an insurance program promote or fail to promote efficiency?
5. Consider Figure 14-2. Suppose that a hospital has the ability to be a perfect price discriminator in the private market, that is, it can charge every private patient the maximum that the patient is willing to pay. Explain how this will affect the number of private patients the hospital will take and the impact on the hospital's profits.

# The Physician's Practice

- A Benchmark Model of the Physician's Practice
- Physician Agency and Supplier-Induced Demand
- Modeling Supplier-Induced Demand
- Diffusion of Information and Small Area Variations
- Other Physician Issues and Policy Puzzles
- Conclusions

Physicians both provide important labor in medical care production and make key decisions in the health system. In Chapter 16, we examine the characteristics of physicians as labor in the production of health, including the investment decisions they must make to become physicians. In this chapter, we focus on the physician as decision maker, whether in solo practice or in a group practice. Since the physician largely controls and directs the use of medical inputs as an agent for the patient, physician practice decisions heavily influence the quantity, quality, and cost of the health care system.

Physician practice behaviors, however, pose many challenges for economic understanding. Physicians make decisions that determine the content of hospital care and these hospitals are often called the “doctors’ workshops” (Pauly, 1980). As agents for the patients, physicians potentially neglect their roles as representatives of the community. The doctor’s large information advantage over the patient leads researchers to question whether the typical physician is really a “perfect” agent, or perhaps, at times, even induces patient demand in order to increase income.

Patients expect scientifically based medical practice, and this raises yet another information question: Is each doctor well informed regarding current best medical practice? Variations in surgery rates across small, seemingly identical, geographic areas have caused many observers to be skeptics in this regard.

With so many departures from the usual economic model of the firm, health economists have offered a variety of physician practice models to capture these departures. Instead of a review of these models and variants, we first develop a general “benchmark model” which will show the important hypotheses as special cases.

### A BENCHMARK MODEL OF THE PHYSICIAN'S PRACTICE

The following model, introduced by McGuire and Pauly (1991), captures most of the observed physician behaviors as special cases of a utility-maximizing physician decision maker. Describing a physician as a utility maximizer just means that the physician values items besides profit. We can conceptualize a profit-maximizing firm in conventional theory as a utility maximizer who emphasizes profits far beyond other goals. In this benchmark model, the physician gets utility from (1) net income and (2) leisure, and disutility from (3) inducement, the physician's own efforts to induce patients to buy more care than appears medically necessary. This last factor introduces the supplier-induced demand (SID) controversy into the model. As we will see, the question regarding SID is whether physicians use their knowledge advantage to abuse their agency role for monetary gain.

Let the physician's utility function be:

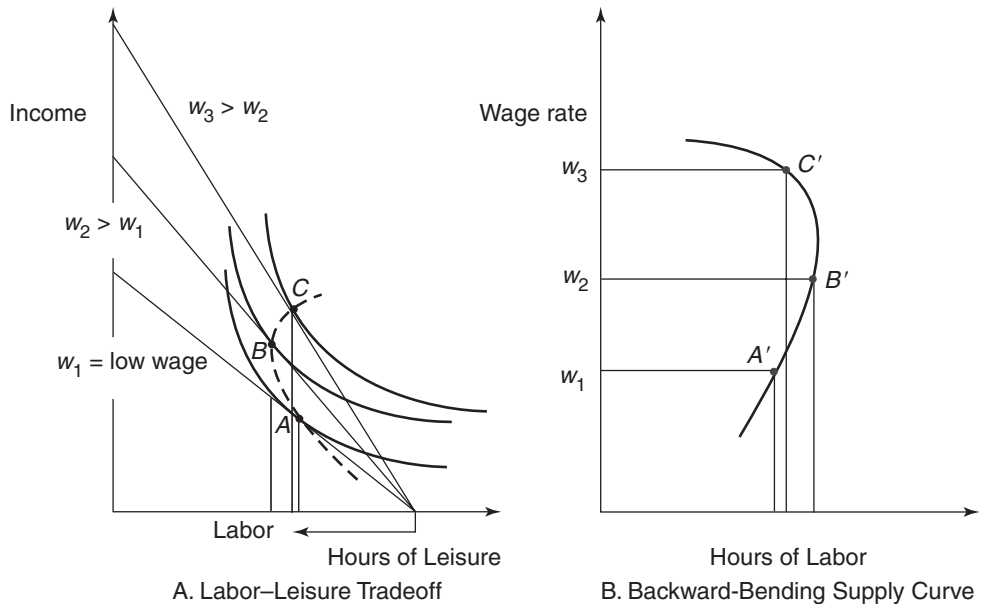
$$U = U(\pi, L, I) \tag{15.1}$$

where  $\pi$  is the net income from the practice;  $L$  is the physician's leisure time, and  $I$  is the degree of inducement. The physician can choose any amount of labor and inducement effort consistent with the profit level implied by these choices.

With three variables to consider, the physician must consider the willingness to trade off between three pairs of goals:

- Net income  $\pi$  and Leisure  $L$ ;
- Leisure  $L$  and Inducement  $I$ ;
- Net income  $\pi$  and Inducement  $I$ .

The overall decision involves all three variables, but a good understanding of the model focuses on two pairs. First, consider the tradeoff between income and leisure. In Panel A, Figure 15-1, assume that the work hours return a constant revenue,  $w$ , for each hour worked. As depicted, a physician's



**FIGURE 15-1** Supply of Physician Labor

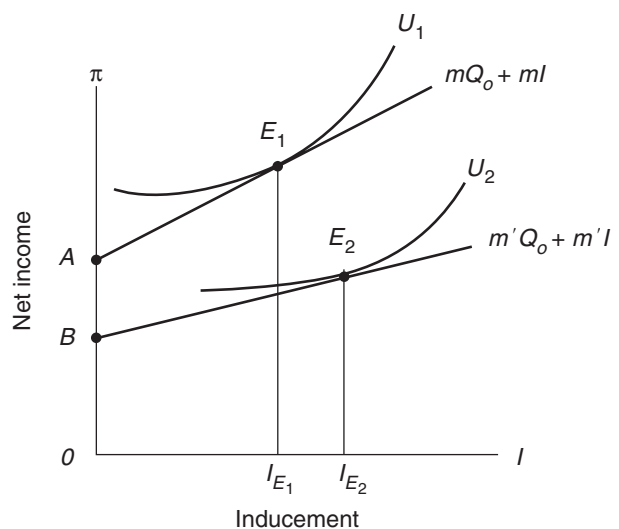
“wage,”  $w$ , determines the slope of this labor–leisure tradeoff. Starting from the maximum possible leisure (the horizontal intercept in the graph) each hour worked represents one hour less leisure. Corresponding to this, income rises by  $w$ , the net payment per hour worked. It follows that higher wage levels  $w_2 > w_1$ , and then  $w_3 > w_2$ , will result in steeper rising income lines. The physician chooses the optimal points on each income line; the indifference curves in Figure 15-1 illustrate this process.

Note that the path of optimal points—from  $A$  to  $B$  to  $C$ —at first heads to the northwest and then bends back and upward to the right. This pattern is recast in Panel B, where wage levels ( $w_1$ ,  $w_2$ , and  $w_3$ ) and labor are on the axes, in this case forming the “backward-bending labor supply curve.” As conventionally explained, in the region from  $A'$  to  $B'$ , the physician is primarily motivated by higher wages to substitute labor for leisure; but, in the region from  $B'$  to  $C'$ , the physician’s income effect dominates this substitution effect. The physician becomes rich enough to wish to spend more time enjoying the income.

These graphs show that the physician’s income and especially the income effect play critical roles in determining how much he or she is willing to work. In Panel B, suppose a physician is at point  $C'$ , and then the wage rate falls. The physician shown would move along the curves toward point  $B'$ . This implies a wish to compensate for lost income by working more. Later in this chapter, we will consider whether and how the physician would be able to find the additional patients. Notice now, however, that a rise or fall in income, for any reason, will cause the physician to re-evaluate the choice about how much to work.

Consider next the tradeoff between net income  $\pi$  and inducement  $I$ . The model proposes that physicians dislike inducing patient demand, viewing such activity as “less than professional.” With each unit of induced patient care, the physician experiences a decline in utility that must be offset by the extra utility gained from the extra income that inducement brings. Figure 15-2 represents the process of choosing levels of net income  $\pi$  and inducement,  $I$ . The indifference curves represent the physician’s preferences. The indifference curves slope upward because one of the two “goods,”  $I$ , is really a “bad.” To remain on the same indifference curve, the physician must gain additional net income to offset the disutility of engaging in a higher level of  $I$ . As is the usual case, however, higher curves are preferred.

In Figure 15-2, net income  $\pi$  with zero inducement is at point  $A$ . This point represents  $mQ_o$ , where  $m$  is the assumed profit rate (related to, but not necessarily identical to the physician’s wage) from each unit of patient care, and  $Q_o$  is the amount of patient care with zero inducement. Net income  $\pi$  increases by rate  $m$  with each unit of inducement  $I$  along the income line  $mQ_o + mI$ . The physician’s initial equilibrium is determined by the tangency of the net income line and the indifference curve, shown at  $E_1$ , where the physician induces  $OI_{E_1}$  of extra patient care.



**FIGURE 15-2** Physician’s Response to Reduced Rate of Profit

With a lower profit rate  $m' < m$ , the net income line becomes flatter and lower, or line  $m'Q_o + m'I$ . The physician now chooses equilibrium point  $E_2$ . In this case, it results in  $OI_{E_2}$ , a higher level of inducement than  $OI_{E_1}$ , even though the doctor had a higher profit at  $E_1$ .

We will apply the model to explain the economic controversies over the nature of physician practice. The model explains that physicians can induce demand but that they dislike doing so. We will apply it to understand two of the most prominently discussed controversies surrounding physician behavior: supplier-induced demand and small-area variations.

### Do Physicians Respond to Financial Incentives?

The benchmark model assumes that physicians seek profits among other goals. There is ample evidence that suggests physicians do respond to profit incentives. One sort of evidence describes physician output changes when facing clearly different reimbursement methods. For example, when physicians are paid per service provided, they provide more services than when they are given *capitation*, a fixed total payment (Nassiri and Rochaix, 2006). Studies also suggest that physicians respond to income pressures on their practice by striving to increase their incomes. In Norway, patients are required to register with a physician, creating a patient list for each physician. Physicians with short lists tend to grow their lists more than average during the next five years to make up for reduced income (Iversen, 2004).

Analysts have found that OB/GYN physicians will recommend caesarean section, a more lucrative treatment, when their practice incomes are threatened by competition (Gruber and Owings, 1996). Others find that more profitable settings for surgery tend to be used more (Plotzke and Courtemanche, 2011). One study finds that patients whose physician receives reimbursement under a capitation system may get fewer services and thus perhaps lower quality care (Quast, Sappington, and Shenkman, 2008). There is evidence that physicians have personal goals for income and adjust their practice prices and qualities when these goals exceed their current income (Rizzo and Zeckhauser, 2003, 2007). There is also evidence that when government provides incentives for physicians to choose more cost saving methods, physicians will tend to save costs (Ho and Pakes, 2011).

These studies support the idea that physicians respond to financial incentives. They also raise deeper policy questions, the foremost of which is whether the physician responses harm the social welfare of the community. To answer this, we need to learn whether physicians in fact can induce demand. Moreover, are physicians appropriately responsive to patients' preferences? The evidence regarding this question relates directly to the issue of "supplier-induced demand," which turns on the question of whether physicians are violating agency relationships with their patients.

### PHYSICIAN AGENCY AND SUPPLIER-INDUCED DEMAND

On becoming ill, consumers hire health care professionals to serve as *agents*. Agency is common in fields like auto mechanics, law, and medicine where professionals have a great deal more knowledge of the subject at hand than the consumer. When parties have unequal knowledge, we refer to the problem as *asymmetric information*. In medicine, we identify the physician as the *agent*, and the patient as the *principal*.

The policy concern is that out of self-interest physicians may violate their roles as agents. Health economists have defined the "perfect agents" as ones who make those choices and recommendations on behalf of the patients that the patients themselves would have made if they had the same information. We will describe physicians who knowingly induce their patients to consume other than this optimal amount of care as being in violation of agency. There are matters of degree, certainly. The physicians who induce their patients to have an unneeded and risky heart surgery, we would say, are outright fraudulent. However, if a physician recommended an unneeded follow-up visit, while technically a violation of agency, it would not warrant public oversight.

It is logically possible to observe supplier inducement that entails no violation of agency. For example, a physician may encourage a patient to exercise more or undergo diagnostic screening



more frequently. Inducing more care does no one harm if it encourages a move toward the patient optimum. Inducements such as better office décor or more personal care represent complements to the physician output. Such complements provide utility and are part of the package among which a well-informed patient, hypothetically, would evaluate his optimum. More personal care, for example, may even increase the probability of good health; in this case the complement to care in question is an increase in the quality of care.

## MODELING SUPPLIER-INDUCED DEMAND

Health economists have modeled supplier-induced demand for at least two reasons. First, one wishes to understand the motivations of physicians, how their incentives affect their practice. Second, models help to understand the data we observe. Furthermore, the history of this subject once centered on the question of whether “inducement” evidence was or was not consistent with the neoclassical model of markets. To begin, let us acknowledge that the asymmetric information advantages give the physicians the power to misuse their agency relationship with the patient to personal advantage.

Reconsider Figure 15-2 and note that the physician depicted at equilibrium  $E_1$  has chosen to induce demand by  $OI_{E_1}$ . If such inducement is very common, it may be very difficult to detect because when everyone induces, it becomes normal. Economists have focused instead on identifying changes in consumer demand caused by a reduction in the physician’s profit rate, usually stemming from increased competition. In this case the theory explains the increased inducement in that physicians seek to recover lost revenue by promoting procedures to patients. It also provides a test to detect inducement.

In Figure 15-2, assume that new competition brings the initial profit rate  $m$  down to  $m'$ . As seen, the physician’s response is to increase inducement from  $OI_{E_1}$  to  $OI_{E_2}$ . Thus, we predict that physicians will increase their inducement of demand in response to more competition.<sup>1</sup> Testing for changes ignores the underlying question: “At what level was the physician inducing demand to begin with?” Empirical studies invariably test for changes in inducement, leaving the underlying level unknown (Phelps, 1997).

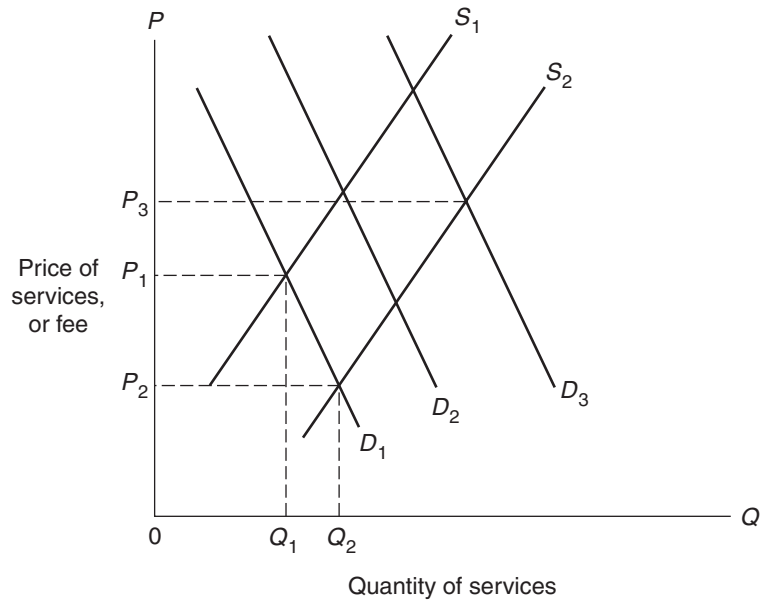
**THE SUPPLY AND DEMAND MODEL** Notice that if all physicians in the market behaved like our individual physician, an increase in the supply of physicians would lead to an increased aggregate quantity of care. Note, however, that one need not assume SID to predict aggregate demand increases in response to increased competition. A simple market supply and demand model also predicts this. From Figure 15-3, we note that an increase in supply from  $S_1$  to  $S_2$  implies an increased quantity consumed from  $Q_1$  to  $Q_2$ , and a decreased price (or fee) from  $P_1$  to  $P_2$ .

Uwe Reinhardt proposed a “fee test” to address this confusion, claiming that a rise in physician fees to a level higher than before the supply increase could be caused only by inducement. This is an interpretation of the shift in demand  $D_3$ , which is sufficient to cause fees to rise from  $P_1$  to  $P_3$ . However, Feldman and Sloan (1988) have also shown that a model incorporating quality as a demand determinant can explain the same phenomenon. If physicians respond to competition by increasing their quality, and if that higher quality earns a higher price, then SID is again not needed as an explanation.

### The Target Income Hypothesis

Economists who first proposed the SID criticism of physician behavior formulated the “target income hypothesis” (Evans, 1974). This argues that physicians have desired incomes that they strive to achieve or to restore whenever actual income falls below the targets. This target income model is a special case of the benchmark model, though a relatively extreme one.

<sup>1</sup> This response, however, is not a necessary outcome, given the assumptions of the theory. Students can experiment with alternative shapes of well-behaved indifference curves to find instances where physicians reduce inducement instead.



The supply and demand model shows quantity increases resulting from a supply increase. The Reinhardt fee test argues that SID is identified when demand increases sufficiently to cause the price (fee) to rise.

**FIGURE 15-3** The Supply and Demand Model of SID

Target income behavior suggests that for the physicians in question, income becomes not merely the main thing but the only thing. This extreme focus on an income target, as well as the disinterest in further income in excess of the target, constitute features that have caused many health economists to question the target income idea.

Would physicians adjust their price and quantity decisions if their target income changed? Suppose we asked physicians their desired income. Do physicians who express levels much higher than their current income choose different and more profitable price and quantity combinations than others? Evidence (Rizzo and Blumenthal, 1996; Rizzo and Zeckhauser, 2003) suggests that they do. Furthermore, physicians appear to change their price and quantity choices when they make gains in income from sources unrelated to their practices (for example, stocks and bonds). For one physician's perspective, see Box 15-1.

### BOX 15-1

#### SID and Target Income: A Physician's Perspective

The models of SID and target income necessarily abstract from the care process rendered by skilled professionals who are seeking to provide the most appropriate treatment for their patient. This personal perspective comes from an assignment in a health economics course at a local university. The writer is a physician who is director of pulmonary critical care at a large hospital.

I would, however, agree that physician behavior does and will change in response to income fluctuation but I would lend some insight into the current thought process. Recall, always, that the physician does not determine charges or the criteria for reimbursement for inpatient or outpatient services. The physician does not "approve" of the levels of service for an inpatient or

*(continued)*

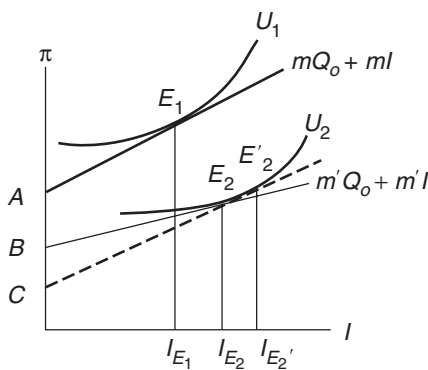
the criteria to an MRI, CT, or PET scan. Insurance companies and Medicare determine all these rules. By and large, most physicians attempt to provide the best available care, but now being cognizant of the myriad of work needed to support a claim.

Let me give an example of how this works in my private practice. I am asked to see a patient in consultation. The patient may have been post-op with respiratory failure. Without regard to patient insurance, I provide care and dutifully document my activities with the patient. After the patient recovers, I post a billing, let's say to Medicare. There was an initial consult and 10 days of care in the ICU and step-down unit. Based on my provider ID number (PIN) my claim is categorized. Routinely, I get 10–15 percent rejection due perhaps to incomplete data transmission or another doctor who also billed a similar “category” code. I then must use my staff to copy the progress notes in medical record and provide these documents to the payer. The delay is now perhaps 90–120 days.

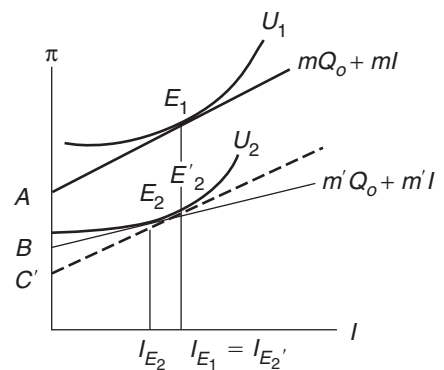
Have I put my income interest above the patient's best interest and welfare? A resounding No!! However, what physicians are actively pursuing is the details and methodology to code and document properly so as to receive payment for what has been actually done . . . I favor full prosecution of fraudulent behavior, but there is a distinct philosophical difference between billing what was actually performed and understanding the nuances of the billing coding, reimbursement world as contrasted to a doctor that would purposely “up-code” billing for services not rendered and deliberately mislead a patient. Institutions are also working aggressively to accurately document patient care and co-morbidity to increase (read legally!) reimbursement and document a higher case-mix index. Continued efforts by payers to monitor services, prescription patterns, and key benchmark care points will not only help rein in any outliers but also improve patient care and safety by eliminating wasteful care.

### The Benchmark Model as a Synthesis

We can merge the various SID models using the benchmark model. Figure 15-4 compares target income behavior in panel A with profit-maximizing behavior in panel B. In panel A, once competition forces the profit rate  $m$  to a lower level,  $m'$ , the new equilibrium is tangency point  $E_2$ . This much is identical to the graph in Figure 15-2. Now examine the broken line drawn parallel to line  $mQ_o + mI$  (the original budget line) and just tangent to the lower indifference curve,  $U_2$ , at point  $E_2'$ .



A. Physician responds to lowered profit rate  $m'$  by increasing influence on patient demand.



B. When there is no income effect, as in profit-maximizing behavior, then the model predicts a decline in supplier influence will follow a decrease in the profit rate.

**FIGURE 15-4** The McGuire-Pauly Synthesis of SID Benchmark Models

The broken line identifies the physician's changed inducement in a different way. It hypothetically removes income from the physician who is at  $E_1$  until the physician attains the equivalent utility to  $E_2$ . The resulting change in inducement caused is the *income effect* and it is measured in this case as  $I_{E_1 E_2'}$ .

Panel B shows the contrasting profit-maximizing behavior. Notice that the income effect in this case is zero; we see this in panel B because the equilibrium inducement does not change when income is removed. Similar to the previous analysis, the broken line shows this tangency at  $E_2'$ , which equals the inducement level at  $E_1$ . Why is a zero income effect like the profit maximizer? The profit maximizer gains utility only by the net dollars brought in, and the decisions he or she makes are unaffected by other matters, such as his or her income. As a result, the profit maximizer's income effect will always be zero. Notice in addition that the physician in panel B would reduce inducement when faced with greater competition and a lower  $m$ . The new equilibrium is at  $E_2$ .

The McGuire-Pauly synthesis tells us that the size of the income effect is critical to understanding and identifying SID behavior. A lower profit rate,  $m$ , has two offsetting effects on inducement:

*Substitution effect:* If inducement is less profitable (smaller  $m$ ), providers would do less inducement, that is, substitute away from it.

*Income effect:* Decreased income would make inducement more desirable.

For inducement to increase, a positive income effect on inducement must be large enough to overcome the negative substitution effect on inducement. This fact has led analysts to search for evidence of a substantial positive income effect.

### The Parallel Between Inducement and Marketing

Analysts will find it easy to conflate inducement with one or more of the many marketing practices common to virtually all firms. Marketing seeks to match consumer wants with the supplies. The "four P's" of marketing are: Product, Place, Price, Promotion. Any firm facing a setback in net revenues might choose to improve the product quality, provide its service at a new location, or use one of several ways to improve its promotion. Unfortunately, it can also provide misinformation, and it could engage in narrowly self-interested behavior, or even in fraud. The research problem is that it is difficult to distinguish socially harmful SID from the more benign forms of the inducement practices. Even advertising, which is often criticized, can be beneficial when it provides truthful information, or at least benign when it is merely complementary to the advertised product.

### What Do the Data Say About Supplier-Induced Demand?

Two criticisms were raised about much of the earlier SID work. First, many of those studies could not distinguish between the SID model and the conventional supply and demand model. Second, many estimates of the SID effect proved to be statistically flawed, meaning that the econometric coefficient thought to be evidence of SID could not be disentangled from other coefficients. In Box 15-2 Dranove and Wehner provide an example of such flaws by looking at the "supplier-induced pregnancies."

### Physician Fees, Fee Tests, and Fee Controls

Reexamining the competitive model in Figure 15-3, we see that with a sufficient degree of SID the physician's fee level can rise in response to greater competition. Recall the Reinhardt fee test for SID, which detects the presence of SID via its effect on physician fees. The premise of this

**BOX 15-2****Supplier-Induced Pregnancies**

Dranove and Wehner (1994) challenged the accuracy of standard statistical methods used by health economists to test for SID in an unusual and revealing way. They deliberately assumed—counter to all logic—that obstetricians have the power to influence demand for the delivery of babies. They further “assumed” that an increase in obstetricians per capita will lower the expected incomes of these obstetricians and impel them to use their superior knowledge vis-à-vis the patient to induce demand for childbirths. The two then applied a typical SID approach to investigate.

First they showed that availability of obstetricians and gynecologists per capita and childbirths per capita were positively correlated, an interesting result but one consistent with the ordinary long-run operations of supply and demand. They then measured the “pure SID” effect. Their estimated SID elasticity for obstetricians per capita was 8 percent and significant. Using this figure in their further calculations they found that a one standard deviation influx of obstetricians will induce an additional 7 percent in childbirths per capita. Can we conclude that the obstetrician newcomers were, let’s say, socially very active? Hardly.

Gruber and Owings (1996) also applied the fact that pregnancies are determined by parents and not by the doctors. Reasoning that the fall in fertility among U.S. women was unrelated to their model of obstetrical care, they measured the change in caesarean deliveries relative to vaginal deliveries, C-sections being more lucrative for obstetricians. They concluded that the drop in fertility and consequent effects on physician income led to an increase in C-sections. Physicians were able to recoup about 10 percent of their income drop by encouraging the C-sections.

approach is that price cannot rise in response to increased competition unless there is SID. However, Figure 15-3 applies precisely only to the perfectly competitive market structure. Many prefer to describe physician markets as monopolistically competitive with the firm having a downward sloping demand curve. McGuire (2000) showed that the implications of availability on fees in that case are not so clear. In addition, if physicians can adjust their quality in response to increased competition, then higher fees could result even when there is no SID (Feldman and Sloan, 1988).

In a more promising approach, Rizzo and Blumenthal (1996) use surveys of physicians to compare their desired incomes to their actual current incomes. They found that physicians with greater “gaps” demanded greater price increases. Similarly, Rizzo and Zeckhauser (2003) found that physicians whose current incomes fall below their reference incomes show greater income growth than the average of other physicians in subsequent periods. Evidence like this supports the idea of a target income.

Addressing somewhat similar questions, Norwegian researchers investigated the physicians’ responses when they have shorter lists of patients than the average. These physicians tend to grow their patient lists faster than average in subsequent periods (Iversen, 2004; Grytten and Sorensen, 2007, 2008). However, the physicians with short lists did not increase service production per consultation.

Nguyen and Derrick (1997) found that Medicare fee reductions led to increased volumes of care for the physicians in their sample who incurred the greatest fee reduction. Yip (1998) likewise found that Medicare fee reductions (in New York state) led to increased volumes.

One study reports on patient’s utility. If we rely on patient self-reports, we get another provocative contrast. Carlsen and Grytten (2000) found that patients in Norway enjoyed *greater* consumer satisfaction after increases in the market area’s availability of physicians. Plausibly, each physician has a somewhat smaller workload and makes up for this in part by spending more time with each patient. The two authors reason that if reported consumer satisfaction adequately reflects patient utility, then SID may not matter to patient welfare.

This section asked the question “What do the data say about SID?” It is clear from these studies that reductions in net income led physicians to induce demand in many cases. However,

none of these cases necessarily demonstrates physicians violating their agency roles. For example improving quality to attract more patients would not necessarily violate his trust with current patients.

## DIFFUSION OF INFORMATION AND SMALL AREA VARIATIONS

We have seen that what patients don't know can hurt them, and their relative lack of knowledge is especially worrisome because of the physician's dual role as provider and agent. But are physicians themselves always well informed? Do they know the true productivities of the medical procedures they provide? For health conditions with alternative treatments, and where substantial uncertainty prevails regarding productivity, physicians may vary in their preferred treatment practice styles. Do these information problems cause the substantial variations in medical and surgical use rates per capita seen across and within small geographic areas?

Economists took interest in the study of small area variations (SAV) upon recognizing its significance for social welfare. Whereas SID stems from the information gap between physicians and patients, SAV probably occurs because of the physician's uncertainty and lack of knowledge. In this case, even if all physicians acted as perfect agents for their patients, their own lack of knowledge of the true productivities of available procedures would create welfare losses for the patient, an economic problem stemming from the inadequate diffusion of medical information (Phelps, 2000).

Modern SAV research stems from the pioneering work of John Wennberg and colleagues, who studied New England hospital markets. Later studies corroborated their evidence of wide variations across small service markets. Why did women in one New England town undergo hysterectomies at more than twice the rate of another apparently similar New England town? Understandably the variations proved worrisome, and many researchers focused on discovering their sources.

The favored measures of small area variations have been the coefficient of variation (CV) and the systematic component of variation (SCV). The coefficient of variation  $CV = \frac{sd}{mean}$  relates the standard deviation  $sd$  of observed medical use rate to the mean of the same measure. Dividing  $sd$  by the mean adjusts for the size of the rate being studied. For example, the unadjusted rate for treatment of Guillain-Barré syndrome (a relatively rare disorder, afflicting only about one person in 100,000) will be much smaller than the rate for treatment of the common flu. Division by the mean rate of each serves to put these on a common basis.

The SCV is a twist on this theme. Here, the researcher first removes from the observed treatment rate that portion of its variation that we can attribute to the random nature of the disease. The remaining portion then becomes a better measure of the effects of systematic factors, such as physician practice styles, and supply and demand factors.

Table 15-1 excerpts data from a study by Wennberg (1990) featuring the CV statistic. Phelps (1997) suggests that a CV in range 0.00 to 0.10 indicates low variation, 0.10 to 0.20 is moderate, while over 0.20 is high. Of the 10 ailments reported, most of these procedures would be considered highly variable based on the size of their CV statistic. The *extremal* ratio, also reported in the table, while not as reliable a measure of variation, is easier to visualize. It is the ratio of the largest rate observed across the small areas to the smallest. Thus, the rate of mastoidectomy across the study areas varies by a factor of over four.

## Contributions to These Variations

Researchers have investigated these variations. Much of the SAV work focuses on the contribution of socioeconomic characteristics of the population and the role of the availability of supplies of hospital and physician services (see Alexander et al., 1999). The studies together reached two

**TABLE 15-1 Variations by Medical Procedure Category**

Surgical Procedure	No. of Cases	Coefficient of Variation	Extremal Ratio
Colectomy	3,190	.116	1.47
Open heart surgery	1,439	.232	2.29
Appendectomy	5,381	.305	2.86
Thyroidectomy	949	.342	3.35
Total hip replacement	1,717	.353	2.99
Diaphragmatic hernia	2,178	.369	3.45
Coronary bypass surgery	3,744	.383	3.62
Mastoidectomy	569	.461	4.03
Spinal fusion w/wo disc excision	1,234	.520	5.20
Total knee replacement	998	.525	7.42

Source: Based on information from Wennberg (1990).

conclusions: (1) Supply variables are important and demand characteristics play a somewhat lesser role though both are statistically and materially significant; and (2) such variables do not seem to suffice, as much variation is unexplained (Folland and Stano, 1990).

### The Physician Practice Style Hypothesis

Wennberg (1984) argued that much of the observed variation relates closely to the degree of physician uncertainty with respect to diagnosis and treatment. When there is little consensus about the effect and value of a medical procedure, a wide range of physician treatment choices lies within the bounds of accepted practice. The physician's habits of treatment choices, beliefs about efficacies, and patterns of practice are seen as "style."

Practice style probably varies among physicians due to an incomplete diffusion of information on medical technologies. The true production function may not be widely known. Some physicians may be optimistic about the good effect of a given procedure, yet others pessimistic.

Furthermore, as medical science uncovers better information about the true production function, efforts to disseminate this information will help reduce unnecessary surgery. Reducing variations related to misconceptions about the effectiveness of given treatments is a chief cost reducing method of the Patient Protection and Affordable Care Act (PPACA) in the United States. Reliable quantitative estimates are hard to come by, but some experts claim that the portion of ineffective treatments is high.

**FORMULATION OF PRACTICE STYLE** We assume throughout that physicians have a practice style created and altered by the irregular diffusion of information. Epstein and Nicholson (2009) model the sources of relevant information. These authors find that variations between physicians within a small area are much greater than variations between areas. Contrary to prior hypotheses they find that a physician's residency has relatively little influence on his practice style. Stronger influences are his peers with the hospital where he practices as well as his peers in the other hospitals in his region. Their study's indicator of OB/GYN practice style was the portion of deliveries conducted by caesarian section. By controlling for patient and physician characteristics, they found that the practice of physician peers were important to the choice of caesarian birth.

**EDUCATION, FEEDBACK, AND SURVEILLANCE** Studies show that information programs directed at physicians can alter their behaviors and thus presumably their practice styles. One study (Wennberg & Fowler, 1977) found that an informational program significantly affected

the tonsillectomy rates in 13 New England areas. Another (Dyck et al., 1977) found that the rate of “unjustified” hysterectomies dropped by two-thirds subsequent to a review program introduced in the Canadian health system. Yet another (Chassin and McCue, 1986) found reductions in unneeded use of pelvimetry following a physician information program.

**COMPARING UTILIZATION RATES IN HOMOGENEOUS AREAS** A second empirical approach was developed in an influential study by Wennberg and Fowler (1977). The authors studied use rates and population characteristics for six areas in Vermont. After finding that morbidity and many socioeconomic variables were not sufficient in explaining the variations in this region, they concluded that variations in use rates probably are due largely to practice style differences across the small areas.

Though practice style in principle can determine the observed variations across small areas, it is invalid simply to infer that a given set of observed inter-area differences is necessarily due to practice style. For example, Green and Becker (1994) studied small areas in Michigan where acute cardiac ischemia (ACI) discharge rates ranged from high to low by a factor of three. The large variations proved to be mainly due to differences in the underlying rate at which patients chose to use the hospital emergency room rather than to differences in physician decision making. They concluded that area variation does not necessarily mean variation in practice style.

### Multiple Regression Approaches

It may be best to approach the problem with multivariate techniques. As more variables are added to the analysis, one learns better how much of the observed variation can be explained by standard socioeconomic variables without resorting to the practice style hypothesis. The residual unexplained variation from the regression then represents an upper bound on the contribution of practice style (Folland and Stano, 1989).

Multiple regression studies using standard socioeconomic variables explain substantial amounts of the inter-area variations, especially when they focus on aggregated utilization rates (commonly represented by a weighted average of many procedures) in larger geographic areas. Multiple regression analysis of individual procedures, as opposed to aggregates of procedures, is generally less successful in explaining the variations. Phelps and Parente (1990) found that standard demand and supply variables typically account for between 40 and 75 percent of the variation in their study of 134 separate diagnostic categories. The authors pointed out that nevertheless these results leave a substantial amount of unexplained variation that in principle could be due to practice style.

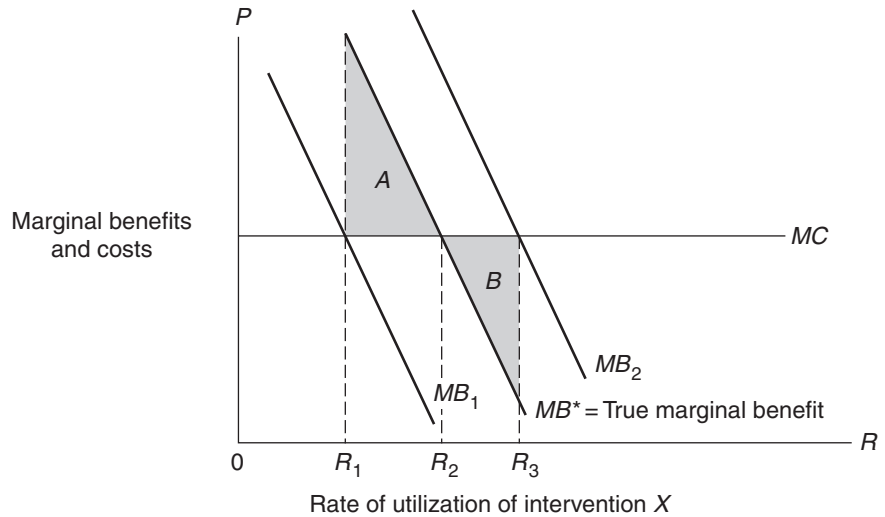
Elsewhere, Escarce (1993) explained about 43 percent of the variation in cataract surgery rates for the Medicare population with socioeconomic variables. This still leaves a potential 57 percent that may be due to practice style. However, it is clear that the influence of practice style is well below Escarce's upper bound. He could identify several nonpractice style influences. After purging the surgery rates of any possible practice style influence, Escarce found that the variation in cataract surgery rates reduced by only a small amount.

These contributions indicate that practice style could be important for many procedures, but that we cannot overlook the contribution of the many standard variables. Surely the inadequately informed physician is out there, but the question of to what degree has proved more difficult to answer.

### SAV and the Social Cost of Inappropriate Utilization

The most important issue in the SAV literature is the proposition that substantial variation in utilization rates is an indication of inappropriate care. The potential savings from improving the quality of care and reducing excessive utilization has led to a national focus on research into patient outcomes. Indicative of this priority, Congress in 1989 created the Agency for Health Care Policy and Research “to determine what works and to develop practice guidelines and standards.”





**FIGURE 15-5** The Inefficiency of Misinformation About the Marginal Benefit of Health Care

The economic cost of physician misinformation can be theoretically identified in Figure 15-5. Suppose that the marginal benefit curve  $MB_2$  represents the true patient benefit given full information; then the efficient quantity of services is  $R_2$ . If physician lack of knowledge led them to believe that  $MB_1$  represented the true benefits, then the area,  $A$ , would describe the value of the patients' foregone opportunity, their welfare loss. This area represents the inefficiency or welfare loss to society from underutilization; the patient's true willingness to pay (represented by  $MB_2$ ) exceeds marginal cost for each unit of care from  $R_1$  to  $R_2$ . In a similar fashion, one can measure the welfare loss of inefficiently excessive utilization at rate  $R_3$  by triangle  $B$ .

Despite the simple and direct application of consumers' surplus concepts to the welfare losses of practice variations, investigators have found it difficult to generate reliable estimates of these costs. Phelps and Parente (1990) and Phelps and Mooney (1992) approached the problem by using regression analysis to estimate true marginal benefit curves. Phelps and Parente find that the welfare loss due to variations from "true" practice in the nation total \$33 billion. This may understate the actual loss were these intermarket variations accompanied by substantial within-market variations. In work by Epstein and Nicholson (2009), the within-market variations proved to be substantial. However, the approach has drawn criticism in several respects. Regression always predicts the dependent variable imperfectly, leaving a residual. Distinguishing between how much of the residual is due to inappropriate use and how much is due to legitimate omitted factors, such as patient preferences, can be difficult.

Furthermore, Dranove (1995) has pointed out that patients also make errors when predicting the true value that medical care will actually deliver. This additional "noise" also disturbs our estimates. Most economists studying the issue probably would agree that while inappropriate use occurs commonly and is often costly, we do not yet have confident estimates of the total costs of inappropriate care.

Despite the empirical challenges, many researchers in the medical community now recognize the importance of using an economic concept of benefits, as opposed to benefit measures determined strictly by medical science, and they urge physicians to involve patients in treatment decisions (see Box 15-3). Differences between economists and medical researchers over SAV, practice style, and appropriate care thus appear to be narrowing. In particular, the consensus is growing in the health services literature that cost-effectiveness analysis and other forms of program evaluation (Chapter 4) are the best means to identify and rank treatments through their effects on social welfare.

## Other SAV Applications

Another common usage of SAV is to study inequalities in environment and in health care availabilities. Wheeler (2004) provides an example of the SAV framework used to study inequalities in environmental hazards and finds that environmental hazards affect disproportionately those who are materially less well off. Folland's 2005 study of East European economies in transition provides a SAV example focused on inequalities of health care across small areas of command systems. Though a comparison of health care availability across small areas in the command systems versus American areas favored the U.S., the (former) command economies experienced smaller variations, and thus a lesser inequity, in mortality rates.

## OTHER PHYSICIAN ISSUES AND POLICY PUZZLES

### Malpractice

On occasion providers make mistakes. The website StateLawyers.com provides a useful description and definition:

**Medical Malpractice** occurs when a negligent act or omission by a doctor or other medical professional results in damage or harm to a patient.

Negligence by a medical professional can include an error in diagnosis, treatment, or illness management. If such negligence results in injury to a patient, a legal case for medical malpractice can arise against:

- The doctor, if his or her actions deviated from generally accepted standards of practice;
- The hospital for improper care or inadequate training, such as problems with medications or sanitation;
- Local, state or federal agencies that operate hospital facilities.

*Source:* [http://www.statelawyers.com/Practice/Practice\\_Detail.cfm/PracticeTypeID:63](http://www.statelawyers.com/Practice/Practice_Detail.cfm/PracticeTypeID:63), accessed August 5, 2011.

Although legal cases are fraught with nuances, proof of malpractice requires evidence of harm to the patient and negligence by the physician. Most mistakes in patient care go unnoticed by the patient, but a small number of patients sue for malpractice, and some of these suits will be justified.

The problem for physicians is that many judgments are very large and malpractice insurance premiums can be very high for the most suit-prone specialties. For physicians in total, the premiums are less than 10 percent of physician revenues. Health economists take interest in malpractice insurance costs when they threaten to change physician geographic movements in a manner that might reduce social welfare.

Polsky and colleagues (2010) studied hospital discharge data from Pennsylvania, Florida, and New York. They found that higher malpractice premiums tended to increase the rate of exit and reduce the rate of entry of obstetricians. Premium increases averaging 20 percent per year led to a decline in the obstetrician supply of 5.3 percent.

Helland and Showalter (2009) employing national data on physician liability (a measure closely related to malpractice premiums), found that a one percent increase in liability leads to a -0.285 percent decline in work hours for physicians. Among physicians 55 or older this elasticity was -1.224.

Robert and Hoch (2007) asked whether increased malpractice premiums led to increased medical expenditure for the consumer. Using Medicare data, they found that greater rates of malpractice law suits increased medical expenditures per enrollee. They presumed that "defensive medicine" (where physicians overtreat to ward off lawsuits) was the cause. In some jurisdictions the extra cost exceeded 25 percent.

### Direct-to-Consumer Advertising of Drugs

Those old enough to remember when direct-to-consumer advertising of drugs started, are often stunned by the frequency of pharmaceutical ads on today's TV. Direct-to-consumer drug advertising became legal in 1985, and to date (2011) occurs only in the United States and New Zealand. Economists ask what effects this has on physician practice and on community welfare. David Bradford and colleagues (2010) studied physician data from 27 states, reaching the encouraging conclusion that direct advertising reduced the time from when patients perceived the symptoms to when they are treated by the physician: a benefit to patient welfare. However, Hai Fang et al. (2008) instead complained that "consumerist" patients, those "educated" by both advertising and the Internet sources, often take an excessive amount of physician time, creating an external cost to other patients. Compared to a consumerist-free world, these authors found high levels of consumerism associated with lower levels of perceived quality.

### Paying for Outcomes

When a customer gets a car repaired, the mechanic usually guarantees the work, and if the outcome is not satisfactory the customer can go back and have the mechanic make things right. Why can't we do the same with physicians?

Dranove and White (1987) have argued that the common physician form of contract stems from both the difficulty of evaluating the health status of the returning patient as well as the fact that unobservable patient behavior is very important to the outcome. How can one tell if the patient's claim of pain is true, and how can one be sure that the patient has followed the treatment regime faithfully including taking his medicine as prescribed?

Leonard and Zivin (2005) argue that outcome-based contracts will most likely succeed when both patients and practitioners play important complementary roles. This means that they must work together (the provider in diagnosis and the patient in compliance with treatment regimens) in managing or treating diseases, such as diabetes, asthma, or HIV infections. Effort-contingent contracts are likely to be successful when either the patient's or the provider's effort is necessary, but not both; surgery is a good example.

#### BOX 15-3

#### Clinical Decision Making and Patient Preferences

David Eddy is a pioneer in studying the role of information, medical science, and preferences in clinical decision-making. Eddy (1990) breaks the treatment decision into two main components—evaluating the outcome of the alternatives and then comparing and choosing from among the options. The first component is largely a scientific one determined primarily through clinical studies of patient outcomes. However, for many patient conditions, even a fully informed physician will face the problem of inadequate scientific information.

The second component, comparing the alternatives, involves judgment and patient preferences. Eddy emphasizes that "It is the patient's preferences that should determine the decision" (p. 442). A failure to make appropriate decisions can arise from failures in both stages. Clearly, physicians' inadequacies in education and training, as well as deliberate attempts to misrepresent outcomes (e.g., to engage in SID), will distort information needed at the second step. Similarly, if patients misrepresent their preferences or physicians misinterpret their preferences, or if inadequate information is communicated to patients, the second step will be flawed.

The medical community is increasingly being alerted to the role of the patient through the development of information technology to improve the decision-making process. For example, Barry et al. (1995) developed an interactive videodisc that helps patients with prostate cancer choose from among several treatment options. To help with the decision, patients answer questions designed to measure their attitude toward risk and their ability to tolerate discomfort.

The authors take advantage of a unique opportunity to observe patient choices between fee-for-service and fee-for-outcome of care in Cameroon, investigating patients who chose between the two types of payments. Payment by outcome is rare among physicians in the developed world, but it is common among traditional healers in many African countries. In rural Cameroon, patients may choose mission-based physicians, compensated by physician effort, or traditional healers, compensated by patient outcomes, with both types of providers covering a variety of illnesses. Leonard and Zivin find that when illness requires large amounts of effort by *both* patients and providers, the patients are more likely to seek treatment from traditional healers who are paid based on outcomes. When the disease is not particularly responsive to one of the two types (physician or patient) of effort, patients visit effort-compensated physicians.

## CONCLUSIONS

Our benchmark model depicts the physician as someone who positively values net income and leisure, and dislikes inducing patient demand. It shows that a physician may respond to increased competition by greater effort at inducement. The model also shows that to support the supplier-induced demand (SID) hypothesis, the physician's income effect must be positive and substantial.

Small area variations can be understood to result from uneven diffusion of medical information to these same physicians. The evidence indicates that substantial variation in physician practice style exists even after accounting for the standard supply and demand variables. Physician practice also responds to new information. Removing all practice style variation that it is possible to remove may create improvements to social welfare, although it is not clear how large they will be.

Some independent issues close the chapter. Malpractice litigation has effects on entry and exit patterns of physicians. Direct-to-consumer advertising of drugs can be beneficial in some cases. Though payment by outcome is rare in physician practice, we observe it among traditional healers in Africa, where it is beneficial for patient cases that require the complementary effort of both physician and patient.

---

## Summary

1. A benchmark model describing the physician as seeking to maximize utility over income, leisure, and inducement, captures most observed physician behavior.
2. In the benchmark model, the physician faces trade-offs with income and leisure but also with income and the disutility of inducement.
3. Supplier-induced demand (SID) includes the possibility of physicians' deviating from their agency responsibilities to provide care to their self-interest rather than their patients' interests.
4. Inducement behavior can be understood as the physicians' having strong income effects in response to reduced practice profitability.
5. The target income hypothesis suggests that physicians use their discretionary advantage to achieve a target level of income or an improvement in their relative income.
6. One profit maximization approach develops the analogy between inducement and marketing.
7. Small area variations (SAV) refer to the frequently wide inter-area and intra-area variations in the per capita use rates found for many medical and surgical procedures. Researchers have found such variations in the United States and many other countries.
8. The practice style hypothesis, referring to physician uncertainty in diagnosis and treatment, has been proposed as the major explanatory factor for small area variation.
9. We can model practice style as an information problem based on the physician's incomplete

knowledge of the productivities of his or her treatment alternatives.

10. Most of the evidence for the practice style hypothesis is indirect, coming from studies showing the following:
  - Changes in practice patterns following physician education and monitoring
  - Wide variations in utilization rates across small, homogeneous areas
11. SAV suggests that there is considerable unnecessary or inappropriate care being provided. There is no way to infer, however, whether a particular area provides too much care or just the right amount.
  - High, unexplained residuals in multiple regression analyses of inter-area utilization rates.

---

## Discussion Questions

1. Suppose that insurers monitored all health care payments to determine whether the services were appropriate. Would you expect to see more or less tendency toward SID?
2. Figure 15-2 shows how increased competition can lead to a higher degree of inducement at point  $E_2$ . This suggests that providers try to induce more usage to compensate for lower profit margins. Suppose, however, that the physician ends up at a different point, call it  $E'_2$ , where there is less inducement than at  $E_1$ . Could convex indifference curves be drawn so that the change to the  $m'$  rate of profit would lead to this reduced inducement?
3. If physician fees are fixed so that they do not adjust to changes in supply, explain how a firm whose demand curve slopes downward would react if the fixed fee were lowered, perhaps by the government.
4. What are some criticisms of the target income approach to SID? In the target income model, what determines the physician's target income? Would target incomes differ between physicians in practice? Would some physicians choose instead to be income maximizers?
5. In the utility maximizing model, what forces limit a provider's ability and willingness to engage in SID? In the profit-maximizing model of SID, what are the costs to the physician of inducement? Give examples of inducement costs.
6. Assuming that SID is prevalent and substantial, what are the implications for policy? A policy to reduce Medicare payment rates? A policy to increase the number of medical school graduates?
7. What is SAV? What are some economic forces that can help explain SAV? What are some demographic and other considerations? How might physician uncertainty lead to SAV? What evidence supports the practice style hypothesis?
8. Do high utilization rates necessarily indicate the provision of unnecessary care? If not, why not?
9. If the cause of SAV is lack of information about the productivities of various procedures, would you expect SAV to have increased, decreased, or stayed the same over time? Why?
10. As new technologies become available more rapidly for given procedures, would you expect an increased or decreased amount of SAV?
11. Suppose large variations occur in use rates within a typical small area, as well as among small areas. What would this mean for the practice style hypothesis?

---

## Exercises

1. Explain why the indifference curves in Figure 15-2 are positively sloped.
2. In Figure 15-2, suppose that the initial profit is \$1 per unit of inducement ( $m = 1$ ). Suppose then, that increased competition lowers  $m$  from 1 to 0.5.
  - (a) Draw the new profit line  $\pi = mQ_o + ml$ .
  - (b) Demonstrate the case where the change in profit increases the level of inducement.
  - (c) Demonstrate the case where the change in profit decreases the level of inducement.
3. In comparing SAV among diseases and diagnoses, would more complicated diseases suggest greater or lesser variation?
4. Consider the approximation of the welfare loss due to inter-area deviations from the correct rate of care. All else equal, which procedures would yield the largest welfare losses—those with low price elasticities in absolute value or those with high price elasticities in absolute value? Why is this the case?
5. In Figure 15-5, suppose for each patient that marginal cost = \$100, that  $R_1 = 1$ ,  $R_2 = 2$ , and  $R_3 = 3$ . Suppose also that one-third of the physicians use treatment levels  $R_1$ ,  $R_2$ , and  $R_3$ , each. What would be the “cost per patient” of SAV, if all of the demand curves have slopes of  $-50$ ? (Hint: Use the parameters above to figure out the sizes of triangles  $A$  and  $B$ ).

# Health Care Labor Markets and Professional Training

- The Demand for and Supply of Health Care Labor
- Factor Productivity and Substitution Among Factors
- Health Manpower Availability and the Meaning of Shortages
- Medical Education Issues and the Question of Control
- Licensure and Monopoly Rents
- Other Physician Labor Issues
- Conclusions

Commensurate with its high share of GDP and the labor-intensive nature of its output, the health care economy employs a large number of workers. We use the term *labor* here in the general economic sense of an input in production that is distinct from capital and that is provided by human beings. Like other goods and services, the production of health services requires both labor and capital.

While many health sector workers are relatively unskilled, the health sector also requires large numbers of highly trained professionals. Most physician specialists obtain years of postgraduate education past their medical school education. This chapter addresses labor issues that range from general supply and demand principles applied to all health care occupations, to specialized topics involving shortages of doctors and registered nurses, medical education and licensure, and various practice decisions of physicians.

## THE DEMAND FOR AND SUPPLY OF HEALTH CARE LABOR

Table 1-3A in Chapter 1 shows the magnitude and variety of health care occupations, and the importance of labor not only to the health sector, but to the overall economy. In 2009, 15.5 million people, representing 11.1 percent of all employed civilians, worked at health services sites. These data do not include many other workers in the pharmaceutical and health insurance industries, as well as those in industries providing supplies, capital goods, and services for people providing direct patient care. Workers in some health-related occupations, such as pharmacists employed in drugstores, also are not included.

These numbers have increased substantially with the growth of the health economy. Between 1970 and 2009, the number of physicians rose from 334,000 to 972,000, an increase of 191 percent; the number of registered

nurses more than tripled from 750,000 to over 2.5 million. Table 1-3B also indicates that the per capita ratios have grown substantially since 1970 for the major health care occupations.

How labor is used—and how it is combined with other factors of production—helps determine both the amount of health care provided and the wages and salaries of the providers. The productivity and training of health care providers are important to the working of labor markets and to the demands and supplies of labor.

We begin by describing the determinants of labor demand. Demand for a factor of production, either labor or capital, is derived from the demand for health. We demand health care providers because we demand health care, and we, in turn, demand health care because we demand health.

### Production Functions and Isoquants

Recall that the production function describes the relationship of factors of production (the inputs) to the resulting goods and services produced (the outputs). Under the existing technology and know-how, it shows the maximum sustainable output obtained from all possible combinations of inputs, such as labor, materials, buildings, and equipment.

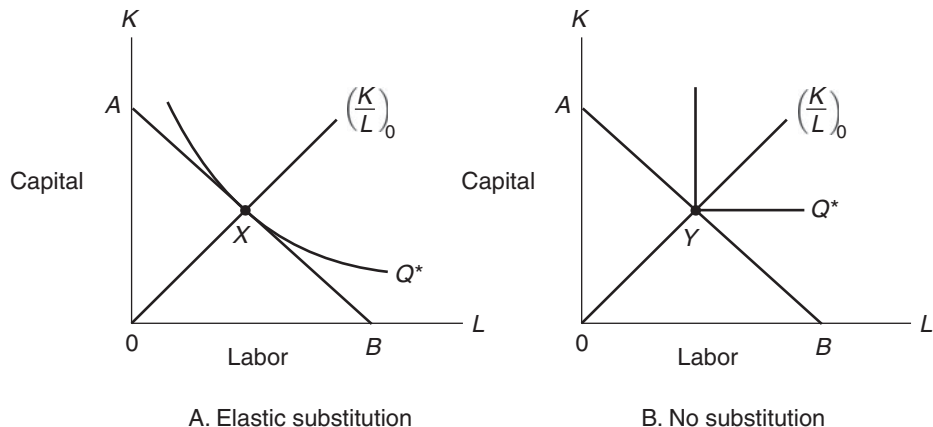
Economists often simplify the production relationship as follows:

$$Q = f(L, K) \tag{16.1}$$

where  $Q$  represents output over a period of time, and  $L$  and  $K$  represent the quantities of labor and capital inputs over the period. We can illustrate many features of a production function graphically through isoquants. Recall also that an isoquant represents all combinations of inputs (e.g., labor and capital) that result in a given level of output. Figure 16-1 (panels A and B) describes isoquants for two different situations.

In panel A, the isoquant shows a technology in which labor and capital are fairly good substitutes for each other so that labor and capital can be combined in many different proportions to produce output. The budget line,  $AB$ , reflects the trade-off between capital and labor, and point  $X$  is the location at which the costs of producing  $Q^*$  units of output are minimized.

With the given set of input prices (wages and payments to capital) the optimum (cost-minimizing) capital-labor ratio  $(K/L)_0$  is the slope of a ray from the origin through point  $X$ . Due to the curvature of the isoquant, a change in relative input prices changes the optimum capital-labor ratio. In panel A, labor and capital are good substitutes as we move along a given isoquant. Lower



**FIGURE 16-1** Production with Varying Rates of Substitution

prices of labor (i.e., a flatter budget line) will lead to a relatively large substitution of labor for capital, and vice versa.

In contrast, panel B shows a technology in which labor and capital are not good substitutes; as drawn, they must be used in fixed proportions to one another. Although point *Y* represents the same ratio ( $K/L$ ) as point *X*, changes in the factor prices will not change the capital-labor ratio. Specialized surgeries, for example, may require specific ratios of labor to capital with little substitution available. The degree to which substitution among inputs is possible, either between health care labor and capital, or among different types of health care labor, is a key issue in manpower planning and in determining the efficiency of production exhibited by health care firms.

The demand for any type of health care labor depends in part on these substitution possibilities. The demand for a factor of production also depends on the price of the output. Consider an example. Suppose you were working as a skilled worker in a bicycle manufacturing plant, and suppose that bicycle riding was becoming more popular. The increase in demand would result, at least temporarily, in a higher price for bicycles. More bicycle workers would be demanded as a consequence.

### Marginal Productivity of Labor

Consider that the demand for an input, and consequently the wage paid to the input, will depend at least in part on the input's productivity. This is one explanation why college-educated workers earn more money than others. If college-educated workers are more productive than others, then the demand for them will be greater. Although this is only one theory that explains the earnings advantage of college graduates, it illustrates the premise that more productive laborers are in greater demand.

Reconsider the idea of marginal product—the increase in output when an input is increased by one unit. Suppose that labor in equation (16.1) represents the number of laboratory technicians employed. If we add one lab technician to the production process, holding the available lab equipment and materials (i.e.,  $K$ ) constant, we expect to increase lab output. This increase is the marginal product of lab technicians.

Although additional lab technicians in the production process will likely increase total output, they likely will add incrementally less output. As we increase an input, holding all others constant, output will tend to increase but at a decreasing rate. This illustrates the law of diminishing returns.

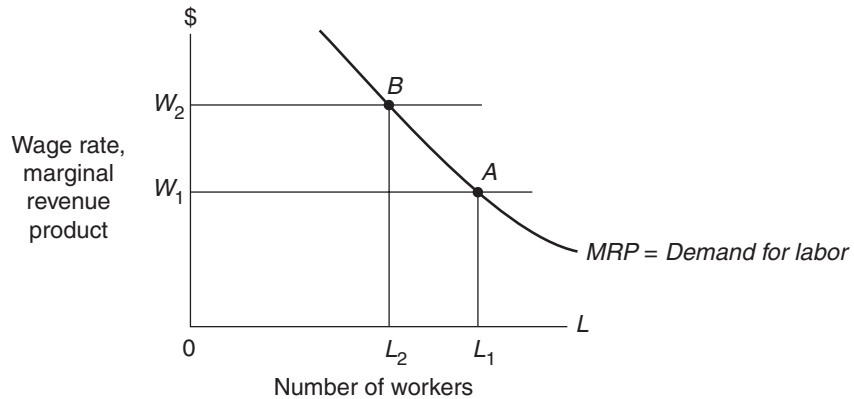
The number of lab technicians to hire depends also on the price of output. Suppose that lab tests sell for \$100 per unit. Suppose also that an additional technician would increase output by four tests per day. If these tests sell for \$100 each, the technician is bringing in an additional \$400 in revenue. The extra revenue generated is called the marginal revenue product (*MRP*).

Would it pay to hire this extra technician? Clearly the answer depends on the wage per day. If technicians earn \$150 per day, the technician nets the hospital a \$250 gain (\$400 marginal revenue product less \$150 wage), so it pays to hire another one. It always pays to hire laborers whose marginal revenue products exceed their wage.

Would it pay to hire still another technician with a marginal product of three tests per day? Because output (the tests) sells for \$100 each, this next technician is netting the hospital \$150 because the marginal revenue product, \$300 in this case, exceeds the wage, \$150. It will pay the firm to continue to hire more workers up until the point where the marginal revenue product equals the wage.

Figure 16-2 represents the marginal revenue product curve for lab technicians. The curve is downward sloping, reflecting the law of diminishing returns. In competitive markets, we find the *MRP* curve in Figure 16-2 by multiplying the marginal product curve (not shown) by the price of output. The optimal number of lab technicians depends on the wage rate. At wage  $W_1$ , the optimal input demand at point *A* is  $L_1$ . At a higher wage,  $W_2$ , the firm would demand fewer technicians,  $L_2$ , at point *B*.





**FIGURE 16-2** Firm Hiring Decisions at Different Wage Levels

This analysis shows that the demand for labor is precisely the marginal revenue product of the labor curve. Through the marginal product, this curve is closely related to the production function for the laboratory test. It is also directly related to the price of laboratory tests. If their price increases, the marginal revenue product increases, and more labor is demanded. If better machines make the lab technicians more productive, the demand for their services will also increase.

### Factor Substitution and Labor Demand

At this point, recall the meaning and importance of the substitutability of one input for another. Suppose, for example, a newly invented machine allows lab technicians to perform certain functions previously performed by radiologists. This makes technicians better substitutes for radiologists. As a result, the demand for technicians will tend to increase. This change will also tend to shift the demand for radiologists and probably make it more elastic—that is, flatter and more responsive to their wage rate. As a result, if the firm finds it can substitute more easily between inputs, it will become more resistant to input price changes, replacing increasingly expensive inputs with cheaper substitutes.

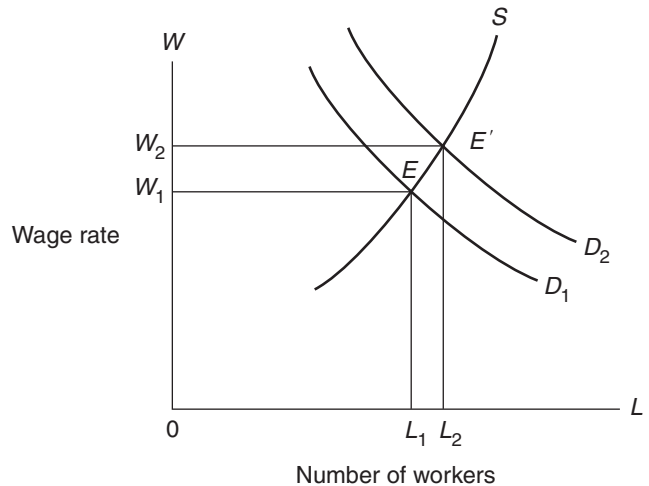
In recent decades, firms and policy makers have looked closely at substitution possibilities in their efforts to control health care costs. To the extent that nonphysician providers can substitute for physicians, firms can hire lower-cost workers, hence cutting costs and possibly increasing the amount of the service provided. We address this issue in more detail later in the chapter.

We find the market demand for various occupations by horizontally adding the demands of the individual firms. The the market demand for laboratory technicians in Figure 16-3 is the downward-sloping curve labeled  $D_1$ . Under competitive conditions, the labor market equilibrium, and consequently the equilibrium wage, depend on the interaction of both demand and supply.

### The Supply of Labor

The supply of labor tends to slope upward, implying that a higher wage rate in a given market will attract more workers or labor time. First, those workers currently employed may choose to work more hours if higher wages are offered. Second, similar workers may be attracted into the market from elsewhere. We show the labor supply curve for lab technicians in Figure 16-3 as  $S$ .

The equilibrium market wage,  $W_1$ , for technicians, along with the market-clearing number of technicians,  $L_1$ , occurs at the intersection of the supply and demand curves. Consider that the market equilibrium wage will tend to increase in response to any demand-increasing event. These events include increases in the firms' desires to substitute lab technicians for other



**FIGURE 16-3** Market Demand and Supply for Laboratory Technicians

laborers, increased productivity of the technicians, and increases in the price of the lab outputs. Here, demand increases to  $D_2$ , increasing the equilibrium wage to  $W_2$  and employment to  $L_2$ , at point  $E'$ .

Similarly, the market wage will tend to fall in response to events that increase the labor supply. These include increased graduations of trained technicians or influxes of technicians into the market from other professions or other locations.

Changes in supply can take place relatively quickly for those health care occupations requiring minimal education or training. For physician specialists and others requiring the highest levels of education and training, the number of new professionals is determined by the admission decisions of medical schools as well as the decisions of applicants, both made many years earlier. Physicians must be licensed by a state in order to practice in that state. Requirements for licensure include graduation from an accredited medical school, passing a licensure examination, and completing one to two years of internship or residency in an accredited graduate medical education program. Many graduates, nevertheless, complete three- to four-year residency programs. Many physicians also become board-certified specialists. The requirements typically include advanced residency training for three to six years, practice in the specialty, and passing the board examination.

Through education and training, medical students make investments in their human capital. Economists treat the decision to invest in human capital with the same tools used to analyze investments made by businesses in physical capital. The decision maker will consider the revenues associated with investment along with all costs, including any opportunity costs. For medical students, the forgone earnings associated with the time it takes to complete their medical education is an important opportunity cost, but monetary values must also be imputed to nonpecuniary gains, such as the satisfaction from helping the ill and the prestige associated with being a physician. The internal rate of return is the rate that equates the present value of revenues with the present value of costs (see the Appendix to Chapter 4 for a discussion of discounting). Prospective medical students, in principle, will compare the return from medical education with those of other possible occupational choices and select the highest one.

Information regarding the rate of return to a medical education also is useful to the policy analyst. Suppose we discovered that the average rate of return to medical education is high and rising. This information might mean that physicians are becoming increasingly scarce, suggesting a shortage. Numerous studies have attempted to estimate the rate of return to a medical education, and we will describe some of them later in this chapter.

## FACTOR PRODUCTIVITY AND SUBSTITUTION AMONG FACTORS

The supply of health services, and consequently health care prices, depends on the number of workers. However, productivity of labor inputs represents a critical determinant of supply. Productivity improvements increase output or alternatively, the same output can be produced with fewer inputs. In a macroeconomic sense, general improvements in productivity represent a major source of economic growth and rising standards of living. At a microeconomic level, productivity gains in an industry can lead to lower prices for the goods and services produced in that industry and possibly to higher rates of remuneration for workers.

The productivity of a factor of production can be measured as the average product of the factor—that is, the ratio of total output ( $Q$ ) to the amount of a particular labor input ( $L$ ):

$$\text{Average product of labor} = Q/L \quad (16.2)$$

This definition corresponds to the concept of average product used in microeconomics and is distinguished from marginal product, which we define as the change in output associated with a one-unit increase in the input, holding all others constant.

Despite the simplicity of the concept of average product, difficult problems occur in measuring it. The product may be heterogeneous, consisting of many different outputs. In such cases, one often uses the dollar value of output for the numerator,  $Q$ . Similarly, many kinds of labor are used in the production process. In such cases, a weighted sum of related inputs is often used in the denominator term,  $L$ .

### Measurement of Physician Productivity

Reinhardt (1972) undertook a classic study of physician productivity. He examined general practitioners in private practice for three measures of output: total patient visits, office visits, and patient billings. In addition to physician time, he considered the use of various auxiliary personnel. Reinhardt estimated the marginal product of physician time—the increment to output resulting from the addition of one hour of physician time to the production process.

He found that the marginal product tended to increase up to where the physician was working a total of about 25 hours per week; the marginal product eventually declined to zero at about 110 hours per week. He found that starting from a base of 60 hours per week, a 1.0 percent increase of physician input would result in an increase of 0.8 percent in the number of patient visits produced.

In addition to studying the physician's productivity, Reinhardt examined the substitution possibilities between physician and other labor inputs. Consider physician aides. The aides' marginal products were highest when about one aide was present per physician. Physicians could improve productivity of their practices and increase profits if they doubled the number of aides from two aides per physician (the average he found) to four aides per physician.

### The Efficient Utilization of Physician Assistants: Substitution Among Inputs

The possibility that physicians were underutilizing aides was a provocative one. In 1988, Brown refined Reinhardt's work. Brown's marginal product estimates of physician time and other inputs, calculated at mean values of the variables, are in Table 16-1. The marginal products of auxiliary workers are in the columns labeled  $MP$  for data from physician offices of various categories. These include all physicians, solo physicians, and group practice physicians.

By dividing  $MP$  by the wage rate ( $W$ ) of each input to get the marginal product per dollar spent on each factor, we can draw inferences about whether physicians underutilize or overutilize various categories of workers. The marginal product per dollar is the relevant measure when determining which input to increase. To increase profits, one should hire the extra input that has the greatest  $MP/W$ . If this marginal product per dollar is not equal for each category of worker, the firm can save money by replacing lower-producing workers per dollar (low  $MP/W$ ) with higher-producing ones (high  $MP/W$ ).

**TABLE 16-1 Marginal Products and Efficiency of Input Use**

Input	All Physicians		Solo Physicians		Group Physicians	
	MP	MP/W	MP	MP/W	MP	MP/W
Physician	2.967	0.114	2.686	0.102	2.793	0.110
Secretary	0.192	0.043	0.253	0.058	0.105	0.023
Registered nurse	0.585	0.104	0.628	0.109	0.625	0.114
Practical nurse	0.542	0.129	0.533	0.132	0.485	0.109
Technician	0.320	0.067	0.321	0.059	0.278	0.057
Physician assistant	0.231	0.040	-0.014	-0.003	1.082	0.192

Note: MP = Marginal product

MP/W = Marginal product per dollar wages spent on input

Source: Reprinted from the *Journal of Human Resources*. Douglas M. Brown, "Do Physicians Underutilize Aides?" Originally published in the *Journal of Human Resources* 23 (1988): 342-55. © 1998 by the Board of Regents of the University of Wisconsin System. Reproduced by the permission of the University of Wisconsin Press.

Brown concluded that physicians were underutilizing nursing inputs. Consider, for example, the data for practical nurses in all physicians' offices. These practical nurses have a higher marginal product per dollar, 0.129, than do physicians, 0.114, so the offices would become more profitable if one substituted practical nurses for physicians.

In addition, Brown found that physicians in group practices were on average 22 percent more productive than those in solo practices. He suggested that this figure, much higher than the 5 percent estimated by Reinhardt, resulted from advantages that group practices have in employing physician assistants. From Table 16-1, we can see that the marginal product of physician assistants (PAs) for solo practices was estimated to be negative. In contrast, PAs are productive on the margin in group practices. Even so, group practices are underutilizing PAs.

Other research provides considerable evidence on the substitutability between physician time and other labor inputs. Escarce and Pauly (1998) found that each hour of time for an office-based internist substitutes for \$60 in nonphysician costs or vice versa. Elsewhere, the growth of managed care organizations has heightened interest in PAs and nurse practitioners (NPs) to improve productivity and lower costs. Jacobson and colleagues (1998/1999) report that PAs/NPs can perform 50 to 90 percent of the tasks of primary care physicians without compromising quality when they work collaboratively with physicians. They also found that PAs/NPs have greater scope of practice and autonomy as the proportion of managed care patients in a health care organization increases. Despite legal impediments and other limits on the use of PAs/NPs, the delivery of primary care is likely to rely increasingly on these skilled physician substitutes as organizations try to move toward their optimal input mix. Box 16-1 summarizes some of the most recent work on productivity.

## HEALTH MANPOWER AVAILABILITY AND THE MEANING OF SHORTAGES

Table 1-3B presented data on four large and important health manpower groups.<sup>1</sup> For those with available data since 1970, the total number of professionals has increased, substantially. The population of ordinary citizens (the potential consumers) also increased during the period. Table 1-3B

<sup>1</sup> Although regrettably sexist, the term *manpower* has evolved as the standard term for planning health care needs.

**BOX 16-1****Recent Productivity Studies**

The basics of productivity theory outlined above follows the standard textbook model in which a production function shows the maximum sustainable output associated with any given level of inputs. In theory, the inputs are homogeneous and the production function should not be affected by financial or other incentives. It is purely a technological phenomenon. Two recent contributions highlight some of the nuances of dealing with productivity for health care services where both outputs and inputs are difficult to measure.

Kantarevic and colleagues (2011) compared various measures of productivity for primary care physicians in the Canadian province of Ontario under two different payment systems. One was traditional fee-for-service (FFS); the other was an enhanced FFS model, introduced in 2003, that provides financial rewards for improving quality and access through preventive care; managing chronic diseases; and providing after-hours services. The study found that those who joined the enhanced FFS plan had significantly higher productivity (by about 6–10 percent) than comparable FFS physicians in terms of services, visits, and number of patients. However, a substantial portion of these gains was attributed to an increase in the supply of physician labor rather than increases on output from any given level of work effort.

Elsewhere, Doyle, Ewer, and Wagner (2010) take advantage of a natural experiment in a large U.S. Veterans Administration (VA) hospital where 30,000 patients were randomly assigned to clinical teams from two academic affiliates. The randomization ensures that patients have very similar conditions and characteristics. One of the affiliates is considered a top medical school while the other has a lower rating. The teams had access to the same VA facilities, nursing staff, and other support staff.

Although patient health outcomes were essentially the same, there were large differences in treatment costs. Teams from the highly ranked affiliate incurred considerable lower costs—10 percent overall and up to 25 percent for more complex conditions. The differences were attributable to higher use of diagnostic testing by teams from the *lower* rated affiliate. The study indicates that high quality physician teams can be much more productive, i.e., attain the same output with fewer total inputs, than lower quality teams. Physician quality matters a great deal and this realization has significant policy implications relating to standards at medical schools and hospitals.

made clear that the number of professionals has been increasing more rapidly so the number of professionals per 100,000 population also has increased. As noted previously, the rate of increase for physicians and registered nurses per capita has been sharp. A casual examination of Table 16-2 indicates that the sum of active physicians more than doubled between 1975 and 2008, from 340,000 to 784,000.

Despite these increases, a recurrent policy concern has been the availability of various critical categories of health manpower. The issue usually hangs on whether we have or will have enough of these professionals; that is, whether there will be shortages or surpluses.

**Availability of Physicians**

Manpower analysis often focuses on the availability of personnel to provide needed or demanded patient care. Many federal physicians (approximately 3 percent of the total) provide services only to the portion of the population eligible, such as veterans. Furthermore, many physicians, including researchers and administrators, do not provide care for patients. Because the portion providing office-based patient care to the public has not changed much over recent decades, the rapid increase in physicians per capita also has meant a greater number of patient care physicians per capita.

Of the 784,000 active medical doctors in the United States in 2008, 741,000 provided direct patient care (Table 16-2). Of those, three-fourths were office based and the rest were hospital based,

**TABLE 16-2 Physicians by Type of Practice: 1975–2008**

	Number in Thousands				
	1975	1985	1995	2005	2008
All Physicians	394	553	720	902	954
Professionally Active	340	497	625	762	784
Nonfederal	312	476	604	NA	NA
Patient Care	288	432	564	718	741
Office-Based	213	329	427	563	557
Hospital-Based	75	102	137	155	184
Other Active	24	44	40	44	43
Federal	28	21	20	NA	NA
Inactive/Not classified/Unknown Address	53	56	95	140	170

NA = Not available

Source: U.S. Department of Commerce, *Health United States, 2010* (Table 107) and earlier issues.

including residents. As we noted previously, the pattern of the portion of physicians available for patient care has not been changing rapidly.<sup>2</sup>

Physicians form a large number of specialties rather than a homogeneous group. About 50 percent of office-based physicians practice in primary care, and 20 percent are in general surgery or the surgical specialties, with the remainder having other specialties. Analysts have long expressed concerns about specialization as well as about uneven distributions between rural and urban areas in the United States.

Planners and policy makers often worry about having adequate quantities of health manpower and avoiding serious shortages, especially of physicians. In the mid-twentieth century, a need-based method, based entirely on medical considerations, became the dominant approach to determining health manpower requirements. This approach, illustrated by the classic study of Lee and Jones (1933), calculated the number of physicians required to serve in a given market area from the needed number of procedures that in turn related to the incidence of morbidity (illness) in the population. Under what Fuchs (1974) has criticized as the monoteknic approach, health services planners assumed that a single technique that includes a fixed amount of physician time is required to treat each particular type of illness. Total physician time, or physician equivalents, was determined by aggregating over a broad range of medical conditions. Economic considerations, such as the potential substitution of other inputs for physician time, changes in technology, differences in patients' preferences, and even costs were generally ignored. For example, Box 16-2 investigates the impacts of the organization into large group practices as opposed to more traditional physician-owned practices.

### Economic Definitions of Shortages of Health Professionals

Economic definitions of manpower shortages usually differ from those based solely on medical grounds. Economists apply definitions based on considerations of how characteristics of a given market for professionals deviate from those found in an ideal, highly competitive market. As a result, several approaches for determining shortages occur in the literature.

<sup>2</sup> There are also about 67,000 doctors of osteopathy (DOs). Although we often combine DOs with MDs when considering physician supply and access to physician care, there have been major differences in the historical development and organization of the two groups. Unless otherwise indicated, our discussion will focus specifically on MDs.

**BOX 16-2****Keeping up with the News**

Articles covering health care professionals appear regularly in the popular press. Many relate directly to material covered in our text; while others may be harbingers of important issues that have not yet reached the scholarly literature. Some items from the *New York Times* caught our attention over a recent period.

Journalist Gardiner Harris describes how a “quiet revolution” is changing the practice of medicine in the United States. Since 2005, the proportion of medical practices owned by physicians has dropped from about two-thirds to less than half. (The 184,000 hospital-based physicians shown in Table 16-2 for 2008 represented a 9 percent increase over 2007.) On the one hand, large health systems, by buying and consolidating smaller private practices, can improve coordination of care and lower costs through improved efficiency. On the other hand, the erosion of the traditional physician-patient relationship as well as the increased monopoly power held by large health systems constitute serious concerns. Harris also draws attention to the political ramifications (e.g., will organized medicine become more liberal?) as the majority of physicians become employees rather than remain as independent entrepreneurs.

While the aging of the U.S. population has received a lot of attention, the aging of physicians is another trend that has slipped under the radar. Tarkan reports on the serious adverse consequences of this phenomenon, such as higher disciplinary rates and higher mortality rates for complicated procedures among older practitioners, and the generally lax oversight of physician competency. Physicians must complete only minimum numbers of continuing education credits to maintain their licenses and older physicians have usually been “grandfathered” in those cases where specialty boards have adopted tougher standards. Some hospitals have introduced formal programs to deal with aging physicians but the vast majority have not yet faced up to this difficult challenge, in part, because of the opposition of other staff doctors.

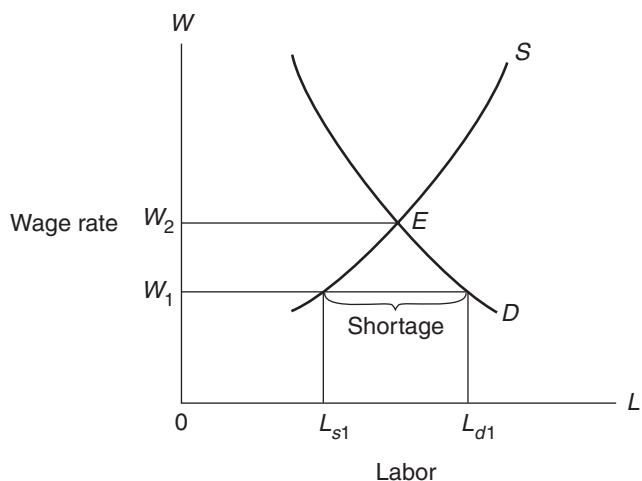
Finally, Chen reports on an interview with Howard Brody—a well-known physician who also holds a Ph.D. in philosophy. Dr. Brody’s (2010) editorial in the *New England Journal of Medicine* sparked a firestorm with his criticism of organized medicine’s failure to exercise moral leadership in health reform. To address costs, Dr. Brody proposed a seemingly simple and appealing remedy. Under his “Top Five” approach, each medical specialty would identify and agree on five widely used and expensive treatments and tests that have limited benefits. The specialties would then develop and endorse a plan to limit their use.

*Sources:* Gardiner Harris, “More Doctors Giving Up Private Practices,” *New York Times*, March 26, 2010 ([nytimes.com/2010/03/26/health/policy/26docs.html?pagewanted=2](http://nytimes.com/2010/03/26/health/policy/26docs.html?pagewanted=2)); Laurie Tarkan, “As Doctors Age, Worries Grow About Their Ability,” *New York Times*, January 24, 2011 ([nytimes.com/2011/01/25/health/25doctors.html](http://nytimes.com/2011/01/25/health/25doctors.html)); Pauline W. Chen, “The Doctor’s Failure to Cut Costs,” *New York Times*, March 3, 2010 ([google.com/search?q=The+doctor’s+failure+to+cut&ie=utf-8&oe=utf-8&aq=t&rls=org.mozilla:en-US:official&client=firefox-a](http://google.com/search?q=The+doctor’s+failure+to+cut&ie=utf-8&oe=utf-8&aq=t&rls=org.mozilla:en-US:official&client=firefox-a)).

**EXCESS DEMAND** Begin with the conventional economic definition of a shortage: the excess of the quantity demanded over the quantity supplied at market prices. Figure 16-4 depicts a shortage defined in this way: The health manpower shortage at the wage  $W_1$  is equal to  $L_{d1} - L_{s1}$ . If the wage instead had been  $W_2$ , no shortage would exist. This is because at the higher wage, less labor services would have been demanded and more would have been supplied.

This definition, though valid, raises critical questions in the case of health manpower. Why didn’t the wage rise to equilibrium, thus automatically eliminating the shortage? The usual case of persistent excess demand is associated with stickiness in wages or prices imposed by law or regulation. A common example in some American and European cities is legally enforced rent control in the housing market.

What would cause wage stickiness in health manpower markets? It seems doubtful that health manpower wages are sticky in the sense of administered rents or prices, or that health manpower shortages in terms of excess demand are a serious policy problem. Some analysts have argued that excess demand shortages are not serious concerns for most categories of professionals.

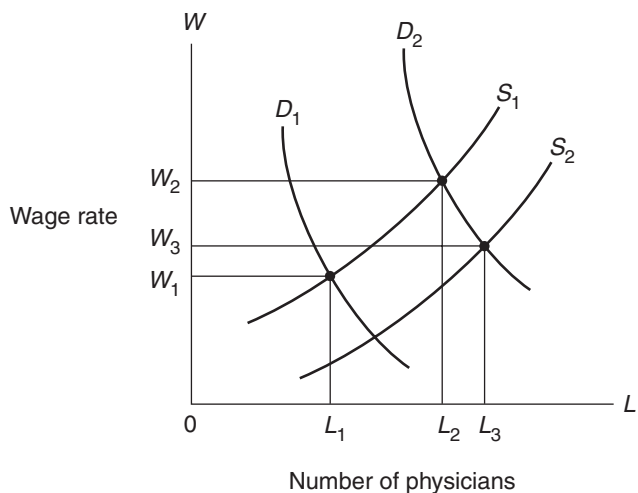


**FIGURE 16-4** An Economic Shortage

**RELATIVELY RAPID INCREASES IN WAGES: DYNAMIC SHORTAGES** An unnecessary focus on excess demand also obscures the fact that economically meaningful shortages of professionals may well exist even when supply and demand are in short-term equilibrium. In particular, a shortage may occur when demand and supply conditions change over time. Suppose, for example, that demand for a category of health professionals expands over time, and that the supply of these professionals is slow to respond or even perhaps faces barriers in responding. The result may be a large rise in wages relative to the wage gains of other professionals.

To illustrate, consider Figure 16-5, which depicts the demand for physicians at two points in time: an initial Period 1 and a subsequent Period 2. We compare the equilibrium quantity in Period 1 with that in Period 2. The wage increase from  $W_1$  to  $W_2$  may indicate a shortage, even though quantity supplied equals quantity demanded (at  $L_1$  and  $L_2$ ) in both periods. We would say a shortage exists if the relative wage of physicians has risen sharply relative to that of other professionals. The measure of shortage under this approach is the relative wage and the direction of its movement.

Several variations on this general approach have been described. The pattern of wages over time may be more complex than the movement described from equilibrium at  $W_1$  to equilibrium at  $W_2$ . We might find, for example, that the initial market response to increased demand for the professionals



**FIGURE 16-5** Changes in Equilibrium over Time Depending on Supply Adjustments



would be to raise wages to  $W_2$ . Only after these high wages had induced the expansion of supply to  $S_2$  would we observe market wages adjusting to  $W_3$  (and quantity supplied to  $L_3$ ). Under this scenario, the professional's wage for a time falls, here from  $W_2$  to  $W_3$ . Thus, a decline in relative wage during a given period may reflect a long-run adjustment offsetting a shortage and not necessarily an indication of excess supply.

**RELATIVE RATES OF RETURN** How should we measure the monetary gains from professional training? Hansen (1964) provides a classic measurement approach that is both plausible and consistent with theory. He proposes that the relevant measure of monetary gains to a given health professional group must take into account the various opportunity costs incurred by professionals in obtaining their training. The internal rate of return is a measure that attempts to accomplish this conceptual task.

The internal rate of return is the discount rate that equates the present value of the stream of costs to the stream of revenues from education. The higher the rate of return, the greater the financial rewards are to investment in the human capital attained through education. To determine whether a given health professional group is in relatively short supply, we can compare the rate of return to that of other professionals and examine these comparative data over time.

High, even excessive, rates of return may occur whenever the supply of labor fails to respond quickly to changes in demand. In some instances, the underlying reason for this slowness in response may be barriers to entry faced by potential health professionals. A barrier to entry exists, in this case, when a potential health professional faces higher entry costs than incumbents faced.

This definition includes cases where potential entrants are completely barred from entry. Such barriers would occur if controls on slots in health professional schools limited entry. They also occur to some degree whenever entry to the profession is limited by licensure laws. The issue of licensure laws is of special interest here because it is common in the health professions. Because of this connection, we treat the empirical literature on rates of return to physician education in the discussion of licensure in a later section of this chapter.

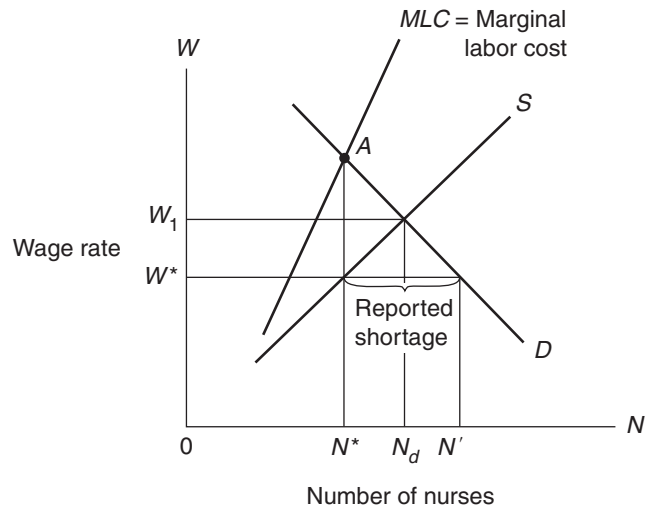
### The Role of Monopsony Power: Shortages of Registered Nurses

Practitioners who describe health manpower availability often rely on reported percentages of unfilled, budgeted positions. One explanation of this measure (using Figure 16-4) is excess demand. Excess demand, however, is generally temporary; as long as prices are not rigid, price increases will tend to cure the problem. More plausibly, the analyst will focus on changes in the percentage of unfilled, budgeted positions, analogous to dynamic definitions of manpower shortages. First, however, we must address the problem of interpreting reported data for markets that are monopsonistic. A monopsony is a market that in theory has only one buyer; for example, a monopsony would be one hospital that hires virtually all registered nurses in the market. In Canada's single-payer health system, individual provinces have monopsonistic market power in paying hospitals and professionals.

**MONOPSONISTIC LABOR MARKETS** Under monopsony, a firm may report unfilled, budgeted positions, for example, for registered nurses, even when the firm is actually in equilibrium. The paradox is that such a monopsony firm may announce that it wishes to hire more nurses even though it is unlikely to take the necessary steps to do so.

The paradox is explained by the monopsony firm's upward-sloping supply curve. Because it is a big employer, it has the power to influence nurses' wages and, thus, to induce more nurses to work by raising the average nurse wage level. The monopsony firm is willing to hire more nurses at the current wage, but it has no intention of paying a higher wage in order to hire more nurses.

These ideas can be illustrated with the help of Figure 16-6. Imagine that one hospital is the only demander of nurse labor in the market. The hospital's demand curve for nurse labor, labeled  $D$ , represents the marginal revenue product curve for nurses employed at that hospital.



**FIGURE 16-6** Nurse Shortage Under Monopsony Conditions

Under monopsony the supply curve for labor will no longer represent the marginal labor cost,  $MLC$ , to this hospital. If this hospital had been a competitive hirer, competition would have meant that the hospital could have hired as many nurses as it wanted at any given wage. In a competitive market,  $N_d$  workers will be hired at wage  $W_1$ .

Consider the monopsonist's marginal labor cost curve,  $MLC$ . When the monopsonist hospital seeks to add one nurse to its labor force, it must pay a higher wage than before in order to induce this marginal (extra) nurse to work. But it then must pay all of its employed nurses a higher wage. The problem arises entirely because the labor supply curve faced by this firm is rising as is typical under monopsony, instead of being flat as is typical under competition.

The extra (marginal) labor cost incurred for that one nurse is not just the wage it pays. It also includes the extra wages it must pay all its other nurses. As a result, the marginal labor cost curve,  $MLC$ , will lie above the labor supply curve,  $S$ . The monopsonist will hire fewer nurses and pay a lower wage than will the competitor. Early evidence indicated that nursing markets were monopsonistic.

Sullivan (1989) found the typical hospital's labor to be upward sloping to a degree that suggested monopsony power in the nursing labor markets. However, Hirsch and Schumacher (1995) note that an upward-sloping labor supply curve for nurses is a necessary but insufficient condition for monopsony effects. In examining RNs' wages relative to wages for females with a college degree across a large number of markets, they find that relative wages were not related positively to the number of hospitals per square mile as predicted by the monopsony model. The same authors (1998) also show that unionization is not a countervailing force to hospital monopsony power. More recently (2005), they found very modest decreases in RN wages in response to increases in hospital concentration, but this effect cannot be sustained over the long run due to the relatively high mobility of nurses.

**REPORTED SHORTAGES** Suppose now that the hospital acts to maximize its profits. It would find it profitable to hire an additional nurse as long as the extra revenue the new nurse brought in, the marginal revenue product, exceeds the extra cost of the nurse to the hospital, the marginal labor cost. Thus, as long as the demand curve (the marginal revenue product) lies above the marginal labor cost curve for a given level of nurse employment, it always will pay to hire more nurses. The hospital achieves its profit-maximizing complement of nurses where  $D = MLC$ , an employment level of  $N^*$  nurses in Figure 16-6. The equilibrium wage is found on the corresponding point of the labor supply curve; here it is  $W^*$ .

At this equilibrium wage,  $W^*$ , the hospital would desire to hire  $N'$  nurses. It may well budget for these nurses and effectively report a shortage of  $(N' - N^*)$ . The hospital acts as if it did not realize that if it wants more nurses to work, it must pay its nurses more. At any rate, a shortage in this case has a limited meaning; it only means additional nurses are desired at the current wage level.

This theoretical point suggests the hazards of interpreting data on unfilled budget positions. The possibility of monopsony power in labor markets suggests that unfilled positions data may overstate the problems of nursing availability. The problem is somewhat mitigated by the fact that these data typically represent unfilled budgeted positions because a hospital is unlikely to budget for all the positions it might desire at the current wage. Furthermore, the analyst might in any case be interested in changes in the percentage of unfilled budgeted positions. If the number of unfilled positions were rising rapidly, it would suggest an increasingly relative scarcity of nurses.

## **MEDICAL EDUCATION ISSUES AND THE QUESTION OF CONTROL**

From their authority in treatment decisions, physicians are the dominant providers in the health economy. Most other health care workers carry out their tasks under the direction of physicians. As a result, economists have concentrated on the training and practice of physicians even though they represent a minority of workers providing patient care. Many professions require a considerable length of time for education and training, but the time period for formal training of physicians is among the longest. In addition, medical education poses the question of who has control. Does the medical profession itself exercise control over access to medical education in order to improve its own profitability? It is helpful to study medical schools and their funding, as well as information about the supply of potential medical students.

### **Sources of Medical School Revenues**

In the United States, as elsewhere, the government heavily subsidizes medical school education. This situation comes from public concern for the adequacy of the supply of physician labor. The rapid growth in medical school enrollments did not arise by accident. Several actions by Congress, beginning with the Health Professions Educational Assistance Act in 1963, provided grants to medical schools and financial assistance to students. Some of this federal support hinged on enrollment increases. In 1971, however, federal support to medical schools increased substantially and came in the form of capitation grants, which were grants that rewarded the medical schools for expanding their enrollments by giving money on a per-student basis, initially about \$3,000 per year per student. With fears of physician surpluses emerging by the late 1970s, assistance under the Health Professions Educational Assistance Act was phased out, and enrollments stabilized.

Nevertheless, medical education is still subsidized heavily. Tuition represents a relatively small source (about 4 percent) of medical school revenues so the student pays only a small portion of the true cost of the investment in education. Governmental support for operating revenues and through grants and contracts is typically about 30 percent of total revenues. The largest share (about 50 percent) comes from reimbursements for health services provided to patients (Jones et al., 1998).

### **Capital Market Imperfections Justify Subsidies**

The policy that students pay for a relatively small share of their training raises questions in light of the fact that physicians earn high incomes and receive relatively high rates of return on their investment. However, at least part of this subsidy reflects a policy that attempts to overcome imperfections in capital markets where potential medical students would have to go to get the loans needed to pay for their education. The economic imperfections stem from the students' difficulties in getting these markets to lend funds, up to the full value of the investment.

Even though there is a perception that a medical education pays off handsomely, it is difficult for a potential lender to secure the loan adequately. The investment in education is an investment in human capital. Human capital cannot be bought and sold in the same way as physical capital.

Most plainly, one cannot repossess a human being. Although the bank knows that the typical medical graduate easily can afford to repay the loan, it is difficult to guarantee this repayment in an individual case, and potential medical students may not get the needed loans even though the investment is worthwhile. From society's point of view, this problem may result in insufficient numbers of medical students.

Leffler and Lindsay (1981) conclude that with such capital market imperfections, reliance on private markets leads to an underinvestment in medical education. Thus, government support can be justified on economic grounds. These authors estimated that an optimal level of support would be about 36 percent of the costs of medical education, which is a substantial figure.

### Teaching Hospitals, Medical Schools, and Joint Production

Medical education is a good example of joint production. That is, medical schools produce at least three products jointly:

- Medical education
- Patient care
- Research

To reimburse for patient care or to fund medical education appropriately, it is necessary to determine the pure costs and the joint costs of these activities. An example taken from Newhouse (1978a) illustrates these terms.

In Table 16-3, the total annual cost for a medical school that produces only education and patient care is shown to be \$60 million. If the school produced only education with only the minimum patient care needed to do this, its costs would be \$50 million. If it produced only its present volume of patient care and no medical education, its costs would be \$30 million.

Incrementally, the cost of patient care raises the school's budget from \$50 million to \$60 million. Thus, the *pure* cost of patient care is the extra \$10 million. Reasoning in a similar fashion, adding education to the cost of patient care raises the budget from \$30 million to \$60 million. Thus, the pure cost of education is \$30 million.

Notice that the difference between total cost of this hypothetical medical school and all the pure costs is \$20 million. This \$20 million is called the joint cost. It follows that if the school were

**TABLE 16-3 Hypothetical Example of Joint Production at a Medical School**

	(in millions of dollars)
Total cost of school	60
Cost if school produced only patient care	30
Cost if school produced only education	50
"Pure" cost of education <sup>a</sup>	30
"Pure" cost of patient care <sup>a</sup>	10
Joint costs <sup>a</sup>	20

<sup>a</sup>The pure cost of education is total cost (60) less the cost of producing only patient care (30). The pure cost of patient care is total cost (60) less the cost of producing only education (50). Joint costs are total costs (60) less all pure costs (30 + 10).

Source: Based on information from Newhouse (1978a).

reimbursed only for pure costs, it would run a deficit. Much of the controversy with respect to funding revolves around the problem of who will pay for the joint costs.

The issue of joint production has centered on the teaching hospital, which also jointly produces patient care and graduate medical education (GME) by providing residency and medical research. In particular, with the substantial cost differences between teaching and nonteaching hospitals,<sup>3</sup> third-party payers are concerned about whether they are implicitly subsidizing GME. Medicare, which provides most of the explicit funding for GME—\$9 billion in 2008 or about \$100,000 annually for each of the 90,000 residents that Medicare supports under a cap that was established in 1997—is also concerned about the lack of accountability in the existing payment system. In 2010, the Medicare Payment Advisory Commission (MedPAC, 2010) recommended a major overhaul of GME funding. It would establish a performance-based incentive structure where programs have financial incentives to prepare graduates with the skills needed to improve quality of care while helping to contain costs.

### Foreign Medical School Graduates

Physician supply in the United States depends to a significant degree on foreign medical school graduates (FMGs), and reliance on them continues to grow. As a proportion of the total number of active physicians, FMGs increased from less than 14 percent of the total in 1963 to 25 percent in 2008. Critics often argue that the United States and other rich nations drain valuable talent from many poor countries that have inadequate health care systems.

Nevertheless, the availability of physicians from other countries can have important policy and planning implications. Foreign national FMGs can increase the responsiveness of physician supply in the United States to changes in the physician wage.

Rapid increases in physician wages send a market signal to potential physicians, increasing the estimated rate of return to an investment in medical education. It takes a long time, however, for new applicants among American college graduates to get to medical school, get training, and enter practice. As a result, the supply of new American physicians will respond slowly to the wage signal. Foreign national FMGs, already trained but currently practicing elsewhere, can respond more quickly so the availability of foreign national FMGs makes total physician supply in the United States, in principle, more elastic.

Policymakers can apply these facts about FMGs during periods of shortages and increasing fees. Immigration policy can be relaxed to admit more foreign national FMGs during periods of physician shortages, and vice versa during periods of perceived physician surpluses. This happened during the 1960s, when preferential treatment was given to foreign national FMGs. Requirements subsequently were tightened by a 1976 act once shortages were no longer perceived.

In addition to immigration policy, the number of FMGs, including U.S. graduates of foreign medical schools, reflects certification and licensure requirements. All FMGs must take certification examinations required for admission into an approved graduate medical education program. The 1976 act also placed restrictions on the access of FMGs to graduate medical education.

A frequent concern about FMGs is quality of care. Advocates for tightening standards for FMGs usually claim that these graduates are inferior in quality to those educated in U.S. and Canadian medical schools. The claims are based on comparisons of examination performance and other measures of the credentials and personal attributes of FMGs and their U.S. counterparts.

Studies of the quality of care provided by FMGs, however, find little difference between the two groups. Some have argued that differences will more likely arise in ambulatory care settings

---

<sup>3</sup> An examination of the cost differences between teaching and nonteaching hospitals shows that nonphysician costs per day are 21 percent higher in teaching hospitals. However, sorting out the causes for cost differences and making appropriate statistical adjustments refine these data. After this is done, nonphysician costs, though still higher in teaching hospitals, show a difference that is typically less than 10 percent (Sloan, Feldman, and Steinwald, 1983).

because there is less organizational control than in hospital settings. A study that examined more than 14,000 patient episodes by nearly 1,200 physicians in three specialty groups found little difference in performance. FMGs, in fact, sometimes outperformed U.S. medical school graduates (Rhee et al., 1986).

### **The Control of Medical Education**

In 1974, Victor Fuchs wrote that “most economists believe that part [of physicians’ high incomes] represents a ‘monopoly’ return to physicians resulting from restrictions on entry to the profession and other barriers to competition” (p. 58). Fuchs refers to the claim that physicians restrict entry to their profession in order to drive up prices for their services and make larger incomes for themselves.

Do physicians control entry to their profession in order to earn above-normal returns on their investment? To answer this question, we first ask whether physicians do, in fact, earn above-normal returns. Historically, physicians often have earned above-normal returns. Distinguished economists found this to be the case using data from early in this century, and believed that control of entry was the cause. Subsequent studies found high returns in at least some more recent historical periods.<sup>4</sup>

We must further ask how physicians can control entry. Our present ideas and beliefs about the role of organized medicine in controlling entry owe much to Kessel (1958), who argued that organized medicine attained monopoly power through the licensure of physicians and the control of access to medical education. The first, licensure, is explored later in this chapter. At present, consider Kessel’s account of the control over medical education exerted by physicians primarily through the American Medical Association (AMA).

### **Control over Entry**

Shortly after the founding of the AMA in 1847, the organization campaigned state by state to get the medical profession controlled through licensure. Having largely achieved this goal by the turn of the century, the AMA turned its attention to the control of medical schools, which had proliferated in number. In 1906, the Council on Medical Education of the AMA inspected the 160 medical schools existing at that time, declaring only slightly more than half of them to be acceptable. The council sought support for this position through the Carnegie Foundation, which in 1910 issued the Flexner Report calling for substantial reductions in the number of medical schools and control on their quality. Kessel argued:

If impact on public policy is the criterion of importance, the Flexner Report must be regarded as one of the most important reports ever written. It convinced legislators that only the graduates of first class medical schools ought to be permitted to practice medicine and led to the delegation to the AMA of the task of determining what was and what was not a first class medical school. (p. 28)

Following this report, the number of medical schools fell to 85 by 1920 and to 69 by 1944. Kessel pointed out that giving the AMA charge over determining the supply of physicians was like “giving the American Iron and Steel Institute the power to determine the output of steel” (p. 29). The AMA also was able to gain control over the internship/residency process through its ability to certify hospitals for such training. It also maintained control over the process through which physicians become board-certified. The picture is one of significant power and means to control entry.

The AMA also was able to exercise control over substitute providers (e.g., optometrists, podiatrists, chiropractors) by influencing licensure to limit their scope of practice and later to limit third-party reimbursement for their services. Starr (1982) wrote of a survey of 9,000 families conducted

---

<sup>4</sup> See Burstein and Cromwell (1985). For earlier work, see Friedman and Kuznets (1945) and Leffler (1978).

between 1928 and 1931, which found that nonphysician providers treated only 5.1 percent of all attended cases of illness. He concluded that “physicians had medical practice pretty much to themselves” (p. 127).

Kessel’s argument is historical and written over 50 years ago, yet many analysts point to anomalies in recent medical school data as continuing evidence of control of entry by the medical profession. The large excess demand for medical school slots by qualified applicants to medical schools has been used to support the claims. Over time, large fractions of medical school applicants in the United States have been rejected; that is, there appears to be a substantial excess demand for medical school slots. Is this evidence of monopolizing control by the medical profession? Kessel’s view certainly suggests that it is.

However, this simple story of professional control can be challenged. Hall and Lindsay (1980) argue that medical schools do not take larger proportions of applicants and medical school enrollments respond only partially to applicant demand because the administrators of medical schools are responding rationally to their economic incentives. As we have seen, medical school revenues come not so much from tuition paid by students, but from “donors”—that is, sources such as government agencies, alumni, businesses, and research organizations. For the most part, these donors are the true demanders of the output of medical schools—trained physicians. The donors may be especially interested in applicants from certain racial or ethnic groups, females, those with specific specialization interests, and those who indicate a willingness to return to shortage areas.

In summary, we may say that organized medicine historically exerted considerable influence over the supply of trained physicians. Such influence is consistent with a view of the profession as one seeking above-normal returns by trying to control entry of new physicians. However, data in recent decades indicate that medical school enrollments respond to market forces. These data further suggest that continuing to view medical education as controlled by a monolithic or conspiratorial medical profession is implausible.

## LICENSURE AND MONOPOLY RENTS

Licensure is not unique to the health care professions. However, licensure of physicians has received unparalleled attention. The first licensure requirements for prospective doctors (though they had little effect and were minimally enforced) were passed in New York City in 1760.<sup>5</sup> Subsequently, many states introduced licensing, often through state medical societies. After the 1820s, however, many of the same states modified or abolished licensure. It was not until after the founding of the AMA in 1847 and the last decades of the nineteenth century that stronger licensure laws were widely promulgated.

The primary controversies with respect to licensure relate to its role in limiting competition and the role of professional societies on state licensure boards. The conventional view held by many economists is that organized medicine has used control of licensure for self-interest by limiting entry (and by influencing the licensure requirements of potential competitor providers to the advantage of physicians). Some, however, have advanced a public interest argument for licensure—that is, as a result of information imperfections, the public demands quality controls. Licensure and certification help fill these information gaps.

Many economists believe that licensure and professional control over medical education ensure that physicians earn economic rents, which are payments to factors over and above those necessary to induce them to provide their services. These views were heavily influenced by the early work of Friedman and Kuznets (1945) and Kessel (1958). Friedman and Kuznets examined the relative return of physicians and dentists. After adjusting for training differentials, they estimated that about half of the 33 percent excess earnings of physicians between 1928 and 1934 represented economic rents.

---

<sup>5</sup> Starr (1982, p. 44). Starr provides a fascinating history of licensure legislation in the United States.

Much has changed in the health industry since these earlier articles. However, the broader issues of monopoly rents and motivation for licensure remain controversial. Despite other empirical estimates supporting the rent hypotheses, several critiques of these studies have appeared. Leffler (1978) argued that many earlier studies failed to take into account some important economic considerations that tend to reduce estimates of the return. These adjustments include the high number of hours worked by physicians, their expected mortality rates, and the progressive income tax structure (which took an increasing share of their incremental incomes).<sup>6</sup>

A subsequent study undertaken by Burstein and Cromwell (1985) compared the internal rates of return of physicians to dentists and lawyers. The authors incorporate many adjustments into their estimates, including length of physician training, length of working life, and the earnings of medical residents. The rates of return were high compared to lawyers; for example, 12.1 percent versus 7.2 percent in 1980, the last year included in the study. The returns were high also for specialization based on board-certification requirements. This was true despite the rapid growth in physician supplies and the constraints imposed by third-party payers to contain costs over the study period. The authors concluded that “the conventional picture of medicine as a financially attractive profession is strongly confirmed” (p. 76).

This strong conclusion is further supported by a more direct test of physician pricing. Seldon and colleagues (1998) examined physicians’ price-cost margins, defined as  $(P - MC)/P$  where  $P$  represents price and  $MC$  represents costs. Under highly competitive conditions and marginal cost pricing (i.e., where  $P$  approaches  $MC$ ), the price-cost margin is zero. If physicians have monopoly power and the ability to maintain price above marginal cost, the margin will be positive. The researchers estimated the margin at 23 percent overall (and from 13 to 54 percent across the nine regional markets in the study). These estimates indicate “nontrivial” levels of monopoly power that produced a welfare loss to the U.S. economy of about \$8 billion in 1996 dollars.

**PUBLIC INTEREST AND SELF-INTEREST THEORIES OF REGULATION** The specific issue of licensure is but a part of the broader issue of regulation reflecting the two competing theories: public interest versus self-interest motives. The public interest motive is based on theories of market failure, such as information failure. According to the public interest view, the demand for regulatory measures, such as licensure, stems from the limited information patients have about quality and the relatively high costs of obtaining information. Drawing on Akerlof’s lemons principle introduced in Chapter 10, Leffler (1978) argued that asymmetric information will lower quality. Thus, a “state-enforced minimum quality standard is claimed to be an efficient response to costly quality information” (p. 173).

In contrast, the self-interest motives for licensure and other forms of regulation to reduce competition have long been accepted in economics but only relatively recently have been formalized (Stigler, 1971; Peltzman, 1976). In this theory, discussed in more detail in Chapter 19, regulation is a return to special interests that provide financial and political support in return for favored legislation. Thus, a demand for political favors arises from the rent-seeking behavior of special interest groups. The effort and amount of resources expended by a special interest group are limited by the rents that would accrue from the favored legislation.

**EVIDENCE OF PUBLIC INTERESTS VERSUS SELF-INTERESTS** Paul (1984) tested the public interest versus self-interest theories using data on the initial decisions by states to license physicians, and he rejected the public interest theory. His findings show a strong negative association between the year of initial licensure and the number of AMA-associated physicians in a state per capita.

---

<sup>6</sup> In addition to the adjustments suggested by Leffler, Headen (1990) separated the labor from entrepreneurial returns for self-employed physicians in family practice. He found that about 16 percent of the reported incomes are associated with their entrepreneurship.



Graddy (1991) also tested the competing hypotheses by estimating the probability (and type) of regulation by states of six health care professions. Variables representing the public-interest view included the profession's educational requirements in years (to capture the complexity of the service) and professional liability insurance rates (to represent the potential risk to consumers in the absence of regulation). The legislative environment was represented by variables, such as the strength of the majority party and the rates of turnover of legislators. Variables, such as the size of the occupation and its proportion of members belonging to a professional association, represented private interests.

The statistical approach was possible because licensure practices for many professions vary substantially across states. Graddy found important roles for each of the categories of explanatory variables in determining the kinds of regulation. She found a higher probability of a stricter form of regulation as the profession's educational requirements are higher—a finding consistent with a public-interest motive. The overriding conclusion, though, is that no single dominant motive can be found for regulation. Legislators respond to organized interests, the public interest, and their own legislative environments.

### Licensure and Quality

The Graddy evidence supports, in part, a public demand for regulation. We also know that physician board certification, or even board eligibility, increases remuneration, meaning consumers are willing to pay more for those with additional training and credentials. For example, O'Halloran and Bashaw (2006) simultaneously estimated the decision to become board certified and the returns to board certification. The likelihood of "investing" in certification is greater for physicians who stand to gain the most including those who practice in more competitive markets (in order to stand out) and those with lower explicit and implicit costs associated with certification. Minority physicians earn a smaller reward from certification and they are less likely to become board certified. Overall, the authors conclude (p. 641) that "physician decisions to become board certified in their respective specialties rigidly follows a pattern consistent with human capital theory."

Does licensure actually improve the quality of care? Gaumer's (1984) review of the empirical evidence questions whether the goals of protecting the public and ensuring minimal standards of competency are being achieved. He found that (1) in spite of licensure, a substantial amount of deficient care occurs; (2) quality of care would not be impaired if the scope of practice of secondary (nonphysician/dentist) providers were increased; (3) the licensing process may "not accurately assess the practice competence of applicants" (p. 397); and (4) fees and provider incomes are higher in states with more restrictive licensure requirements (supporting the self-interests motive for regulation).

More specifically, with respect to the quality of physician care, he cites studies indicating that 5 percent of physicians are "unfit to practice," 8 to 22 percent of obstetrics patients and 61 to 65 percent of well-care patients received deficient care, and that 7.5 percent of all cases in two hospitals indicated physician-inflicted injury (p. 395).

Brennan and colleagues (1991) provide additional evidence on the quality of medical care in hospitals. Licensure is just one of many regulatory requirements intended to ensure quality. From a large number of randomly selected admissions, the researchers found that nearly 4 percent produced "adverse events," defined as injuries caused by medical management. Nearly 14 percent of these injuries led to death. The authors concluded that "there is a substantial amount of injury to patients from medical management, and many injuries are the result of substandard care" (p. 370).

Though no one is suggesting that eliminating licensure and other requirements will reduce such negative outcomes, it is clear that regulation does not ensure quality care. The national concerns with health care quality, as described in Chapters 10 and 14, and the policies promoted by

Congress to improve quality, are clear examples of the limits of regulation. Nevertheless, one could still ask whether tighter regulation would help. From his review, Gaumer concludes:

Research evidence does not inspire confidence that wide-ranging systems for regulating health professionals have served the public interest. Though researchers have not been able to observe the consequences of a totally unregulated environment, observation of incremental variations in regulatory practices generally supports the view that tighter controls do not lead to improvements in quality of service. (p. 406)

As a result of the questionable effects of licensure on quality, changes in the health care environment, and the anticompetitive effects of restrictions on entry and restrictions on the scope of practice of potential competitors (e.g., podiatrists, nurse practitioners), the benefits of licensure are being re-examined. Svorny (1992) suggests that the benefits have been weakened by, among other things, the added liability that courts have placed on hospitals and HMOs for the negligent conduct of independent physicians and by the increased use of salaried physicians. In a stronger attack, Safriet (1994) concludes:

Clearly these barriers serve no useful purpose, and in fact contribute to our health care problems by preventing the full deployment of competent and cost-effective providers who can meet the needs of a substantial number of consumers. (p. 315)

Have these conclusions changed over the past two decades—a period in which many states have expanded the scope of practice of nonphysician clinicians and in which the influence of organized medicine has been thought to wane? Not according to Svorny (2008) who describes the politics of licensure and the “turf wars” between medical physicians and other providers, writing:

Medical licensure fails to meet expectations in the area of discipline and consumer protection. State medical boards’ disciplinary efforts can arguably be said to protect clinicians more than consumers. (p. 11)

## **OTHER PHYSICIAN LABOR ISSUES**

The prominence of physicians and their dominating role in treatment decisions have led to important research on a variety of labor issues. Many are associated with physician earnings, and we examine two of these issues below.

### **Specialization**

Studies of physician specialty selection are especially important because of widespread beliefs that quality health care requires access to an appropriate mix of specialists. Policy effort in recent years has also been directed toward encouraging more physicians to go into primary care, especially in underserved areas. Some medical schools have responded to the challenge by favoring applicants who are committed to primary care.

The primary care challenge, however, is formidable due to the wide gap in earnings between specialists and generalists. For example, the U.S. Department of Labor reported median compensation of \$338,738 for medical specialists in 2008 compared to \$186,044 for primary care physicians.

Economists recognize the role that nonpecuniary rewards, such as status and social responsibility, may play in decisions to specialize. Nevertheless, the economic focus is usually on the degree to which physicians respond to financial incentives. Decisions to specialize normally occur early in the physician’s education and training career, so Nicholson (2005) was curious about medical students and their knowledge of physician incomes. He examined surveys of first- and fourth-year

students conducted annually by large medical schools and found systematic biases in the responses. The students overestimated incomes in the 1970s but underestimated incomes by about 25 percent in more recent years. The study also showed a significant learning pattern with estimates that were much more accurate for fourth-year students than for those in their first year. The estimates were also more accurate for a specialty that a student was more likely to select.

How strong is the response to earnings potential, especially in light of the wide income variations by specialty? Economic theory suggests that a rational decision ought to be based on expected lifetime income, not simply the current earnings within a specialty, and several studies indicate substantial responsiveness. Estimated elasticities of entry into specialties with respect to changes in expected lifetime earnings are usually greater than one, leaving little doubt that physicians respond to income when making their specialty decisions. Nevertheless, one study of specialty choice provides some unique results on differences in specialty earnings.

Bhattacharya (2005) describes four possible explanations for the wide income disparities across specializations: (1) differences in hours worked, (2) differences in length of residency and other required training, (3) difference in the attributes and skills needed to perform in a specialty, and (4) barriers to entry into some specialties. The first three possibilities, which reflect competitive labor market forces, account for only one-half the observed differences in lifetime earnings. He concluded that the remainder of the differences results from differential entry barriers, suggesting that some specialist fees are excessive relative to competitive fees. From a policy perspective, a strong case can be made to pressure those specialty boards that impose relatively high entry barriers to increase their number of residency slots.

### Physician Income by Gender

Women now account for about one-half of new medical school graduates. They also represented 29 percent of professionally active physicians in 2008 compared to just 11 percent in 1980. This dramatic shift toward gender balance will likely continue, and sociologists have eagerly studied a wide range of phenomena including the female physician's approach to patient care relative to the male physician. Economists have focused more narrowly on labor market issues, such as differences in earnings, job status, and hours worked.

Female physicians earn considerably less than male physicians. A survey by *Medical Economics* indicated that male compensation in primary care was 23 percent higher than female compensation in 2003; it was 54 percent higher among all respondents. Among the usual reasons for the gender gap is that women are more likely to choose the lower-paying specialties, and to work fewer hours than male physicians due to disproportionate burdens in raising a family. Sasser (2005) describes several mechanisms through which family responsibilities affect the gender gap. These include the fewer number of years in active practice that female physicians may anticipate and, thus, a reduced willingness by the female physician or her employer to invest in human capital. Greater household responsibilities could affect specialty selected and characteristics of the job environment sought. After controlling for specialty, practice setting, and demographic and professional characteristics, Sasser finds that female physicians "earn 11 percent less for being married plus 14 percent less for having one child and 22 percent less for having more than one child." The main determinant of these earnings differences are personal choices to reduce working hours sharply upon marriage and having children.

Despite Sasser's strong conclusions, the story of the gender earnings gap remains incomplete and complex. (See Box 16-3 for a recent study.) Timothy Hoff (2004) provides a rich example using hospitalists, a relatively new specialty consisting of hospital-based general physicians who focus on the care of hospitalized patients. He found that female hospitalists earn significantly less per year than their male counterparts even after controlling for demographic, professional, and job-related characteristics. Hoff also controlled for marital status and children. Furthermore, male and female hospitalists worked similar schedules and had similar levels of commitment. Thus, he concluded that the pay gap, at least in this new specialization, is real and not due mainly to personal and career

**BOX 16-3****The \$16,819 Unexplained Gender Income Gap**

Lo Sasso and colleagues (2011) analyze earnings data for physicians leaving residency programs in the state of New York over the ten-year period, 1999–2008. The survey data provide information on many observable factors that might influence earnings including specialty, hours worked, age, practice type, and employer location. By examining starting pay, the investigators avoid the difficult problem of controlling for some determinants of compensation, e.g., productivity, that are known over a period of time only after the initial hire.

Over the period covered by the study, an increasing proportion of females entered specialized fields. Nevertheless, the unadjusted female-to-male earnings ratio dropped from 87.4 percent in 1999 (\$151,600 for females vs. \$173,400 for males) to just 83.1 percent in 2008 (\$174,000 vs. \$209,300). After ordinary least-squares regression was used to control for the observable factors, the unexplained pay gap of \$35,400 in 2008 was reduced to \$16,819. This unexplained gap is substantial especially as compared to the statistically insignificant differential estimated for 1999.

The investigators consider a variety of explanations including the possibility that unobservable factors account for the widening adjusted pay gap. One of the unobservable factors considered is the possible change in employment practices resulting from the influx of a large number of female graduates. In particular, the authors speculate (p. 198) that employment practices “may now be offering greater flexibility and family-friendly attributes that are more appealing to female practitioners but that come at the price of commensurately lower pay.”

choices. Hoff urges the medical establishment and policy makers to take pay inequality seriously and to develop mechanisms to address the problem.

**CONCLUSIONS**

In this chapter, we used basic economic tools to provide important insights into a variety of health care labor issues, including the demand and supply of labor, optimal input decisions and factor substitution and labor shortages. We examined two earnings issues as they related to specialization and the gender gap. The chapter has also addressed several aspects of medical education. In particular, we have tried to examine whether various characteristics of physician training and licensure are designed to increase barriers to entry into the profession, producing higher-than-normal rates of return.

We caution that rapid restructuring of the U.S. health care system, especially through managed care and post-managed care initiatives, is creating profound changes. As previously noted, physicians are increasingly organized in or affiliated with large groups that compete for managed care contracts. The widespread purchase of physician practices by hospitals provides another example of a dramatic change. The adoption of pay-for-performance and growth of consumer-driven health plans are other recent innovations that may have major effects on health care delivery. As a result of these developments, the economic power and professional influence of physicians have undoubtedly been affected in ways that are still largely unknown. Clearly, these represent new challenges for the health economist.

**Summary**

1. In 2009, 15.5 million people, representing 11.1 percent of all employed civilians, worked at health services sites.
2. The demand for labor is precisely the marginal revenue product of labor curve. It is closely related to the production function and is directly related to the price of the output.
3. The supply of labor tends to slope upward, implying that the higher the wage rate is in a given market, the more laborers will be forthcoming.

Workers currently employed may choose to work more hours if higher wages are offered; other workers may be attracted from elsewhere by the higher wages.

4. There are basically two types of shortages: need shortages and economic shortages. Need shortages use a nonmarket, or noneconomic, definition of shortage.
5. One definition of an economic shortage is the excess of quantity demanded over the quantity supplied at the market wage rate. Stickiness in wages helps explain why the wage does not rise to equilibrium, thus automatically eliminating the shortage.
6. Meaningful shortages of professionals may exist even when supply and demand are in short-term equilibrium. If demand for a category of health professional expands over time and supply is slow to respond, the result may be a wage increase that is large relative to wage gains of other professionals.
7. Under labor monopsony conditions, a firm may report unfilled, budgeted positions, for example, for registered nurses, even when the firm is in equilibrium. A monopsony firm may announce that it wishes to hire more nurses even though it is unlikely to take the necessary steps to do so.
8. Medical education is heavily subsidized. Tuition is a relatively small source of revenues; thus, the student pays only a small portion of the true costs of the investment in education.
9. Medical education is a good example of joint production. Medical schools produce medical education, patient care, and research.
10. Kessel argued that monopoly power was attained by organized medicine through licensure of physicians and control of access to medical education.
11. According to an alternative view of medical education, the donor-preference hypothesis, medical school revenues come not so much from tuition paid by students but from donors, such as government agencies, alumni, businesses, and research organizations. For the most part, it is these donors who demand and control the output of medical schools: trained physicians.
12. Licensure is a prominent example of the controversy of self-interest versus public interest views of regulation.
13. It is generally believed that licensure has given physicians economic rents. Licensure has not led to obvious improvements in quality.
14. There are wide differences in the earnings across physician specialties. Usual labor market explanations account for only one-half the variations, suggesting differences in barriers to entry among specialties.
15. There are also substantial gender differences in earnings. It is not yet clear whether these differences can be fully explained by the personal and professional decisions made by female physicians.

## Discussion Questions

1. Give examples of ways in which labor and capital can be substituted for each other in the production of health services.
2. In the text, we considered only forgone income and tuition as costs of going to medical school. Enumerate other monetary and nonmonetary opportunity costs. Why are these opportunity costs relevant?
3. Why might demand for nurse labor by hospitals or other organized health providers be monopsonistic?
4. What is the marginal product of an input? Marginal revenue product? Why does the demand for a factor correspond to the marginal revenue product curve? What will determine whether the demand for a factor will be elastic or inelastic?
5. Why will a profit-maximizing physician firm want to equalize the marginal product per dollar spent across all inputs?
6. What is meant by the term *barriers to entry*? What are some entry barriers for someone who wants to be an obstetrician? For someone who wants to be a nursing assistant?
7. If barriers to entry into a profession were absolute so that entry would not be possible, what would the supply curve look like? What would the supply curve look like if entry into an occupation were free and easy? Thus, what role do barriers to entry play in explaining relative rates of return to an occupation?
8. Define *monopsony* and *marginal labor cost*. Why is the marginal labor cost in the case of monopsony above the supply (average labor cost) curve? What is the nature of the inefficiency or misallocation associated with monopsony power? Is there any inefficiency when the supply curve facing the monopsonist is perfectly elastic?
9. Some argue that imperfect capital markets constitute one reason for subsidizing medical education. Discuss why the market for human capital differs from, say, the market for housing capital (that is, buying a house).
10. If there were no subsidies for medical education, would enrollments be larger or smaller? Would the return to medical education be larger or smaller? If physician education was not subsidized, would the economically warranted supply of physicians tend to emerge?

11. What are the social benefits and costs behind regulating the number of medical schools?
12. What is joint production? What does the term *joint production costs* mean? Given that medical schools engage in joint production of education, patient care, and research, what inferences can be drawn about the economies of scope in producing these three outputs?
13. In contrast to medical education, numerous night and part-time law schools have been established. Compare and contrast the various aspects of training that have led to these different educational systems.
14. What are some factors that help explain earnings differences across specialties? Why might the earnings differences persist over long periods of time?
15. The rate of return on investment in medical education exceeds that for other professions. What are arguments for and against government subsidies?
16. Female physicians earn considerably less than their male counterparts. Discuss some of the reasons that account for the differences. What kind of evidence would lead one to conclude that at least some of the difference is due to bias or discrimination?

---

## Exercises

1. Consider the firm's demand (*MRP*) for labor, such as in Figure 16-2. If the demand elasticity is  $-0.5$ , what will be the effect of increased wages on total labor earnings?
2. Using Figure 16-3, graph and analyze the impact of an increase in the price of lab tests on the labor market.
3. Consider the market for highly skilled laboratory technicians. Graph the impacts on market wages if limitations on immigration were lifted. Would more or fewer services be provided? What would happen to the price?
4. In this chapter, we discuss how physicians' marginal products rise up to 25 hours and then slowly fall to zero at 110 hours. Graph both marginal and total products from this statement.
5. Using supply-and-demand analysis, model the equilibrium level of physicians' wages. What would be the impact on physicians' wages of more stringent policies on the employment of foreign medical school graduates?
6. Suppose that a medical school provides three outputs—patient care, education, and research—and that the total cost of the school is \$100 million per year. If the school produced only education, its costs would be \$60 million. If the school produced only patient care, its costs would be \$30 million. If it produced only research, the costs would be \$20 million. Joint costs for each pair would be \$10 million.
  - (a) What are the pure costs of education, patient care, and research?
  - (b) What are the joint costs?
7. Suppose that the licensure requirements for health care providers were eliminated. Use supply-and-demand analysis to predict what may happen to the price and quantity of health care services. Are there other considerations—in particular, mechanisms—that could evolve to replace licensure?

# The Pharmaceutical Industry

- Structure and Regulation
- The Production of Health and Substitutability
- Drug Pricing and Profits
- Research and Development (R&D) and Innovation
- Cost Containment
- New Drugs and Health Care Spending
- Conclusions

Prescription drugs and the pharmaceutical industry occupy increasingly important places in the health economy. Drug therapies traditionally have supplemented nutrition, sanitation, and medical care as methods for preserving health. Vaccinations for diseases such as smallpox represented early public health initiatives that saved thousands of lives. Insulin, developed in the 1920s, prevented the certain deaths that once accompanied diabetes, and a world without antibiotics, introduced in the 1940s, or the polio vaccination, in the 1950s, would be unthinkable.

Drugs are used to treat many diseases and conditions. Examples include chemotherapy for cancer, steroids for skin diseases, psychotropic drugs for mental health problems, beta-blockers for heart disease, clot busters for stroke, and protease inhibitors for AIDS. Some drugs prevent disease; some substitute for more invasive surgical procedures; some are used in conjunction with surgical and radiation treatments; while others provide treatment for conditions where no treatment was available previously. In recent years alone, observers view the introduction and widespread use of cholesterol absorption inhibitors to reduce the amount of cholesterol delivered to the liver for at-risk populations as a major breakthrough in the fight against coronary heart disease.

Despite these successes, the U.S. pharmaceutical industry has encountered intense media and legislative scrutiny. Pharmaceutical firms have been among the largest and most profitable businesses in the United States. As recently as 2001, the drug industry ranked first in various measures of profitability among *Fortune's* industry groupings. Negative publicity, litigation problems, widespread efforts to contain drug spending, and loss of patent protection for several major drugs since then (see Box 17-1) are serious threats to profitability. Nevertheless, the 12 pharmaceutical firms among the *Fortune 500* in 2009 reported a median profit of 20 percent on revenues and 23 percent on stockholders' equity. These were among the highest of all industries.

**BOX 17-1****Patents and Media Attention**

As we have noted, the pharmaceutical industry has a long history of superior financial performance. On account of perceptions of “exorbitant” prices and other questionable practices, drug companies are often the subject of unflattering media coverage. Here is one prominent example from early 2011.

In February 2011, the FDA granted KV Pharmaceutical of St. Louis exclusive rights for seven years for the injectable form of a drug marketed as Makena. Makena reduces the risk of preterm delivery for expectant mothers with a previous premature delivery. Early delivery is a serious and growing problem in the United States so FDA approval sounded like welcome news.

However, Makena is chemically the same as another drug that was produced for years by a different firm and then withdrawn from the market. It was subsequently made by “compounding” pharmacies (those pharmacies that actually mix prescription lotions, creams, or doses for injection) at a cost of about \$10 to \$20 per shot (an expectant mother receives about 20 injections over the first 4-5 months of pregnancy). Some worry about the quality and consistency of the drug made by these pharmacies, but they also worry about the \$1500 price that KV initially established for Makena. KV also warned specialty pharmacists that compounding the mixture would be unlawful. Following the media attention, KV announced a price cut to \$690 per injection. The FDA also indicated in March 2011 that it will not take enforcement action against pharmacists that compound the drug.

Many important issues relate to patent protection (granted by the patent and trademark office) and exclusivity (granted by the FDA), and we will examine some in later sections of this chapter. At this time, we note that, despite its historical success, the pharmaceutical industry has recently faced unprecedented challenges to replace expiring patents with new revenue streams. In 2011 alone, patents expired on a large number of blockbusters including Lipitor (to lower cholesterol), Advair (to prevent asthma symptoms), and Zyprexa (to treat schizophrenia). The number of new drugs approved is running at a low rate and the Food and Drug Administration (FDA) appears less likely to approve new products if there are significant side effects and if good therapies are currently available. An article in the *New York Times* concluded that the industry is “now under pressure to reinvent itself and shed its dependence on blockbuster drugs.” The article also quoted another source which warned that “The operating environment for pharma is worsening rapidly.”

*Sources: Bloomberg Business Week, “FDA Approves First Drug to Prevent Premature Births,”* February 6, 2011, [businessweek.com/lifestyle/content/healthday/649631.html](http://businessweek.com/lifestyle/content/healthday/649631.html); Gardiner Harris, “Drugs’ Cost and Safety Fuel a Fight,” *New York Times*, April 4, 2011, [nytimes.com/2011/04/05/health/05FDA.html?\\_r=1&pagewanted=print](http://nytimes.com/2011/04/05/health/05FDA.html?_r=1&pagewanted=print); Duff Wilson, “Drug Firms Face Billions in Losses in ’11 as Patents End,” *New York Times*, March 6, 2011, [nytimes.com/2011/03/07/business/07drug.html](http://nytimes.com/2011/03/07/business/07drug.html).

Rapid growth in drug spending has made the pharmaceutical industry a convenient target for the budgetary challenges facing patients and insurers. The introduction of new and expensive drugs, increasingly promoted through direct marketing to consumers, has only added to the expenditure burden and heightened criticism of the industry.

This burden becomes especially severe for those who suffer disproportionately from chronic and other conditions that fuel the use of drugs. Policies to deal with these pressures include the 2006 expansion of Medicare (Part D) to include outpatient prescription drug benefits (see Chapter 21), and proposals to regulate prices as well as to permit the re-importation of drugs from Canada and other countries. Private insurance initiatives include higher patient copayments, increased emphasis on generic products, and new strategies, such as the development of drug formularies. Most managed care plans adopted formularies, that is, approved lists of drugs, by the late 1990s. Patients may have difficulty in obtaining reimbursement for any drugs that are not on the list.

Pharmacoeconomics, which includes cost-benefit, cost-effectiveness, and cost-utility analyses, plays an increasingly important role in pharmaceutical decisions, but policy must



address other questions. This chapter selects several of the most general interest to health economists. After describing the structure and regulation of the pharmaceutical industry, we focus on the following areas:

1. The role of pharmaceutical products in the production of health, patient choices of drugs under various insurance schemes, and the effects of technological change on the use of drugs
2. Drug pricing issues, including price discrimination by sellers and price regulation by the government
3. Pharmaceutical research, the determinants of innovation, and the effects of price regulation on innovation
4. Cost containment through use of generic products and other measures

We conclude with recent evidence on the effects of drugs on health derived from international comparisons on drug utilization.

## STRUCTURE AND REGULATION

In 2009, spending on prescription drugs amounted to \$250 billion or 10.1 percent of national health expenditures (NHE), up from 8.8 percent in 2000 and just 4.7 percent in 1980. Consumers' out-of-pocket costs for drugs represented 21 percent of total spending on drugs in 2009, and accounted for 18 percent of all out-of-pocket costs.

Although prescription drug spending has stabilized at about 10 percent of NHE in recent years, its high long-run growth rate relative to other categories of health spending created considerable interest in the sources of these increases. Analysts have found that most of the increases have been due to greater use of drugs and to new products rather than to higher prices of existing products. Nevertheless, rising expenditures and high out-of-pocket costs help create public perceptions that something is seriously wrong with the conduct of pharmaceutical firms, and that stiff measures are needed to contain drug costs.

With its long history of relatively high profits and rich set of features—patent protection, high research and development spending, intense product promotion, and heavy regulation—the pharmaceutical industry always has drawn the attention of economists in the field of industrial organization. Scholars describe levels of competition in an industry; how the competitive environment influences decisions on prices and other decision variables, such as advertising, research and development (R&D), and quality; and the consequences of these decisions for socially efficient allocations of resources.

### Competition

The level of competition often holds the key to firm and industry behavior. To measure competition, economists need to look at meaningful industry groups. Prior to 1997, the Standard Industrial Classification (SIC) served as the standard. The Department of Commerce has since replaced the SIC codes with the North American Industry Classification System (NAICS). The NAICS codes range from two to six digits, with each successive digit representing a finer degree of classification. The “pharmaceutical preparations” industry, NAICS Code 325412, formally consists of “establishments primarily engaged in manufacturing in-vivo diagnostic substances and pharmaceutical preparations (except biological) intended for internal and external consumption in dose forms, such as ampoules, tablets, capsules, vials, ointments, powders, solutions, and suspensions.”

The four-firm ( $C_4$ ) and eight-firm ( $C_8$ ) concentration ratios for any selected six-digit NAICS codes indicate the share of industry output produced by the four or eight largest firms. Analysts use these concentration ratios, shown in Table 17-1, for pharmaceuticals and several other six-digit industries, to gauge competition. The pharmaceutical industry is one of the largest manufacturing industries. As measured by concentration ratios, it also appears to be much more competitive than many others.

**TABLE 17-1 Concentration in Selected Manufacturing Industries: 2002**

NAICS Code	Industry	C <sub>4</sub>	C <sub>8</sub>	HHI	N	Shipments (in \$ billions)
325412	Pharmaceutical preparation mfg.	36	53	530	731	114.7
311230	Breakfast cereal mfg.	78	91	2,521	45	9.1
324110	Petroleum refineries	41	64	640	88	193.5
334111	Electronic computer mfg.	76	89	2,662	934	32.3
334220	Radio & TV broadcasting & wireless	43	55	584	823	32.1
325510	Paint and coating mfg.	37	55	505	1,149	19.9
325611	Soap and detergent mfg.	61	72	2,006	699	16.6
336111	Automobile mfg.	76	94	1,910	164	88.1
336112	Light truck & utility vehicle mfg.	96	100	W	69	137.1
336411	Aircraft mfg.	81	94	W	184	64.3

Note: W = withheld to avoid disclosure of individual firm data. The undisclosed HHI value will undoubtedly be very high.

Source: U.S. Bureau of the Census, 2002 Economic Census, "Concentration Ratios: 2002," Report EC02-31SR-1 (May 2006).

Another method used to measure competition is the Herfindahl-Hirschman Index (HHI). Compare an industry with four firms, each with 25 percent of the market, to a second industry, also with four firms, but where one firm has 85 percent of the market and the other three each have 5 percent. Both industries have a four-firm concentration ratio of 100 percent. However, one might guess that the one in which the leading firm has 85 percent of the market is more monopolized. The HHI incorporates differences in the size distribution of firms by squaring the market shares of each and adding them together, so that the lowest value approaches 0 (thousands of tiny firms) and the highest value approaches 10,000 (with a pure monopoly). In the previous example, the respective HHIs are 2,500 (the four equal sized firms) and 7,300 (the very large firm, with three smaller ones).

Table 17-1 shows that compared to other well-known industries, in addition to relatively low concentration ratios, there are a relatively large number of firms (*N*) and the HHI (for up to the 50 largest firms) is relatively low. Do these data indicate substantial competition? Most analysts would argue that for pharmaceuticals they could be especially misleading. Drugs in different therapeutic categories usually are not substitutes for each other. Concentration ratios for narrower drug classes are better indicators. When such data are available, they still can show considerable competition. In many cases, however, the concentration ratios will be higher, sometimes much higher. Schweitzer (1996) illustrates this with a class of drugs used to control hypertension. The top four firms controlled 91 percent of the market in 1992. Why? Patents and other barriers to entry often restrict competition.

### Barriers to Entry

A barrier to entry is any factor that impedes the entry of new firms into an industry or product market. Patent protection granted by government represents a classic example. To gain further protection, pharmaceutical firms adopt a common business strategy of surrounding a product with patents on many variations of that product.

A patent forms a legal barrier. Advertising and promotion also can create economic barriers when they successfully increase brand loyalty. Pharmaceutical promotion differs from that of typical consumer goods because pharmaceutical firms direct much of their marketing at physicians rather than patients, the end users through "detailers"—pharmaceutical representatives who directly visit physicians' offices. Critics of this practice believe that detailing may lead to

questionable financial arrangements that encourage the physician to prescribe a particular product, possibly in place of cheaper drug or nondrug substitutes.

Pharmaceutical firms also reach physicians by distributing samples, by direct mail, and by advertising in medical journals. Following the FDA's relaxation of rules governing advertisements through the media in 1997, the industry responded by increasing advertising in newspapers, on radio, on television, and even on freeway billboards, all aimed directly at patients (see Box 17-2). The medical community and other critics of DTC have raised concerns about the misinformation, confusion, and unnecessary or even harmful treatment that could result from such advertising.

As a last example of protection from competition, the regulation of drugs itself can create entry barriers. The Food and Drug Administration (FDA) approval process for a new drug is costly and time consuming. A new firm will find it difficult to marshal the financial and expert resources needed to go through the process and especially to have a portfolio of products under development to spread risks. According to the industry, only "five in 5,000 compounds that enter preclinical testing make it to human testing" and only one of these five ultimately is approved as a drug (PhRMA 2010, p. 16). Fewer still ever become profitable. Not surprisingly, such long odds create formidable deterrence to new drug development, and new pharmaceutical firms often concentrate on generic products.

## BOX 17-2

### Direct-to-Consumer (DTC) Advertising

Prior to 1951, the distinction between over-the-counter and prescription drugs was not as well defined as it is today. The FDA did require that certain highly potent and potentially dangerous drugs be available only through prescription, but the decision for many others was left to the producer until the 1951 Durham Humphrey Amendment to the Food, Drug, and Cosmetic Act of 1938. In the following years, an increasing proportion of medications were available only through prescriptions. However, until the 1990s, the pharmaceutical industry overwhelmingly concentrated its promotion efforts on doctors, largely through "detail men" who would visit physician offices. This strategy reflected the prevailing view of the medical decision-making model as one based on the authority of the physician over a passive patient.

It was not until the 1990s that marketing managers began to re-evaluate the potential of DTC. There were two important developments in this process: (1) the growth of managed care, which constrained consumer choices and put downward pressure on drug prices, and (2) the growth of consumerism in general, but especially in health care. Nevertheless, DTC marketing amounted to only \$363 million in 1995, with just 15 percent directed to the broadcast media. The major impetus came in 1997 after the FDA made it easier for broadcast ads to meet requirements regarding a summary of the risks and benefits of the advertised product. For example, the ad could now direct consumers to a toll-free number or to a Web site for such information. DTC advertising, especially on television, grew rapidly, reaching \$4.8 billion in 2006 for the research-based pharmaceutical firms (out of \$12 billion spent on all marketing and promotional activities).

DTC television advertising remains one of the most controversial and visible practices of the pharmaceutical industry. It is allowed only in the United States and New Zealand among advanced countries. The drug industry maintains that DTC advertising "creates awareness of diseases and treatment options and empowers patients with information." However, the considerable fallout from Merck's 2004 withdrawal from the market of Vioxx spurred 2007 legislation strengthening the FDA. It also led to a re-assessment of DTC advertising. Vioxx was a heavily promoted and widely prescribed nonsteroidal anti-inflammatory drug that was associated with thousands of heart attacks and strokes. After Vioxx was withdrawn, other firms declared voluntary moratoriums on DTC ads for some new drugs. In 2007, a federal advisory panel rejected Merck's proposal to market a new drug that is similar to Vioxx.

Sources: Donohue (2006) and PhRMA (2008).

## Regulation

The pharmaceutical industry is one of the most heavily regulated of all industries. Governments regulate most firms for worker safety and health concerns, but pharmaceutical products face further oversight by the FDA. Following a public scandal over adulterated food products and dangerous medicines with unknown contents, the federal government introduced the Food and Drug Act of 1906. The act did nothing to prevent the public from dangerous medicines. It did not even require formal testing but dealt mainly with labeling. Requirements for testing and safety were introduced with the Federal Food, Drug, and Cosmetic Act of 1938. However, these requirements were left mainly to the drug companies.

Two events accelerated regulatory change. Exposure of questionable drug industry practices in hearings held by Senator Estes Kefauver in 1959 was soon followed by the thalidomide tragedy. Thalidomide, a tranquilizer widely used in Europe to treat morning sickness in pregnancy, was discovered to cause severe defects in babies, who were sometimes born with deformed, flipper-like limbs. The drug was available on an experimental basis in the United States at the time. Fortunately for the United States, the number of thalidomide babies was relatively small. The FDA delayed approval, and the distributor withdrew the product quickly after reports of the European experience.<sup>1</sup>

Although the thalidomide tragedy was averted in the United States, Congress nonetheless approved amendments in 1962 that gave the FDA increased control over the introduction of new products. The new legislation required much more testing and extended the FDA's authority to regulate premarket testing (including generic drugs). Equally important, the legislation for the first time required evidence of efficacy.<sup>2</sup>

FDA review has become a lengthy, complex process. Following the discovery stage during which new chemicals are synthesized, the firm conducts preclinical animal studies involving short-term toxicity and safety tests. The drug firm next must file an application with the FDA to conduct clinical trials. If approved, the trials are conducted in three phases. Phase I begins with small groups of healthy volunteers and focuses on safety and dosage. Phase II trials involve a larger number of subjects, often several hundred, who have the targeted condition, and concentrates on the drug's efficacy.<sup>3</sup> Phase III trials usually are conducted on thousands of patients in different settings so that safety and efficacy can be determined more precisely.

If these trials indicate safety and efficacy, and the drug's safety is supported by long-term animal studies, the company submits a New Drug Application (NDA) containing all the data and results to the FDA. The FDA review usually takes more than a year. Total development time for a new product stands at about 14 years, nearly double the eight-year period in the 1960s (DiMasi, 2001).

These requirements provoke considerable controversy and provide obvious trade-offs between the goals of protecting the consumer and rapid innovation. The economic approach is to weigh the gains in safety and efficacy against the cost of delaying patients from utilizing useful products. Economists also express concern about the potential stifling of innovation caused by regulation and its adverse effects on competition.

In a classic study of the 1962 amendments, Peltzman (1974) found a sharp decline in new product development, especially of innovative drugs, after 1962, as well as higher prices from the decreased competition. These consequences far outweighed the benefits of reduced spending on ineffective drugs, creating a net welfare loss of about 6 percent of total drug sales.

---

<sup>1</sup> Thalidomide still is marketed with a warning on pregnancy. It has shown promise in treating a variety of conditions, including autoimmune disorders for patients with advanced AIDS. In 2006, the FDA approved thalidomide for treatment of multiple myeloma, a cancer of plasma cells. Women who take it, however, must agree to take contraceptive measures.

<sup>2</sup> The legislation also gave the FDA authority over the manufacturing process and extended testing requirements to include generic drugs and drugs that are similar to available products ("me-too" drugs). It also transferred regulation of drug advertising for prescription products from the Federal Trade Commission to the FDA.

<sup>3</sup> The terms *efficacy* or *effective*, as used by the FDA, mean that a drug has positive effects compared to a placebo. Efficacy does not imply that the product is cost-effective or that it meets other economic criteria for efficiency.

The FDA recognized these problems and in the mid-1970s developed policies to accelerate the review of “important” drugs. Dranove and Meltzer (1994) found that important drugs reach the market about three years sooner than other drugs. Thus, they argue that the losses resulting from delays in the approval process have been overestimated. A 1984 act also eliminated the full range of tests for generic products that were required by the 1962 amendments.

To expedite the review process, 1992 legislation and the Modernization Act of 1997 provide the FDA with additional resources derived from user fees levied on the industry. This has considerably reduced approval times. The Food and Drug Administration Amendment Act of 2007 included components that enhanced FDA authority and gave it significant increases in users fees to conduct comprehensive reviews of drugs and medical devices.

Philipson et al. (2008) found a very favorable trade-off between approval times and safety of legislative changes between 1992 and 2002. More rapid access to drugs saved between 140,000 to 310,000 life-years compared to an upper bound of 56,000 life-years lost due to harmful effects of drugs before they were withdrawn from the market. However, Olson’s (2008) work reminds us of the risk. She found that a reduction in review time of one standard deviation increased serious adverse drug reactions by 21 to 23 percent, and hospitalizations and deaths from these reactions by about 20 percent each.

## THE PRODUCTION OF HEALTH AND SUBSTITUTABILITY

We have seen that spending on prescription drugs is increasing rapidly and that drug firms have some monopoly power. Before we examine the exercise of that power, we turn to the role of prescription drugs in producing health and their relationship to other medical inputs using the concept of a health production function.

Recall that the patient’s demand for health leads to a demand for health inputs such as drugs and medical care. Consider the following production function for a patient with chronic and severe low-back pain—one of the most common reasons for physician office visits and hospitalization:

$$HS = f(D, M) \quad (17.1)$$

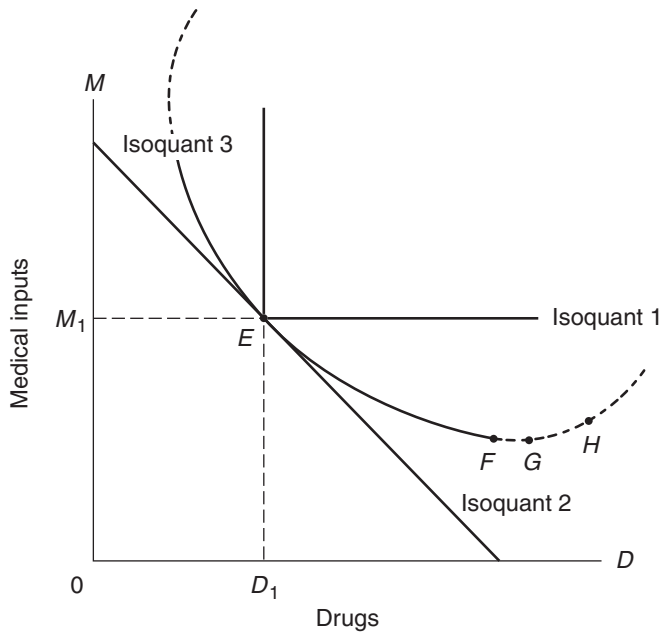
where  $HS$  represents the individual’s health status in the current period,  $D$  represents prescription drugs, and  $M$  represents all other medical inputs in this period, given existing technology and medical know-how. If no drugs or medical inputs are applied, the patient might experience considerable pain and be unable to perform many normal tasks, including work.

Assume that this patient’s health can be improved by medical intervention. Suppose an individual consumes the amount of drugs,  $D_1$ , and the amount of other medical inputs,  $M_1$ , as noted at point  $E$  of Figure 17-1. How do the drugs and the medical inputs substitute for each other?

Figure 17-1 shows three different effects of drug products and their relationship to other medical inputs. Isoquant 1 shows that drugs (e.g., narcotic analgesics or muscle relaxants) must be used in a fixed proportion to other inputs (e.g., physical therapy, counseling, and surgery in some cases). Here, inputs  $D$  and  $M$  are perfect complements with no substitutability between them.

At the other extreme, isoquant 2 reflects a production function where inputs are perfect substitutes: The marginal rate of technical substitution (MRTS) is constant, meaning that drugs substitute for the other inputs at a constant rate. (Depending on the prices of each, a patient would use one or the other, but not both.) Finally, the solid portion of isoquant 3 reflects an intermediate situation where  $D$  and  $M$  can substitute for each other, but where the MRTS is diminishing. Fewer and fewer amounts of  $M$  are needed to substitute for a unit increase in  $D$  as more drugs are used.

It seems unlikely that  $D$  and  $M$  are either perfect complements or perfect substitutes. Many conditions likely involve some substitutability, although the extent may vary widely among



**FIGURE 17-1** Substitution Between Drugs and Other Medical Inputs

conditions and even patients. Pharmacological and other clinical studies must determine not only substitutability, but also the “uneconomic” portions of an isoquant, meaning those combinations that never should be selected.

Such combinations arise when additions of *D* over some range (e.g., dashed segment *FG*) will not benefit or harm the patient. Here, the marginal product of *D* is zero and the isoquant becomes horizontal. It is also possible that increases in *D* beyond some point may harm the patient and require more medical intervention to maintain the same health status. In this region (arc *GH*), the isoquant will become positively sloped. Similar logic may apply to the vertical portion of an isoquant with increases in *M*. Patients will not want to be in the regions shown by the dashed segments.

**Least-Cost Production**

How will the patients and their providers choose? To abstract from the many possible levels of health, as well as the inherent uncertainty of medical practice, assume that the patient/provider believes that it is reasonable to attain *HS*<sub>1</sub> in Figure 17-2. The rational patient seeks to find the combination of *D* and *M* on *HS*<sub>1</sub> that minimizes spending. Without insurance coverage for either *D* or *M*, the total cost (*C*) of care can be written as:

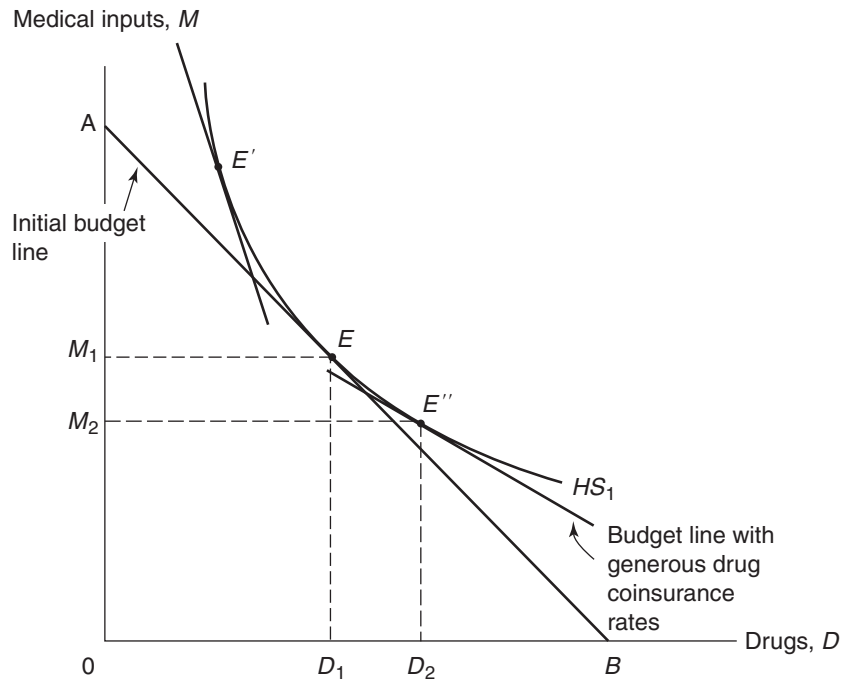
$$C = P_D D + P_M M$$

or

$$M = C/P_M - (P_D/P_M)D \tag{17.2}$$

where *D* and *M* are quantities of drugs and other inputs, and *P<sub>D</sub>* and *P<sub>M</sub>* are their respective prices. If, for example, *P<sub>D</sub>* = \$50 and *P<sub>M</sub>* = \$100, the slope of the budget line in equation (17.2) is  $-(50/100) = -0.5$ .

The cost-minimizing combination is at *E*, where the isoquant, *HS*<sub>1</sub>, is tangent to the budget line, *AB*. The optimal inputs are *D*<sub>1</sub> and *M*<sub>1</sub> and we calculate total spending by multiplying these



**FIGURE 17-2** Cost Minimization

quantities by their respective prices. At  $E$ , the numerical slope of the budget ( $P_D/P_M = 0.5$ ) equals the MRTS, the slope of the isoquant. Suppose, for example, that  $D_1 = 4$  and  $M_1 = 6$ , so total costs for  $HS_1$  equal:

$$\text{Costs} = (\$50 \times 4) + (\$100 \times 6) = \$800$$

If prescription prices increase above \$50, the budget line will become steeper and the rational patient will try to substitute more medical care by moving to the left on the isoquant to a point such as  $E'$  (and vice versa to  $E''$  if medical care prices increase).

### Insurance and Substitutability

Assume now that like most Americans the patient has insurance coverage. Begin with a policy that covers a constant proportion (e.g., 80 percent) of spending on either  $D$  or  $M$ . Out-of-pocket patient costs are \$10 for each prescription (20 percent of \$50) and \$20 for a medical visit (20 percent of \$100). Because the slope of the patient's budget line does not change, the optimal choice remains at point  $E$ , and the patient continues to buy four units of drugs and six visits. Total drug costs will continue to be minimized with the patient paying 20 percent (\$160) and the insurer paying 80 percent (\$640) of the \$800 total bill.

However,  $D$  and  $M$  often are not treated uniformly under traditional health insurance. Consider a policy that pays 80 percent of medical costs but requires a deductible of only \$5 (copayment) for each prescription. The patient's drug price is the \$5 deductible regardless of the actual price of the medication. If the patient's out-of-pocket drug costs diminish, the numerical slope of the cost-minimizing budget line diminishes (in our example, it is now  $-5/20 = 0.25$ ). The patient will have an incentive to substitute  $D$  for  $M$  at  $E''$ . Continuing with the example, let  $D_2$  increase from 4 to 5, and  $M_2$  fall from 6 to 5.75.

The patient's cost burden diminishes from \$160 to \$140 or:

$$(5.75 \text{ visits} \times \$100/\text{visit} \times 0.2 \text{ coinsurance rate}) + (5 \text{ units of drugs} \times \$5/\text{unit})$$

However, the total cost of care (patient plus insurer) increases from \$800 to \$825. We know this is true because we already determined that  $E$  is the least costly combination to provide  $HS_1$ .

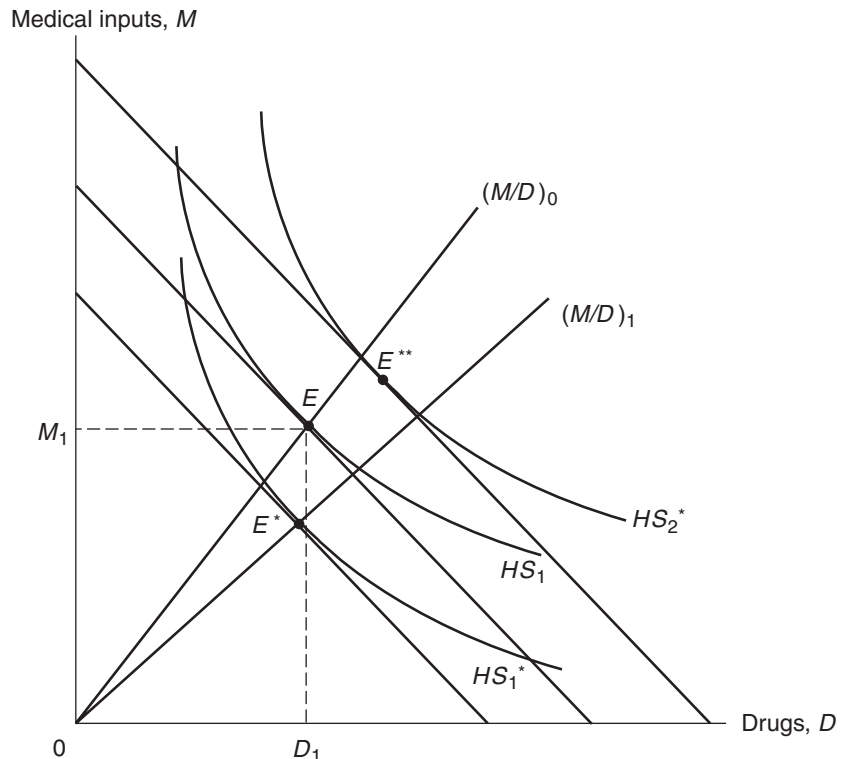
Similarly, if prescription prices (to the insurer) increase, say to \$100, the patient still pays \$5 and will remain at  $E''$  with the insurer picking up the increased drug costs. Patients have no incentive to economize by making substitutions and moving toward  $E$ . The higher the prescription price, the greater is the distortion.

A similar distortion toward excessive levels of  $M$  and greater total costs occurs when the patient's coverage excludes or limits drug benefits. Here, the savings from reducing  $M$  will more than offset the additional drug spending from improved drug coverage.

### Technological Change

At the turn of this twenty-first century, a new category of experimental drugs, blood vessel inhibitors, generated extraordinary excitement in the medical community by fighting both cancer and heart disease. Technology often is associated with major breakthroughs. More often, however, new drugs are similar to existing drugs, but they may produce somewhat better outcomes (if only for some patients) or reduced side effects.

With technological improvements, fewer inputs are needed to produce a given health outcome, or outcomes that were previously unattainable are now attainable. For example, begin with  $HS_1$  in Figure 17-3 and let  $HS_1^*$  represent all combinations of inputs with a new drug that leads to



**FIGURE 17-3** Technological Change



the same health status as  $HS_1$ . If the cost-minimizing ratio of inputs at a given price ratio remains unchanged, so that it lies along the ray  $OE$  (denoted  $(M/D)_0$ ), the innovation represents a neutral technological change. As drawn, the new drug saves a relatively high amount of the medical input at any given price ratio. That is, drug utilization increases relative to medical care as the patient moves to  $E^*$  and substitutes  $D$  for  $M$ . As noted in the figure, the lower ratio of  $M$  to  $D$  is reflected in the less steeply sloped ray,  $(M/D)_1$ .

New technology can increase costs for two reasons. First, it can routinely provide health levels that were unattainable previously (e.g.,  $HS_2^*$  at point  $E^{**}$ ). It may require much more drug use and possibly increased use of  $M$ , as well. When health improvements are dramatic or when drugs treat serious conditions that were not treatable previously, cost concerns are likely to be far less troublesome than those leading to only marginal improvements in health.

The second cost pressure comes from insurance. As we have seen, a patient with a constant copayment will not face any price increases of the new drug. Assume that the slope of the budget line in Figure 17-3 reflects the patient's copayment of \$5 and the patient moves from  $E$  to  $E^*$ . If the price of the new drug is high enough, the total cost of care at  $E^*$  could be substantially higher than total costs at the original equilibrium at  $E$ .

For conventional goods, where the consumer pays the entire price out of pocket, such technological changes will not be introduced because they will not be demanded. With insurance, the determination and elimination of cost-inefficient technology are far more difficult. A drug-maker may market a socially cost-inefficient drug successfully simply because it is more convenient for the patient to reduce other services and take more medications at  $E^*$ .

## DRUG PRICING AND PROFITS

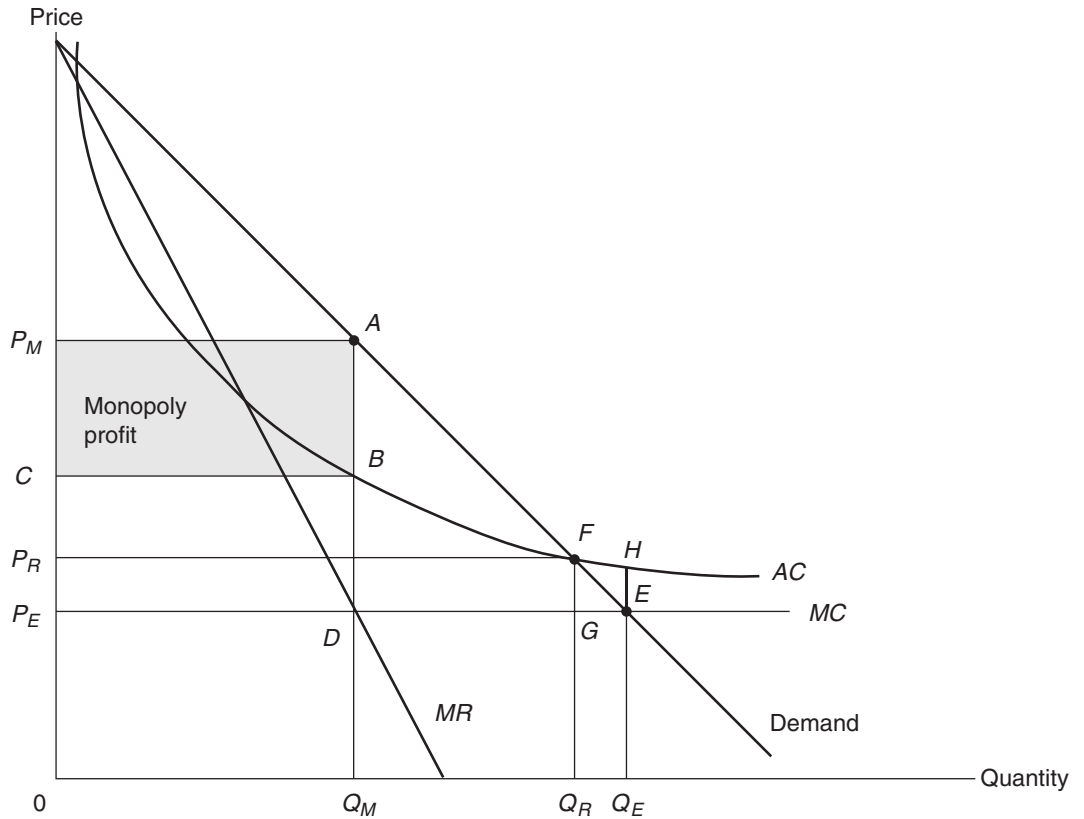
Drug pricing and profitability undoubtedly generate the strongest reactions among the public and the media. News reports paint pharmaceutical companies as exploiting patients through patents and other strategies that reduce competition. Many studies have found that pharmaceutical profits, as reported in financial statements, are consistently among the highest of all industries.

This dramatic difference often is attributed to monopoly power, but the profit picture is far less clear. Conventional accounting methods treat R&D and advertising and promotion as current expenses to be "expensed" even though, like physical investment, they provide returns in future years. Expensing can be thought of as an extreme form of accelerated depreciation where all of the "R&D capital" is used up in one year. It raises rates of return by reducing taxes. When Clarkson (1996) made adjustments to capitalize and depreciate these "investments," the industry's return remained higher than average but well below the adjusted returns for the highest industries.

Others argue that if drug R&D is riskier than other types of investments, it requires a higher rate of return to attract capital into the industry. Without trying to sort through all the measurement nuances, it seems reasonable to conclude that pharmaceutical firms earn above normal rates of return but that their profitability has been exaggerated by simple accounting comparisons.

### Monopoly Pricing

We begin with a firm selling a single product (or a composite of products) at a uniform price to all buyers. Figure 17-4 shows the demand and cost conditions facing the firm. Demand is a negatively sloped curve for several reasons. Even though patients with fixed copayments do not face higher out-of-pocket prices and will have a perfectly inelastic demand, others have more limited coverage or no drug coverage at all. Patients with limited (or no) coverage will likely substitute generic or over-the-counter products as a drug's price increases. Some may reduce utilization by not complying with the medication regimen. The drug supplier also must consider purchasing decisions by managed care organizations and other large buyers, such as hospitals, which can be sensitive to price changes.



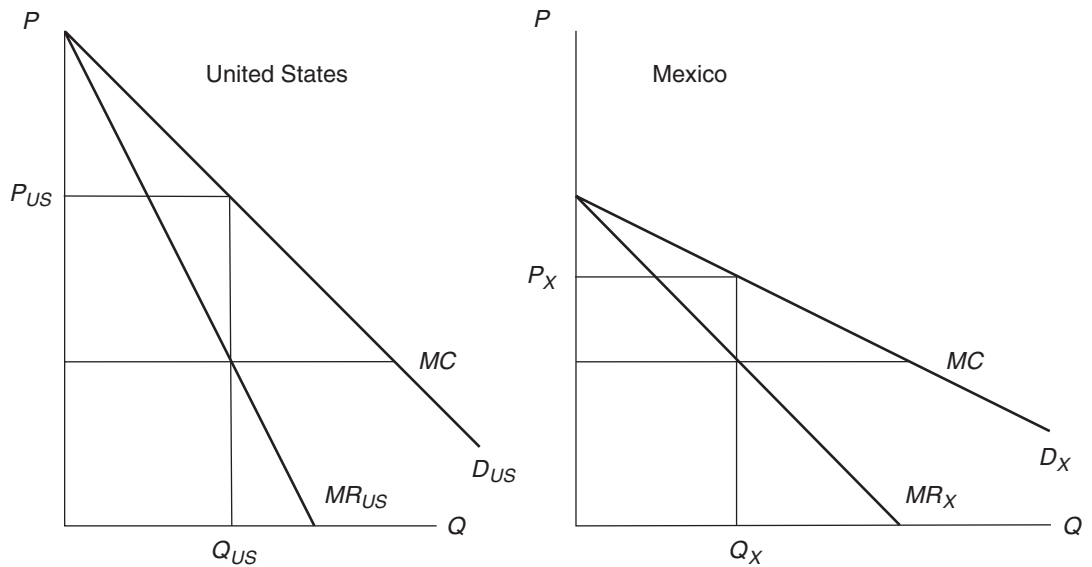
**FIGURE 17-4** Drug Pricing

On the cost side, the marginal cost of manufacturing and distributing the product is usually relatively low—about half the total cost. R&D and various promotion costs are substantial. In Figure 17-4, we show the marginal cost ( $MC$ ) as constant and the average cost ( $AC$ ), due to high fixed costs of R&D and promotion, as a downward-sloping curve. The profit-maximizing output occurs where  $MC$  equals  $MR$ , resulting in  $P_M$  and  $Q_M$ . Assuming that the drug supplier earns economic profits, the price must lie above average cost. The shaded rectangle  $P_MABC$  shows economic profits. Consistent with the hypothesized demand and cost structures, the gap between price and the low marginal cost will be large.

The profit-maximizing model also predicts that the difference between price and marginal cost varies inversely with the elasticity of demand. Lu and Comanor (1998) examined pricing decisions on new products, and their findings support profit-maximization. Initial (launch) prices are considerably higher for products that represent large, therapeutic gains than prices for new “me-too” drugs that are similar to available products. Why? Demand will be relatively inelastic for a product that provides significant benefits as compared to other products. Launch prices are also much higher when few branded substitutes exist. This factor similarly reduces a product’s demand elasticity and requires a higher price for profit maximization.

### Price Discrimination

A firm may be able to increase profits beyond the level described in Figure 17-4. One of the most interesting features of the pharmaceutical industry is third-degree price discrimination (also known as market segmentation) where different groups of buyers are charged different prices. Before the



**FIGURE 17-5** Price Discrimination

enactment of Medicare Part D drug benefit legislation, the media routinely ran news stories about American seniors flocking to Mexico, where prices are much lower, to buy drugs. Hospitals or managed care groups are often charged less than retail pharmacies, and prices for drugs used in veterinary medicine can be much lower than prices for similar products packaged for human use.

What accounts for the wide variations in price? One explanation is straightforward. If a firm can distinguish between markets with different demand characteristics, and can also limit arbitrage (third-party resale at lower prices in higher-priced markets), it can increase profits by charging different prices. Assume, for simplicity, that the firm described in Figure 17-4 sells only in the United States and Mexico. Figure 17-5 separates the total demand into the U.S. and Mexican demands. With higher incomes and better insurance, the demand is relatively inelastic in the United States. Assume further that the marginal costs of production and distribution remain constant and are equal in both countries, and that prices are not regulated in either market.

Profit maximization occurs where  $MR$  equals  $MC$  in each market, resulting in quantities  $Q_{US}$  and  $Q_X$ . Even though marginal revenue will be equal in the United States and Mexico, the price is higher in the market with the less-elastic demand (United States).<sup>4</sup> Total profits must be greater than those obtained under uniform pricing.

### Monopsony Pricing and Price Controls

Price discrimination is not the only possible explanation for price differentials. Prices in some foreign countries can be lower because their governments regulate prices or their national health plan serves as a monopsony buyer. Continuing with Figure 17-5, suppose that the Mexican government

<sup>4</sup> A convenient formula for profit maximization is:

$$P = \frac{MC}{\left[1 + \frac{1}{E_p}\right]}$$

where  $E_p$  is the algebraic value of the price elasticity of demand. With the same  $MC$  in both markets, price must be higher when demand is less elastic (e.g.,  $-2$  versus  $-4$ ).

imposes price controls. Conceptually, it can drive price as low as the marginal cost, further increasing the price differential with the United States. Critics charge that by failing to control prices in a similar manner, consumers in the United States bear the burden of the development costs and that the United States subsidizes other countries.

Proposals have been made to limit prices in the United States or even to treat pharmaceutical firms as regulated utilities. Return to Figure 17-4. It is theoretically possible to reduce price to  $P_R$ , raising quantity to  $Q_R$ , and enabling the firm to just cover its costs and earn a normal rate of return. However, the administrative complexity of regulating prices of multiproduct firms that are continuously introducing new products is enormous. Consider just the variations of a single product available in different strengths, forms (tablet, liquid), and delivery systems (oral, intravenous, patch, inhaler). In addition, Abbott (1995) found that pharmaceutical firms often set much higher introductory prices under regulation.

The complexity and potential perverse effects of regulation prompt many analysts to conclude that cost containment is better left to private initiatives. They also worry about the damaging effects of price or rate-of-return controls on innovation. We cover both topics later in this chapter.

### Competition and Generic Entry

Once a patent expires, other firms can enter the market. Entry barriers are considerably lower than those for new product development, and the 1984 Drug Price Competition and Patent Restoration Act further eased requirements. The act replaced the safety and efficacy testing under the 1962 amendments with much less costly bioequivalence tests. If the generic is approved, the FDA certifies it as “therapeutically equivalent” to the branded version. The new legislation has greatly increased generic applications, and as one would expect, firms target those markets with the greatest opportunities, in particular large markets and those where drugs treat chronic conditions (Bae, 1997).

What happens to prices and market shares after generic entry? Wiggins and Maness (2004) estimated an 83 percent drop in prices of anti-infectives (e.g., penicillins, tetracyclines) as the number of sellers increases from 1 to between 6 and 15, with further drops in price as more firms enter the market. This conventional finding on the impact of entry runs counter to a more complex story that had been developed for pharmaceutical pricing. Previously, Grabowski and Vernon (1992) examined 18 drugs that first experienced generic competition after the 1984 act. Generics captured one-half of their markets within two years. Surprisingly, though, as generic prices were falling, brand producers were raising theirs and widening the price gap over time. This phenomenon, corroborated by Frank and Salkever (1997), suggested that generics were not viewed as close substitutes by some patients or their providers. Pioneer firms can retain some monopoly power by capitalizing on the brand loyalty and relatively inelastic demand of this group. That is, as generics siphon off price-sensitive patients, the price-insensitive ones are left. The pioneer takes advantage of this market segmentation by raising brand-name prices for its loyal customers. (Pioneers can even introduce their own generic versions to compete in the generic segment.) We will return to the demand for generic substitutes in the section on cost containment.

## RESEARCH AND DEVELOPMENT (R&D) AND INNOVATION

Estimates of the drug industry’s spending on R&D vary widely (Golec and Vernon, 2007), but there is no doubting the large amounts. Domestic R&D expenditures for members of the Pharmaceutical Research & Manufacturers of America (research-based pharmaceutical firms) rose from just over \$1.5 billion in 1980 to \$37.4 billion in 2010 (PhRMA, 2011), with another \$12 billion spent abroad. Between 1980 and 1994, their share of domestic sales devoted to domestic R&D increased from 13.1 to 21.9 percent, stabilizing since then in the range of 18–21 percent. Firms would not undertake these investments unless they could gain protection for their work. The patent system provides one method for providing protection.

A patent gives the holder the right to “to exclude others from making, using, or selling the invention.” The usual term for a patent is 20 years, but there are exceptions for drug products through exclusive rights granted by the FDA. To offset partially the delays posed by the long testing and regulatory review period, the Drug Price Competition and Patent Restoration Act of 1984 allows extensions for up to 5 years so long as the total effective patent life does not exceed 14 years. In 1983, Congress also passed the Orphan Drug Act, permitting extensions (and providing other benefits) for drugs designed to treat rare conditions that might otherwise not be profitable.<sup>5</sup> The exclusive right granted to KV Pharmaceutical, described in Box 17-1, falls under this Act.

On the one hand, patents and other legal protections, such as trademarks and copyrights, can lead to monopoly power, which is not in the public interest. On the other hand, a firm would be much less willing to expend millions of dollars on research if others can become free riders by mimicking its innovations. By being first, firms still would invest in R&D but at reduced levels. The question of just how much less is clearly an empirical issue.

Mansfield (1986) estimated that 60 percent of pharmaceutical drugs between 1981 and 1983 would not have been developed without patent protection. This figure is especially dramatic compared to the 11 other industries he sampled. The chemical industry was the only other industry with a substantial impact (30 percent).

Although patent protection has relatively small impacts on most industries, it is critical to pharmaceutical innovation. We, therefore, turn our attention to the determinants of pharmaceutical R&D and a conceptual framework to examine the effects of FDA regulations and patent law on innovations.

### Investment Decisions

Net present value analysis provides a simple yet powerful approach to investment decisions. Letting  $R_t$  and  $C_t$  represent the revenues and costs in time,  $t$ , the net present value  $NPV$  of a project is given by:

$$NPV = \sum_{t=1}^{t=T} (R_t - C_t)/(1 + r)^t \quad (17.3)$$

where  $r$  is the discount rate or cost of capital and  $T$  is the life of the project. Following some of the discussion in the appendix to Chapter 4, under the standard decision rule, a project is accepted if the net present value is positive.

Several characteristics of pharmaceutical R&D become apparent if we break  $NPV$  into three components representing:

- a. The research, testing, and review period ( $m$  years)
- b. The effective period of patent protection ( $n$  years) after the product is launched
- c. The period following patent expiration ( $s$  years, where  $m + n + s = T$ )

$$NPV = \sum_{t=1}^{t=m} (R_t - C_t)/(1 + r)^t + \sum_{t=m+1}^{t=m+n} (R_t - C_t)/(1 + r)^t + \sum_{t=m+n+1}^{t=m+n+s} (R_t - C_t)/(1 + r)^t \quad (17.4)$$

a.
b.
c.

In the first component, **a**, the firm will not have any revenue and there will be large, negative net cash flows reflecting the high R&D costs.

<sup>5</sup> Schweitzer (1996) describes the incentives provided by the act and how they have led to a large number of products given orphan status. He argues, however, that the act does little to help populations in developing countries who suffer from diseases that are rare in the developed world. Together with the high cost of many existing drugs, the responsibility of industrialized nations and the pharmaceutical industry to poorer nations is becoming a major issue. Under a proposal to encourage the development of vaccines against diseases concentrated in low-income developing countries, the developed countries would commit to purchasing vaccines from pharmaceutical companies.

To offset these costs, a successful project will require even larger positive net cash flows in later years, particularly over the period represented by the second component, **b**, the effective patent life. The ability to charge high prices and/or reach large potential markets will have strong positive effects on *NPV*. Despite potential competition from generics and significant erosion of sales, products still may capitalize on brand recognition, marketing efforts, and new uses to remain successful after patent expiration, the last component, **c**.

This framework further tells us that regulations and testing procedures that increase costs in the first component reduce *NPV* and make an investment less attractive. Similarly, *NPV* is reduced by an increase in the length of the research, testing, and review period, because it must reduce the length of the patent protection period **b**. Conversely, changes such as reduced regulation or fast-track laws to lower initial costs and speed up the review process as well as extensions of patent rights each serve to increase *NPV*.

Finally, the risks are important. Projects with higher risks should be discounted at a higher rate or, put another way, high-risk projects need a high rate of return to be viable. To the extent that a firm can reduce risks, for example, by supporting a portfolio of diverse projects or sharing risks through joint ventures with other firms, the discount rate,  $r$ , diminishes and the likelihood of investment is increased. This also suggests that large firms have an advantage over smaller firms in R&D.

## R&D Spending

We have noted the substantial total industry spending on R&D. Despite the spending, and a substantial development pipeline, the FDA approved only 21 new medicines in 2010, the fewest since the 19 approvals in 2007 (which itself was the lowest total since 1983).<sup>6</sup>

Firm-level analyses of R&D provide some startling figures on costs and their recent growth rates. Focusing on the more significant innovations, DiMasi and colleagues (1991) estimated total costs, computed as capitalized expected costs and discounted at 9 percent, at \$231 million in 1987 dollars per new chemical entity that was marketed. Because there is substantial attrition as projects move to successive stages of development, about two-thirds of the cost is attributable to the preclinical phase. In a controversial update covering the late 1990s, DiMasi, Hansen, and Grabowski (2003) estimated average out-of-pocket R&D costs for new chemical entities at \$403 million, in year 2000 dollars. This figure reaches \$802 million when capitalized at 11 percent.<sup>7</sup>

Although an accompanying editorial by Frank (2003) supported the study's high quality, DiMasi's findings were attacked even prior to their formal publication. Two former editors of the prestigious *New England Journal of Medicine* (Relman and Angell, 2002) raised serious questions about the innovativeness of the pharmaceutical industry and many of its marketing practices. They further argued that DiMasi misrepresents R&D costs for new drugs, in that new chemical entities account for a minority of newly approved drugs. More recently, Light and Warburton (2005, 2011) have voiced concern about biases and other limitations of the proprietary and confidential survey data used by DiMasi. The accuracy and consistency of such data cannot be independently verified—an important caveat if one believes that pharmaceutical firms have an incentive to overstate development costs.

How do firms recover these formidable amounts? Our understanding of the returns to investment has been aided greatly by the work of Grabowski and Vernon (1994, 1996), which follows sales over the life cycle of a product. Grabowski and Vernon found that a product has an

<sup>6</sup> As reported in an analysis undertaken by Bloomberg News, December 30, 2010, [bloomberg.com/news/2010-12-30/new-drug-approvals-fell-in-2010-as-safety-concerns-slow-u-s-fda-decisions.html](http://bloomberg.com/news/2010-12-30/new-drug-approvals-fell-in-2010-as-safety-concerns-slow-u-s-fda-decisions.html).

<sup>7</sup> See also DiMasi and colleagues (2005) and the Tufts Center for the Study of Drug Development ([http://csdd.tufts.edu/news/complete\\_story/internal\\_news](http://csdd.tufts.edu/news/complete_story/internal_news)—accessed April 21, 2011) for responses to these concerns. The industry itself reports the 2005 cost per drug at \$1.3 billion (PhRMA, 2011).

effective patent life of about 9 to 13 years and a market life of about 20 years. Cash flows do not become positive until the third year after launch, and sales peak in the tenth or eleventh year. The most significant finding is that a substantial portion of a company's revenue and profits come from a few big winners. Only the top 20 percent of new drugs have substantially positive *NPV*; the *NPV* of the representative new drug is actually negative. The highly skewed distribution of returns to investment reinforces the firm's need to diversify by having a large number of drugs under development to reduce risks. It also suggests difficulties for smaller firms that cannot sustain large R&D programs.

### **Firm Size and Innovation**

Henderson and Cockburn (1996) ask: "Are the research efforts of larger firms more productive than those of smaller rivals, and if so, why?" They search for evidence of the effects of size on "important" patents granted and find that the returns to size are significant. This result is not surprising, but their contribution lies in distinguishing between "economies of scale" and "economies of scope." Recall that a firm experiences economies of scale when its long-run average costs decrease with higher output. It has economies of scope if the cost of producing two or more different products is less than the costs of separate production. Earlier work, which did not distinguish between these effects, showed mixed results on the effects of firm size on innovation.

For innovation, these concepts need clarification in that output is not easily or uniquely measured. One approach measures output through patents. Economies of scale arise if patents granted increase more rapidly than the size of an R&D program. Economies of scope occur if a joint program to develop different drugs is more productive than distinct programs.

Henderson and Cockburn found that size is important and that larger programs are more productive. However, it is due more to economies of scope than to economies of scale. There is little gain in increasing the size of an individual program beyond some minimal threshold level. Economies of scope arise when different research activities can share inputs, thus lowering cost. More important, spillovers of knowledge among projects raise overall productivity. The authors also found significant spillovers of information among firms.

The industry also is responding in other ways to the growing cost of R&D and conducting clinical trials. There have been a large number of mergers and acquisitions but this activity may be motivated by the desire to eliminate excess capacity, rather than decisions to exploit potential R&D economies (Danzon et al., 2007). Strategic alliances have risen sharply. The alliances, often between pharmaceutical and biotech firms, seek to pool efforts to innovate or bring products to the market more successfully. As part of this process, less pharmaceutical R&D is being conducted "in house" as firms look for R&D partners to reduce costs and spread risks. These partners include contract research organizations that are used to conduct clinical trials and, in some cases, to provide a broad range of drug-development services.

### **Prices, Price Regulation, and Innovation**

Arguably, no issue is more important to drug policy than the effects of prices on innovation and, by implication, the effects of drug price regulation on innovation and the availability of drugs. Price regulation is often proposed as a means of limiting expenditures on drugs and, as we will describe, many other countries have adopted price controls or other forms of regulation. Should the United States rely largely on markets, even if imperfectly competitive, to determine drug prices and R&D activity, or is there an important policy role for price controls?

The theoretical framework represented by equation (17.4) suggests that higher drug prices and larger potential markets should spur R&D and consequently the rate of innovation. Research provides strong evidence supporting these predictions. Vernon (2005) estimates that a price control policy that would lower pre-tax pharmaceutical profit margins to the average of those in non-U.S. markets would

lower industry R&D investment by between 23 and 33 percent. Other studies find a consistent and substantial direct relationship between higher real drug prices and increased innovation.<sup>8</sup>

In analyses of drug launches, Danzon and colleagues (2005) investigated the number of launches and launch delays for 85 important new drugs in 25 countries over the period 1994–1998. The United States led all countries with 73 launches (Japan was lowest with 13). Higher expected prices and greater market size increase the number of launches and reduce launch delays.

The literature is clear on the adverse effects of price regulation on R&D investment, innovation, access to new drugs, and delays in availability. Are there benefits from regulation, such as increased access due to lower prices and reductions in expenditures, which could offset these adverse effects? Or, better yet, are there structural changes to the drug industry that could promote marginal cost pricing while maintaining high rates of innovation? There is considerable ongoing effort to answer these important questions.

## COST CONTAINMENT

The rapid growth in drug expenditures has led to great policy interest in cost containment. President Clinton's proposed 1993 health care reform plan included a mechanism to regulate prices through caps geared to prices in other countries and to producer costs. Despite questionable methods, a series of government reports in the early 1990s, indicating that U.S. prices were higher than prices in Canada and the United Kingdom, intensified interest in drug price controls.

We already have addressed some of the difficulties of regulating prices for large numbers of constantly changing products. We also have addressed the problems of recovering common costs. If prices are driven toward marginal costs, R&D investment and the resulting innovation and access to new drugs may decrease, to the detriment of the public's welfare. Many countries have introduced various forms of regulation to rein in spending. Western European countries tend to control either producers' prices or reimbursement rates while England has a profit control system. In the United States, Medicaid programs "discount" prices and impose restrictions on utilization. The federal government also discounts prices for drugs purchased by the Department of Veterans Affairs and other federal agencies.

U.S. consumers purchase most outpatient drugs either out of pocket or with private insurance (Medicare's coverage for outpatient prescriptions started only in 2006). We, therefore, turn to the cost-containment efforts of the private sector, especially those introduced through managed care. These efforts include price discounting and the exercise of monopsony power, much like their public insurance counterparts. To narrow our discussion, we will describe three other strategies: higher copayments (often through multiple tiers of cost sharing), use of generic drugs, and the adoption of drug formularies.

### Copayments

A higher copayment seems simple and straightforward, intended to shift a larger share of the cost burden to the patient and to decrease consumption of marginally beneficial drugs. Moreover, under many current copayment plans, prescription drugs cost only a small fraction of their over-the-counter substitutes.

However, a higher copayment may produce other results. Recall from Figure 17-2 that an increase in the out-of-pocket cost of a drug will decrease its use as patients substitute other inputs for drugs.

With no substitutability, use of drugs remains unchanged; only the distribution of the costs between patients and insurers is affected. As we noted earlier in the chapter, if significant substitutability among treatments is possible, the total health care cost could increase as inefficient cost-minimization combinations are selected.

---

<sup>8</sup> The elasticity of innovation with respect to drug prices is about 0.6 (Giacotto, Santerre, and Vernon, 2005, and Lichtenberg, 2007a).



However, when generic (multiple-source drugs) substitutes are available, strategies involving differential copayments for brand products offer the potential for substantial shifts toward lower-priced generics without affecting overall utilization. Many managed care plans now have three or more tiers of cost sharing in which the patient pays, for example, \$7 for a generic drug, \$15 for a brand-name product on an approved list (formulary), and more if it is not found on the formulary.<sup>9</sup> Patients and providers who consider the generic a close substitute will choose the generic.

The key question is whether there will be a large substitution toward generics. Motheral and Henderson (1999) examined two plans with tiered systems that increased brand-name copayments more than copayments for generics. They found little effect on total drug utilization. However, utilization of brand-name products decreased about 18 percent relative to a control group that had no price increases. This substitution produced substantial savings to the health plans.

Other evidence for a wider set of drug benefit arrangements is even more revealing. From a sample of over 400,000 working-age adults, Joyce and colleagues (2002) found that doubling copayments for all drugs from \$5 to \$10 reduced average drug spending by 22 percent, and by about one-third when copayments were doubled in two-tier plans. Adding a second \$20 tier for brand-name drugs that previously had a \$10 copayment for all drug purchases reduced spending by 19 percent. Drug spending was also reduced by 8 percent in two-tier plans that mandated generic substitution as compared to those that did not mandate such substitution. Numerous analyses of the Medicare Part D plan (discussed in Chapter 21) show that 100 percent copayments in the so-called “doughnut hole” lead to substitution of generic drugs for brand name drugs, and on occasion for the patient to reduce frequency or even stop taking the drugs.

Finally we caution that higher copayments and other cost-sharing schemes not only reduce costs, but also change treatment. Goldman et al. (2004) found substantial decreases in utilization within the most common drug classes from a doubling of copayments. Reductions ranged from a low of 25 percent for antidiabetics to highs of 44 percent for antihistamines and 45 percent for non-steroidal anti-inflammatory drugs (NSAIDs). Utilization for those with chronic illnesses was less responsive to the copay increases. Nevertheless, the authors were concerned about the health effects of the large reductions, especially for patients with diabetes.

The cost-sharing literature reinforces this concern. Gibson and colleagues (2005) concluded that these arrangements generally work as intended—by encouraging generic use and limiting over-use. But their study also found reports that higher cost sharing can also disrupt treatment through lower levels of adherence, lower use of essential medicines, and, in some cases, drug discontinuation.

Philipson and colleagues (2010) highlight the potential for such problems by studying patients with acute coronary syndrome who underwent stent implantation. Patients with high cost-sharing were less likely take antiplatelet drugs (a highly effective therapy) following the implantation and more likely to discontinue their use within the first year. As a result, these patients experienced worse outcomes and had higher total costs due to increased re-hospitalization.

### Generic Substitutes

With the expiration of patents on some important drugs and the cost-containment efforts made by many plans, about 67 percent of the prescriptions written in 2007 were filled with generic drugs—from 50 percent in 2001. The percentage of generics for multiple-source products was likely to be much higher. Many top-selling drugs now have generics. With generic prices considerably below their brand-name counterparts and FDA certification of their therapeutic equivalence, one would expect little resistance to generic versions from physicians and their patients. This has not always been so.

---

<sup>9</sup> Many Medicare Part D plans, and some commercial plans, have introduced four-tiered systems in which the fourth tier is reserved for certain expensive drugs including biologic drugs used to treat cancer. The fourth tier is handled as coinsurance; for example, the patient pays 25 percent of the prescription’s cost with no limits, in some plans, on the out-of-pocket amounts. See also Box 8-2.

Interest in promoting generic products has a long history. Most states passed ant substitution laws after World War II, prohibiting pharmacists from substituting a generic for a prescribed brand, but mounting cost pressures led to reforms. In 1970, Massachusetts became the first state to legalize drug product selection and, by the end of the decade, most others followed. Despite the potential for cost saving, an early study of Michigan's substitution laws was revealing (Goldberg et al., 1979). After 1974, substitution by a pharmacist was allowed unless the doctor wrote "dispense as written" or "DAW," but physicians wrote relatively few prescriptions with this restriction. However, pharmacists provided substitutions for less than 2 percent of all multiple-source prescriptions.

Substitution has increased well beyond the levels of the 1970s, but stood at just 19 percent in 1984, when the Hatch-Waxman Act of 1984 passed. The legislation allows a generic producer to file an abbreviated new drug application and to use the pioneer company's clinical research. Since then, efforts by managed care and other third-party payers have greatly increased generics' share of the prescription drug market.

### **Drug Formularies**

Managed care's strong financial interest in cost containment has led to policies that go well beyond copayment strategies to promote generics. Many plans monitor physicians and require substitution when generics are available. Many also use pharmacy benefit managers to negotiate discounts and improve the efficiency of their claims-processing and pharmacy operations. They are increasingly adopting other methods such as drug-utilization review programs and lower-cost, mail-order sources for prescriptions. One of the most ambitious and controversial strategies, however, involves the use of formulary committees to develop a list of approved drugs. A positive formulary restricts the choice of drugs to those on the list. A negative formulary excludes drugs on a list.

The formulary review and approval process can be elaborate, dealing not only with generic substitution but also with recommendations of different drugs to treat a condition. Ideally, the review committee will periodically review all drugs in each therapeutic class for their clinical effectiveness, safety, and cost. Mather (1999) observes that, when properly implemented, the formulary can be an effective and well-accepted tool. If drug-product decisions are based largely on cost instead of clinical outcomes, Mather suggests that "the health system may experience higher overall costs and the pharmacy benefit may be sharply criticized by health plan providers, enrollees, and suppliers wishing to see their products on the list" (p. 277).

The widespread adoption of formularies has elevated the importance of pharmacoeconomic analyses. Pharmaceutical firms face strong pressure to provide clinical and economic data that justify their inclusion in the formulary. Formulary committee members require the skills to compare, interpret, and analyze data from many different sources and to justify their decisions. They are under especially great pressure from both patients and providers when new and improved drugs come onto the market.

### **NEW DRUGS AND HEALTH CARE SPENDING**

It might appear from the material presented that drug spending in the United States has gone beyond any level that analysts might consider reasonable or efficient. If so, then the United States is not alone among developed countries. As measured by the percent of health spending devoted to drugs (see Table 22-1), the United States actually ranks below many other industrial countries (though it is the biggest spender in absolute dollars). Many factors account for the wide variations in shares across countries in addition to price differences. These include differences in physician practice patterns, cultural attitudes toward drugs, licensure laws, and insurance and other financial arrangements (physicians are permitted to dispense drugs in some countries). All nations are dealing with the same pressures and the need to find mechanisms to promote cost-efficient use of drugs and other health care inputs. After all, the most important economic question is the value of drugs in producing health.

Fortunately, economists have addressed this question. Frech and Miller (1999) took advantage of the wide differences in drug utilization and health status across 21 OECD countries to investigate the productivity of drugs, other medical inputs, wealth, and lifestyle. They used regression analysis with alternative measures of a nation's health status as the dependent variable. Independent variables included a country's wealth, pharmaceutical spending per capita, other health spending per capita, and various lifestyle indicators.

Some of the findings are consistent with our analyses in previous chapters. Wealth, measured by GDP per capita, showed significant and positive effects on life expectancy. A doubling of GDP increases life expectancy by 6 percent effect at age 40 and by 9 percent at age 60. Lifestyle (measured by smoking, alcohol, and consumption of animal fat) also has important effects. In particular, the consumption of animal fat, the most important lifestyle variable, has positive effects on life expectancy at birth, age 40, and age 60 up to certain levels of consumption (92, 78, and 70 grams per day, respectively). Nonpharmaceutical inputs showed no statistically significant effects at any of the three age levels.

The most interesting and surprising results relate to pharmaceutical spending per capita, with strong positive effects on life expectancy. A doubling of drug spending increases life expectancy by 2 percent at age 40 and 4 percent at age 60. The study also estimated the lifetime pharmaceutical cost of extending life by one year. For the United States, these costs for males and females at age 40 are \$21,000 and \$23,000, and about \$17,000 and \$19,000 at age 60. Analyses of infant mortality show that drug consumption has a small, positive (harmful), but statistically insignificant effect. The consumption of animal fat, with a negative (beneficial) impact, is the most important variable.

More recent evidence on the value of new drug development is also very informative. For the U.S. population, Lichtenberg's (2007b) analysis of individual medical conditions shows that newer drugs are well worth the extra cost. Reducing the age of drugs used in treatment reduces nondrug spending by a factor of 7.2 (8.3 for the Medicare population) relative to the increased drug spending. The bulk of the savings (i.e., offset effects) come from reduced hospitalization. Lichtenberg (2008) also examined data for 20 OECD countries specifically for cardiovascular disease. Cardiovascular disease is the leading cause of hospitalization and death in many nations. The adoption of newer drugs reduced the cardiovascular age-adjusted mortality rate and number of hospital days (through lower hospitalization rates and lengths of stay). The per capita savings on hospital stays (\$89 in 2004) from adoption of the new drugs was 3.7 times as large as the increase in per capita spending for cardiovascular drugs (\$24).<sup>10</sup>

## CONCLUSIONS

Our overview of the pharmaceutical industry focused on individual provider and patient decisions, firm pricing and investment decisions, and public policy issues in regulation and cost containment. Throughout, we have emphasized the use of basic economic tools to simplify and make sense of complex problems.

Pharmaceutical companies and their profits are highly visible and they make convenient targets. However, newer drugs appear to lower mortality and produce net reductions in health care spending. Careful research thus cautions against ill-conceived proposals to limit the availability of drugs and to contain drug spending.

This caveat is especially important in light of the politics surrounding drug policy. Many supporters of health care reform in the United States favored a bill that would have allowed the Department of Health and Human Services to negotiate the prices of all drugs purchased by Medicare beneficiaries. In exchange for the industry's support of reform, and to gain support from a

---

<sup>10</sup> Civan and Köksal (2010) similarly found that newer drugs reduce total health care spending with the largest reduction occurring for hospital care. However, Law and Grépin (2010) describe some serious biases in the methodology used by Lichtenberg and others. The controversy over the offset effects of new drugs is not a completely settled issue.

major organization representing the elderly, Congress dropped any provision from the legislation that President Obama ultimately signed in 2010. Instead, the legislation improves Medicare prescription drug coverage beginning in 2011 with a 50 percent discount to seniors for brand name drugs purchased in the coverage gap (the “doughnut hole”). The doughnut hole itself will shrink and disappear by 2020.<sup>11</sup> Thus, while the drug industry made some concessions, it escaped more serious threats to its profitability. Nevertheless, opponents of “big pharma” remain committed to limit the industry’s power and curtail some of its practices.

---

## Summary

1. In 2009, prescription drug spending accounted for 10.1 percent of national health care spending, up from 8.8 percent in 2000 and 4.7 percent in 1980.
2. The pharmaceutical industry is characterized by significant barriers to entry and substantial regulation. The industry has been regulated since 1906, but 1962 legislation had the most profound effects.
3. Pharmaceuticals substitute for health and non-health inputs in the production of health. A rational patient will select least-cost input combinations. Relative changes in a patient’s out-of-pocket costs resulting from insurance will lead to substitution of drugs for other health care inputs or vice versa.
4. Technological change, even when the benefits are marginal, is often cost-increasing due to insurance.
5. Pharmaceutical firms earn higher-than-normal profits, but the extent of their profits is exaggerated by conventional accounting data.
6. Opportunities are substantial for price discrimination. Markets, or groups of buyers, with inelastic demand will pay higher prices.
7. Generic products often capture a significant share of the market following patent expiration. Nevertheless, through effective marketing and promotion, trade names still can retain a monopoly premium.
8. R&D spending is substantial, and a firm’s profits often depend on a few big winners. Firms need to earn substantial amounts over the effective life of a patent to justify the risks.
9. Larger R&D programs are more productive than smaller programs. Economies of scope play an important role.
10. Price regulation reduces investment in R&D, the rate of innovation, and the number of drug launches.
11. Price regulation, rate-of-return regulation, and national formularies are used in other countries to contain costs. In the United States, higher copayments, the promotion of generic substitutes, and other managed-care strategies, such as drug formularies, have been adopted.
12. Higher copayments encourage substitution toward generics as well as a decrease in overall utilization. In some cases, they may also lead to disruptions in treatment.
13. Spending on drugs in the United States is not out of line compared to other developed countries.
14. Drugs are highly productive in improving health compared to other medical inputs. New drugs reduce health care costs, largely through reduced hospitalization, relative to the additional cost of these drugs.

---

## Discussion Questions

1. Explain and distinguish between the “concentration ratio” and the “HHI.” What are the limitations of these measures within the context of the pharmaceutical industry?
2. Direct-to-consumer (DTC) advertising of prescription drugs has grown rapidly. List several products with which you have become familiar as a result of such advertising. Discuss the pros and cons of DTC advertising from the perspective of physicians and patients.
3. Use Figure 17-1 to explain how an isoquant can be positively sloped. Under what circumstances may a patient actually end up in the positively sloped region (e.g., at point *H*)?

---

<sup>11</sup> There are other provisions affecting the drug industry such as the imposition of an annual fee on manufacturers and importers of branded drugs, and expanded power for the FDA. The health care reform legislation is more fully described in Chapter 23.

4. Regulation is often proposed (and widely used in other countries) to limit prices or profits. Discuss possible adverse effects of regulating prescription prices. In light of your discussion, what accounts for the strong pressure in many countries, including the United States, to regulate prices?
5. What are barriers to entry? Describe three potential barriers in the pharmaceutical industry. What are some consequences of these barriers?
6. In 2004, Congressman Dennis Kucinich proposed the Free Market Drug Act. This legislation would have removed patent protection on drugs developed with public funds and given control over pharmaceutical R&D to the National Institutes of Health (NIH). Evaluate this type of proposal in terms of the effects on price, competition, and level of innovation.
7. There are wide differences across countries in the share of health resources spent on drugs. Describe possible economic and noneconomic factors that may contribute to the variation.
8. Direct-to-consumer advertising has been criticized for possibly misleading patients and for increasing spending on drugs. Discuss the benefits and costs of DTC advertising.
9. Media reports often show much higher drug prices in the United States than in other countries. Analyses by economists often show that the price differential is not as large. Describe some possible problems in comparing domestic with foreign drug prices.

## Exercises

1. Use Figure 17-2 to explain why cost minimization through a tangency between an isoquant and a budget line does not apply in cases where  $D$  and  $M$  are either perfect complements or perfect substitutes. Describe the cost-minimization process and the effects of changes in the relative prices of  $D$  and  $M$  in these special cases.
2. Assume that a patient has 80 percent coverage for medical services but no coverage for prescription drugs. An 80 percent drug benefit is added. Show graphically what will happen to the relative utilization of  $M$  and  $D$ , and total spending on health care, to attain a given health status. What happens to the amount spent on drugs (insurance plus patient amounts)? Why will total spending on health care diminish when the 80 percent drug benefit is added?
3. A pharmaceutical firm faces the following monthly demands in the U.S. and Mexican markets for one of its patented drugs:
 
$$Q_{US} = 300,000 - 5,000P_{US} \text{ and } Q_X = 240,000 - 8,000P_X$$
 where quantities represent the number of prescriptions. Assume that resale or arbitrage among markets is impossible and that marginal cost is constant at \$2 per prescription in both markets. Monthly fixed costs are \$1 million in the United States and \$500,000 in Mexico.
  - (a) Draw the demand, marginal revenue, and marginal cost curves for each market. Estimate the profit-maximizing prices and quantities graphically and/or determine the solutions algebraically. What are the firm's total profits?
  - (b) Determine the quantity in each market and maximum possible total profits if the firm engages in perfect (first degree) price discrimination. Is this behavior possible?
4. Assume that the firm in Exercise 3 cannot prevent resale and is forced to set the same price in both markets. Find the price graphically and/or algebraically and show that total profits are less than those from part 3a.
5. For your answer in 3a:
  - (a) Calculate the price elasticity of demand in each market at the optimal price.
  - (b) Verify that the prices and elasticities are consistent with the profit-maximizing formula given in footnote 4.
  - (c) Why are both elasticities fairly close to unity? (Hint: Think about the requirement for profit maximization when marginal cost is zero.)
  - (d) If a firm finds that its price elasticity is numerically less than one, what advice would you have?
6. Consider only the U.S. market from Exercise 3. Graph solutions to parts 6a and 6b using the demand, average cost, and marginal cost curves. Also try to develop the answers algebraically.
  - (a) Price regulation is proposed. Find the regulated price that enables the firm to cover all its costs. Caution! There are two mathematical solutions. Which one will regulators prefer?
  - (b) Find also the economically efficient price (i.e., one that is consistent with marginal cost pricing). What subsidy per prescription is required to enable the firm to cover all its costs?
7. Compare your results in Exercises 6a and 6b with the profit-maximizing solution for the United States obtained in Exercise 3a. Explain which of the three alternatives you would prefer if you were responsible for public policy. Be sure to consider some of the problems of regulating prices.
8. Many insurance companies increased premiums by 15 percent or even more in the early years after 2000, blaming soaring pharmaceutical costs for their premium increases. Evaluate the validity of this justification. (Hint: Consider spending on pharmaceuticals as a share of national health expenditures.)

# Equity, Efficiency, and Need

- Efficiency and Competitive Markets
- Deviations from the Competitive Model in the Health Care Sector
- Promoting Competition in the Health Care Sector
- An Economic Efficiency Rationale for Social Health Insurance
- Need and Need-Based Distributions
- Horizontal Equity and Need
- Theories of Social Justice
- Conclusions

Are health care markets healthy? Are they efficient, and do they provide the care that people need? These questions occupy the minds of the many people who study health care. Efficiency questions arise because of the high costs that people must pay for health care. Are these costs too high? Likewise, equity questions occur because many people, certainly including the uninsured, face barriers in obtaining health care.

Compared to Canadians and Europeans, Americans will more likely find experts who favor competitive market solutions to health care system problems, though most here (as elsewhere) will argue that this approach is often ill-suited to the nature of health care markets. Canadians and Europeans are more willing to use government interventions, although the success of the Patient Protection and Affordable Care Act (PPACA) in 2010 suggests that the United States may have moved in that direction, albeit with some substantial political opposition. Similarly, Medicare reform remains centrally important, and many Americans find equity and efficiency in the health care system to be the fundamental issues.

A solid background in these issues requires a study of the economics of efficiency, the departures of many health care markets from the competitive model, the role of equity concerns, and issues of social justice theory. Because of the central role of “need” in health equity discussions, we must also investigate the meaning of health care need. These subjects are the themes of the present chapter.

The chapter focuses on welfare economics, the study of normative issues that bear on economics. “Normative issues” deal with how people believe the economic world *should be*, as opposed to “positive issues” that deal with how the world of economics functions *in practice*.

So, welfare economics would encompass those that are critical of existing markets and question the distribution of goods and services. This understanding, however, is itself disputed by some health economists (Hurley, 2000; Culyer, 1989) who argue that an “extra-welfarist” viewpoint is required, rejecting some or all of the philosophical principles on which welfare economics is based. Yet other theorists find within welfare economics the tools to understand the concepts of welfare and efficiency that concern us the most (Absolo and Tsuchiya, 2004).

While we focus on standard welfare economic theory, we will explain sources so students can explore the extra-welfarist view more fully. We first describe the standard results for competitive markets, but also the many market flaws that cause markets to deviate from competition, causing many competitive efficiency propositions to fail. We will also explain and describe the role of need and need-based distributions in the health economy. Finally, we will present theories of social justice and explain why welfare economic claims must be grounded in a philosophical position on justice.<sup>1</sup>

## EFFICIENCY AND COMPETITIVE MARKETS

We clarify the meaning of economic efficiency within the context of the Edgeworth box for exchange. This approach derives theorems in a graphical framework that theorists have also developed in more sophisticated mathematical models. The analysis here generates the First Fundamental Theorem of Welfare Economics, and illustrates the Second Fundamental Theorem as well. The First Theorem demonstrates that competitive markets under certain conditions are economically efficient. The Second Theorem establishes that a society can achieve any desired economically efficient outcome by competitive markets if it starts from the appropriate initial endowments.

### The Concept of Pareto Efficiency (Optimality)

A century ago, the economist Vilfredo Pareto defined the concept of efficiency most frequently used by economists today. According to Pareto, an economically efficient (optimal) outcome in society is one under which it is impossible to improve the lot of any person without hurting someone else. Pareto efficiency also implies that no further exchanges would be found that could improve the lot of everyone to some degree. An efficient economy necessarily would have exhausted all means for mutual gains.

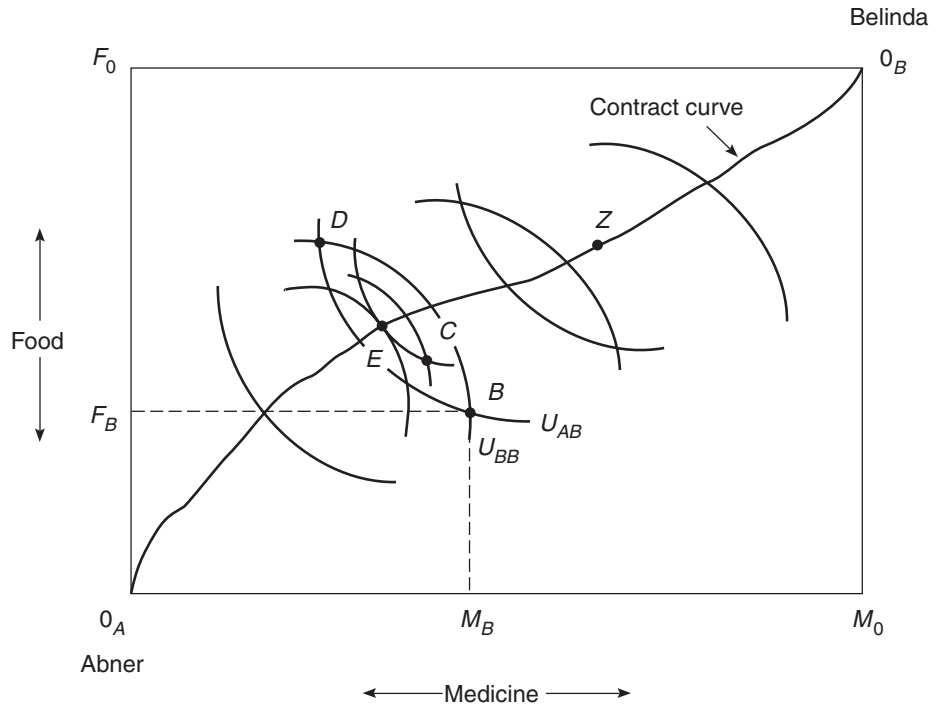
The Edgeworth box, using a hypothetical two-person economy and showing exchanges between these two people, provides a context in which to make the idea of Pareto efficiency clear. The box also is convenient for describing the mutual gains from trade and for defining the Pareto concept of efficiency.

Suppose that persons A and B, say Abner and Belinda, inhabit a desert island, forming a two-person economy. Further suppose that only two goods are available on the island. Food,  $F$ , is gathered and is available in a fixed total amount,  $F_0$ , and medicine,  $M$ , is likewise available in a fixed amount,  $M_0$ .

To form the Edgeworth box, consider Figure 18-1. Abner’s preference map (indifference curves) starts from the southwest corner. There is no reason to draw the axes out further than  $M_0$  and  $F_0$ , which represent the total amounts of medicine and food available on the island. Belinda’s preference map is similar to Abner’s except that it starts at the northeast corner. It is also constrained by amounts  $M_0$  and  $F_0$ .

Any point in the box represents a complete and exhaustive distribution of the island’s endowment of food and medicine. For example, point  $B$  represents a distribution in which Abner has  $M_B$  units of medicine and Belinda has  $M_0 - M_B$  units of medicine. Similarly, at  $B$ , Abner has  $F_B$  units of food and Belinda has  $F_0 - F_B$  units of food. With this orientation, we ask whether point  $B$  is an economically efficient distribution.

<sup>1</sup> For an excellent alternative exposition of welfare economic issues, see Williams and Cookson (2000).



**FIGURE 18-1** Edgeworth Box for Exchange; Pareto Efficient Points Lie on the Contract Curve

The answer to this question must be no. To see this, examine by comparison point  $C$ . Point  $C$  lies on an indifference curve that is above (to the northeast of) indifference curve  $U_{AB}$  and, therefore,  $C$  is superior to  $B$  in Abner's view. Similarly, point  $C$  lies on an indifference curve that is above (to the southwest of) indifference curve  $U_{BB}$  and, therefore,  $C$  is superior to  $B$  from Belinda's view. Because point  $C$  is attainable and improves the lot of both persons while harming neither, it follows that the original point  $B$  is not economically efficient.

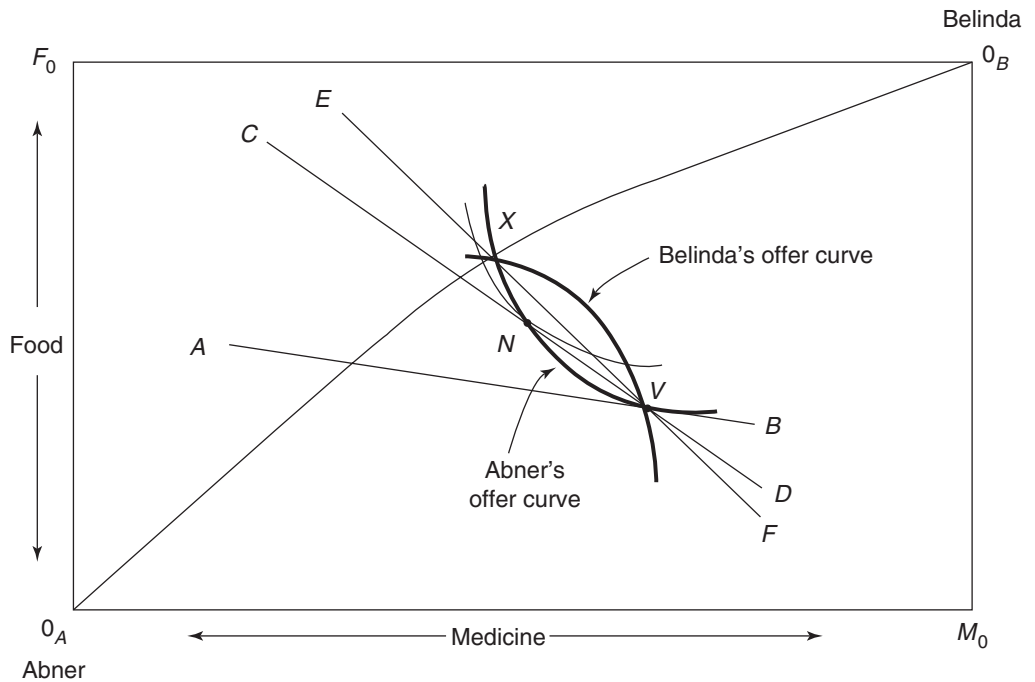
Geometrically, we can repeat the analysis regarding point  $B$  for any point that forms a "lens" from the indifference curves passing through it. A lens is formed by the indifference curves  $U_{AB}$  and  $U_{BB}$  from point  $B$  to point  $D$ . Whenever we can find such a lens, we can identify one or more other points superior to the initial point. Reapplying this reasoning, point  $C$  is also not Pareto efficient. Pareto-superior moves, where the welfare of both improves, can also be made from point  $C$ . In contrast, a Pareto-efficient point in the box is a point of tangency between two indifference curves, such as point  $E$ . It is impossible to move from a point of tangency without harming the lot of one of the two persons.

Each of Abner's indifference curves will have a point of tangency with one of Belinda's indifference curves. We call the collection of all Pareto efficient points in the box the *contract curve*, which is so labeled in the figure. For example, at point  $O_A$ , Belinda has all of both goods, and even if many or most people consider this inequitable, it *is* Pareto efficient because giving any of either good to Abner would make Belinda worse off.

### Trading Along the Budget Line

Having defined efficiency in the context of the Edgeworth box, we next ask whether the competitive market generates an efficient equilibrium in exchange. In a competitive market, each person treats prices as given and responds to prices by choosing the utility-maximizing bundle subject to his or her resource constraint. The resource constraint depends on the person's initial endowment





**FIGURE 18-2** The Intersection of Offer Curves Determines the Competitive Equilibrium

of food and medical care. Let point  $V$  in Figure 18-2 represent the initial endowment for this two-person economy. Either person may trade away from his or her initial endowment at the market prices. Thus, Abner's resource constraint will be represented by a budget line passing through point  $V$ . As with any budget line, the slope of this line is the negative of the ratio of the price of medical care to food.

The slope of the budget line represents the rate at which one can trade one good for another at market prices. The steeper the budget line is in Figure 18-2, the greater the price is of medical care relative to food. For example, budget line  $AB$  represents a relatively lower price of medicine relative to food than does budget line  $CD$ .

### The Competitive Equilibrium

To find the competitive equilibrium, we must identify how much each person would be willing to trade. Abner's offer curve, for example, is the collection of points representing his offer for trade at each possible set of prices. Start at point  $V$ . Given budget line  $AB$ , Abner stays at point  $V$ , the point of tangency between budget line  $AB$  and the highest indifference curve that is attainable.

Suppose the price of medical care were higher relative to food, leading to the steeper budget line  $CD$ . Given budget line  $CD$ , Abner would trade some medicine for some food to go from point  $V$  to point  $N$ . With budget line  $EF$ , Abner would trade to point  $X$ . Connecting all such points generates Abner's offer curve.

The figure also shows Belinda's offer curve, beginning at endowment point  $V$ . The two heavily shaded offer curves represent voluntary trades for the two parties. For trade, as in a competitive market, to be mutually voluntary, the offers of the two persons must agree. The offer curves agree only at their point of intersection, labeled point  $X$  in the figure. Point  $X$  thus constitutes the competitive market equilibrium in exchange for this two-person economy, starting with the endowment of  $V$ .

### The First Fundamental Theorem of Welfare Economics

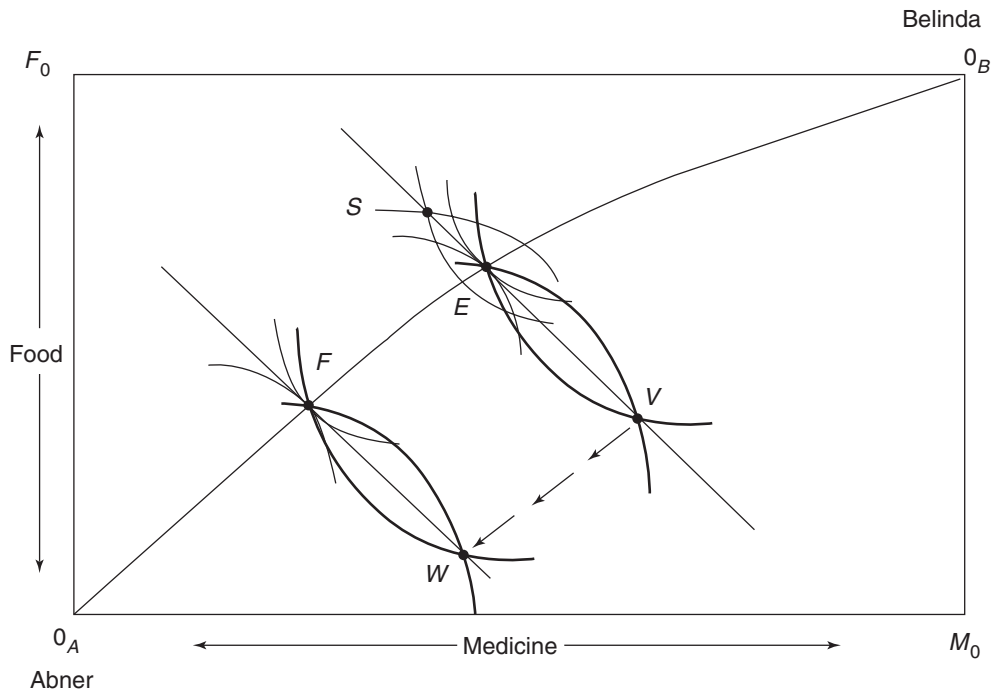
Is the competitive equilibrium,  $X$ , Pareto efficient? Yes, and there are two reasons that it must be so for every competitive equilibrium. The intersection of two offer curves represents a trade made at competitive prices starting at point  $V$ . Each person is at a point of tangency between the budget line and the highest attainable indifference curve. At point  $X$ , Abner's indifference curve (not shown) is tangent to the budget line. Likewise, at point  $X$ , Belinda's indifference curve (also not shown) is tangent to the budget line. Because these indifference curves are tangent to the same budget line at the same point, they must be tangent to each other. Because they are tangent to each other at point  $X$ , this point is Pareto efficient. The same argument applies for any competitive equilibrium; therefore, we have shown the First Fundamental Theorem in this context, namely that the perfectly competitive market equilibrium is Pareto efficient.

The theorem makes the competitive market solution attractive. If we can establish perfect competition, then the market forces left to their own workings will generate an efficient outcome—an invisible-hand solution. However, the theorem evokes several serious questions: Can we achieve competitive markets in health care? Is the context of this theorem appropriate for health care? Would the competitive market solution be equitable or would it leave too many people without adequate health care? We will address each of these questions. However, we begin this process by exploring the issue of equity within the context of the Second Fundamental Theorem.

### Redistribution of the Endowment

We extend the applicability of the First Fundamental Theorem with the Second Fundamental Theorem, which states that given an appropriate endowment, any Pareto efficient outcome can in principle be achieved by a competitive market. Figure 18-3 illustrates the significance of this theorem.

In Figure 18-3, suppose that the initial endowment is  $V$ , and suppose that this endowment results in the competitive outcome represented by point  $E$ . Point  $E$  is only one of an infinite number



**FIGURE 18-3** A Preferred Efficient Outcome May Be Achieved by an Initial Redistribution of Resources

of Pareto efficient points. It may be an outcome that many view as inequitable, here either Abner or Belinda, or both. In real life, the society may have millions of members, and plausibly a majority of people may perceive this market outcome to be inequitable.

The Second Theorem, however, defines a central role for competitive markets, even in cases where many view some of the competitive outcomes as inequitable. Suppose, in Figure 18-3, that society prefers outcomes in the vicinity of point  $F$  to outcome  $E$ . By the Second Theorem, a competitive market can achieve the desired outcome, but it requires a different initial endowment from point  $V$ . As shown, the endowment point  $W$  is a suitable point from which to achieve an equitable market outcome, point  $F$ . Seen this way, redistribution combined with competitive markets generates an efficient and equitable outcome. This contrasts with command systems that reject free markets, as well as with alternative schemes, such as price discrimination.

### Price Discrimination

Some propose achieving a more equitable outcome by providing certain services to the poor at reduced, subsidized prices. Readers may be surprised to learn that such systems are not consistent with Pareto efficiency. Consider the proof of the efficiency of competitive markets. It was crucial that both parties achieve a point of tangency to the same budget line. If the poor are charged different prices than the rich, the two groups face different slopes of their budget lines. The result would be a position such as point  $S$  in Figure 18-3, a point that is not Pareto efficient.

For an intuitive argument, a program subsidizes the poor in purchasing bread. The poor will adapt to the subsidized price until the rate at which they were willing to trade bread for other goods equals the rate at which they could exchange the goods at the subsidized price. The result is that the poor will undervalue bread in comparison to the wealthy. It will be more efficient for the poor to buy up bread and sell it to the rich. Such a side market, which would improve efficiency in the bread example, is not possible for many forms of medical care (treatment for broken legs, for example), which are not easily transferrable. Thus, subsidized prices for medical care will likely generate an inefficient equilibrium.

The two theorems, along with the inefficiency of price discrimination, suggest the superiority of income transfers as a solution to equity problems in health care markets. In Figure 18-3, the following situation takes place: Transferring initial resources between the two persons and then allowing the market to work, will achieve an efficient outcome within the equitable range.

### Trade-offs Between Equity and Efficiency

The theoretical superiority of redistribution of income to programs, such as price subsidies, has led many analysts to favor income maintenance programs as policy tools to offset the problems of poverty, including the problems of access to health care. Income maintenance programs are government programs designed to provide cash subsidies to the poor to maintain their incomes at or above a preset floor. Despite continuing interest in such programs, policy makers often have hesitated to use large-scale income redistribution.

Economists explain a major criticism of income maintenance by appealing to Okun's (1975) analogy of the leaky bucket. The act of transferring wealth from one group to another in society may generate disincentives that discourage productive effort. The taxpaying group incurs a tax burden that may reduce work incentives, and the recipient group receives subsidies that may reduce incentives to work and to self-help. By analogy, when we transfer income, our task is similar to transferring water in a leaky bucket. The amount of income available for redistribution may decline as a result.

Blank (2002) challenges the equity–efficiency trade-off idea, arguing that situations exist where the efficiency costs of improving equity may be very small, such as when the group receiving the benefit is unlikely to change its behaviors. She further posits that in some cases equity and efficiency are complementary.

In the 1960s and 1970s, the federal government sponsored large-scale experiments to investigate the degree of work loss induced by the incentives inherent in income maintenance programs. These experiments reported reductions in work effort on average of between 5 and 10 percent. However, the work reduction estimates were considerably higher for certain subgroups, such as “male nonheads (of families)” and women. Also, results generated in an experimental situation make it difficult to predict the results if the program were to become universal and permanent.

## **DEVIATIONS FROM THE COMPETITIVE MODEL IN THE HEALTH CARE SECTOR**

Another major criticism of the applicability of our theoretical analysis concerns the question of whether health care markets are sufficiently competitive or whether we can make them sufficiently competitive to obtain competitive outcomes. Substantial differences exist between most health care markets and the theoretical model of competition. For the results to hold, several assumptions must apply.

### **The Assumptions Under Perfect Competition**

The First and Second Theorems apply to competitive markets. To be perfectly competitive, a market must have free entry and exit, perfect information, a homogeneous product, and numerous buyers and sellers each with no power over price. Furthermore, we derive the theoretical efficiency of competitive markets under conditions where no significant externalities, public goods, or natural monopolies exist. Finally, the actors in the competitive markets are alternatively consumers maximizing their utility, or producers maximizing their profits.

Many have criticized the application of the theorems to the health care sector, claiming that health care markets are typically not perfectly competitive. Health economists have recognized most of these criticisms as having validity. The health care markets depart from competition in several ways:

1. Barriers to entry exist in health care markets. Such barriers include licensure laws and health planning controls on prices and facility construction.
2. There are often few enough firms that those in the market have some degree of monopoly power.
3. Health care services are not uniform in quality or other characteristics.
4. Motivations other than pure profit are common in health care.
5. The model depicts the operation of markets under conditions of certainty. However, health events entail a considerable degree of uncertainty.
6. Information problems exist.
7. Externalities are prevalent in health care.

Several of the seven listed deviations need little further explanation. However, we consider three for extended discussion: the role of uncertainty, the role of information, and the role of externalities.

**THE ROLE OF UNCERTAINTY** The uncertain nature of health status gives rise to the demand for insurance coverage among persons who are risk-averse. In the present context, insurance creates problems for the efficient functioning of health care markets. We note four issues:

1. Insurance changes the price of care to the insured person, which in turn leads to the distortions described under price discrimination.
2. Insurance causes the price paid to suppliers to differ from the price paid by the consumer, and this distorts the efficient matching of production to consumption.
3. Large insurance companies and government programs negotiate payment rates, thus removing price determination, at least in part, from the market.
4. In some health care markets, insurance coverage is so complete as to distort the health care producer’s incentives to be efficient.

**THE ROLE OF INFORMATION** The efficiency results for competitive markets depend on all parties having complete information available. As we have shown elsewhere, it is particularly problematic for markets to function when information is imperfect and asymmetrically available to the parties in the market. Potential problems of information and efficiency arise either when the physician has much more information about the appropriateness and effectiveness of treatments and techniques than does the consumer, or when the consumer has more knowledge of his or her health status and health habits than does the potential insurer.

**THE ROLE OF EXTERNALITIES** Finally, health care markets may involve externalities. A prominent externality will occur whenever participants in the market are significantly concerned about the health care received by others, not just about their own health care. This externality may be difficult to internalize in private charity markets, and it arguably causes health care markets to be inefficient. Because some analysts have identified externalities as the most important efficiency argument for social insurance programs in health care, we develop an extended discussion later in this chapter.

## **PROMOTING COMPETITION IN THE HEALTH CARE SECTOR**

If we could manipulate real-world markets as easily as we can change the assumptions of theory, then it would follow from our theoretical discussion that we should promote competition in health care markets whenever possible. Often the promotion of competitive elements in health care markets will prove useful. However, further theoretical grounds exist to qualify our statements.

### **The Theorem of the Second Best**

One qualification involves the Theorem of the Second Best in welfare economics. Consider an economy with more than one departure from the conditions of perfect competition. Consider further any policy that corrects one or more of these departures from perfect competition but does not correct all of them. The Theorem of the Second Best shows that such a policy may not necessarily improve society's welfare.

An intuitive understanding of why this result comes by considering a market with a pure monopolist (a departure from the conditions of perfect competition) who is also a polluter (a departure from the conditions under which competition is efficient). Basic theory shows that a monopolist will produce less output than would a competitive industry under otherwise similar conditions. A policy that hypothetically converts the industry to perfect competition would resolve one discrepancy but not both because both output and pollution would increase. Societal valuations of the extra output versus the extra pollution could in principle determine whether the change worsened society's well-being. Thus, correcting some economic "wrongs," but not all of them, may not necessarily improve welfare.

This classic example of the monopolist polluter illustrates the idea of the theorem, but it does not make clear its applicability to the health sector. Consider a somewhat more controversial health-related example. Laws requiring that physicians go to medical school and pass additional exams grant licensed physicians a degree of monopoly power, a distortion from the competitive conditions. At the same time, however, health consumers have imperfect information on therapies and prices and less information than the physicians. This, too, is a departure from competitive conditions. If one eliminated the imperfection caused by licensure but did not simultaneously address the information problem, leaving patients less informed, patient welfare could decrease. This could happen because, without licensure restrictions on physicians, poorly informed consumers could be fooled by quacks or by possibly dangerous treatments. This example also illustrates the Theory of the Second Best.

It would be a misapplication of the Theorem of the Second Best to conclude that all health care policies that increase health care market competition are incorrect. More properly, the theorem

states that we cannot assume competitive policies will improve welfare. We necessarily operate in the world of second best because it will be impossible to convert all health care markets into the model of perfect competition. A competitive policy may improve the functioning of health care markets in a manner that improves society's well-being. Each policy must be considered on its own merits, not solely on the grounds that it promotes competition.

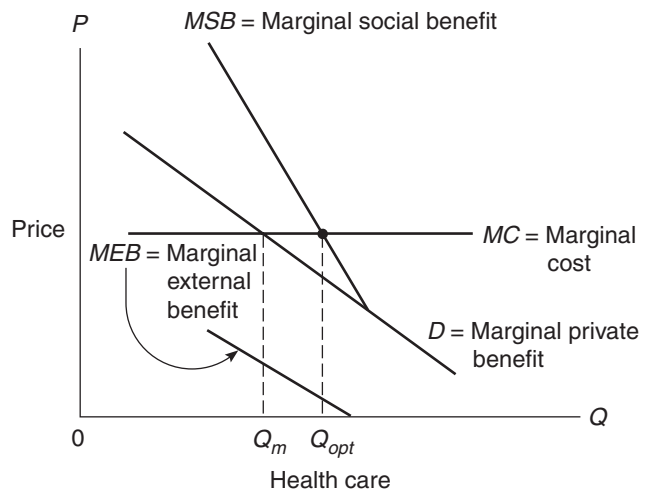
A number of health economists not only point out reasons why health care markets do not qualify as competitive, but they also criticize the very assumptions that underlie the efficiency claims of welfare economics. We will discuss these "extra-welfarist" claims in the section on Need and Need-Based Distribution later in this chapter.

## AN ECONOMIC EFFICIENCY RATIONALE FOR SOCIAL HEALTH INSURANCE

An externality occurs when someone external to the market transaction—that is, someone who is neither the buyer nor the seller—is affected directly by the transaction and does not receive compensation. A common example in health care occurs in the case of immunization for contagious diseases. Here, people outside the market transaction—those not presently immunized—benefit from the immunization because the immunized person will less likely become a carrier of the disease and threaten their health. This situation is an example of a beneficial consumption externality.

In the presence of a beneficial externality, the competitive market will tend to produce an inefficiently low level of output. Within a single market, the Pareto efficiency definition leads to the condition that marginal benefit equals the marginal cost in equilibrium. Individuals in a well-functioning, perfectly competitive market in theory will use medical care until the marginal benefits, measured through the demand curve, equal marginal costs, which in equilibrium will equal the price. In Figure 18-4, this leads to an efficient level of consumption,  $Q_m$  in the absence of externalities.

A marginal external benefit to people in society must be added to the marginal private benefit, which is measured by the demand curve, leading to the marginal social benefit. In Figure 18-4, the marginal external benefit curve is *MEB*. The marginal benefit to society as a whole is the vertical sum of the *MEB* curve and the demand curve. The result is the marginal social benefit curve, *MSB*. Efficiency for society occurs at output level  $Q_{opt}$ , whereas the market would achieve an inefficiently low level of output,  $Q_m$ . Thus, on efficiency grounds alone, society may be justified in subsidizing immunizations.



**FIGURE 18-4** The Socially Efficient Equilibrium in the Presence of a Beneficial Externality

Although immunization for contagious diseases illustrates the logic and role of beneficial consumption externalities in justifying subsidies for (or possibly public provision of) care, such as the U.S. polio immunizations of the 1950s and 1960s or modern-day immunizations in less-developed countries, it represents a fairly minor problem and could not in itself be used to justify large social insurance programs. However, an alternative health care externality, one that we have identified elsewhere as a charitable externality, can in principle be sufficiently important to justify such programs.

This externality would occur, for example, whenever people feel that some segment of society is receiving insufficient care in the sense that the charitably minded person would be willing to pay to help these people get care. Willingness to pay means here that they would pay if contribution would help the poor to acquire health care. Such charitable feelings are probably widespread in most societies. As Pauly (1971) argued:

The desire to eliminate the diseconomy that the presence of curable but uncured disease or injury may exert on others does appear, in general terms, to be a common characteristic of human beings. At least at some levels, most of us would be willing to give up some of our income to help a suffering fellow. Some may, of course, be immune to such feelings, but individuals may also be immune to contagious disease, and this should cause no insurmountable theoretical problems. (pp. 10–11)

This charitable externality has a different source than the case of immunization and is probably more important, but the economic argument is of the same form. In Figure 18-4, consider the curve *MEB* to measure this externality. Then, as before, the efficient level of output,  $Q_{opt}$ , exceeds the market output,  $Q_m$ . Under certain circumstances, this efficiency may be sufficient grounds to intervene, for example, by providing a program of social insurance.

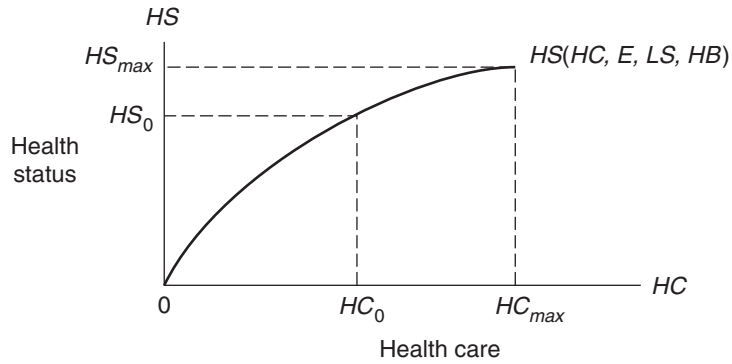
## NEED AND NEED-BASED DISTRIBUTIONS

Even if we can identify accurately the efficient allocation of health care, we nevertheless may find many people in society dissatisfied with the outcome because many people will not get the health care they need. That is, there will be additional concerns, over and above efficiency concerns, regarding equity. In the health care literature, the concerns for equity most often center on the question of whether people are getting the health care they need. Unfortunately health care need is often either undefined or variously defined. The definitions employed may take either extreme, either maximally or minimally. For example, Culyer and Wagstaff (1993) review several definitions of health care need; they conclude that a maximal need definition is superior, defining need as:

the expenditure required to effect the maximum possible health improvement or, equivalently, the expenditure required to reduce the individual's capacity to benefit to zero. (p. 436)

In contrast, some analysts or policy makers treat health care need as a minimal requirement or standard of adequacy. Federal health planning efforts in the 1970s sought to control the perceived proliferation of health care in order to control costs. Health planners at the time, as well as most health cost control advocates in any era, argued implicitly that consumers were getting more health care than they really needed.

Often the discussion of needs gets disconnected from the fact that the output and distribution of health care to meet people's needs are chosen in the context of society's choices of all its public goals. We present a construction of need that brings this to mind. Let health care needs be defined within the context of the choice of society's goals for population health status as well as in the contexts of other goals, such as education and defense. To illustrate, we identify in Figure 18-5 the production function for health defined over the levels of a variable input, health care, given the



**FIGURE 18-5** The Production Function for Health

conditions of environment,  $E$ ; lifestyle,  $LS$ ; and human biological endowment,  $HB$ . Here the technically maximal health status is  $HS_{max}$ . Achieving this health status level, requires a health care level of  $HC_{max}$ . However, society may choose a lesser health status goal than the maximum achievable, using the savings to further other goals. For example, if society through its choice-processes selects health status goal  $HS_0$ , then health care level  $HC_0$  is needed.<sup>2</sup>

The choice of a health goal implies a needed level of health care, that is, a level of health care “utilization.” In the early 1990s, economists debated whether utilization or “access” was the superior choice for defining needs. The words *equity of access*, which frequently appear in public documents in various countries, suggest an equal opportunity, especially a financial opportunity (Mooney et al., 1992). Many health economists, however, find *access* difficult to define. While health care utilization is more easily measured, it is usually also the ultimate reason for our concern about access (Culyer, van Doorslaer, and Wagstaff, 1992a, 1992b).

### Health Care Needs and the Social Welfare Function

How, then, is the health status goal selected? We depict this choice using the concept of a social welfare function.

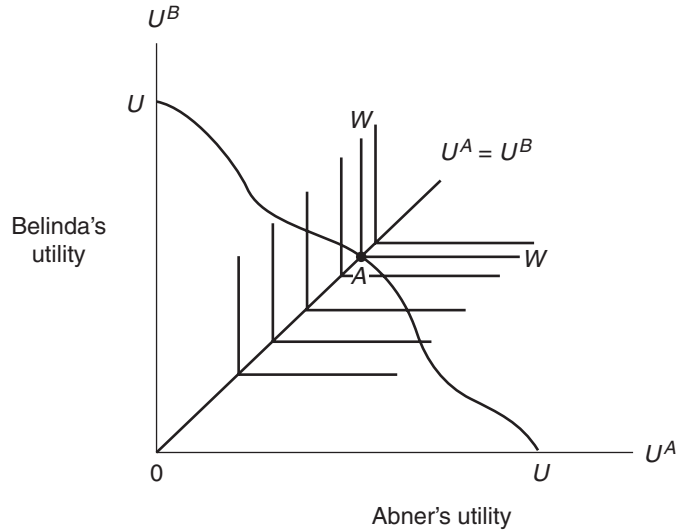
**THE UTILITY-POSSIBILITY FRONTIER** The Edgeworth box shows the efficient choices available to society in allocating resources among people. Yet it is also apparent that at many points on the contract curve, Abner or Belinda get few or no resources. By the Pareto criteria, these points are unquestionably economically efficient, yet they may be indefensible within any definition of a humane society.

Economic theory suggests that a social welfare function reflecting society’s overall preferences is necessary to determine which of the efficient points to choose. We can trace out a utility-possibility frontier,  $UU$ , from the points in the Edgeworth box in Figure 18-1. Begin at the allocation where Belinda has everything and Abner has nothing. In Figure 18-6, we can assign Abner a utility of zero and evaluate Belinda’s utility as the intercept of the vertical axis.

We can then draw a  $UU$  curve by reallocating resources to Abner from Belinda as we move along the contract curve in Figure 18-1. Recall that the fundamental property of Pareto efficient distributions is that Abner’s utility can come only at the expense of Belinda’s utility. Therefore, the  $UU$  curve in Figure 18-6 must be downward sloping. The horizontal axis intercept summarizes the point at which Abner has all of the resources.

<sup>2</sup> The exposition of need focuses on the consequences. Hurley (2000) describes this interpretation as strongly “consequentialist” and further describes alternative views.





**FIGURE 18-6** The Utility–Possibility Frontier and Social Welfare Maximization: Egalitarian Preferences

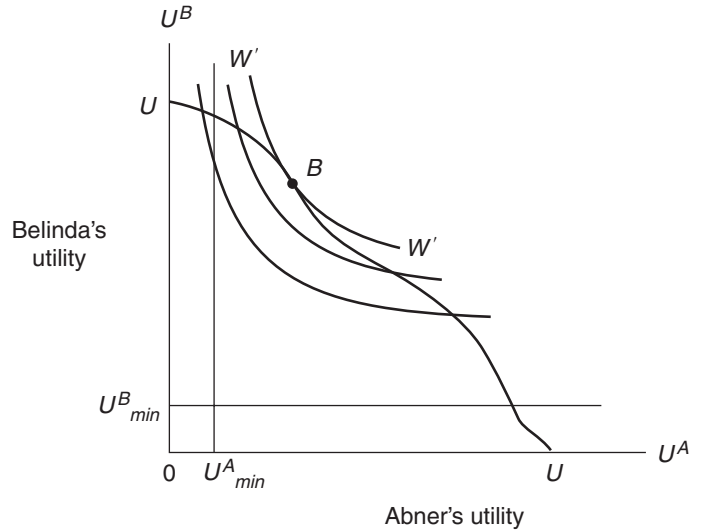
**THE SOCIAL WELFARE FUNCTION** Society must now decide which point on  $UU$  to choose according to the rules by which societies operate—through debate, consensus, and maybe even dictatorial power. Economists define this set of rules as a social welfare function. Consider an unusual society in which the consensus was that everyone’s utility level should be exactly the same, with no variations tolerated. We would recognize this as a set of right-angled social indifference curves along a 45-degree line from the origin. The optimum allocation would be at point  $A$ , which is a tangency between the  $WW$  curve and the  $UU$  curve. This would indicate equal utility levels for Abner and Belinda. From point  $A$ , we can then return to the one point on the contract curve in Figure 18-1 in which the utility levels are equal, denoted  $Z$ . Selecting this point leads to the unique allocation of the two goods to Abner and Belinda (although not necessarily the same amounts to each, as shown in Figure 18-1).

Many would find the specific social welfare function with equal levels of utility highly questionable and almost impossible to define or obtain. Humane societies might agree, however, that everyone should receive with enough resources for at least a minimal standard of living; for Abner this would be  $U_{min}^A$ , and for Belinda this would be  $U_{min}^B$ .

We can characterize this condition as two constraints on the  $UU$  curve in Figure 18-7. Such societies would consider no social welfare function that would provide Abner with less utility than  $U_{min}^A$ ; this would be similar for Belinda regarding  $U_{min}^B$ . Even here, however, the answer depends on the exact social welfare function chosen. Societies in which people like Belinda are most influential may have social welfare functions with social indifference curves like  $W'W'$ . In this case, people like Belinda would get larger shares of the resources than people like Abner. The optimum at point  $B$ , although providing subsistence living for citizens like Abner, would leave citizens like Belinda better off than the optimum in Figure 18-6.

**THE SOCIAL WELFARE FUNCTION AND HEALTH CARE NEEDS** Within this framework, we now discuss the various social choice processes that scholars and policy makers have proposed for the equitable provision of health care. Let the social welfare function of society ( $SW$ ) represent the preferences of society as a whole. The function in a commonly used form:

$$SW = f(U_1, U_2, \dots, U_n) \tag{18.1}$$



**FIGURE 18-7** The Utility Frontier and Social Welfare Maximization: Preferences Favoring Belinda

where social welfare is characterized as a function of the utility levels of each of the  $n$  persons in the society. Utility for each person as usual depends on his or her consumption of the available goods in society, including health care. In a variation, we may suppose that each individual to some degree perceives an external benefit from the consumption of health care by others.

We may perceive society to be efficient when it acts as if it were choosing among its variables to maximize the social welfare function. Consider the choice process graphically in Figure 18-7 as choosing the highest social indifference curve attainable given the utility possibility frontier. This leads to the appropriate level of health care. The health care needed by each person in society is that level which maximizes  $SW$ . Social welfare will be maximized when society chooses its optimal health status goals in conjunction with optimal levels of other goals.

The social welfare function has proved flexible within health economics to formulate other conceptions of health equity. Yet a lot of what economists do when they are advising the public lies outside standard welfare economics. In Box 18-1, we have suggested several “extra-welfarist” criticisms. It may now be beneficial to define and contrast the relevant terms more thoroughly.

### BOX 18-1

#### The Extra-Welfarist Critique

Werner Brouwer and colleagues (2008) do an excellent job of drawing distinctions between welfare economics and extra-welfarist views, and we consider their work here. Some extra-welfarist criticisms go well beyond the usual criticisms of economics, directed more at the basis of welfare economics itself:

1. Consumers may not be rational.
2. Individuals may not be the best judges of their own well-being.
3. Social welfare may depend on more than individual utilities.
4. Consumer tastes are not fixed but are often learned and malleable.

These issues, recognizable to economists, raise serious objections for the analysis we have described. For example, if the assumption of rationality fails to approximate behavior, then most microeconomic theory would need to be re-evaluated.

*(continued)*

While items 2 and 3 further show how challenging the normative issues are, they are not foreign to America. U.S. society often makes choices that imply that individuals are not the best judges of their own well-being. Examples include motorcycle helmet laws, criminalization of drugs, and mandatory retirement contributions.

However, controversy always arises over where to draw the line. In applying welfare economics to the uninsured, Mark Pauly suggests that voters could probably be convinced that the value of certain reforms aimed at reducing the rate of uninsurance is worth the costs. However, he asserts, “If we cannot convince the decisive voters of the value of what we value, then I think we need to accept the verdict of democracy” (p. 14). Uwe Reinhardt responds in contrast:

I, for one, believe that, if this nation is ever to have truly universal health insurance coverage and a truly humane safety net all around, an elite espousing those goals would have to impose that state of affairs on generally confused plebs that has quite unstable, often logically inconsistent and utterly malleable preferences on the matter (Pauly and Reinhardt, 1996, p. 24).

Reinhardt echoes the frustration of many health economists on this point, but many also question it. *Plebs* here means the general populace, but it shares the same root with the word *plebiscite*, a vote of the general populace. Would health policy “impose(d) . . . on a generally confused plebs” pass a vote, or plebiscite?

**WELFARIST HEALTH ECONOMICS** Thus the social welfare function (SWF), which we have used to choose the “best” social option in Figure 18-7, though the dominant approach, is disputed by many health economists. Called “welfarism” by its critics it is contrasted with “extra-welfarism.” It will be beneficial to define and describe it more thoroughly here.

The SWF represents welfarism when based solely on the utilities of the individuals who make up society. As in much of economics, we assume that these individuals are rational and that they know what is best for themselves. Everyone counts. If John would improve his utility when society moves from *A* to *B*, and if no one else is harmed, then society must choose *B*. Thus the Pareto Principle, which you recognize here, applies in the SWF.

This welfare economics contains two other salient features that matter when we compare the extra-welfarist argument. First, note that the mathematical form of the SWF and any equity weights that may be incorporated in it come from “outside” of the assumptions described previously. Perhaps they come from a societal advisor, or, more in keeping with the individualistic and democratic nature of the approach, perhaps society could determine preferences over these features by surveys of the public.

Second, the concept of utility is in some places variously defined. The introductory classroom will find it defined as “a measure of satisfaction” or even “happiness.” The more sophisticated definition describes utility as merely an index of preferences; where preferred bundles get higher utility numbers. The latter definition puts an emphasis not on emotion but on what the individual would choose under the right conditions.

**EXTRA-WELFARIST ANALYSIS** In contrast, the extra-welfarist approach allows valuations other than the individual’s utility in the determination of social choice, and the sources of valuation may be other than the affected individual’s. Therefore it is less individualistic and tends to be less democratic, at least in the sense that unlike the “methodological individualism” of standard welfare economics, extra-welfarism makes no assumption that the individuals’ preferences are the best guide to what is best for themselves.

Sen (1980), for example, questions whether a person’s utility is a reliable guide to his or her well-being. He argues that utility may merely result from, or represent, the emotions of the moment. He proposes, instead, that individuals are entitled to an acceptable level of “capability,” which includes health and normal functioning. Cookson (2005) has proposed that the quality of life-years

**BOX 18-2****The “Fair Innings” Proposal**

Williams (1997) proposes that people would generally agree to the principle that everyone is entitled to a normal span of life at a reasonable level of quality. The Williams metaphor fits on both sides of the Atlantic: Cricket has one or two innings and baseball has nine innings—generally that is all anyone gets—but everyone should get that much. This idea applies with most force to trade-offs in life years between the old and the young; in this context, it implies a rationing by age.

Consider two individuals, each of whom stands to gain eight good-quality extra years of life after being “rescued” from a lethal disease by medical technology costing \$100,000. Let one of these people be 80 years of age and the other 30. If resources are scarce, which individual should get the treatment? The “fair innings” concept would clearly require the treatment go to the younger person since the 80-year old person has had his or her fair innings already.

Valuations by society in this fair innings scheme are generally not so simple. This central issue lies in how much society prefers to help the one versus the other. If we agree that the elderly facing illness and the young facing illness are not equal, then the question for Williams becomes “To what degree are we as a community averse to this inequality?”

(QALYs) measure provides an acceptable approximation to Sen’s capabilities. An interesting twist of this theme is Alan Williams’s Fair Innings approach; a description of his reasoning is in Box 18-2.

**Norman Daniels’s Concept of Health Care Need**

The social welfare function framework implicitly treats health care as just one of many commodities that provide utility, something determined by the same social choice mechanism by which society makes all its choices of social goals. Some analysts argue instead that health is special and that health care needs have a more objective and independent basis.

To address the special role of health, the philosopher Daniels (1985) has developed a health care needs definition, and we outline his approach here.

1. **Health care is special.** This view separates health care from other goals. To make this separation plausible, we must view health care as a primary good that is special because it is central to the task of attaining or restoring a fair equality of opportunity. In this sense, it comes prior to other considerations.
2. **Species-typical functioning.** Daniels argues further that the human species has a range of functioning that is typical and appropriate to it. *Disease* here means the absence of health; *health* is the ability to attain a functioning level typical to the species.
3. **Fair equality of opportunity.** Given the nature of society and the human species, there is a range of behavior opportunity that every person in society should have available. The range will vary somewhat from person to person inasmuch as we each have different endowments of skills and abilities, but all are entitled to their fair share.

To use the fair equality of opportunity standard of health care need, we must imagine that there is a degree of objectivity behind its construction, and that need, so constructed, would be observable in common by most people. We do not solve this matter here, but instead point out what Daniels’s project entails. If we accept Daniels’s view as correct, then health care need is separate from the other goals for society. We leave as a discussion question the issue of whether this is the case.<sup>3</sup>

<sup>3</sup> The issue can be pursued further with Daniels’s book (1985) and articles (1981, 1982).

## Economic Criticisms of Need-Based Distributions

Mainstream economic ideas often clash with need-based conceptions of the appropriate distribution of health care resources. The economic criticisms are directed to particular conceptions of need and can be understood as saying “If by need you mean X, then the following criticism applies. . . .” Several criticisms of this sort exist in the literature.

1. **The bottomless pit.** If the technical maximum health status goal is chosen, it will likely be the case that the cost of bringing all inputs to bear for some patients, even when these inputs have little effect on the patient’s health, could exhaust society’s resources. This would constitute a bottomless pit.
2. **Needs should not be chosen independent of costs.** Society’s health status goals should depend in part on the costs of health and the price of health care. Health is undoubtedly subject to diminishing marginal returns in production. If some other societal goal offers greater marginal utility per dollar than health care, society could improve its well-being by transferring the money value of the marginal health care unit to the service of that other good, for example, education.
3. **The role of scientific medicine in determining needs.** From similar reasoning, it follows that health needs cannot be determined solely on the basis of scientific medical knowledge. The role of medical experts is critical in needs analysis, inasmuch as scientific data are required to determine the medical inputs needed to attain a given health goal. However, the appropriate health status goals themselves must be chosen with knowledge of society’s economic constraints and its values. In some form, the political process is required to identify the trade-offs that society is willing to incur to attain any given goal.
4. **Monotechnic needs.** Finally, Fuchs (1974) has pointed out that when needs analyses are stated in terms of needed health resources per capita, they often implicitly (and usually incorrectly) assume that only one available technique exists for pursuing a health status goal. More plausibly, many opportunities exist for substitution, not only among health care inputs but also among health care inputs and other inputs into the production function of health.

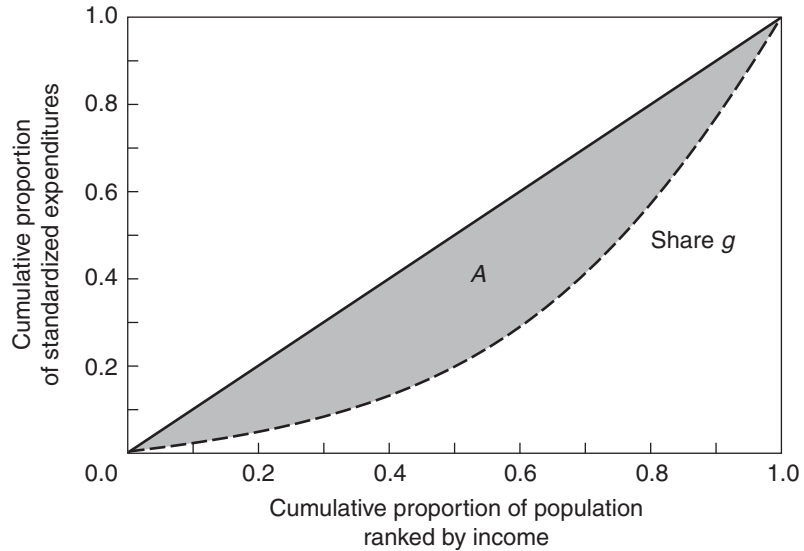
## HORIZONTAL EQUITY AND NEED

Horizontal equity is the requirement that equal people be treated equally. Analysts have often compared health care equity across countries using a Gini Index, or at least one modified from the original Gini’s use, though recently researchers have developed measures with more attractive features. Consider first what the traditional Gini Index is.

Figure 18-8 depicts the cumulative portion of the population ranked by income (on the horizontal) graphed with the cumulative portion of earned income on the vertical. The diagonal line indicates the “perfect distribution.” Along this line each income group is earning an equal portion of the income. The broken curve line represents the Lorenz Curve, the actual situation. For this example, the bottom 20 percent of the income distribution is earning 5 percent of the income. Income inequality favoring the rich occurs whenever the actual curve lies below the diagonal. The Gini Index is the area marked *A*, and the Gini Coefficient is the ratio of area *A* to the total area under the diagonal.

We see that by these definitions the Gini Coefficient must always lie between zero (perfect equality) and one (complete inequality); in other words the Lorenz Curve must always lie below the diagonal. For example, it would be illogical to say that the lowest 20 percent of the income people received more than 20 percent of the income. But things are different when health care inequality is the issue.

Suppose we wish to compare the cumulative proportion of health care use with the cumulative distribution of income. In contrast to income, it is not illogical to say that the lowest 20 percent of the income distribution receive *more* than 20 percent of the health care. Such a distribution would be “biased” in favor of the poor. We shall see the distribution of health care favors the lower income



**FIGURE 18-8** The Gini Coefficient

groups in many of the European countries. We can imagine in such a case a figure somewhat like Figure 18-8 except that the actual distribution curve has portions that lie above the diagonal. By convention we associate bias toward the poor with negative numbers and bias toward the rich with positive numbers.

While the Gini analysis still embodies the basic ideas of horizontal equity analysis, health economists (van Doorslaer, Koolman, and Jones, 2004; Koolman and van Doorslaer, 2004) have developed further measures and extensions of the analysis, which they find more useful. These authors prefer a Concentration Index, stated as

$$C_M = \frac{2}{\bar{y}} \text{cov}(y_i, R_i) \quad (18.2)$$

where  $y_i$  is the health care utilization of income group  $i$ ,  $\bar{y}$  is the mean health care use in the population, and  $R_i$  is the cumulative fraction of the population in fractional income group  $i$ . The equation (18.2) states the unweighted version of the index, which is correct when all of the groups are of equal size. If the groups have different sizes, a weighted version must be used, as in van Doorslaer, Koolman, and Jones (2004). Under the present assumptions, the unweighted covariance of  $y_i$  and  $R_i$  is

$$\text{cov}(y_i, R_i) = \frac{\sum_{i=1}^{i=n} (y_i - \bar{y})(R_i - \bar{R})}{n} \quad (18.3)$$

Equation (18.2) shows that a bias favoring the rich will yield a positive covariance and thus a positive value for  $C_M$ . Consider this statement in more detail. Note that each term under the summation sign in (18.3) is a product of the individual's relative amount of care times his or her relative position in the income scale. If richer than average,  $(R_i - \bar{R}) > 0$ , and if at the same time receiving more care than average,  $(y_i - \bar{y}) > 0$ , then that product will be positive. Similarly, if poor and receiving less than average care, the corresponding product would also be positive.<sup>4</sup>

<sup>4</sup> Aaberge and colleagues (2010) provide an analysis of distributional measures in the presence of non-cash values.

If the predominance of care likewise tends to favor the rich and disfavor the poor, the covariance will tend to be positive. Conversely, a bias in favor of the poor will tend to result in a negative covariance. Returning to (18.2), we understand that a positive value for  $C_M$  suggests a bias in favor of the rich and a negative value for  $C_M$  suggests a bias for the poor.

In the data that follow, the authors have created the Concentration Index for several European countries, and they have also created a Health Inequity ( $HI$ ) index. The  $HI$  is calculated by first creating a Concentration Index for health need, denoted  $C_N$ . They then subtract  $C_M - C_N$  to control for the need-based variation. The reason behind this step, in the authors' view, is that we should not view health care visits to the doctor that respond to immediate need as treatment caused by inequity of the health system itself.

To devise their estimate of health need, the authors regressed doctor visits on "need indicators." For this purpose they chose health status measures, morbidity (illness), and demographics. They then created  $C_N$  by replacing the values for  $y$  in the equation (18.2) with this measure of an individual's health care need.

The Health Inequity index is

$$HI = C_M - C_N \quad (18.4)$$

Both indexes are reported in Table 18-1. Index  $C_M$  is the Concentration Index as we have previously described it. Index  $HI$  may be thought of as health care inequality after removing the underlying variation that can be attributed to need. Note that  $C_M$  and  $HI$  estimates in the table are predominantly negative for these countries for General Practitioner visits, indicating a bias in favor of the poor.

However, for the specialty visits the reverse is true almost throughout. Table 18-1 gives us two means to assess the degree of bias in these countries. First, the sign of the indices indicates a bias toward the rich if the index is positive and toward the poor if the index is negative. Second, we measure the strength of the bias by the size of the index and whether the index estimate differs significantly from zero at the 5 percent level.

**TABLE 18-1 Health Care Inequality Measures Across Several Countries**

	$C_M$	$HI$	$C_M$	$HI$
	GP Visits Total	GP Visits Total	Spec Visits Total	Spec Visits Total
Ireland	-0.1323*	-0.0696*	0.0770*	0.1388*
Belgium	-0.1145*	-0.0508*	-0.0269	0.0255
Spain	-0.0906*	-0.0492*	0.0267	0.0740*
Luxembourg	-0.0918*	-0.0406*	-0.0704*	-0.0282
Italy	-0.0649*	-0.0349*	0.0179	0.0537*
Greece	-0.1258*	-0.0308*	-0.0418*	0.0492*
Germany	-0.0636*	-0.0268*	0.0158	0.0517*
UK	-0.1006*	-0.0240*	-0.0234	0.0524*
Netherlands	-0.0535*	-0.0113	-0.0178	0.0413*
Denmark	-0.0831*	-0.0008	0.0223	0.0844*
Portugal	-0.0692*	-0.0051	0.0971*	0.1604*
Austria	-0.0499*	-0.0146	0.0345	0.0740*

Note:  $C_M$  is the Concentration Index and  $HI$  is the Health Inequality Index, both of which are described in the text. The table is created from data published in van Doorslaer, Koolman, and Jones, *Health Economics*, 2004, Tables 1 and 2, pp. 637–38, with permission. An asterisk indicates that the estimated value is significant at the 5 percent level or better. "GP" stands for general practitioner, and "Spec" stands for specialist.

## THEORIES OF SOCIAL JUSTICE

Inevitably, understanding what health care distribution is equitable and choosing what health care needs should be met in a society depends on ethical theory. An ethical theory serves to identify a context and reasoning by which to determine what we *ought* to do, as opposed to mere positive analysis which describes what we do. Ethical theories that serve to determine a fair or just distribution of economic resources are sometimes called theories of social justice. Seen this way, any notion of equity or need in health care, to be complete, must be connected to an ethical theory, perhaps to a theory of social justice.

Although there are several theories of social justice, there is no consensus-accepted theory. Even without a consensus, however, such theories help to illuminate issues to address in order to achieve a consensus. We offer a brief overview of three social justice theories, along with a selection of criticisms of each theory.

### Utilitarianism

Utilitarianism became prominent in the nineteenth century and is still current in modified forms. It can be understood as the greatest good for the greatest number. In its classical form, it identified the social optimum coincident with maximization of the sum of utilities of all persons in society. Classical utilitarianism essentially defines the social welfare function as the sum of individual utilities.

Followers of utilitarianism promoted by Jeremy Bentham in the eighteenth and nineteenth centuries, conceived of an individual's utility as cardinally measurable, at least in principle, and comparable among individuals (your "utils" compare with my "utils"!.) The utilitarian ethic originally was conceived somewhat literally as maximization of society's total satisfaction level.

Utilitarianism captures the idea of trade-offs among goals. Under this construction, society may choose to accept some harm for a few members in return for a greater good for the many. As such, it avoids the bottomless pit criticisms mentioned earlier. Health status would not generally be maximized for every individual in society under this view.

**SOME CRITICISMS OF UTILITARIANISM** Economists came to criticize classical utilitarianism early in the twentieth century. They generally rejected the idea that utility could be cardinally measurable and comparable among people. Economists view it as unscientific to suppose that one individual's level of satisfaction could somehow be added to that of another person. Modern social welfare theory in economics has proceeded along ordinal utility lines. Utility in these theories retains the role of ranking preferences among alternatives, but we have discarded the notion of a fixed quantitative measure of happiness in most modern theory.

Two other criticisms of utilitarianism illustrate some of the weaknesses identified in the theory. One is the question of domain—that is, whose utilities are to count? Utilitarianism does not itself identify where to draw the boundaries of membership in the society. Are foreign people or noncitizens to count? If not, why not? Are animals to count? Unborn future generations? Is the utility of the fetus to count or only that of the already born?

A second criticism raised by Nozick (1974) poses a similar question regarding possibly malevolent individuals in society. For example, suppose an individual, because of bigotry or sheer malevolence, gets satisfaction out of the suffering of some other group in society. Is the malevolent utility of such a person also to count?

### Rawls and Justice as Fairness

John Rawls (1971) approached the concept of social justice from a different viewpoint. Here, a primary principle of justice is that social choices must be fair. Rawls views it as unfair for people with economic or political power who often have vested interests because of their circumstances in society to dominate the social choices. Instead, according to Rawls, to be fair we should make our



choices from a position divorced from arbitrary special interestedness. Such a position, it is proposed, is one from behind the “veil of ignorance.”

The Rawlsian veil of ignorance is a hypothetical situation in which we can think rationally but for which our particularities of self and economic situation are as yet unknown. It is as if we could somehow contemplate life in society before we are born and before we know whether we will be rich or poor, black or white, male or female, tall or short, and so on. Rawls believes that, so divorced from our vested interests of life, we would generally come to a consensus about principles of social justice, and specifically we would agree to the Rawlsian “maximin” principle.

Under the maximin principle, we would each reason that without knowing who we were to become in society, we would presume that we could be the person worst off. Under such circumstances, we would agree, argues Rawls, only to a system of justice in society that maximized the position of the worst off. This need not result in complete equality of incomes and resources including health care, but it would permit departures from equality only if the lot of the worst-off would improve. Health care under a Rawlsian system of social justice would presumably also be provided if the needs of the worst-off were regarded as a priority.

**SOME CRITICISMS OF RAWLSIAN JUSTICE** Rawls’s theory of justice also has drawn criticism. Critics have noted that Rawls assumes that each of us behind the veil of ignorance is extremely risk-averse. Suppose that an alternative situation, A, offered everyone an income of \$10,000, while alternative B offered one person \$9,000 but everyone else \$100,000. Under the maximin principle, persons behind the veil would choose alternative A, the alternative with the higher income for the worst off. Would people really be so risk-averse as to forgo even extremely good odds of a large gain? The Rawlsian theory of justice also appears subject to the bottomless pit argument. The instance of health care provides a good example of the problem in the views of Arrow (1973):

Thus there could easily exist medical procedures which serve to keep people barely alive but with little satisfaction and which are yet so expensive as to reduce the rest of the population to poverty. (p. 251)

Despite the criticisms, Rawlsian justice provides a prominent example of a theory of social justice that entails a strong presumption in favor of equality—a presumption that permits inequalities to arise only if they contribute to the lot of the worst-off.

### **Liberalism, Classical and Modern**

Classical liberalism refers to the political philosophy developed largely during the Age of Enlightenment, which centers on the eighteenth century. Led by the principles of John Locke (1690), it emphasizes the rights of the individual to his property and to himself. In this view, people enter the state voluntarily, and are free to choose what they deem best for themselves and their families. The tradition was the intellectual guide to the American Constitution, as is well-known to American schoolchildren. These “classical liberals” also included Adam Smith and, later, John Stuart Mill and Friedrich Hayek. Liberty in this tradition was largely a matter of emphasis as opposed to a fixed constraint. These philosophers largely supported taxation and often mentioned favorably things, such as government programs, to improve the well-being of the community; Locke (1697, p. 4), for example, favored government aid to the poor by way of the English Poor Laws, Smith favored public schooling for working-class children (1776, pp. III, II, 303), Mill mixed liberty values with social concerns (1975), and Hayek, whose life spanned the creation of the New Deal, spoke favorably of social health insurance itself (1960, p. 298). Would classical liberals, in sum, support, for example, a modern universal social health insurance plan? Folland (2005) reasons that they might do so.

In contrast to the older, verbal tradition of these economic philosophers, Nozick (1974) departs both in analytical style and by drawing stronger implications. Nozick asserts that government must limit itself to maintaining only the necessary services: the “minimal state.”

Nozick proposes that natural rights suggest the necessity for a *libertarian constraint*. By a libertarian constraint, he means that any system of social organization should prohibit the coercion of others, and that people are entitled to keep any property received through a voluntary transaction. From these principles, Nozick justifies the existence of a minimal state. He argues that these principles of justice necessarily limit the role of the state, and in consequence, this would exclude social programs beyond the minimal functions of government in providing public police protection services. It follows that social programs providing for health care also would not be warranted, and the health care needs of one person would not place obligations on any other person other than for what he or she voluntarily is willing to accept.

**SOME CRITICISMS OF THE MODERN LIBERTARIAN THEORY** A central focus of criticism is the assumed libertarian constraint itself. For those who do not accept the constraint as an implication or necessity of natural rights, it will be hard to see why we are not free to trade off some degree of liberty in order to make gains in efficiency. For example, we do not allow slavery even if the would-be slave would voluntarily accept it. In the health setting, the Food and Drug Administration restrictions on the availability of certain drugs represent restrictions on liberty, but at least, in some cases, they may be supportable on efficiency grounds. Pauly (1978) has argued that if the costs of conveying information to the public are sufficiently large, substantial efficiency gains may accrue simply by restricting access to a potentially dangerous drug. The restriction of access may in some cases require a physician's prescription; in other cases, it may require banning the drug from the market.

Members of society frequently accept rules that restrict liberty somewhat but that are expected to improve outcomes. Wittman (1982) has discussed the potential efficiency of simple rules in day-to-day life and in sports. A traffic light restricts liberty but promotes the efficient and safe flow of traffic. If we accept the principle that liberties can be traded off to gain certain efficiencies, this raises the question of whether we must accept the libertarian constraint.

## CONCLUSIONS

The three theories of social justice described here attempt to ground our understanding of the distribution of goods and services—including health care—in a system of ethics. The brief review of these theories cannot account for all arguments and rebuttals nor does it constitute an exhaustive coverage of the many normative models available. Those interested in ethics, as it concerns economic distribution, should consult the original sources.

The theories, however, serve to show how widely people's understanding of appropriate distributions of health care differ, and their consideration suggests that we have as yet no consensus. Nevertheless, these issues of justice are raised whenever society wishes to modify the distribution of health resources on grounds of need and equity. The issues are no less important because there is disagreement.

The meaning of efficiency is more sharply defined. Here the controversial issues involve the degree of efficiency attainable either with existing health care markets or with health care markets as modified by new policies. The theoretical model of perfect competition generates a Pareto efficient outcome. Health care markets in practice, however, differ in many respects from those that have perfect competition. Perhaps the most notable discrepancies of actual health care markets from the theoretical model arise because of the role of uncertainty, the problems of information, and the presence of externalities.

Perfection in either direction is not attainable. We must inevitably accept approaches to health care distribution that are second-best, evaluating each proposal on its merits. While perfect competition is unattainable, proposals that improve the degree of competition in health care markets may nevertheless improve society's well-being. On the other hand, proposals for providing social insurance programs to at least some segments of the population can be supported in principle on efficiency grounds whenever substantial charitable externalities are present. Alternatively, social insurance programs may be justified on the basis of one or more systems of social justice.

## Summary

1. Pareto efficiency defines a situation where it is no longer possible to make mutually beneficial changes. It is Pareto efficient to exhaust all avenues for gains that benefit someone and harm no one.
2. Under theoretical conditions of perfect competition, the competitive market is Pareto efficient. In the Edgeworth box for exchange, regardless of the initial endowment position, a competitive free exchange will lead to a Pareto efficient point, a point on the contract curve. This is the First Fundamental Theorem of Welfare Economics.
3. The Second Fundamental Theorem of Welfare Economics also is illustrated by the Edgeworth box for exchange. The theorem holds that we can achieve any Pareto efficient outcome, in principle, by a competitive market, given a suitable initial distribution of resources.
4. Price discrimination is Pareto inefficient. This result includes that form of price discrimination arising when a favored segment of the population is provided reduced prices for health care to improve access.
5. The health care sector deviates from the conditions of perfect competition in many respects. These include major issues of the role of uncertainty, the role of information, and the role of externalities.
6. The Theorem of the Second Best suggests that promoting competitive features in health care markets is not welfare enhancing *per se*. Nevertheless, many competitive proposals in practice may be welfare enhancing.
7. Social programs for the subsidization or provision of health care can be theoretically rationalized on grounds of efficiency as well as equity. The usual efficiency argument posits the existence of a charitable externality in health care. The presence of externalities may, in principle, justify market interventions.
8. Need-based distributions of health care resources tend to be based on equity grounds and usually imply a rejection of market outcomes. Health care need may be understood as health care resources required to attain a given health goal chosen by society.
9. Society's optimal choice of goals may be summarized by the social welfare function, defined over all possible combinations of the individual utilities of society members. This leads to the optimal choice by selecting the point on the utility-possibility frontier that maximizes social welfare.
10. Daniels bases his concept of health care need on the fair equality of opportunity. He argues that health care needs may be identified separately from other social decisions.
11. Several need-based distributions can be criticized on economic grounds. These grounds argue that health care needs should not be (1) the technical maximum, (2) independent of cost, (3) chosen solely on technical medical criteria, or (4) monotonic.
12. A philosophical theory of justice is needed to provide an ethical grounding for a proposed distribution. There is no present consensus on such a theory of justice.

## Discussion Questions

1. At point  $0_A$  in Figure 18-1, Belinda has all of both goods. Is this point Pareto efficient? Is it equitable? Discuss.
2. If society could clearly choose an equitable point reflecting a distribution of the two goods, is this point inevitably going to lie on the contract curve?
3. Choose an example of a health care market and identify ways in which it differs from the perfectly competitive model. Do you think that these deviations from competition could each be repaired by appropriate policies? Discuss.
4. We describe several economic criticisms of need-based distributions. Do any of these criticisms apply to Daniels's conception of health care need?
5. Speculate on how each of the three described theories of social justice would view government programs designed to provide infant and child care to the poor using tax dollars.
6. Under utilitarianism, one maximizes the total utility of society. What does this imply about the marginal utility for each person? What does it imply about the total utility for each person?
7. Suppose society determined that it must provide a minimal sustained level of health to everyone. What would this imply regarding society's expenditures on health?
8. Insufficient health care for some often is seen as a problem of insufficient income to purchase health care. Discuss two alternatives to social programs that provide health care.

## Exercises

1. Prove that point  $B$  in Figure 18-1 is not Pareto efficient.
2. Draw an Edgeworth box like the one in Figure 18-2 but with only these details inside: the point  $V$  and budget line  $AB$  through  $V$ . Using indifference curves, depict the utility-maximizing choice for Abner. Now let the budget line rotate to  $CD$ , drawn to reflect a higher relative price of medical care. Again identify the utility-maximizing choice for Abner.
3. Is it possible to find a point on the contract curve that is not a competitive equilibrium?
4. Let  $Q_{opt}$  in Figure 18-4 represent the optimal level of health care in society under the external benefits rationale for social health care programs. Would members of society necessarily view  $Q_{opt}$  as the equitable amount of health care?
5. If all taxpaying members of society became “hard-hearted,” feeling no external benefit in the health care provided to others, then what would be the optimal health care output under the external benefit rationale in Figure 18-4?
6. In Figure 18-7, which depicts the utility-possibility frontier, would society ever choose an inefficient point (inside  $UU$ ) as the optimal point?
7. Suppose Fred has an income of \$5,000 per year, and Harry has an income of \$105,000 per year. If we tax \$50,000 from Harry to give to Fred, will this represent a Pareto improvement for society? Why or why not? Would this improve society under some other criteria?
8. Suppose Fred has an income of \$5,000 per year, and Harry has an income of \$105,000 per year. If we tax \$5,000 from Harry to give to Fred, will this represent a Pareto improvement for society? Why or why not? Would this improve society under some other criteria?

# Government Intervention in Health Care Markets

- Economic Rationale for Government Intervention
- Forms of Government Intervention
- Government Involvement in Health Care Markets
- Government Failure
- Conclusions

Previous chapters have revealed the scope of government involvement in health care delivery. We have seen that (1) government spending accounts for a substantial portion of all health care spending, (2) governments are deeply involved in producing as well as financing health care services, and (3) governments regulate the health care industries. Also, we have found that governments are at the center of most contemporary health care issues. These include major policy issues, such as national health insurance and initiatives intended to control costs and increase access to care, and also less publicized issues, such as those involving federal government support for education and research.

Chapters 20 and 21 will provide detailed coverage of the main regulatory and social insurance programs in the United States. Chapter 22 will examine health care systems in other countries, while Chapter 23 will focus on recent reform efforts in the United States. The current chapter provides a framework for assessing the role of governments in health care markets. We begin with the conventional approach favored by economists, which emphasizes market failure as the rationale for government intervention. We follow with a historical review, including examples of federal, state, and local involvement in the health economy. So that we can evaluate government programs with a more critical eye, we close this chapter with a discussion of government failure to identify impediments in developing effective programs.

## **ECONOMIC RATIONALE FOR GOVERNMENT INTERVENTION**

As we have seen, efficiency is one common standard for evaluating the desirability of economic allocations. Inefficient allocations are associated with various distortions that lead to market failure. The previous chapter introduced the distortion resulting from beneficial externalities as one justification for social programs in health care.

## Monopoly Power

Monopoly power provides the classic example of market failure. A profit-maximizing monopolist produces to the level at which marginal revenue equals marginal cost. Because the marginal revenue lies below the demand curve, the monopolist will reduce production below competitive levels, and the price charged by the monopolist will exceed the marginal cost of production. The reduced production and the price-marginal cost gap together create the welfare loss.

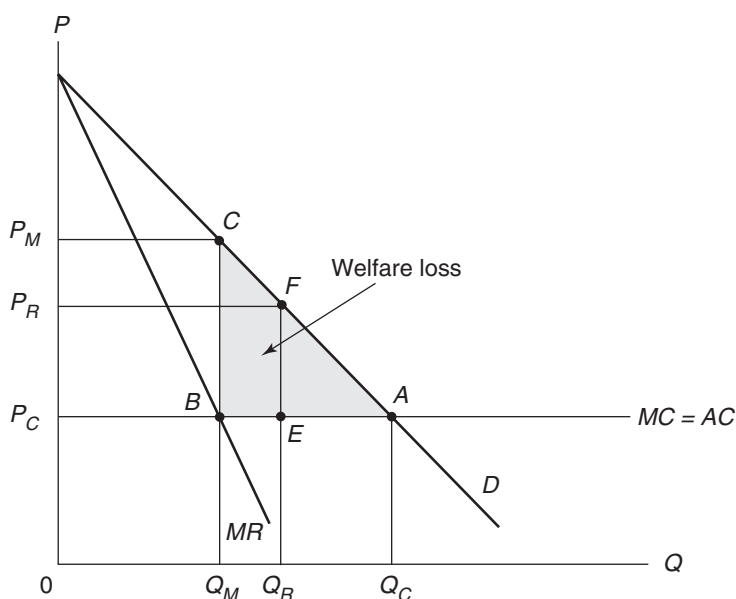
Figure 19-1 summarizes these standard results. With a perfectly elastic competitive supply in the long run, which also represents the industry's average costs ( $AC$ ) and marginal costs ( $MC$ ), the competitive price and quantity are  $P_C$  and  $Q_C$ . If this industry is monopolized and no changes occur in demand or costs, the profit-maximizing output is given by  $Q_M$ , where  $MR = MC$ . The higher monopoly price is  $P_M$ , and triangle  $ABC$  represents the welfare loss.

However, monopoly power need not be associated solely with pure monopoly. The monopoly model is applied commonly to markets in which one or a small number of sellers are dominant. Several health care markets seem to hold a potential for the exercise of monopoly power. Examples include hospital services in markets with few hospitals, pharmaceutical products protected by patents, and some health insurance markets dominated by Blue Cross and Blue Shield associations.

The potential for monopoly power exists even in markets characterized by a large number of sellers, as in the markets for doctor and dental services. Licensure laws and other forms of regulation restrict entry into some professions. Furthermore, professional associations may be able to reduce price competition by setting minimum fee schedules or by inhibiting the flow of information to buyers.

Several issues arise regarding monopoly power. First, some barriers to entry result from government intervention itself. These include licensure and patent laws. Licensure intends to ensure minimal standards of quality; patent laws seek to promote innovative activity.

Second, monopoly power may be inevitable in some situations and does not necessarily lead to economic profits. In a small market, for example, demand may be sufficient for only one hospital to survive while it just covers costs. If demand were to diminish, even the one existing hospital might not be able to survive unless it either received subsidies or donations, or cut its costs.



**FIGURE 19-1** Welfare Loss of Monopoly

Third, the proposed cure to monopoly inefficiencies may be worse than the problems posed by the existence and exercise of monopoly power. Some have argued that direct intervention through public provision or price controls could worsen the situation because of government failure. These critics suggest that countervailing forces and other constraints on the full exercise of monopoly power will tend to arise in private markets, especially where antitrust laws are enforced vigorously.

In the simplest case and in the absence of government failure, price controls can theoretically reduce the welfare loss caused by monopoly. In Figure 19-1, suppose a price ceiling of  $P_R$  is established. The monopolist's marginal revenue is constant, equal to the price,  $P_R$ , up to an output of  $Q_R$ . Because marginal revenue exceeds marginal cost up until output  $Q_R$ , the monopolist will produce at least to  $Q_R$ . Beyond  $Q_R$ , the marginal revenue comes from the usual  $MR$  curve because the monopolist would have to lower prices to attract more buyers. Marginal revenue then will be below marginal cost and eventually will become negative. The profit-maximizing output is  $Q_R$ , and the welfare loss now falls from area  $ABC$  to area  $AEF$ . While it appears in this case that price regulation can be an effective instrument in reducing prices, monopoly profits, and welfare losses, consider that a typical hospital or physician provider may produce many different services. Consider also that demand and technology change constantly and that it may be difficult to monitor quality. As we will discuss later, price regulation under such circumstances becomes far more difficult.

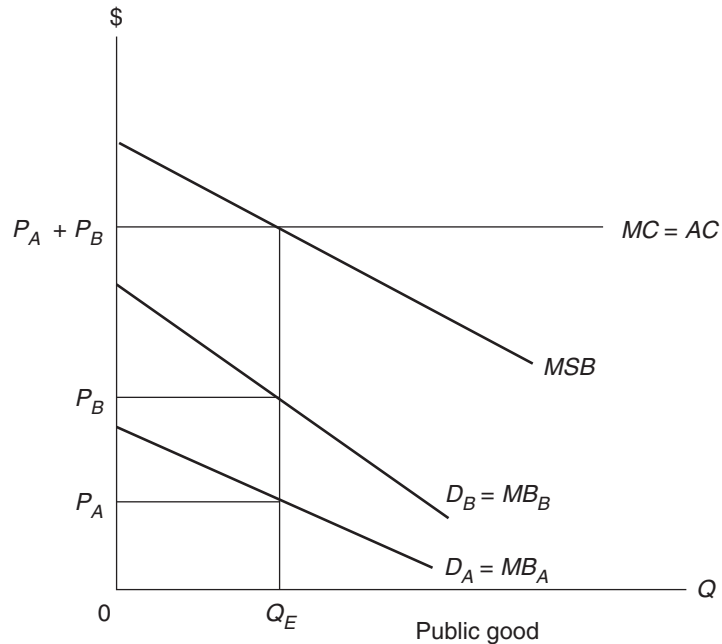
### Public Goods

A public good should not be confused with the public provision of a good. The postal service and garbage collection service are examples of public provision of private goods. Government may provide such goods because of natural monopoly, or a desire to subsidize certain users (e.g., rural postal customers who might not otherwise be served by a profit-maximizing postal service). In contrast, a pure public good is one for which consumption is nonrival (i.e., consumption by one individual does not reduce someone else's consumption) and nonexcludable (i.e., a consumer cannot be excluded from consuming the good either by having to pay or through some other mechanism). Economists often use national defense as an example of a pure public good. Other examples of goods having some degree of "publicness" include highways, bridges, and parks.

Market failure arises because an inefficiently small quantity of pure public goods typically will be provided without government intervention. Figure 19-2 develops this proposition. Begin with a simple case involving two persons whose demands and marginal benefits for a public good are represented by  $D_A$  and  $D_B$ . By definition, a public good is nonrival and nonexcludable. Each person must consume the identical quantity that he or she buys or that is made available. Thus, the marginal social benefits ( $MSB$ ) are the sum of the two individuals' marginal private benefits. Since, unlike private goods, each will get the same amount of the public good, the combined willingness of the two consumers to pay for the public good, the  $MSB$ , is the vertical summation of  $D_A$  and  $D_B$ .

Assuming constant costs, the efficient quantity is  $Q_E$ , where  $MSB$  equal the marginal costs of production,  $MC$ . Will the efficient quantity become available without government intervention? Knowing that their contributions are important, the two consumers may decide to cooperate by making voluntary contributions to fund at least some of the good. In fact, if each contributes according to his or her marginal private benefit at  $Q_E$  by paying  $P_A$  and  $P_B$ , respectively for every unit, enough money will be collected to cover the cost of providing the optimal quantity  $Q_E$ . It is not certain, however, whether this solution will be reached.

More realistically, public goods usually involve a large number of individuals. In principle, determining the optimal quantity, using a vertically summed marginal social benefit curve, follows the two-person example. The major difference is that these people are unlikely to cooperate to fund efficient amounts of the public good through voluntary contributions. Instead, more are likely to become free riders (i.e., make no contributions) because they cannot be excluded from consuming the



**FIGURE 19-2** The Optimal Quantity of a Public Good

good and because any voluntary contribution one makes to the provision of the good will have a negligible impact on the good's availability. The predicted undersupply of public goods in private markets has led many to conclude that governments should be responsible for making them available.

Are health care services public goods? Health services provided to one person are not at the same time consumed by others. Also, those who do not pay can be excluded from receiving care. Therefore, health care services are private goods even though they may involve public provision (e.g., through the Department of Veterans Affairs) or public financing (e.g., through Medicare and Medicaid). Thus, the public goods rationale for government provision of health care is not immediately apparent.

Despite this caveat, economic theories of public goods are highly relevant to certain health care issues. Consider the following cases:

**INFORMATION** Information can be thought of as an economic good with a large public content. One consumer's obtaining information does not reduce the information available to another. Although those who do not pay often can be excluded from receiving information, the marginal cost of providing information to another individual is often relatively small. Because of this, one can argue that information will be underproduced in private markets and that government intervention is needed to increase its availability.

The government may take on two distinct roles. The first is to help disseminate existing knowledge to the public, either through direct provision or through subsidizing private sector activities. Second, governments may expand the stock of knowledge by taking an active role in scientific research, again by direct provision or through subsidizing private sector research. The federal government has pursued both strategies.

**REDISTRIBUTION** Voluntary giving also has a public goods dimension. Donations help raise lower-income persons' standards of consumption, including their consumption of health care. By letting others donate and knowing that one's contribution will have little impact on total



contributions, individuals often become free riders. To help offset this effect, we may need mandatory programs to correct the undersupply of voluntary giving.

However, considerable disagreement can arise over the form of the redistribution. Should redistribution occur through cash transfers to let recipients spend their additional incomes as they see fit? Or should some programs take the form of in-kind transfers (e.g., Medicaid) where recipients must use the transfers to purchase health care services? The many issues relating to equity and mechanisms for redistributing income are discussed in Chapters 18 (“Equity, Efficiency, and Need”) and 21 (“Social Insurance”).

## Externalities

In contrast to pure or nearly pure public goods, another group is composed of those goods that have third-party effects, also known as externalities. Recall that externalities arise when a third party is affected by another party’s consumption or production of a good. If a neighbor’s loud music or smoke from burning leaves bothers you, it constitutes a negative externality.

Moreover, to be considered an externality, the effects must be transmitted outside the price system. A situation where an increase in demand for lower-cholesterol meats raises their prices, adversely affecting consumers of these products, is not an externality. The higher prices ration the supply of low-cholesterol products.

The major problem is that the prices of the goods and services may not fully reflect many negative or positive externalities. Thus, even when competitive forces drive prices to the marginal private cost of production, social efficiency requires that marginal social benefits equal the marginal social costs. Marginal social benefits sum the marginal private benefits and any marginal external benefits that might exist, while marginal social costs similarly sum the marginal private costs and any marginal external costs.

When a negative externality, such as pollution, creates a marginal external cost, a competitive market tends to overproduce the polluting good relative to the socially efficient quantity. The market price will not reflect the external cost. Conversely, competitive markets tend to undersupply goods that create beneficial (positive) externalities. Marginal social benefits exceed the price at the competitive output. Positive externalities can be important in health care, as when a charitably minded person derives satisfaction from knowing that the sick, poor, or uninsured consume more health care. More tangible externalities occur when others are inoculated against communicable diseases.

We should not confuse positive externalities with health benefits that are largely private. Subsidies that lead to improved health are often supported on the grounds that recipients will benefit society by being more productive. However, the gains from an individual’s increased productivity are largely private. Thus, the effects of the consumption of health care on productivity are not externalities.

## Other Rationales for Government Intervention

Several other arguments favor government intervention. An important responsibility of the federal government is to stabilize the economy through macroeconomic policies. Although macroeconomics does not usually focus on specific sectors of the economy, changes in monetary, fiscal, and debt policy can have major effects on federal and state health care programs, as well as on private health care spending through changes in taxes and interest rates.

Another distinct category involves government’s role in promoting the consumption of merit goods. Merit goods are commodities thought to be good for someone regardless of the person’s own preferences. Supporters of the arts, compulsory education, and mandated consumption of other goods argue that individuals do not always know what is in their best interests. Undoubtedly public policy with respect to public health interventions, such as seat belts and alcohol, tobacco, and drug use, has reflected the merit goods idea.

A final role for government involves incomplete markets which occur when private markets fail to meet an existing demand. Certain insurance markets, such as those for patients with cancer, HIV/AIDS, or other pre-existing conditions who seek new insurance, may be examples of incomplete markets in the sense that those patients may be unable to buy insurance at any price. Government may have a role in filling these gaps by providing insurance or requiring insurers to do so.

It is important to determine, however, whether some of these markets are truly incomplete. Is there sufficient demand by those willing to pay actuarially fair rates so that a market would emerge? Because premiums would, on average, match insurance payments, they would be very high. Are such patients seeking subsidies by having legislation guaranteeing access to insurance at lower than actuarially fair rates?

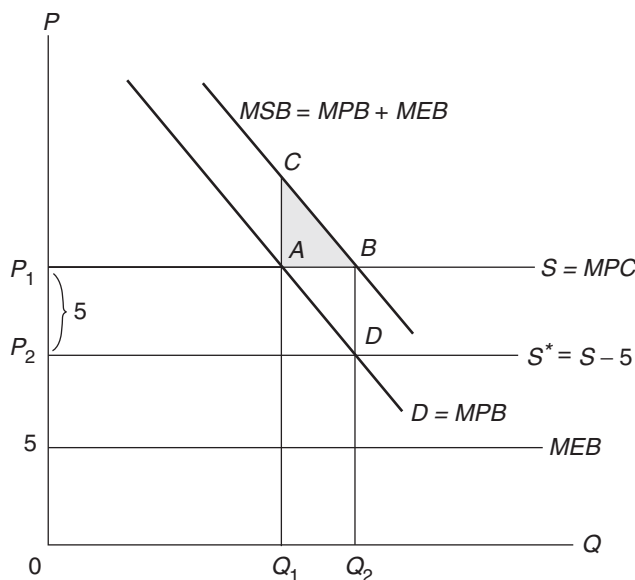
### FORMS OF GOVERNMENT INTERVENTION

Governments can adopt a variety of policies and instruments to influence the allocation of resources or the distribution of income. The principal categories relevant to health care are selective commodity taxes and subsidies, public provision of health care, transfer programs, and regulation.

#### Commodity Taxes and Subsidies

We already have established that a competitive market is inefficient when beneficial (positive) externalities result from the consumption of a commodity. Using inoculations against infectious disease as an example, we extend our previous work to show how taxes and subsidies can, in principle, correct for the externality.

Let demand,  $D$ , and supply,  $S$ , in Figure 19-3 reflect the marginal private benefits ( $MPB$ ) and marginal private costs ( $MPC$ ) of inoculations.  $MPB$  equals  $MPC$  at the equilibrium quantity,  $Q_1$ . However, the competitive allocation is inefficient when those inoculated confer a marginal external benefit ( $MEB$ ) on others. Too little is produced at  $Q_1$  because marginal social benefit, the vertical sum of the  $MPB$  and  $MEB$ , exceeds marginal social costs (here equal to  $MPC$  because there are no external costs). Pareto efficiency and the elimination of the welfare loss shown by triangle  $ABC$  require output  $Q_2$  where  $MSB$  equals  $MPC$ .



**FIGURE 19-3** Using Commodity Subsidies to Correct for Positive Externalities

The imposition of a commodity tax or subsidy provides a method of correcting for the externality. Assume for simplicity that the marginal external benefit is constant at \$5 per inoculation. To correct for the externality, Congress can provide producers with a subsidy of \$5. The supply curve facing consumers will shift down by the amount of the subsidy to  $S^*$  because producers will need to receive \$5 less than before from consumers to produce the quantities shown along the original supply,  $S$ .<sup>1</sup> With the new supply, the equilibrium price paid by patients decreases to  $P_2$ , and the equilibrium quantity rises to the optimum  $Q_2$ . Governments pay producers  $\$5Q_2$ , or rectangle  $P_1BDP_2$ .

This simple example illustrates several important features of a commodity subsidy. First, the price paid by consumers will fall by the full amount of the subsidy only when the supply is perfectly elastic, as in Figure 19-3, or when the demand is perfectly inelastic. In other cases, as with a positively sloped supply, the price paid by consumers will not fall by \$5. Including the subsidy, producers will receive a higher price than before. Both consumers and producers share the benefits of the subsidy even though the government pays it to producers.

As the elasticity of demand increases, or elasticity of supply decreases, the equilibrium price falls by smaller amounts, and the equilibrium quantity increases by larger amounts. In the long run, the competitive market supply will be highly elastic (perfectly elastic in a constant-cost industry). Thus, the benefits of a subsidy will accrue mainly to consumers in competitive markets.

Second, to achieve efficiency, decision makers need accurate estimates of both the monetary value of the externality and elasticities of demand and supply. The costs of administering taxes and the difficulties of accurate estimation are strong arguments against the use of a corrective subsidy unless the positive externality is substantial.

Finally, we note that policy makers can achieve the results described in Figure 19-3 by taxing those who are not inoculated rather than subsidizing those who are (see Exercise 4). By imposing a cost on those who refuse inoculations, the opportunity cost of an inoculation is its price minus the tax. The lower effective price increases quantity. The administrative complexity of monitoring whether people are inoculated argues against this approach.

When consumption of a good, such as smoking, leads to harmful (negative) externalities, the marginal external cost must be added to the marginal private cost to determine the efficient solution. The competitive output will be greater than the optimum. To reduce consumption, price must be raised. A price increase can be achieved either through a corrective tax to shift the supply curve up (Box 19-1) or through a subsidy to nonsmokers that raises the effective price of smoking.

## Public Provision

Roads, education, water, and police and fire protection are just a few examples of the many goods and services provided by governments. Most of these goods are not pure public goods. Although national defense often serves as an example of a good that is both nonrival (my use does not prevent your use) and nonexcludable (I cannot keep you from enjoying the good even if you don't pay for it) we cannot say the same about water, education, and most health care services. Inoculations are rival and excludable even though they may generate substantial positive externalities.

Public provision of health care is a complex process requiring a decision for each of the three basic economic questions (what? how? and for whom?) faced by every society. The “what” question relates to the types of health care to be provided (e.g., limited services, such as inoculations or comprehensive health care) as well as their quantity and quality. Whether governments themselves

<sup>1</sup> A \$5 subsidy to the consumer will shift the demand curve up by the \$5 so that the new demand mirrors  $MSB$  and passes through point  $B$ . The solution remains the same. Consumers buy quantity  $Q_2$  at price  $P_1$ , paying  $P_2$  net of the subsidy. It will generally be easier to administer commodity taxes and subsidies through producers than consumers.

**BOX 19-1****Is There a Case for a “Junk Food” Tax?**

Chapter 7 introduced the health care consequences of obesity. To the extent that individual insurance premiums do not fully reflect the additional costs of treating diseases associated with obesity, and recognizing the burden that obesity places on publicly funded programs, the obese population unquestionably imposes negative externalities on others. For obvious reasons, legislating a corrective tax on individuals considered obese, i.e., Body Mass Index (BMI) > 30, is not practical or realistic (though, in this spirit, United Airlines began enforcing a controversial policy in 2009 that required some obese passengers to purchase a second seat). A growing number of employers have adopted wellness programs that penalize employees who fail to take steps to manage their weight, or reward those that do so. For example, in 2008, Alabama required its state workers to undergo free screenings. Those who refused the screenings or who refused to take action for problems such as high blood pressure, high cholesterol, or obesity faced higher insurance premiums beginning in 2011.

In addition to wellness programs, health care experts have directed considerable attention to “junk food.” Junk food is generally associated with calorie-dense food that has high levels of sugar, glycemic starch, and saturated fat. The experts often associate such products with obesity, especially childhood obesity (e.g., Harris et al., 2009), but we caution that a causal relationship between junk food and obesity is more difficult to establish (Collins and Baker, 2009).

Powell and Chaloupka (2009) review the literature on food price policy and taxes in the United States. Food is subsidized for those who qualify for SNAP (formerly known as the Food Stamp Program) and several other programs such as the National School Lunch and Breakfast Program. For the most part, there are no restrictions on the types of food that are allowed under these programs. As for taxes, many states impose sales taxes on at least one category from among soft drinks, candy, and snacks (Kim and Kawachi, 2006). However, with relatively inelastic demands for these products, even tax rates as high as 20 percent would have little impact on annual consumption.

From a review of nine scholarly articles that met certain selection criteria, Powell and Chaloupka reach a sobering conclusion regarding the effects of food and restaurant prices on BMI and weight. The relatively small taxes on junk food which we have experienced have not produced significant changes in weight outcomes. More substantial price increases might lead to some measurable effects, especially for children and adolescents.

In 2011 several states proposed bills to levy additional sales taxes on junk food, e.g., \$0.01 per ounce on drinks sweetened with sugar or corn syrup in California, and an extra 5 percent on junk food purchased in convenience stores or restaurants in Nevada. While these taxes may raise significant revenues, they are not likely to have large effects on obesity levels.<sup>2</sup> As a result, there is interest in other interventions to replace or supplement tax policy. These include “traffic-light” systems that rely on food labeling with visual cues on nutritional content. The United Kingdom adopted a voluntary red-amber-green color program to reflect the nutrient levels of fat, sugar, saturates, and salt in processed foods. Using a choice experiment survey to gather data on willingness-to-pay for reductions in these nutrients, Balcombe et al. (2010) found traffic-light labeling to be promising. Econometric analysis of the survey data indicated a high willingness-to-pay for a shift from red to green for all four nutrients with much lower willingness-to-pay for shifts from amber to green, i.e., consumers are especially concerned about high levels of these nutrients.

<sup>2</sup> In September 2011, Hungary imposed a substantial tax on foods with high contents of fat, sugar, and salt. The following month, Denmark introduced a tax of about \$1.30 per pound of saturated fat in a product. Policymakers will be watching to see how these high tax rates will affect consumption. For more information see Cheney (2011).

produce the services and how they do it, or whether they contract with the private sector, is a part of the “how” question. The “who” question deals with the financing and distribution of the services. Will the program be created as a universal entitlement, as an entitlement for some groups, or as one with other eligibility requirements? Furthermore, should the program be “free” at the point of service and funded mainly by tax revenues, or should the beneficiaries be charged user fees? The funding mechanism could have large impacts on the resources allocated to health and on the possible redistributions as from rich to poor and young to old.

### **Transfer Programs**

Cash transfer programs usually are intended to meet society’s equity concerns by redistributing income, with recipients free to spend their income in any way they want. Social Security for the elderly and some disabled is the principal example, but income supports also are provided for the poor. Cash transfers to the poor include Temporary Assistance to Needy Families (TANF), funded by the federal government and administered by the states.

In-kind transfers (benefits other than cash) also redistribute income, but their main purpose is to increase a recipient’s consumption of specific goods or services. Important in-kind transfers include Medicare for the elderly, food through the SNAP program, housing, and Medicaid for those who pass means tests. Medicare and Medicaid are described in detail in Chapter 21.

### **Regulation**

Governments influence the allocation of resources by establishing rules and regulations. In the extreme, governments can prohibit certain goods or activities entirely, such as the production and consumption of illicit drugs. More commonly, governments regulate the form or terms under which goods are produced or consumed. Regulation may appeal to legislators because it appears to tackle problems without incurring substantial government spending in the process. For example, regulating managed care to prevent “drive-through” deliveries appears to respond to a public concern at little direct cost to government. However, as described in Chapter 12, the legislation imposes significant costs on others.

Regulation in health care markets can take many forms: licensure laws; mandates; and regulation of price, quality, and quantity of services. The following sections contain several examples of regulation. Chapter 20 will focus exclusively on the principal regulatory mechanisms in health care markets.

## **GOVERNMENT INVOLVEMENT IN HEALTH CARE MARKETS**

Hardly any economic activity is free of government intervention. Intervention comes mainly through three activities: provision of goods and services, redistribution, and regulation. Governments have pursued each of these activities in the health economy. Through public or VA hospitals, and other programs, they provide substantial amounts of health care, though this activity is overshadowed by social insurance programs for the elderly and many poor. The provision of health care and of health insurance is also the major means used to redistribute income to lower-income groups from higher-income groups.

Less obvious to many is government’s role as a regulator. At the federal level, the Securities and Exchange Commission (SEC), the Environmental Protection Agency (EPA), and the Occupational Safety and Health Administration (OSHA) are regulatory agencies that affect nearly every business and working individual. In addition, states and localities impose various requirements such as those governing building and safety codes.

However, when economists and others speak of regulating or deregulating the health care industries, they are not referring to the kinds of social and commercial controls cited earlier but rather to regulations such as HIPAA (Box 19-2) targeted specifically at the health care industries.

**BOX 19-2****What Is HIPAA?**

The Health Insurance Portability and Accountability Act (HIPAA), that passed in 1996 and required compliance by October 2003, is one of the most far-reaching health care regulatory measures ever approved. A principal goal of HIPAA is to protect workers for loss of coverage when they change jobs, especially to reduce “job lock” for those with pre-existing conditions. It also prohibits discrimination by insurers based on health status. Among other goals, including standards for electronic data interchange, the Act also was intended to guarantee the security and privacy of patient health data. Patients must be able to access their records and know how their personal information will be used. Changes in 2009 introduced guidelines on how to protect health information appropriately. Also, amendments to HIPAA provided new guidelines for dealing with breaches of protected information.

HIPAA requirements are complex, and the penalties for failure to comply are stiff. As a result of the complexities, a cottage industry consisting of legal experts, information specialists, and consultants has emerged to help physicians, hospitals, insurers, and other providers to achieve and maintain HIPAA compliance. The economic effects of HIPAA requirements on costs and other consequences, including restrictions on the ability of researchers to access health care data, have not yet been fully evaluated.

Government involvement in the health economy takes on many forms, some of which are developed elsewhere in this text. Here we provide examples to highlight the variety and scope of government intervention in health care markets.

**Support of Hospitals**

As described in Chapter 14, the modern hospital did not begin to emerge until the confluence of several developments in the late nineteenth century, such as major improvements in anesthesia, antisepsis, and in the invention of X-rays. Temin (1988) characterizes hospitals prior to this period as being more like municipal almshouses funded by taxes or voluntary contributions. Hospitals “existed for the care of marginal members of society, whether old, poor, or medically or psychologically deviant” (pp. 78–79).

In retrospect, one can argue that public support for hospitals reflected a redistribution motive and by a desire to deal with the negative externalities associated with living with the insane and those harboring communicable diseases, such as tuberculosis. With the improvements in physicians’ abilities to diagnose and to treat patients surgically, hospitals grew rapidly in the first decades of the twentieth century. Public hospitals continued to serve the poor but also focused their attention on the growing middle classes. Ultimately, patient payments and insurance became the primary sources of funds for many of these institutions.

**The Hill-Burton Act**

Federal support for private hospitals was minimal until the passage of the Hill-Burton Act in 1946. At that time, about 40 percent of U.S. counties did not have a community hospital, with many existing hospitals considered substandard. The Hill-Burton Act sought to expand rural health facilities by providing for matching grants to nonprofit institutions. It established a standard of 4.5 beds per 1,000 population. The program, which helped finance about one-third of all hospital projects, contributed substantially to the rise in hospital bed per capita between 1947 and 1975 when funding under the Act ended.

The Hill-Burton Act also required a state “to survey its hospital needs, to develop a statewide plan for the construction of public hospitals and health centers, and, finally, to construct needed hospitals and health centers.” This process of planning was carried a step further with the enactment of other health-planning legislation. As governments increasingly emphasized cost issues in the late 1960s, the health planning agencies turned to capital expenditures controls to limit hospital growth

and costs. The resulting Certificate-of-Need (CON) requirements represented a major turning point toward an increasingly regulated health care system.

### **Department of Veterans Affairs and Department of Defense**

In addition to direct subsidies for hospital construction, governments are also major providers of health care. State, county, and municipal hospitals account for approximately 20 percent of total hospital beds in the United States. The largest public provider is the Department of Veterans Affairs (VA). In 2009, the VA spent \$42 billion for providing, among other services, hospital care for 640,000 patients and almost 74 million outpatient visits. The VA's primary purpose is to provide care for service-related injuries through institutions that specialize in providing and undertaking research for such care. However, it will treat veterans for other conditions unrelated to service injuries if facilities are available and the veteran indicates an inability to afford treatment from other sources. As a result, most patients in VA hospitals are lower-income people treated for conditions not related to their military injuries. A large portion of the substance abuse treatment in the U.S., for example, comes through the VA system.

In addition to direct care provided by the VA, the Department of Defense (DOD) is responsible for providing coverage to active-duty service members as well as reservists and various other categories of military personnel, their families and survivors. Coverage comes through its TRICARE program (formerly known as the Civilian Health and Medical Program of the Uniformed Services or CHAMPUS). TRICARE provides several options including HMO and fee-for-service plans. In 2009, the DOD spent over \$36 billion on health care. Health spending, as a share of total DOD spending, increased by nearly 50 percent over the decade ending in 2010. While much of this increase is due to growth in the number of beneficiaries, Congress has been concerned about the cost burden facing DOD.

### **Food and Drug Administration**

As noted in Chapter 17, the U.S. has passed several pieces of drug legislation beginning with the Food and Drug Act of 1906. Legislation adopted in 1962 required increased testing and evidence of efficacy and gave the Food and Drug Administration (FDA) considerably more control over the introduction of new products than previously. Among many provisions found in the most recent amendment (in 2007), the FDA's authority and resources to conduct reviews were considerably expanded.

The pharmaceutical industry believes that regulatory requirements lengthen the period to FDA approval to about 10 to 14 years and raise the cost of introducing new drug products, estimated by DiMasi and colleagues (2003) to be as high as \$800 million to the point of marketing. The economic issues of drug regulation pit the relative gains in drug safety and efficacy against the discouragement to innovation and the delays in availability attributable to the approval process.

### **Mandated Health Insurance Benefits**

The insurance industry is regulated mainly at the state level. The states' traditional roles are to ensure financial solvency of insurance companies, but in both auto and health insurance, the states have expanded their roles considerably. One relatively recent phenomenon in health insurance is the proliferation of mandated benefits. Mandates can require coverage for specific health care services, e.g., breast reconstruction; provider types, e.g., physical therapists; or even who is eligible for coverage, e.g., domestic partners (Table 11-4 provides additional examples). Such laws had grown from 48 in 1970 to 2,156 in 2010. They often pass after lobbying by previously excluded special interest groups, or after highly publicized cases involving persons who find they are not covered for heart transplants or certain expensive experimental therapies.

Gabel and Jensen (1989) found that these requirements prevent insurers from offering lower-cost alternatives that exclude the mandated benefits. This effect could be substantial for many lower-income individuals who have been priced out of the market.

## Tax Policy

Federal and state governments provide substantial tax subsidies for the consumption of health care and purchases of insurance. In particular, employer contributions to group health plans are not included in an individual's taxable income (escaping federal, state, and Social Security taxes). Individual payments for health insurance as well as many health care expenses can be itemized if they exceed threshold levels (7.5 percent of adjusted gross income in 2011).

Some have argued that the substantial reductions in the after-tax prices of insurance and health care represent a major reason for the growth of insurance and consequent growth of health care spending that began in the 1960s and 1970s. Thus, a proposed cost-containment strategy is to reduce such subsidies. Without discussing the merits of this argument here, we wish only to point out that there seems to be little justification for the extensive tax subsidies. Ironically, the tax advantages are greatest to those in the highest marginal tax rates, meaning that the subsidies do not promote equity.

## Public Health

We have thus far focused largely on the study of personal health and health care decisions. Public health, in contrast, deals with communicable diseases, epidemics, environmental health issues, and other activities with significant third-party health effects, such as smoking and the use of drugs and alcohol.

Until the latter part of the nineteenth century, public health was a state and local responsibility. However, in 1878, the U.S. government created the Marine Hospital Service, which became the U.S. Public Health Service in 1912. Including state and local public health spending, total government spending on public health activities was \$77 billion in 2009. Odin Anderson (1990) provides an excellent history of public health in the United States; Mays and Smith (2011) provide current estimates on the significant reductions in preventable causes of death associated with increased public health investments.

## Other Government Programs

Numerous other channels for government involvement exist. Some are apparent—such as support for medical education and medical schools, which will influence the supply of providers. Similarly, support for health care research undertaken directly by a government agency, such as the National Cancer Institutes or undertaken by other public and private organizations, can have a substantial impact on the spread of technology and the direction of the health care system. For example, critics of the U.S. system claim that it has placed excessive emphasis on high-tech medicine at the expense of preventive medicine and that this bias stems in part from the kinds of research projects supported.

Other forms of intervention are less obvious. Changes in immigration policy can affect the supply of health care personnel. Also easily overlooked is the role of the federal government as the nation's largest employer in providing fringe benefits to its employees. The Federal Employees Health Benefits Program (FEHBP) offers numerous insurance options to employees and their dependents. Because of its size and its ability to experiment with alternatives, the FEHBP has the potential to influence and serve as a model for the private sector.

To summarize, in the United States governments at all levels have intervened heavily in the health care sector. The intervention has taken the forms of direct provision of care, financing of health insurance (especially of Medicare and Medicaid), subsidizing of medical education and the construction of health facilities, subsidizing of purchases of health insurance through the tax structure, and the regulation of health care industries. This intervention has major effects on the total resources devoted to health care as well as on the distribution of resources within the health care industries.

Following new accounting methods adopted in 2009, the Centers for Medicare and Medicaid Services (CMS) reported that governments were responsible for \$1.08 trillion, or 43.6 percent, of the



**TABLE 19-1 National Health Expenditures, by Payer and Sponsor, 2009**

	Total	Out of Pocket	Private Insurance	Medicare	Medicaid	Other Insurance & Third Party <sup>1</sup>
National Health Expenditures	\$2,486.3	\$299.3	\$795.3	\$487.7	\$384.8	\$519.2
Business, Households & Other Private Revenue	1,403.1	299.3	645.1	239.1		219.5
Private Business	518.3		397.5	77.7		43.1
Household	708.4	299.3	247.6	161.5		
Other Private Revenue	176.4					176.4
Government	1,083.2		150.2	248.6	384.8	299.7
Federal government	678.4		26.8	237.0	254.3	160.3
State and local government	404.8		123.4	11.6	130.5	139.3

<sup>1</sup>Includes the Department of Defense, Department of Veterans Affairs, Children's Health Insurance Program, workers' compensation, and other private and government programs.

Source: United States, Department of Health and Human Services, Centers for Medicare & Medicaid Services, *National Health Expenditures Tables*, [www.cms.gov/nationalhealthexpenddata/02\\_nationalhealthaccountshistorical.asp](http://www.cms.gov/nationalhealthexpenddata/02_nationalhealthaccountshistorical.asp). Accessed April 11, 2011.

nearly \$2.49 trillion in national health expenditures (Table 19-1).<sup>3</sup> The largest programs are Medicare and Medicaid. The federal government paid for about half of total spending under Medicare (\$488 billion) and for nearly two-thirds of the total for Medicaid (\$385 billion). Governments accounted for another \$300 billion in spending for many other activities including public health (\$77 billion), the departments of Defense and Veterans Affairs (\$79 billion), and the Children's Health Insurance Program (\$11 billion).

## GOVERNMENT FAILURE

Market failure is a necessary condition for government intervention. As we have seen, economists associate market failure with monopoly power, externalities, and public goods, including the public goods characteristics of redistribution and information. Government policies, in principle, can correct misallocations resulting from market failure. To do so, governments can use specific commodity taxes and subsidies, public provision of goods and services, transfer programs, and regulation.

We have to ask whether government in practice can improve efficiency and better meet society's equity objectives. Many might argue that this question has already been answered. Governments in each of the 34 countries that belong to the Organization for Economic

<sup>3</sup> CMS prepares the official National Health Expenditures (NHE) accounts which it reviews and updates every five years. Major changes affecting the classification of private and government spending began in 2009 to deal with the complexity of programs that have combinations of private, federal, and state and local funding. For example, workers compensation was previously included in government spending (state and local) even though the ultimate payers are private firms. Thus the proportion of government to total health care spending reported for 2009 is not consistent with the 47.3 percent share CMS reported for 2008 or values we have cited in previous editions of this text. Following passage of PPACA, analysts at CMS (Keehan, et al., 2011) projected that additional federal and state obligations will increase government's share to 49.2 percent by 2020.

Cooperation and Development (OECD) are heavily involved in their health economies, to the point where nearly all have universal health care coverage. The United States had the lowest public share of total health care spending in 2008 among the 34 countries, but the combined federal, state and local share is still substantial—46.5 percent—according to OECD data.<sup>4</sup>

The real issues are the extent and forms of government involvement. The difficulties of agreeing on objectives, choosing from many different policy instruments, and selecting the correct values of these instruments create many opportunities for “government failure.”

The literature on public choice illustrates many of the problems in developing and implementing policy. Public choice attempts to model how decisions are made through the political process. While many models of public choice have been created, we limit our discussion to two features that are relevant to efficiency: special interest groups and bureaucratic behavior.

### Who Does the Regulator Represent?

Wherever we have addressed regulatory activities such as licensure or other quality controls, we have assumed that the regulator knows the right actions to take. Most often we treat the regulator as representing an omniscient but benevolent despot who knows what is good for the economy and regulates accordingly. Yet students of railroad regulation might argue that the railroads played major roles in influencing their regulators. Likewise, some observers believe that the American Medical Association has greatly influenced laws on licensure and legislation on Medicare. Can the theory of regulation address issues such as these?

It can. An influential school of economists often associated with the University of Chicago has argued that the regulation process, like many others that we have examined, is a “maximizing” process, in which a regulator seeks “votes” from a group of potential beneficiaries. These votes would allow the transfer of wealth, such as monopoly profits, from those regulated to those who benefit from the regulations.

In this model formulated by Peltzman (1976), the regulator seeks supporters by imposing a percentage “tax” on those who are regulated, with the tax receipts being transferred to the beneficiaries. A higher tax rate on those who are regulated (more stringent regulation) may win more supporters, but it also may mobilize more opponents. The successful regulator seeks a certain transfer of wealth while gathering votes to effect this transfer efficiently.

Suppose the regulator starts with a modest tax rate. The imposition and increase of this rate yields favorable votes for two reasons:

1. The revenues increase the probability of support from the potential beneficiaries.
2. The revenues increase the value of the transfer to the beneficiaries.

The marginal benefit to the regulators is shown as the left-hand side of equation 19.1. As the tax rate rises, the marginal benefits fall for two reasons. There may be fewer supporters to attract (term [1]), and the taxation or regulation may be reducing the wealth left for the beneficiaries to tax (term [2]).

Opposition also will be heard. Those who would be regulated will oppose a transfer because it is coming from them. Their opposition will grow the higher the tax is (the more stringent the regulation is), that is, the more that is being taken from them.

---

<sup>4</sup> The OECD provides comprehensive annual health care data for its members. For the United States, these data are somewhat different than those reported in its NHE accounts. Mexico closely followed the United States with a public share of 46.9 percent. Denmark and Norway topped the list at 84.5 and 84.1 percent, respectively ([oecd.org/document/16/0,3343,en\\_2649\\_34631\\_2085200\\_1\\_1\\_1\\_1,00.html](http://oecd.org/document/16/0,3343,en_2649_34631_2085200_1_1_1_1,00.html)—Accessed April 12, 2011). See Paris (2010) for a comprehensive overview of health financing and other characteristics across most OECD countries.

Expressing this opposition effect in dollars, the regulator's optimum is found from the usual condition where the marginal benefits from regulation equal the marginal costs, or:

$$\begin{array}{rcccl} \text{marginal} & & \text{marginal} & & \\ \text{increase in} & & \text{revenue} & & \\ \text{probability of} & \times & \text{product from} & = & \text{marginal} \\ \text{support} & & \text{those regulated} & & \text{opposition from} \\ [1] & & [2] & & \text{increased taxes} \\ & & & & [3] \end{array} \quad (19.1)$$

The two terms on the left-hand side of equation (19.1) suggest that regulation will be sought if there is a large group of supporters (term 1) or if there is a smaller, well-organized group that perceives large expected gains (term 2). Thus, one may see regulations imposed that benefit special interest groups, that is, relatively small but well-organized groups of potential gainers.

While this analysis does not negate the economic arguments in favor of regulation, it provides sobering insights into how regulations may be implemented. Regulators may respond to similar economic incentives as do other economic men and women. The resulting regulations, although possibly addressing issues of economic efficiency, most likely will reflect the particular preferences of the beneficiaries, moderated by the opposition of those who stand to lose under the regulations.

### Bureaucracy and Efficiency

Lawmakers pass legislation, but government bureaucrats are responsible for implementing programs. Do bureaucrats try to minimize the costs of their activities? We know relatively little about the technical efficiency of the bureaucracy, but it is clear that government operations do not face the discipline of the marketplace. No income statements, balance sheets, and stock market prices are available to evaluate the success of a government program and motivate the bureaucracy, even though it is not completely immune from oversight (from legislatures, public watchdogs, and even government agencies, such as the U.S. Government Accountability Office). Nevertheless, the hypothesis that governments operate at lower levels of technical efficiency than the private sector is a plausible one.

Aside from the inefficiency associated with failure to minimize costs, government programs are also subject to allocative inefficiencies. That is, there may be too little or too much of a program. The more likely scenario is one where programs become too large because their supporters (lobbyists, legislators who support it, top bureaucrats who will run it) are motivated by self-interests to overstate the benefits and understate the costs.

These arguments warn us against the danger of concluding that government intervention necessarily improves social welfare when there is market failure. Equally serious government failure can result from inappropriate policies and high levels of technical and allocative inefficiencies. Just as the fact that many countries that have national health insurance programs suggests limitations of the mixed private-public system that has evolved in the United States, the fact that most countries face serious problems with their health care systems demonstrates the limits to higher levels of government intervention. A recent report by the OECD (2010, p.3) that examined the performance of different systems concluded:

There is no health care system that performs systematically better in delivering cost-effective health care. It may thus be less the type of system that matters but rather how it is managed. Both market-based and more centralized command-and-control systems show strengths and weaknesses.

## CONCLUSIONS

This chapter emphasizes market failure as the economic rationale for government intervention. Monopoly power provides the classic example, but public goods and externalities are two additional categories that are relevant to health care. Government policies, in principle, can correct misallocations resulting from market failure. To do so, governments can use specific commodity taxes and subsidies, public provision of goods and services, transfer programs, and regulation.

Through a wide variety of programs, governments at all levels have become major players in the health economy, accounting for 44 percent of national health expenditures in 2009. However, government activities also are associated with government failure. Information deficiencies, the efforts of special interest groups, and bureaucratic behavior can lead to socially undesirable programs or inefficient levels of these programs. Nonetheless, despite concerns about the effectiveness of government programs, governments will almost certainly continue to dominate the health economy through their various spending programs and heavy regulation.

At the same time, the United States has prompted competitive strategies to deal especially with the cost and access concerns. With the strong political pressure from some groups to downsize the role of government and even rescind PPACA, there is likely also to be continuing interest in market-driven changes to the private insurance system as well as to Medicare and Medicaid.

The following four chapters will further examine the role of governments in health care. Chapter 20 will describe and evaluate the principal regulatory mechanisms used in the United States, while Chapter 21 will concentrate on the main public insurance programs: Medicare, Medicaid, and the Children's Health Insurance Program (CHIP). Following an examination of health systems in other countries in Chapter 22, and the lessons learned from them, we will address health system reform in Chapter 23. Our focus will be on reforms directed at attaining universal or near-universal coverage in the United States, with special attention given to the Patient Protection and Affordable Care Act that President Obama signed into law in 2010. Because many respected scholars and political leaders embrace market solutions to the challenges of access, costs, and quality, we will also evaluate competitive approaches to health reform.

---

## Summary

1. Federal, state, and local governments accounted for about 44 percent of total health care expenditures in 2009. The share is projected to increase to 49 percent by 2020.
2. Governments' share of total health care spending in the United States is the lowest among the OECD countries.
3. The traditional rationale for government intervention is market failure. Sources of market failure include monopoly power, externalities, and public goods.
4. There is a public goods aspect to information and redistribution that can be used to justify a role for government in health care.
5. Commodity taxes and subsidies, public provision, transfer programs, and regulation are the principal policy instruments used by governments.
6. Regulation refers to the use of nonmarket means to affect the quality, price, or quantity of a good or service. The principal categories of regulation include fee controls and rate regulation, quantity and capacity controls, and quality controls.
7. Governments have been involved in a wide variety of activities, including the direct provision of health care, subsidizing the production of health care, the provision of social insurance, public health, and regulation of health care products and providers.
8. Traditional economic descriptions depict regulation as a process in which the optimal policy is determined and imposed by an omniscient regulator. Alternative theories suggest that regulations result from political processes. Thus, the regulations will reflect the particular preferences of the beneficiaries, moderated by the opposition of those who stand to lose if the regulations are imposed.

## Discussion Questions

1. What is meant by market failure? What is the potential role of government in each instance of market failure found in the health care sector?
2. In what sense can information and redistribution be thought of as public goods? Explain whether private markets will oversupply or undersupply these goods.
3. What are some examples of government providing health care? Of subsidizing the production or consumption of health care? Of providing insurance for health care? Of regulating health care markets?
4. Why is government needed to provide a public good? Under what circumstances might the voluntary contribution model, described in Figure 19-2, work reasonably well?
5. The standard monopoly-competition comparison describes the welfare loss. Develop arguments to support the view that in the real world:
  - (a) the welfare loss is exaggerated.
  - (b) the welfare loss is understated.
6. Why does air pollution represent a public health externality? How do regulations operate in addressing the air pollution? How do markets operate?
7. Mandated health benefits have proliferated since 1970. Discuss the pros and cons of a mandate that requires coverage of hearing aids for minors (17 states mandated this benefit in 2010).
8. What is meant by the tax subsidy of employer-paid health insurance? Explain why the subsidy is very likely to increase health care spending and thus the cost of such insurance. Give an example as to why a \$5,000 health insurance benefit provides a greater monetary benefit to someone in the 33% marginal tax bracket than someone in the 15% marginal tax bracket.
9. Some economists propose a tax policy that would allow individuals who purchase their own insurance to deduct these costs as well as all out-of-pocket costs for health care from taxable income. Explain how this policy could help offset the bias toward more comprehensive insurance resulting from the preferential tax treatment of employer-paid insurance.
10. The pharmaceutical industry has been subject to considerable regulation in bringing drugs to the marketplace. Discuss the benefits and costs to society from such a policy.

## Exercises

1. What is meant by the welfare loss of monopoly? Who bears this loss? Determine the regulated price in Figure 19-1 that will eliminate the welfare loss. Why will it be difficult in practice to adopt this solution?
2. Suppose that the public good in Figure 19-2 is associated with increasing costs of production (this would occur if  $AC$  is positively sloped.) Will the two individuals be able to fund the optimal amount if each pays according to marginal benefits? Repeat if there are decreasing costs of production.
3. Assume a positively sloped, short-run supply curve in Figure 19-3 and a constant \$5 per unit marginal external benefit. Show what happens with a \$5 subsidy given to producers. Who gains the benefits of this policy?
4. Using Figure 19-3, explain how a \$5 tax on those who are not inoculated will result in the efficient output. Why is this approach unlikely to work in practice?
5. Take a commodity, such as cigarettes or “junk food,” associated with negative externalities. Assume that the marginal private cost of production and the marginal external cost per unit are both constant. Graph the actual and efficient quantities. Show how a tax on producers (or consumers) can result in the efficient output. Explain whether it is possible to attain efficiency with a subsidy.
6. Draw a graph showing the marginal revenue ( $MR$ ) and marginal cost ( $MC$ ) of increasing the tax rate ( $t$ ) in the Peltzman model of regulation represented by equation 19.1. Place the tax rate on the horizontal axis and  $MR$  and  $MC$  on the vertical axis. Show the optimum tax rate.

# Government Regulation: Principal Regulatory Mechanisms

- Do the Laws of Supply and Demand Apply?
- Objectives of Regulation
- Regulation of the Hospital Sector
- Prospective Payment
- The Theory of Yardstick Competition and DRGs
- Regulation of Physician Payment
- Antitrust
- Conclusions

We have demonstrated that conventional economic tools can play a powerful role in explaining how health care markets work. While we must account for the special features of the health economy and modify the analysis where appropriate, we have found considerable empirical evidence attesting to the rationality of decision makers. Such considerations guide the evaluation of principal regulatory mechanisms applied to the health care economy.

## DO THE LAWS OF SUPPLY AND DEMAND APPLY?

Supporters of competitive strategies believe that market forces allocate resources to their best uses regarding health care, and that it would be unwise to ignore concerns about the strength of market forces. However, two articles separated by 15 years indicate problems that face those who would participate in the health care market. From the *Wall Street Journal* (1992):

In a conventional market, a high-cost, low-quality producer would be doomed, or, at least, worried. But the laws of economics have been repealed in the health business. Consumers don't comparison shop and often lack the information to do so. Many hospitals compete to buy the latest technology rather than provide the lowest-cost quality care. And providers resist outsiders' attempts to gauge quality. The result: a seemingly unstoppable increase in the cost of health care. ("Laws of Economics Often Don't Apply in Health Care Field," *Wall Street Journal*, January 22, 1992)

From the *New York Times*, 15 years later:

Patrick Fontana twisted his left knee last spring while hitting a drive down the fairway on a golf course in Columbus, Ohio. But what really pained him was the \$900 bill for diagnostic imaging ordered by his doctor. Mr. Fontana, a 42-year-old salesman, has a high-deductible health plan coupled to a health savings account. Since he was nowhere near meeting his deductible, he was on the hook for the entire bill. So he did something that insurance companies routinely do: he forwarded the bill to a claims adjuster, in this case a Web-based company that reviewed doctor and hospital bills for consumers.

After concluding that Mr. Fontana was not getting the best possible price, the company's representatives called the imaging facility and demanded a lower one, promptly saving him \$200—minus a 35 percent collection fee. "I asked before I went in to the clinic how much it would cost, and they just will not tell you," he said later. "I didn't know until I got the bill, and at that point I figured I had nothing to lose."

The savings are possible for one reason: medical care is often priced with the same maddening, arbitrary opacity as airline seats and hotel rooms. "The average provider—doctors or hospitals—has between 5 and 100 reimbursement rates for the exact same procedure," said Timothy Cahill, then-president of the web-based company. "A hospital chain with multiple locations may have 150 rates for the same procedure. Consumers don't know this." ("Bargaining Down That CT Scan Is Suddenly Possible," *New York Times*, February 27, 2007)

*Source:* Based on Mason, Michael, "Bargaining Down That CT Scan Is Suddenly Possible," <http://www.nytimes.com/2007/02/27/health/27cons.html>, accessed June 10, 2011.

Although a well-informed consumer like Mr. Fontana can try to negotiate claims, the type of price information that economists view as essential to informed health care market shopping is still very difficult to obtain.

How do market processes work in such circumstances? We saw in Chapter 19 that the fundamental story relates to monopoly power. We have established that most health care providers (hospitals, clinics, physicians) face portions of downward demand curves. In other words, they have some monopoly power because patients do not or cannot shop among providers, and do not immediately abandon the provider who raises its price. As such, the desirable outcomes of competitive markets become much more difficult to achieve. Using competitive markets as a comparison group, consider three major categories of policies:

1. Recognize providers' monopoly powers and try to control them. These policies include some traditional features, such as utilization review, capital controls such as Certificate-of-Need (CON) regulations, or rate controls.
2. Make monopolistic firms act like competitors. These policies include prospective payment plans for physicians and hospitals.
3. Prevent the accumulation of monopoly power. Most antitrust policies fall into this category.

In Figure 19-1 in the previous chapter, the monopolist faced a downward-sloping demand curve and a constant marginal cost curve. In equilibrium, price exceeded marginal cost. As a result, at least some consumers who would have willingly paid more than it cost to produce the health care service were unable to get it. The shaded triangle *ABC* in Figure 19-1 represented the loss to society from monopoly.

The consequences of real-world monopoly and production are more complex. We have ignored the possibility of quality degradation under monopoly and also the possibility that if producers cannot win the monopoly power of a patent they may delay technological innovation. Nevertheless, Figure 19-1 illustrates that competitive markets often will be more efficient than markets in which firms have monopoly power.

## OBJECTIVES OF REGULATION

Early U.S. regulation focused on the transportation and utilities sectors. Concern with excessive monopoly power and natural monopoly arising from economies of scale and high capital expenditures led to the 1888 establishment of the Interstate Commerce Commission, which regulated railroads and shipping. Other federal and state regulatory commissions came to regulate a variety of commodities (some of which have since deregulated) including electricity, gas, telephone service, oil pipelines, airlines, and broadcasting. The overriding objective in regulation was, and continues to be, rate setting. It generally seeks to limit the high price-cost margins that would result from monopolistic pricing.

Despite concerns with possible abuse of monopoly power in certain health care markets, advocates of increased regulation do not rely on the traditional monopoly power argument. They focus on the absence of discipline in markets brought on by a lack of consumer information and the prevalence of comprehensive insurance.

Policymakers have used lack of consumer information, as discussed elsewhere, to justify licensure and other quality controls as means to ensure minimal standards of care. Similarly, the role of insurance may serve as a source of market failure, by reducing the incentives to both patient and provider to shop around or otherwise conserve resources. Perhaps more importantly, traditional forms of insurance contribute to moral hazard. Unlike other sources of market failure, market failure associated with insurance has led to regulation of fees and to regulation of hospital revenues and costs.

### Regulatory Policy

Regulation refers to the use of nonmarket means to address the quantity, price, or quality of the good brought to market. Because regulations typically address the quantity of services used, the quality of the services, or the amount charged to payers (either individuals or insurers), consider total expenditures in a given health care market.

By definition, expenditures,  $E$ , equal the price of health care,  $P$ , multiplied by the quantity of services,  $Q$ , or  $E = PQ$ . Price of care,  $P$ , also depends on quality of care,  $q$ . Quality of care in this framework is any aspect of a service for which someone is willing to pay extra. More-experienced physicians, private rooms, or better food, may or may not help patients recover more quickly. Irrespective of whether they improve outcomes, patients may prefer these features of care, and hence be willing to pay more for them.

Suppose, for example, that Hospital 1 has quality index of  $q_1 = 1$ , whereas Hospital 2 has a quality index of  $q_2 = 1.1$ , or 10 percent higher. Because quality is valued, price  $P_2$  at Hospital 2 will be greater than price  $P_1$  at Hospital 1, although not necessarily 10 percent higher. Expenditures at Hospital 2 will also be greater than at Hospital 1, although not necessarily by the full price difference ( $P_2 > P_1$ ), if the higher price at Hospital 2 reduces the demand for services.

The key feature of this framework is the idea that within the health economy, regulation uses nonmarket means to affect  $P$ ,  $Q$ , or  $q$ , or combinations. For example, licensure requirements may raise quality, which will likely increase price. Licensure requirements that restrict the supply of providers may also increase price. Other regulations, such as the Medicare Prospective Payment System, may tend to decrease the amount of services used, lowering expenditures. We will address these aspects elsewhere in this chapter. The point here is that most regulations are directed toward changing price, quantity, or quality.

However, the objective of regulation is not merely to reduce spending; otherwise, rationing or even prohibiting certain services may meet this objective successfully. Rather, its goals are to promote minimal quality levels while eliminating the inefficient components of spending.

### Regulatory Instruments in Health Care

In the post-World War II U.S. health care system, regulatory instruments concentrated either on fee controls (the  $P$  in the equation), or quantity/capacity controls (the  $Q$ ), generally relating to the



hospital sector. State approval of rate changes was generally based on hospital costs per day or per admission, or on a hospital's overall costs.

Reimbursement for hospital care generally occurred on a retrospective basis, determined by allowable costs incurred. A number of states, however, adopted prospective forms of reimbursements beginning in the 1970s. Some were mandated at the state level while other programs were voluntary. The general goal of these arrangements has been to remove hospital disincentives to produce efficiently under cost-based retrospective reimbursement, and to limit the amount of care they provide hospitalized patients. Dranove (2000) and Salkever (2000) provide good historical descriptions of the development and implementation of such arrangements.

Prospective reimbursement, adopted by Medicare in 1983, marked an important turning point in regulatory efforts to contain the growth of hospital costs. Medicare's Prospective Payment System (PPS) limits hospital revenues per Medicare patient. Under PPS, hospitals receive a fixed amount per admission, determined by the patient's diagnostic condition, regardless of the hospital's actual costs. The strategy behind PPS is simple. Limiting revenues provides hospitals with strong incentives to limit lengths of stay and to refrain from providing unnecessary or marginal care. We provide a detailed discussion of PPS and its effects later in this chapter.

**QUANTITY AND CAPACITY CONTROLS** The controls on quantity tend to be indirect rather than direct, with the most prominent example being Certificate-of-Need (CON) laws. These laws require that hospitals obtain approval from planning agencies for capital expenditures in excess of various threshold levels. The operational premise is that unnecessary duplication of facilities as well as the mere availability of facilities leads to higher costs (i.e., fixed costs spread out over fewer users) and to higher rates of utilization (due to access and/or physician discretion).

Utilization review (UR), sometimes called *utilization management*, focuses primarily on quality, but proponents also expect it to reduce quantity. To the extent that third parties choose not to reimburse for services determined to be inappropriate, UR could plausibly deter the provision of questionable services.

## REGULATION OF THE HOSPITAL SECTOR

Given the hospital industry's large share of the U.S. health economy, policymakers have addressed cost inflation through regulation, primarily at the state level. In this section, we discuss briefly the three forms of traditional hospital regulation: utilization review, rate regulation, and Certificate-of-Need.

### Empirical Findings on Regulation

**UTILIZATION REVIEW** While "utilization review" once distinguished between fee-for-service and various forms of managed care, it has spread through all health care delivery systems. Morrissey (2008, Chapter 8) summarizes recent UR studies, including work by Lessler and Wickizer (2000) and Kapur, Grisenz, and Studdert (2003). Lessler and Wickizer look at readmission rates for medical diagnoses. For those with *no days denied*, 9.5 percent were re-admitted within 60 days, whereas 9 percent with *two or more days denied* were subsequently re-admitted—an insignificant difference. Morrissey cites further evidence that quality of surgical care may have been impaired for those for whom utilization review had limited lengths of stay.

Kapur et al. reviewed the records on all coverage requests of two large, multispecialty group practices in California in the late 1990s. Looking at emergency care services, they conclude that in the case of almost every denial the enrollee's medical condition deemed not to be an emergency according to the "prudent layperson standard."

Morrissey finds "remarkably little evidence" regarding the effectiveness of ambulatory utilization review. His pessimism extends to the impacts of primary care gatekeeping (in managed care

organizations) and newer forms of utilization management such as disease management and intensive case management.

**RATE REGULATION** Much of the early research investigating the impact of rate regulation (RR) centered on Connecticut, Massachusetts, New Jersey, New York, Maryland, and Washington states with RR in the 1970s and early 1980s. These studies looked at costs or changes in costs in these states, or in hospitals within these states. The typical method estimated a regression like:

$$\log C_t - \log C_{t-1} = b(X_t - X_{t-1}) + dD + \varepsilon$$

where the left-hand side reflects the change from year to year in the logarithm of hospital expenses (or the logarithm of the percent change in expenses) as a function of changes in explanatory variables,  $X$ , and imposition of rate regulation,  $D$ . If coefficient  $d$  is negative, then regulation  $D$  reduced the growth in costs.

Dranove and Cone (1985) seek to explain (1) expenses per admission, (2) expenses per day, and (3) expenses per person. They find that those states with regulation laws exhibited approximately 1.32 percent smaller increases in expenses per admission than other states, 1.41 percent smaller increases in expenses per day, and 1.04 percent smaller increases in expenses per person.

**CERTIFICATE-OF-NEED** CON legislation continues in at least 35 states (plus the District of Columbia and Puerto Rico), so it is important to consider its impacts. The arguments in favor of CON laws focus on three areas: (1) control of costs, especially unneeded capital costs; (2) assurance of quality for selected services; and (3) maintenance of access, particularly for underserved populations.

Conover and Sloan (1998) assess the impact of CON regulation on hospital spending per capita, hospital supply, diffusion of technology, and hospital industry organization for the years 1976 through 1993. Long-term CON programs were associated with what they termed a “modest (5 percent)” long-term reduction in acute care spending per capita, but not with significant reduction in total spending per capita. The authors doubt that CON regulations have had much effect (either positive or negative) on quality of care.

Nonetheless CON has continued, and Dobson and colleagues (2007) review its impacts in a report to the state of Illinois. They examine other states’ experiences to consider the risks and benefits of terminating CON programs and conclude that the evidence on cost containment is weak, but that the CON process does affect spending patterns. Expecting the CON process to reduce overall expenditures, they assert, is unrealistic.

Do CON laws increase quality of care? Dobson et al. note that even CON’s strongest supporters agree that the range of activities where CON can directly influence quality is narrow. Research shows a positive correlation between volume and quality in certain tertiary procedures such as cardiac surgery and transplant programs. Restricting new services certainly leads to fewer providers to perform a given number of procedures. However the impact of CON laws on quality and care is limited.

The remaining argument, maintenance of access, particularly for the underserved, deserves careful consideration. Community hospitals and academic medical centers that, by virtue of their location and/or reputation, are able to maintain a high proportion of well-insured patients tend to fare very well financially, and those who cannot are at risk of failure.

Competition can compromise the financial health of safety-net hospitals. For providers who may be struggling to survive already, competitive pressures could lead to failure. This failure could force the remaining providers to serve larger numbers of less profitable patients, which could lead to failures, potentially starting in the inner city and radiating out to more distant areas and rural communities. Dobson et al. note that CON laws in Illinois and other states have helped to protect those hospitals. They argue that the greatest effect that CON laws have is to delay the

shift of relatively profitable services from the inner city into the suburbs. They find no evidence that safety-net hospitals are financially stronger in CON states than other states.

## PROSPECTIVE PAYMENT

The 1965 enactment of Medicare and Medicaid brought the federal government into the reimbursement of services. Program costs have increased substantially over the years, and as costs have grown, so has interest in cost-containment policies.

Hospital payments and physician payments comprise the two major cost centers. Hospitals and physicians often bill patients separately and receive payment separately. We discuss hospital payment in this section, and we will look at particular physician payment later on in the chapter.

Here, we focus on the Medicare program that converted the financing of Medicare hospital care to a prospective payment system based on Diagnosis Related Groups (DRGs). We discuss the system in the present context because its widespread use stems from its adoption by the Medicare program. However, the PPS under DRGs has wider significance to other segments of the health sector, as well as to the regulation of industry in general.

### Description of PPS

Contrast the current prospective payment to Medicare's previous retrospective reimbursement system. Under retrospective payment, a hospital submitted its bill to Medicare after the care had been provided and the costs to the hospital were known. Retrospective payment allowed the hospitals to recover their expenses as allowed by Medicare rules regardless of whether these expenses were high or low, excessive or efficient. Retrospective reimbursement provided only modest incentives for hospitals to control costs.

Prospective payment sets payment rates *prior* to the period for which care is given. By setting a fixed reimbursement level per admission, prospective payment provides economic incentives to conserve on the use of input resources. Hospitals that use more resources than covered by the flat rate lose the difference. Those with costs below that rate retain the difference.

We can group Medicare's payment systems as follows:

1. Inpatient acute care in short-term hospitals and psychiatric facilities
2. Ambulatory care furnished by physicians, hospital outpatient departments, ambulatory surgical centers, and clinical laboratories
3. Post-acute care furnished by skilled nursing facilities, home health agencies, inpatient rehabilitation facilities, and long-term-care hospitals
4. Dialysis services furnished in outpatient centers and hospice care
5. Ambulance services and products furnished by durable medical equipment suppliers
6. Services furnished by private health plans under the Medicare Advantage program

DRGs define one of the primary mechanisms for prospective payment. The years 2007 and 2008 marked major changes in the definition and usage of DRGs. Previously DRGs reflected hospital charges. The new policies aimed to:

- Improve the accuracy of Medicare's inpatient hospital payments by using hospital costs rather than charges to set rates
- Adjust payment to recognize better the severity of illness and the cost of treating Medicare patients by increasing payment for some services and decreasing payment for others
- Eliminate biases that had provided incentives for physician-owned specialty hospitals to treat the healthiest and most profitable cases, leaving the sickest and least profitable patients to general acute care hospitals
- Refine the payment system to ensure that hospitals are provided with incentives to invest in service areas based on the clinical needs of their patients rather than financial incentives

**TABLE 20-1 Fiscal Year 2010 DRGS by Weight—5 highest and 5 lowest**

MS-DRG	TYPE	MS-DRG Title	Weights	Geometric mean LOS <sup>1</sup>	Arithmetic mean LOS
Five Highest					
001	Surgical	Heart transplant or implant of heart assist system w MCC	24.8548	31.5	43.9
003	Surgical	Ecmo or trach w MV 96+ hrs or PDX exc face, mouth & neck w maj O.R.	18.2667	31.6	38.5
927	Surgical	Extensive burns or full thickness burns w MV 96+ hrs w skin graft	13.7351	24.4	32.7
215	Surgical	Other heart assist system implant	12.8304	7.0	14.4
002	Surgical	Heart transplant or implant of heart assist system w/o MCC	11.7540	16.4	21.2
Five Lowest					
779	Medical	Abortion w/o D&C	0.4386	1.6	2.1
778	Medical	Threatened abortion	0.4229	1.9	3.0
894	Medical	Alcohol/drug abuse or dependence, left AMA	0.4021	2.1	2.9
780	Medical	False labor	0.2023	1.2	1.3
795	Medical	Normal newborn	0.1617	0.0	0.0

<sup>1</sup> The *geometric mean* or average is an alternative measure of central tendency, the  $n^{\text{th}}$  root of the product of the observations. For example, the arithmetic mean of observations  $a$  and  $b$  is  $(a + b)/2$ , while the geometric average is  $\sqrt{ab}$ .

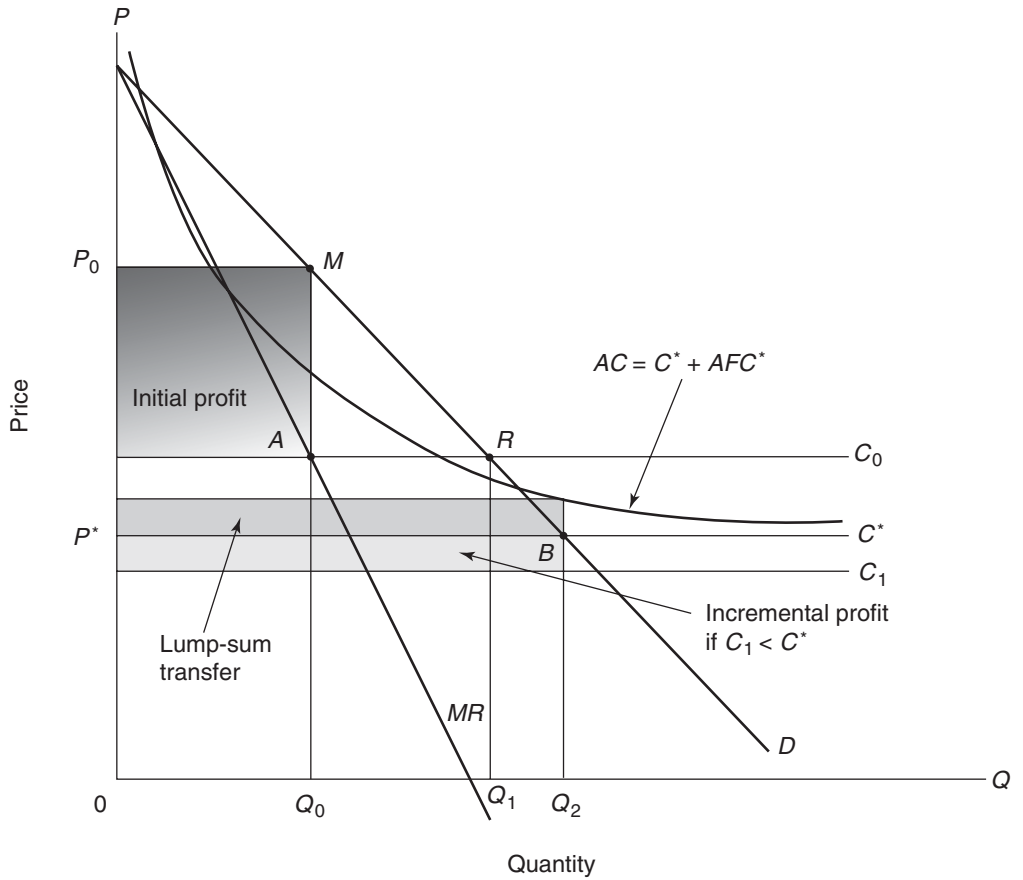
As of 2007–2008, CMS adopted a new set of 745 Medicare Severity Long-Term Care Diagnostic Related Groups (MS-DRGs) that replaced the existing 538 DRGs with ones that better recognized illness severity.

As before, each DRG has a flat payment weight calculated on the basis of costs incurred for that DRG nationally. For example, based on Table 20-1, DRG 1, for a heart transplant with a larger weight, is about 36 percent more costly ( $24.8548/18.2667$ ) than DRG 3, the use of an artificial lung (membrane) located outside the body (extracorporeal). It is almost 154 times as costly as a normal newborn birth (0.1617).

McClellan (1997) notes that DRGs are very complex, like the production systems that they regulate, and that some features of the DRG system do allow reimbursement to vary with actual treatment decisions during an admission. Moreover, a look at alternative methods available on the Internet for calculating hospital specific DRG adjusted payments shows how truly complicated the procedure is, with adjustments for differences in hospital labor costs, disproportionate shares of low-income clients, and “hold harmless” clauses (reflecting prior practices). Most importantly, the rates are flat in the sense that they do not change for hospitals that spend more than the rate or, for that matter, less.

## THE THEORY OF YARDSTICK COMPETITION AND DRGS

Shleifer (1985) describes the theory of yardstick competition, a close approximation to PPS under DRGs. We can think of yardstick competition as the ideal form, while the actual Medicare payment system is a real-life approximation. Shleifer considers markets where firms have monopoly power. Most medical providers face downward-sloping demand curves so they possess some degree of monopoly power.



**FIGURE 20-1** The Impact of Yardstick Competition

Recall that monopolistic firms reduce outputs and correspondingly charge higher prices, compared with competitive firms. Moreover, many analysts believe that secure monopolistic firms, without the market discipline of competition, may lack some of the incentives to be cost-efficient, that rigorous competition would provide. In U.S. hospital markets prior to PPS, one of the most serious cost-efficiency disincentives was the retrospective payment system already discussed. Shleifer’s yardstick competition describes a regulatory scheme, much like Medicare’s PPS, that restores cost-consciousness incentives.

Consider Figure 20-1 in which the hospital faces demand curve  $D$ , and where marginal and average costs equal  $C_0$ . By equating marginal revenue and marginal cost, the monopoly hospital will provide  $Q_0$  units of output and charge  $P_0$  indicated by point  $M$ , with an initial economic profit as indicated in the shaded box. If instead, the hospital received reimbursement prospectively at a rate of marginal cost  $C_0$ , the hospital would produce quantity  $Q_1$  at the intersection of demand  $D$  with  $C_0$ , at point  $R$ .

Suppose the hospital believes that it could produce at lower cost if it would hire a team (a *fixed cost* to the hospital because it is unrelated to output) of efficiency experts and carry out its advice (also a fixed cost). Recall that efficiency requires a firm to produce a quantity at which marginal cost (value in production) equals the market price (value in consumption). The problem for those who design a yardstick competition mechanism is to set up a payment scheme so that these firms have the incentives to spend just the right amount of money and effort on reducing production costs.

What is just the right amount of expense to incur in the effort to reduce production costs? Suppose the hospital discovers that it could reduce its marginal costs of production by \$100 per case

treated if it were to spend \$150,000 on cost-reduction efforts, and that the hospital treats 5,000 cases per year. Would the cost-reduction effort be worth it? Consider the first line of the following schedule:

	Marginal Cost of Efficiency Effort		Marginal Revenue Generated
Step (1)	\$150,000	<	$\$100 \times 5,000 = \$500,000$
Step (2)	\$270,000	<	$\$100 \times 5,000$
Step (3)	\$490,000	<	$\$100 \times 5,000$
Step (4)	\$750,000	>	$\$100 \times 5,000$

Each step will reduce costs per case by \$100 and step (1) represents the first of four possible steps. With step (1), reducing per-unit costs by \$100 costs a total of \$150,000 in cost-efficiency efforts. The extra \$1 saved for every case treated generates \$500,000 in extra revenue.

This step of cost saving is worthwhile because it costs less (\$150,000) than it saves (\$500,000). By similar reasoning, one more step of cost-saving also would be worthwhile, costing \$270,000, but saving another \$500,000. Step (3) is worthwhile, too, but that is where we would stop. The problem, from society's viewpoint, is that if competition provides incentives to cut costs, unregulated monopolists or retrospectively reimbursed firms may not have the incentives to take these three steps. It would be good to design a payment system that would induce them to do so.

How does yardstick competition provide this inducement? Return to Figure 20-1. With yardstick competition, prices for hospital reimbursement are set beforehand at fixed rates. These rates—that is, these regulated prices—are set equal to the averages of the marginal costs of *all other hospitals* in the market.

Suppose the regulator “knows” the marginal cost of each hospital. Consider Hospital 1. Recall that the cost-reduction efforts represent a fixed cost to the hospital, and the average fixed cost ( $AFC$ , graphed in Figure 20-1) must be covered at least in part by additional revenues. Yardstick competition assigns Hospital 1 service price  $\bar{C}$ , equal to the average marginal cost of *all of its competitors*, Hospitals 2 through  $N$ , and lump-sum subsidy  $\bar{R}$  (dark shaded area), equal to the average efficiency investment of its competitors. With its revenue per patient now fixed prospectively, Hospital 1 has an incentive to invest in cost-cutting technologies or practices, although its optimal investment will not necessarily equal the lump-sum subsidy (see problem 5 in the study questions to examine this).

Recall what the “yardstick competition” does. It induces the hospital to reduce its marginal costs below  $C_0$  and to compete with the other hospitals by pricing its output at the “average” marginal cost of its competitors (but *not* its own marginal cost). We treat the cost-cutting innovation as a fixed cost for the hospital output. As a result, the hospital's new average cost curve  $AC$  is the sum of the (now-lower) marginal cost  $C^*$  and the optimum average fixed cost,  $AFC^*$ . The new output occurs where  $C^*$  equals the demand price, or point  $B$ . The optimal output becomes  $Q_2$ , and the price falls from  $P_0$  to  $P^*$ , which equals  $C^*$ .

Suppose that the hospital's new marginal cost  $C^*$  fortuitously equals  $\bar{C}$ . Under yardstick competition, it will lose an amount equal to the fixed costs of reducing its marginal costs, or the dark shaded area in Figure 20-1. The lump-sum subsidy to the hospital, will defray some or all of the fixed costs required for its cost-reducing efforts.

If the hospital has succeeded in lowering its marginal costs to  $C_1$ , which is less than  $\bar{C}$ , then it will earn a profit equal to  $Q_1(\bar{C} - C_1)$ , the lightly shaded area. If the hospital cannot lower its costs as low as  $\bar{C}$ , then it may lose money if the lump sum transfers do not fully defray the excess costs per unit. In either case, because the hospital's actual costs *do not* enter into the price that it receives, and because the hospital can earn a profit if it reduces costs, the hospital has a considerable incentive to reduce its costs.

In this yardstick model, hospitals are not provided additional subsidies if they cannot reduce their costs to  $\bar{C}$  (that is, subsidized for inefficient behavior). Under the yardstick mechanism, hospitals

compete, and this competition leads to an optimum solution discussed earlier. The competition, known as a “Nash equilibrium” (after Nobel Laureate John Nash), refers to a market solution in which each firm does the best that it can, given the decision of others. It is an *equilibrium* because once the choices are made, no firm has any motive to change its action. Shleifer views it as

. . . essential for the regulator to commit himself not to pay attention to the firms’ complaints and to be prepared to let the firms go bankrupt if they choose inefficient cost levels. Unless the regulator can credibly threaten to make inefficient firms lose money . . . cost reduction cannot be enforced. (p. 327)

The model as presented implicitly assumes that all hospitals are the same but Shleifer shows how to construct a statistical multiple regression that would generate a solution in the more realistic case where hospitals differ from each other. Furthermore, he demonstrates a closely related scheme of pricing based on average industry costs that leads to virtually the same results in the realistic case where the theoretically needed lump-sum subsidies would be difficult to calculate or difficult to pass politically.

Returning now to DRGs as practiced, consistent with Shleifer’s formulation a hospital’s actual costs do not enter into its formula for payment rates. As a result, hospitals must become price takers in the strictest sense. If it costs the hospital more to provide the service than the DRGs allow, the hospital either loses money on the service and is forced to stop offering it, or cross-subsidizes the service from other services that may be produced at costs lower than their DRGs. Thus, the cost-cutting incentives are strong.

Although yardstick competition applies to the hospital, many of the costs reflect orders from physicians who are generally not even their employees. Box 20-1 provides an example of how physicians and hospitals might align incentives.

How accurate are the yardstick payments generated for DRGs? Although designed to elicit cost-cutting behavior, DRGs would be expected on average to equal provider costs. Ginsburg and Grossman (2005) report that Medicare and other payers have found it difficult to devise payment

## BOX 20-1

### The Alignment of Provider and Hospital Incentives

According to physician Kathryn Stewart, most physicians do not connect their local hospital’s financial position and their everyday actions in managing inpatients. They hear hospital executives “fussing” about saving money and assume it is the hospital’s problem, not theirs. Dr. Stewart offers several ways for physicians to help their hospitals address costs. These include:

1. **Change your mind-set about the use of acute-care hospitals.** Acute-care (inpatient) hospitals are for patients who are unstable in terms of vital signs or mental status, who require complicated surgery or procedures, who might suffer death or severe morbidity without 24-hour-a-day monitoring, or who require a rapid diagnosis to prevent death or severe morbidity. Problems that are less severe can be handled on an outpatient basis.
2. **Don’t include the kitchen sink.** When a patient must be admitted to the hospital, make every attempt to limit care to the reason for admission. If another, unexpected problem surfaces while the patient is in the hospital, ask, “Would I admit this patient to the hospital for this problem if I saw him or her in the office?” If the answer is no, then don’t treat the problem while the patient is in the hospital.
3. **Use other levels of care.** Patients can receive physical and occupational therapy, get X-rays and lab work, and even receive intravenous (IV) drugs and total parenteral (through a vein) nutrition at home or in skilled-care nursing homes. Once stabilized and with a treatment plan, the patient can be moved out of the acute-care hospital to another place to finish treatment. It is simply too expensive today to use acute-care hospitals for these purposes.

**4. Bring costs down.** General rules include:

- *Pay attention to medication costs.* If a medication can be given by mouth (e.g., it is formulated that way and the patient has a functional gastrointestinal tract), then give it by mouth as soon as possible. Many drugs work as well or better that way. Intravenous drugs are expensive so change from IV to oral as soon as possible, but don't simply choose the cheapest drug on the market. A once-a-day IV antibiotic may be less expensive than a cheaper twice-a-day antibiotic when staff costs of administering the drugs have been considered.
- *Don't do a test in the hospital that will not be used to make decisions about the rest of the patient's stay.* If a test is not directly related to the reason for admission and it will not change the course of the hospital diagnosis and treatment plan, don't do it in the hospital.
- *Remember, most hospitals count their census at midnight.* If the patient is in that bed at midnight, you have just added another day to the length of stay. Consider sending the patient home tonight unless there is something critical to be done in the morning.

Aligning provider and hospital incentives saves money. A careful examination of the recommendations suggests that it may also improve quality of care!

*Source:* Based on Kathryn Stewart, "Seven Ways to Help Your Hospital Stay in Business," [www.aafp.org/fpm/20030500/27seve.html](http://www.aafp.org/fpm/20030500/27seve.html), accessed July 23, 2008.

rates that closely follow relative costs. Medicare prospective payment for inpatient care, based on DRGs, determines the relative payment for each DRG on the basis of average charges for patients in that DRG across all hospitals. To the degree that the pattern of charges systematically diverges from the pattern of unit costs, then relative payments for different DRGs will not reflect relative costs. As a result, patients in some DRGs are more profitable than others.

Newer models of equipment are often more productive or have lower costs than what they replace. For example, recent advances in CT scanning equipment have reduced the time to do a scan, meaning that a machine and technician staff can perform more scans per day. Similarly, the physician time required to perform many cardiovascular surgical procedures has fallen over time as technology has evolved. In addition, with new services provided to more patients, equipment and dedicated space and personnel run at higher capacity, another factor behind productivity gains over time. For supplies such as prepackaged instruments and supplies, entry by additional manufacturers leads to lower prices. These cost declines often mean that payment rates soon become out of date, leading to profit margins that are higher than those for services not experiencing these trends.

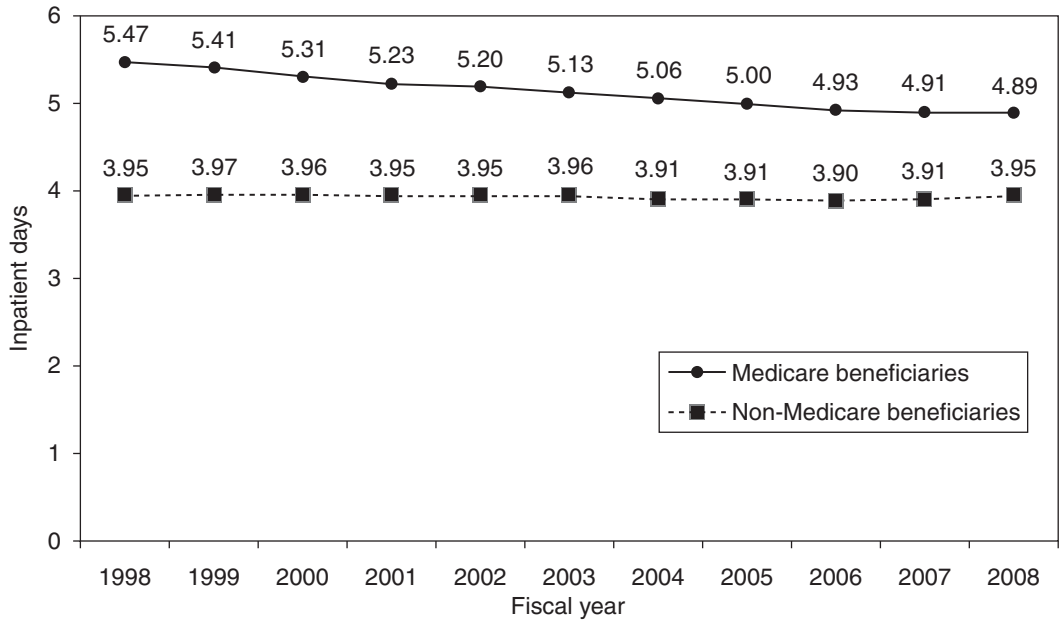
### **On the Effects of Medicare's Prospective Payment System**

Yardstick competition in theory and the actual Medicare PPS in practice change firms' economic incentives considerably. The Medicare PPS began in 1983, and studies over the next 10 to 15 years compared outcomes with what might have occurred in the absence of PPS. Expected length of stay began to fall and continued falling. Studies on access to care showed mixed results, there was no definitive evidence that quality of care had decreased. At least at the beginning, hospital fixed-payment rates proved sufficiently generous that average operating margins of hospitals were substantial. The next section concentrates on more recent studies of the performance of the hospital sector under PPS.

**AVERAGE HOSPITAL LENGTH OF STAY** Decreases in length of stay have been pervasive and long lasting. MedPAC data tracing hospital length of stay have shown continual decreases. Length of stay for all hospitals fell by 32.2 percent from 1989 to 1999, with annual declines exceeding 5 percent from 1993 to 1996.

As noted in Figure 20-2, length of stay for Medicare inpatients was nearly one day longer than non-Medicare patients in 2008. Length of stay for Medicare inpatients had continued to fall, from





**FIGURE 20-2** Trends in Medicare and Non-Medicare Lengths of Stay, 1998–2008

Source: MedPAC Databook 2010, Chart 7-9.

5.47 to 4.89 days, or by 10.6 percent, from 1998 to 2008. It is important to note that because inpatient stays cannot fall below one day, further large decreases in length-of-stay will be hard to achieve. It is clear, however, that prospective payment has effected major reductions in length-of-stay since its inception.

**ACCESS AND QUALITY** MedPAC (2000, p. 61) reported a net reduction of 340 short-term acute-care hospitals (440 closing; 100 opening or re-opening) from 1990 through 1999. The capacity reduction was much smaller than the 6.5 percent decline in hospitals because the closed hospitals were generally small and had low occupancy rates and very low volumes. The pattern appeared to stabilize between 2000 and 2008 with a net increase of 164 hospitals (368 closing; 532 opening or re-opening).

Further MedPAC analyses report generally positive results regarding quality of care. Table 20-2 shows that for all of the conditions noted, mortality measures improved, most of them significantly. For example, 30-day post discharge mortality for stroke fell from 22.96 per 100 eligible discharges in 2005 to 19.98 in 2008. Even those marked “no difference” show movement in the right direction, although not at the (very conservative) 1 percent significance level.

In contrast, Table 20-3 shows safety indicators to have mixed trends. While death from treatable serious complications fell significantly, three others were unchanged, and safety from postoperative respiratory failure and pulmonary embolism and/or deep-vein thrombosis worsened.

**SEVERITY OF THE CASE MIX** Another observed change following the onset of Medicare’s PPS has been the increase in common case-mix indexes. A case-mix index is a numerical measure of the assortment of patient cases treated by a given hospital so that a higher index value indicates a greater average degree of complexity of the cases and consequently a greater need for input resources. Some of the increase in the indexes has probably occurred because hospitals are being more careful about which diagnosis they choose (for patients with multiple

**TABLE 20-2 Most hospital inpatient and 30-day mortality rates improved from 2005 to 2008**

Condition or procedure	Risk-adjusted rate per 100 eligible discharges, 2005	Risk-adjusted rate per 100 eligible discharges, 2008	Directional change in rate, 2005-2008
<b>In-hospital mortality</b>			
Acute myocardial infarction	9.51	7.36	Better
Congestive heart failure	4.21	3.10	Better
Stroke	11.42	9.14	Better
Hip fracture	3.22	2.31	Better
Pneumonia	5.17	3.93	Better
Esophageal resection	8.21	3.73	No difference
Pancreatic resection	6.36	5.15	No difference
Abdominal aortic aneurysm repair	7.45	6.07	Better
<b>30-day postdischarge mortality</b>			
Acute myocardial infarction	15.37	12.83	Better
Congestive heart failure	9.95	8.26	Better
Stroke	22.96	19.98	Better
Hip fracture	8.71	6.81	Better
Pneumonia	10.97	8.86	Better
Esophageal resection	10.13	7.42	No difference
Pancreatic resection	8.49	6.73	No difference
Abdominal aortic aneurysm repair	8.93	8.22	No difference

*Note:* Rates are calculated based on the discharges eligible to be counted in each measure. Rates do not include deaths in non-inpatient prospective payment system hospitals or Medicare Advantage plans. “Better” indicates that the risk-adjusted rate decreased by a statistically significant amount from 2005 to 2008 using a  $p \leq 0.01$  criterion. “No difference” indicates that the change in the rate was not statistically significant from 2005 to 2008 using a  $p \leq 0.01$  criterion.

*Source:* MedPAC analysis of CMS Medicare Provider Analysis and Review data using Agency for Healthcare Research and Quality Inpatient Quality Indicators Version 3.2.

diagnoses) to categorize the patient for Medicare payment purposes. This maneuver is sometimes called “DRG creep.”

Many analysts have expressed concern that the PPS under DRGs would induce hospitals to save on costs by cutting corners, reducing quality, refusing to treat costly patients, or closing treatment units. That is, the system would reduce access to care and the quality of care. While early studies found little statistical impact, hospital treatment practice patterns clearly changed following PPS. Also, average severity of the cases treated on an inpatient basis in hospitals increased subsequent to PPS suggesting that less severe cases now use outpatient treatment.

Silverman and Skinner (2004) revisit the DRG creep issue. They found that between 1989 and 1996, the percentage point share of the most generous DRG for pneumonia and respiratory infections rose by 10 points among not-for-profit hospitals, 23 points among for-profit hospitals, and 37 points in hospitals converting to for-profit status. After 1996, however, there was a dramatic decline in upcoding ratios, with the greatest drop among for-profit hospitals.

Why did this occur? Differential upcoding by for-profits was not because of sicker patients or attenuation in “down-coding” caused by more efficient billing systems. A variety of other audits conducted by the government and by nongovernment researchers have found ex post that chart data could not support the severity of disease actually billed for. Upcoding took the form of presumptive diagnosis; while there was no documentation that the patient actually had the disease, it could not definitely be ruled out.

**TABLE 20-3 Trends in hospital patient safety indicators are mixed from 2005 to 2008**

Patient safety indicator	Risk-adjusted rate per 100 eligible discharges, 2005	Risk-adjusted rate per 100 eligible discharges, 2008	Directional change in rate, 2005-2008
Postoperative PE or DVT	0.80	0.95	Worse
Accidental puncture or laceration	0.38	0.41	No difference
Postoperative respiratory failure	1.12	1.29	Worse
Iatrogenic pneumothorax	0.05	0.05	No difference
Death among surgical inpatients with treatable serious complications	13.15	10.75	Better
Postoperative wound dehiscence	0.24	0.29	No difference

*Note:* PE (pulmonary embolism), DVT (deep vein thrombosis). “Better” indicates that the risk-adjusted rate decreased by a statistically significant amount from 2005 to 2008 using a  $p \leq 0.01$  criterion. “Worse” indicates that the risk-adjusted rate increased by a statistically significant amount from 2005 to 2008 using a  $p \leq 0.01$  criterion. “No difference” indicates that the change in the rate from 2005 to 2008 was not statistically significant using a  $p \leq 0.01$  criterion.

*Source:* MedPAC analysis of CMS Medicare Provider Analysis and Review data using Agency for Healthcare Research and Quality Patient Safety Indicators Version 3.2.

In explaining the higher propensity of for-profits to upcode, the authors appeal to Pauly and Redisch’s (1973) distinction, as noted in Chapter 13, between for-profit hospitals, where physicians either share residual hospital revenue or received salaries, and not-for-profit hospitals, where “no administrator can afford to incur the displeasure of the medical staff . . .” (p. 89). For upcoding to occur, administrators must be willing to engage in ex ante risky but potentially profitable behavior, and physicians must acquiesce by approving (and until 1995, signing) the DRG claims submitted by the hospital to Medicare.

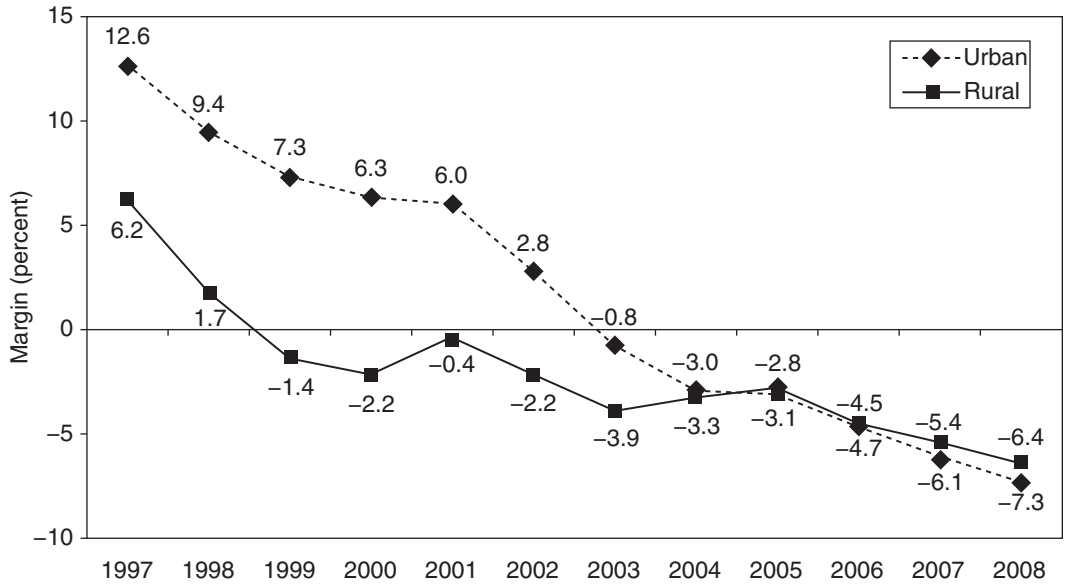
**HOSPITALS’ FINANCIAL CONDITION** Most analysts agree that PPS can financially stress hospitals, particularly on those with higher costs. This, after all, is the way in which PPS alters financial incentives. However, hospitals run on the principles of yardstick competition might well increase their wealth positions subsequent to PPS. In any case, at the beginning of Medicare’s PPS, the fixed-payment rates proved sufficiently generous that average operating margins of hospitals were substantial, thus reducing the cost-saving incentives. Payment rates did not subsequently keep up with inflation, perhaps due to congressional concern about windfall profits, and operating margins fell steadily. By the late 1980s and extending through the early 1990s, average operating margins for the Medicare segment of patients tended to be negative.

Analysts define hospital margins as revenue minus costs, divided by revenue. Overall Medicare margins cover the costs and payments of acute hospital inpatient, outpatient, inpatient psychiatric and rehabilitation unit, skilled nursing facility, and home health services, as well as graduate medical education and bad debts.

Figure 20-3 shows that overall Medicare margins have historically been higher for urban hospitals than for rural hospitals. The difference in margins between the two groups grew between 1997 and 2000 but has since narrowed, with rural hospital margins almost identical to those of urban hospitals since 2004. Policy changes made in the Medicare Prescription Drug, Improvement, and Modernization Act of 2003 targeted to rural hospitals helped to narrow the difference in overall Medicare margins between urban and rural hospitals. Further changes aimed at helping rural hospitals occurred in 2008.

We note, however, that overall hospital margins are not negative. The total hospital margin for all payers, Medicare, Medicaid, other government and private payers, reflects the relationship of all hospital revenues to all hospital costs, including inpatient, outpatient, post-acute, and non-patient services, and non-patient revenues such as investment revenues.

Figure 20-4 shows that the total hospital margin peaked in 1997 at 6.4 percent. From 2002 to 2007, total margins increased to the highest level in a decade. In 2008 the total margin declined to

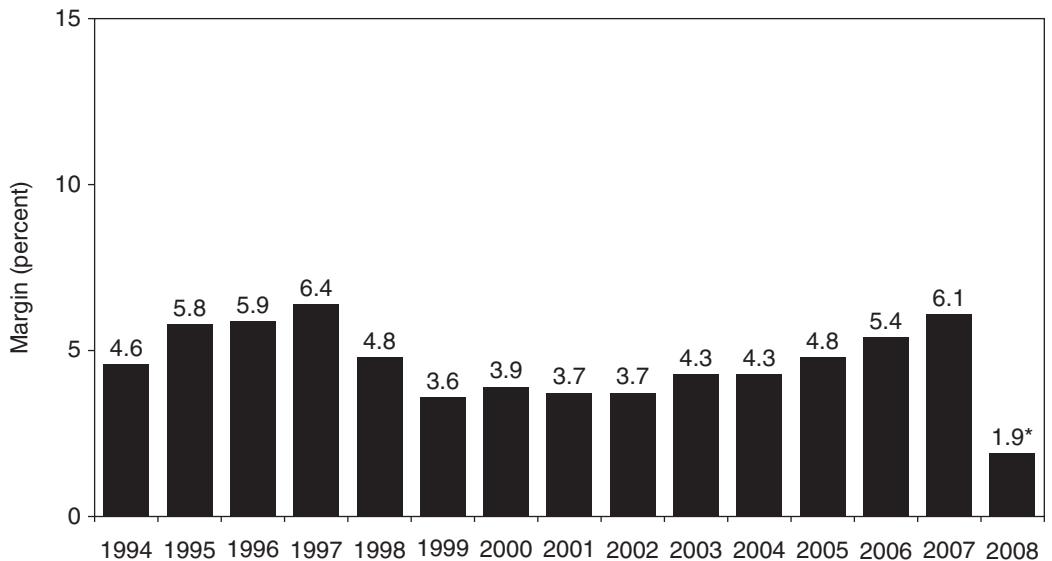


**FIGURE 20-3** Overall Medicare Margin, by urban and rural location 1997–2008

Source: MedPAC Databook, 2010, Chart 7-15.

1.9 percent, but this stemmed largely from investment losses due to the precipitous decline in the stock market. Operating margins did not fall in the same manner in 2008.

**EARLY EVIDENCE OF EFFECTS ON COSTS** Medicare’s goal was to control costs. The ultimate question, then, is whether PPS succeeded, and we can address it from at least three perspectives. First, did individual hospitals respond by reducing their expenditures? Evidence indicates that they did so, at least in the first year. Zwanziger and Melnick (1988) observed that for a sample of California hospitals, those under the strongest pressure from PPS responded by reducing expenditures.



**FIGURE 20-4** Hospital total all payer margin, 1994–2008

Source: MedPac Databook, 2010, Chart 7-16.

Second, did PPS reduce expenditures in Medicare's budget? Hospital savings seem likely. Medicare outlays for hospital care increased at an annual rate of 3.5 percent from 1984 to 1987, compared to 7.1 percent from 1977 to 1982, and hospital care fell as a proportion of the Medicare budget from 70 percent to 57 percent by 1988. Subsequently, PPS payments per case rose only slightly faster than the consumer price index so that by 1993, after the first 10 years of PPS, Medicare's share of hospital payments remained at 28 percent despite a 23 percent growth in Medicare enrollments.

Lastly, did per-capita expenditures for society as a whole fall as a result of Medicare's PPS? This is a much more difficult question to answer. Feinglass and Holloway (1991) argue that the savings from PPS were one-time-only events, occurring in the early years of PPS. This possibility is evidenced by increasing late 1980s inflation rates for U.S. health care expenditures, which in real terms outstripped the inflation rates of 1980 to 1985.

In a statistical study of the issue, Sloan, Morrissey, and Valvona (1988a) find a significant reduction in hospital patient-generated revenues per capita subsequent to PPS, but they note that the savings stem primarily from reductions in admissions, a category PPS does not provide specific incentives to reduce. They note the possibility that utilization review under peer review organizations (PROs)—a utilization review effort begun along with Medicare's PPS—could conceivably be the source of per capita cost reductions for hospital care. Shifts to outpatient care, which similarly do not provide specific cost-reduction incentives, also may occur.

**MORE RECENT EVIDENCE ON COSTS** Medicare developed PPS in part to address perceived cost increases in the health care sector, but early analyses indicated mixed effectiveness. We examine here analyses provided by MedPAC, rewriting costs per inpatient stay, indicated by discharge data, as the following equation:

$$(\text{cost/discharge}) = (\text{days/discharge}) \times (\text{service intensity/day}) \times (\text{cost/service unit})$$

So, a hospital stay that lasts eight days, with two treatments per day at a cost of \$500 per treatment, will cost:

$$(\text{cost/discharge}) = (8 \text{ days}) \times (2 \text{ treatments}) \times (\$500), \text{ or } \$8,000.$$

It is easy to apply this formula to changes over time. Some manipulation leads to:

$$\begin{aligned} \% \text{ Change in } (\text{cost/discharge}) &= \% \text{ Change in } (\text{days/discharge}) + \% \text{ Change in } (\text{services/day}) \\ &+ \% \text{ Change in } (\text{cost/service}) \end{aligned}$$

Table 20-4 provides results of the formula.

We compare cost changes to general price inflation, so the table suggests that from 1985 to 1989, for example, costs increased by 2.5 percent more per year than goods outside the hospital sector. During that period, the most important variable was length-of-stay reduction, which went from an increasing factor (0.8 percent per year) to substantial decreases of 1.3 percent per year from 1989 to 1992 and 3.3 percent per year from 1992 to 1996. This was somewhat offset by increased services per day, although overall services per discharge (the sum of percentage changes in days/discharge and services/day) appear to have dropped.

One of the hospital industry's key tools for cutting costs was reducing length of stay, largely by discharging patients to various forms of post-acute care earlier in their stays. But they also cut costs in other ways, such as reducing staffing levels, providing smaller increases in compensation, and substituting less-skilled workers (such as nurse aides for RNs).

**TABLE 20-4 Average Annual Hospital Cost Changes for Medicare Inpatient Services**

Years	% Change Costs per Discharge	% Change Days per Discharge	Services per day	Costs per Service
1985 to 1989	2.5	0.8	2.2	-0.3
1989 to 1992	1.3	-1.3	1.4	1.3
1992 to 1996	-2.7	-3.3	0.9	-0.3
1996	-1.2	-5.9	u	u
1997	-0.6	-4.7	u	u
1998	1.1	-3.3	u	u
1999	2.6	0.0	u	u
2000	2.4	-1.7	u	u
2001	5.1	-1.7	u	u
2002	8.1	0.0	u	u
2003	6.6	-1.8	u	u
2004	5.6	-1.8	u	u
2005	5.1	-1.8	u	u

u, unavailable

Sources: MedPAC (March 1999), Appendix D, Table D-1, MedPAC DataBook 2007, Charts 7-5, 7-10, and 7-22.

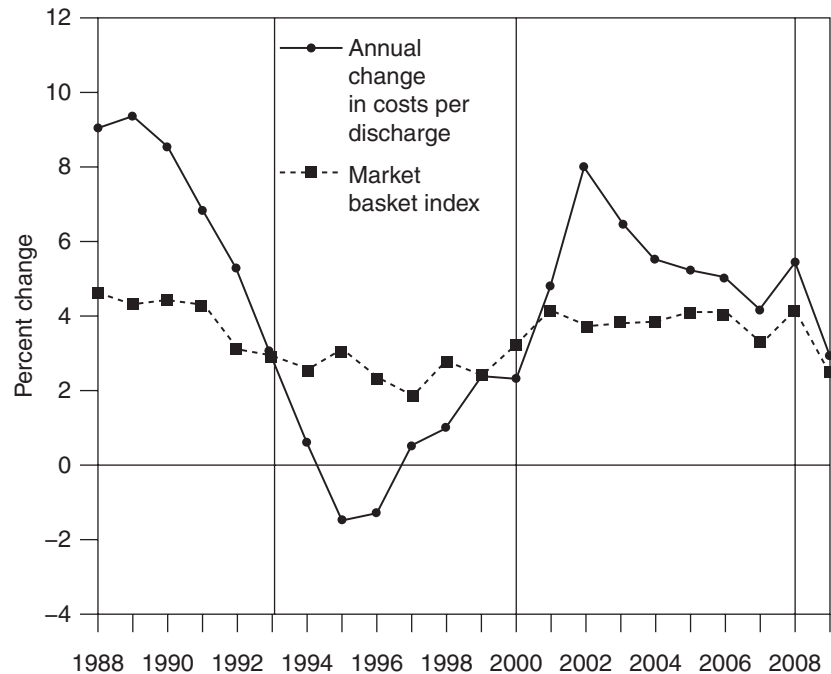
In 2000, private payer payments to hospitals once again began rising faster than costs. The payment-to-cost ratio for private payers rose 1.3 percentage points per year through 2003—almost as rapidly as in the late 1980s—and hospitals’ profits from privately insured patients rose by 6 percentage points. Health plans continue to negotiate prices with hospitals, but many providers gained the upper hand in these negotiations. Payers’ primary leverage in price negotiations is the threat of selective contracting, but their use of this tool was limited by both hospital consolidation and consumers’ reluctance to accept limitations on their choice of providers, preferences for so-called “managed care light.”

Figure 20-5 shows how annual cost increases exceeded the market basket index of inputs used to deliver care starting in 2001. The Medicare trustees note that with high hospital margins of the middle years of the first decade, hospitals felt little cost pressure. In response, however, to heavy stock market losses in late 2008, hospitals pulled back from the high levels of capital expenditures and employment growth seen in 2007 and 2008 to more moderate levels of capital expenditures and employment growth. The result was the drop in cost growth between 2008 and 2009 from 5.5 percent to a more moderate 3.0 percent. However, the trustees believe that if hospitals’ financial condition continues to improve and their expectation of future revenue growth does not decline, they expect to see increased cost growth in 2011 and thereafter.

## REGULATION OF PHYSICIAN PAYMENT

Physician payment constitutes a second major cost center in the U.S. system. U.S. physicians traditionally received payment on a fee-for-service (FFS) basis. Under FFS, each physician has a schedule of fees and expects reimbursement for each unit of service provided.

As commercial insurance proliferated following World War II, insurers used benefit schedules to determine reimbursement. These schedules contained the maximum amounts to pay for a particular service. If the provider charged less, insurers would pay the lower amount. Typically, also, individuals paid the providers’ charges and were reimbursed (indemnified) by insurers according to the benefit schedule.



**FIGURE 20-5** Cost growth falls in 2009 as financial pressure increases

*Note:* The market basket index measures annual changes in the prices of the goods and services hospitals use to deliver care. Cost growth refers to annual change in inpatient allowable costs per discharge.

*Source:* Medicare Trustees Report, March 2011, Figure 3-6, p. 54

## UCR Reimbursement, Assignment, and Alternative Payment Mechanisms

Following the introduction of Medicare in 1965, another system for reimbursing providers became more prevalent. The usual, customary, and reasonable (UCR) payment mechanism uses fee screens that are determined by the maximum of the doctor's median charge (i.e., usual charge) and the customary fees charged by the other physicians in the area. In addition, higher charges may stand, if deemed reasonable in light of any unusual circumstances that would justify the higher payment. Under UCR, the maximum reimbursement is the lowest of the actual charge or the two fee screens.

The UCR method is far more complicated and difficult to administer than the benefit schedule because it involves data collection on every provider. Supporters of this approach argue, however, that it is flexible and fair because it allows for differences across different specialty groups and market areas. Once Medicare adopted the UCR, the Blues Cross insurers and many other third-party payers followed.

However, several concerns with UCR soon became apparent. Many critics argued that UCR is inflationary. Providers have incentives to raise their charges, thus increasing the usual and customary screens and resulting in increased future payments. Furthermore, there is no incentive to compete by charging less than the screens. To deal with this problem, Medicare and many Blue Shield plans introduced various mechanisms to limit the increases in the screens.

## Relative Value Scales

Relative value scales (RVSs) were developed by state medical societies and other organizations (the California RVS, developed in 1969, was one of the most widely used). An RVS establishes a weight (or multiple) for each procedure. The weight may reflect the time needed to perform a procedure and its complexity. Although the RVS does not provide fee information, if providers adhere to the guide, the price they set for any one procedure will determine the prices charged for all other procedures.

Any system of relative value scales depends on a set of given technologies. When, for example, the method of treating heart disease changes, the RVS for that category also must change in order to remain valid. If the technological advance causes the treatment to become cheaper, the treatment may become more profitable to the provider as long as the treatment retains its old place in the relative value scale.

Over the years, perceived inequities in Medicare's UCR system between primary care and surgical specialties led to attempts to change Medicare's physician payment system. Harvard University researchers, led by William Hsiao, undertook the task of developing a "resource-based" RVS, or RBRVS, as an alternative method of reimbursing physicians. Based on their work, a new Medicare payment system phased in over a five-year period beginning in 1992.

The RBRVS has three components, corresponding to three types of resources used to provide physicians' services:

1. *Physician work*, including the time, intensity of effort, skill, and risk to the patient associated with each service. On average, this accounts for 52 percent of the fee.
2. *Practice expenses*, including the cost of nonphysician staff, office space, equipment, and supplies. This accounts for 44 percent of the fee.
3. *Professional liability insurance* (malpractice) expense relative values amount to approximately 4% of the physician fee schedule payment. CMS replaced the cost-based professional liability insurance relative values with resource-based professional liability insurance RVUs in 2000. The end result was to retain the same total professional liability insurance RVUs as they were under the charge-based system.

*Source:* 2011 RBRVS, What is it and how does it affect Pediatrics? [www.aap.org/visit/rbrvsbrochure.pdf](http://www.aap.org/visit/rbrvsbrochure.pdf)

The system sought a major restructuring of fees away from specialists and toward primary care providers, through conversion factors that reflect resources used. The conversion factors have been updated under the so-called sustainable growth rate (SGR) system, in which CMS (Medicare and Medicaid) must target a rate of growth. Target expenditures for each year equal target expenditures from the previous year increased by the SGR. The formula comprises the following four factors: (1) the estimated percentage change in fees for physicians' services, (2) the estimated change in the average number of Medicare fee-for-service beneficiaries, (3) the estimated 10-year average annual growth in real gross domestic product (GDP) per capita, and (4) the estimated change in expenditures due to changes in law or regulations.

The SGR system addresses only input price inflation and productivity growth and, therefore, does not fully account for changes in the cost of providing physician services. Moreover, annual updates penalize spending above the target and reward spending below the target, thus leading to payments that diverge from the costs of efficiently provided care. Recent rapid growth in the volume and intensity of physicians' services per beneficiary is driving the growth in Medicare physician spending and resulting in the consistently negative annual updates. Physicians naturally oppose these negative updates (decreases in reimbursement) and threaten to stop accepting new Medicare patients if their fees are reduced.

As of early 2012, the impasse continues. The Trustees note (Trustees, 2011, P. 70) that the current law under the SGR system calls for Medicare to cut fees for physician and other health professional services by more than 30 percent over the next several years. They acknowledge that "dramatic" fee cuts would hinder beneficiary access to care, and that Congressional overrides of the SGR have successfully averted payment cuts in recent years.

The Trustees note that proposals to change the updates are quite costly (from a budgetary scoring perspective) because they eliminate some or all of the scheduled fee cuts. However, they recognize the "mounting frustration" of physicians, other health professionals, and their patients if substantial Medicare fee cuts continue to loom large in future years. They conclude that the Commission "plans to continue to work on SGR payment policies and consider various approaches



for updating Medicare’s physician fee schedule.” In short, physician compensation under Medicare in the U.S. remains a serious economic, medical, and political problem.

## ANTITRUST

We have addressed the issues of regulating monopolistic firms, as well as trying to make them act like competitors. Antitrust policy, a third form of regulation, attempts to keep monopolies from forming. A discussion of federal antitrust law in the United States begins with the Sherman Act, passed in 1890 following widespread perceptions of abuse of consumers in several industries as a result of trust arrangements. In a trust, a central board votes all the stock of the members of the trust. The best known, Standard Oil Trust, was able to monopolize the markets for crude and refined oil, achieving a 90 percent share by combining more than 100 firms into the trust.

The Sherman Act has two main sections. Section 1 specifies that “every contract, combination in the form of a trust or otherwise, or conspiracy, in restraint of trade . . . is declared to be illegal.” The main target of Section 1 is price fixing and market division. Under Section 2, “every person who shall monopolize, or attempt to monopolize, . . . shall be deemed guilty of a felony.”

Enforcement of the Sherman Act was a problem, and because specific antitrust practices were not stated, two additional actions by Congress strengthened the antitrust laws. The 1914 Clayton Act addressed practices such as price discrimination (when the price discrimination restrains trade), interlocking directorates, mergers, and joint ventures. The Federal Trade Commission Act, also passed in 1914, created an independent agency to help enforce the laws. The act also addressed “unfair” methods of competition and “unfair” practices, such as false advertising. The Robinson-Patman Act of 1936 modified the price discrimination provision of the Clayton Act. The 1950 Celler-Kefauver Act focused on vertical mergers (for example, when a firm merges with its input suppliers).

### Enforcement

Two agencies enforce the federal antitrust laws. The Antitrust Division of the Department of Justice (DOJ), formed in 1903, alone has the power to seek criminal charges. Shortly after its creation, it launched a major wave of “trust-busting,” culminating with the breakup of the Standard Oil and American Tobacco companies following court decisions in 1911.

The second agency, the Federal Trade Commission (FTC), is an independent agency with exclusive jurisdiction over the FTC Act. Because the two agencies are small, they are selective about the cases they pursue, seeking to deter anticompetitive behavior and to clarify the application of antitrust laws with precedent-setting reviews.

### Exemptions

In the United States, professional baseball and labor unions have received special exemptions from federal antitrust laws. For example, organized baseball (representing the Yankees and the Mets) can prevent a third major league team from entering the very profitable New York metropolitan market. Similarly, if the UAW-represented employees of General Motors, Chrysler, and Ford organize to take wages out of competition (set a standard union rate) and/or set up job demarcations (specify who does which work), without an antitrust exemption, the union would be considered a cartel. Many other industries are at least partially exempt. These include agriculture (which can form marketing co-ops), fishing, the many regulated industries, and national defense contractors.

There are some important limits relevant to the health industries. First, federal antitrust applies only to entities involved in interstate or foreign commerce. Many health care providers such as hospitals and physicians would seem to be immune from federal antitrust because their activities affect mainly local, within-state markets. However, the courts have ruled it sufficient to show that the seller is receiving substantial reimbursement from Medicare (or other out-of-state payers) or purchasing supplies from out-of-state vendors. As a result of this interpretation, there is a broad applicability of federal antitrust laws to health care markets (the relevant case is *Summit Health, Ltd. v. Pinhas*, 59 U.S.L.W.4493).

Nevertheless, two important exemptions, though not unique to health care, are particularly important in health care antitrust. The McCarran-Ferguson Act of 1945 exempts the insurance industry to the extent that individual states regulate the industry. With the monopsony power enjoyed by Blue Cross and Blue Shield in many states, a broad interpretation of this exemption would give the Blues and other insurers a free hand in exercising that power. The courts, however, have narrowed this exemption by carefully defining insurance activities and eliminating antitrust protection for the many noninsurance activities that may be undertaken by insurance companies.

The second exemption is the Noerr-Pennington doctrine. Through two decisions handed down by the Supreme Court, private parties and professional organizations have broad immunity from antitrust for lobbying activities, even where those lobbying activities lead to anticompetitive legislation. For example, lobbying efforts designed to tighten licensure requirements or limit the scope of practice allowed for other groups of providers could be protected under Noerr-Pennington even though the legislation reduces competition. This doctrine is especially important in health care as a result of the intense political lobbying by professional and trade associations. It is also worthwhile to note that the Noerr-Pennington doctrine is based on First Amendment constitutional law, which provides for guarantees to freedom of speech. The link to constitutional grounds gives the doctrine more force than is often the case regarding antitrust law. Professional advertising also receives First Amendment protection so that advertising freedoms have a force beyond antitrust and competitive grounds.

Because the basic premise of antitrust policy points toward competition with its result of pricing at marginal cost, it is useful to address a potential criticism that relates to health insurance. Figure 20-6 revisits the issue of insurance-induced moral hazard. The optimal health quantity occurs where demand  $D$  equals marginal cost  $MC$  or quantity  $Q^*$ . If coinsurance rate  $r$  lowers the price facing the consumers from  $P_2$  to  $rP_2$ , then the demand curve rotates clockwise to  $D_1$  and quantity demanded increases to  $Q_2$ , leading to a welfare loss equal to the sum of areas  $(A + B)$ . It would appear that if providers have monopoly power and charge a higher price, say  $P_1$ , consumers would demand less care, here  $Q_1$ , thus reducing the welfare loss from area  $(A + B)$  to area  $A$ . It would follow that forcing providers to charge the lower price  $P_2$ , corresponding to the marginal cost of providing care, would actually increase the welfare loss of excess consumption to area  $(A + B)$ .

Gaynor, Haas-Wilson, and Vogt (2000) expose a flaw in the “conventional” analysis, which assumes the coinsurance rate to be constant at all prices of care. A *competitive* insurance industry will adjust the coinsurance rate depending on the price of the medical services, rather than leave it fixed. Lower medical service prices will lead to lower coinsurance rates with both leading to increased purchases. Curve  $D_{true}$  represents the quantity demanded if insurance companies were to adjust coinsurance rates at all prices.  $D_{true}$  bows inward because decreases in price from  $P_1$  allow consumers increasingly larger amounts of services at lower out-of-pocket prices and coinsurance rates.

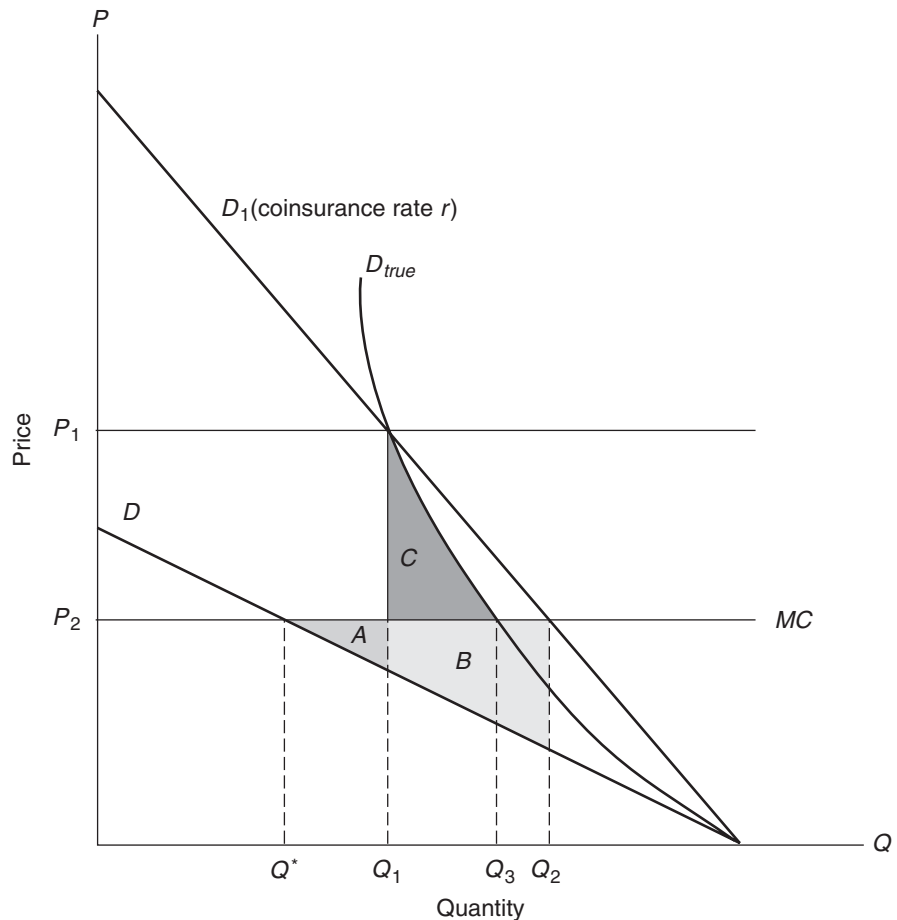
With this coinsurance rate flexibility, demand schedule  $D_1$  is not relevant for evaluating welfare, and areas  $A$  and  $B$  are not relevant to calculating welfare losses. Using  $D_{true}$ , decreasing price from  $P_1$  to marginal cost  $MC$  will result in welfare gain  $C$ . If insurers set the premiums competitively, reducing the price of medical services toward marginal cost improves consumers’ well-being!

### Measuring Monopoly Power

Perhaps the most natural measure of monopoly power, yet the most difficult to implement, is a direct comparison of the price to the marginal cost. Abba Lerner (1943) proposed the following measure,  $d$ , of monopoly power.

$$d = \frac{P - MC}{P} = 1 - \frac{MC}{P}$$

where  $P$  = market price and  $MC$  = marginal cost. Clearly, in the theoretical limit under perfect competition,  $d = 0$  because price equals marginal cost. Because  $MC$  is less than price when monopoly exists, then  $d$  will increase and will approach 1 if  $MC$  is low relative to price.



**FIGURE 20-6** Welfare Optimum in the Presence of Insurance and Regulation

Lerner's measure is elegant, and it captures the essence of monopolistic behavior and its economic inefficiencies very well. Unfortunately, it is difficult if not impossible in many cases to measure marginal cost, especially for the modern multiproduct firm with its complicated production processes. Consequently, economists also have developed other measures that may indicate market power.

**MEASURES OF CONCENTRATION** Chapter 17 introduced measures of concentration for the pharmaceutical industry. The concentration ratio describes the percentage of total industry sales accounted for by the largest firms in an industry. The logic behind this approach is the premise that the greater the market share held by a few firms, the more likely it is that firms are colluding. The concentration ratio itself does not indicate inefficiency (as does a gap between price and marginal cost), but high concentration ratios offer a greater possibility that firms collude to engage in anticompetitive behavior. The maximum value for a concentration ratio is 100 in the case of a monopoly, and often the four-firm concentration ratio (i.e., share accounted for by the four largest firms) is used. A four-firm concentration ratio of 80 percent indicates more monopoly power than a four-firm concentration ratio of 20 percent.

A second measure of market monopolization, the Herfindahl-Hirschman Index (HHI), calculates the sum of the squares of the market shares, expressed as a percent, held by each firm in an industry. The maximum value is 10,000 (i.e.,  $100^2$ ), and the minimum approaches zero. As noted

in Chapter 17, analysts view the HHI as superior to the concentration ratio because it captures the size distribution of firms by giving larger firms more weight. Under the Merger Guidelines issued by the Department of Justice in 1984, a market with an HHI value above 1,800 is considered “highly concentrated,” while one below 1,000 is considered “unconcentrated.”

As an example, Emmons, Guardado, and Kane (2010) look at 2008 concentration measures for the health insurance industry for the HMO and PPO markets at the Metropolitan Statistical Area (MSA) level. They report that 357 of the 359 combined HMO+PPO markets are highly concentrated ( $HHI > 1,800$ ), with only Colorado Springs, Colorado, and Poughkeepsie, New York below this level. Likewise, they report that 99 percent of the HMO markets and 100 percent of the PPO markets also have HHI ratings exceeding 1,800.

Many economists argue that the concentration ratio and HHI concepts say little about competitive behavior. Some industries with few firms may compete vigorously, whereas others with many firms may not. It is also critical to define the relevant geographic and product markets carefully. One may see many hospitals in a given state, implying a low concentration ratio. However, if separated geographically, they may have monopoly power in the relevant geographic area or in the market for a particular service if it is impractical to receive that care elsewhere. We will see later in this chapter how the definition of the relevant market is crucial in the practice of antitrust policy.

### Antitrust Procedures

Gaynor and Vogt (2000) show that in deciding whether anticompetitive behavior should be addressed under the Clayton or Sherman Acts, the courts and the enforcement agencies go through a fairly routine sequence of steps. Using hospital mergers as an example, the steps (Miles, 1998) are:

1. Defining the relevant product market
2. Defining the relevant geographic market
3. Identifying the competitors in the relevant product/geographic market
4. Calculating the market shares of the competitors and HHI
5. Calculating the merging firms’ postmerger share and the postmerger HHI, and determining the likely competitive effect of the merger
6. Considering any factors that mitigate or exacerbate anticompetitive effects

In defining the relevant market, analysts must find the smallest group of products for which no close substitutes exist. A product is to be included in the relevant market if buyers easily could find a substitute for it in response to a price increase of the merging firms. In addition, a supplier is to be included in the relevant market if it does produce the same product or does not produce the relevant product currently but could easily come to produce it. DOJ’s 1984 Merger Guidelines say that the relevant market, both product and geographic, is the smallest market in which the sellers, acting as a cartel, could implement profitably a “small but significant and nontransitory price increase.” They go on to say that, most often, this means a 5 percent increase for a period of at least one year.

Defining the relative geographic market has become critical. If a market is defined too narrowly, the true impact of a merger on market concentration will be overestimated. Conversely, a market that is defined too broadly will lead to an underestimate of the impact on the concentration of a merger.

The relevant geographic area for antitrust purposes, then, is the smallest area in which buyers could not switch easily to providers elsewhere in order to escape attempts to exercise monopoly power. In effect, this definition revolves around the ability of producers to raise prices above competitive levels. This ability reflects the demand price elasticity of facing the producers. Relatively low price elasticities increase the ability of producers to raise prices. Reliable measures, or even indirect evidence, of elasticity are often not available so the courts will look at a variety of approaches.

**BOX 20-2****How the Department of Justice Works: A Marin County Consolidation**

The DOJ reviews proposed hospital mergers and publishes findings on a Web site. This 1997 example provides the economic reasoning used to evaluate mergers for potential antitrust violations.

Marin County, California, is directly north of San Francisco across the Golden Gate Bridge. Two hospitals, Marin General and Ross, proposed to consolidate their inpatient mental health services. While the two hospitals competed in providing inpatient and other psychiatric care to adults, Marin General did not provide the chemical dependency programs and the inpatient psychiatric services for children and adolescents that Ross provided. The hospitals stated that they would continue to compete in the sale of the consolidated services and would not jointly determine prices for the consolidated services, other than for Medicaid and indigent patients covered under the county's program. Joint pricing for Medicaid and county program patients would not eliminate competition, however, since the hospitals did not compete for that business.

The DOJ approved the consolidation. Note the explicit economic analyses and considerations. The original argument has been quoted verbatim, with the particular features numbered to clarify the issues.

**APPROVED:**

1. The proposed consolidation will not result in *per se* illegal conduct, and under a rule-of-reason analysis the Department is not prepared to say that the consolidation is likely to have a net anticompetitive effect. While these are the only hospitals in Marin County providing inpatient psychiatric care, the venture explicitly preserves the potential for price competition between the hospitals and includes protections against the unnecessary sharing of confidential business information.
2. The venture may lower the cost of adult mental health services by eliminating duplicative costs and spreading fixed costs over a larger population. The consolidation may thus permit the hospitals to offer competitive rates for the care of Medicaid patients and indigent patients covered by a Marin County program.
3. On the other hand, the venture has significant potential for eliminating competition in quality of care or other nonprice areas, and joint pricing for Medicaid patients could facilitate collusion on the pricing for other patients.
4. On balance, and on the facts presented, the Department does not have a present intention to challenge the venture, but that view could change depending on how the venture actually operates.

*Source:* Original source no longer available. Related document at <http://www.justice.gov/atr/public/busreview/1044.htm>, June 14, 2011.

**The Elzinga-Hogarty (EH) Criterion**

One common standard for determining market areas involves measurements of product flows into and out of an area—an approach consistent with the economic concept of a market. Here, the work of Elzinga and Hogarty (1978) has been influential. The authors define a geographic market as one in which the outflow of goods to other areas is relatively small and the inflow (imports) from elsewhere is also relatively small.

The logic is simple. If producers exported significant amounts elsewhere, they could not raise prices without jeopardizing their external markets. Thus, those external markets must be included. Similarly, if imports are substantial, price increases would result in users substituting more imports. Therefore, the geographic markets supplying the imports must be included. Criteria of 75 percent and 90 percent were suggested; that is, both imports and exports must be less than 25 percent (10 percent) of the product flows. To satisfy the 90 percent criterion, hospitals must provide at least 90 percent of their services to patients residing in the market, and patients residing in the market must receive no more than 10 percent of their care from hospitals located in other market areas.

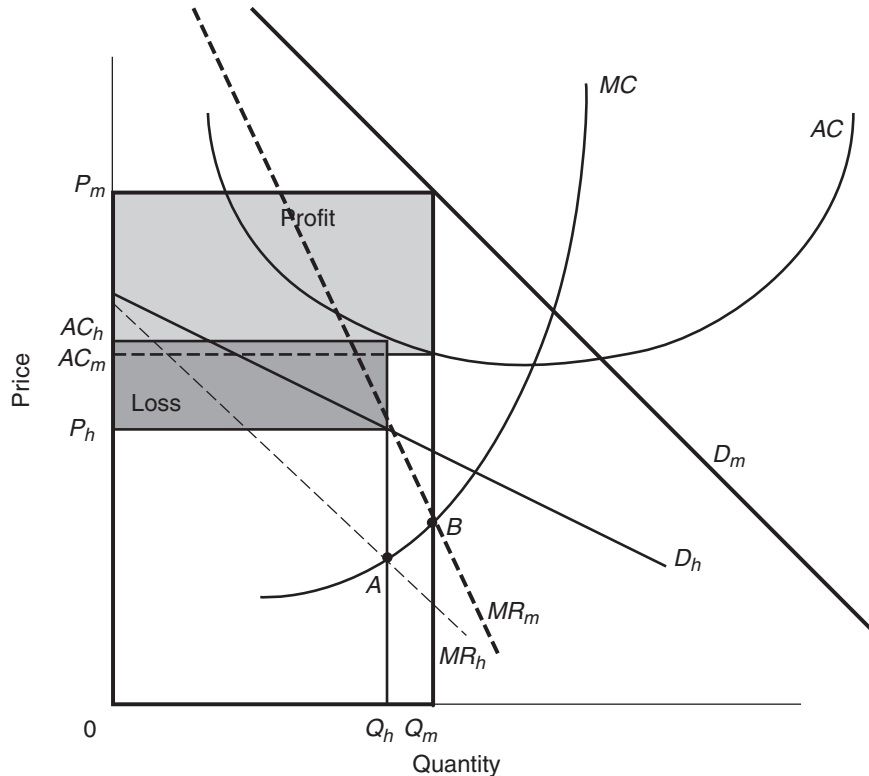
Despite advances made by economists in defining markets, court decisions in merger cases have been contradictory. Some cases have accepted the EH method, while others have relied on various methods including ad hoc approaches, such as defining markets by the 15-mile radius around a hospital. Thus, the law remains unsettled on this important issue.

Much of the analysis has involved hospital mergers. Consider a proposed merger of two neighboring hospitals. Antitrust analysts look at two fundamental concerns:

1. Hospital services technology involves large fixed costs, and requires spreading them out over a large demand. We see very few 10-bed hospitals, because they would not create enough revenue to cover the very large fixed costs.
2. Without a merger, one or both of the hospitals might go bankrupt, leaving a monopoly anyhow.

To demonstrate the economics behind the rules and practices, consider Figure 20-7. Curve  $AC$  reflects average costs, and the representative hospital's service demand is  $D_h$ . If each representative hospital equates marginal revenues and marginal costs at point  $A$ , it loses money, because  $P_h$  (determined by the demand curve) is less than average cost  $AC_h$ . The loss for each hospital is indicated by the (darker shaded) box with height  $(AC_h - P_h)$ .

If regulators allow the hospitals to merge, the (thick) demand curve facing the single new entity shifts to the right because there are more people to serve. It also becomes less elastic there are fewer competitors. As drawn, the merged hospital equates marginal revenues and marginal costs at point  $B$  and earns a profit indicated by the box with height  $(P_m - AC_m)$ . Note, however, that this is achieved with a higher price, because  $P_m$  exceeds  $P_h$ , and  $Q_m$  of services provided by the newly monopolized hospital is far less than the amount ( $2Q_h$ ) that was provided by the two hospitals. Antitrust policy makers must address issues such as these.



**FIGURE 20-7** Potential Impacts of Hospital Mergers

## CONCLUSIONS

This chapter discusses and analyzes specific regulatory mechanisms introduced into the health care economy. Early regulatory mechanisms, including rate regulation, utilization review, and Certificate-of-Need programs sought to control markets that seemed otherwise impervious to competitive forces. Rate regulation was somewhat effective in reducing expenses per admission, per day, and per capita; CON regulation (addressing capital expenses) was largely ineffective; and utilization review was somewhat effective in reducing admissions, inpatient days, and expenditures. With the growth of managed care throughout the economy in the 1990s, utilization review has come to characterize almost all health delivery systems, whether formally regulated or not.

The imposition of PPS in the 1980s represented a form of regulation that promotes competition. We see how payment by DRG promotes a form of yardstick competition among monopolistic firms. Medicare's PPS has reduced length of stay in hospitals subject to the regulation but has not led to reduced quality of care or access to care. With the particular limitations on inpatient care, many large U.S. hospitals have come to face major financial challenges, and analysts will be monitoring governmental responses to the considerable political pressure to adjust reimbursement rates.

With the challenges facing hospitals, interest in antitrust approaches toward regulation was renewed. Regulators continue to face the trade-off between the potentially increased efficiency of larger-scale firms, at the potential cost of decreased competition and higher prices for the delivered services.

---

## Summary

1. Economists feel that regulation often is needed when competitive market pressures are not present. One can categorize the policies as those that:
  - Recognize the monopoly power and try to control it
  - Try to make monopolistic firms act like competitors
  - Attempt to prevent the accumulation of monopoly power
2. Regulation refers to the use of nonmarket means to affect the quality, price, or quantity of a good or service. Principal categories of regulation include fee controls and rate regulation, quantity and capacity controls, and quality controls.
3. Studies of rate regulation suggest that this regulatory device is effective in reducing expenses per admission, per day, and per capita.
4. The Medicare Prospective Payment System (PPS) based on DRGs predetermines a flat fee per case. Hospitals that exceed this rate suffer losses, while hospitals with case costs below the rate receive profits.
5. Medicare's PPS approximates yardstick competition. By setting the payment rates according to industry average marginal costs, yardstick competition induces the firms to choose the socially efficient level of cost-containment expenditure.
6. Medicare's PPS has reduced length of stay in hospitals subject to the regulation. It is doubtful that the regulation has led to reduced quality of care or access to care. Finally, while it is likely that the system has helped control Medicare's budget, it is unclear whether a reduction has occurred in per capita costs.
7. Relative value scales (RVSs) originally were developed by state medical societies and other organizations as guides in establishing physician fees and reimbursement rates. The weights under the resource-based RVS (RBRVS) introduced under Medicare's physician payment reforms are determined by the inputs needed to provide procedures, including physician work, practice expenses, and professional liability insurance. They have had only limited effectiveness in addressing per provider reimbursement and aggregate provider costs.
8. Antitrust refers to the use of government policy or legislation aimed at deterring collusive arrangements and certain other practices that restrict competition.
9. If insurers set the premiums competitively, moving the price of medical services toward the marginal cost improves consumers' well-being.
10. The Lerner index is a measure of a firm's monopoly power. Concentration ratios and the HHI often are used to measure market monopolization.
11. Antitrust policies must often weigh the benefits of increased competition against the need for entities with relatively large fixed costs to cover those costs.

## Discussion Questions

- Describe the major regulatory strategies that in the United States.
- How does prospective payment change the incentive to hospitals as compared to retrospective reimbursement? What predictions would one make as a result of the adoption of reimbursement based on DRGs?
- Some argue that controls on some payers lead to a shift in cost to other payers (and that a single-payer system is needed). What evidence would suggest cost shifting to other third-party payers? What evidence suggests that not all the costs are shifted?
- CON legislation seeks to control costs by limiting capital expenditures. However, CON also may influence competition in the hospital market. Discuss at least two ways in which this might occur.
- Under Shleifer's theory of yardstick competition, why does the firm have an incentive to reduce its costs? If all firms respond by reducing their costs, will the payment rate also subsequently fall? (Students with training in game theory may wish to examine the elegant logic in the original article.)
- What is the Medicare PPS program under DRGs? How has the Medicare PPS payment under DRGs affected hospital practices? Length of stay? Quality of care? Financial condition?
- Does the Medicare PPS payment under DRGs reduce costs? Discuss.
- Market power means that providers face downward-sloping demand curves. Give three examples of health services providers who have considerable market power. Give three examples of health services providers who have little or no market power.
- Why is the determination of relevant markets so important to antitrust cases? Why will plaintiffs want narrow definitions and defendants want broad definitions?
- Consider where you live or go to school. Have there been hospital mergers or closings in the last several years? How would you explain them in the context of material presented in this chapter?

## Exercises

- In Figure 20-1, suppose the demand for the good was summarized by the equations:
 
$$P = 100 - 0.5Q$$

$$MR = 100 - Q$$
 and that the marginal cost equals the average costs at \$10 per unit.
  - Calculate the optimum market quantity in a competitive market. (Hint: Set price equal to marginal cost.)
  - Calculate the quantity brought to market by the monopolist.
  - Calculate the monopolist's profit.
  - Calculate the deadweight loss to society from the monopoly.
- Suppose that a hospital has a production function of the type:
 
$$\log Q = a \log S + b \log B$$
 where  $Q$  is quantity of output,  $S$  is level of services, and  $B$  is number of beds, with  $a$  and  $b$  as parameters. The hospital buys  $S$  at price  $p_s$  and  $B$  at price  $p_b$ .
  - If regulators force the hospital to decrease the amount of  $B$  by 10 percent, what must the hospital do to maintain quantity  $Q$ ?
  - If the hospital maintains quantity  $Q$ , what will be the effect of the regulation on total hospital expenditures? Why?
- Suppose that Hospitals A through E have the following marginal costs for a given procedure:
  - Hospital A—\$2,000
  - Hospital B—\$2,200
  - Hospital C—\$1,800
  - Hospital D—\$2,700
  - Hospital E—\$2,300
 Calculate the yardstick price that would be assigned to each hospital. Which two hospitals will be assigned yardstick prices that do not cover their current marginal costs?
- Examine the analysis in Figure 20-1. Suppose that the hospital's initial equilibrium price,  $P_0$ , is \$10,000, the initial equilibrium quantity is 1,000 procedures, and the initial marginal cost per procedure is \$5,000.
  - What is the initial profit level for the hospital?
  - If it costs \$500,000 in fixed costs to bring in a consultant who will reduce marginal costs to \$4,500, what does the new average cost curve look like?
  - Suppose that the hospital is given a yardstick price of \$4,600, and it raises its production to 2,000 procedures. Assume that the lump-sum subsidy is \$525,000. Calculate the following:
    - Hospital's profit or loss
    - Increase in consumer surplus brought about by the imposition of yardstick pricing
- Here is a more complex yardstick problem. A monopoly hospital faces the following demand curve
 
$$q = 400 - 10p$$



and the following marginal cost (with no fixed costs)

$$c = 22$$

- (a) Calculate the profit-maximizing values of  $p^*$  and  $q^*$ , the maximized profit  $\pi^*$  and the consumer surplus  $CS^*$ . Suppose that the firm could reduce its costs according to the formula
- $$R = 40d^2, \text{ where } d = \text{the original cost (here, 22) - the new (reduced) cost.}$$
- A yardstick regulator assigns the hospital the following parameters:
- Lump sum subsidy = 300;  
Yardstick price = 20;
- (b) Give the profit maximizing condition for the yardstick regulation.
- (c) Calculate the profit maximizing values of  $p^*$  and  $q^*$ , cost reduction expense  $R^*$ , maximized profit  $\pi^*$  and consumer surplus  $CS^*$
6. Suppose that the town of Greenville has three hospitals. Hospital A has a 30 percent market share, Hospital B has a 35 percent market share, and Hospital C has a 35 percent market share. Hospital A and Hospital B have proposed to merge. The government argues that the merger would constitute monopoly power, but the two hospitals disagree.
- (a) Provide a clear argument in favor of the hospitals' position.
- (b) Provide a clear argument in favor of the antitrust position.
7. Consider an industry with six firms with the following market shares:
- Firm A—38 percent  
Firm B—28 percent  
Firm C—13 percent  
Firm D—11 percent  
Firm E—6 percent  
Firm F—4 percent
- (a) Compute the four-firm concentration ratio. What is the value of the HHI?
- (b) Suppose Firm D proposes to merge with Firm E. Calculate the new four-firm ratio and the new HHI.
8. Consider the equation for costs per hospital discharge:
- $$\text{cost/discharge} = (\text{days/discharge}) \times (\text{service intensity/day}) \times (\text{cost/service unit})$$
- (a) Suppose that cost/service unit increased by 10 percent and days/discharge stayed constant. If we know that cost/discharge increased by 15 percent, then what happened to the service intensity? Why?
- (b) In part a, what can we say about quality of care?
- (c) Can we be sure that if we reduced the number of days per discharge that costs would fall? Why or why not?

# Social Insurance

- Social Insurance Policies and Social Programs
- Historical Roots of Social Insurance
- The Patient Protection and Affordable Care Act (PPACA) of 2010
- Medicare and Medicaid in the United States
- Public Insurance and Health
- The Effects of Medicare and Medicaid
- Criticisms of the U.S. Health Care System
- Conclusions

Throughout this book, we have emphasized the role that markets can play in providing health care. We have also generally used economic efficiency, provided in theory by perfectly competitive markets, as a standard against which to judge the costs and benefits of policies. In this chapter, we address instead the issues that arise when a society considers providing for health care by offering health insurance, to some significant degree, at the public's expense. Such insurance programs provided through taxes or regulations are called social insurance programs. Having provided an overview of the rationales for social insurance in health care, we now turn to an examination of social insurance. We begin by considering the history of health care social insurance throughout the world, with emphasis on the United States. We then examine Medicare, Medicaid, and CHIP, and their effects, and we close with prospects and issues for the future.

## SOCIAL INSURANCE POLICIES AND SOCIAL PROGRAMS

Social insurance programs can be broken down into five categories. While not mutually exclusive, programs of the following five types exist in the United States and in other countries.

1. **Poverty:** Poor people lack purchasing power to buy the amounts of goods considered necessary to provide the minimal standards of decent life. Programs directed toward persons experiencing poverty involve either cash (more recently, debit cards) or goods “in kind,” such as rent vouchers or food stamps.
2. **Old Age:** The elderly have attained a certain age, generally coinciding with retirement from active employment. Programs include income maintenance, such as Social Security, as well as services and considerations (such as assisted housing, Meals on Wheels, or transportation) that may address their generally decreased mobility.

3. **Disability:** The disabled have either temporary or permanent inability to work because of illness or work-related injuries. Programs generally provide cash benefits. Disability programs were among the earliest social insurance programs available.
4. **Health:** Programs cover illness or well-care financing and/or provide facilities for various groups. In the United States, most programs have targeted children, the elderly, and/or the poor, with the government financing the individual's health care either entirely or in part. In other countries, governments have more direct involvement in the financing and delivery of health services for larger segments (or all) of the population.
5. **Unemployment:** The unemployed receive assistance due to a temporary loss of work. While unemployment-related programs generally provide short-term cash benefits, in many countries, longer-term unemployment may lead into poverty-related welfare programs.

Analysts find several other definitions useful in considering social insurance programs. Some programs, termed *entitlements*, are available to all who qualify. Food assistance in the form of stamps or debit cards, for example, is available in the United States to all households whose incomes fall below specified levels, related in part to income, family size, and geographic location. Medicaid, or financed health care for the poor (discussed later in this chapter), is also an entitlement program for all who meet particular qualifications. In contrast, housing programs in the United States are not entitlements. Only limited numbers of subsidized units (on the supply side) or vouchers (on the demand side) are available.

Many programs are means-tested in that they are available only to individuals or households who meet certain income criteria. Households that receive aid for poverty-related problems may lose some or all of the aid as their incomes increase. Such reductions in aid may have the unintended effects of discouraging efforts by low-income households to find jobs. For example, formulas that reduce poverty-related aid by \$1 for each \$1 earned on the job constitute taxes on job earnings at rates approaching 100 percent.

Finally, aid may take various forms. Programs often tie aid to the purchase of certain items, such as food or housing. This procedure ensures that the people use the aid to buy items that the legislators have deemed important. Under certain circumstances, however, it may be more economically efficient to provide a cash subsidy rather than one in kind.

Answer the question, "Which would you prefer, \$100 in cash or \$100 worth of physician care?" Most would answer "\$100 in cash!" Legislators and the voting public, however, often seem to prefer subsidies *in kind* rather than in cash so that they can monitor or control the purchases of those receiving the subsidies. Food subsidy rules limiting the purchase of "non-food" items such as cigarettes or liquor, deemed undesirable by many, also prevent the purchase of laundry detergent or toilet tissue, which most would view to be desirable.

### Program Features

We discuss certain common features to characterize health-related social insurance programs in the United States. The first three relate to receipt of care:

1. Contributions—taxes, deductibles, and coinsurance
2. Benefits—how much, who is included, and what types of treatment are included
3. Length of coverage.

The latter two describe the provision of care, as well as the political problems involved in initiating plans:

4. Means of reimbursement to providers
5. Methods of determining payment levels to providers.

Although supported by government, most social insurance plans also impose costs on their recipients. Many are funded by tax collections, and care recipients are often taxpayers. In some cases, the taxes in question may be regressive. By definition, a regressive tax is one for which lower income people pay

higher fractions of their incomes to the tax than do richer people. In contrast, a progressive tax is one in which lower-income people pay lower fractions of their incomes to the tax than do richer people.

In the United States tax payments into the Social Security program have always been somewhat regressive. Since its beginning in 1935, the tax has been a constant percentage of wage income, up to a ceiling at which the marginal tax rate becomes zero.<sup>1</sup> This means that on average, people in higher-income groups pay a smaller proportion of their income in payroll taxes. Both workers and their employers pay 6.20 percent, and the maximum wage base was \$106,800 in 2011. Since workers with wages of \$120,000 pay the same amount, 6.20 percent of \$106,800, or \$6,621.60, the tax percentage for the more affluent is 5.52 percent, rather than 6.20 percent below the \$106,800 cap. The effective tax rate obviously falls as wage income rises above \$106,800. In contrast, the Medicare tax rate is now 1.45 percent of all wage income, a constant percentage.<sup>2</sup>

To determine whether a social insurance program is redistributive—whether it in net causes a transfer of money from the rich to the poor—one must consider not just the tax payments made but also the benefits received. For example, data for the Social Security program in the United States tend to show that Social Security is redistributive—that the poorer people tend to gain more in net than do the rich.

In addition to tax payments, eligible recipients often must pay deductibles or coinsurance. Analysts also consider time costs with respect to paperwork or waiting time for appointments or treatment.

Benefit levels and length of coverage are similar to the workings of private insurance. Given the equity considerations of social insurance, political considerations may affect both. Determining who and which treatments are covered is also important. Coverage of individuals may involve children and spouses. Coverage of treatments may mandate coverage of certain diseases (i.e., end-stage kidney disease) and exclude others (i.e., optometric or chiropractic services).

Health-related social insurance also has supply-related characteristics. In some programs patients may pay directly for expenses and then be reimbursed. In others, government may pay providers directly. In some countries, all physicians who participate in the national health care program are government employees.

## HISTORICAL ROOTS OF SOCIAL INSURANCE

### European Beginnings

Prior to the passage of the Patient Protection and Affordable Care Act of 2010 (PPACA), the United States was often characterized as the only industrialized country lacking a comprehensive health-related social insurance system. Historians date the pioneering legislation for a system of compulsory national health insurance to Germany in 1883. National health insurance spread to other European countries at the end of the nineteenth and the early part of the twentieth centuries.

The German system and the other European systems extended already-existing voluntary associations, often guild or mutual aid groups whose benefits to members included the pooling of insurable risks. The political impetus for converting or merging voluntary local groups in a system of national coverage was often its usefulness as a stabilizing influence. Paul Starr notes:

Political discontent precipitated the introduction of social insurance in both Germany and England. . . . In Germany, Bismarck introduced social rights to avoid granting wider political rights; in Britain, Lloyd George sought social rights within the context of existing rights to political participation. But both were basically defensive efforts to stabilize the political order . . . (Starr, 1982, p. 239)

<sup>1</sup> From 1937 to 1949, the tax rate was 1% on incomes up to \$3,000, or a maximum of \$30 per year!

<sup>2</sup> In 2010, to stimulate the economy Congress lowered the employee's share from 6.2 percent to 4.2 percent for the year 2011, while keeping the employer's share at 6.2 percent. As of late 2011 it is not clear whether this difference between the two payers will continue past 2011.

The German laws of 1883 set up a highly decentralized program that covered workers in mining, transportation, construction, manufacturing, mechanical trades, and establishments using power machinery. Austria (1888), Hungary (1891), Sweden (1891), Denmark (1892), and Belgium (1894) followed. With Spain's adoption of a plan in 1929, nearly every European country had enacted health insurance laws. Some were compulsory, as in Germany, but other countries, such as Belgium, Denmark, Sweden, and Switzerland, provided government subsidies to voluntary mutual insurance funds.

The United Kingdom established its first social health insurance system in 1911. That legislation helped prepare for the establishment of the British National Health Service in 1946, the most prominent example in the Western countries. Government provision was also common in Eastern European countries and the former Soviet Union, which began its system in 1926.

### **Early Experience in the United States**

Compared with the European countries, the United States came late to social insurance and to governmental health insurance in particular. In the late nineteenth and early twentieth centuries, Americans, like the Europeans, established voluntary group purchasing arrangements, and mutual benefit associations. However, in the United States, government did not take up the funding of these voluntary societies, and from the advent of the German system in 1883 to the beginnings of the British system in 1911 governmental compulsory insurance for health care was not really a U.S. issue.

Health insurance plans began to gain supporters in the early part of the twentieth century. The American Association for Labor Legislation supported such a plan, and the American Medical Association initially responded positively to health insurance proposals. Political consensus was lacking, however. For example, Samuel Gompers, president of the American Federation of Labor, opposed compulsory health insurance.

The major advance of social insurance legislation in the United States occurred with the establishment of the Social Security program in 1935. Despite the social insurance thrust of the program and the reform-minded support for it, the legislation made concessions to political opposition to the New Deal, including the omission of governmental health insurance. The omission of health insurance from the Social Security Act was by no means the act's only conservative feature. It relied on a regressive tax and provided no coverage to some of the very poor, such as farm laborers and domestics.

Proponents of compulsory health insurance plans were no more successful through the 1940s and 1950s. President Harry Truman proposed "a single health insurance system that would include all classes of the society, even those . . . not covered by Social Security" (Starr, p. 282). During the public debate over these issues, opponents of compulsory health insurance called it "socialized medicine," a term that greatly weakened its support in the political climate of the Cold War. Though Truman had won the 1948 election, his success did not translate into a health insurance program.

### **The Establishment of Medicare and Medicaid**

The social insurance debate in the United States has often ranged between those who believe in voluntarism and voluntary insurance, as won in negotiations between self-reliant industries and unions, and those who believe that only a compulsory insurance program would provide the insurance that was necessary for the larger population. The major social insurance programs for health care in the United States, Medicare and Medicaid, passed in 1965. President Lyndon Johnson had supported health care for the aged, and by winning a landslide victory in the 1964 election, he was able to push for these programs.

While Medicare and Medicaid have achieved widespread political support since the 1960s, fundamental concerns stem from their rising costs. To the present day, with growing elderly and poor populations as well as continual health care cost inflation, many observers perceived the total

costs of the social insurance as having grown out of control. Politicians have sought to ensure that Medicare would have a sufficient trust fund to meet future needs, while balancing these needs against the increased taxes and payments for services necessary to support them.

The wider issue of social health insurance for the population as a whole has not yet been decided. The experience of 1993 and 1994 with the Clinton health plan, however, was a “massive political upheaval” (Skocpol, 1995). President Bill Clinton won the 1992 election with a campaign promising health system reform. There had been signs of growing interest among the electorate in health system reform including social health insurance. The Clinton plan emerged during January through May 1993 through a large task force of government officials, health policy experts, congressional staffers, and others. With many perceiving the task force to be secretive and unresponsive to the public, plan supporters were unable to develop a coalition of interests around clearly defined features of reform, while insurers lobbied and advertised against the plan. During 1994, the Clinton administration re-worked the plan to respond to critics but the public came to believe that it would entail considerable government bureaucracy, cost, and inefficiency. It was scrapped in late summer 1994.

## **THE PATIENT PROTECTION AND AFFORDABLE CARE ACT (PPACA) OF 2010**

The 2008 U.S. Presidential election set the stage for potential health policy reform. With worries about rising health care costs, and with increasing numbers of the uninsured due to the “Great Recession,” the major Democratic candidates (Barack Obama and Hilary Clinton), and the Republican candidate John McCain all spoke to the need for health care reform. Obama’s election and increased Democratic majorities in the Congress led to passage in March 2010. While the details of the Act and the legislation will be given more detailed treatment in Chapter 23, the Act as passed does the following (Kaiser Family Foundation [KFF], 2011):

- It requires most U.S. citizens and legal residents to have health insurance, the so-called *individual mandate*.
- It assesses a fee against employers with 50 or more full-time employees that do not offer coverage as a premium tax credit; this fee is \$2,000 per full-time employee, excluding the first 30 employees from the assessment. It requires employers with more than 200 employees to enroll employees automatically into health insurance plans offered by the employer.
- It expands Medicaid to all non-Medicare eligible individuals under age 65 (children, pregnant women, parents, and adults without dependent children) with incomes up to 133% of the Federal Poverty Level (FPL) with a benchmark benefit package.
- It creates state-based American Health Benefit Exchanges and Small Business Health Options Program (SHOP) Exchanges, administered by a governmental agency or non-profit organization, through which individuals and small businesses with up to 100 employees can purchase qualified coverage.

The Act includes many other important provisions such as the imposition of a “Cadillac” tax on high cost employer-provided policies, and the elimination of underwriting policies that enable insurers to deny individuals insurance or charge them higher premiums based on pre-existing conditions. Some provisions of the Act took effect on enactment of the legislation; other provisions will be phased in through 2018.

Political opposition immediately arose, focusing on the individual mandate, as well as concerns about the projected costs. However, implementation has proceeded, and by 2020 the PPACA is expected to insure at least 32 million of the 50 million currently uninsured (non-citizens and illegal immigrants will not be allowed to participate). This represents the biggest expansion of U.S. health policy since the passage of Medicare and Medicaid in 1965.

## MEDICARE AND MEDICAID IN THE UNITED STATES

While PPACA will change large parts of the federal health care policy presence, Medicare and Medicaid continue to form the foundation. Enacted in 1965, Medicare is a national program that primarily provides compulsory hospital insurance to the elderly plus optional medical coverage to which nearly all elderly subscribe. In contrast, Medicaid is a program operated by the states with matching federal dollars. It primarily provides health care coverage to people who are poor.<sup>3</sup>

### Medicare

Medicare traditionally consisted of two parts: Hospital Insurance (HI), also known as Part A, and Supplementary Medical Insurance (SMI), also known as Part B. A third part of Medicare, sometimes known as Part C, the Medicare Advantage program, was established as the Medicare + Choice program by the Balanced Budget Act (BBA) of 1997 (Public Law 105-33) and subsequently renamed and modified by the Medicare Prescription Drug, Improvement, and Modernization Act (MMA) of 2003 (Public Law 108-173). The Medicare Advantage program expanded beneficiaries' options for participation in private-sector health care plans. In 2006 the MMA established a new prescription drug benefit, also known as Part D.

When Medicare began on July 1, 1966, approximately 19 million people enrolled. By 2010, approximately 47.1 million people were enrolled in one or both of Parts A and B of the Medicare program, and almost 12 million of them have chosen to participate in a Medicare Advantage plan. Figure 21-1 displays the growth of the Medicare program since its inception.

**MEDICARE COVERAGE** Medicare Part A generally goes automatically to persons age 65 and over who are entitled to Social Security or Railroad Retirement Board benefits. Similarly, those who have received such benefits based on a physical disability for at least 24 months also are entitled to Part A benefits. In 2010, Part A benefit payments totaled \$244.5 billion to 47.1 million enrollees. The average annual benefit per enrollee was \$5,187.

Part A coverage includes:

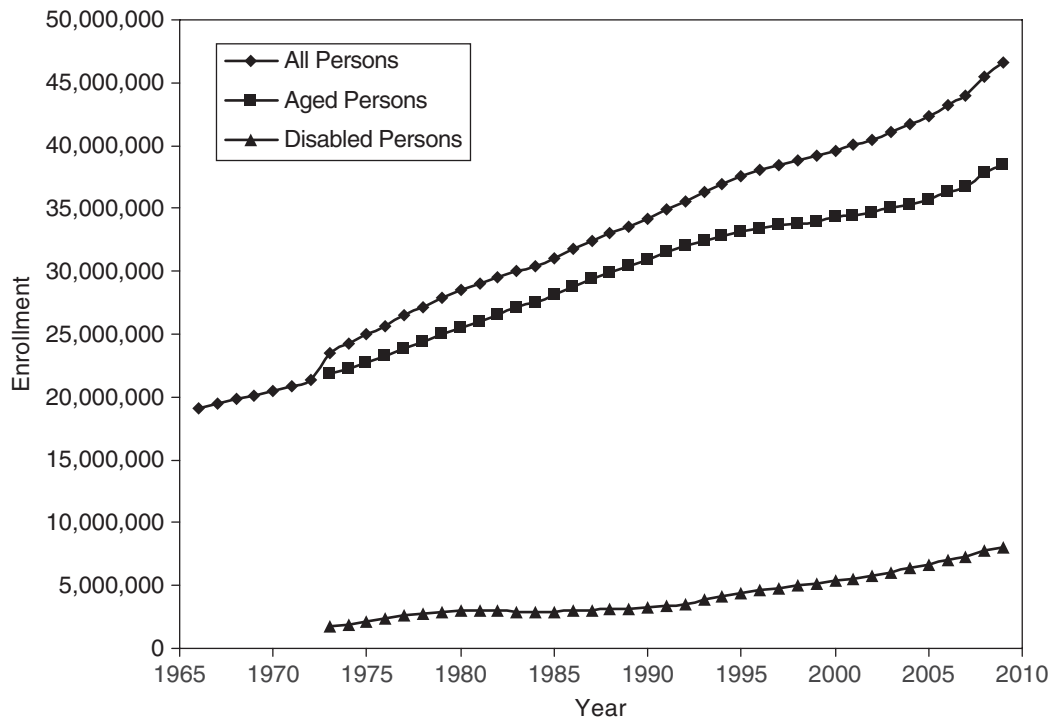
- Inpatient hospital care coverage, requiring an initial deductible payment, plus copayments for all hospital days following day 60 within a benefit period
- Skilled nursing facility (SNF) care, which generally is covered by Part A only if it is within 30 days of a hospitalization of three or more days and certified as medically necessary
- Home Health Agency (HHA) care, including care provided by a home health aide
- Hospice, which is provided to those terminally ill persons with a life expectancy of six months or less and who elect to forgo standard Medicare benefits and receive only hospice care

SMI benefits (Parts B and D) are available to almost all resident citizens age 65 and over. Part B coverage is optional and requires payment of a monthly premium. Part B covers

- Physicians' and surgeons' services (in both hospital and nonhospital settings)
- Some covered services furnished by chiropractors, podiatrists, dentists, and optometrists
- Services in an emergency room or outpatient clinic, including same-day surgery, and ambulance services

<sup>3</sup> The most current information on these programs are:

- *2011 Annual Report of the Boards of Trustees of the Federal Hospital Insurance and Federal Supplementary Medical Insurance Trusts Funds Communication* from The Boards of Trustees, Federal Hospital Insurance and Federal Supplementary Medical Insurance Trust Funds (2011) <https://www.cms.gov/ReportsTrustFunds/downloads/tr2011.pdf>, accessed June 7, 2011.
- *2010 Actuarial Report on the Financial Outlook for Medicaid*, Washington DC, December 2010, Centers for Medicare & Medicaid Services United States Department of Health & Human Services, accessed June 7, 2011.



**FIGURE 21-1** Medicare Enrollment by Year, 1966–2009

Source: Graph developed by authors through data from Centers for Medicare and Medicaid Services data compendium.

Part B also covers other services including clinical laboratory tests, X-rays, diagnostic tests, ambulance services, and blood that are not supplied by Part A. Almost all persons entitled to Part A also choose to enroll in Part B. In 2010, the Part B program provided benefits of \$209.7 billion to about 43.8 million people (36.7 million aged and 7.1 million disabled). The average Part B benefit per enrollee was \$4,786.

While all Medicare beneficiaries can receive their benefits through the original fee-for-service program, most beneficiaries enrolled in both Part A and Part B can choose different forms of service delivery by participating in a Medicare Advantage (Part C) plan instead. Organizations that seek to contract as Medicare Advantage plans must meet specific organizational, financial, and other requirements. Following are the primary Medicare Advantage plans:

- Coordinated care plans, which include HMOs, PSOs, PPOs, and other certified coordinated care plans and entities that meet the standards set forth in the law.
- Private, unrestricted fee-for-service plans, which allow beneficiaries to select certain private providers. For those providers who agree to accept the plan's payment terms and conditions, this option does not place the providers at risk, nor does it vary payment rates based on utilization.

These Medicare Advantage plans are required to provide at least the current Medicare benefit package, excluding hospice services. Plans may offer additional covered services and are required to do so (or return excess payments) if plan costs are lower than the Medicare payments received by the plan.

Medicare does not cover everything. Noncovered services include long-term nursing care, custodial care, and certain other health care needs, such as dentures and dental care, eyeglasses, and hearing aids. These services are not a part of the Medicare program unless they are a part of a private health plan under the Medicare Advantage program.



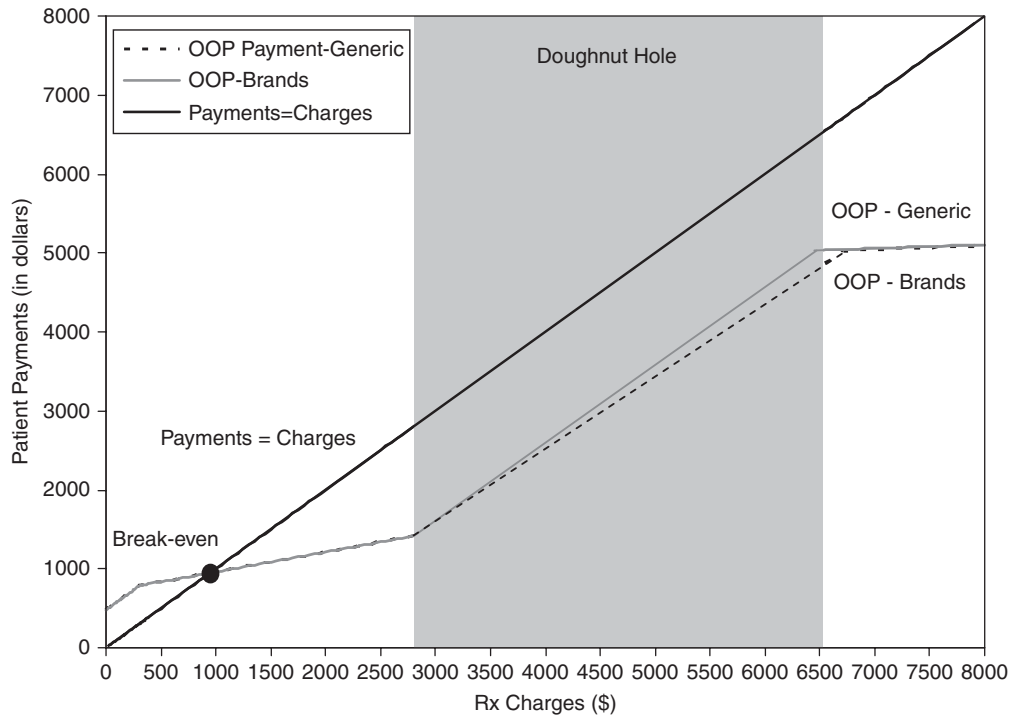
### Part D Prescription Drug Insurance

Beginning in 2006, Part D provided subsidized access to prescription drug insurance coverage on a voluntary basis upon payment of a premium, to individuals entitled to Part A or Part B, with premium and cost-sharing subsidies for low-income enrollees. Part D coverage includes most FDA-approved prescription drugs and biologicals. For an additional premium, plans might also offer supplemental coverage exceeding the value of basic coverage. To encourage employer and union plans to continue to offer prescription drug coverage to Medicare retirees, Part D provides for certain subsidies to those plans that meet specific criteria.

In 2010, Part D provided \$61.7 billion in benefits to 34.5 million enrollees. The average benefit per enrollee was \$1,789. Parts A, B, and D together provided \$515.8 billion dollars to 47.5 million enrollees.

Figure 21-2 presents the current features of the Part D coverage. Annual premiums vary by plan, but consider a typical moderate coverage at \$40 per month or \$480 per year. In 2011, there is a \$310 annual deductible. After the deductible, Part D covers 75 percent of all incremental expenditures up to \$2,840.

A controversial feature of Part D has been the so-called “doughnut hole”. As noted in the shaded part of Figure 21-2, charges above \$2,840, into the \$6,000 range have been subject to very high copayment rates, originally 100 percent. This means that after an initial subsidy, the enrollee would have to pay dollar per dollar for several thousands of dollars of drugs.<sup>4</sup>



**FIGURE 21-2** Medicare Part D Prescription Drug Benefit - 2011 Payments vs. Drug Charges

<sup>4</sup> In 2011, there is a 93 percent coinsurance (7 percent plan benefit) for generic drugs. The 93 percent coinsurance counts against the out-of-pocket threshold as does 100 percent of the cost of brand drugs (approved brand drugs qualify for a 50 percent subsidy). Thus, the total amount of spending required to reach the OOP threshold and catastrophic coverage will depend upon whether spending is on generic drugs, brand drugs or a combination. (Source: [www.buckconsultants.com/.../FYI-04-08-10a-CM-Releases-2011-Medicare-Part-D-Benefit-Parameters.pdf](http://www.buckconsultants.com/.../FYI-04-08-10a-CM-Releases-2011-Medicare-Part-D-Benefit-Parameters.pdf), accessed June 6, 2011).

The 2011 Part D plan has a \$4,550 catastrophic threshold. Upon having spent \$4,550 on drugs (in addition to the monthly premiums), enrollees have to spend only 5 percent out of pocket on expenditures past that threshold.

The initial Part D financing plan left open conjecture as to who would participate. When the program began, the break-even point, where the participant was no worse off than not participating (that is, premium + out-of-pocket expenditures equal to total charges), occurred at a value of \$850 per year (about \$70 per month).<sup>5</sup> While Part D initially appeared generous for people with small (up to about \$210 per month) or large levels of expenditures (initially \$425 per month, now over \$500 per month), participants in the middle would derive little or no additional benefit as charges rose.

Levy and Weir (2010) provide an excellent evaluation of the Part D Benefit. Their “before-and-after” study framework allows them to examine the impact of an intervention (that is, the initiation of Part D) on a group seen prior to the intervention. They analyze data from the 2002, 2004 (both before Part D), and 2006 (after Part D was initiated) Health and Retirement Study on senior citizens’ take-up of Medicare Part D and an associated Social Security Administration Low-Income Subsidy to help the lower-income elderly pay for Part D.

They find that economic factors—specifically, demand for prescription drugs—drove the decision to enroll in Part D. For the most part, individuals who already had employer-sponsored coverage kept that coverage, as they should have. Take-up of Part D among those without previous (2004) drug coverage was high; about 50 to 60 percent of them had Part D coverage in 2006. Only 7 percent of senior citizens lacked drug coverage in 2006 compared with 24 percent in 2004. Many of those who remained without coverage in 2006 reported that they did not use prescribed medicines, and the majority had relatively low out-of-pocket spending.

When Part D began, many felt that the program was too complicated for the elderly to use, but Levy and Weir report that the majority of those interviewed had little or no difficulty with the Part D enrollment decision and were confident that they made the right decision. For the most part, then, despite the complexity of the program, Medicare beneficiaries were able to make economically rational decisions in which they had confidence. This too is not surprising. For those without any coverage, the decision to buy Part D coverage was hardly a marginal decision—almost any type of Part D plan was better than nothing.

The coverage gap, or doughnut hole, did induce some substitution behaviors. Hoadley and colleagues (2007) examined nationwide retail pharmacy claims data for 2007 and found that about 74 percent of the enrollees (excluding those enrollees who received low-income subsidies and nonusers) did not reach the coverage gap, about 22 percent remained in the coverage gap, and about 4 percent reached the catastrophic coverage level. Among eight drug classes, the majority of enrollees who reached the coverage gap made no detectable change in their medication use for the drug (or drugs) they were taking within the class when they reached the gap.<sup>6</sup> However, averaged across the eight classes, 20 percent of those who reached the gap made some change in their use of drugs within the selected class, while others may have stopped taking a drug in another class to continue taking medication in the studied class. In particular:

- 15 percent stopped taking their medication within the particular class,
- 5 percent switched to another medication (most often a generic drug) in the same class, and
- 1 percent reduced the number of separate medications they were taking in the class.

<sup>5</sup> With the parameters in the paragraph above, the current break-even point is about \$950.

<sup>6</sup> These classes were (1) Angiotensin-Converting Enzyme (ACE) Inhibitors, generally used to treat hypertension; (2) Alzheimer’s disease treatments; (3) Anti-Depressants; (4) Angiotensin Receptor Blockers (ARBs), generally used to treat hypertension; (5) Oral Anti-Diabetics; (6) Osteoporosis treatments; (7) Proton Pump Inhibitors (PPIs) for heartburn, gastroesophageal reflux disease (GERD), and ulcers; and (8) HMG-CoA Reductase Inhibitors (Statins) to treat high cholesterol.

In sum Part D has provided comprehensive prescription drug insurance for those who previously had none. The program has been workable, and enrollment has been substantial. The coverage parameters have led some enrollees to spend and to substitute drugs in predictable ways. Under PPACA, Part D faces changes. While it will not provide a brand benefit in the doughnut hole until 2013, there will be a separately calculated 50% brand drug discount provided by the manufacturer of the brand drug prescription to most Part D enrollees. Planners expect that by the end of the decade the Part D doughnut hole will disappear, through the combination of the additional Part D benefit and brand discount.

**MEDICARE PROGRAM FINANCING** The Medicare Part A program is financed primarily through a mandatory payroll deduction (FICA tax). The FICA tax is 1.45 percent of earnings (paid by each employee and also by the employer) or 2.90 percent for self-employed persons. This tax is paid on all covered wages and self-employment income without limit.

The SMI trust fund differs fundamentally from the HI trust fund with regard to financing. SMI is now composed of two parts, Part B and Part D, each with its own separate account within the SMI trust fund. The nature of the financing for both parts of SMI is similar, in that both parts are primarily financed by beneficiary premiums and contributions from the general fund of the U.S. Treasury.

Part B is financed through premium payments and contributions from the general fund of the U.S. Treasury. In 2011, new beneficiaries pay \$115.40 per month.<sup>7</sup> The premiums are indexed according to income, so those with incomes between \$85,000 and \$107,000 pay \$161.50 per month, with the rate topping off at \$369.10 per month for individual incomes over \$214,000. Beneficiary premiums are generally set at a level that covers 25 percent of the average expenditures for aged beneficiaries. Therefore, the contributions from the general fund of the U.S. Treasury are the largest source of Part B income.

Similarly, Part D is financed primarily through premium payments and contributions from the Treasury general fund, with general fund contributions accounting for the largest source of Part D income, since beneficiary premiums are to represent, on average, 25.5 percent of the cost of standard coverage (as described in the next section). The premiums and general fund contributions for Part D are determined separately from those for Part B.

**BENEFICIARY PAYMENT LIABILITIES** Parts A and B beneficiaries are responsible for charges not covered by Medicare and for various cost-sharing features of the plans. These liabilities may be paid by the beneficiary, by a third party, such as a private “Medigap” insurance policy purchased by the beneficiary, or by Medicaid, if the person is eligible. Medigap refers to private health insurance that, within limits, pays most of the health care service charges not covered by Parts A and B of Medicare.

For hospital care covered under Part A, the beneficiary’s payment share includes a one-time deductible at the beginning of each benefit period (\$1,132 in 2011). This covers the beneficiary’s part of the first 60 days of each spell of inpatient hospital care. If continued inpatient care is needed beyond the 60 days, additional coinsurance payments (\$283 per day in 2011) are required through the 90th day of a benefit period.

For Part B, the beneficiary’s payment share includes the following: one annual deductible (\$162 in 2011), the monthly premiums, the coinsurance payments for Part B services (usually 20 percent of the medically allowed charges), a deductible for blood, certain charges above the Medicare-allowed charge (for claims not on assignment), and payment for any services that are not covered by Medicare. For outpatient mental health treatment services, the beneficiary is liable for 50 percent of the approved charges.

<sup>7</sup> Those in the system before 2010 pay \$96.40 per month in 2011; those entering in 2010 pay \$110.50.

**PROVIDER PAYMENTS** Before 1983, HI paid providers on a “reasonable cost” basis. Since 1983, Medicare payments for most inpatient hospital services have been made under prospective payment, or PPS. As discussed in Chapter 20, PPS pays a specific predetermined amount for each inpatient hospital stay, based on each stay’s DRG classification. In some cases, the payment the hospital receives is less than its actual cost for providing the HI-covered inpatient hospital services for the stay; in other cases it is more. The hospital absorbs the loss or makes a profit. Certain payment adjustments exist for extraordinarily costly inpatient hospital stays, and payments for skilled nursing care and home health care, and rehabilitation and psychiatric care, are currently reimbursed on a reasonable cost basis, but prospective payment systems are expected in the near future.

Before 1992, under SMI, physicians were paid on the basis of “reasonable charge,” initially defined as the lowest of (1) the physician’s actual charge, (2) the physician’s customary charge, or (3) the prevailing charge for similar services in that locality. Changes beginning in 1992, defined “allowed charges” as the lesser of (1) the submitted charges or (2) the amount determined by a fee schedule based on a relative value scale (RVS). SMI reimburses most hospital outpatient services on a prospective payment system, with home health care reimbursed under the same prospective payment system as Part A.

Doctors or suppliers who agree to accept the Medicare-approved rate as payment in full (“take assignment”) may not request any added payments, or “balance bill” (beyond the initial annual deductible and coinsurance) from the beneficiary or insurer. If providers do not take assignment, they will charge the beneficiary for the excess (which may be paid by Medigap insurance). Limits now exist on the excess that doctors or suppliers can charge. Physicians are “participating physicians” if they agree before the beginning of the year to accept assignment for all Medicare services they furnish during the year. Since Medicare beneficiaries may select their doctors, they have the option to choose those who participate.

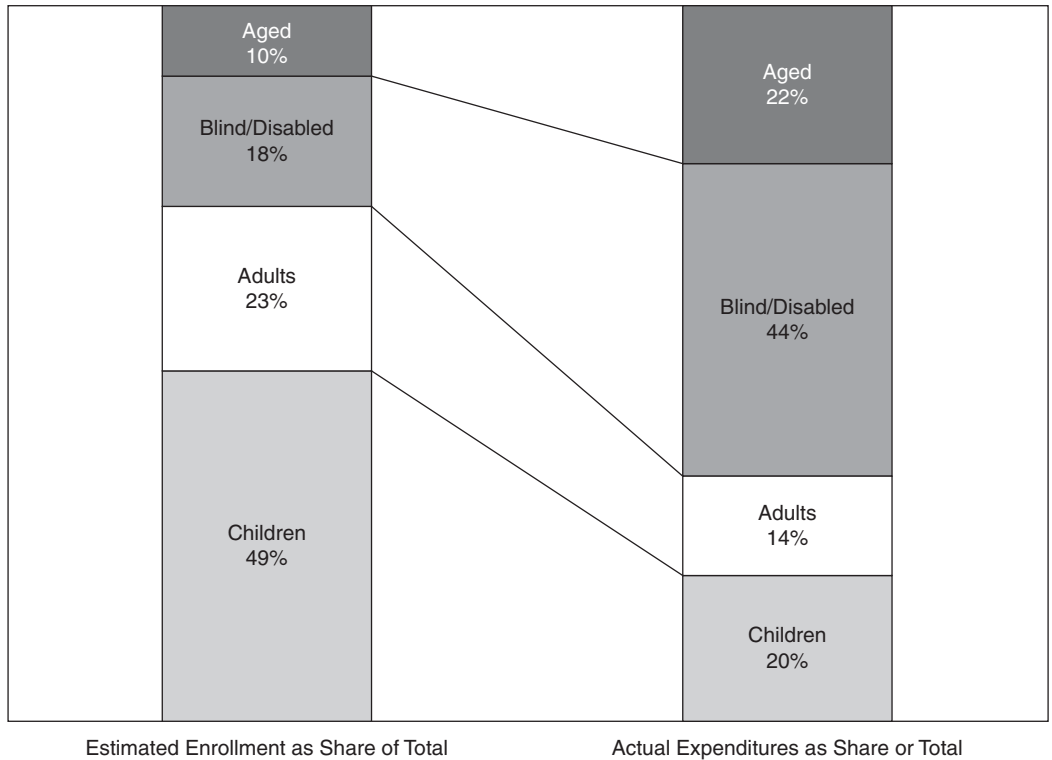
**MEDICARE SUMMARY** The Medicare program covers 95 percent of our nation’s aged population, as well as many people who are on Social Security because of disability. In 2010, Part A covered about 47 million enrollees with benefit payments of \$244 billion, and Part B covered about 44 million enrollees with benefit payments of \$210 billion. Administrative costs for both Parts A and B were \$6.7 billion, or approximately 1.5 percent of disbursements.

## Medicaid

Medicaid, referring to Title XIX of the Social Security Act, is a federal-state matching entitlement program that pays for medical assistance for certain vulnerable and needy individuals and families with low incomes and resources. This program is the largest source of funding for medical and health-related services for America’s poorest people. In 2007, it provided health care assistance to more than 50.1 million persons on average, with a total of 62.9 million people enrolled for at least one month. Total expenditures (including CHIP) for fiscal year 2009 were \$380.6 billion.

Gruber (2002) described Medicaid as four public insurance programs in one, a description that is still valid. The first provides coverage of most medical expenses for low-income women and children families. The second provides public insurance for the portions of medical expenditures not covered by Medicare for the low-income elderly, and the third covers most medical expenses for the low-income disabled. The fourth pays the nursing home expenditures of many of the institutionalized elderly. Figure 21-3 summarizes the enrollment and expenditures as shares of the total.

Each state, within broad national guidelines established by federal statutes, regulations, and policies, (1) establishes its own eligibility standards; (2) determines the type, amount, duration, and scope of services; (3) sets the rate of payment for services; and (4) administers its own program. Medicaid policies for eligibility, services, and payment vary considerably even among similar-sized and/or adjacent states and the services provided by one state may differ considerably in amount, duration, or scope from services provided in a neighboring state.



**FIGURE 21-3** Medicaid Enrollment and Expenditures, by Enrollment Group, as share of Total, FY 2009. *Source:* Truffer et al (2010), p. 14.

### Medicaid Eligibility

Medicaid was not designed to provide medical assistance for all poor persons. Even under the broadest provisions of the 1965 federal statute, it may not provide health care services for very poor persons unless they are in one of the designated groups. Low income is only one test for Medicaid eligibility for those within these groups; their resources also are tested against threshold levels (as determined by each state within federal guidelines).

To be eligible for federal funds, states must provide Medicaid coverage for certain individuals who receive federally assisted income-maintenance payments, as well as for related groups not receiving cash payments. Although there is a long list, Medicaid “categorically needy” eligibility groups for which federal matching funds are provided to states include:

- Low-income families with children
- Children under age six and pregnant women whose family income is at or below 133 percent of the federal poverty level (FPL)
- All children born after September 30, 1983, who are under age 19, in families with incomes at or below the FPL
- “Dual eligible” Medicare beneficiaries

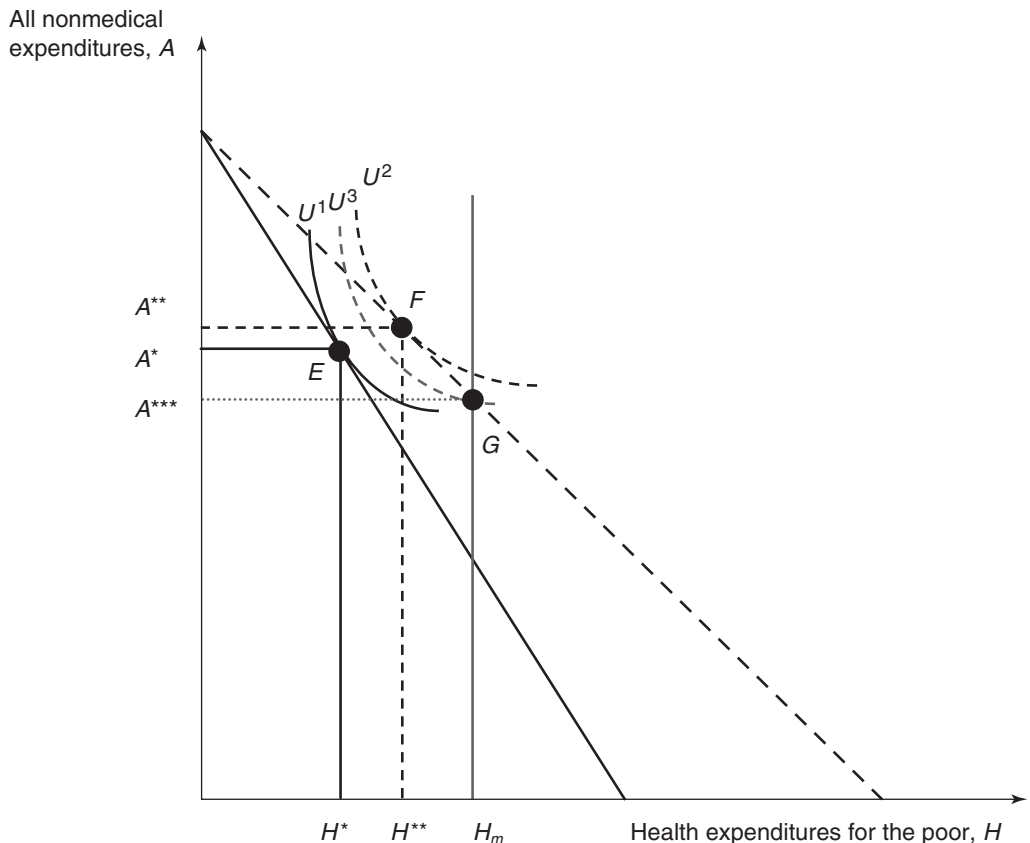
Outside of these categories, however, states generally have broad discretion in determining which groups their Medicaid programs will cover and the financial criteria for Medicaid eligibility.

The Patient Protection and Affordable Care Act of 2010 creates a new mandatory Medicaid eligibility category for all individuals with income at or below 133 percent of the Federal Poverty Level (FPL) beginning January 1, 2014. Over the following years, prescription drugs and mental health services would be added to the list of services that must be covered.

Medicaid is a cost-sharing partnership between the federal government and the states. The federal government pays a share of the medical assistance expenditures under each state's Medicaid program. That share, known as the Federal Medical Assistance Percentage (FMAP), is determined annually by a formula that compares the state's average per-capita income level with the national income average. States with higher per-capita income levels are reimbursed smaller shares of their costs. By law, the FMAP cannot be lower than 50 percent or higher than 83 percent. In 2011, the FMAPs varied from 50 percent (13 states) to 74.7 percent (Mississippi), with the median federal share among all states being 60.5 percent.

This means that a state with an FMAP of 50 percent is matched 50 cents for every 50 cents that it contributes. Hence that state is paying at a rate of one-half ( $\$0.50/\$1.00$ ) of the actual price. In contrast, a state with an FMAP of 74.7 percent is matched 74.7 cents for every 50 cents it contributes, giving a rate of 40.1 percent ( $\$0.50/\$1.247$ ) of the actual price.

Figure 21-4 shows how this formula works. Suppose that a state was previously providing health services expenditure level  $H^*$  for the poor, and expenditure level  $A^*$  for everything else, at point  $E$ , leading to utility level  $U^1$ . These expenditure patterns presumably reflected the preferences of the public for taxing themselves to spend for various items for their residents. The slope of the budget constraint reflects the relative costs of providing the services. The Medicaid cost share reduces the relative cost of health care for the poor, thus rotating out the  $x$ -axis of the diagram, as shown by the dashed line. If unconstrained, the state might be able to increase expenditures on both  $A$  and  $H$  and get to point  $F$ , and utility level  $U^2$ .



**FIGURE 21-4** The Impact of Medicaid Cost Sharing on State Expenditures

However, we should analyze Medicaid's requirements as constraints that require states to provide (and tax their residents for) a mandated expenditure level  $H_m$  of health services which exceeds  $H^{**}$ . Medicaid is thus not a "blank check" for the states; residents must provide their own shares of expenses through taxes, and provide services to specified groups of the needy in order to participate. The requirement that states provide stipulated levels of services in order to receive Medicaid funding constrains state behavior, and may reduce the utility of the representative voter. This is shown by the utility level  $U^3$  at point  $G$ . Although Medicaid program participation certainly increases the utility of the Medicaid beneficiaries, program mandates reduce utility level from level  $U^2$  at point  $F$ . Nonetheless, the fact that all states choose to participate in Medicaid with its cost-sharing and mandated benefits indicates the overall desirability of this program to the states' residents.

**THE SCOPE AND DURATION OF MEDICAID SERVICES** The Medicaid program allows considerable flexibility within the states' Medicaid plans (see Box 21-1 for a particular example related to Oregon). Because the states do vary, analysts can compare state programs to determine how differing program features might work. However, a state's Medicaid program must offer medical assistance for certain basic services to most categorically needy populations, including inpatient hospital services, outpatient hospital services, prenatal care, vaccines for children, physician services, nursing facility services for persons age 21 or older, and family planning services and supplies.

Within broad federal guidelines and certain limitations, states determine the amount and duration of services offered under their Medicaid programs. States may limit, for example, the number of days of hospital care or the number of physician visits covered. States must provide sufficient levels of services to achieve the purpose of the benefits, and benefit limits may not discriminate among beneficiaries based on medical diagnosis or condition.

### BOX 21-1

#### Oregon Medicaid's Doctor-Assisted Suicide

Perhaps nowhere is the state-level autonomy in the U.S. Medicaid system more apparent than in the items various states choose to cover. In late February 1998, the state of Oregon's Health Services Commission voted 10-1 to include doctor-assisted suicide on the list of "treatments" covered for Medicaid patients, reported Peter Steinfels of the *New York Times*. This decision joined doctor-assisted suicide to other forms of "comfort care" for any "terminal illness, regardless of diagnosis." Residents of Oregon had voted twice, in 1994 and again in 1997, to legalize doctor-assisted suicide but neither vote had dealt with the public financing of the procedure.

Dr. Alan Bates, who headed the commission, acknowledged the divisive nature of the decision. He noted, however, that if dying people with private insurance could pay for medical help in taking their own lives, why should poor people be deprived of the same opportunity?

The federal government bars all federal support, direct or indirect, for assisted suicide, so Oregon must segregate carefully the medical services involved in assisting suicide and pay for them with the state's own dollars. Reporter Steinfels noted that in effect, the federal government is saying that taxpayers throughout the country should be spared "moral complicity" in Oregon's innovation.

The controversy continued. In 2002, then-U.S. Attorney General John Ashcroft went to court to challenge Oregon's practices. After numerous appeals, the U.S. Supreme Court agreed to hear the case, and in January 2006 it ruled, in a 6-3 vote, that a federal drug law could not be used to prosecute Oregon doctors who prescribed overdoses intended to facilitate the deaths of terminally ill patients. Writing for the majority, Justice Anthony Kennedy noted, "He [Ashcroft and his successor, Alberto Gonzales] is not authorized to make a rule declaring illegitimate a medical standard for care and treatment of patients that is specifically authorized under state law."

*Source:* Steinfels, Peter, "Oregon Medicaid's Doctor-Assisted Suicide," *New York Times*, March 7, 1998, National/Metro Section.

**PAYMENT FOR MEDICAID SERVICES** Under Medicaid, states may pay health care providers directly on a fee-for-service basis or through various prepayment arrangements, such as HMOs. Each state has broad discretion in determining the payment methodology and payment rate for services. Generally, payment rates must be sufficient to enlist enough providers so that covered services are available at least to the extent that comparable care and services are available to the general population within that geographic area. Providers participating in Medicaid must accept Medicaid payment rates as payment in full. States must make additional payments to qualified hospitals that provide inpatient services to a disproportionate number of Medicaid beneficiaries and/or to other low-income or uninsured persons under what is known as the “disproportionate share hospital” (DSH) adjustment.

States may impose nominal deductibles, coinsurance, or copayments on some Medicaid recipients, for certain services, but pregnant women and children under age 18, are excluded from cost sharing. All Medicaid recipients must be exempt from copayments for emergency services and family planning services.

**MEDICAID SUMMARY** Medicaid started as a medical care extension of federally funded programs providing cash income assistance for the poor, with an emphasis on dependent children and their mothers, the disabled, and the elderly. Legislation in the late 1980s assured Medicaid coverage to an expanded number of low-income pregnant women, poor children, and some Medicare beneficiaries who are not eligible for any cash assistance program. Legislative changes also focused on increased access, better quality of care, specific benefits, enhanced outreach programs, and fewer limits on services.

Most Medicaid recipients require relatively small average expenditures per person each year. The 2009 data indicate mean Medicaid payments for non-disabled children of about \$2,848 per child. Per capita spending for non-disabled children (\$2,848) and adults (\$4,123) was much lower than that for aged (\$15,678) and disabled beneficiaries (\$16,563). This reflects the differing health status and use of services by the members of these groups.

In 2008 Medicaid paid for 40.6 percent of the total cost of nursing facility care and 34.7 percent of home health services care. With the elderly or disabled percentage of the population increasing faster than the younger groups, the need for long-term care is expected to increase.

### **The Medicaid-Medicare Relationship**

The Medicare and Medicaid programs work jointly for many beneficiaries, called “dual eligibles.” The Centers for Medicare and Medicaid Services estimate that Medicaid provides some level of supplemental health coverage for about 9 million Medicare beneficiaries. Medicare beneficiaries who have low incomes and limited resources also may receive help from the Medicaid program. For those eligible for full Medicaid coverage, the Medicare health care coverage is supplemented by services available under their state’s Medicaid program, according to eligibility category. These additional services may include, for example, nursing facility care beyond the 100-day limit covered by Medicare, as well as eyeglasses and hearing aids. For persons enrolled in both programs, Medicare pays first for services because Medicaid is always the “payer of last resort.”

### **Medicare and Medicaid: Conflicting Incentives for Long-Term Care**

The structures of Medicare and Medicaid can create conflicting incentives regarding dually eligible beneficiaries, without coordination of their care. Both programs have interests in limiting their own costs, but neither has an incentive to take responsibility for the management or quality of care.

David Grabowski (2007) explains that Medicare beneficiaries who meet Medicaid’s (low) income and resource eligibility standards may become dually eligible (for both programs). Under federal rules, most states are required to offer Medicaid coverage to recipients of the Supplemental Security Income (SSI) program. In 2011, the SSI income limit for an unmarried individual is



\$1,133. However many states' Medicaid programs cover elderly people who have incomes up to 100 percent of the federal poverty level and assets that do not exceed the SSI threshold. In regard to income, the states adopted two broad sets of rules that expand eligibility:

- “medically needy” programs, and
- special income rules.

If individuals' incomes exceed the state's income test, medically needy programs subtract medical and long-term care expenses from their incomes in calculating Medicaid eligibility. Other states have special income rules for people in nursing homes and in home- and community-based services (HCBS) waiver programs, which extend eligibility up to 300 percent of the SSI income limit.

Either Medicare or Medicaid may have the incentive to shift costs to the other. According to the rules, Medicare is the primary payer for dual eligibles' hospital, physician, and other acute medical care; Medicaid (according to the states' discretions) can choose to pay the often considerable Medicare copayments for the dual eligibles. If the states are seeking to reduce their Medicaid expenditures, they may restrict their cost-sharing paying. This may result in less access and less treatment for beneficiaries in states with more restrictive policies.

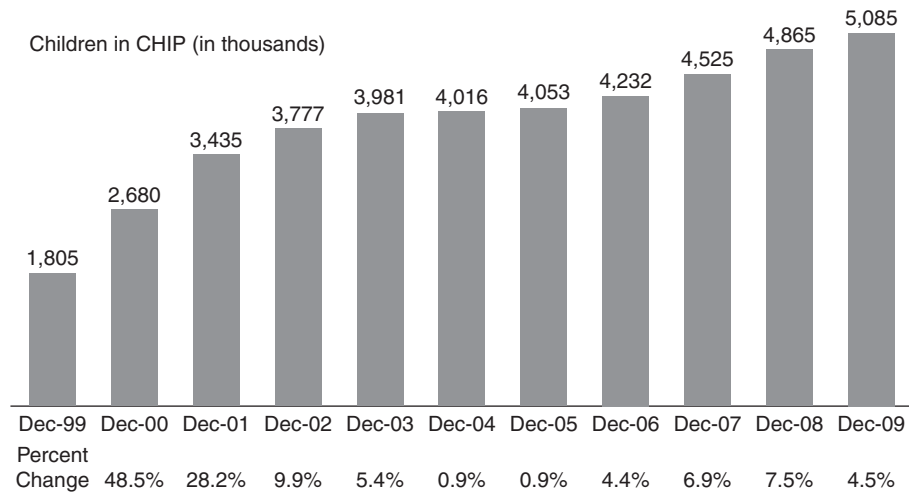
The adverse incentives can also go in the other direction. Most analysts believe that Medicare's 1983 adoption of DRGs for hospital care, led to patients' being discharged “sicker and quicker.” This change in payment contributed to the growth in Medicare-covered post-acute nursing home care in the years following prospective payment (Dalton and Howard, 2002).

The transfer of patients from the hospital to the nursing home also raises issues related to the coordination of care and the beneficiaries' health. Under the Medicare hospital prospective payment, discharge planners have more incentive to discharge patients as soon as (safely) possible but less incentive to consider the long-term cost and health implications of the initial discharge placement. Given the high number of Medicare nursing home stays that ultimately become Medicaid nursing home stays, it is desirable that the nursing home to which a hospitalized patient is discharged participate in Medicaid, even if Medicare finances the initial stay. Such placements would remove the need to transfer patients when their Medicare coverages end, thereby avoiding the adverse health consequences of transfers. Similarly, hospital discharge planners would ideally avoid transfers to nursing homes when adequate home care is available to support a community-based placement. This could improve the patient's welfare and lower Medicaid's spending, but under the current Medicare payment system, discharge planners are not rewarded for placing patients in the most appropriate setting, having little incentive to consider the long-term implications of the discharge placement for either the beneficiary's long-term health or Medicaid's budget.

What can be done about the conflicting incentives? The 2010 PPACA established the Federal Coordinated Health Care Office (CHCO), known as the “office of the duals.” This office will work closely with the Center for Medicare and Medicaid Innovation to streamline care for people dual eligibles. The office will address the cost-shifting and inconsistencies that can contribute to fragmentation of care, particularly as patients move back and forth from hospital, home, rehab and long-term care—with some services and settings under Medicare's purview and some under Medicaid's.

### **Children's Health Insurance Program—CHIP**

The State Children's Health Insurance Program, or SCHIP, was established in the Balanced Budget Act of 1997. Now called CHIP, and designed as a federal-state partnership, similar to Medicaid, it seeks to provide health insurance to children whose families earn too much money to participate in Medicaid, but not enough money to purchase private insurance. CHIP is the largest expansion of health insurance coverage for children since the initiation of Medicaid in the mid-1960s.



**FIGURE 21-5** U.S. CHIP Monthly Enrollment, 1999–2009

Source: <http://www.kff.org/medicaid/upload/7642-05.pdf>, accessed June 8, 2011. This information was reprinted with permission from the Henry J. Kaiser Family Foundation. The Kaiser Family Foundation is a non-profit private operating foundation, based in Menlo Park, California, dedicated to producing and communicating the best possible information, research, and analysis on health issues.

Figure 21-5 shows that in December 2009 monthly “point in time” enrollment was 5.1 million; the number of children enrolled any time during that year was close to 7.7 million.<sup>8</sup>

CHIP aims to provide coverage to “targeted low-income children.” A “targeted low-income child” is one who resides in a family with income below 200 percent of the FPL or whose family has an income 50 percent higher than the state’s Medicaid eligibility threshold. Some states have expanded CHIP eligibility beyond the 200 percent FPL limit, and others cover entire families and not just children.

If a state elects to establish an expanded Medicaid program using CHIP, the eligibility rules of Medicaid apply, and the services provided under CHIP mirror the Medicaid services provided by that state. Regardless of the type of health benefits coverage provided by a state, they must provide coverage for well-baby and well-child care, immunizations, and emergency services.

## PUBLIC INSURANCE AND HEALTH

How does public insurance affect health? In part, this depends on how effective the public insurance programs are in reaching their targeted populations. Janet Currie (2006) argues that countries with universal programs seek to maintain a minimum level of service for all individuals, at a reasonable cost to government. In the United States, the goal is to maintain such a standard for selected groups of vulnerable or “deserving” individuals, such as children, the elderly, and the disabled. Targeting will always be imperfect. Some of those who take up benefits will not “deserve” them, and some of those who are eligible for benefits will not take them up. If take-up by eligible individuals is low,

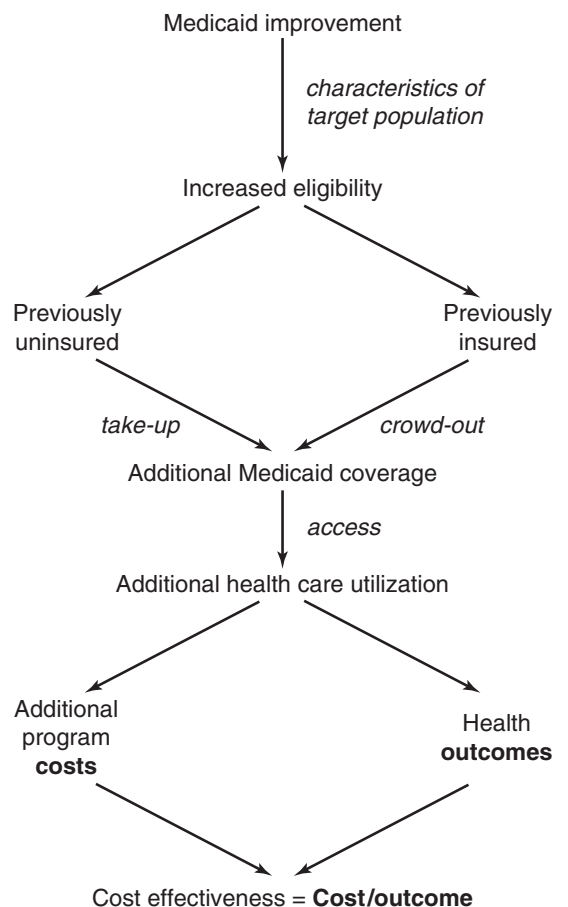
<sup>8</sup> Tabulations of those unemployed or uninsured often confuse “ever” with “point-in-time.” KFF (2011) notes that the annual count of children ever-enrolled will always exceed the number enrolled at any point-in-time, as long as new enrollments and departures occur during the year. The greater the number of new enrollments and departures, the greater will be the difference between the point-in-time and annual ever-enrolled counts. In the years leading up to 2009, over one-third of CHIP enrollees enrolled at any time during the year were not enrolled at the end of the year.

then the program may fail to reach its main goal of helping the targeted group. If take-up by ineligible is too high, then government revenues will be diverted from other productive uses.

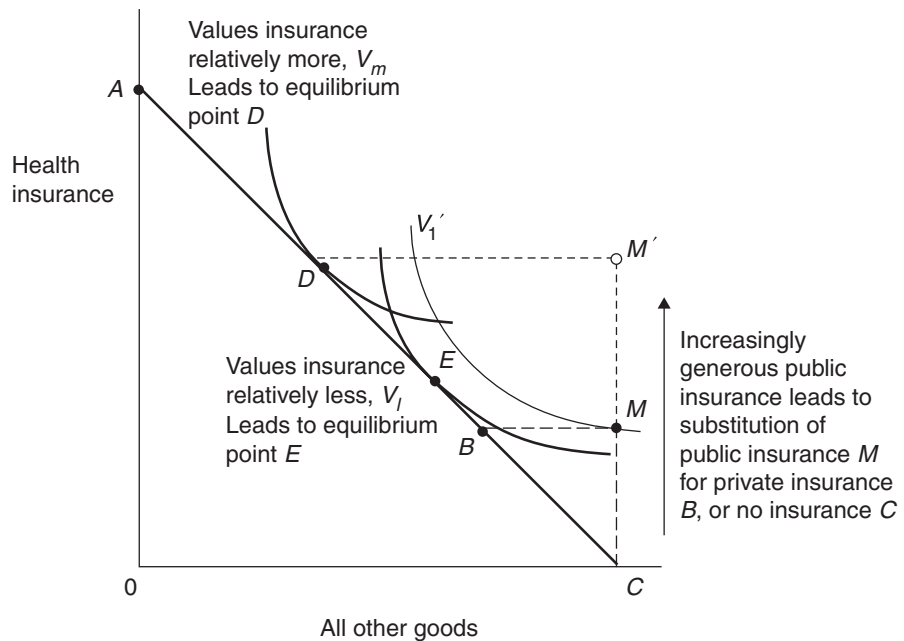
Researchers have identified two categories of impediments to program take-up. The first is program stigma, the premise that some people are embarrassed or afraid to apply for programs, even though they might benefit greatly. Second, individuals face costs of learning about and applying for programs, and these costs may deter some from using them. Moreover, the costs may be highest for precisely those individuals in greatest need, and in cases where the beneficiaries are young children or infirm adults, the costs may be borne by individuals other than the beneficiaries, such as parents or caretakers. The costs that these impediments impose on would-be participants may be sufficiently large to prevent them from enrolling in potentially beneficial programs.

Gruber (2002) traces the potential effects of a generic Medicaid improvement in Figure 21-6. The adoption or enhancement of a program, such as Medicaid, depending on the population characteristics, will lead to increased eligibility of the poor or the young. Some of them may have been previously uninsured. To the extent that they find the public insurance attractive, they “take up” coverage. For those who were previously insured by other means, some may choose to substitute the public insurance. In the literature, this impact is referred to as “crowd-out” since the public insurance has replaced the private insurance. Either one of these will have some measurable impact on coverage.

The increased coverage would be expected to affect health care utilization. Analysts have found that this impact depends on access to the health care, which may relate to the availability of



**FIGURE 21-6** Analyzing the Impacts of Improved Medicaid Coverage  
 Source: Adopted from Gruber (2002).



**FIGURE 21-7** Analyzing the Economics of Crowd-Out

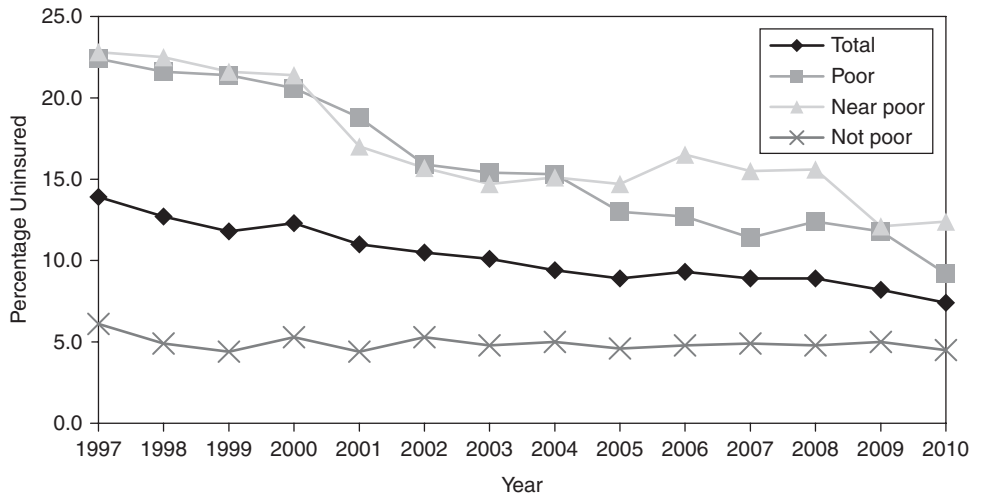
providers and the distance, cost, or convenience of dealing with the providers. Increased utilization increases costs and presumably improves outcomes, which are typically measured in terms of reduced morbidity (illness) or mortality (death). The incremental cost per unit of outcome is often summarized in terms of costs per illness day prevented, or costs per death prevented—measures of the program's cost efficiency.

Cutler and Gruber (1996) examine the economics of crowd-out. Consider a household choosing between health insurance and all other goods, and assume that more generous plans offer a greater range of providers or cover a wider set of services. As noted in Figure 21-7, households valuing insurance highly (e.g., those demanding the highest quality providers) will exhibit utility function  $V_m$  (more services), and select point  $D$ . Those valuing insurance less highly will exhibit utility function  $V_l$  (less services), and select point  $E$ .

Suppose the government introduces free public insurance with generosity  $M$ . It may have a lower value relative to the private policies for a couple of reasons. Because of low Medicaid reimbursement rates, some providers may be reluctant to treat Medicaid patients. Some households may prefer to avoid public programs because of the stigma of being enrolled. Households cannot purchase a supplement to Medicaid; if they want higher insurance, they must return to the original budget constraint. Hence, the budget constraint is the kinked set of segments  $ABMC$ . Responding to the public coverage, people with low values of private insurance, such as those at point  $E$ , will enroll in the public sector, because utility level  $V_l'$  (passing through point  $M$ ) is higher than  $V_l$ . Households with a high valuation of insurance will retain their private insurance at point  $D$ . As the value of the public insurance (point  $M$ ) rises (say to  $M'$ ), the households are more likely to drop private insurance and enroll in Medicaid.

To the extent that the Medicaid coverage provides insurance where none has been available, one observes take-up; to the extent that it replaces existing insurance, one observes crowd-out. One may even see a household choosing less coverage (point  $M$  provides less generous coverage than existing point  $E$ ) because it is free. Cutler and Gruber estimate that under Medicaid expansions about half of the increase in Medicaid eligibility has been associated with a reduction in private insurance coverage (crowd-out), with the remainder representing take-up by those who were previously uninsured.

The impact of children's health programs has been substantial. Figure 21-8 traces the impact of CHIP on the percentage of children without health insurance. We note that while the percentage



**FIGURE 21-8** Percentage Uninsured, 1997 - March 2010, Age Under 18 years  
 Source: Health Insurance Coverage: Early Release of Estimates From the National Health Interview Survey, January-March 2010, from Centers for Disease Control, September 21, 2010, <http://www.cdc.gov/nchs/data/nhis/earlyrelease/insur201009.htm#table1>, accessed June 8, 2011.

of children above 200 percent of the poverty level (the “not poor”) stayed roughly constant from 1997 to 2010, the percentage of those below 200 percent, and particularly those below the poverty line, fell from well over 20 percent to less than 12 percent over the 13-year period. This is particularly notable given the increase in poverty that accompanied the hard economic times in the U.S. toward the end of the first decade of the twenty-first century.

Another example illustrates issues of take-up, crowd-out, and program costs in recent economic times.<sup>9</sup> In October 2008 Hawaii announced that it was dropping the only state universal child health care program in the country just seven months after it was launched. Governor Linda Lingle’s administration cited budget shortfalls and other available health care options for eliminating funding for the program. A state official said that some families had dropped existing private coverage so their children would qualify for the subsidized plan, an example of crowd-out.

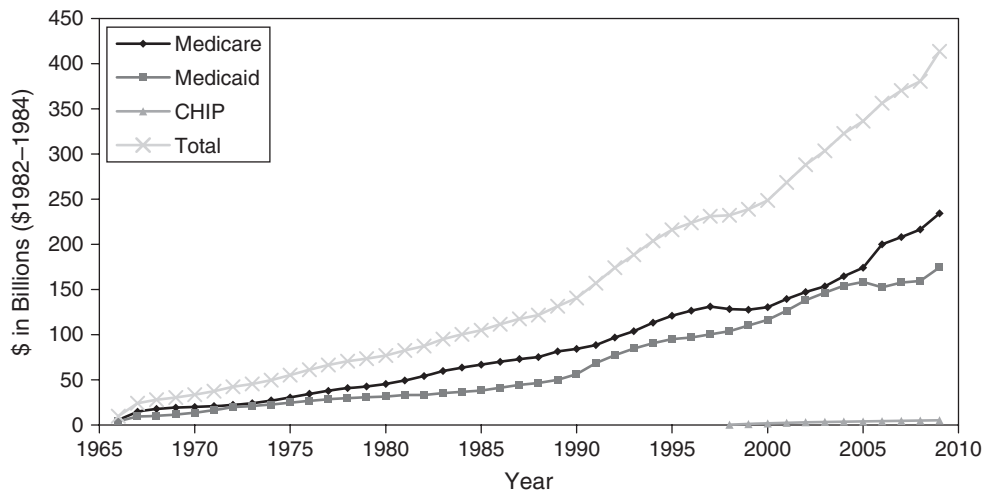
The Keiki (Child) Care program had sought to cover every child from birth to 18 years old who didn’t have health insurance—mostly immigrants and members of lower-income families. Estimates of those lacking health coverage ranged from 3,500 to 16,000 in a state of about 1.3 million people. All were eligible for the program. The universal health care system was free except for copays of \$7 per office visit. Governor Lingle had signed Keiki Care into law in 2007, but it and many other government services faced cuts as the state dealt with a projected \$900 million general fund shortfall.

## THE EFFECTS OF MEDICARE AND MEDICAID

Though we can be certain about the provisions of Medicare and Medicaid, we are necessarily less certain about their effects. We consider here a selection of findings on the effects of Medicare and Medicaid on:

- Health care costs
- Access to health care
- Health status

<sup>9</sup> The source of the discussion is “Hawaii Ends Universal Health Care for Kids,” <http://www.foxnews.com/story/0,2933,439607,00.html>, accessed June 8, 2011.



**FIGURE 21-9** Inflation Adjusted Expenditures for Medicare and Medicaid, 1966–2009  
*Source:* Centers for Medicare & Medicaid Services, Office of the Actuary, National Health Statistics Group; U.S. Department of Commerce, Bureau of Economic Analysis; and U.S. Bureau of the Census., <https://www.cms.gov/NationalHealthExpendData/downloads/tables.pdf>

## Costs and Inflation

The implementation of Medicare and Medicaid coincided with a considerable increase in health care costs in the United States. While health care costs had been rising before 1965, both in simple percentage terms and in comparison to the general rate of inflation, the hospital care inflation rate increased somewhat after the implementation of Medicare and Medicaid.

The expenditure levels of the two programs increased much more rapidly than most people had expected. Figure 21-9 shows the pattern of expenditures over time. In monetary terms, percentage increases in expenditures on both programs were in double digits for most of the years. Even accounting for the substantial rate of inflation since the late 1960s, Medicare outlays were 11.8 times higher in 2009 than in 1970, and Medicaid outlays were 12.8 times higher.

The populations at risk also grew for both Medicare and Medicaid. From 1970 to 2006, the number of Medicare enrollees grew at a compounded rate of about 2 percent—including disabled enrollees beginning in 1973. This rate exceeded population growth and was due in part to the aging of the population. Growth in the enrollee population often is cited as the primary reason for growth in Medicaid payments. Responding to the availability of Medicaid, which provided cost sharing for many programs, many states changed their needs standards for several types of programs, allowing more families to qualify for cash assistance.

**WHY SPENDING HAS RISEN: INCREASED COVERAGE, TECHNOLOGICAL IMPROVEMENT, AND INCREASED INEFFICIENCY** The increase in the eligible population covered by Medicare and Medicaid clearly helps explain why program expenditures have risen, but it does not fully account for the inflationary effects. Newhouse (1978a) suggested three ways through which insurance programs, such as these, could affect prices and costs, even without growth in the population served.

First, Medicare and Medicaid both tended to increase the insurance coverage of the populations eligible. An increase in insurance expands the demand for care. Second, insurance coverage may induce technological improvements. If so, then the cost per unit of care may rise.

Finally, Newhouse proposed a third theory for the effect of insurance on costs and quantity used. This may be called the “increased inefficiency” theory. The idea is that when insurance covers

a substantial portion of the health care bill, institutions, such as hospitals, have less incentive to control costs. It is not clear from this theory that the advent of Medicare and Medicaid, for example, would cause the level of inefficiency in hospitals to increase over time, but such a pattern is at least consistent with the theory.

**THE EVIDENCE** What do we know about the patterns of health expenditure inflation subsequent to the adoption of Medicare and Medicaid, and what do we know about the sources of this inflation? One approach partitions the observed rise in expenditures into its logical components: changes in population, in quantity per capita, and in the nature of services provided per visit or per admission.

Cutler and Meara (1997) find a dramatic change in medical spending over time, and disproportionately so for the very young (those younger than 1 year old) and the old (those age 65 or over). From 1963 through 1987, per-person spending on infants increased by 9.8 percent per year, and per-person spending on the elderly increased by 8.0 percent per year (compared to 4.7 percent per year for the others).

Moreover, they find that essentially all of the disproportionate growth of spending for the very young and the old was accounted for by high-cost users within those groups, and that a substantial amount of high-cost medical use is associated with the increasing technological capabilities of medicine. Among infants, high-cost users are premature babies with substantial respiratory or other acute conditions. For the elderly, high-cost users are generally patients with severe cardiovascular problems or cancer.

More recently, Finkelstein (2007) suggests that the impact of Medicare on hospital spending is substantially larger than what the existing evidence from individual-level changes in health insurance would have predicted. She argues that the introduction of Medicare was associated with an increase in the rate of adoption of then new medical technologies. A back of the envelope calculation based on the estimated impact of Medicare suggests that the overall spread of health insurance between 1950 and 1990 may explain at least 40 percent of the increase in real per capita health spending over this time period.

Irrespective of the considerable costs, the predominant evidence seems to suggest that both Medicare and Medicaid have succeeded in addressing problems of access. That there was a change in health care use rates among the lower-income groups and the elderly following the beginning of Medicare and Medicaid is evident from the data.

Table 21-1 investigates two dimensions of health care access and utilization: the interval since the last physician contact and the number of hospital discharges per 1,000 people. In 1964, just prior to the passage of Medicare and Medicaid, 69.7 percent of those aged 65 and older had seen a physician within the past year. This was 4.2 percent higher than the general population and 8.1 percent higher than the group aged 45 to 64. By 1990, those aged 65 and older were 11.4 and 12.7 percent more likely than the respective comparison groups to have seen a physician within a year.

Another measure of access for the elderly involves hospital discharges per 1,000. In 1964, those aged 65 and older had 190.0 compared with 146.2 for the group aged 45 to 64, or a 30 percent differential. By 1990, the differential had grown to 83.3 percent, and by 1995 to 118.1 percent.

Similar comparisons are appropriate when comparing the less affluent to the more affluent. In 1964, those with incomes less than \$15,000 were 79.6 percent as likely to have seen a physician in the past year as those with incomes higher than \$50,000. By 1990, they were 94.6 percent as likely.

In 1964, those with incomes less than \$15,000 had 102.4 hospital discharges per 1,000, compared with 110.7 discharges for those with incomes higher than \$50,000, or only 92.5 percent as many. By 1990, the lower-income people had 96.1 percent more discharges per 1,000 people, and by 1995 they had 128.4 percent as many.

Clearly, many other factors in the health care system have changed since the enactment of Medicare and Medicaid. One might ask whether poor people overuse inpatient care, seeking hospital emergency care instead of more appropriate outpatient care. Nonetheless, by these gross measures, the poor and the elderly improved their access dramatically since 1965, with a boost from Medicare and Medicaid.

Would we find that the social insurance programs increased access to care in a study that controlled among other things, for the level of health status or health care need, rather than just

**TABLE 21-1 Indirect Impacts of Medicare and Medicaid**

<b>A. Age Comparisons</b>						
<b>Interval since last physician contact</b>						
	<b>Percent Less Than One Year</b>			<b>Age 65+ Relative to:</b>		
	<b>1964</b>	<b>1990</b>	<b>1995</b>	<b>1964</b>	<b>1990</b>	<b>1995</b>
Total	66.9	78.2	79.5	1.042	1.114	1.132
Age 45–64	64.5	77.3	79.9	1.081	1.127	1.126
Age 65+	69.7	87.1	90.0			
<b>Hospital discharges per 1,000</b>						
	<b>1964</b>	<b>1990</b>	<b>1995</b>	<b>1964</b>	<b>1990</b>	<b>1995</b>
Total	109.1	91.0	86.2	1.742	2.734	3.096
Age 45–64	146.2	135.7	122.4	1.300	1.833	2.181
Age 65+	190.0	248.8	266.9			
<b>B. Income Comparisons</b>						
<b>Interval since last physician contact</b>						
	<b>Percent Less Than One Year</b>			<b>Income &lt; \$15,000 Relative to:</b>		
	<b>1964</b>	<b>1990</b>	<b>1995</b>	<b>1964</b>	<b>1990</b>	<b>1995</b>
Total	66.9	78.2	79.5	0.876	0.988	0.984
Income < \$15,000	58.6	77.3	78.2			
Income > \$50,000	73.6	81.7	83.5	0.796	0.946	0.937
<b>Hospital discharges per 1,000</b>						
	<b>1964</b>	<b>1990</b>	<b>1995</b>	<b>1964</b>	<b>1990</b>	<b>1995</b>
Total	109.1	91.0	86.2	0.939	1.563	1.632
Income < \$15,000	102.4	142.2	140.7			
Income > \$50,000	110.7	72.5	61.6	0.925	1.961	2.284

Source: Derived from *Health United States* (1998), Tables 77 and 87.

comparing data before and after the event? The answer appears to be yes. Davis and Reynolds (1976) examine the effect of access to these public assistance programs in a statistical context that controls for other factors. For people with the same health status, they compare average health care use by those eligible for assistance with those low-income people not receiving assistance. Examine Table 21-2 excerpted from their study. These data indicate a fairly strong impact of improved financial access on care.

**TABLE 21-2 Annual Predicted Utilization for Low-Income Persons by Health Status and Welfare Eligibility, Adjusted for Other Characteristics, 1969**

	<b>Health Status</b>					
	<b>Public Assistance Recipients</b>			<b>Other Low-Income Persons</b>		
	<b>Good</b>	<b>Average</b>	<b>Poor</b>	<b>Good</b>	<b>Average</b>	<b>Poor</b>
Physician Visits	4.09	4.95	7.10	2.29	3.36	5.12
Hospital Admissions	0.14	0.16	0.21	0.09	0.11	0.15
Hospital Days	2.40	2.72	3.47	1.18	1.42	2.04

Source: Karen Davis and Roger Reynolds, "The Impact of Medicare and Medicaid on Access to Medical Care" in Rossett (ed.), *The Role of Health Insurance in the Health Services Sector*, New York: Neal Watson, 1976, with permission.



Before concluding the discussion on access to health care, we identify a characteristic of access specific to the Medicaid program. Because Medicaid is state operated, as we have noted, characteristics vary somewhat across the nation. As recently as 1987, Medicaid was fundamentally restricted to children in very low-income, single-parent families. Since then, a series of legislative initiatives has extended Medicaid more broadly among children. By 1996, all pregnant women and all children under age 6 who had family incomes below 133 percent of the poverty line, and all children age 13 and younger with family incomes below 100 percent of the poverty line, were eligible for Medicaid. Coverage is being phased in for all children born after September 1983 until poor children of all ages are covered.

The net result is that the proportion of the child population enrolled in Medicaid increased from 15.7 percent in 1989 to 23.2 percent in 1995. For children under the age of 6, the percentage grew from 20.3 to 29.6. Interestingly, at the same time that Medicaid eligibility increased, there was a reduction in children's private insurance coverage. There are two plausible reasons. First, the 1990s saw a leveling-off in the variety and generosity of employer-provided health insurance, affecting both low- and high-income households. To the extent that lower-income households have been affected disproportionately by changes in job-related insurance, they may have substituted Medicaid. Second, newly eligible families may have voluntarily dropped private insurance coverage in order to enroll in Medicaid.

Coincident with the expansion of Medicaid since the late 1980s, analysts have seen an increase in physician contacts for the poorest segment of the population. Table 21-3 shows the access to physicians over a six-year period. Those in good or excellent health increased their average number of contacts by 11.1 percent, and those in fair or poor health increased their average number of contacts by 31.5 percent. Although increased federal coverage under Medicaid comes with increased state matches, Medicaid scholars have noted that unlike cash assistance welfare, health care is viewed as a "positive good" that everyone should receive. Holahan, Wiener, and Wallin (1998, p. 58) noted that "declining Medicaid caseloads are matters of concern for state officials while falling cash welfare caseloads are matters of pride."

## Health Status

The ultimate question regarding Medicare and Medicaid is whether these programs improve the health status of the eligible populations. This question is a specific application of the production function of health. Early econometric studies of the effects of health care on mortality rates suggested small and insignificant effects. The RAND Health Insurance Experiment with its randomized format found little or no effect on health status in most cases for increased health care consumed by people with better insurance coverage. Into the mid-1980s, the evidence seemed to weigh against the belief that health care made a significant difference in the health status of populations generally, and the insured populations in particular.

Card, Dobkin, and Maestas (2009) look at the impact of Medicare by examining over 400,000 hospital admissions to the emergency room for "non-deferrable" conditions— diagnoses with the

**TABLE 21-3 Average Number of Physician Contacts in Last Year for Children under Age 15**

Family Income	Good/Excellent Health			Fair/Poor Health		
	1987–1989	1993–1995	Percent Change	1987–1989	1993–1995	Percent Change
Below Poverty Line	3.6	4.0	11.11	10.8	14.2	31.48
Poverty Line to Twice Poverty Line	3.8	3.9	2.63	15.2	16.2	6.58
Above Twice Poverty Line	5.0	4.9	–2.00	22.6	20.7	–8.41

Source: Derived from Table 3-7, *Economic Report of the President* (1998).

same daily admission rates on weekends and weekdays. There is no discernible rise in the number of admissions at age 65, suggesting that the severity of illness is similar for patients on either side of the Medicare threshold. The insurance characteristics of the two groups differ, however, with a large jump at 65 in the fraction who have Medicare as their primary insurer, and a reduction in the fraction with no coverage. These changes are associated with significant increases in hospital list charges, in the number of procedures performed in hospital, and in the rate that patients are transferred to other care units in the hospital. The authors estimate a nearly 1 percentage point drop in 7-day mortality for patients at age 65, implying that Medicare eligibility reduces the death rate of this severely ill patient group by 20 percent. The mortality gap persists for at least two years following the initial hospital admission.

Finkelstein and McKnight (2008) remind us that Medicare is a form of insurance against risk. They calculate that the welfare gains from reductions in risk exposure alone may be sufficient to cover between half and three-quarters of the costs of the Medicare program. They view these findings as underscoring the importance of considering the direct insurance benefits from public health insurance programs, in addition to any indirect benefits from an effect on health.

Favorable impacts come with incremental costs. Currie and Gruber (1996) measure the impacts of increased Medicaid eligibility (throughout the United States) for pregnant women between 1979 and 1992. Certain groups saw substantial improvements. For example, a 30 percentage point increase in eligibility among 15- to 44-year-old women was associated with a decrease in infant mortality of 8.5 percent. However, even the most carefully targeted changes in Medicaid eligibility cost the Medicaid program \$840,000 per infant life saved, raising important questions of cost effectiveness. In a similar study, Joyce (1999) finds reductions in newborn costs associated with Medicaid participation (this time in New York) to be between \$100 and \$300 per recipient, insufficient to offset program expenditures.

## Medicare: Recent Changes and Future Prospects

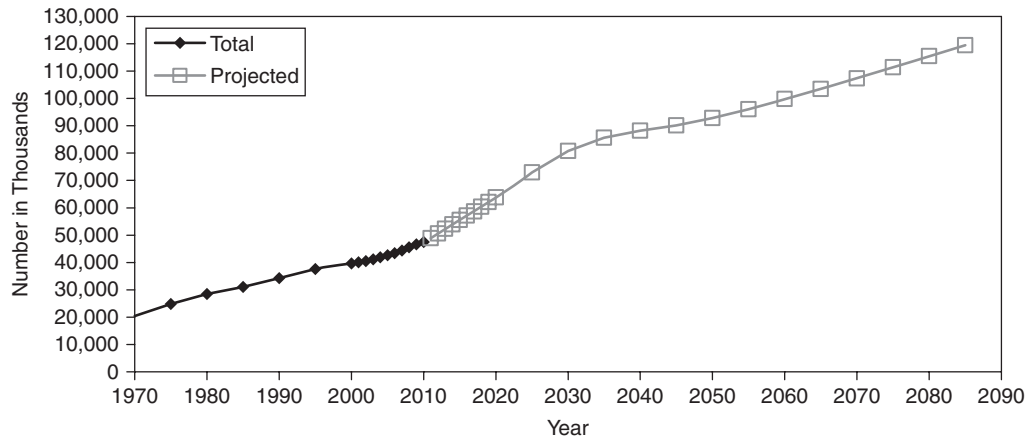
**RECENT CHANGES** In 1996, trustees of the Hospital Insurance (Part A) Fund predicted that the Part A Trust Fund would have a zero balance by 2001. With little appetite for increased payroll taxes, the U.S. Congress chose to make major changes in how Medicare paid health care providers through the Balanced Budget Act (BBA) of 1997.

The BBA increased the incentives for efficient production by mandating the development of prospective reimbursement systems for post-acute care. For hospital outpatient departments, it ended cost-based reimbursement. These two changes virtually ended cost-based reimbursement throughout the Medicare system. Payment formulas for new entrants and for home health services were adjusted downward, and the BBA reduced physician payments.

Representatives of hospitals and physicians argued that the measures were unexpectedly harsh. Since the low Medicare outlays in 1998 and 1999 were well below projected outlays, Congress provided givebacks amounting to about 3 percent of Medicare spending.

**FUTURE PROSPECTS** Despite the major changes in the 1997 BBA it is clear that the U.S. Medicare system will become much larger over the next quarter century. Figure 21-10 displays projections of the Medicare-eligible population, starting in 2010. Projections into the future can be risky, but this one is a safe bet. All those who will be 65 years of age in 2040 are already over 35 years old. To project future populations, demographers statistically “age” the various population cohorts by predicting deaths between now and then. Immigration and emigration generally provide only small adjustments at the national level.

Any way that one looks at things, Medicare will grow. As noted in Figure 21-10, the number of Medicare beneficiaries, largely those ages 65 and over, will increase to 79 million, the result of the baby boom starting in the late 1940s and extending through the early 1960s. In addition, the number of workers (who are also paying into the fund) per beneficiary is projected to fall, from 3.4 in 2010 to 2.3



**FIGURE 21-10** Number of Medicare Beneficiaries, 1970–2085 *Source:* Medicare Trustees Report 2011.

in 2030. This decrease in workers per beneficiary suggests that there will undoubtedly be more financial pressure on providers to cut costs and on payers to pay for services.

We must treat projections of future spending with caution, and with the uncertainty surrounding the implementation of PPACA, this caution must be re-emphasized. The aging population and expected increases in health care costs suggest a major increase in the Medicare share of the GDP. Table 21-4 shows how analysts expect Medicare spending which was 3.1 percent of GDP in 2007, to jump to 3.99 percent of GDP by 2020 and to 5.94 percent by 2050.

**TABLE 21-4 Medicare Expenditures as a Percentage of GDP**

**HI and SMI Incurred Expenditures as a Percentage of the Gross Domestic Product**

Calendar year	HI	SMI		Total
	Part A	Part B	Part D	
Historical data:				
1970	0.52%	0.22%	—	0.74%
1975	0.73	0.30	—	1.03
1980	0.91	0.41	—	1.32
1985	1.12	0.56	—	1.68
1990	1.14	0.76	—	1.90
1995	1.58	0.90	—	2.47
2000	1.31	0.94	—	2.25
2001	1.38	1.01	—	2.39
2002	1.42	1.06	—	2.48
2003	1.41	1.12	—	2.52
2004	1.43	1.17	0.00%	2.61
2005	1.45	1.21	0.01	2.68
2006	1.45	1.27	0.33	3.05
2007	1.46	1.31	0.36	3.13
2008	1.54	1.28	0.38	3.20
2009	1.67	1.46	0.41	3.54
2010	1.69	1.46	0.43	3.58

**TABLE 21-4 (continued)**

Calendar year	HI Part A	SMI		Total
		Part B	Part D	
Intermediate estimates:				
2011	1.71	1.50	0.44	3.65
2012	1.71	1.38	0.47	3.56
2013	1.69	1.40	0.50	3.59
2014	1.68	1.43	0.51	3.62
2015	1.64	1.45	0.54	3.62
2016	1.63	1.47	0.56	3.66
2017	1.64	1.50	0.59	3.72
2018	1.65	1.53	0.61	3.79
2019	1.67	1.57	0.64	3.87
2020	1.70	1.63	0.67	3.99
2025	1.86	1.91	0.83	4.59
2030	2.03	2.15	0.98	5.16
2035	2.19	2.29	1.08	5.56
2040	2.27	2.34	1.15	5.77
2045	2.30	2.35	1.21	5.87
2050	2.30	2.36	1.28	5.94
2055	2.28	2.37	1.35	6.00
2060	2.26	2.40	1.42	6.09
2065	2.25	2.42	1.49	6.16
2070	2.24	2.44	1.55	6.22
2075	2.21	2.44	1.61	6.25
2080	2.16	2.43	1.66	6.25
2085	2.11	2.42	1.70	6.24

Source: Medicare Trustees, *Annual Report*, 2011, P 49

These projections are based on projected annual growth of GDP of close to 5 percent until 2020, and 4.6 percent thereafter (Medicare Trustees, P. 55), increasing the denominator of the fraction relating to the percentage of GDP (Percentage =  $100 \times \text{Expenditures}/\text{GDP}$ ). Many economists, irrespective of political stripe, would find such projections to be optimistic based on historical perspective. Changing the projected GDP growth rate from 4.6 percent to even a 4.0 percent growth rate, high by historical standards, would raise the 2050 projected share from 5.94% to 7.06%.

Victor Fuchs (2000) argues that there are three major economic approaches to the crisis implicit in elderly health spending projections:

1. Slow the growth rate of health expenditures. As previously noted, this may not be desirable or feasible: undesirable because many medical advances have improved quality of life for the elderly; infeasible because the elderly want all of the care that might do them some good.
2. Impose higher taxes on the young to pay for the care for the old. Fuchs views such tax hikes as highly unlikely, as they will add to an already high burden of support that the young are asked to pay in support of the elderly.
3. Provide more of their own income by increases in work and saving. In earlier work, Fuchs (1999) examined elderly retirement savings, concluding that, “most low-income elderly could have saved more prior to age 65.”

## CRITICISMS OF THE U.S. HEALTH CARE SYSTEM

Critics of the Medicare and Medicaid delivery systems have expressed concerns about the sizable personal liabilities involved for individuals entering hospitals under Medicare. As noted earlier in this chapter, beneficiaries are responsible for sizable out-of-pocket expenditures on Medicare Parts A, B, and D. Service such as dental care, eyeglasses, and most long-term care are not covered (although most Medicare patients' out-of-pocket liabilities are reduced by private supplemental insurance or various forms of public assistance). Are these personal liabilities under Medicare too burdensome? Certainly, some Americans think so.

Because Medicaid refers to 51 separate programs, there are numerous variations in coverage. Holahan and Liska (1997) find that a major source of variation in Medicaid spending among states comes from their differences in coverage of their low-income populations. In 1994, less than half (46 percent) of all Americans below 150 percent of the poverty line were covered by Medicaid.

Gruber (2002) follows Medicaid program take-up in the 1980s and 1990s. He notes that take-up was close to full in the early 1980s, but fell over time as Medicaid expanded. By 1996, 31 percent of children were eligible for Medicaid, but only 22.6 percent were enrolled, for an average take-up rate of 73 percent, but a marginal rate that was much lower. Eligibility varied substantially among the states; from 1983 to 1996 eligibility rose by over 50 percent in New Mexico, by 45 percent in Hawaii, and by 39 percent in Arizona, but by only 1 percent in Utah and 6 percent in Rhode Island, and it actually fell by 2 percent in Alaska. There are no estimates of eligibility for Medicaid and take-up of the program by the elderly and disabled. These would be complicated because many elderly or disabled who are not currently eligible could become eligible by spending down enough of their resources to qualify.

Aside from these gaps in the Medicare and Medicaid programs, many critics of the American system characterize its main weakness as the problem of the uninsured. Approximately 50 million people at any point go without health insurance. While this does not mean that they go entirely without care, the uninsured consume only half as much health care on average as those who are insured.

Holahan and Ghosh (2005) argue that without Medicaid the numbers of uninsured might have been much higher. Medicaid spending increased by about one-third between 2000 and 2003, largely due to enrollment growth. The program saw rapid increases in enrollment of children and nondisabled adults. Many more Americans were covered by Medicaid at the end of the period than at the beginning, even though there were few expansions of and even some reductions in eligibility standards. Thus, the growth in enrollment was largely due to the economic downturn, which led to declines in incomes and made more people eligible under existing eligibility standards.

The large increase in Medicaid and SCHIP enrollment for children offset the decline in enrollment in employer-sponsored insurance. As a result, the number of uninsured children declined slightly. Medicaid enrollment among adults also rose, but the increase only partially offset the decline in employer-sponsored coverage, and there was a large increase in the number of uninsured adults.

Health spending per Medicaid enrollee grew at a rate below that experienced in the private sector, in part because states made policy choices that affected benefits and provider reimbursement rates. The large growth in Medicaid spending, 10.2 percent per year, nonetheless represented rapid growth. Medicaid played its role as a safety net, providing coverage to those facing economic declines and loss of employer-sponsored coverage, but the result was a sharp increase in program costs.

Toward the end of the first decade of the twenty-first century, with a sluggish U.S. economy, Dorn, Garrett, Holahan, and Williams (2008) examined the policy responses through Medicaid and CHIP during an economic downturn. Increased unemployment wields a double-edged sword. On one edge, they estimated that 2.5 million people (1.75 million adults and 0.75 million children) lost employer-sponsored insurance. About 1 million gained insurance through Medicaid and CHIP, about 400 thousand bought non-group coverage, and about 1.1 million became uninsured.

The other edge of the sword involved states' abilities to make up for the insurance loss through Medicaid. The authors noted that a 1 percentage point increase in the unemployment rate

**BOX 21-2****States Target Medicaid for Deep Cuts in Response to Budget Deficits**

In January 2011 the National Partnership for Women and Families reported, drawing on news from the *New York Times*, that state governments planned significant cuts to Medicaid as they faced ongoing budget deficits and federal restrictions against limiting the program's eligibility requirements. In 2014 PPACA will begin to require states to expand Medicaid eligibility, leading to an increase of 16 million beneficiaries by 2019. The federal government will cover the cost of the expansion through 2016, but states will be responsible for 10% of expansion costs by 2020.

Meanwhile, additional federal aid for Medicaid granted by the 2009 economic stimulus package was scheduled to end in July 2011, causing an increase of one-fourth to one-third in each state's share of program costs. At the same time, states were facing an estimated \$125 billion in combined deficits.

States planning significant Medicaid cuts included:

- Arizona, where Gov. Jan Brewer requested a waiver from a reform provision so that the state could drop 280,000 adults from its Medicaid program;
- California, where Gov. Jerry Brown proposed cutting \$1.7 billion from Medicaid through measures that would limit beneficiaries to 10 physician visits annually and six prescriptions monthly, among other things;
- Georgia, where Gov. Nathan Deal proposed ending coverage of dental, vision, and podiatry treatments for adults; and
- South Carolina, where lawmakers considered ending hospice care.

Meanwhile, states such as California, Texas and others were weighing further reductions of up to 10 percent in payments to providers, despite the fact that some physicians already have left the program because of low reimbursement. Some states also might increase copayments for beneficiaries or expand managed care plans to control costs. In fact, certain states, like Texas, were considering withdrawing from Medicaid entirely. Texas lawmakers, however, decided that the loss of federal matching funds made the strategy impractical.

*Sources:* National Partnership for Women and Families, January 31, 2011; [http://www.nationalpartnership.org/site/News2?page=NewsArticle&id=27551&news\\_iv\\_ctrl=0&abbr=daily2\\_](http://www.nationalpartnership.org/site/News2?page=NewsArticle&id=27551&news_iv_ctrl=0&abbr=daily2_); Sack, Kevin, "For Governors, Medicaid Looks Ripe for Slashing," [http://www.nytimes.com/2011/01/29/us/politics/29medicaid.html?\\_r=1](http://www.nytimes.com/2011/01/29/us/politics/29medicaid.html?_r=1), both accessed June 17, 2011.

leads to a 3 to 4 percent decline in state revenues. Assuming states must balance their budgets and that they cut all spending proportionately, Medicaid and CHIP faced 3 to 4 percent cuts, leading to similar reductions in the federal matches. Thus, revenue loss might be a bigger fiscal problem than increased enrollment.

The authors' 2008 analyses have proven prescient. We noted in the analytical discussion of Medicaid in Figure 21-4, that receiving Medicaid funds requires a state match and the individual states, coming out of the "Great Recession" have found that match difficult to meet. Some examples are noted in Box 21-2. Many states have sought to reduce coverage, and at least one state, Texas, has considered (although rejected) withdrawing from Medicaid entirely.

**CONCLUSIONS**

In this chapter, we have presented a general discussion of social insurance and its application to the health care sector. Almost all modern industrialized countries provide fairly comprehensive health care social insurance, and we have reviewed the history of these developments and the pattern in the United States.

The major health care social insurance programs in the United States are Medicare, Medicaid, and CHIP. These programs increase health care costs in theory and have been increasingly costly in

practice. It is clear that they have had a beneficial effect on access to care among the elderly and low-income groups, and recent studies suggest that they have a beneficial effect on health status. They leave, however, a substantial number of the poor or uninsured without health care coverage.

Whether and how the United States should move to a national health insurance program must reflect the values of the public. Comprehensive social insurance for health care in the United States would directly address and presumably solve the widely perceived problem of providing for the uninsured, a group that often includes people in the poverty, near-poverty, and other lower-income groups.

These problems can be addressed in many ways. We have noted the salient policy features or decision issues in the plans considered by Congress in 1993 through 1994 and in the vigorous debate accompanying the passage of PPACA in 2010. What can economics say about these features, whose ultimate choice or the choice of no change depends on the values of the voting public? There are areas of theory and evidence in economics that shed light on these questions. It is also useful, however, to look at health reform beyond the borders of the United States. We begin by comparing the features of health systems across countries in the next chapter.

---

## Summary

1. Several types of social insurance policies and social programs exist, usefully grouped into poverty programs, old-age assistance, disability, health, and unemployment.
2. Social program features include contributions, benefits, length of coverage, means of reimbursement to providers, and methods of determining payment levels to providers.
3. Social insurance originated in nineteenth-century Europe. Groups in the United States have supported health care social insurance programs throughout the twentieth century. This history led to the adoption of Medicare and Medicaid in 1965. Until 2010, the United States remained one of the few developed countries that had not adopted a comprehensive health care social insurance program.
4. Medicare is a national program that provides hospital insurance to the elderly, along with optional supplemental physician care insurance. The Balanced Budget Act of 1997 established several new categories of Medicare options, and prescription drug coverage (Part D) passed 2003, and started in 2006.
5. Medicaid programs are funded through matching state and federal funds and run by the states. They provide health care to certain categories of the needy and are the primary providers of nursing home aid.
6. Medicare and Medicaid expenditures have increased rapidly since the programs began, due to increased medical care prices, population covered, and quantity of care per capita consumed by the population, as well as due to changes in the nature of the services provided.
7. Medicare and Medicaid accompanied clear improvement in access to care by the lower income population, as evidenced by increased utilization rates by lower-income groups, both absolutely and relatively, to the higher-income groups.
8. With the aging of the baby boom cohort and the improvements in health care technologies in the early twenty-first century, Medicare must determine how best to structure, provide, and finance the benefits that it is providing to this growing segment of the population.

---

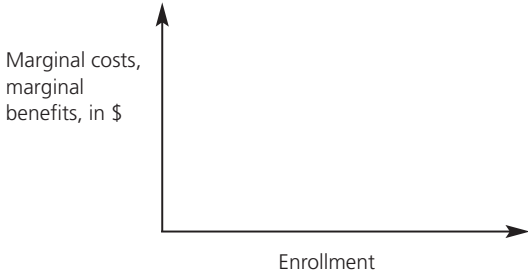
## Discussion Questions

1. In what ways does social insurance differ from private insurance?
2. Of the five types of social insurance programs described, which types characterize Medicare? Which types describe Medicaid?
3. What are the similarities between Medicare and Medicaid? What are the differences?
4. What factors contributed to the historical growth in Medicare spending?

5. Describe how Medicare has affected access to care for the elderly.
6. Is health status affected by the access to health care provided through social insurance programs? Discuss the evidence.
7. What are some possible reasons that other industrially advanced countries have far more comprehensive social insurance programs for health care than does the United States?
8. Historically in the United States, what groups have supported social insurance for health care, and what groups have opposed it? Why do you think this is the case?
9. Reductions in federal stimulus plans and decreasing state resources have affected the Medicaid program. Discuss the impact on Medicaid and state responses in the state where you live or go to school.

## Exercises

1. Calculate the average tax rate for Social Security at incomes of \$25,000, \$50,000, \$75,000, \$100,000, and \$125,000. Do the same for Medicare. (Hint: You may choose to do each graphically.) Characterize each tax as being progressive, regressive, or neutral.
2. Figure 21-2 describes the Medicare Part D prescription drug benefit. Look at the Web page <http://www.partd-medicare.com/> and determine the marginal coinsurance rates, applicable in each segment. Then calculate the average amount spent at the following levels of charges: \$2,000, \$4,000, \$6,000, and \$8,000. Discuss the “burden” of payments under this schedule.
3. Consider Currie’s discussion of take-up of social programs, where the  $x$ -axis is program enrollment and the  $y$ -axis refers to monetary costs and benefits.
 



  - (a) If we measure the number of people enrolling in a program on the  $x$ -axis, why would the “demand” for these programs be downward sloping? Draw a demand curve.
  - (b) Why would the costs of establishing a program be upward sloping? Draw a supply curve.
  - (c) What is meant by the equilibrium where supply equals demand?
  - (d) How can one model program “stigma,” and what does it do to equilibrium enrollment? Why?
4. Figure 21-6 traces the economics of take-up and crowd-out. Consider a population of 1,000 families: 200 had Medicaid insurance, 700 had some other type of insurance, and 100 were uninsured. Suppose now that Medicaid increases eligibility rules that would allow an additional 100 families to get coverage. After one year, 250 families now have Medicaid, 675 now have some other type of insurance, and 75 are uninsured.
  - (a) Calculate the average take-up and crowd-out both in numbers of families and in rates.
  - (b) Calculate the marginal take-up rates occurring due to the eligibility change.
  - (c) Has insurance coverage for the population increased? Why or why not?
  - (d) Has insurance coverage for all families increased? Why or why not?
5. Consider the analysis described in Figure 21-7. Tom and Dick each earn \$15,000 per year. Tom has a spouse and two children, and Dick is unmarried. Health insurance and other goods trade off dollar for dollar (there is no tax advantage to health insurance).
  - (a) Where would each of the two be located on the budget constraint, and why?
  - (b) Which of the two would more likely take up a health insurance program, such as Medicaid?
  - (c) How would your answers to the first two parts change if health insurance were subsidized (as it is) relative to all other goods?
6. Table 21-3 shows the effect of enhanced Medicaid programs on children who are below the poverty line. The table shows how physician utilization has increased for the poor children.
  - (a) Compare the levels of utilization between children who are below the poverty line and children who are above twice the poverty line. Who had more care in 1987? In 1993? Who is likely to be sicker?
  - (b) Compare the relative amount of utilization between the two groups. What happened to these measures from 1987 to 1993? Discuss your answer.



# Comparative Health Care Systems

- Contemporary Health Care Systems
- The United Kingdom—The National Health Service
- China—An Emerging System
- The Canadian Health Care System
- Different Systems: The Public’s Evaluation
- Differences in Health Care Spending Across Countries
- Conclusions

We now consider the experiences of other countries in providing large-scale health care. We do so for two reasons. First, many other countries have constructed programs that predate U.S. programs by decades. There are rich variations in programs and experiences that worth discovering. Second, and more importantly, the U.S. system has some gaping holes compared with the coverage extended by many other systems. Understanding the approaches used by other countries helps us to assess our own system.

## CONTEMPORARY HEALTH CARE SYSTEMS

Many industrialized countries either provide health care directly through the government or provide publicly funded health insurance with comprehensive coverage. Rather than describing details about the health care programs of dozens of countries, we will characterize the basic types of systems employed and develop a few examples in detail.<sup>1</sup>

### A Typology of Contemporary Health Care Systems

Consider a broad framework for describing different approaches. For this purpose, Margaret Gordon (1988) developed a useful typology of four health benefit systems.

1. **Traditional sickness insurance**, as provided in Germany, is fundamentally a private insurance market approach with a state subsidy.

<sup>1</sup> The Social Security Administration maintains an archive called *Social Security Statistics Throughout the World*, which describes the health, unemployment, and poverty programs of more than 100 countries. It is at [ssa.gov/policy/docs/progdsc/ssptw/](http://ssa.gov/policy/docs/progdsc/ssptw/), accessed June 27, 2011.

2. **National health insurance** plans like Canada's involve a national-level single-payer health insurance system.
3. **National health services** like the United Kingdom's have the state providing the health care.
4. **Mixed systems**, as seen in the United States, contain elements of both traditional sickness insurance and national health coverage.

In comparing economic data across countries, Table 22-1 shows per capita health expenditures expressed in U.S. dollars in many countries for 2009–2010. We adjust these figures by the purchasing powers of the local currencies (known as *purchasing power parity* or *PPP*). Other columns show each country's health care spending as a percent of GDP for selected years.

The countries vary substantially. Many European countries (including Italy, Iceland, Norway, Belgium, Austria, France, Finland, and Germany) spend larger percentages on inpatient care than does the United States. It may surprise readers to discover that in percentage terms, for 2008, the U.S. pharmaceutical expenditures were among the lowest percentages of spending, although they do constitute large absolute amounts, almost \$950 per person, due to the overall size of U.S. expenditures.

Several countries have lower crude death rates than the United States rate of 8.0 deaths per 1,000 population, including Canada with a rate of 7.1. Many countries also have higher life expectancies at birth. The United States has the largest expenditures per capita (\$7,960). It is also the biggest spender as a share of GDP (17.4 percent by OECD figures). These figures, as well as concerns about access to health care, are the sorts of indicators that have led many to question what Americans are getting for their spending. However, high expenditures may have three meanings:

1. High average level of services
2. High resource costs for services
3. Inefficient provision of services

In examining cross-country differences, we note that high levels of services reflect at least the possibility that populations have chosen to spend their incomes in this fashion. We have noted previously that higher income levels lead to higher consumption levels of all normal goods, including health care. Cross-national studies indicate a substantial responsiveness of health care expenditures to increased income (relatively large income elasticity). U.S. expenditure levels reflect in part the higher per-capita income level in the United States.

As shown by the comparative data in Table 22-2, the resources available across countries can vary widely. All of the countries in the table have more inpatient beds per 1,000 population than the United States (3.1 beds per 1,000), and France (6.9), Germany (8.2), Hungary (7.1), Japan (13.8), and South Korea (7.8) have more than twice as many beds. Germany, Hungary, and the United Kingdom have more practicing physicians per 1,000 than the United States, while South Korea and Japan have fewer. Although current data for the United States are not available, there is a wide range of practicing nurses across the countries that provide data.

For a better perspective of the relative success of various health systems in controlling inflation, examine Figure 22-1. The upward trends in expenditures continued into the early 1990s for the United States, and eased some through the 1990s. U.S. expenditures accelerated in the first years of the twenty-first century and jumped in the “Great Recession” of 2008–2009, as they did in Canada, France, Germany, and the United Kingdom. The recent jump is due in part to a fall in the denominator (GDP per capita) for these countries. Nonetheless, comparing the United States with these other countries shows an increasing spread in expenditure shares.

## THE UNITED KINGDOM—THE NATIONAL HEALTH SERVICE

This section examines the national health system of the United Kingdom in detail, and the following section looks at China. After that, we will look at Canada and contrast Canada's plan, a national health insurance system, with that of the United States, an example of a mixed system.

**TABLE 22-1 Health Care Spending and Outcomes in Selected Countries**

	GDP/Capita US\$ PPP (2010 <sup>a</sup> )	Tot. Expend. Health/Capita US\$ PPP (2009 <sup>a</sup> )	Inpatient Care as % of NHE (2008 <sup>b</sup> )	Pharm and Others as % of NHE (2008 <sup>b</sup> )	% of GDP Spent on Health Care			Crude Death Rates/1,000 Population (2007 <sup>b</sup> )	Life Expectancy at Birth	
					(1970 <sup>b</sup> )	(1990 <sup>b</sup> )	(2009 <sup>a</sup> )		Female (2009 <sup>a</sup> )	Male (2009 <sup>a</sup> )
Australia	40,349	—	—	—	—	6.9	—	6.6	83.9	79.3
Austria	39,881	4,289	40.0	13.3	5.2	8.4	11.0	9.0	83.2	77.6
Belgium	38,412	3,946	41.4	16.4	3.9	7.2	10.9	9.5	82.8	77.3
Canada	39,574	4,363	27.3	17.2	6.9	8.9	11.4	7.1	—	—
Chile	14,846	1,186	—	—	—	—	8.4	—	80.9	75.6
Czech Republic	26,063	2,108	32.1	20.4	—	4.7	8.2	10.1	80.5	74.2
Denmark	38,864	4,348	—	—	—	8.3	11.5	10.2	81.1	76.9
Estonia	21,163	1,393	33.0	20.7	—	—	—	13.0	80.1	69.8
Finland	36,670	3,226	35.0	14.4	5.5	7.7	9.2	9.3	83.5	76.6
France	34,500	3,978	37.3	16.4	5.4	8.4	11.8	8.4	84.4	77.7
Germany	37,526	4,218	34.1	15.1	6.0	8.3	11.6	10.0	82.8	77.8
Greece	28,067	—	—	—	5.4	6.6	—	9.8	82.7	77.8
Hungary	21,036	1,511	29.6	31.6	—	—	7.4	13.2	77.9	70.0
Iceland	36,104	3,538	45.4	13.9	4.7	7.8	9.7	6.2	83.3	79.7
Ireland	39,169	3,781	—	17.3	5.1	6.1	9.5	—	82.5	77.4
Israel	28,615	2,164	—	—	—	—	7.9	—	83.5	79.7
Italy	33,807	3,137	45.6	18.4	—	7.7	9.5	—	—	—
Japan	34,049	—	—	—	4.6	6.0	—	8.9	86.4	79.6
Korea, Republic of	28,236	1,879	29.4	23.9	—	4.0	6.9	5.1	83.8	76.8
Luxembourg	91,377	4,808	—	—	3.1	5.4	7.8	8.1	83.3	78.1
Mexico	15,196	918	15.2	28.3	—	4.8	6.4	—	77.6	72.9
Netherlands	42,151	4,914	—	—	—	8.0	12.0	8.1	82.7	78.5
New Zealand	29,458	2,983	31.8	9.4	5.2	6.9	10.3	6.9	82.7	78.8
Norway	56,137	5,352	42.4	7.6	4.4	7.6	9.6	8.9	83.2	78.7
Poland	19,695	1,394	32.0	22.6	—	4.8	7.4	9.9	80.0	71.5
Portugal	25,547	—	—	—	2.5	5.9	—	—	82.6	76.5
Slovak Republic	24,066	2,084	21.1	28.1	—	—	9.1	10.0	78.7	71.3
Slovenia	28,446	2,579	33.7	18.7	—	—	9.3	9.2	82.3	75.8
Spain	32,625	3,067	28.2	20.5	3.5	6.5	9.5	8.6	84.9	78.6
Sweden	39,316	3,722	29.2	13.2	6.8	8.2	10.0	10.1	83.4	79.4
Switzerland	46,019	5,144	—	—	5.4	8.2	11.4	8.1	84.6	79.9
Turkey	15,258	—	—	—	—	3.6	—	—	76.1	71.5
United Kingdom	36,083	3,487	—	11.8	4.5	6.0	9.8	9.6	82.5	78.3
United States	47,184	7,960	24.5	11.9	7.0	11.9	17.4	8.0	80.6	75.7

Source: <sup>a</sup>OECD Health Data 2011, June; and <sup>b</sup>OECD Health Data 2010, June 2011.

**TABLE 22-2 Health Care Resources**

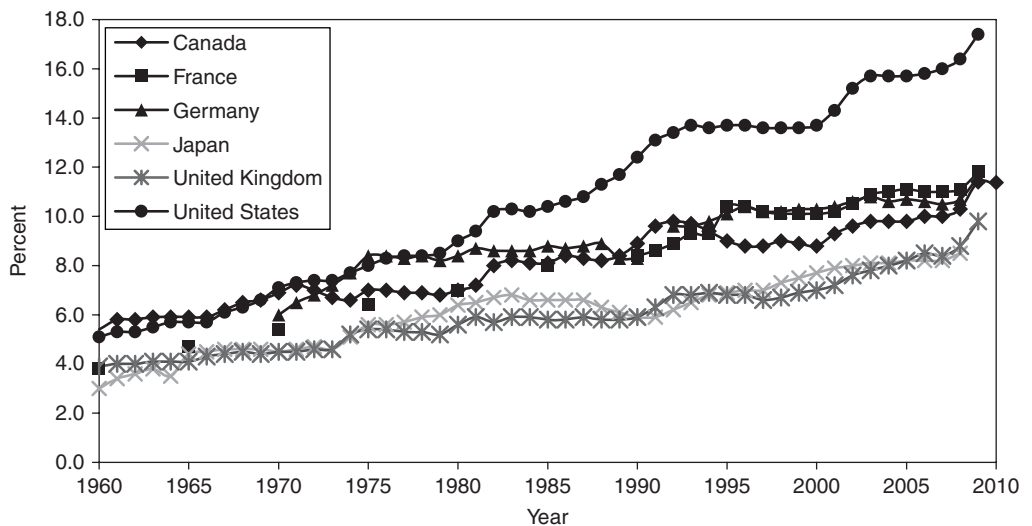
	Inpatient Beds/1,000		Practicing Physicians/1,000		Practicing Nurses/1,000
	2008	% of U.S.	2009	% of U.S.	2009
Canada	3.3	106.5	—	—	9.4
France	6.9	222.6	—	—	—
Germany	8.2	264.5	3.6	150.0	11.0
Hungary	7.1	229.0	3.1	125.0	6.2
Japan	13.8	445.2	2.2	91.7	—
S. Korea	7.8	251.6	1.9	79.2	4.5
UK	3.4	109.7	2.6	112.5	9.7
USA	3.1	100.0	2.4	100.0	—

Source: OECD Health Data 2011, June 2011.

## The National Health Service<sup>2</sup>

Great Britain established its National Health Service (NHS) in 1946, and it provides health care to all British residents. About three-quarters is funded by general taxation, with about 20 percent from national insurance and about 3 percent each from user charges and other sources of income. Capital and current budget filter from the national level down to the regional and then to the district level. The plan pays general practitioners on a capitation basis and hospital physicians largely on a salaried basis. In addition to the NHS, there is also a private-sector health system. About 11 percent of Britons purchase private health insurance.

Services are not entirely free. English patients pay £7.40 (about \$11.70 at the November 2011 exchange rate of about \$1.58 per £1) for each prescription, but close to 90 percent of prescriptions are exempt from charges, and patients in Scotland, Wales and Northern Ireland are not charged. As



**FIGURE 22-1** Percent of GDP Spent on Health Care, 1960-2010

Source: OECD, 2011.

<sup>2</sup> Updates and summaries are from Boyle (2011).

of 2011, patients pay no more than £204 (about \$322) for each “course of [dental] treatment.” This maximum, called Band 3, includes crowns, dentures and bridges—others treatments are far less. Those receiving means-tested benefits and their adult dependents, children under age 16 (under age 19 if a student), pregnant women, and nursing mothers are exempt from dental and prescription charges.

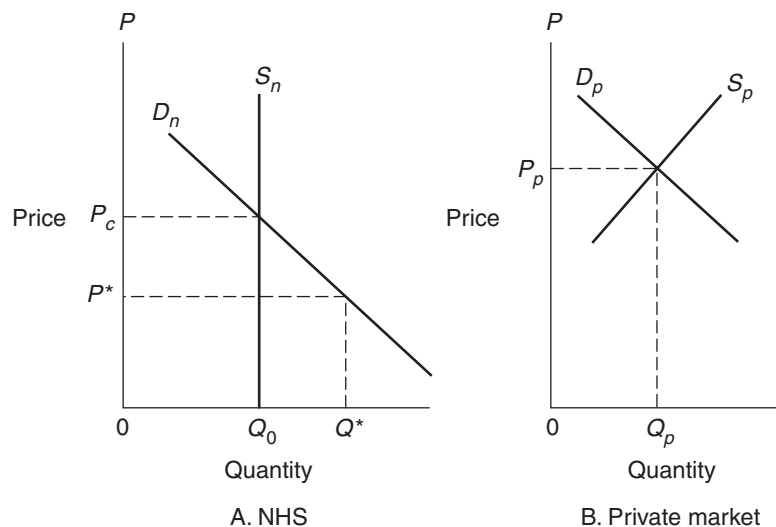
The general practitioner (GP) serves as the gatekeeper to the health care system. GPs are not government employees, but are self-employed and receive about half their incomes from capitation contracts. GPs typically treat routine conditions and refer patients to hospitals for more specialized care. The referral usually will be to a district hospital. Once at the hospital, the patients are under the care of physicians (consultants) who are allocated staffed beds and junior hospital staff to work under their direction.

Table 22-1 shows the U.K. spending per capita (\$3,487) in 2009 as 43.8 percent of the U.S. level (\$7,960), and a little more than half, when expressed as a ratio of GDP (9.8 percent as opposed to 17.4 percent). How does the United Kingdom keep its health care expenditures this much lower while providing universal access to health care? Though patients have relatively easy access to primary and emergency care, specialty care is rationed through waiting lists and limits on the availability of new technologies. A relatively simple model illustrates this phenomenon.

**A MODEL OF RATIONED HEALTH CARE AND PRIVATE MARKETS** We can examine the practices of an NHS-type of organization diagrammatically.

In panel A of Figure 22-2, we treat the supply of health services as totally price inelastic. Why? The supply curve, reflecting what the government provides irrespective of price, is a vertical line. This indicates that the quantity supplied is not responsive to the price of the services. Furthermore, the money price of the services is set by the government at  $P^*$ , which is less than  $P_c$ , the market clearing price. Predictably, we see excess demand ( $Q^* - Q_0$ ) at the administered price  $P^*$ . Because most health care cannot be bought and resold, other forms of rationing, largely time-related, become important. For many ailments, the waiting period for treatment by the NHS is months, or even years.

Again, as might be predicted, a private market for services has developed for those who choose to enter the private market without governmental aid, either due to strong preferences for private care or due to the ability to pay more than the NHS price. Returning to Figure 22-2B, with



**FIGURE 22-2** Prices and Quantities in a Controlled Market

**BOX 22-1****NHS Bars Woman after She Saw Private Doctor**

We have examined the distinction between private and NHS practices in Great Britain. Isabel Oakeshott, writing in *The Times (London)*, discusses the problems that confront some patients. Ms. Oakeshott wrote in April 2010 how Jenny Whitehead, a breast cancer survivor, paid £250 for an appointment with an orthopedic surgeon after being told she would have to wait five months to see him on the NHS.

Whitehead, 64, had gone to her GP in December 2009 for back pain. Because of her breast cancer history, she was immediately offered an MRI scan to check that the disease had not returned. It revealed a cyst on her spine, pressing against her sciatic nerve. Her GP referred her to a consultant (the surgeon) at Airedale NHS hospital.

She was told the next available NHS appointment was in May 2010, five months later, but because the back pain was so severe, she accepted the offer of a *private slot* to see him the following week. After seeing her, the surgeon promised to add her to his NHS waiting list for surgery. After two months, however, hospital managers told her she could not be on the waiting list because she had seen the surgeon privately. Now her only alternative to paying £10,000 privately was to go back to her GP, seek another referral to the same specialist, this time on the NHS, and face another 18-week wait.

The Bradford and Airedale NHS trust said it was looking into the case “as a matter of urgency” but added, “Anyone who chooses to pay for a private outpatient consultation cannot receive NHS treatment unless they are then referred on to an NHS pathway by their consultant.”

*Source:* Oakeshott, Isabel, “NHS bars woman after she saw private doctor,” London: *The Times*, April 18, 2010, [http://www.timesonline.co.uk/tol/life\\_and\\_style/health/article7100968.ece](http://www.timesonline.co.uk/tol/life_and_style/health/article7100968.ece), accessed June 27, 2011.

excess demand at  $P^*$ , that excess demand represents, in part, people who are queued and who may wish to pay in the private sector to avoid the long waits. Those who participate in the private market, shown in panel B, will pay  $P_p$  for the quantity of services,  $Q_p$ . The two markets exist simultaneously, although as Box 22-1 indicates, not always comfortably.

**PERFORMANCE UNDER THE NHS AND MORE RECENT REFORMS** On the one hand, a system such as the NHS that depends on queuing in line for access to care often leads participants to postpone or simply not purchase certain services. On the other hand, the NHS devotes considerable resources to such high-return services as prenatal and infant care. To these populations served, and to the larger public concerned with equitable provision of care to these segments of the population, the universal nature of the service is particularly beneficial.

In addition, although the United Kingdom has spent considerably less on health care than the United States and many other countries, by most measures of mortality and morbidity the United Kingdom does about as well. Many nonmedical factors are involved in determining disease and death rates in a population and these factors will vary across countries.

Despite universal access to care in the United Kingdom, historically there have been considerable regional disparities in funding and in the use of health care. Evidence has shown that upper-class patients have received substantially more care for a given illness than have lower-class patients (Maynard, 1990). Thus, even where access was universal the results were not necessarily equal.

The NHS has always been popular with the public. Nevertheless, 1991 reforms under Prime Minister Margaret Thatcher replaced the centralized, hierarchical NHS bureaucracy with a quasi-market mechanism. It created a purchaser/provider split, where “purchasers” of health care, regional health authorities, and general practice fund holders were allocated budgets to purchase services for their populations. These purchasers were distinct from providers—mainly hospitals. The incentives for efficiency and a responsive system came as providers competed for contracts with purchasers.

Le Grand (2002) notes that following Tony Blair’s Labour party victory in 1997, the government remained committed to the purchaser/provider split but emphasized “cooperative” rather than

competitive arrangements. As a last resort, purchasers could still shift their purchasing away from providers. According to Le Grand:

The quasi-market was largely abolished and market-based competitive pressures mostly eliminated. Despite an ostensible commitment to collaboration and the introduction of some measures designed to foster it, central control was maintained and indeed, as illustrated by the proliferation of central agencies and control instruments, greatly increased. (p. 121)

Since 2000, there have been two major problems. The first has related to capacity constraints—shortages of doctors and nurses, as well as relatively small levels of acute hospital beds. Also, a shortage of nursing home beds has meant difficulties discharging elderly patients from the hospital, preventing hospitals from taking on new admissions. Second, incentive problems pervaded the system. NHS providers were paid salaries to work 11 sessions per week in the NHS. If NHS providers were willing to work (and be paid) for only 10 of the 11 sessions, they were allowed to work as much as they liked in the private sector—where they were paid on a fee-for-service basis. The longest NHS waiting lists occurred in the specialties in which specialists had the highest private earnings.

The NHS has made major efforts to reduce patient waiting times. In 2005, the Healthcare Commission reported that overall satisfaction of patients with most NHS services was high. The number of people waiting more than six months for admission as inpatients in England decreased by 85 percent from March 2000 to March 2005. There was also a significant drop in the number of people waiting more than 13 weeks for an appointment as outpatients—down by 92 percent over the same period. For specialties with high inpatient death rates, the number waiting less than six months increased by 8 percent between 1999 and 2005.<sup>3</sup>

However, by March 2007, one in eight NHS hospital patients still had to wait more than a year for treatment. A Department of Health analysis of 208,000 people admitted to the hospital in March showed 48 percent were wheeled into the operating theatre within 18 weeks of a GP sending them for hospital diagnosis. However, 30 percent waited more than 30 weeks and 12.4 percent more than a year. Many people also experienced problems gaining access to NHS dentists, with nearly two-thirds of all dental practices not taking on new NHS patients.

As of 2011, the NHS actively seeks to limit waiting times to 18 weeks. A 2011 visit to the NHS website shows:

“The NHS is making sure that you are seen as soon as possible, at a time that is convenient for you. To do this, the NHS Constitution gives you the right to access services within maximum waiting times, or for the NHS to take all reasonable steps to offer you a range of suitable alternative providers if this is not possible”

and:

“What this means for patients

- You have the right to start your consultant-led treatment *within a maximum of 18 weeks* [emphasis added] from referral.
- You have the right to be seen by a specialist *within a maximum of two weeks* [emphasis added] from GP referral for urgent referrals where cancer is suspected.”

<sup>3</sup> See Commission for Healthcare Audit and Inspection, *State of Healthcare 2005*, [www.healthcarecommission.org.uk/healthcareproviders/nationalfindings/publications/publications2004-05](http://www.healthcarecommission.org.uk/healthcareproviders/nationalfindings/publications/publications2004-05), accessed August 2, 2008.

Performance appears to have improved since 2007, but the target remains elusive. NHS data reported by *The Guardian* in July 2011 indicated that in April 2011, over one in ten (10.2%) of NHS patients had waited more than 18 weeks for treatment. This represented an increase of 24% over the 8.2% facing similar waits in the same month in 2010 (that is, 10.2 divided by 8.2). Despite 29,000 fewer procedures carried out in April 2011 compared with a year earlier, an additional 2,387 patients (of 241,000) had waited more than 18 weeks.<sup>4</sup>

The experience of the NHS in the area of cost containment seems clear. Rationed care cuts money costs. Even with increased expenditures from the health care reforms, analysts expected total U.K. expenditures to remain well below the European Union and the United States.

## CHINA—AN EMERGING SYSTEM

The Chinese health economy has undergone substantial changes since the formation of the People's Republic in 1949. Governmental policies moved from a doctrinaire political system with administered prices in the first three decades, to more market-oriented processes since the 1980s, affecting coverage and focus. We begin by examining the organization of health care, and the role of the private sector. We then consider some government policy initiatives and measures of system performance. We finish with observations about the future of the Chinese health economy.<sup>5</sup>

Although China is a large world economy, its per capita income is small compared to many of the Western countries. Table 22-3 compares the Chinese health economy to developing nations India and Indonesia, as well as to Japan, with its more advanced economy. According to the World Health Organization (WHO), China spends considerably more on health per capita than do India and Indonesia, but less than Japan. Measures of life expectancy at birth and probability of dying are more favorable than India and Indonesia, but less favorable than Japan.

Eggleston and colleagues describe the development of separate three-tiered urban and rural systems starting in the early 1950s. In urban areas, the three-tier network was composed of street clinics, district hospitals, and city hospitals. In rural areas it consisted of village clinics, township health centers (THCs) and county hospitals. Provincial and central hospitals provided high-level referral care. Under this system, the Ministry of Health or the local Bureau of Health managed the majority of the provider organizations.

**TABLE 22-3 Comparative Health Services Data: Four Asian Countries, 2009**

Statistics	China	India	Indonesia	Japan
Total population (in thousands)	1,353,311	1,198,003	229,965	127,156
Gross national income per capita (PPP international \$)	6,010	2,930	3,600	35,190
Life expectancy at birth male/female (years)	72/76	63/66	66/71	80/86
Number dying under age five (per 1,000 live births)	19	66	39	3
Probability of dying between 15 and 60 years m/f (per 1,000 population)	142/87	250/169	234/143	86/42
Total expenditure on health per capita (\$ 2009)	309	132	99	2,713
Total expenditure on health as % of GDP (2009)	4.6	4.2	2.4	8.3

Source: World Health Organization, <http://www.who.int/countries/en/>, accessed June 26, 2011

<sup>4</sup> Sources: [http://www.nhs.uk/choiceintheNHS/Rightsandpledges/Waitingtimes/Pages/Guide to waiting 20times.aspx](http://www.nhs.uk/choiceintheNHS/Rightsandpledges/Waitingtimes/Pages/Guide%20to%20waiting%20times.aspx), accessed July 19, 2011, and James Ball and Denis Campbell, "NHS Waiting Times – Stable or Critical?", <http://www.guardian.co.uk/society/2011/jul/10/nhs-waiting-times-andrew-lansley>, accessed July 19, 2011.

<sup>5</sup> The organization of this section follows Eggleston et al (2008a).



The goal under Maoist Communist rule through the 1970s was to assure access to care. Providers received direct budgetary support to cover the difference between costs and revenues earned from the nominal fees that were paid. The government financed preventive and other public health services and provided anti-epidemic stations at province, prefecture, and county/district levels, as well as at THCs and village clinics.

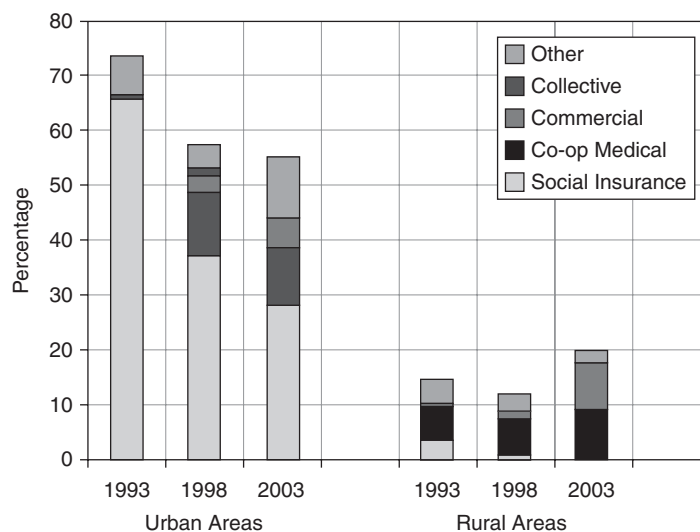
The authors note that the three tiers that characterized the original system remain today. Since the early 1980s, however, the government has allowed providers to generate, retain, and manage surpluses, with subsidies to providers constituting smaller and decreasing shares of provider financing.

When the government routinely subsidized the providers, the variation of administered prices from costs had little impact because deficits could be made up from the subsidies. Without government subsidies, however, providers have tended to favor high-technology diagnostics at the expense of less-profitable basic services.

The combination of rapid private sector growth, and decreased organized financing, have made health care less affordable for many. According to National Health Surveys, between 1998 and 2003, the proportion ill in the previous two weeks who did not seek care for financial reasons increased in both urban and rural areas (Ministry of Health, 2004). Ma, Lu, and Quan (2008) note that health insurance coverage fell by about 20 percentage points in urban areas between 1993 and 2003, while rising slightly in rural areas. (See Figure 22-3.) Some 700 million rural Chinese must pay out of pocket for virtually all health services, leading to the deferral of care and untreated illness.

The 1990s saw the initiation of several new policies in both urban and rural areas. In urban areas, municipal risk pooling for employees, known as Basic Medical Insurance (or BMI) was established. The government also established a series of medical savings accounts, but they did not stipulate the means of provider payment. As a result, most people purchase treatment under a fee-for-service (FFS) model. In rural areas, the government established a new cooperative medical scheme (NCMS), which combines household contributions with central and local government subsidies. It was piloted in 2003, and Chen and colleagues (2011) report that 95 percent of the counties were implementing the scheme by 2008.

How has the Chinese health delivery system performed? Eggleston and colleagues report that between 1990 and 2002, spending has risen much faster than per capita income and prices, with nominal per capita health spending increasing by a factor of eight in urban areas and by a factor of almost seven in rural areas. Health spending as a percentage of GDP increased from 3.2 percent in 1980 to 5.6 percent in 2003, and current (2011) estimates put it at 4.6 percent. As noted by Ma and



**FIGURE 22-3** Health Insurance Coverage in Urban and Rural Areas in China, by Program, 1993, 1998, 2003

colleagues, rapidly rising health-care costs and limited insurance coverage have made health care increasingly unaffordable for China’s poor families.

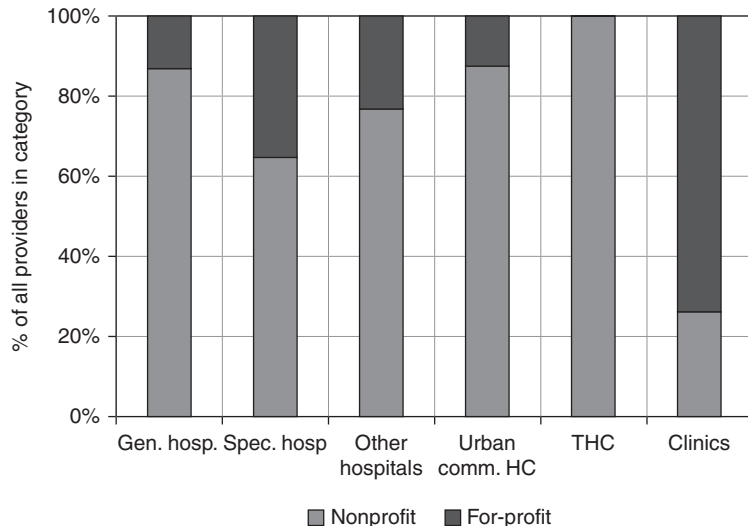
Providers receive payments from three main sources. Over 60 percent comes from out-of-pocket payments, based on a regulated fee schedule. Social insurance agencies such as NCMS and BMI, mentioned above, rather than private insurers such as China Life, account for another 20 percent or so. The remaining funds come largely from government subsidies from provincial and county governments, although these account for a decreasing share of providers’ revenue.

Ownership of facilities remains largely nonprofit. As noted in Figure 22-4 private for-profit ownership is common among clinics (74 percent private), and outpatient departments (48 percent private). By 2005, almost 16 percent of the hospitals, particularly specialized hospitals, were registered as for-profit. Most urban community health centers (HC) and all township health centers (THC) are nonprofit.<sup>6</sup>

While a well-functioning referral system could lower cost and enhance equity (Gerdtham and Jonsson, 2000), many believe that China lost this opportunity in its transition to a more market-based system. Patients now choose the level of provider they can afford and this means that more affluent Chinese overutilize higher-level hospitals such as provincial and county hospitals, while lower level hospitals such as township hospitals are underutilized and serve mostly low-income patients. Some regions have discussed reinstating referral systems or tiered co-payment requirements, and some Basic Medical Insurance patients must designate a specific hospital or clinic as their “appointed” provider (*dingdian yiyuan*).

What can one conclude about the Chinese system? Eggleston and colleagues suggest that the current system “leaves much room for improvement” in terms of responsiveness to patients, efficiency, and equity. They urge strengthened provider-payment reforms, and improved provider management. They advocate “sector-neutrality” with supervision and regulation applying to both public and private providers, and purchasing from higher-performing providers on equal terms, regardless of ownership.

Observers also note major issues of Chinese population health. The Chinese smoke heavily—the estimated 350 million smokers constitute one of three in the entire world.<sup>7</sup> Thirty-six percent of



**FIGURE 22-4** Who delivers health services in China?  
 Source: From Eggleston et al. (2008).

<sup>6</sup> The “nonprofit” category mainly consists of organizations owned by government and enterprises (available data do not permit researchers to disaggregate the nonprofit category by ownership).

<sup>7</sup> These figures are from The Quit Smoking Guide, <http://www.quitguide.com/smoking-facts.html>, accessed June 26, 2011.

the population smoke, including 70 percent of all Chinese men. More than one million people a year die in China from tobacco-related diseases, including lung cancer and heart disease. Advocacy groups argue that most Chinese have no knowledge of the facts about smoking or any awareness of the consequences they face. Regulation of advertising can play an important role in protecting population health and reducing information asymmetries— particularly in relation to the use of tobacco.

Chinese economist Jian Wang (2011) highlights five priorities for Chinese health policy reform. These include:

- expanded coverage and improved basic health insurance benefits for both the urban and rural populations;
- full coverage of essential medicines;
- reformed and improved capabilities for the primary health care institutions;
- more efficient provision of and access to public health programs; and
- improved capacity and quality of traditional Chinese medicine care and further containment of health care costs.

These underscore an ambitious set of tasks for the Chinese economy.

## THE CANADIAN HEALTH CARE SYSTEM

The rapid increases in U.S. health care costs and growing concern over the large number of uninsured have led many in the United States to look at Canada's health system as a model for reform. Many Americans perceive that Canada has developed a comprehensive and universal national health insurance program that is cost-effective and highly popular.

### Background

Canada and the United States share a long border and similar heritage in terms of language, culture, and economic institutions.<sup>8</sup> The health care systems evolved similarly until the 1960s. As recently as 1971, both countries spent approximately 7.5 percent of their GDPs on health care.

Since 1971, however, the health care systems have moved in different directions. While Canada has had publicly funded national health insurance, the United States has relied largely on private financing and delivery (although governments have been heavily involved through Medicare, Medicaid, and numerous regulatory programs). During this period, spending in the United States has grown much more rapidly despite large groups that are either uninsured or minimally insured.

The Canadian system of financing and delivering health care is known as Medicare, not to be confused with the U.S. Medicare program for the elderly. In Canada, each of the ten provinces and three territories administers a comprehensive and universal program partially supported by grants from the federal government.<sup>9</sup>

Various federal government criteria regarding coverage must be met. Coverage must be universal, comprehensive, and portable, meaning that individuals can transfer their coverage to other provinces as they migrate across the country. There are no financial barriers to access, and patients have free choice in the selection of providers.

Canada's Medicare should also be distinguished from Britain's NHS. Most Canadian physicians are in private practice and have hospital-admitting privileges. Although traditional remuneration was by fee-for-service, several provinces now remunerate a majority of their physicians by systems other

<sup>8</sup> Updates and summaries are from Watson and Allin (2011).

<sup>9</sup> There is direct federal government funding (services do not fall under provincial programs) for health care services for First Nations (i.e., aboriginal) people on reserves, and some services to the Royal Canadian Mounted Police, Correctional Services, the Armed Forces, and Veterans.

than fee-for-service, and this trend has accelerated. According to a 2007 physician survey, about half of family physicians received at least 90 percent of their incomes through fee-for-service, and about 30 percent received at least 90 percent through blended payment.

Canadian hospitals are private not-for-profit institutions, although their budgets are approved and largely funded by the provinces. However, Watson and Allin report that some provinces have introduced activity-based funding to pay for additional services that were targeted in strategies to reduce waiting times. Ontario, for example, adopted activity-based funding for cataract surgery, joint replacement surgery, and cardiac bypass surgery, and has been successful in reducing waiting times.

There was considerable support for national health insurance in both the United States and Canada up to the late 1940s, with much of it coming from organized labor. The Canadian system originated in 1947, when the Saskatchewan government introduced the first provincial hospital insurance program. Organized medicine in both countries strongly opposed national health insurance proposals, but the American Medical Association in the United States was more successful than its Canadian counterpart in linking national health insurance to “socialized medicine,” and preventing its adoption. The main impetus in Canada came from federal legislation adopted in 1957 that provided cost sharing for inpatient hospital services, and the Medical Care Act of 1966, which provided cost sharing for physician care. Since 1972, every province and territory has provided universal coverage for hospital and physician care.

The Canada Health Act of 1984 defines the criteria and conditions for the provinces and territories to satisfy in order to qualify for their full share of the federal transfers under the Canada Health Transfer (CHT) cash contribution.

1. **Public administration:** The administration of the health care insurance plan must be carried out on a nonprofit basis by a public authority;
2. **Comprehensiveness:** All medically necessary services provided by hospitals and doctors must be insured;
3. **Universality:** All insured persons must be entitled to public health insurance coverage on uniform terms and conditions;
4. **Portability:** Coverage for insured services must be maintained when an insured person moves or travels within Canada or travels outside the country; and
5. **Accessibility:** Reasonable access by insured persons to medically necessary hospital and physician services must be unimpeded by financial or other barriers.

Two key provisions of the act guide Canada’s Medicare:

- No extra billing by medical practitioners or dentists for insured health services under the terms of the health care insurance plan;
- No user charges for insured health services by hospitals or other providers under the terms of the health care insurance plan.

The provinces and territories also provide coverage to certain groups of people (e.g., seniors, children, and social assistance recipients) for health services that the publicly funded health care system does not generally cover. These supplementary health benefits often include prescription drugs, vision care, medical equipment and appliances (prostheses, wheelchairs, etc.), independent living, and the services of podiatrists and chiropractors. Dental services are much like services in the United States—either uninsured, privately insured, or group insured through place of employment. The level of health coverage varies across the country. Many Canadians have supplemental private insurance coverage, through group plans, which covers the cost of these supplementary services.

Table 22-4 provides comparative data on the two countries. While geographically larger than the United States, Canada has about 11 percent of the U.S. population. Canada’s GDP per capita is about 83.5 percent of the U.S. level. With a national health system providing universal coverage, public funds account for over 70 percent of total health spending. Canada has maintained substantially

**TABLE 22-4 Comparative Data: Canada and the United States**

	Canada	United States
Population—2010 in millions <sup>a</sup>	33.6	309.1
Population over 65 (2010, %) <sup>a</sup>	14.1	13.1
GDP—2010 (trillions of 2010 \$US) <sup>b</sup>	1.33	14.66
GDP per capita—2010 (2010 \$US) <sup>b</sup>	39,400	47,200
Government expenditures as % of GDP (2008) <sup>c</sup>	39.8	38.8
Health spending per capita—2009 (\$US PPP) <sup>a</sup>	4,363	7,960
Health spending—2009 (% of GDP) <sup>a</sup>	11.4	17.4
Percent of total health spending (2008 <sup>c</sup> ):		
Public Expenditures	70.2	46.5
Inpatient care	27.3	24.5
Outpatient care	25.4	44.2
Pharmaceuticals	17.2	11.9
Acute care inpatient beds/1,000 population (2007 <sup>a</sup> )	1.8	2.7
Average length of stay (acute care days) (2007 <sup>c</sup> )	7.5	5.5
Uninsured population in percent (2009 <sup>a</sup> )	0.0	18.7
Out-of-pocket payments per capita (\$US)—2009 <sup>a</sup>	636	976
Private insurance % expenditure on health—2009 <sup>a</sup>	13.4	34.4
Tobacco (% population older than 15)—2009 <sup>a</sup>	16.2	16.1
Alcohol consumption (liters/capita 15+)—2008 <sup>a</sup>	8.2	8.8
Life expectancy (in years) at birth—females (2007 <sup>a</sup> )	83.0	80.4
Life expectancy (in years) at birth—males (2007 <sup>a</sup> )	78.3	75.4

Source: <sup>a</sup>OECD *Health Data 2011*, June 2011; <sup>b</sup>CIA *World Fact Book* (last updated: Jan 26, 2011); <sup>c</sup>OECD *Health Data 2010*, June 2010.

lower health spending and share of GDP per capita than the United States, despite its universal health insurance system and its longer lengths of stay.

Although data are not available for the most recent years, Canada's physician-population and nurse-population ratios have been 10 to 20 percent lower than the U.S. Even though about 15 percent of the U.S. population goes without insurance coverage at any time, Americans spend 82 percent more per capita on health care (\$7,960 versus \$4,363). Canadians drink a little less alcohol and smoke slightly more than do Americans. Despite lower spending, Table 22-4 shows that commonly compared health status indicators—such as life expectancy (about 4.7 years longer for women and 5.0 years longer for men)—are more favorable in Canada than in the United States. Finally, public opinion polls indicate that Canadians support their system more than Americans support theirs and are concerned about any threats to it. Given the Canadian record on cost savings, health care scholars, policymakers, and politicians have shown great interest in determining the sources for its apparent success.

### Physician Fees and Quantity

An early study by Fuchs and Hahn (1990) sought to break down Canadian and U.S. expenditures by specific services, while separating their price and quantity components. The authors estimated that spending on health care per capita was 38 percent higher in the United States in 1985. More striking was the disparity in spending on physician services; it was 72 percent higher in the United States, and 178 percent higher for the procedures component.

With aggregate spending equal to the product of prices and quantities, the authors sought to identify differences in fees (prices) and utilization per capita (quantities). Overall, fees were 239 percent higher in the United States for 1985. Though there were variations in the ratios across service categories, U.S. fees were considerably higher in each category. The net incomes of U.S. doctors were also substantially higher than were their Canadian counterparts.

A large reason for the reduced fees in Canada is that provincial governments constitute monopsonies (single buyers) of physician labor. Monopsony means that the provincial governments face upward-sloping supply curves for physicians, so that the marginal labor cost of raising the fees for one physician requires raising the fees for all others. This results in lower fees than with competitive buyers, and in hiring fewer workers than in a competitive market (readers can look ahead to Figure 22-5C for a monopsony analysis). Negotiations with the local medical societies reflect this monopsony power as compared to the United States with its myriad buyers.

The differences in service volume found by Fuchs and Hahn were perhaps more surprising than the fee differentials. Despite the much higher spending per capita for physician care, the quantity of care per capita was considerably lower in the United States. Thus, the savings in Canada, at least for physician care, did not come from reduced volume of care.

The Fuchs and Hahn findings provided provocative insights and led readers to numerous questions about the two systems. After discussing the Canadian system in more detail, we return with a study by June and Dave O'Neill (2008) that revisits some of the questions.

### **Why Are Fees and Hospital Costs Lower in Canada?**

Hospital patients in Canada have longer lengths of stay, in part because of the greater use of Canadian hospitals for chronic long-term care. Nonetheless, after adjusting for differences in case mix between the two countries, Newhouse, Anderson, and Roos (1988) found that the cost per case-mix adjusted unit was roughly 50 percent higher in the United States. Several reasons may explain this phenomenon.

In Canada, unlike the United States, physician fees result from negotiation between physicians' organizations and the provincial governments, as well as from other limits on total spending. Physicians cannot evade the fee controls by charging extra (sometimes called *balance billing*) to patients who can afford it.

The provinces also regulate hospital costs are similarly through approval of hospital budgets. Hospitals and provinces negotiate operating budgets financed by the provincial governments. The capital budget may include other sources of funding, but provinces still must approve capital expenditures. Thus, a centralized mechanism allocates resources to the hospital sector and determines the distribution of resources among hospitals. Occupancy rates are higher in Canadian hospitals. Also, the provinces have limited the capital costs associated with expensive new technologies.

Table 22-5 compares several relatively recent and expensive technologies among Canada, the United States, and the United Kingdom, with other OECD countries available for context. Analyses from the 1980s indicated that the United States had greater availability of many of the technologies, but at the end of the first decade of the twenty-first century, other countries are catching up. With respect to CT scanners, several countries compare favorably to the U.S. measure of 34.3 per million, including Greece (34.4), Australia (42.5), and Japan (97.3). Japan also has a much higher availability of MRI units (43.1 per million v. 25.9 in the United States).

The U.S. is more "high-tech" than Canada. The U.S. figure of 34.3 CT scanners per million people is well over twice as large as Canada's 14.4. The U.S. figure of 25.9 MRI units per million is over three times as large as Canada's 8.4.

### **Administrative Costs**

The centralized system of health care control in Canada has led to theories about the possible economies associated with administrative and other overhead expenses. Almost all U.S. patients have experienced problems due to extensive paperwork and complex billing practices. For providers

**TABLE 22-5 Availability of Selected Technologies**

	CT Scanners		MRI Units		Radiation Therapy		Lithotriptors		Mammographs	
	Number	Per Million	Number	Per Million	Number	Per Million	Number	Per Million	Number	Per Million
Australia	949	42.5	130	5.8	205	9.2	21 <sup>b</sup>	1.0 <sup>b</sup>	533	23.9
Austria	245 <sup>a</sup>	29.3 <sup>a</sup>	154 <sup>a</sup>	18.4 <sup>a</sup>	42 <sup>a</sup>	5.0 <sup>a</sup>	16 <sup>c</sup>	1.9 <sup>c</sup>	—	—
Canada	484	14.4	281	8.4	—	—	14	0.4	—	—
Czech Republic	148 <sup>a</sup>	14.1 <sup>a</sup>	60 <sup>a</sup>	5.7 <sup>a</sup>	89 <sup>a</sup>	8.5 <sup>a</sup>	31 <sup>a</sup>	3.0 <sup>a</sup>	133 <sup>a</sup>	12.7 <sup>a</sup>
Denmark	153	27.6	85 <sup>a</sup>	15.4 <sup>a</sup>	70	12.6	—	—	107	19.3
Estonia	20 <sup>a</sup>	14.9 <sup>a</sup>	10 <sup>a</sup>	7.5 <sup>a</sup>	3 <sup>a</sup>	2.2 <sup>a</sup>	2 <sup>a</sup>	1.5 <sup>a</sup>	—	—
Finland	113	21.1	104	19.4	48	8.9	3	0.6	163	30.4
France	766	11.8	451	7.0	—	—	—	—	—	—
Greece	388	34.4	255	22.6	66	5.8	20	1.8	632	56.0
Hungary	72 <sup>a</sup>	7.2 <sup>a</sup>	28 <sup>a</sup>	2.8 <sup>a</sup>	41 <sup>a</sup>	4.1 <sup>a</sup>	49 <sup>a</sup>	4.9 <sup>a</sup>	146 <sup>a</sup>	14.6 <sup>a</sup>
Iceland	12	37.7	7	22.0	4	12.6	1	3.1	5	15.7
Ireland	70	15.7	56	12.5	38	8.5	6	1.3	66	14.8
Israel	70	9.2	14	1.8	4	0.5	4	0.5	—	—
Italy	1870 <sup>a</sup>	31.7 <sup>a</sup>	1272 <sup>a</sup>	21.6 <sup>a</sup>	386 <sup>a</sup>	6.5 <sup>a</sup>	—	—	1888 <sup>a</sup>	32.0 <sup>a</sup>
Japan	12420 <sup>b</sup>	97.3 <sup>b</sup>	5503 <sup>b</sup>	43.1 <sup>b</sup>	—	—	—	—	3792 <sup>b</sup>	29.7 <sup>b</sup>
Korea	1743	34.5	985	19.5	269	5.3	724	14.3	2414	47.8
Luxembourg	13	25.9	7	13.9	2	4.0	1	2.0	10	19.9
Mexico	467 <sup>a</sup>	4.3 <sup>a</sup>	209 <sup>a</sup>	1.9 <sup>a</sup>	222 <sup>a</sup>	2.1 <sup>a</sup>	157 <sup>a</sup>	1.5 <sup>a</sup>	718 <sup>a</sup>	6.7 <sup>a</sup>
Netherlands	186 <sup>a</sup>	11.3 <sup>a</sup>	181 <sup>a</sup>	11.0 <sup>a</sup>	—	—	35 <sup>a</sup>	2.1 <sup>a</sup>	—	—
New Zealand	68	15.6	46	10.5	39	8.9	4	0.9	110	25.2
Poland	473 <sup>a</sup>	12.4 <sup>a</sup>	141 <sup>a</sup>	3.7 <sup>a</sup>	107 <sup>a</sup>	2.8 <sup>a</sup>	161 <sup>a</sup>	4.2 <sup>a</sup>	544 <sup>a</sup>	14.3 <sup>a</sup>
Portugal	276 <sup>c</sup>	26.0 <sup>c</sup>	94 <sup>c</sup>	8.9 <sup>c</sup>	106 <sup>c</sup>	10.0 <sup>c</sup>	32 <sup>c</sup>	3.0 <sup>c</sup>	376 <sup>c</sup>	35.4 <sup>c</sup>
Slovak Republic	72 <sup>a</sup>	13.3 <sup>a</sup>	33 <sup>a</sup>	6.1 <sup>a</sup>	73 <sup>a</sup>	13.5 <sup>a</sup>	33 <sup>a</sup>	6.1 <sup>a</sup>	78 <sup>a</sup>	14.4 <sup>a</sup>
Slovenia	26	12.8	9 <sup>a</sup>	4.5 <sup>a</sup>	12	5.9	5 <sup>a</sup>	2.5 <sup>a</sup>	36	17.8
Switzerland	255	32.6	—	—	129	16.5	—	—	259	33.1
Turkey	838 <sup>a</sup>	11.6 <sup>a</sup>	647 <sup>a</sup>	8.9 <sup>a</sup>	134 <sup>a</sup>	1.8 <sup>a</sup>	205 <sup>a</sup>	2.8 <sup>a</sup>	728 <sup>a</sup>	10.0 <sup>a</sup>
United Kingdom	510.5	8.3	365	5.9	319.6	5.2	—	—	543	8.9
United States	10335 <sup>c</sup>	34.3 <sup>c</sup>	7810 <sup>c</sup>	25.9 <sup>c</sup>	3495	11.3	—	—	12215 <sup>b</sup>	40.2 <sup>b</sup>

Note: Values without any superscript are for year 2010, and those with superscripts a, b, and c are for years 2009, 2008, and 2007, respectively.

Source: OECD Health Data 2011, June 2011.

and third-party payers, the paperwork is more than an inconvenience as it involves major administrative expenses.

An oft-cited 2003 article by Woolhandler, Campbell, and Himmelstein compared 1999 Canadian and United States administrative costs and calculated U.S. excess per capita administrative costs of \$752, or \$209 billion in aggregate. This implied that a single-payer, Canadian-style health system for the United States would save \$0.71 out of every \$1 of U.S. administrative costs.

Re-examining their data, Aaron (2003) argues that looking at per capita expenditures overstates the difference because it depends on arbitrary assumptions relating to currency values and wage rates. Aaron's preferred comparison observes that administrative costs in the United States accounted for about 31 percent of total health care spending compared to 16.7 percent in Canada. This 14.3 percentage point differential, if applied to the United States, would save 46.1 percent (i.e., 14.3 divided by 31) of U.S. administrative costs (compared to Woolhandler et al.'s 71 percent), or \$489 per capita. However, even this more conservative calculation pointed to excess spending at that time of \$159 billion per year!

Recent analyses verify this result. Pozen and Cutler (2010) break down the \$1,589 difference in 2002 health expenditures per capita between the U.S. and Canada. Adjusting for population size, there are 44 percent more administrative staff in the U.S. system than in the Canadian system. The authors find that higher administrative costs in 2002 accounted for \$616 or 39 percent of the difference. Multiplying this by 310 million Americans, and accounting for the approximately 20 percent rate of inflation from 2002 to 2011, yields a total of \$232 billion dollars in "excess" administrative costs, or between 8 and 9 percent of total U.S. health expenditures. This large cost does not appear to bring commensurate benefits along with it.

## A Comparison

The foregoing data suggest that the Canadian system is more effective than the U.S. system in several respects. Costs are lower, more services are provided, financial barriers do not exist, and health status as measured by mortality rates is superior. Canadians have longer life expectancies and lower infant mortality rates than do U.S. residents.

However, the comparisons do not tell the whole story, nor do they necessarily imply that the United States should adopt the Canadian approach. The Canadian system has had its own financial problems. As a result of unprecedented federal deficits in the 1990s, the Canadian government substantially reduced its cash transfers to the provinces. Despite considerable improvement in the federal government's fiscal health in recent years, the provincial governments face the numerous options to cope with their increased burdens. These include finding new sources of tax revenue, imposing more stringent fee and budgetary controls on health providers, finding ways to increase efficiency in health care delivery, scaling back on benefits by no longer insuring some previously covered services, and imposing user fees.

Similar to the shift we have seen for the United States, the provinces have forced large reductions in hospital capacity with a corresponding substitution of outpatient care for inpatient care. Regional boards with budgetary authority have replaced centralized provincial departments.

Watson and Allin report that sub-national cost-control measures include mandatory annual global budgets for hospitals and health regions, negotiated fee schedules for health care providers, drug formularies, and reviews of the diffusion of technology. Further, many governments have developed pricing and purchasing strategies to obtain better drug prices. In July 2010, the ten provinces and three territories agreed to establish a "pan-Canadian" public sector purchasing alliance of common drugs and medical equipment and supplies.

In short, Canada faces difficult choices in finding a balance between quality and cost while maintaining the principles of its universal health care system. Canada has been able to limit spending and health care's share of GDP despite sluggish recent growth of its economy. To deal with the fiscal constraints, Canada is taking a more directed approach than one involving market reforms.



Critics of the Canadian system charge that health care is rationed in the sense that all the care that patients demand, or would be provided to meet their best interests, cannot be supplied on a timely basis. As noted in the model on the British NHS, rationing below market price leaves some people, who would be willing to pay more, unable to purchase any of the good at all.

Box 22-2 discusses the landmark 2005 *Chaoulli v. Quebec* lawsuit, which found in favor of the plaintiff, who argued that Canadians should have the right to buy rationed goods in the private market. Responding to a 2011 query by the authors, University of Toronto economist Eric Nauenberg notes that *Chaoulli* has yet to have had an impact on buying rationed goods in the private

## BOX 22-2

### *Chaoulli v. Quebec: The Future of Private Health Insurance in Canada*

Georges Zeliotis, a resident of Quebec (province), had suffered from several health conditions requiring that he undergo medical treatment, including heart surgery and several hip operations. He went on a hospital waiting list in Montreal for hip surgery for nearly a year before he was able to receive his treatment. While on the waiting list, Zeliotis sought to determine if he could pay to obtain hip surgery through a private health care facility. He also wanted to purchase private health care insurance should he require similar treatment in the future, but Quebec law prohibited obtaining hip surgery by means of a private facility and purchasing private health care insurance.

Jacques Chaoulli, a physician in Quebec, provided medical services to many of his patients at their homes. Chaoulli appealed to the Quebec government to cover the costs of the home medical treatment offered to his patients and requested that the province grant him the right to establish a private and autonomous hospital. The province denied both of Chaoulli's requests.

Starting in 1997, Zeliotis and Chaoulli challenged specific sections of Quebec's laws. They challenged Article 15 of Quebec's Health Insurance Act, which prohibited private insurance for services covered by the government's insurance plan.

They also challenged Article 11, which prohibited private contracting for medical services in hospitals by physicians who were nonparticipants in the government insurance plan. Whereas the first law concerned how patients could pay for medical services, Article 11 concerned doctors and how they could charge for medical services. It prohibited those doctors who chose not to participate in the government's public insurance plan from setting up private hospitals and then charging directly for their medical services.

In June 2005, after a lengthy court battle, the Canada's Supreme Court (voting 4–3) ruled:

The evidence in this case shows that delays in the public health system are widespread and that, in some serious cases, patients die as a result of waiting lists for public health care. The evidence also demonstrates that the prohibition against private health insurance and its consequence of denying people vital health care result in physical and psychological suffering that meets a threshold test of seriousness.

Where lack of timely health care can result in death . . . where it can result in serious psychological and physical suffering, the . . . protection of security of the person is triggered. In this case, the government has prohibited private health insurance that would permit ordinary Quebecers to access private health care while failing to deliver health care in a reasonable manner, thereby increasing the risk of complications and death. In so doing, it has interfered with the interests protected by . . . the Canadian Charter.

*Section 11 HOIA and s. 15 HEIA are arbitrary and the consequent deprivation of the interests protected by s. 7 [of the Canadian Charter] is therefore not in accordance with the principles of fundamental justice (italics added).*

The Court found that restricting access to private medical services was illegal vis-à-vis the significance of waiting lists for treatment under Quebec's public health system. It set off fierce debate among policymakers about the future of health care in Canada, and six years later (2011), they are still debating its impact.

Sources: [www.mapleleafweb.com/features/medicare/charter-health-care/background-chaoulli.html](http://www.mapleleafweb.com/features/medicare/charter-health-care/background-chaoulli.html), accessed August 14, 2005; <http://scc.lexum.org/en/2005/2005scc35/2005scc35.html>, accessed July 20, 2011.

market. In particular, this decision applied only to Quebec, and not the nine other provinces. Secondly, Quebec introduced legislation limiting private insurance and private service provision to a set of services (such as hip and knee replacements, cataracts, and diagnostic radiology) such that it would be difficult for a private market to emerge for such services.

Speaking to the authors, University of Toronto law professor Colleen Flood terms the 4–3 Chaoulli decision as “poorly written.” Although numerous lawsuits are in process, Professor Flood advises health policy experts to follow an British Columbia expected court decision. She predicts the development of a parallel private insurance system comparable to the United Kingdom or Ireland that would open the private health care market to those able to afford it.

Though specific estimates of service shortages are not available, the consensus is that the limits on capacity and on new technology result in longer waiting periods for hospital services. The “safety valve” of a private system, as in the United Kingdom, for those who are willing to pay more is not readily available, although some Canadians (particularly those near large U.S. border cities such as Buffalo and Detroit) use U.S. facilities for this purpose.

Defenders of the U.S. approach claim that the waiting and queues found in Canada would be unacceptable to many U.S. patients. They view the greater level of amenities in the United States, and the greater availability of specialized care, together with high-tech medicine, as indicators of superior quality.

June and Dave O’Neill (2008) raise three questions regarding American and Canadian differences:

1. What differences in health status can be attributed to the two systems?
2. How does access to needed health care resources compare?
3. Is inequality in access to resources different?

They use a data set from the Joint Canada/U.S. Survey of Health (JCUSH), designed and conducted jointly by Statistics Canada and the U.S. National Center for Health Statistics, which asked the same set of questions under similar conditions to representative samples of U.S. and Canadian residents.

Regarding differences in health status, the authors examine various aggregate data sources to break down differences in life expectancy (Canadians live longer) and causes of mortality (Americans are more susceptible to nondisease determinants such as accident and homicide). Americans tend to have lower birthweight babies (with higher mortality rates), and Americans tend to be more obese. The authors argue that although health care systems can have impacts (including better prenatal care), they have much less direct effects on important factors such as obesity.

With respect to unmet needs, the authors use the JCUSH to examine difficulty in receiving health care. Those with an unmet need were asked the reason for the unmet need—had to wait too long or service not available; cost (i.e., could not afford service); or a reason other than those two. As Table 22-6 shows, the “wait too long/service not available” reason (56.3 percent) dominates among the Canadians who had an unmet need, while for U.S. residents cost (54.7 percent) was the major factor and “waiting too long” (13.2 percent) was relatively minor.

The authors also examined the effect of unmet needs on the Health Utility Index (HUI) of health status and found that in Canada unmet needs reduced the HUI by 0.097 (compared to a mean of 0.898) when the individual cited waiting as a reason for unmet need. The effect was much smaller and not significant for Americans.

Inequality in access relates individuals’ scores on the HUI to income. The HUI provides a description of an individual’s overall functional health based on eight attributes—vision, hearing, speech, mobility (ability to get around), dexterity of hands and fingers, memory and thinking, emotion, and pain and discomfort. If a single-payer system equalized health irrespective of income, one would expect income to show a zero impact. Comparisons of subjects in the U.S. and Canada showed the relationship of health to income to be *roughly similar* in the two countries.

**TABLE 22-6 Percent of Those with Self-Reported Unmet Health Need and Reason for Unmet Need**

	Ages 18–64		Ages 65+	
	Canada	U.S.	Canada	U.S.
% with unmet need	11.3	14.4	7.4	6.4
Reason for unmet need (%):				
Wait too long or not available	56.3	13.2	51.9	24.6
Cost	8.6	54.7	2.2	35.9
Other reasons	36.6	33.2	45.9	42.6

\*Categories with reasons are not mutually exclusive and therefore will not add to a hundred

Source: Reprinted from O'Neill, June E. and O'Neill, Dave M. (2008). "Health Status, Health Care and Inequality: Canada vs. the U.S.," *Forum for Health Economics & Policy: Vol. 10: Issue 1* (Frontiers in Health Policy Research), Article 3. <http://www.bepress.com/fhep/10/1/3>, with permission.

The authors conclude that the U.S. and the Canadian systems provide similar results. The need to ration "free" care may ultimately lead to long waits or unavailable services and to unmet needs. In the United States, costs are more often a source of unmet needs, but costs "may be more easily overcome than the absence of services." When those ages 18 to 64 were asked about satisfaction with health services and the ranking of the quality of services recently received, more U.S. residents than Canadians responded that they were fully satisfied (51.5 percent vs. 41.3 percent) and ranked quality of care as excellent (40.4 percent vs. 37.7 percent). Satisfaction and quality of care may relate to expectations as well as to objective measures.

The authors finish with the caveat that they do not address the differential in per capita health care expenditures, which are over 80 percent higher in the United States. They ask, "Is the U.S. getting sufficient additional benefits to justify these greater expenditures and where should we cut back if cutbacks must be made? Alternatively, what would Canada have to spend to increase their technical capital and specialized medical personnel to match American levels or to eliminate the longer waiting times? And would it be worthwhile to them to do so?"

Duclos and Échevin (2011) address in more detail the income-health relationship brought up by the O'Neills. They rank Canada and the U.S. using data from the Joint Canada/United States Survey of Health. Using "stochastic dominance" methods, they find that Canada dominates the United States over the two groups of lower health statuses in terms of the bi-dimensional distribution of health and income. This occurs because Canada has better health distribution, a lower correlation between income and health, and lower income inequality.

## DIFFERENT SYSTEMS: THE PUBLIC'S EVALUATION

Decisions about health care systems ultimately reflect the attitudes of the public with regard to satisfaction, cost of care, and quality of care. Schoen and colleagues (2010) surveyed citizens of Australia, Canada, France, Germany, the Netherlands, New Zealand, Norway, Sweden, Switzerland, the United Kingdom, and the United States who had had recent experience with their countries' health care systems, regarding general satisfaction, access to care, cost of care, and quality of care. All of the countries are economically advanced, but they have a wide range of insurance and care systems.

We have described the United States and the United Kingdom health systems in detail. In the German universal coverage system, competing insurers offer a standard comprehensive benefit package, with higher-income households allowed to opt out of statutory “sickness funds” to purchase private coverage (10 percent of the population in 2009). Switzerland and the Netherlands require residents to purchase a standard, comprehensive health insurance package, offered by non-profit private insurers in Switzerland and by a mix of nonprofit and for-profit insurers in the Netherlands. All of the countries in the study allow some role for private insurance that can pay for extra benefits, or for some part of patients’ cost sharing.

Comparative international studies can be very difficult to do, and this one takes special care to use similar methods and questionnaires across the eleven countries. The survey used a common questionnaire translated and adjusted for country-specific wording. It was conducted by computer-assisted telephone interviews of random samples of adults age eighteen or older. The final country samples ranged from 1,000 (New Zealand) to over 3,500 (Australia).

Although the researchers evaluate many dimensions of the health economy, we focus on issues of access and costs. As we noted with the O’Neills’ study, satisfaction with a system comes both from expectations and system performance. Different people may register different levels of satisfaction with the same service and same outcome, depending on their expectations.

Table 22-7A shows that access varies according to the service needed. Those participating in the Swiss system had almost instant access to a doctor or a nurse (93 percent seeing a provider either the same or the next day), whereas Canadians were less than half as likely (45 percent) to see a provider either the same or the next day, and 33 percent reported waiting 6 or more days. In contrast, Germans (78 percent) and Americans (68 percent) had the shortest waits for surgery,

**TABLE 22-7A Access – Schoen et al (Exhibit 3)**

**Adults’ Experiences With Access To Health Care In Eleven High-Income Countries, 2010**  
Percent of adults who

Country	Saw a doctor or nurse last time they needed care		Needed after-hours care or used the ED and reported		Wanted to see specialist <sup>a</sup>		Wanted for elective surgery <sup>b</sup>	
	Same or next day	Waited 6 days or more	Somewhat/very difficult to obtain care after hours <sup>c</sup>	ED use in past 2 years	Less than 4 weeks	2 months or more	Less than 1 month	4 months or more
AUS	65	14	59	33	54	28	53	18
CAN	45	33	65	44	41	41	35	25
FRA	62	17	63	27	53	28	46	7
GER	66	16	57	22	83	7	78	0
NETH	72	5	33	26	70	16	59	5
NZ	78	5	38	29	61	22	54	8
NOR	45	28	45	26	50	34	44	21
SWE	57	25	68	35	45	31	34	22
SWI	93	2	43	22	82	5	55	7
UK	70	8	38	25	72	19	59	21
US	57	19	63	37	80	9	68	7

Source: 2010 Commonwealth Fund international health policy survey in eleven countries.

Notes: Sample sizes for each country are reported in Exhibit 2. Significance tests are available in the Technical Appendix, which can be accessed by clicking on the Technical Appendix link in the box to the right of the article online. ED, emergency department. <sup>a</sup>If they needed to see a specialist within the two previous years. <sup>b</sup>If they had elective surgery within the two previous years. <sup>c</sup>If they answered the question and had needed after-hours care.

**TABLE 22-7B Costs – Schoen et al (Exhibit 2)**

Adults' Confidence In And Cost-Related Experiences With Health Care In Eleven High-Income Countries, 2010  
Percent of adults who

Country (sample size)	Were confident/very confident that if seriously ill they would		Had problems with access because of cost in previous year			Had out-of-pocket medical spending in previous year		Had serious problem paying or were unable to pay medical bills in previous year
	Receive most effective treatment, including drugs, diagnostic tests	Be able to afford care needed	Did not see doctor when sick or did not get recommended care	Did not fill Rx or skipped doses	Had either access problem	\$200 or less	\$1,000 or more	
AUS (3,552)	76	64	18	12	22	35	21	8
CAN (3,302)	76	68	8	10	15	51	12	6
FRA (1,402)	85	73	9	7	13	47	4	9
GER (1,005)	82	70	23	6	25	41	8	3
NETH (1,001)	88	81	4	3	6	39	9	4
NZ (1,000)	84	75	12	7	14	61	7	6
NOR (1,058)	70	69	8	6	11	33	16	5
SWE (2,100)	67	70	6	7	10	50	2	5
SWI (1,306)	89	78	9	4	10	20	25	6
UK (1,511)	92	90	5	2	5	76	1	2
US (2,501)	70	58	28	21	33	31	35	20

Note Significance tests are available in the Technical Appendix, which can be accessed by clicking on the Technical Appendix link in the box to the right of the article online.

Sources: Table 22-7A from Schoen et al Exhibit 3; Table 22-7B from Schoen et al Exhibit 2. "Copyrighted and published by Project HOPE/Health Affairs as Schoen, Cathy et. al., "How Health Insurance Design Affects Access To Care And Costs By Income In Eleven Countries", *Health Affairs*, 29, no.12 (2010):2323–2334 (published online November 18, 2010; 10.1377/hlthaff.2010.0862)," with permission.

with Canadians, Norwegians, Swedes, and British all having over 20 percent of those sampled waiting 4 months or more for elective surgery. See Box 22-3 for a Canadian example.

Costs are another matter. Swedes, Americans, and Norwegians lag behind the other countries in their confidence about receiving the most effective treatment. UK respondents (92 percent—see Table 22-7B) have the most confidence. 76 percent of the UK respondents reported spending \$200 or less out-of-pocket, with only 1 percent reporting spending \$1,000 or more. Twenty percent of the US respondents reported "serious problems" or inability to pay medical bills in the previous year. No other country in the group reported even 10 percent of the population with this concern.

### BOX 22-3

#### "Someone Else Needed It More than I Did"

While vacationing in Florida, one of the authors (Goodman) played golf with a Canadian man who remarked that this was his first round after having had his hip replaced. The surgery had incurred no out-of-pocket costs, and he felt fine. When asked how long he had to wait for surgery, he responded that he had waited 18 months. How did he feel about waiting? "It didn't bother me . . . someone else needed it more than I did."

## DIFFERENCES IN HEALTH CARE SPENDING ACROSS COUNTRIES

Different countries have different incentive systems, and, in fact, have differing shares of national product in the health care sector. Having described the systems, and examined the health sector shares of national product, it is appropriate now to explore why the shares differ.

### A Model of Health Expenditure Shares

Consider a model of health expenditures and call total expenditures on health care  $E$ . By definition, these expenditures equal the price of health care multiplied by the quantity of health care consumed, or  $E = PQ$ . Defining the share of national income spent on health care as  $s$ , we calculate  $s$  as the ratio of  $E$  to national income,  $Y$ , or:

$$s = PQ/Y \quad (22.1)$$

The share,  $s$ , can increase because either the price or quantity has increased, or because the national income has decreased. In fact, mathematically it can be shown that:

$$(\% \text{ Change } s) = (\% \text{ Change } P) + (\% \text{ Change } Q) - (\% \text{ Change } Y) \quad (22.2)$$

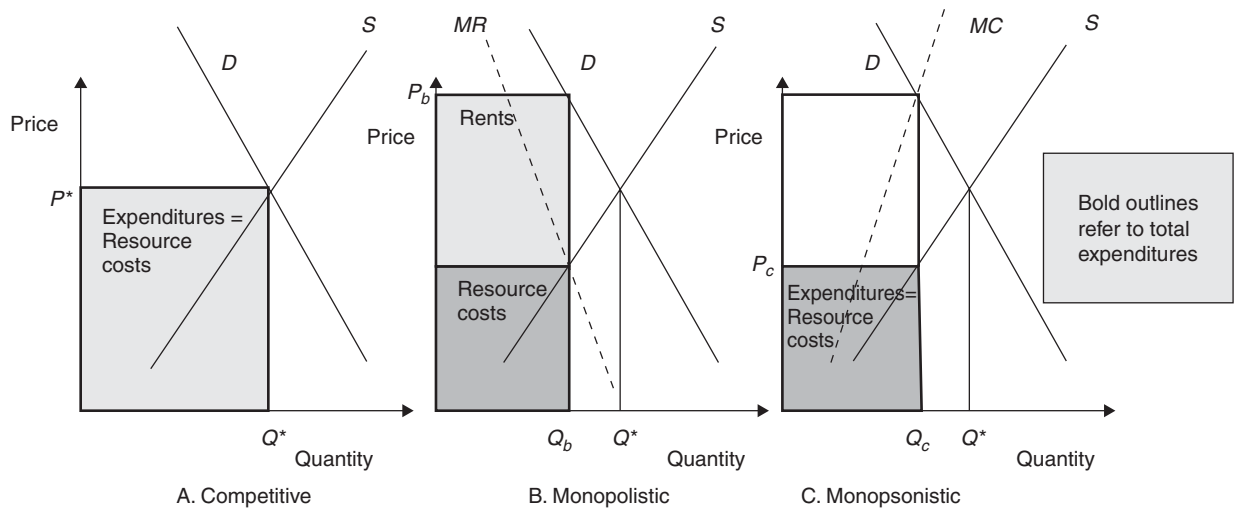
The preceding expression is an identity, mathematically true by definition. Even so, it can provide useful insights. If the price of health care,  $P$ , increases by the same rate as the price of all other goods, so that  $Y$  increases at that same rate, then the health care share of national income does not change.

**APPLYING THE MODEL** Rather than looking just at the percentage changes that occur, we try to examine why. Consider, for example, that prices of health care relate to the kind of health system the country has or to the social insurance scheme. Also, consider that the quantity of health care used,  $Q$ , tends to increase when national income,  $Y$ , increases. Note further that the quantity of health care,  $Q$ , is negatively related to the price of health care,  $P$ , through the demand relationship.

Consider several ideas in turn:

1. An increase in the price of health care would increase the share if there were no consumer response. The extent to which consumers reduce quantity demanded (in response to price changes) will offset the increase in prices.
2. An increase in the share of population who use health care would tend to increase health care expenditures.
3. An increase in national income,  $Y$ , unaccompanied by an increase in health care demand would decrease the share. However, if increased income leads to increased demand, the effect depends on the demand elasticity. A one percent increase in national income that leads to a one percent increase in expenditures (that is, the income elasticity equals +1.0), will result in a constant share.

It is important to examine the differences among countries with respect to payment mechanism. Economists often implicitly view expenditures in the context of perfectly competitive markets. If this view is valid, as noted in Figure 22-5A, then the total health expenditures figure (the numerator of fraction of GDP going to health care, indicated by the box with bold outlines) accurately reflects the resource costs  $P^*$  of health care at the margin. Anderson and colleagues (2003), however, note that the markets for the health workforce (especially physicians) are still largely national and even local within countries. Moreover, many markets related to health care within localities do not satisfy the rigorous conditions of the textbook model of competition.



**FIGURE 22-5** Health Expenditures by Market Structure

We find varying degrees of monopoly power on the “sell” side of the market and varying degrees of monopsony power on the “buy” side. Because monopolists (Figure 22-5B) equate marginal costs to marginal revenues, they are able to raise prices above those they would obtain in perfectly competitive markets. This earns them “rents,” defined as the excess of the prices actually received by sellers above the minimum prices the sellers would have to be paid to sell into the market. Figure 22-5B shows that the resource costs (the box defined by the supply curve) are considerably less than the total expenditures (the sum of the resource costs and the monopoly rents), with the difference going as rents to providers. Monopoly quantity  $Q_b$  is also less than  $Q^*$ , under competitive markets because in order to increase prices, monopolistic providers must sell less.

Countries differ in the degree to which they try to reduce the rent earned on the supply side through the creation of market power on the buy (monopsony) side of the market. A single-payer system (similar to the one used by Canadian provinces) would be called a “pure monopsony.” Because a pure monopsonist (Figure 22-5C) must pay increased resource costs to all supply factors, the monopsonist faces a market marginal cost curve, not unlike the monopolist’s marginal revenue curve. Here, the producer provides quantity  $Q_c$ , but expenditures are much smaller than in Figure 22-5B.

Note again that in either the monopolistic or the monopsonistic case, the quantity of services provided falls short of the optimum  $Q^*$ . We have intentionally drawn the Figure 22-5 monopoly and monopsony quantities  $Q_b$  and  $Q_c$  to be identical, but in the monopolistic case, extra resources are *transferred* as monopoly rents from the buyers to the sellers.

In the United States, the Medicare program and Medicaid programs do possess some monopsonistic purchasing power, and large private insurers may enjoy some degree of monopsony power in some localities, but the highly fragmented buy side of the U.S. health system is relatively weak by international standards. This is one factor, among others, that might explain the relatively high prices paid for health care and for health professionals in the United States.

In comparison, the government-controlled health systems of Canada, Europe, and Japan allocate considerably more market power to the buy side. As noted earlier in the section on the Canadian system, in each Canadian province, the health insurance plans operated by the provincial governments constitute pure monopsonies. They purchase (pay for) all of the health services that are covered by the provincial health plan and used by the province’s residents. However, note that even pure monopsonists are ultimately constrained by market forces on the supply side—that is, if

fees are too low, health care providers will not supply their goods or services. However, monopsonistic buyers may enjoy enough market power to drive down the prices paid for health care and health care inputs fairly close to those reservation prices characterized by the supply curve.

For the United States, most measures of aggregate utilization, such as physician visits per capita and hospital days per capita, typically lag below the medians of other Western countries. Since spending is a product of both the goods and services used and their prices, this implies that U.S. consumers pay much higher prices than consumers elsewhere. However, U.S. policymakers must reflect on what Americans are getting for their greater health spending. The authors conclude that the answer lies in the higher prices paid by U.S. health consumers.

In a follow-up analysis, Anderson and colleagues (2005) revisit the high level of U.S. health expenditures, examining two commonly proposed explanations. The first is that other countries have constrained the supply of health care resources, particularly for elective services, which has led to waiting lists and lower spending. The second is that the threat of malpractice litigation and the resulting defensive medicine in the United States adds to malpractice premiums and, more importantly, the practice of defensive medicine, hence increasing costs.

Regarding the first explanation, if consumers in other countries must wait for procedures that U.S. consumers can get immediately, then the international consumers are bearing waiting time costs that do not enter national accounts. The researchers argue, however, that the procedures for which waiting lists exist in some countries represent a small part of total health spending. Using U.S. survey data, they calculated the amount of U.S. health spending accounted for by the 15 procedures that account for most of the waiting lists in Australia, Canada, and the United Kingdom. Total spending for these procedures in 2001 was \$21.9 billion, or only 3 percent of U.S. health spending in that year.

Mello and colleagues (2009) provide updated analysis to address the claim that the U.S. medical liability system leads to unneeded care and extra expenses. They examine indemnity payments, administrative costs, and the identifiable hospital and physician costs due to defensive medicine. They estimate these costs to be \$55.6 billion, in 2008 dollars, or about 2.4 percent of total health care spending. This leads to two inferences. First, not all of these costs represent waste—some of them almost certainly provide positive benefits to the patients, or appropriately deter potential malpractice. Second, even eliminating all defensive medicine would have only a minor impact on overall health care spending.

## CONCLUSIONS

In this chapter, we have examined a variety of health care systems found elsewhere. Variations exist in terms of financing, provider payment mechanisms, and the role of government, including the degree of centralization. The United States stands out as the country with the highest expenditures on health care as well as the highest percentage of the GDP devoted to health care.

Systems that ration their care by government provision or government insurance incur lower per-capita costs. In the largely private U.S. system, however, waiting times tend to be shorter than in rationed systems, a conclusion that follows from theory as well as from observation. Americans have been more dissatisfied with their health system than Canadians or Europeans have been with theirs. The study of comparative systems suggests several features of other systems that may be worth adopting. It also suggests that cultural differences among countries could dictate that systems tailored to the local culture continue to differ even in the long run.

Countries have sought to control costs in a variety of ways. Strategies include a global budget, increased cost sharing, and various market incentives. Single-payer plans, as in Canada, offer theoretical economies of administration, but it may be difficult to identify whether the observed cost advantages in Canada would survive translation into a reformed U.S. system.

The United States has fundamentally left cost containment to managed care. Although managed care achieves cost savings and may have contributed to the decline in the U.S. health cost



growth rate, its potential will be limited to the extent that employers fail to offer true financial advantages to consumers who choose the low-cost health plans. American-style HMOs, for example, probably would not transfer unchanged to other countries because of cultural and system structure differences.

There is a growing agreement that reform of the U.S. health system must address four critical elements:

- A health “safety net” for all residents, irrespective of age, health status, or employment status
- Mechanisms that promote cost containment
- Choice for patients and providers
- Ease in administration

The next chapter examines these reform elements in more detail.

---

## Summary

1. A useful typology of health benefit systems provides four categories:
  - Traditional sickness insurance: the private insurance market approach, with state subsidy
  - National health insurance: the state establishes a national-level health insurance system
  - National health service: the state provides the health care
  - Mixed system: mixed elements of the previous systems
2. Among all countries, the United States is by far the biggest spender in absolute per-capita terms. It is also the biggest spender as a share of GDP.
3. The United Kingdom’s NHS provides relatively easy access to primary and emergency care. It rations elective services either through long waiting lists or by limiting the availability of new technologies. The NHS devotes considerable resources to high return services, such as prenatal and infant care.
4. The United Kingdom has reformed its health system to include elements of competition. The United States has fewer practicing physicians per capita than the United Kingdom and inpatient beds per capita as well. Health care spending per capita in the United Kingdom, however, is only 44 percent of the U.S. level.
5. The Chinese system has moved from a more “command-based” system to a more market-based system since the early 1980s. These changes, however, have made health care less affordable for many as the proportion of those ill who did not seek care for financial reasons increased in both urban and rural areas.
6. Compared to the U.S. system, the Canadian system has lower costs, more services, universal access to health care without financial barriers, and superior health status. Canadians have longer life expectancies and lower infant mortality rates than do U.S. residents.
7. Canada’s single-payer system appears to have substantially lower administrative cost burden than the United States.
8. National health systems appear to reduce health spending. However, careful analysis across alternative systems must impute the additional time costs, as well as differential quality of care in NHS systems, before deciding conclusively on the full costs of alternative systems.
9. Comparisons of health care systems feature varying degrees of monopoly power on the “sell” side of the market and varying degrees of monopsony power on the “buy” side. Because monopolists equate marginal costs to marginal revenues, they can raise prices above those they would obtain in perfectly competitive markets, thus earning “rents,” the excess of prices actually received by sellers above the minimum prices the sellers would have to be paid to sell into the market.
10. Analysts believe that a monopolistic model characterizes the U.S. system more than systems (Canada, Europe, or Japan) that allocate more market power to the buy side.

## Discussion Questions

1. Discuss the factors that may lead one nation to spend more per person on health care than another nation. What are the implications of finding health care to be income elastic in cross-national studies? When health care is income elastic, will richer countries tend to have a higher or lower proportion of GDP spent on health care?
2. In countries in which there is nonprice rationing for care, waiting time costs may be substantial. How could you evaluate the economic costs of the waiting time?
3. Create a table comparing the British, Chinese, and Canadian health care systems with respect to financing, availability, and costs of care. How do they compare with the system in the United States?
4. Suppose that the price of health care services rises and the quantity demanded falls. Under what conditions might the health care share of GDP fall? Rise?
5. It is important to compare items under the rubric of “all else equal.” What are some of the crucial factors that must be adjusted when comparing health expenditures across countries?
6. Distinguish between an NHI system and an NHS. Provide examples of each. What kind of a system does the United States have?
7. As noted in Table 22-7, consumer satisfaction varies among a number of measures of access and cost. Are these measures useful indicators of the performance of health care systems?
8. A common thesis in economics is that markets are efficient unless characteristics are present that lead to market failure. What sorts of market failure in the health economy can be used to justify adoption of universal NHI? What kinds of government failure can be used to argue against this proposal?
9. Do countries with more comprehensive national programs for the provision of health care tend to have lower average costs than the United States? Do they have lower rates of growth in costs? Discuss.
10. Speculate about the level of technology available across countries. Do you think that better health care is available in the United States than in Canada? Do international health indices suggest this? What are the complicating issues?
11. What ideas discussed in this chapter would be suitable to recommend to a country just now revising its health system? To pursue equity? To pursue cost containment?

## Exercises

1. Consider the allocation of services in the United Kingdom’s NHS, as noted in Figure 22-2. If the government raises the administered price up from  $P^*$ , trace what would happen to expenditures in the NHS and in the private sectors.
2. Define income elasticity of health care demand. If income increases by 1 percent and the income elasticity of health care demand is +0.5, does the share of income going to health care increase or decrease? Why?
3. Suppose the price elasticity of health services is  $-0.5$ . What will happen to the share of health care expenditures, given a 10 percent decrease in health care prices?
4. For more advanced students, Table 22-1 provides data for at least rudimentary estimates of income elasticity of health care expenditures. Estimate a regression equation of the following form:
 
$$\text{Log (Expenditures per capita)} = a + b \text{ log (GDP per capita)}$$
 What is the implied income elasticity of expenditures across countries?
5. From the data in Table 22-1, estimate a regression equation of the following form:
 
$$\text{Log (Death Rates per 1,000)} = c + d \text{ log (Expenditures per capita)}$$
 What does your resulting equation say about the “effectiveness” of expenditures per capita (without adjusting for any other factors)?
6. Figure 22-5 shows various types of national health insurance systems.
  - (a) Compare the total expenditures in panels A and B. Which set of expenditures is larger? What determines which will be larger? Why?
  - (b) Compare the total expenditures in panels B and C. Which set of expenditures is larger? In which are resource costs larger? Why?
  - (c) If the demand curves truly reflect consumer preferences, which of the three panels is economically efficient? Show the economic losses and the transfers for those panels that are not economically efficient.
7. The Organization for Economic Cooperation and Development (OECD) provides some of the best data available for comparative international work. Its website is [www.oecd.org](http://www.oecd.org). Use the OECD data to examine the health care system of Mexico along the following dimensions:
  - financing
  - expenditure
  - technology
  - coverage

# Health System Reform

- Goals of Reform
- Ensuring Access to Care
- Competitive Strategies
- Health System Reform and International Competitiveness
- Quality of Care
- The Patient Protection and Affordable Care Act (PPACA) of 2010
- Conclusions

Chapter 22 examined national health insurance (NHI) programs in several major industrialized countries. Although the United States established a more comprehensive health care system in 2010, many issues (as well as court challenges) remain. In 2011, 50 million Americans were still without health insurance at any given time. Many favor universal health care coverage as a solution to the problems of access and costs. Others argue that we can meet health care objectives more effectively through reforms that reduce the role of government and instead take advantage of market forces. The basic issues in health system reform and alternative reform proposals are the focus of this chapter. We finish the chapter by evaluating the 2010 Patient Protection and Affordable Care Act (PPACA).

## GOALS OF REFORM

Most would agree that a national health system reform needs to address these four elements:

- A health “safety net” for all residents, irrespective of age, health status, or employment status
- Mechanisms that promote cost containment
- Choice for patients and providers
- Ease in administration

Consider the four elements in order:

**Safety net**—Large portions of the U.S. population receive inadequate health care by almost any criteria. While Medicare provides almost universal health care for those over age 65 and Medicaid/CHIP are making great inroads into the population under age 18, millions of Americans lack access to levels of health care that even the most conservative analysts would view as adequate.

**Cost containment**—The United States spends well over one in six dollars of its GDP on health care, and expanded coverage will almost certainly increase that ratio. While some analysts have argued that this amount may reflect consumer preferences for high quality health care, there are clearly avenues relating to administrative costs, and ineffective treatment, that could reduce overall health care costs. Further, few Americans would desire cost containment at the expense of the quality of the health care.

**Choice for patients and providers**—The failure of President Clinton’s reform initiative in 1994 made it clear that Americans will reject any national health insurance policy that can be characterized as “one size fits all.” A successful plan must provide choices of providers and treatments.

**Ease in administration**—Consider the weekly trip to the supermarket. The decision as to where to shop and what to buy, while constrained by budgets and the prices of the goods, is administratively simple. People go where they shop, buy what they need, and need not deal with bureaucrats. Contrast that to current health insurance systems, with different application forms, insurance forms, cards, and a myriad of questions about who pays for what, and whether what one has paid will be reimbursed. While purchasing health care is obviously more complicated than purchasing food, any national health care policy that simplifies the process would be desirable.

Many reform proposals face the dilemma whether to fund coverage by *individual* mandate, *employer-employee* mandate, or *general revenues*. An individual mandate is a law that requires individuals to buy health insurance for themselves, with subsidies for those who cannot afford it. The subsidies usually would be funded out of general revenues. Employer-employee mandates would require taxes on wages for the employee’s share. The employer’s share may also fall on the employee in the form of lower wages. Subsidies out of general revenues would provide for the unemployed.

Those who advocate a Canadian-style system seek a single-payer system with government revenues providing most of the NHI. Yet another alternative is the medical savings accounts method of payment that would allow people to set up a tax-free savings account out of which they can pay the out-of-pocket costs of their health care. Usually, these systems entail health insurance with high deductibles and coinsurance.

Some plans define a minimum acceptable insurance level, with people allowed to purchase more extensive coverage if desired and if they can afford it. Others define one plan to fit all. In early debates, many plans featured universal coverage. As support for universal coverage waned in Congress, more plans offered lesser goals for reaching the population.

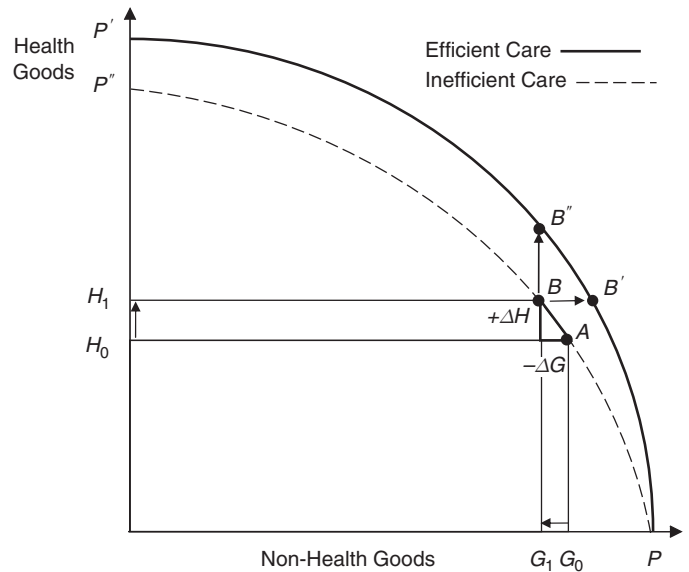
## Basic Issues in Reform

Any reform program must difficult questions. One fundamental question is service coverage. Clearly, as more services are covered or mandated and more provider types are included, the costs will increase.

Figure 23-1 shows a health care system that allocates its resources to goods and health at Point A. It would be best, of course, if A were on the production possibility frontier of efficient production for health and all goods, the solid line  $PP'$ , but there are many reasons that it is probably not. Ineffective treatments, needless tests, and excessive paperwork, may all provide less health (and other goods) than possible, so that we see an interior frontier indicated by the dashed line  $PP''$ .<sup>1</sup>

Assume that the society determines to provide a safety net for all residents, increasing the amount of health goods provided from  $H_0$  to  $H_1$ . The *economic cost* of providing  $\Delta H = H_1 - H_0$  of health is the amount of  $G$  given up, or  $\Delta G = G_0 - G_1$  at point B. If we could control costs, or provide

<sup>1</sup> For simplicity on the dashed curve we have drawn the production of all other goods as efficient at point P (on curve  $PP''$ ), although there is no reason to believe that other goods are produced more (or less) efficiently than health.



**FIGURE 23-1** The Resource Costs of Health Reform

health more efficiently, society might plausibly reach a point like  $B'$  or  $B''$ , on the efficient frontier. One of the underlying goals of reform would be to move to a more efficient production of health from health care. A related issue is whether there will be cost-sharing for covered services and, if so, what type of cost-sharing arrangement will occur.

The question of who is covered can be equally difficult to address. At any time in the United States, there are millions of foreign students, visitors, and temporary workers as well as millions of illegal immigrants. The difficulty of determining the covered population in some cases is evidenced by the acrimonious debate over care provided (and paid by governments) to undocumented workers and their families.

A third major issue is how to fund health reform. Will it rely on general tax revenues or will funding come from mandates on businesses and/or individuals? In either case, where will the burden of funding ultimately rest?

The most challenging issue is to determine whether health reform will build largely on the existing framework of government programs and private employment-based insurance with most of the reform effort aimed at cost-containment and reducing the pool of uninsured. Other proposals, especially market-oriented proposals, attempt to attain these goals by weakening the link between private insurance and employment.

### The Costs of Universal Coverage

National Health Insurance programs that guarantee universal coverage certainly cost a great deal of money, but it is important to distinguish which costs are incremental. In other words, what are the *additional* costs to society from the imposition of NHI?

From society's point of view, the incremental cost of NHI in the United States is the extra total expenditure on health care incurred if we switched to national health insurance. Inasmuch as most people already have insurance for almost all hospital care and most physician care, the extra cost of NHI would be much smaller than many expect.

The truly incremental costs stem from several sources. First, the major reason for switching to a NHI plan is to extend coverage to the 50 million uninsured. It should be understood that the uninsured already consume health care. Zero insurance does not necessarily mean zero care.

Using data from the 2002–2004 Medical Expenditure Panel Surveys (MEPS), a nationally representative survey of the civilian, non-institutionalized population, Hadley and colleagues

(2008), in what remains the most comprehensive study to date, estimated that uninsured Americans consumed \$86 billion worth of health care in 2008. This total consisted of \$30 billion in out-of-pocket costs and \$56 billion in uncompensated care. Governments picked up about \$43 billion of the latter.

The authors then projected that the incremental cost of providing full-year coverage for all uninsured would amount to \$123 billion, so that total spending of those currently uninsured would rise from \$86 to \$209 billion. This incremental cost represented 5.1 percent of total health care spending in the United States and slightly less than 1 percent of its GDP. The authors base their estimates on the utilization patterns of lower-income and lower-middle income individuals. More or less generous plans as well as higher or lower payment rates to providers would raise or lower estimated costs accordingly.

Second, the insured population will cost more to the extent that an NHI plan provides greater typical coverage than people already choose to buy or have provided to them by other sources. Third, any tax-supported system of financing care potentially entails a deadweight loss to society, as taxpayers respond to the changed incentives. The deadweight losses that accompany tax increases mean that some efficiency loss will result, caused by the disincentives to work and invest. This is true even if the program is of the employer-mandated type, because a law forcing employers to incur expense is really a tax.

The incremental costs constitute real costs to society, because, as noted in Figure 23-1, society must divert resources from elsewhere to pay them. In contrast, differences in financing methods (determining who pays) mean less in economic terms. It may be politically more palatable to choose a plan that does not greatly expand the government budget, and employment-mandated plans may be attractive politically for this reason. Nonetheless, society incurs the cost irrespective of whether it finances it through the government or through mandates to individuals or employers by law.

## ENSURING ACCESS TO CARE

In this section, we group reforms by their two main motivations: the desire to see that sick people get health care, and the desire to control the rising cost of health care.

### Employer versus Individual Mandates

The country that wishes to provide universal coverage for health care must choose one scheme or another to extract resources from its households. Schemes for employers or government to pay the bills are only mechanisms by which households ultimately pay. The U.S. debate features and contrasts two mechanisms: employer versus individual mandates.

*Employer mandates* form the backbone of the health systems in Europe, Latin America, and Asia (Krueger and Reinhardt, 1994). Under employer mandates, employers must procure health insurance for their employees and their dependents. Although the employer writes the check, the firm undoubtedly will pass on as much of this cost as it can to customers in the form of higher prices or to employees in the form of lower wages. The *individual mandate*, in contrast, obligates all residents to purchase health insurance for themselves and their families, either from private insurance (individually purchased) or through a group, such as a work group, professional organization, or religious group. The government subsidizes the poor in their purchases through taxation of those who have more money.

In Chapter 11, we showed that a lower market money wage rate leads an employer to hire more workers. Assuming at the outset that there are no benefits and that the market wage is \$20 per hour, employers will hire workers as long as the marginal revenue from the goods those workers produce exceeds the \$20 per hour wage. To begin, assume that the employer hires 1,000 workers.

Suppose that an NHI mandates the provision of a health benefit for all workers that costs \$1 per hour of work. If the mandated benefit is worth at least \$1 per hour to the workers, and costs

exactly \$1 per hour for the employer to provide, employers who were previously willing to pay \$20 will now pay \$20 less the \$1 cost to provide the mandated benefit. Other points on the demand schedule will also change by the \$1 cost of the benefit.

Workers previously willing to accept a wage of \$20 will now be willing to supply their labor for \$1 less since they value the mandated benefit at \$1. The net wage (money wage + the value of the benefit) remains unchanged at \$20, but the equilibrium money wage falls to \$19, or by exactly the amount of the benefit. Workers accept lower money wages, and the same 1,000 workers are employed at the same net wage, \$19 in money wages plus the \$1 benefit. The workers are no worse off at a wage of \$19 with the mandated benefit than at \$20 without the mandated benefit because the benefit is worth the \$1 that it cost in reduced wages.

Business leaders often complain that employer mandates either will reduce profits or force firms out of business. Such responses implicitly assume that their firm is the only one affected by the mandate. If all firms faced the same labor costs, it is doubtful that closings would result. In the short run, firms would pay workers less, take less in profits, and/or raise prices to consumers. “Economists are convinced, however, that in the longer run more and more of the cost of the employer mandate would likely be shifted backward to employees . . . through smaller real (inflation-adjusted) increases in wages than would have been warranted by long-run productivity gain” (Krueger and Reinhardt, 1994, p. 44).

If the labor supply is very unresponsive to the wage rate, the employer’s lower wage expenditures will offset extra health benefit costs regardless of whether the laborers value the benefit highly or not at all. Most economists would agree that the aggregate labor supply, at least in the long run, is nearly vertical (totally inelastic) for men, and also highly inelastic for women. In this scenario, the mandate has little effect on producers, their competitive position, or their customers. Whether the program helps or harms the well-being of society under conventional economic analysis depends largely on whether workers value their health insurance as much as or more than they did their foregone wages.

The *individual mandate* provides the same result with a clearer pathway, because its costs fall on the beneficiary who pays them directly. Pauly (1994b, 1997) describes an individual mandate, enforced by employers and subsidized for the poor, requiring all individuals to purchase a minimum health plan or better. He argues that this approach is desirable so that people can relate their taxes to what they are paying to obtain benefits.

In this scheme, individuals are required to purchase health insurance. They may in fact acquire it through their workplaces, or they may buy it explicitly in a market setting.

During the U.S. debate of 1993–1994, disputes arose frequently over the fraction that the employer pays as opposed to the fraction paid by the individual. The presumption in these arguments is that the chosen fraction reflects the burden. Economists, however, tend to agree that the fraction chosen does not matter. The discussion presented above (regarding the \$20 per hour wage) says nothing about fractions. The economic logic suggests that those who are least able to avoid a tax will bear its burden, irrespective of who writes the check. Some argue that it is a political necessity to overlook the economics, but others insist that an open public discussion of the genuine issues would improve the quality of national debate.

## Separation of Health Insurance from Employment

Those seeking to redesign a health system can make a good argument for revising or replacing the prevailing system of employer-provided insurance with either a single-payer system or an individual mandate. The advantages of employer provision stem from long-term practices that cause economic distortions. During World War II, the U.S. government froze prices and wages. Competing for workers, firms expanded their fringe benefits, which were not subject to the freeze. After World War II, employer contributions to health insurance were, and continue to be, tax-exempt, providing workers with a substantial discount and inviting inefficiencies of over-insurance and moral hazard.

Meanwhile, many of the unemployed, as well as many low-wage employed, have gone without health insurance.

Health insurance problems also occur when workers change jobs. When leaving their previous employer's health coverage behind them, workers have little choice but to buy an individual policy, a "continuation of benefits" or COBRA, from the previous employer, or do without insurance entirely. Individual policies are often more expensive, sometimes pose administrative problems, and sometimes comprise a lower financial priority for people out of work. Workers often find pre-existing conditions such as heart disease to be uninsurable.

### **Single Payer versus Multiple Insurers**

A move in the United States toward universal coverage also entails the option of a single insurer, presumably the federal government. In the United States, multiple private companies insure a majority of the population. The most prominent single-payer proposal discussed in the U.S. debates has been the Canadian Medicare system.

Economic theory suggests that consumers value variety. Numerous restaurants serve different foods, prepared in different ways. American auto manufacturer Henry Ford said (at least apocryphally) that one could have any color of his pioneering Model T, so long as it was black—his company lost its market prominence to General Motors who provided a wider variety of cars (and colors). In principle, a variety of insurers may provide different types of coverage, pool different groups of people, and provide products that more closely match the variety of consumers.

However, the multiple-insurer system, as it has evolved, has led to multiple forms and policy rules that face hospitals, clinics, and nursing homes. Patients, as insurance clients, must often provide the same information numerous times, with commensurate possibilities of error. With hundreds of different health insurers, the difficulty of coordinating different policies falls on hospitals and physicians. It is external to the insurance companies and as a result, they do not see the need to reduce it. Nevertheless, it is a real economic cost both to patients and to providers, and the government as single payer could reduce it with fewer and standardized forms.

In principle, a consolidation of insurers could reduce such administrative costs if there are economies of scale in administration, or if gains could obtain from pooling those insured. Many economists have tried to estimate the excess administrative costs. Cutler and Ly (2011) partition the \$1,589 difference in per capita health care spending between the United States and Canada in 2002. Higher administrative costs accounted for \$616, or 39 percent, of the difference. The authors argue that this figure probably underestimates the amount and share, because nurses also spend substantial time on administrative tasks, but accounts typically consider nursing time as clinical care rather than administration.

One must be cautious in assuming that a government single-payer system would solve all of the administrative cost problems. The same administrative technology is available to the private sector, and if further economies were possible, and there is appropriate non-monopoly competition, private firms would merge to take advantage of the economies. In addition, the profits that private insurers gain are not a waste to the economy, but rather payments for capital that government also must incur.

Moreover, a switch to a single-payer system would greatly diminish, if not eliminate, the very large private health insurance industry. To put the issue in perspective, Cogan, Hubbard, and Kessler (2010) note that of the approximately \$2.4 trillion in 2008 U.S. national health expenditures, \$830 billion were private insurance premiums. Over one in three dollars of health care expenditures went through private insurance! Private insurers would almost certainly oppose a single payer plan, and they have actively promoted their own cause in the formulation of PPACA.

Do health care system problems warrant a change to a single payer? Other reforms may address specific problems. For example, we insure the uninsured through mandated coverage including subsidies for the poor, and provide coverage for people with pre-existing conditions through the individual mandates. Before PPACA passed, beneficiaries used COBRA legislation to address this issue.



A potential benefit of the single-payer system lies with the possibility of common coverage. We may worry now that some insured people have inadequate policies in terms of the depth and breadth of coverage. The single payer could offer one policy or a small number of variations, each variation determined to be adequate by policy makers representing the public. In contrast, the availability of many policies from many companies offers variety, tailoring policies to the individual preferences for cost-sharing features and coverage.

## COMPETITIVE STRATEGIES

Regardless of any changes in the direction of U.S. health care policy, governments will remain heavily involved in the financing and delivery of health care. The ideological battle is over the superiority of (1) increased government involvement through both expanded regulation and additional government programs to provide or finance health care or (2) an increased emphasis on market mechanisms and market forces with corresponding decreases in the use of regulatory instruments.

This controversy between competition and regulation is not new to the academic and policy-making communities. Those who deny the applicability of the competitive framework will most likely favor regulatory strategies. Proponents of further regulation tend to argue that attempts to promote partial forms of competition will not readily correct information imperfections, flawed agency relationships, and other distortions. Regulation proponents also react to the rapid growth in overall health spending and in spending on Medicare, Medicaid, and other public programs. The spending record in the United States, as compared to other more highly regulated systems, has reinforced beliefs among many that health care markets do not work like most other markets. Many believe that increases in supplies of hospital beds and physicians increase utilization of marginally necessary or even unnecessary services. There are widespread views that physicians can and do create demand and that empty beds in hospitals will be filled through physicians' discretionary decisions to admit patients.

Those who share these views advocate controls limiting hospital capacity and physician supplies, and increasing the monitoring of services provided. We have outlined the general forms of these controls. What is competition in health care? Pauly (1988a) defines competition through the conditions of "free entry and potential entrants willing[ness] to offer goods and services to consumers, with no one firm large enough to have an important influence on levels and quality" (p. 35).

From this definition, in a competitive health care industry we would expect (1) price-elastic demands facing providers; (2) a tendency toward marginal cost pricing; (3) efficiency in production—allocative as well as technical efficiency; and (4) the availability of alternatives in terms of price, quality, and form of delivery.

Others see competition in a more limited way, largely through the form of government policy rather than the structure or performance of health care markets. A competitive health care policy is one that relies primarily on financial incentives rather than controls to achieve goals. Those supporting this approach believe that market participants respond to changes in prices in a predictable and substantial way. Supporters of competitive approaches also argue that even imperfections in their strategies are preferable to the distortions caused by imperfect regulation.

Under either view of competition, certain forms of government intervention such as CON laws, utilization review, and mandated benefits, are anticompetitive because they compel. To be more specific in the distinction between competition and regulation, consider the uninsured. From the competitive point of view, many of the uninsured have insufficient income to purchase insurance. Mandated benefits and other requirements that drive up insurance prices aggravate their dilemma. A competitive solution is to subsidize purchases of insurance for the lower-income uninsured through tax credits, and to deregulate insurers so that lower-priced options become available. In contrast, they view programs that would provide insurance for the uninsured as lacking in incentives for efficient consumer search. Bearing in mind these distinctions, we turn to several broad competitive strategies for reforming the U.S. health care system.

## Development of Alternative Delivery Systems

The dominant competitive strategy, which evolved in the 1970s, has been the promotion of delivery systems that can provide an alternative to traditional fee-for-service with its comprehensive first-dollar insurance coverage. The cornerstone of this strategy has been the promotion of health maintenance organizations (HMOs), preferred provider organizations (PPOs), and other forms of managed care.

As described in Chapter 12, HMOs and other managed care health plans provide comprehensive sets of services for fixed monthly premiums. Typically, there is minimal cost-sharing for covered services. In exchange for expanded coverage, a patient's choice is restricted to specific providers and hospitals. Furthermore, the patient's primary care physician serves as the gatekeeper in that this physician's approval is required for referral to most specialists and nonemergency hospital admissions.

Expansion of the population covered by managed care has been a major policy goal. The primary motive behind the managed care strategy is the view among many policymakers that the traditional form of health care delivery was the primary culprit in rising costs and unnecessary care.

Managed care has reduced costs by lowering the rate of hospitalization and lower payments to providers. However, the rates of managed care cost increases have paralleled the rates of increase elsewhere, limiting the overall potential for cost containment. Nevertheless, federal policy continues to emphasize the managed care strategy by proposing expanded incentives for those who select prepaid, managed care systems, especially Medicare and Medicaid beneficiaries.

## Consumer-Directed Health Plans and Health Savings Accounts

Consumer-directed health plans (CDHPs) paired with health savings accounts (HSAs) represent important new health care delivery alternatives. Does the consumer-directed approach represent a novel strategy that could revolutionize health care delivery, or instead just a passing fad? To understand current developments, we consider earlier proposals for medical savings accounts (Stano, 1981; Pauly, 1994b; Pauly and Goodman, 1995). Under these proposals, employers or public payers would contribute to an individual's *Medical Savings Account* (MSA). The account would allow the holder to purchase relatively low-cost catastrophic insurance with high deductibles. Holders would then use MSA balances to pay out-of-pocket costs due to the deductibles, while providing true catastrophic insurance for large unexpected charges. The account would then distribute the unused portion in the MSA at the end of a designated period or at retirement.

MSA advocates contrast it to comprehensive, tax-subsidized insurance, which creates substantial moral hazard and ineffective incentives for efficient consumption of care. They argue that potential distributions from an MSA, like spending their own dollars, provide individuals with incentives to use care prudently. In principle, patients will be less likely to consume unnecessary or marginally beneficial care, and the stronger market forces will help restrain prices.

After years of debate, the 1996 HIPAA legislation included a provision for a demonstration with up to 750,000 MSAs for the self-employed and small firms with 50 or fewer employees. Individuals would choose plans with high deductibles and other cost-sharing features and they could put a certain percentage of the deductible into a tax-free savings account to pay for out-of-pocket expenses or future health care needs including long-term care. Those individuals under age 65 who withdrew MSA funds for nonmedical purposes, would face taxes and financial penalties. While some analysts projected considerable savings from more efficient levels of utilization, few would-be insurees adopted MSAs relative to the 750,000 authorized, and the MSA demonstration is no longer accepting new enrollees.

Several significant developments led to renewed interest in the MSA approach. In 2002, the Internal Revenue Service authorized the Health Reimbursement Arrangement (HRA)—a restrictive scheme that is completely *employer* controlled. More significant is the HSA legislation passed in 2003 as part of the Medicare prescription drug act. The HSA is a less restrictive MSA, owned by the *employee*, and open to anyone enrolled in a high-deductible health plan and not already covered by

public or private insurance. In 2011, the deductible had to be between \$1,200 and \$5,950 (\$2,400 and \$11,900 for a family plan). The plan allowed individuals with qualified coverage to contribute up to \$3,650 (\$6,150 for families) to their HSAs.

The CDHP strategy, an approach supported by President George W. Bush, was a driving force behind these legislative expansions. In most cases, a CDHP features a high-deductible health plan combined with an HRA or HSA. The motive for the CDHP is the desire to create highly informed consumers and to give them the incentives and the tools so that they take charge of their health care decisions. Their search for price and quality would counter the power of medical providers and the inefficiencies in the current marketplace. In this sense, the CDHP contrasts sharply with managed care under which the patient is a more passive participant and where the managed care plan administrators take responsibility for prices and quality. There were nearly six million enrollees in CDHPs in 2009 and an additional 17 million in other high-deductible plans. In 2007, Medicare also introduced an MSA option.

Evidence on CDHPs and HSAs is still incomplete. Feldman and colleagues (2007) analyzed a three-year window for plans offered by a large employer and find little significant savings for those enrolled in CDHPs. Dixon et al. (2008) examined a large manufacturing company that had large- and low-deductible CDHP options, with premiums lower in the former, as well as a PPO plan. Enrollees in the high-deductible CDHP were more likely to cut back on utilization, but they were more likely to engage in risky cost-saving behavior (e.g., not going to a physician when they should have, or taking less than the recommended dose of a prescription drug). Another report for the same firm found that the high-deductible CDHP enrollees were substantially more likely to discontinue taking some categories of drugs used to treat chronic conditions (Greene et al., 2008a).

Analysts temper their enthusiasm for CDHPs by recognizing the undesired consequences that may result from a system that depends on voluntary enrollments. Healthier individuals will more likely choose high-deductible health plans. They may purchase catastrophic coverage at relatively low rates and will more likely have funds left over in their HSA accounts. This selection phenomenon might also lead to escalating premiums for the sicker populations who remain in managed care and conventional plans so that the net effect could turn out to be largely a redistribution of income toward the healthy. Clancy and Gauthier (2004) provide an excellent collection of articles on CDHPs that includes discussions of this form of adverse selection.

Three other potentially serious problems affect HSAs and high-deductible policies. First, patients in these plans may be tempted to scrimp on preventive health care measures, some of which are often among the most cost-effective. Second, a small proportion of individuals with serious chronic and acute conditions account for a large share of annual health care spending. These patients will have exceeded their maximum out-of-pocket requirements and may not have a strong incentive to economize on their use of health care. Third, HSAs are more difficult to administer, and less sophisticated consumers could have difficulty distinguishing between HSAs and other options (Greene et al., 2008b). Despite these concerns, Cogan, Hubbard, and Kessler (2005) develop a well-crafted defense of HSAs along with other market-based reforms, and Cannon (2008) makes a strong conceptual case for “large HSAs” with the full amount of employer and employee contributions put into an HSA.<sup>2</sup>

### Other Market Reforms

Two other reforms are important to proponents of market-based solutions. (Box 23-1 provides a novel international “free-trade” idea that is generally not part of competitive proposals.) The first deals with the tax subsidy of employer-provided insurance. We have already described the employee gains from

---

<sup>2</sup> Other countries have introduced similar accounts called Medical Savings Accounts or MSAs. Singapore introduced MSAs in 1984 and several other nations, most notably China, have adopted MSA options. Hurley et al. (2008) simulate the effects of a publicly funded MSA system for Ontario, the most populous Canadian province. Their simulations indicate some cost savings but also adverse distributional effects on public spending and out-of-pocket costs.

**BOX 23-1****Could We Benefit from Free Trade in Health Care?**

Medical tourism is a growing industry. The media has abundant stories (horror stories in some cases) of Americans who travel to India and other low-cost countries for certain elective procedures, such as plastic surgery. Some patients in Canada and other countries with wait lists for procedures such as hip replacements are also traveling abroad to low-cost centers. The cost differences—\$45,000 to \$50,000 for hip replacement in the United States compared to \$12,000 in Singapore—are even encouraging American firms to consider expanding their health plans to include some medical tourism.

An op-ed piece in the *Wall Street Journal* by Bhagwati and Madan describes the potential of free trade in health care services and labor. They note four kinds of transactions that could liberalize trade:

1. Services that do not require physical proximity (e.g., claims processing and telemedicine);
2. Services provided to U.S. patients who travel abroad and vice versa;
3. Flows of capital through the establishment of U.S. hospitals and other medical facilities abroad, and vice versa; and
4. Flows of health care professionals between countries.

There are clearly many impediments to completely free trade in each of these categories. These include licensure and other regulations, as well as moral questions regarding importing more physicians and other skilled labor from poor or needy countries. However, Bhagwati and Madan argue that an expansion of trade can annually save the United States billions of dollars in health care costs, and they urge political leaders to consider more seriously the potential of such trade in their health reform proposals.

*Sources:* Based on Jagdish Bhagwati and Sandan Madan, “We Need Free Trade in Health Care,” *Wall Street Journal*, May 27, 2008, p. A19; and McQueen, M. P., “Paying Workers to Go Abroad for Health Care,” *Wall Street Journal*, September 30, 2008, p. B1.

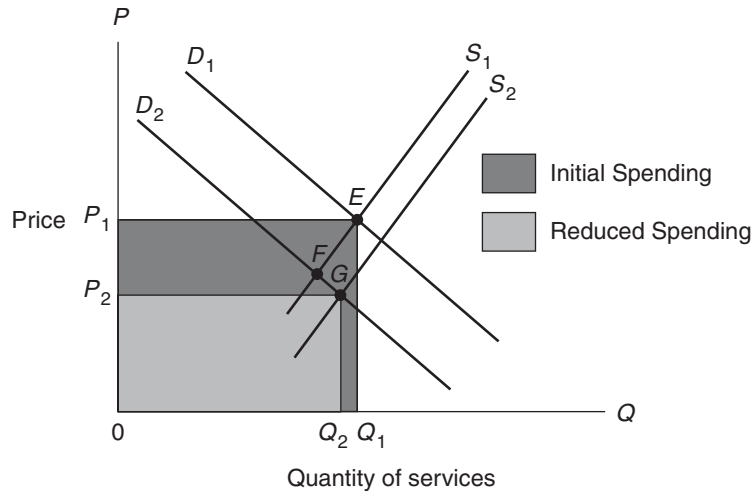
the tax-free income associated with such insurance, and the bias it creates toward deep coverage and the associated increases in utilization. Because proposals to eliminate the tax preference would meet considerable opposition, others have argued for full deductibility from taxable income of individual expenditures on health care and health insurance (Cogan, Hubbard, and Kessler, 2005).

Another reform under the competitive approach would eliminate many mandated benefits to increase the availability of lower-priced insurance policies. We described some of the almost 2,200 mandates that the states have adopted in Chapter 11. Mandates raise insurance premiums and reduce the options available to consumers, especially low-cost policies. To get around the costs imposed by mandates and other state regulations, a competitive strategy would allow individuals to purchase insurance across state lines, now generally prohibited. Parente and Feldman (2008) estimate the reductions in the number of persons without insurance in three scenarios: a national insurance market (which would have the greatest impact), one with competition among states grouped into four regions, and one with competition among the five largest states. Their “moderate” projection for a national market indicates an increase in the number of insured of more than 12 million, if Congress were to remove the interstate insurance barrier.

### Representation of the Competitive Approach

We illustrate the essence of the competitive approach with the help of Figure 23-2. Let  $S_1$  and  $D_1$  represent the existing demand and supply curves for health care. Equilibrium quantity is  $Q_1$ , and total spending is rectangle  $OP_1EQ_1$ .

Competitive strategies have two broad goals. The first is to reduce demand by increasing the number of patients in HSAs and other settings who are sensitive to price and the diminishing marginal benefit associated with health care. Neutralizing the tax subsidy for employer-provided health insurance would decrease demand for health services, especially the relatively less-beneficial



**FIGURE 23-2** The Intended Effects of Competitive Strategies on Demand and Supply

services to  $D_2$ . The equilibrium quantity will decrease, as will prices and expenditures, though the price effect will be small where the elasticity of supply is large.

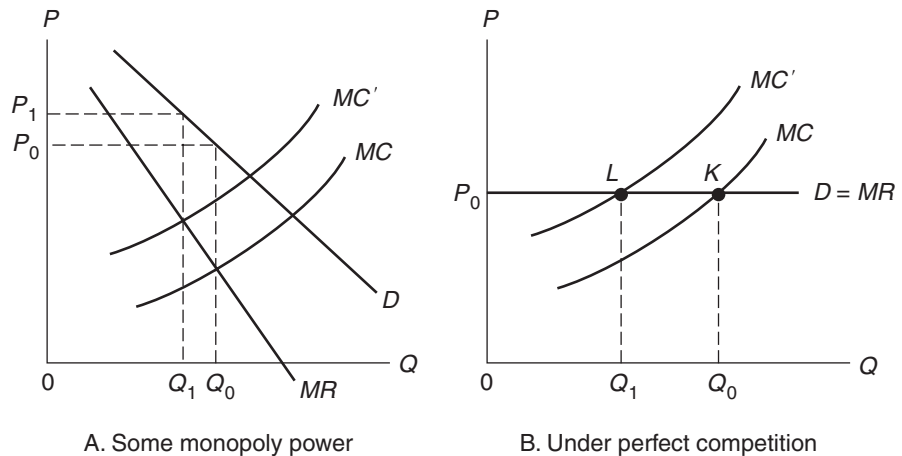
However, another important role for competitive strategies operates through effects on the supply side. Here, a relaxation of regulatory, entry, and capacity controls will reduce producer costs and increase the supply of services. At the same time, competitive pressure introduced through consumer search will push providers to produce care more efficiently, that is, at lower cost, represented by a rightward shift in supply to  $S_2$ . The combined effects, if substantial, would lead to large decreases (from  $0P_1EQ_1$  to  $0P_2GQ_2$ ) in health care spending, as illustrated in Figure 23-2, or to reduced growth rates in spending. Of course, the predicted effects must materialize. Otherwise, the strategies would be ineffective and might even backfire. For instance, if supply does increase but prices remain rigid and large amounts of supplier-induced demand (SID) follow, then both demand and supply may increase, thereby aggravating the utilization and spending problems.

## HEALTH SYSTEM REFORM AND INTERNATIONAL COMPETITIVENESS

Health system reform is increasingly important to business. Many business leaders view the United States at a competitive disadvantage compared to countries with social insurance programs. Would employer-based universal health insurance make the United States uncompetitive internationally? Conversely, would a government-funded single-payer system make American industry more competitive? Industry executives often respond as did Red Polling, then chair of Ford Motor Company, at a 1992 staff meeting of President-elect Clinton:

... health care costs give Japanese automakers an advantage over Ford of \$500 per car. Ford spends as much on health care as it does on steel. Health care providers are our biggest suppliers. (Pauly, 1997, p. 28)

Economists question this analysis because: (1) Health insurance is part of the total labor compensation package, and (2) the incidence of the implied tax falls primarily on the worker. Workers demand health insurance as well as money as part of their compensation. When they bargain for it, their bargaining indicates that they feel buying health insurance as a tax-preferred fringe benefit is worth the inevitable reduction in wages. More health care fringe benefits, or conversely, a reduced level of fringe benefits if government funds health care, do not change a firm's costs and its international competitiveness.



**FIGURE 23-3** The Effect of a Rise in Marginal Cost on the Price of Cars

To analyze further the fundamental issue, consider a situation where health insurance costs increase or employer-based universal coverage is mandated. The adjustments will come to wages, profits, or prices (e.g., cars). The degree to which it falls on reduced wages in theory depends on the elasticities of the labor supply and demand. We will discuss that side shortly, but first suppose that the costs fell entirely on the employer. To consider the ability of the firm to pass along this cost to the buyer, panels A and B of Figure 23-3 depict the demand, marginal revenue, and marginal cost curves for the firm under the alternative assumptions that A, it has some monopoly power, and B, it faces considerable competition.

In panel A, the demand curve slopes downward and the marginal revenue curve lies below the demand curve. If health costs cause the firm's wage bill to increase, this will affect the marginal cost of producing cars, shifting curve  $MC$  upward to  $MC'$ . In response, the equilibrium price will increase from  $P_0$  to  $P_1$ . In this sense, the car buyers will pay part of the increase in marginal costs. However, if that firm faces substantial competition from other U.S. car companies and from abroad, its demand curve is flatter. The more elastic demand curve implies that the firm will lose substantial business if it raises its price. In this case, the demand curve and marginal revenue are close together. The limiting case in panel B shows them to be identical. In panel B, as marginal cost increases, the car price does not increase. The equilibria to compare are points  $K$  and  $L$ .

These analyses indicate that the ability of the firm to pass its health costs on to the consumer depends on its elasticity of demand. If the increased car price includes hundreds of dollars of increased health costs, it is only because the firm had the power to pass these increased costs on to the consumers. The more vulnerable American firms are to both domestic and international competition, the less ability they have to pass extra costs along to consumers. If not consumers, then who pays for them?

As we discussed above, to the extent to which the laborers value a health insurance fringe benefit (at its cost in this case), the labor supply shifts downward, with the entire cost of the benefit falling on the laborers. If the labor supply does not change in response to falling wages, or is inelastic and geometrically vertical, as is likely in the aggregate, then the incidence will inevitably and entirely fall on the workers. This analysis implies that health insurance costs or employer-mandated health benefits become a national tax borne by workers, which will have little effect on the competitiveness of our products internationally. For firms that have so-called legacy costs through promises to retirees who are no longer working, the burden will fall on the retirees, in the form of reduced benefits, and on the firms' shareholders, in the form of reduced share prices and dividends. The U.S. auto industry spent the first decade of the twenty-first century coming to grips with these problems,

and in 2008–2009, saw both General Motors and Chrysler Corporation undergo massive reorganization to address them.

## QUALITY OF CARE

Improving the quality of care has become a vital component of health system reform. Both major party presidential candidates in 2008 stressed the need for higher health care quality. Senator John McCain vowed “to make sure they [patients] get the high-quality coverage they need.” Then-Senator Barack Obama devoted an entire section of his proposal to the goal of “ensuring providers deliver quality care.” Specifics included promoting patient safety and rewarding providers for outcomes rather than for volume. Returning to Figure 23-1, lower quality puts the economy inside health production possibility curve  $PP'$ .

Previous chapters have described clinical studies that detail major gaps and unevenness in quality, and mechanisms such as health care report cards to encourage quality improvements by embarrassing providers who perform poorly and by influencing patient choices. We have also described the use of pay-for-performance (P4P) incentives to raise quality. Medicare has joined this effort. On October 1, 2008, Medicare stopped reimbursing hospitals for eight preventable conditions if they are not present on admission. These conditions included pressure ulcers, catheter-associated urinary tract infections, and objects left in patients following surgery. These conditions obviously reflect lower quality care. The number of cases and costs can be surprisingly high for some of these categories. For example, the 323,000 Medicare cases involving pressure ulcers in 2006, with an average payment of \$40,400 per case, represented total (low-quality) costs of over \$13 billion (Rosenthal, 2007).

Health economists have not yet fully grasped quality issues. They express concern about:

1. Moral hazard and the overutilization associated with insurance (a theme we have stressed throughout the text)
2. Applications of cost-effectiveness analyses to distinguish economically efficient from inefficient procedures, technology and levels of care (see especially Chapter 4 and Box 23-2)
3. The greater use of financial incentives

Giving greater priority to financial incentives is a theme that we have also stressed throughout this text. Health economist David Cutler (2004) summarizes their role.

Medical care providers are paid on the basis of what they do. When they provide intensive services, they are paid well. Providing less intensive services is reimbursed much less well. As a result, medical professionals tend to perform too many intensive surgeries, while failing to do important routine monitoring and follow-up. Reimbursing sophisticated care is important, but paying for it is not a guarantee of better health. The way to get the system to focus on health is to make payments dependent as much on the effectiveness of the services provided as on the quantity and sophistication of services. I propose a system in which doctors would be paid more for meeting care guidelines such as appropriate use of mammograms, cholesterol screenings and flu shots; in which hospitals with better surgical outcomes would earn even more; and in which insurers that took care of people would be better off financially. Good care would earn more than poor care. (p. xiii)

The transition from concept to practice is often very challenging. While the success or failure in healing a broken bone is relatively easy to monitor, quality aspects for many conditions are much more difficult to define and monitor, particularly for chronic conditions related to mental health and substance abuse. Paying for performance requires sophisticated definition of performance measures, and determination of the appropriate incentive amounts needed to influence provider behavior.

**BOX 23-2****Preventive Care and Cost-Effectiveness Analyses**

A greater emphasis on preventive health has become a mantra for many political and thought leaders. Diabetes screening for type 2 (adult-onset diabetes) is one prominent example. The health-related consequences that arise from this disease are staggering. Are widespread screening efforts for this disease cost-effective? Cohen and colleagues (2008) urge caution against sweeping generalization regarding preventive care.

Consider just the following examples of preventive measures they provide (taken from the Tufts–New England Medical Center Cost-Effectiveness Registry). The incremental cost per quality-adjusted life-year (QALY) is in 2006 dollars.

High-intensity smoking relapse program (compared to low-intensity program)	\$ 190
Intensive tobacco use prevention program for 7th and 8th graders	23,000
Screening all 65-year-olds for diabetes (compared to diabetes screening of all 65-year-olds who have hypertension)	590,000

By almost any standard (typically around \$100,000 per QALY), the first two prevention programs meet the threshold for adoption. The third, unlimited screening of all 65-year-olds for diabetes, should unequivocally be rejected.

There are two important lessons here. First, rational decisions cannot be made without reliable cost-effectiveness values and, second, policymakers and third-party payers must discriminate carefully within preventive (and undoubtedly other) categories of health care interventions. Political messages that sound good can reflect bad economics.

P4P is now common in both private and public insurance plans, with mixed evaluations. Pearson et al. (2008) evaluated P4P programs introduced by Massachusetts' leading commercial insurers. The research covered a wide variety of P4P contacts with a large number of physician groups over the period 2001–2003. It showed that the quality improvement, represented by 13 HEDIS measures, for highly incentivized groups was not better than the improvement found in comparison groups.<sup>3</sup>

In contrast, the Tax Relief and Healthcare Act of 2006 mandated a P4P program for Medicare. The program, known as the Physician Quality Reporting System, is still voluntary, but the CMS paid an average bonus of nearly \$19,000 per participating professional practice in 2009. CMS reported an average increase of 10.6 percent over 2008 across 99 measures of performance.

**THE PATIENT PROTECTION AND AFFORDABLE CARE ACT (PPACA) OF 2010**

The U.S. Congress passed the Patient Protection and Affordable Care Act (PPACA) in March 2010. The legislation is complex, with many features scheduled for phase-in through 2018, but PPACA seeks principally to reduce the number of uninsured individuals by focusing on two key areas—the problem of affordability in the individual insurance market due to inadequate incomes or pre-existing conditions and the lack of coverage for employees of small businesses. The most significant provisions of PPACA addressing these two areas will:<sup>4</sup>

- Require most U.S. citizens and legal residents to have health insurance, the *individual mandate*. Families without coverage will face penalties up to \$2,085 per year or 2.5 percent of household income. Individuals and families with incomes as high as 400 percent of the federal poverty level (FPL) will be eligible for refundable credits or an expanded Medicaid program.

<sup>3</sup> The Healthcare Effectiveness Data and Information Set (HEDIS) is a tool, developed by the National Committee on Quality Assurance, used by more than 90 percent of America's health plans to measure performance on important dimensions of care and service.

<sup>4</sup> See the Kaiser Family Foundation (2011) for a comprehensive overview of PPACA.



- Penalize employers with 50 or more full-time employees that do not offer coverage at a fee of \$2,000 per full-time employee. Small businesses that provide health insurance for their employees will be eligible for tax credits. (Effective January 1, 2014).
- Expand Medicaid to all non-Medicare eligible individuals under age 65 (children, pregnant women, parents, and adults without dependent children) with incomes up to 133% of the federal poverty level FPL with a benchmark benefit package.
- Establish state-based health insurance exchanges, where individuals and small businesses can compare policies and buy coverage, administered by a governmental agency, or a non-profit entity. The Act eliminates underwriting policies that enable insurers to deny individuals insurance or charge them higher premiums based on pre-existing conditions.
- Establish a uniform set of benefits, called Essential Health Benefits, with 10 major areas of coverage including prescription drugs and preventive services. All qualified health plans, except employer-sponsored plans and certain individual policies, will be required to offer at least the Essential Health Benefits. (Effective January 1, 2014)
- Eliminate cost-sharing for Medicare-covered preventive services recommended by the U.S. Preventive Services Task Force, and waive the Medicare deductible for colorectal cancer screening tests. (Effective January 1, 2011)

Despite its focus on the uninsured, PPACA affects almost every segment of the health economy. For example, numerous provisions affect Medicare beneficiaries as well as those who provide services to them. In previous chapters, we have described reductions in payments to Medicare Advantage Plans (Medicare Part C) and the gradual elimination of the “doughnut hole” by 2020. Other provisions call for reductions in payments to hospitals and other providers.

To raise revenues and discourage over-insurance, the Act imposes a “Cadillac” tax on high-cost employer-provided policies. In addition to this tax, funding for the program comes mainly from an expanded Medicare tax base that will affect higher income individuals and families, fees on health insurers, and taxes on manufacturers and importers of branded drugs and certain medical devices. These new revenue streams will nevertheless fall short of the increased federal obligations under PPACA.

To close the gap, the Obama administration and supporters of PPACA have emphasized measures to “bend the curve,” that is, slow down the overall rate of growth of health care spending through increased preventive health, administrative simplification, and by reducing inappropriate care. To accomplish the latter, PPACA will create an independent, non-profit Patient-Centered Outcomes Research Institute. Although the authority of the Institute in making recommendations is limited, it will examine and conduct studies comparing the appropriateness of alternative treatments.

A key feature in PPACA is the establishment of state-level “exchanges” which one might think of as an Orbitz®, a popular travel services site, for health care, in which agents (individuals, but most likely employers) can shop for packages of health insurance.<sup>5</sup> This will start January 1, 2014, with hopes of being financially self-sustaining by January 1, 2015. The exchanges will be based on the Massachusetts *Connector* mechanism in Massachusetts’s subsidized health insurance program established in 2006 (Box 23-3 provides further information on the Massachusetts Plan). Under PPACA, those with incomes up to 400% of the U.S. poverty level will see their health insurance subsidized at a decreasing rate using tax credits. This will promote participation by the private insurance industry and consumer choice among plans.

How well does PPACA address the goals of reform discussed earlier in this chapter?

---

<sup>5</sup> States that choose not to create their own exchanges can have the federal government come in and establish it for them.

**BOX 23-3****The Massachusetts Plan**

In 2006 Massachusetts under then-Governor Mitt Romney passed a State Health Care Plan with the following features:

*Individual Responsibility*—Applies to all adults ages 17 or higher, with “penalties” for non-purchase of health insurance—an individual mandate

*Employer Responsibility*—Employers with 11 or more employees must demonstrate a “fair and reasonable” contribution towards employee coverage or face penalties

*Government Subsidies for Low Income Residents*—(1) Expansion of Medicaid (known as “MassHealth”) for children up to 300 percent of the federal poverty level (FPL) and (2) Creation of subsidized insurance (known as “Commonwealth Care”) for adults up to 300 percent FPL offered through the Health Connector (a virtual marketplace that helps match insurance buyers and sellers).

A 2011 evaluation from the Blue Cross-Blue Shield of Massachusetts Foundation presented some results.

- From 2006 to 2009, at a time that the U.S. “uninsured rate” rose from 15.8 to 16.7 percent, the Massachusetts rate fell from 6.4 percent to 2.7 percent (and to 1.9 percent in 2010).
- Fewer than 1.2 percent of tax filers who were subject to the mandate were assessed a penalty on their 2008 return
- There has been no evidence of subsidized coverage “crowding out” employer-sponsored insurance (ESI).
  - Employer offers of coverage have increased.
  - Take-up (participation by those who were previous uninsured) of employer-offered coverage has remained high.

Analysts recognize the dangers in extrapolating any “small” (state) program into a large (national) one. Massachusetts did not share the level of economic pain from the “Great Recession” as did the other parts of the country, and the state (in part due to the influence of the late Senator Edward Kennedy) had enjoyed considerable largesse in receiving federal health care funds. That said, to the extent that the PPACA includes similar features as the Massachusetts Plan, and with two-thirds of adults supporting the state’s reform plan, the Massachusetts experience would seem to provide positive “feedback” regarding potential success of PPACA (for a spirited debate between Douglas Holtz-Eakin and Jonathan Gruber on the lessons from Massachusetts for PPACA, see Joyce, 2010).

*Source:* [http://bluecrossfoundation.org/Health-Reform/~media/D0DDA3D667BE49D58539821F74C723C7.pdf](http://bluecrossfoundation.org/Health-Reform/~/media/D0DDA3D667BE49D58539821F74C723C7.pdf), accessed July 26, 2011.

*Creating a safety net*—The Congressional Budget Office (in a March 2010 document) noted that approximately 32 million people would gain coverage by 2019, leaving about 23 million uncovered. Table 23-1 indicates increases of 16 million through Medicaid and CHIP, and 24 million through the exchanges, with small decreases of those in employer and nongroup insurance. This would leave about 23 million non-elderly residents uninsured (about one-third of whom would be unauthorized immigrants).

*Cost containment*—In the 2010 *New England Journal of Medicine* Jonathan Gruber writes that the best course, is “to ‘cautiously’ pursue as many cost-containment approaches as possible and see which ones work.” These include:

- Reducing consumer demand through the so-called “Cadillac” insurance tax;
- Cutting provider payments by appointing a depoliticized board to make up-or-down recommendations to Congress on changes to Medicare’s provider payments;
- Running pilot programs to test various approaches to revamping provider-payment incentives and organizational structure;

**TABLE 23-1 Congressional Budget Offices Estimated Effects of the Insurance Coverage Provisions of the PPACA**

<b>EFFECTS ON INSURANCE COVERAGE</b> <b>(Millions of nonelderly people,</b> <b>by calendar year)</b>		<b>2010</b>	<b>2011</b>	<b>2012</b>	<b>2013</b>	<b>2014</b>	<b>2015</b>	<b>2016</b>	<b>2017</b>	<b>2018</b>	<b>2019</b>
Current Law Coverage	Medicaid & CHIP	40	39	39	38	35	34	35	35	35	35
	Employer	150	153	156	158	161	162	162	162	162	162
	Nongroup & Other	27	26	25	26	28	29	29	29	30	30
	Uninsured	50	51	51	51	51	51	52	53	53	54
	<b>TOTAL</b>	<b>267</b>	<b>269</b>	<b>271</b>	<b>273</b>	<b>274</b>	<b>276</b>	<b>277</b>	<b>279</b>	<b>281</b>	<b>282</b>
Change (+/-)	Medicaid & CHIP	*	-1	-2	-3	10	15	17	16	16	16
	Employer	*	3	3	3	4	1	-3	-3	-3	-3
	Nongroup & Other	*	*	*	*	-2	-3	-5	-5	-5	-5
	Exchanges	0	0	0	0	8	13	21	23	24	24
	Uninsured	*	*	-1	-1	-19	-25	-30	-31	-31	-32
Post-Policy Uninsured Population											
	Number of Nonelderly Uninsured	50	50	50	50	31	26	21	21	22	23
	Insured Share of the Nonelderly Population										
	Including All Residents	81%	82%	82%	82%	89%	91%	92%	92%	92%	92%
	Excluding Unauthorized Immigrants	83%	83%	83%	83%	91%	93%	95%	95%	95%	94%

Source: Elmendorf, "Estimates" (2010), Table 4

- Investing hundreds of millions of dollars in new comparative-effectiveness research; and
- Launching pilot programs to assess the impact of various re-organizations of the medical malpractice process.

*Choice for Patients and Providers*—In bypassing either single-payer programs or explicit mandated providers, PPACA will preserve both patient and provider choice. The insured would appear to have no less choice in insurance coverage, and payment provisions would seem to remain the same as previously.

*Ease in Administration*—The largest increases in the numbers insured (as noted in Table 23-1) would come through Medicaid/CHIP (an additional 16 million) and the new insurance exchanges (24 million more). Medicaid/CHIP is a well-established program, with state-level relationships already in place. States have not yet established the exchanges, however, and some considerable administrative costs will most likely accompany their creation and operation.

With respect to health care spending, PPACA's complexity, and lack of specificity regarding cost-reducing measures, make it difficult to predict changes in health care spending accurately. Analysts must also deal with a moving target as various provisions of PPACA step in. In an early forecast by analysts at CMS, Sisko et al. (2010) estimated that national health care spending as a share of GDP will be 0.3 percentage points higher (about \$88 billion) in 2019 than without PPACA.<sup>6</sup> Table 23-2 summarizes some of the projections. While the overall estimated impact is

<sup>6</sup> A subsequent analysis by CMS analysts (Keehan et al., 2011) also predicts a relatively small overall effect. The average growth rate in spending over the period 2010–2020 is projected to be 0.1 percent higher as a result of PPACA with the largest impact in 2014 as major features of the legislation kick in.

**TABLE 23-2 Estimated Fiscal Effects of PPACA**

<b>Expenditure Impacts of PPACA</b>								
<b>Source of funds</b>	<b>2009</b>	<b>2010</b>	<b>2011</b>	<b>2012</b>	<b>2013</b>	<b>2014</b>	<b>2019</b>	<b>Total over 10 years</b>
NHE (\$ billions) - PPACA	2,473.3	2,600.2	2,709.8	2,851.6	3,024.8	3,302.4	4,571.5	
NHE (\$ billions) - Otherwise	2,472.2	2,569.6	2,702.9	2,850.2	3,024.7	3,225.3	4,482.7	
Difference (\$ billions)	1.1	30.6	6.9	1.4	0.1	77.1	88.8	206.0
Private funds - PPACA	1,269.9	1,315.8	1,352.1	1,405.7	1,483.5	1,611.0	2,231.6	
Private funds - Otherwise	1,268.8	1,303.9	1,353.0	1,406.6	1,484.1	1,583.7	2,154.4	
Difference (\$ billions)	1.1	11.9	-0.9	-0.9	-0.6	27.3	77.2	115.1
Public funds - PPACA	1,203.4	1,284.4	1,357.7	1,445.9	1,541.2	1,691.4	2,339.9	
Public funds - Otherwise	1,203.4	1,265.7	1,349.8	1,443.6	1,540.7	1,641.6	2,328.3	
Difference (\$ billions)	0.0	18.7	7.9	2.3	0.5	49.8	11.6	90.8
Medicare - PPACA (\$ billions)	507.1	534.4	548.9	585.7	619.8	655.8	891.4	
Medicare - without PPACA (\$ billions)	507.1	514.7	544.4	585.7	626.8	672.8	977.8	
Difference (\$ billions)	0.0	19.7	4.5	0.0	-7.0	-17.0	-86.4	-86.2
Federal Medicaid/CHIP - PPACA (\$ billions)	255.8	285.4	265.2	286.7	310	383.5	541.4	
Federal Medicaid/CHIP - Otherwise (\$ billions)	255.8	284	262.4	281.4	300.8	318.5	451.5	
Difference (\$ billions)	0.0	1.4	2.8	5.3	9.2	65.0	89.9	173.6
Health share of GDP (%) - PPACA	17.3	17.5	17.4	17.2	17.3	17.9	19.6	
Health share of GDP (%) - Otherwise	17.3	17.3	17.3	17.2	17.3	17.4	19.3	

Source: Derived from Sisko et al. (2010).

relatively modest, effects on some programs may be substantial. Analysts expect Medicare to be \$86.4 smaller under PPACA in 2019 than under previous assumptions; they expect Medicaid/CHIP, in contrast, to be \$89.9 billion larger.

At the end of 2011 there is considerable opposition to the mandate, ostensibly because it forces some to purchase a good that they may not wish to purchase. Politics aside, in economic terms, an individual mandate to buy health insurance is no different from a tax to pay for a public good. If a governmental unit passes a tax by legitimate political means to pay for a park with tennis courts, a library addition, or a missile system, taxpayers must pay the tax whether they go to the park, play tennis, read books, or feel more secure due to better national defense.

## CONCLUSIONS

Cost-containment, and reduction or elimination of the number of uninsured, are the principal goals of health system reform in the United States. Other goals include administrative simplicity and choice for providers and patients. Improving the quality of care has also emerged as a national priority. Any reform process requires difficult decisions on the services covered, on who is covered, and on the financing mechanisms.

The most serious obstacle to reform (using the United States as our example) has been the divide over whether to expand the government's role through mandates, additional regulations, and tax subsidies or whether to rely increasingly on markets through deregulation and tax changes that neutralize the current bias toward subsidized, employer-based insurance.

We have examined other important issues to health system reform. We found that the incidence of health premiums under employer-based systems falls on workers. In theory, the burden of increases in health care costs or mandates on employers will fall on workers. Thus employer mandates do not make firms less competitive inter-nationally, nor will movements toward single-payer systems funded by government revenues make them more competitive.

New to the U.S. is the Patient Protection and Affordable Care Act (PPACA), passed in 2010. It is a long-term "fix" in that provisions will step in gradually until 2018. Built on the shoulders of America's private insurance and the Medicaid/CHIP systems, it formulates an individual mandate for consumers to purchase health insurance and provides market-pooling mechanisms to make the insurance available to many who were previously not able to get it. Its cost-containment and quality assessment mechanisms are less concrete than those that provide the insurance. Analysts and policy-makers will seek to address these mechanisms as PPACA phases in through 2018.

---

## Summary

1. There is a growing consensus in the United States on health system reform and the goals of reform. The reforms seek to contain costs and to reduce or eliminate the pool of uninsured.
2. Many difficult decisions in the reform process include determining which services to cover, who to cover, and how to pay the additional costs.
3. From society's viewpoint, the cost of universal coverage is the incremental cost of additional health care purchased by people due to improved insurance coverage.
4. The incremental cost of providing full-year coverage for all Americans would be about 5 percent of current national health care spending. Savings from successful cost controls may reduce this incremental cost.
5. Mechanisms to reduce the uninsured include employer and individual mandates, expansion of

existing public programs, and subsidized coverage for lower income and high-risk households.

6. A switch from the current U.S. health system to a single-payer system would likely save money by reducing administrative costs.
7. For the entire country, labor supply is inelastic. As a result, the incidence of a mandated health insurance program falls mainly on workers.
8. Mandates on employers have the same economic effects as mandates on workers.
9. Some proposals recommend the separation of health insurance from employment in order to eliminate the inefficiency caused by the tax-exempt status of employer-provided health insurance.
10. There is a fierce debate in the United States over whether reform should rely increasingly on an expanded role for government or on competitive strategies.
11. Competitive strategies include the promotion of alternative delivery systems, the expansion of consumer-directed health plans built around various forms of health savings accounts, reductions in mandated benefits and other regulations on the insurance industry, and implementing tax reforms that reduce the bias toward employer-based insurance.
12. The level of health care benefits provided does not generally affect the international competitiveness of U.S. firms.
13. PPACA seeks to reduce the number of uninsured individuals by improving the affordability of insurance, and by improving the availability of coverage for employees of small businesses, in the context of the existing U.S. health insurance industry.
14. In bypassing either single-payer programs or explicit mandated providers, PPACA preserves both patient and provider choice. Consumers maintain choice of insurance coverage, and provisions for provider payment are unchanged.
15. PPACA's cost-containment strategies are less well-developed than those that provide the insurance. In addition to seeking to reduce demand through a "Cadillac tax," it proposes numerous pilot programs addressed toward the revamping of provider-payment incentives and organizational structure.

---

## Discussion Questions

1. Would individual mandates for health insurance be more burdensome to the poor than employer mandates? Would lower-income groups be wise to favor one plan over the other?
2. If the aggregate labor supply curve were elastic instead of very inelastic, how would an employer-mandated health insurance plan affect the country's international competitiveness?
3. How could a single-insurer health insurance system provide cost savings over a multiple-insurer system like that in the United States?
4. What are the major competitive strategies proposed by economists? How, for example, would elimination of the tax subsidy for employer-provided health insurance reduce spending on health care?
5. What are Health Savings Accounts (HSAs)? Why might HSAs decrease health care spending? What are some problems with the HSA concept? Consider an enrollee in a CDHP with a high-deductible HSA who is choosing between the two physicians whose fees vary by \$30 per visit. How might the fee information influence her to choose the lower-priced physician? Why might the patient still choose the higher-priced one? Would your answer change if the difference was \$100 per visit?
6. Compare PPACA passed in 2010 with the Massachusetts health plan adopted in 2006. Use the Internet to obtain more details than we have provided in this chapter. Be sure to note the similarities and differences in the two plans.
7. Improving the quality of health care is becoming a national priority. Discuss the relative merits of using government regulation versus market forces in improving quality.
9. Assume that a brilliant health economist has just developed a plan that will greatly improve the efficiency of the U.S. health care system. Explain why it is highly unlikely that the plan, or anything close to it, would be adopted in the United States. Hint! Think of those who might be interested in preserving the status quo.

## Exercises

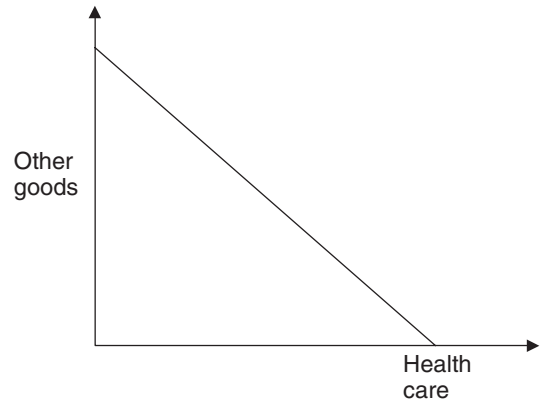
- Suppose that a firm faces a downward-sloping demand curve and offers no health insurance to its employees. Let an employer mandate for health insurance be enacted. If this causes the firm's marginal costs to increase, will the firm pay the full cost of the health insurance out of profits? Is the mandate likely to increase the firm's marginal costs?
- Suppose that a monopolist faced the following demand curve for its goods. Its marginal cost per unit of production is 50, and it faces no fixed costs

Price	Quantity
100	100
90	200
85	250
80	300
70	400
60	500
55	550
50	600
40	700
30	800
20	900
10	1,000
0	1,100

- Calculate the profit-maximizing output and price.
  - Suppose the workers negotiate a health insurance benefit increase that increases marginal cost per unit from 50 to 60. Calculate the new profit-maximizing output and price.
  - Who bears the costs of the benefit increase? Why?
- Use the demand-supply framework in Figure 23-2 to explain how increased cost sharing could lead to lower utilization and spending on health care.
  - Consider two households. They have the same incomes and face the same prices. Household *H* tends to be healthy and

household *U* tends to be unhealthy. Suppose that two insurance plans are available:

- \$2,500 deductible and a 5 percent coinsurance rate after meeting the deductible.
  - \$250 deductible and a 20 percent coinsurance rate after meeting the deductible.
- Using a budget constraint and indifference curves on the diagram below, model the two insurance plans.
  - Assume that a voluntary HSA is made available upon the purchase of a high-deductible policy. Assume that if the money is not used it is lost. Which of the households is likely to participate? Use the diagram below to explain why.
  - Consider part (b) above, but assume now that the unused portion in the HSA can be distributed to the individual at the end of a designated period or at retirement. Would your answer to part (b) change? If so, how? If not, why not?



- (Advanced) For students with computing and statistical skills, the MEPS database is available at [www.meps.ahrq.gov/mepsweb/](http://www.meps.ahrq.gov/mepsweb/). For an individual or class project, try to replicate, or improve, Hadley's estimates of the costs of universal insurance. Be attentive to key assumptions as to how much care those who are currently uninsured might purchase, under the various proposals.

# The Health Economics of Bads

- An Introduction to Bads
- Models of Addiction
- Rationales for Public Intervention
- Advertising Restrictions on Cigarettes and Alcohol
- Excise Taxes and Consumption of Cigarettes and Alcohol
- Conclusions

No discussion of the health economy can be complete without addressing one of the most prominent and controversial issues—the consumption of health bads. In a market economy, we usually regard the consumer’s sovereignty as given, that is, the consumer is “free to choose.” However, we can find many exceptions in any society—cases where society encourages or discourages desired choices. We encourage and promote use of motorcycle helmets, automobile emissions-control devices, old-age pensions, and good prenatal care. We often discourage the purchase of alcohol, street drugs, and cigarettes.

The reason to intervene in private decisions could be paternalistic, but it also could appeal to economic efficiency. Cigarette smoking affects not merely the cigarette buyer and seller—effects that are internal to the cigarette market—but also the health of nonsmokers nearby, an external cost. Since many health insurers do not distinguish between smokers and nonsmokers, perhaps because of high monitoring costs, nonsmokers may pay higher premiums than warranted by their actual health risks. The personal hazards of excessive alcohol consumption are also serious, including disability due to alcoholism and fatality due to liver disease. The external costs include possible harm to family or neighbors as well as the excessive, dangerous, and often-fatal traffic accidents due to drunken driving. Finally, when smokers and drinkers underestimate the probabilities of ill health due to their consumption, the imperfect information provides an efficiency rationale for measures, such as taxes, that tend to curb the behaviors.

Economists as citizens may choose one side or the other of these controversies because their values are not determined by their being economists, but economists as scientists can illuminate the issues that are of interest to the public. For many, the question is not whether to intervene in private decisions to smoke or drink, but how to do so more effectively and unobtrusively. This hinges on many issues although two economic questions are central. First, what is the relationship of price to demand? Second, what is the relationship of product advertising to total consumption?



## AN INTRODUCTION TO BADS

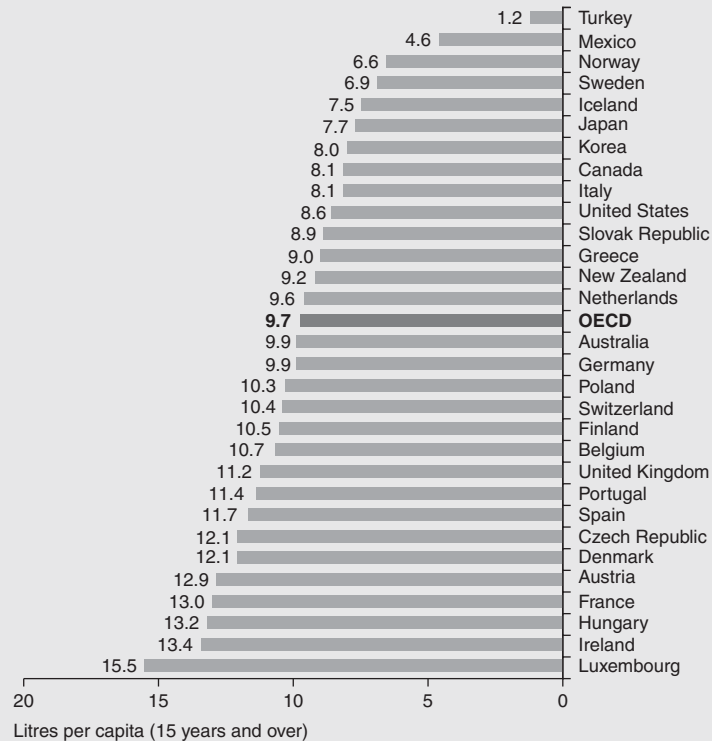
Economic reasoning is helpful for any approach to the analysis of bads, such as cigarette smoking and excessive alcohol consumption. Box 24-1 shows how consumption of alcohol and tobacco varies across countries. Economic models of addiction as well as market failure help determine whether intervention is justified on efficiency grounds. Alternatively, models of consumption, advertising, price, and taxation help provide relatively unobtrusive and cost-effective means to intervene, if we choose to do so.

### BOX 24-1

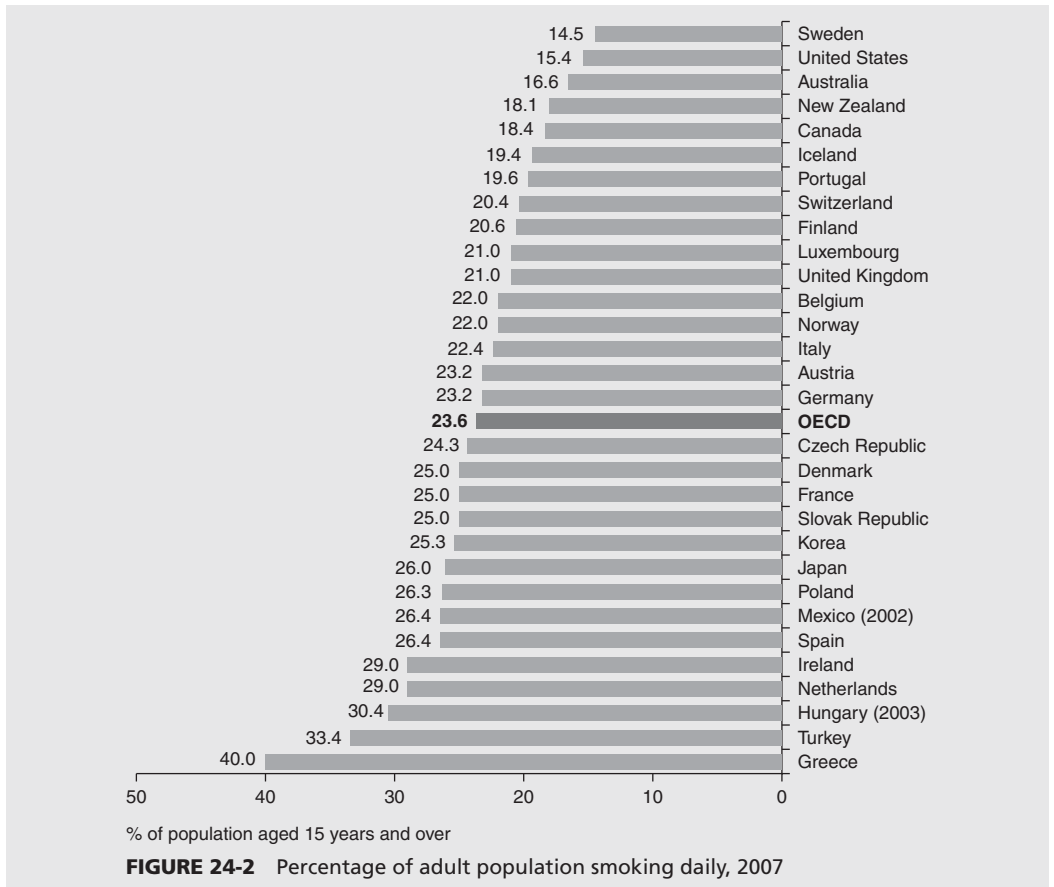
#### Who Smokes and Who Drinks? Cultures and Behaviors

It is common enough to find families who drink wine with dinner, or groups of people who smoke together. A comparison of the consumption across countries gives some perspective. It may surprise some Americans that U.S. consumption per capita of both “bads” is relatively low among the developed countries shown.

In Martin Cruz-Smith’s popular novel, *Gorky Park*, the Russian central character, Arkady, is asked, “Why didn’t you ever go to America?” Arkady answers, “If I went to America, I would have had to quit smoking.” Figures 24-1 and 24-2 show the wide variation in alcohol consumption and cigarette smoking across a range of OECD countries. Comparable Russian data are not available. (Data come from *OECD Health at a Glance 2009*.)



**FIGURE 24-1** Alcohol consumption, population aged 15 years and over, 2007



Cigarettes qualify as bads, as follows from the high death rates attributed to cigarette smoking. Deaths per capita from respiratory cancers rank highest among national death rates from malignancies and rank second only to categories of heart disease overall. Motor vehicle deaths are less than half this rate, and only a handful of U.S. states have HIV/AIDS death rates that exceed one-half of the respiratory cancer rate. Cigarette smoking affects other disease categories as well, such as emphysema and heart disease. There is little doubt that convincing successive generations of youth to choose never to smoke would lower average population age-adjusted mortality rates (Table 24-1) substantially. Econometric studies of health production commonly find cigarette consumption to be a significant and materially important predictor of mortality rates.<sup>1</sup>

In contrast to smoking, moderate alcohol consumption does not necessarily harm people and may (according to some studies) benefit some. Yet, substantial harm occurs with excessive consumption and inappropriate related behaviors, such as drunk driving. Applying a novel means to identify the effect of drinking on traffic fatalities, Levitt and Porter (2001) found that drivers with “alcohol in the blood” are eight times as likely to cause a fatal accident as the sober driver. With “blood alcohol above legally drunk,” the ratio rises to 15 times. Alcohol use among both high

<sup>1</sup> In addition, cigarettes are implicated in low birth weights (Rosenzweig and Wolpin, 1995).

**TABLE 24-1 Mortality Rates**

2009 US Mortality rates per 100,000	
Disease	Mortality Rate
Cardiac Disease	253.9
Respiratory Malignancies	51.5
Diabetes	22.3
Breast Cancer	13.4
Suicide	11.9
Motor Vehicle Accident	11.8
Prostate Cancer	9.2
HIV	3.1

Source: CDC *Vital Statistics*, 2009, [http://www.cdc.gov/nchs/data/nvsr/nvsr59/nvsr59\\_04.pdf](http://www.cdc.gov/nchs/data/nvsr/nvsr59/nvsr59_04.pdf), accessed November 17, 2011.

school and college students also raises policy concerns. College students who drink have poorer academic attainment (Cook and Moore, 1993; Williams, Powell, and Wechsler, 2003). High school drinkers are more likely to drop out before graduation (Chatterji and DeSimone, 2005).

One approach is to assume that society has no grounds to intervene if the consumer chooses rationally and voluntarily, understands the risks, and creates no side effects for others. This view implicitly rejects all arguments not based on economic efficiency, including those grouped together as equity concerns and typically rejects other grounds as paternalistic. However, because alcohol and cigarette consumption are addictive, the issues of rationality, volition, and information take on closer scrutiny.

## MODELS OF ADDICTION

Models of addiction come out of both psychology/medicine and economics. Chaloupka and Warner (1999) identify three types: imperfectly rational, myopic, and rational addiction.

### Imperfectly Rational Addiction Models

Imperfect rational addiction models propose that the addict has stable but inconsistent preferences in the short run as opposed to the long run. As Schelling (1978) described this person:

Everybody behaves like two people, one who wants clean lungs and long life and another who adores tobacco . . . The two are in a continual contest for control. (p. 290)

Are people really like this? Regret, in fact, occurs frequently. In film, there is even the cliché of the hero in comedy who enters the tiger's cage telling the bystanders to ignore him should he change his mind and even scream for help. Of course, he changes his mind. Is he two different people at those two moments?

### Myopic Addiction Models

Nearsightedness about the future harmful effects of the ingested drug provides a variant of the imperfectly rational model. Here, the individuals don't see the facts clearly; they are naïve about the nature of the drug and its side effects. One may imagine someone easily persuaded by cigarette-smoking friends who may denigrate the societal information about cigarettes. One may see the

teenager who only dimly perceives the realities of distant future health events, including cancers, and perhaps fails to connect the experience of smokers in older generations with his or her own behaviors and prospects.

### Rational Addiction

Can addiction in some cases be a rational choice? Becker and Murphy (1988) discarded myopic models and investigated addiction by assuming that people incorporate rationally all information, past, present, and future, into their utility calculations. They showed that rationally choosing to ingest an addictive drug was possible under restrictive yet plausible conditions. The discussion here follows work by Becker and Murphy (1988), Becker, Grossman, and Murphy (1991), and MacDonald (2004).

Addiction researchers usually speak of “reinforcement” and “tolerance.”

*Reinforcement* means that greater past consumption of addictive goods, such as drugs or cigarettes, increases the desire for present consumption. In short, smoking more now may make us smoke more later.

*Tolerance* occurs if the utility from a given amount of consumption is lower when past consumption is greater. This suggests that the future impacts of smoking or drinking or ingesting drugs decrease, when we consume more now. A single glass of wine may make someone tipsy the first time he or she drinks. Over time, with more drinking experience, the first glass of the evening may have little or no impact. We will use smoking to illustrate important model relationships, although drinking, illicit drugs, or even common substances such as caffeine, can provide similar examples.

Becker, Grossman, and Murphy introduce the construct of “addictive capital stock,”  $S$ . For example, with more smoking experience, the smoker’s attitude toward cigarette consumption is likely to change. We assume therefore that addictive stock “reinforces” cigarette consumption,  $C$ , meaning that the more stock, the more one will smoke, leading to curves  $A^1$  and  $A^2$  in Figure 24-3. Though not shown in the figure, the utility function shows the smoker as gaining utility from cigarette consumption,  $C$ , from the addictive stock,  $S$ , as well as from income, which allows the purchase of other goods in addition to cigarettes.

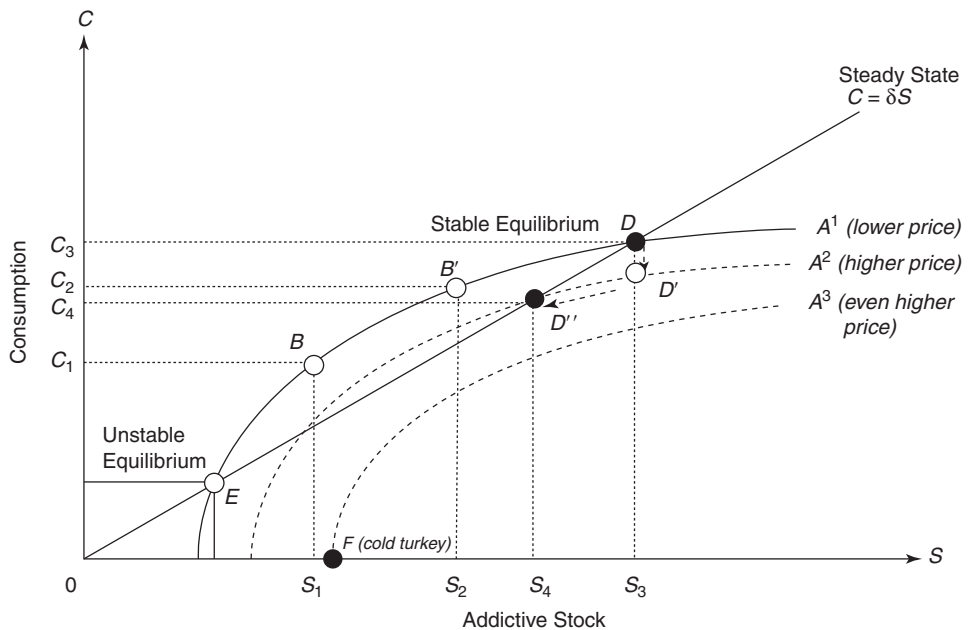


FIGURE 24-3 Rational Addiction Model

The important questions in the model deal with what happens over time. For example, current consumption increases the addictive stock. Listening to Mozart at age 21 will likely increase enjoyment of Mozart at age 22, thus increasing our musical “capital stock.” Most smokers will remember how bad the first cigarette tasted, but similarly smoking or drinking at age 21 may increase enjoyment of smoking or drinking in subsequent years. So, a larger addictive stock makes future consumption more pleasurable.

A myopic, or nearsighted, addict looks solely at the reinforcement effect. A rational addict, however, also considers the future harmful consequences of current addictive behavior. The rationally addicted smoker weighs the present pleasure against both the future health consequences and the beneficial impact of current consumption on future consumption enjoyment.

The rational addiction theorists have drawn several further implications from their analyses.

- Addiction is more likely for people who discount the future heavily, because they pay less attention to the potential adverse consequences.
- Addiction is more likely when the effects of past consumption depreciate less rapidly.
- Expected rises in future prices will have a dampening effect on current consumption, much like increases in current prices.

Models examining people’s behavior over time typically search for a “steady state,” an equilibrium where all “outflows” and “inflows” maintain the system, like the equilibrium of a well-run fish tank. In the steady state equilibrium proposed here, the system will be maintained over time, provided that current cigarette smoking adds exactly enough  $C$  to the addictive stock to replace the depreciation  $\delta S$  of that stock during the period. Mathematically,  $C = \delta S$ , where  $\delta$  is the constant depreciation rate. The  $C = \delta S$  line in Figure 24-3 depicts all the combinations of  $C$  and  $S$  that yield a steady state equilibrium.

Reviewing the elements of the model, we see:

1. **Consumption of cigarettes as a function of addictive stock.** Curve  $A^1$  relates smoking to addictive stock for a person with a given rate of time preference (relating future utility to present utility) and a given level of wealth, and who faces a set of prices for cigarettes and nonaddictive goods. We can think of curve  $A^1$  as a cigarette consumption curve, so the more stock  $S$ , the more consumption  $C$ . In other words, any given stock  $S$  is just sufficient to maintain consumption level  $C$ .
2. **Cigarette consumption needed to maintain addictive stock.** The stock of addictive capital depreciates at a rate of  $\delta$  (between 0 and 1) per year. Depreciated stock is replaced with more smoking. The ray from the origin,  $C = \delta S$  is the *steady state line* where current cigarette consumption just offsets the depreciation of the stock of smoking capital.

The model provides a convenient way to see what happens to the rational addict over time. Consider a smoker who is on consumption curve  $A^1$  (as an exercise, explain why a price increase from “low” to “high” would shift the consumption curve from  $A^1$  to  $A^2$ ) with addictive stock  $S_1$ . This stock implies a cigarette consumption of  $C_1$ , or point  $B$ . Notice, however, that consumption,  $C_1$ , will more than replace the depreciation in  $S_1$  during the period (point  $B$  lies above the steady state line,  $C = \delta S$ ). It follows that addictive stock will grow and exceed  $S_1$  in the next period; for example it may rise to  $S_2$  at point  $B'$ .

Continuing this logic, it follows that whenever consumption,  $C$ , lies above the steady state line, the addictive stock,  $S$ , will grow. Finally, stock  $S_3$  and consumption  $C_3$  give a steady state equilibrium for case  $A^1$ . We label this equilibrium point as  $D$ . Compare steady state equilibrium  $D$  with another place where the two curves cross, point  $E$ . Notice that  $D$  is like a magnet; any stock near  $S_3$  is pulled toward  $D$ . That is, any stock a bit to the left of  $S_3$  will bring growth in stock up to  $S_3$ ; any stock to the right will depreciate down to  $S_3$ . Point  $D$  represents a “stable equilibrium.” Try the same experiment

with the equilibrium at point  $E$ , and see that it is unstable. Any stock slightly to the left of  $E$  will be pulled farther to the left; any stock slightly to the right will eventually increase all the way to  $S_3$ .<sup>2</sup>

This model provides important policy implications about impacts of price changes, often induced through tax policies. Starting at point  $D$  consider a rise in price so that the cigarette consumption curve shifts from  $A^1$  to  $A^2$ . With the price rise, smoking decreases at first from  $C_3$  (at point  $D$ ), to point  $D'$ . It then falls farther over time since  $D'$  is below the steady-state line. Equilibrium smoking level falls to  $C_4$  at point  $D''$ . This shows the model to be consistent with our prior conceptions about price and quantity demanded. The higher the price, all else equal, the lower the quantity demanded. Moreover, the long-run responses to price changes *exceed* short-run responses. Initial decreases in smoking cause a subsequent decrease in the stocks of addictive capital, which then stimulate further smoking decreases.

Second, at some point, a rising price will cause all equilibria to disappear. Starting from point  $D''$  and letting the price rise even more, a new  $A^3$  curve will be everywhere below the  $C = \delta S$  steady state line. This prediction is unique to the rational addiction model, the prediction that some cigarette smokers quit “cold turkey,” without gradually reducing consumption down to zero, landing at a point similar to  $F$  where consumption equals 0.

It also follows that expectations about future prices of cigarettes will affect the addicts' current decisions about smoking. In Figure 24-3, we would interpret this by saying that a permanent price increase shifts the consumption curve downward farther than a temporary price increase. Both the price effect and the probability of going cold turkey are enhanced by permanent price increases. Likewise permanent restrictions on the advertising of cigarettes would have more effect than temporary ones.

## RATIONALES FOR PUBLIC INTERVENTION

Are cigarette smokers well-informed about the risks of smoking so they can make rational choices? Some economists (Lundborg and Lindgren, 2004; Viscusi, 1995) have challenged the common wisdom that smokers are illinformed, reporting smokers' knowledge and responses to risk to be similar to that of nonsmokers. Despite this, findings from various fields regarding smoker and nonsmoker behavior more often suggest that smokers differ from nonsmokers on these bases.

Consider nonsmokers first. We mentioned that another justification for intervention is that smoking has detrimental side effects on others, or “external costs.” Much then depends on the size of these costs. These externalities come from two principal sources: the passive smoking incurred by people nearby, and various other external costs caused by health hazards to the smoker. Manning et al. (1991) estimated the external costs at the equivalent of \$0.33 per pack for a new smoker evaluated in 1995 dollars, though he did not examine passive smoking costs. Viscusi (1995) also estimated the external costs per pack to be in this range, though lower than Manning's, and near zero under some scenarios.

Because economic efficiency is only one economic criterion (the other is equity), because experts may dispute data issues, and because economics is not the only basis to consider, many choose to intervene in tobacco and alcohol use. Economics offers two major tools that may be effective in curbing consumption of a targeted bad: restrictions on advertising and imposition of excise taxes. Advertising can be restricted by tax code revisions, but most often the public issue is whether to restrict advertising by selective or total bans. The excise tax tool is theoretically effective to the degree that demand is more elastic for a given supply. These two principal tools of intervention form the subject of our next investigations. First, however, consider why we have chosen not to address the several other tools.

<sup>2</sup> It can be shown that the  $A$  curve must cross the  $C$  line from above for an equilibrium to be stable. This occurs at point  $D$ . It does not occur at point  $E$ .

### Other Interventions

Two other potential forms of intervention are prohibitions on the consumption of the product and penalties for consumption or misuse of the product. Outright prohibition of cigarettes is unlikely to occur in the United States. Accounts of 1920s Prohibition in the United States seem to point to that clear conclusion. Furthermore, alcohol prohibition of that era probably was not even effective in reducing alcohol consumption (Miron, 1999; Dills, Jacobson, and Miron, 2005).

Lesser prohibitions, however, often have been accepted and proven effective. These include the effect of lower speed limits on rates of fatal accidents involving alcohol, as well as bans on smoking in public places.<sup>3</sup> Many countries, especially in Scandinavia, apply much more severe penalties for drunk driving than does the United States. These include stiff jail sentences for alcohol offenses, offenses that many Americans regard as less serious. Mullahy and Sindelar (1994) showed for U.S. data that legal penalties affect drunken-driving behavior. Increased fines and license revocations significantly reduce the individual's probability to drive drunk.

Regulation of smoking sometimes fails to work, although some research has found reductions in smoking after the passage of clean air restrictions. Tax effects might even be somewhat overestimated if part of the "tax effect" is really due to unmeasured local clean air restrictions. However, a recent Canadian study provides an example of research that finds regulation to be relatively ineffective (Lanoie and Leclair, 1998).

### ADVERTISING RESTRICTIONS ON CIGARETTES AND ALCOHOL

We begin with the role of advertising on cigarettes. At issue is whether advertising can increase the total consumption of bads like cigarettes. The issues we will address are ones on which people differ and hold strong views. In a report on cigarettes, the surgeon general (1989) concluded at one point that:

There is no scientifically rigorous study available to the public that provides a definitive answer to the basic question of whether advertising and promotion affect the level of tobacco consumption. Given the complexity of the issue, none is likely to be forthcoming in the foreseeable future. (pp. 516–517)

Even this conclusion was disputed by parties to both sides of the issue. Economists understand that well-intentioned interventions often have unintended consequences. Before blaming advertising for our ills, we should inquire into its nature.

**THEORIES OF ADVERTISING** Three main theories explain how advertising works and what it does for or to the community. Advertising alternatively is a form of information, a tool for persuasion, or a complementary good. The first two of these represent an old battle in advertising theory with contrasting villains and heroes: Information is generally beneficial, while persuasion is at least more questionable. The most recent addition treats advertising as a complementary good. Finally, Box 24-2 visits the advertisement of worthless goods—patent medicines.

**ADVERTISING AS INFORMATION** Nelson (1970, 1974) studied the implications of advertising as information. Viewed as information, advertising lowers equilibrium prices, creates better access to the market for new entrants, and provides better matches of consumer preferences with feasible consumption bundles. Informed consumers find that their dependence on or loyalty to Brand A will be weakened by their improved knowledge of alternatives. If consumers find it easy to opt for another brand, then they have flexibility, and flexible consumers will more likely resist undesirable changes

<sup>3</sup> See Chaloupka and Saffer (1992) and McCarthy (1993).

**BOX 24-2****Can Advertising Lead Patients Astray? The Case of Medical Quackery**

Excepting the medical heroes of history, such as Hippocrates and Galen, few doctors offered a scientific practice until the twentieth century. Modern medical science did not begin until the mid-nineteenth century and the typical patient confronting the typical physician did not attain even a 50-50 chance of getting better until about 1900. Most early practitioners may have had strong faith in their practices, such as bleeding the patient or applying strong purgatives, but these often may have worsened a patient's chances for recovery.

Oddly, many doctors throughout history have knowingly advertised and sold a worthless bill of goods. These “quacks” brazenly promised marvelous healing powers for such oddities as “patent medicines.” Such practices persist today mixed with generally innocent “alternative medicines,” herbal treatments. However, there still are modernized equivalents to the magical electric boxes and oddly designed “magnetizations.”

The audacity of the advertising medical quack was not lost on Gilbert and Sullivan, the noted creators of comic opera during the late nineteenth century. Their views on quacks were captured by:

MIKADO:

My object all sublime  
I shall achieve in time—  
To let the punishment fit the crime,  
The punishment fit the crime; . . .  
The advertising quack who wears  
With tales of countless cures,  
His teeth, I've enacted,  
Shall all be extracted  
By terrified amateurs.

**Patent Medicines**



Source: Photo by Scott Jordan, with permission.

in the brand, such as a drop in quality or an increase in price. This greater responsiveness to price implies a more elastic demand, and it makes possible lower market equilibrium prices. How can the firm's costly advertising activity help but be passed on to consumers in the form of higher prices? The mistake is that while the price at a given output must rise, the market equilibrium quantity may change due to competition, and the equilibrium price may fall.

**ADVERTISING AS A BARRIER TO ENTRY** In contrast to advertising as information, Bain (1956) argued that advertising differentiates one brand from another, creating increased brand loyalty. By making consumers more resistant to price changes and demand, advertising can result in greater market power and higher equilibrium prices. Comanor and Wilson (1974) added that the persuasive power of advertising may make it asymmetrically effective for the incumbent versus potential new entrants. Consumers have greater experience with established firms and greater recognition of them. The next advertising dollar spent by the established firm will yield a greater return than the same dollar spent by the newcomer.

**ADVERTISING AS A COMPLEMENTARY GOOD** Nobel laureate Gary Becker and colleague Kevin Murphy (1993) proposed a theory to account for these competing claims within a single model—one that appeals to an older theory of complements and substitutes. Let advertising be considered a good that is a complement to the good advertised. A consumer might wade through commercials with irritation during a ball game but enjoy the humorous one featuring a favorite beer. The



commercial increases the consumer's marginal utility from consuming that brand and, under this theory, firm advertising will raise total consumption of the product, just as a reduction in the price of mustard will increase the consumption of hot dogs.

### The Possible Effects of Brand Switching

When an imperfectly competitive firm advertises, it potentially improves its demand in part by inducing consumers to switch brands and in part by inducing consumers to consume more of the product in total. However, other firms also will advertise for these purposes, and the advertising among the firms may be partially or even totally offsetting. Advertising in the industry could have little or no effect on total industry demand, but it also could have a positive effect.

Laypeople often conclude mistakenly that the answer is obvious. Surely cigarette advertising must lead to more aggregate smoking or cigarette companies would not spend so much money on it. Cigarettes were once among the most heavily advertised products. A similar argument is made by some economists who note that the many available brands are really owned by a few companies and that the degree of brand switching (about 10 percent of smokers switch brands in a typical year) is not sufficient to justify the billions spent on advertising (Tye, Warner, and Glantz, 1987).

Others argue that brand switching is a sufficient motivation for heavy advertising. Two points help explain this view: (1) the firm decides to advertise *ex ante* with available information, and (2) the firm's decision criterion is what would happen if it did not advertise. On the one hand, the firm commits money to advertising based on the effect it expects advertising to have. The firm may not have the econometric data published later, it may not believe published data, or it may, in fact, have better information. On the other hand, the number of smokers who actually switch on average is less relevant than the unknown number who would switch if the firm did not advertise.

### Increased Demand or Brand Switching?

Examining the evidence, if we find that advertising increases total consumption, we could reduce consumption by restricting advertising. Studies suggest that cigarette advertising elasticities are very small. Research tends to find (Gallett, 2003; Baltagi and Levin, 1986) that advertising has no significant marginal effect on cigarette demand. Where researchers find the ad effect to be significant, the reported elasticities are generally quite small.<sup>4</sup> These elasticities fall between 0.1 and 0.2, and the most common result falls around 0.1. The Toxic Substance Board of New Zealand (1989), a strong advocate of advertising bans, conservatively uses an advertising elasticity of 0.07 in estimating the effect of an advertising ban.

Suppose that the true advertising elasticity were 0.1 exactly, and we apply this figure to estimate the effect of a total ban on cigarette advertising. We would reduce cigarette consumption by 10 percent, which is the product of a 100 percent price reduction and the 0.1 elasticity. Even this crude estimate might raise an unusual controversy in that ban advocates (arguing that the ban was successful) and ban opponents (arguing that it didn't change much) both might claim that this evidence is support for their case.

The elasticity estimates often are drawn from time trends of aggregate data, sometimes aggregated to the national level. At this level, little variation in advertising expenditures is captured, and the massive advertising levels outside of the cigarette industry, though relevant to cigarette demand, usually are ignored (Saffer and Chaloupka, 1998). It is also econometrically unsafe to extrapolate this far out of sample so we must study the effects of actual advertising bans.

World experience with advertising bans is informative. Several countries have banned cigarette advertising outright, and the United States installed a partial ban (on broadcast media) in 1971.

<sup>4</sup> McGuinness and Cowling (1975), Seldon and Doroodian (1989), and Tremblay and Tremblay (1995). Roberts and Samuelson (1988) simultaneously estimate brand switching and total consumption effects, concluding for their data that total consumption effects dominated.

The picture is clear regarding the U.S. experience. The 1971 ban of cigarette advertising on television and radio somewhat paradoxically may have increased cigarette consumption. Hamilton's (1972) account illustrates this peculiar history. Prior to the 1971 ban, an FTC ruling on the Fairness Doctrine required television and radio stations to give equal time to antismoking messages.

The result was a flurry of memorable commercials showing the harm of smoking. In one, a father appears walking in the forest with his young son. They stop to rest, lean back against a tree trunk, and the father reaches for a cigarette. His son watches in admiration as his dad lights up, but the father sees this, thinks, and stops. He gets the message and so do the viewers. Hamilton showed that these antismoking messages were effective in reducing smoking in the population, while the advertising expenditures had small, if significant, marginal effects. Thus, when the ban eliminated both pro-cigarette and the mandatory anticigarette messages at the same time, the combined effect could have increased smoking.

Perhaps more important, the antismoking messages had a relatively greater effect on young, potential smokers. Lewit, Coate, and Grossman (1981) studied data on more than 6,000 youths. They found that television watching significantly increased the probability that a youth would start smoking, while the Fairness Doctrine ads reduced the probability that a youth would start smoking. Harris and Chan (1999) have also found the clear relation of price elasticity and age; from 15- to 17-year-olds through 27- to 29-year-olds, price elasticity in absolute value starts out high and continually declines.

For many years, Hamilton's 1975 cross-national study showed that advertising bans have no significant effect on cigarette consumption in the country. Later, however, incorporating many new years of data and experience with bans, Laugesen and Meads (1991) reported these bans to be effective. Studying data for 1960 to 1986 and developing an index measuring the advertising restrictions, their time series estimates showed that a country's banning tobacco advertising will reduce tobacco consumption by 6.8 percent. Recent studies corroborate this result (Iwasaki, Tremblay and Tremblay, 2006; Saffer and Chaloupka, 1998), though it warned that partial bans may invite tobacco firms to substitute advertising in unrestricted categories for banned advertising. Keeler and colleagues (2004) report that cigarette companies in the United States acted exactly in this manner in their responses to the U.S. tobacco settlement of 1998. Facing reduced revenues by 8.3 percent due to the consequent price increase, the companies offset about one-quarter to one-half of that through increased advertising.

### **Advertising and Alcohol Consumption**

Although earlier studies reported no effect of advertising on alcohol consumption, Saffer (1991) found that such advertising was a significant factor in drinking. Similarly, Saffer and Dave (2003) found such advertising to be especially effective on youthful drinkers; they estimated that a complete ban on all alcohol advertising could reduce adolescent alcohol consumption by 24 percent with even stronger effects on binge drinking. Taking a different approach, Saffer (1997) found alcohol advertising levels correlated positively and significantly with motor vehicle fatalities, *ceteris paribus*. He estimated that partial bans (bans on broadcast advertising) would reduce annual fatalities by 2,000 to 3,000, while a total ban could reduce fatalities by up to 10,000 nationally.

## **EXCISE TAXES AND CONSUMPTION OF CIGARETTES AND ALCOHOL**

The public commonly believes that producers always pass taxes in their entirety on to consumers, but this is not true. If it were true, then cigarette and alcohol taxes would have no effect other than to raise money for the government coffers. By first examining the theory of excise taxation, we learn the importance of the price elasticity of demand and supply in determining the incidence of the tax and the degree to which it reduces consumption. With this understanding, we will examine the empirical knowledge regarding these elasticities.

### The Consumption-Reducing Effects of Excise Taxes in Theory

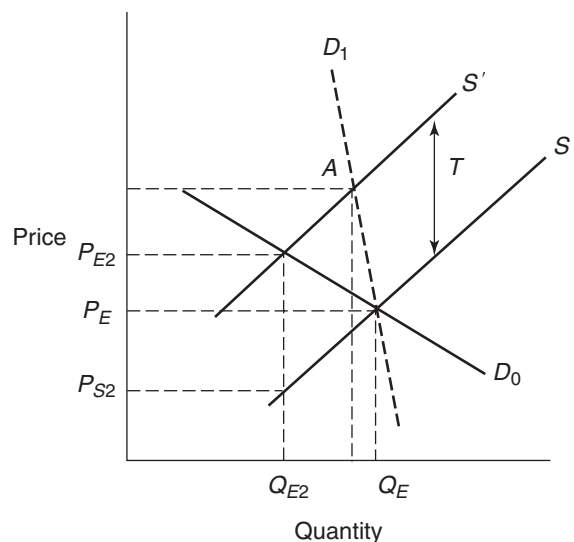
Excise taxes form another major tool that can reduce consumption of bads in populations. Figure 24-4 depicts the supply and demand for alcoholic beverages. If a sufficient number of voters were to agree, the consumption of beer could be reduced from  $Q_E$  to  $Q_{E2}$  by imposing an excise tax increase of  $T$  dollars (recall that the tax shifts the supply curve vertically by the amount of the tax). The analysis also reveals that the tax would not fall entirely on the beer drinker, but the incidence would be shared. Assuming demand curve,  $D_0$ , because the consumer pays a higher price after the tax, we say that the incidence on the consumer is  $(P_{E2} - P_E)$ . Because the producer pockets the lower amount,  $P_{S2}$ , we say that the incidence on the producer is  $(P_E - P_{S2})$ .

The research issue is illustrated by the alternative possibility that the demand curve for beer is  $D_1$ . The alternative equilibrium is at point  $A$ , at which the reduction in consumption (not shown) is less than the original reduction,  $(Q_E - Q_{E2})$ . Thus, the reduction in quantity demanded will depend on which curve represents the true demand curve, the relatively inelastic curve,  $D_1$ , or the more elastic curve,  $D_0$ . Generally, the more responsive demand is to price (the greater the demand elasticity in absolute value), the greater the corresponding reduction effected by the excise tax increase.

From models such as the rational addiction model, we know to distinguish between long-run and short-run effects of price changes. We also learned to distinguish youth from adult consumers, a distinction with important policy consequences. Box 24-3 provides a further discussion related to tobacco policy.

### Excise Taxes and Cigarette Consumption in Practice

The importance of price elasticity is not lost on economists, and most econometric studies of cigarette consumption report price elasticity estimates. Reported cigarette price elasticities in absolute value range rather widely, from as low as  $-0.2$  to occasional estimates greater than  $-1.0$ . There is no doubt that cigarette demand responds significantly to price, but the response is generally inelastic. Building on the earlier work, more recent studies find short-run price elasticities in a narrower band, typical of which is the  $-0.3$  to  $-0.5$  band reported by Keeler et al. (1993).<sup>5</sup> The United States



**FIGURE 24-4** The Consumption of Alcohol Is Reduced by an Excise Tax Increase

<sup>5</sup> See also Tremblay and Tremblay (1995). Smokeless tobacco demand also responds to price in econometric studies, both on price and the price of substitutes (Ohsfeldt, Boyle, and Capilouto, 1998).

**BOX 24-3****Mind If I Smoke?**

On a California billboard, the man's date responds to this question by asking:

"Care if I die?" In a similar spirit, recent American films have depicted the inside of the tobacco industry as darkly menacing. The British news magazine, *The Economist*, ponders whether the tobacco industry has become "the new evil empire."

Meanwhile, other reports suggest that average smoking rates in the population continue to decline. The decline has also occurred among women, with an especially large drop in smoking among women of childbearing age. Tobacco companies generally have raised cigarette prices, and in California, the state legislature's cigarette sales tax increase at the beginning of 1999 was followed by an unexpectedly sharp decline in cigarette sales. The high level of legislative, activist, and consumer discussion and debate over tobacco continues throughout the country.

This scenario forms the backdrop for the tobacco companies' unprecedented November 1998 offer of \$368 billion to the states in return for promises to limit further lawsuit activity. The American Cancer Society and the American Medical Association both gave their approval, albeit with some qualifications, and the settlement began with the highest hopes.

The legal theory applied in these contexts and to be used in future Department of Justice suits holds the tobacco companies liable for health costs incurred by the governments to provide care to cigarette-induced disease victims. The argument distinguishes cigarette issues from others involving voluntary risk taking, such as skydiving, alpine skiing, or work in risky occupations. An important legal question is whether the tobacco companies deliberately withheld critical health information.

Tobacco Settlement of 1998 led to a rise in the price of cigarettes, and consequently it offered a natural experiment on the effects of price on consumption. The results show a decline in cigarette consumption of 8.3 percent, though increased advertising by cigarette companies offset this effect (Keeler et al., 2004). Sheu and colleagues (2004) estimated the price elasticity from these data to be in midrange,  $-0.46$  in absolute value. More recent estimates by Carpenter and Cook (2007) support the previous findings of approximately  $-0.56$ .

Early research found that youth were much more responsive to price increases and more accessible to the excise tax tool. This is an attractive result for anyone wishing to deter youth from taking up the smoking habit: "an excise tax increase . . . might continue to discourage smoking participation by successive generations of teenagers and young adults and gradually impact the smoking levels of older age groups as the smoking-discouraged cohorts move through the age spectrum" (Lewit and Coate, 1982, p. 143). Recent research tends to support the view that "price is a powerful determinant of smoking for high school seniors" (Gruber and Zinman, 2002); they estimate the cigarette price elasticity of these youths to be  $-0.67$ . But the findings of Sen and Wirjanto (2010) are much lower, at elasticities of  $-0.2$  to  $-0.5$ , suggesting that this issue is still in dispute.

Even low elasticities can do some significant good. For example, with  $-0.2$  as the true number, a doubling of the price of a pack would reduce youth smoking by 20 percent. As for an unexpected effect on children, higher cigarette taxes tend to increase the BMI (body mass index) of children of smoking mothers (Mellor, 2011)

As predicted by some models, such as rational addiction, estimated long-run cigarette elasticities exceed short-run elasticities in absolute value. Chaloupka (1991) applied the rational addiction framework and estimated long-run values approximately twice as large as for the short run. Keeler et al. (1993) found them to lie in a narrow range,  $-0.5$  to  $-0.6$ . A more demanding test of the rational expectations model is one that takes seriously its claim that consumers make choices not only based on current cigarette prices but also on their ideas about future cigarette prices. The

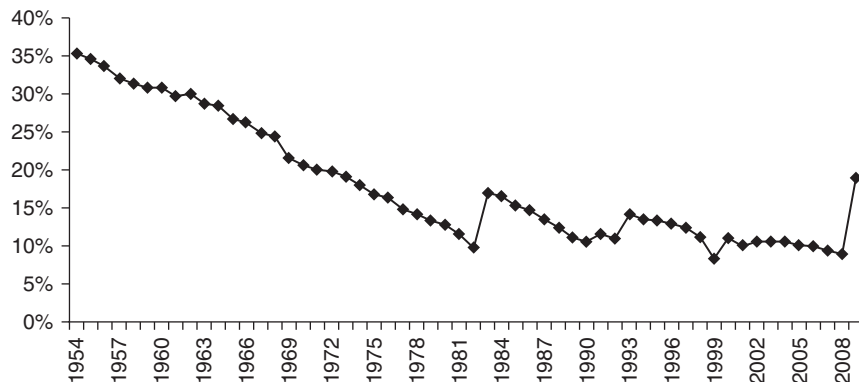
estimation process requires difficult econometrics, and ideal data sets are difficult to come by, but two recent studies confirm this forward-looking prediction (Baltagi and Griffin, 2001; Gruber and Koszegi, 2001).

With the elasticity magnitudes reported in these various studies, what possible effects could we expect excise tax increases to have? To focus the discussion, consider a tax increase of sufficient magnitude to double the price of cigarettes. Assuming an elasticity of  $-0.4$ , the 100 percent price increase, by simple extrapolation, would reduce cigarette consumption by 40 percent, a substantial reduction. As we discussed earlier, where a similar extrapolation for advertising bans suggested a more modest reduction, such out-of-sample extrapolations are risky, but the example serves to illustrate the potency of the tax tool. Furthermore, the elasticity estimates for youth are large enough to double this effect for that age cohort. Likewise, the larger long-run elasticities suggest the greater policy effects as time passes.

Direct measures of the effect of cigarette taxes on mortality also show the effectiveness of tax policy. Moore (1995) tested tax variables in equations to predict mortality from several smoking-related diseases. Higher taxes significantly reduce mortality from lung cancer, cardiovascular disease, and asthma. He reports that a 10 percent cigarette tax hike would save 3,700 lives per year in the United States. Evans and Ringel (1999) studied the effect of cigarette excise taxes on birth outcomes. Their research found the taxes effective in reducing smoking by expectant mothers and generating better, higher birth weights in the newborns.

Although some countries have applied cigarette taxes vigorously to reduce population smoking, the United States has not done so until recently. In 1997, the U.S. average tax rate, at 35 percent of the average price of a pack, was the lowest among 29 countries in one study. The United Kingdom, Ireland, Germany, and all of Scandinavia ranged from 70 percent to 85 percent. The U.S. federal excise tax declined in real terms after the surgeon general's mid-1960s announcement that cigarette smoking causes cancer until recent increases to \$0.39 per pack (in percentage terms it was still lower in 2002 than it had been in 1969). Figure 24-5 reveals that the tax decrease because a nearly constant nominal tax rate fell in percentage terms due to the secular rise in prices per pack.<sup>6</sup>

**MEDICAL AND SOCIAL INFLUENCES ON SMOKING** While health economists often emphasized prices, taxes, and advertising as influences on demand for health bads, there has been a growing interest in medical and social factors. Saffer and Dave (2005) find that people with a history of mental illness are much more likely than the average person to consume alcohol (26 percent more likely), cocaine (66 percent more likely), and cigarettes (89 percent more likely).



**FIGURE 24-5** Federal Taxes as a Percent of the Price per Pack

<sup>6</sup> As of April 1, 2009, the federal excise tax jumped from \$0.39 per pack to \$1.01 per pack, an increase of \$0.62 or 159 percent.

Social influences also play a strong role. DiCicca and colleagues (2006) related an index of local attitudes toward smoking to cigarette consumption levels. The more adverse the anti-smoking sentiments in the local culture, the lower were the smoking levels. Studies by Folland (2008) and by Brown et al., (2006) find that individuals with strong social networks are less likely to smoke. Other studies focus on teenagers, finding that these adolescents are strongly influenced by their social peers regarding smoking behaviors (Katzman, Markowitz, and McGeary 2007; Clark and Loheac, 2007).

### Excise Taxes and Alcohol Consumption

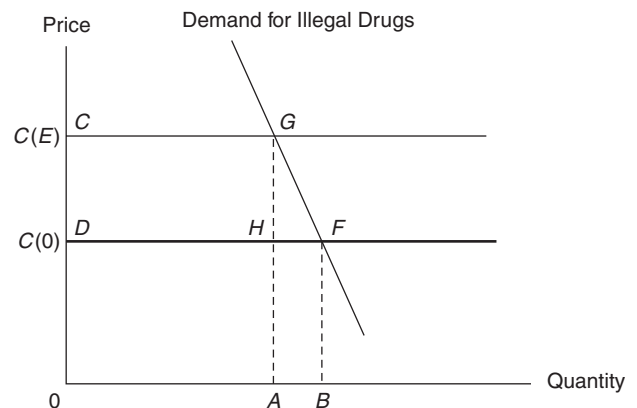
Studies of alcohol consumption, price, and advertising have often focused on youth; these younger age groups exhibit the highest rates of alcohol abuse, such as binge drinking (Cook and Moore, 2000). The beer tax modeled in Figure 24-4 as appeared to be substantially effective in reducing consumption, but would we see this effect in real life? Grossman et al. (1998) find that young adults respond to beer price increases, and the elasticities range from  $-0.2$  to  $-0.4$  in the short run, with long-run elasticities 60 percent higher; other beer elasticity estimates have ranged even higher. Studies of wine and spirits yield elasticities ranging from  $-0.3$  to  $-1.8$  (Saffer and Chaloupka, 1998). More recent results by Sen and Campbell (2010) provide more somber evidence. They find that license-related regulations and excise taxes on alcohol significantly reduce motor vehicle fatalities involving children.

The latter study also reports substantial, though inelastic, estimates for price elasticities of illicit drugs including marijuana, cocaine, and heroin. Thus, taxes should have major impacts on consumption of addictive substances.

**ILLEGAL DRUGS AND PROHIBITION** While many would not wish to prohibit smoking entirely or to return to the 1920s prohibition against alcohol, the subject of prohibition remains highly relevant to consumption of illegal drugs. Since the turn of the twenty-first century, there have been widespread increases in the usage of such drugs as cocaine, ecstasy, and crystal methamphetamines. This increased usage has led to increased illegal drug trafficking, loss of productivity among drug users, and increased violence in both rural and urban America.

America's "War on Drugs" has sought to limit the distribution of such drugs. Should we modify or stop this War on Drugs? We leave the politically contentious issues to a more policy-oriented venue and focus instead on a theoretical issue raised by Becker, Murphy, and Grossman (2004). Which is better, criminalization with enforcement or legalization with an optimal excise tax?

The authors first make the point that when the consumption demand for the drug is price inelastic, increases in enforcement can cause substantial increases in the money that drug smugglers spend to distribute their product to consumers. These extra expenditures do not generally improve social welfare; smugglers must avoid the police, fight off competitors, and distribute the drugs by stealth in criminal networks. Figure 24-6 illustrates this point. Becker and colleagues make the



**FIGURE 24-6** Comparing Prohibition versus Legalization under Conditions of Inelastic Demand

simplifying assumptions that the drug dealers are perfectly competitive and produce at constant unit costs. We associate the unit cost levels,  $C(0)$  and  $C(E)$ , with the legalization case and the criminalization case respectively. When the drugs are legal, there are no police enforcement expenditures and correspondingly no extra expenditures by the drug firms to avoid prosecution, so  $C(0)$  indicates unit costs when enforcement costs are zero. In contrast, when the drugs are illegal, the unit costs will be higher as drug firms must pay not only production costs but extra costs to avoid prosecution and to maintain illegal distribution networks; in this case, the unit costs are  $C(E)$ , where  $E$  equals the enforcement costs. For these reasons,  $C(E)$  is greater than  $C(0)$ .

Suppose that drugs are legalized so that  $E = 0$ , and the market equilibrium is at  $F$ . Drug seller costs are for production and distribution only, represented by the area of  $ODFB$ . In other words, all costs go to production and distribution. Contrast the case where enforcement costs,  $E$ , are positive and the smugglers must evade the law. The new market equilibrium is at  $G$ . Note the implication is that smuggler costs have become the area  $OCGA$ .

These smuggler costs become substantial and the reduction in consumption,  $B - A$ , is quite small. The observation by eye that  $OCGA$  exceeds  $ODFB$  is corroborated by microeconomic principles which state that under inelastic demand a rise in price will increase revenue. Why is the familiar result for revenue relevant here? This is because the authors' assumption that the smugglers are competitive implies that in the long run the firms' costs will equal their revenues. In other words, since we know that revenues have increased and since in competition the firms' revenues equal their costs, we know that drug firm costs have increased, and realistically the illegal drug firm costs will be substantial. Recall that the illegal drug firms' costs are in large part spent avoiding prosecution and related criminal activities. Costs like this are unlikely to provide a social value, and most people would say that they are very harmful to society.

Contrast how society seeks to reduce drug use under legalization. One prominent economic tool of government would be an excise tax on the drug. While not shown here, tax revenues always accrue to the government, which can use them through activities to improve social welfare. The optimal excise tax is familiar to students of microeconomics: one installs a tax high enough to equal the marginal external cost to society of the drug consumption. Although the reader must refer to Becker, Murphy, and Grossman (2004) for the analytic development, the economic theory of externalities and market failures shows that a properly chosen excise tax can, in principle, improve social welfare.

In Figure 24-6, we would show an excise tax with a horizontal line. Becker and colleagues, through an analytical welfare analysis, ask whether the optimal price (with tax) under legalization be higher or lower than the unit costs under criminalization with optimal enforcement costs,  $E$ ? Given the context described in this section, the authors concluded that the price (with optimal tax) under legalization would be higher than the price of illegal drugs under criminalization. This result would make a strong case for legalization. With higher drug prices, consumption would decline, a benefit to society. Second, it would reduce or eliminate the criminal activities and expenses created by illegal drug suppliers.

## CONCLUSIONS

This chapter examines the economic nature of health bads, and it studies the potential of curbs on advertising and increased excise taxes to reduce the consumption of bads. Econometric estimates of the effect of advertising for cigarettes tend to report small and sometimes insignificant elasticities. These are estimates of changes at the margin. In contrast, advertising bans entail by definition large reductions in advertising levels; the most recent work reports that bans have some significant effect when studied on an international basis. Related studies on alcohol advertising also report statistically significant effects.

Excise taxes, however, appear to be more potent. Though cigarette and alcohol price elasticities are in the inelastic range, they are large enough in absolute value to have substantial potential as a curb to consumption if the public chooses to apply them. Consistent with rational addiction models, the long-run price elasticities are greater in absolute value than the short-run elasticities. Cigarette price elasticities for youth tend to be larger in absolute value than those for adults.

---

## Summary

1. The nature of addiction, though yet unresolved, generates behaviors that appear generally consistent with a rational addiction model. Irrational or myopic addiction models provide viable alternatives.
2. Addictive behavior involves both “reinforcement” and “tolerance.” A myopic, or near-sighted, addict looks solely at the reinforcement effect. A rational addict, however, considers the future harmful consequences of current addictive behavior.
3. There are three contrasting theories of advertising. In one, advertising is primarily information, which leads to greater competition and possibly lower prices. The second describes advertising as a potential barrier to entry, which increases monopoly power and prices. The third describes advertising as a complement to the good advertised; it increases marginal utility of the product.
4. Advertising bans and related restrictions appear to have significant effects on consumption of cigarettes and alcohol. Though total effects of advertising are small in econometric studies, they are somewhat more substantial in a recent international study of total bans.
5. Excise taxes work to curb consumption in theory by increasing the price of the product. Then, the effect on consumption depends on consumer responsiveness to price, which is the price elasticity.
6. Increased excise taxes on cigarettes and alcohol appear to be more potent public policy tools for curbing the consumption of bads than are advertising bans. Though estimated price elasticities are small in absolute value, they can be combined with substantial price changes to result in correspondingly large reductions in consumption.

---

## Discussion Questions

1. Many students smoke, drink alcohol, or ingest other addictive substances, such as caffeine (in coffee, tea, or soft drinks). How do the addiction models presented relate to students’ everyday habits?
2. Do cigarette and alcohol ads you have seen contain primarily informative content or persuasive content? What consequences would you predict from your finding?
3. Someone says: “The advertising elasticity of cigarette demand may be small, but it is big enough to warrant policy to ban cigarette advertising.” What would “big enough” mean in this context?
4. What does the evidence on the effects of the partial U.S. ban on cigarette advertising suggest about the relative effectiveness of cigarette advertising versus anti-smoking advertising?
5. How does the responsiveness to cigarette advertising and price differ between youths and adults? Why is this important?
6. Explain how it is possible to advertise too little or too much.
7. Suppose it takes considerable time for the large majority of cigarette smokers to become fully informed about the fact that cigarette prices have risen all over. How would this matter to the effectiveness of a tax hike? What other effect of a cigarette tax hike may take considerable time?
8. How would an excise tax hike in Kansas curb smoking among Kansans if the neighboring states (e.g., Missouri, Nebraska) did not also hike their cigarette excise taxes?
9. When all costs are considered, which public policy tool for curbing the consumption of bads would be the most costly to administer for a given amount of consumption reduction among consumers?



---

## Exercises

1. The analysis accompanying Figure 24-3 investigates the impacts of an increase in cigarette prices. Use the figure to show both the short-run and the long-run impacts of a cigarette price decrease.
2. If the elasticity of aggregate cigarette demand with respect to advertising were 0.15 in absolute value, by extrapolation what effect on cigarette consumption would be caused by a 10 percent reduction in advertising? A 50 percent reduction? A 100 percent reduction? How and why does one's confidence in prediction change over this range of reductions?
3. Suppose the price elasticity of cigarette demand is 0.4. If we increased the prices of cigarettes by 50 percent, what would we expect to happen to the quantity purchased? To total expenditures on cigarettes?
4. In the discussion on rationales for intervention in markets, we note that Manning found external costs of \$0.33 per pack of cigarettes.
  - (a) Draw a supply and demand diagram, and graph Manning's external costs of \$0.33 (in 1995 dollars) based on a market price of \$1.50 (in 1995 dollars) per pack.
  - (b) If a tax of \$0.33 were imposed, what would happen to the market price, and to the equilibrium quantity?
5. Using Manning's estimate of external costs of \$0.33 per pack of cigarettes:
  - (a) Calculate the new market price for cigarettes using a demand elasticity of 0.25. Would this be economically efficient?
  - (b) Calculate the government's revenue from a \$0.33 tax on cigarettes in part a.

# Epidemiology and Economics: HIV/AIDS in Africa

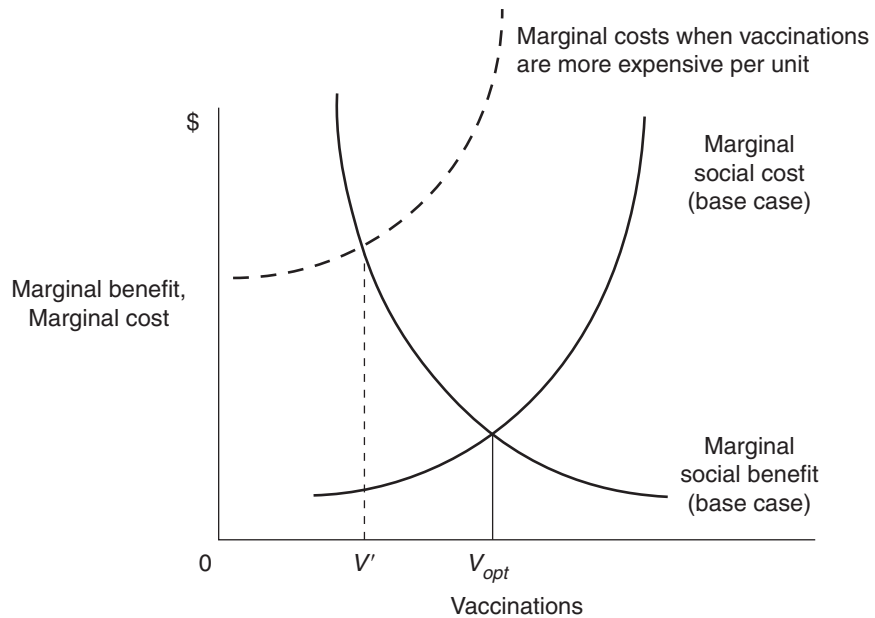
- Concepts from Epidemiology
- Economic Epidemiology
- Case Study: HIV/AIDS in Africa
- Conclusions

Infectious diseases have tormented human history, yet despite our great technological progress, many lethal epidemics remain with us, and some, such as 2009's H1N1 "swine flu," arose only recently. Analysts have long understood the contact of an infecting agent and a human host as necessary for infection to incur. Recently, however, they have come to understand how crucial the behavior of the host is to this process. For example, the uninfected person's demand for prevention may affect choice of geographic location, social milieu, and health information, as well as demand for preventive goods and services. Economic theory shows how these economic behaviors can alter the course of the epidemic. In this chapter, we explain how economists use economic principles to enrich understanding of epidemics. We then apply these and related economic ideas to one of the worst cases, that of HIV/AIDS in Africa.

## CONCEPTS FROM EPIDEMIOLOGY

Analysts describe the magnitude of disease in society in a multitude of ways. The *incidence rate* refers to the number of new cases for a defined population. The *hazard rate*, in comparison, is the instantaneous rate of new cases among the uninfected population in a given period. Like the incidence, it describes one's chances of getting the disease. "Instantaneous" refers to an arbitrarily short period. The hazard rate is common in models of epidemics, though it also applies to engineering, business, and economics. In our discussion, it represents the rate of new infections.

The fraction of the population that is currently infected is the *prevalence rate*. Incidence contributes to prevalence like income contributes to wealth, in that incidence adds new cases to the total pool describing the prevalence of present cases. Entry and exit to this pool of the infected population come from births and deaths, as in all population models, but also changes when people develop immunity or adapting behaviors. The dynamic process of a disease, its path through time, sometimes shows the prevalence growing (perhaps even into a massive epidemic) or, one hopes, declining to a benign level.



**FIGURE 25-1** The Marginal Benefits and Costs of Influenza Vaccination

Alleviating an epidemic can be achieved either by natural means or by public health measures, such as vaccinations. One of the most fundamental economic concepts helps to explain the problem of optimizing the use of vaccinations in a population. Figure 25-1 applies net benefit maximization by illustrating both the marginal social benefits and the marginal social costs of vaccination in what we refer to as the “base case.” As we vaccinate more people, it becomes more difficult and costly to find others to vaccinate; thus, the marginal costs will increase. In contrast, as we vaccinate more, the next vaccinated person provides a lesser external benefit to others (that is, his or her newly achieved immunity is less important to society). Consequently, the marginal benefits will tend to decrease. As we have seen throughout health economics, the maximum social benefit occurs where the marginal benefit equals the marginal cost. This base case is shown in Figure 25-1 as  $V_{opt}$ .

Under some circumstances, the cost of vaccinations may be very high, leading to low levels of vaccination, and correspondingly high levels of disease and potential death, as noted in the dashed line in Figure 25-1, and the implied vaccination level  $V'$ . Epidemic type A meningococcal disease occurs only in West Africa. In 2010, there were at least 88,000 cases of meningitis in sub-Saharan Africa, including more than 5,000 deaths, with more than 224,000 cases reported in the previous 13 years.

When patients contract the disease early and start treatment, up to 10 percent of them die within two days. Up to 20 percent of survivors have long-term problems such as brain damage and hearing loss. The economic impacts of the disease are palpable. In addition to the pain and suffering, a 2007 Burkina Faso study showed that an average family affected by a case of meningitis among one of their family members spends three to four months of their annual income dealing with that case.

Vaccines to deal with the disease have generally cost from \$10 to \$100, far beyond the means of those in countries where the monthly incomes hardly reach those levels. This occurs because larger drug companies seldom find the production of such “orphan drugs” to be profitable. The high cost leads to reduced levels of vaccination, or  $V'$  rather than  $V_{opt}$ , which would occur if the vaccine was less expensive. However, Dr. Richard Besser, a physician and ABC correspondent, reported

that in December 2010 the government of Burkina Faso launched the MenAfriVac vaccine and over a 12-day period, the entire at-risk population (all those between the ages of 1 and 29) was offered vaccine against this disease using a vaccine made *just for Africa*. The new vaccine had resulted from a partnership that began in 2001 between the World Health Organization, the Serum Institute of India, and PATH, an international nonprofit funded by the Bill and Melinda Gates Foundation. It costs less than 50 cents per unit compared to \$10 to \$100 for other vaccine doses. It also offers protection that lasts a decade, compared to three years for others.

This initiative was a rousing success. On May 30, 2011, in a news report on ABC television, Dr. Besser reported that in comparison to the past, where the region averaged thousands of cases of meningitis each year, in 2011 there were only four reported cases of meningitis, all in patients who had not been vaccinated.

Vaccination succeeds by bringing immunity to the disease to the greater population. Immunity, however, can also be acquired through surviving an episode of the disease. It may seem odd to us, but parents in the past would often deliberately expose their children to an infectious disease to gain them immunity (see Box 25-1). For the less virulent diseases that plagued children in recent centuries, groups of young adults had often attained up to 95 percent immunity through surviving the disease. The remaining few young people still uninfected were often naturally immune and would remain so throughout their lives. For an epidemic to result from the process of infection and contagion, it must also be the case that the rate of new infections among the susceptible population be sufficiently high to overcome the healthful offsetting effects of vaccination, naturally acquired immunity, and other protections.

Though fighting infectious disease without the aid of modern medicine might appear to be fruitless, the evidence from medical historians shows that the prevalence of many infectious diseases declined substantially well before the discovery of modern, effective medicines (McKinlay and McKinlay, 1977). As seen in Chapter 5, historians have shown that a good share of the reductions in infectious disease prevalence was due to the improved physical heartiness and resistance of the host, in turn, due especially to improvements in nutrition (McKeown, 1976; Fogel, 1986).

Boulier and colleagues examine the vaccination externality in more detail, starting with the Susceptible-Infective-Removed (SIR) model of epidemiology originally developed by Kermack and McKendrick (1927) and reinterpreted mathematically by Hethcote (2000). This model relates the disease incidence to its infectiousness, the size of the population, and the percentage of the population that is susceptible.

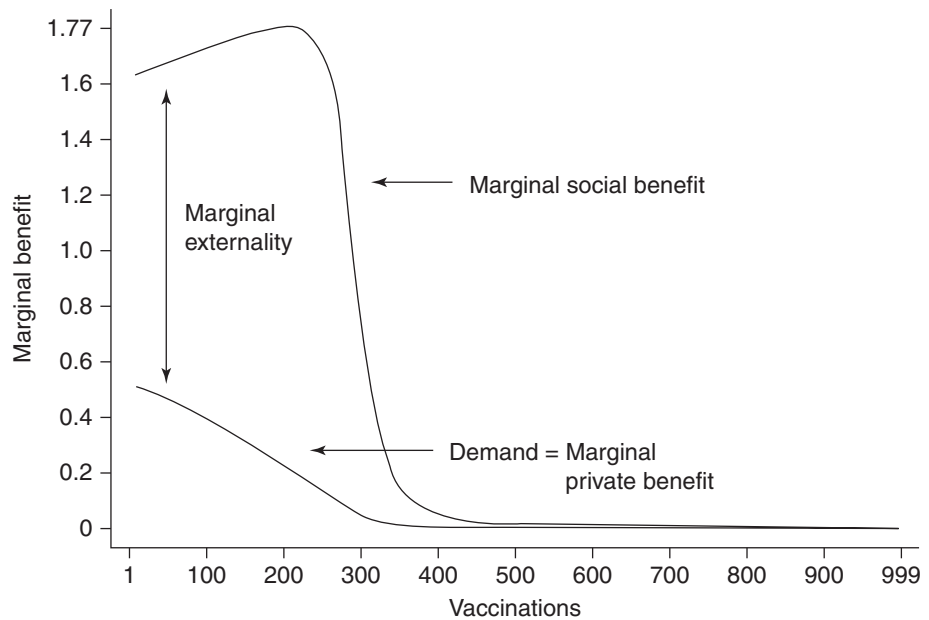
### BOX 25-1

#### Wanting Your Children to Catch Chicken Pox

Edward Jenner conducted the world's first vaccination, for smallpox, in 1796. Jenner's vaccine, derived from cowpox, was different from the vaccine that by 1980 eradicated smallpox from the globe. Each vaccine contains some risk, however small; with eradication the vaccine was finally declared to have become more risky than the chance of contracting the disease. Forty other vaccines have come into use since Jenner's, though none have achieved the eradication success of the smallpox vaccine.

What did parents have at hand when no vaccines were available, and what could they do to protect their children? The answer, often, was "not much." Until well into the twentieth century, the typical child, often one of many, had roughly a 50–50 chance of surviving into adulthood. One perhaps unusual and old-fashioned approach was nevertheless likely to be effective. Parents deliberately exposed their children to infectious diseases! Some held "chicken pox parties" where the uninfected played in the same room with an infected child. Mumps and other childhood diseases were often handled similarly.

How could this possibly help? Many childhood diseases become more virulent as the child gets older, increasing the risks of bad consequences. The exposed younger children gained immunity and were safe through to their teens, during which time susceptible children faced real dangers of more severe consequences including sterility or even death.



**FIGURE 25-2** Marginal Private and Social Benefits of Influenza Vaccination  
 Source: Bryan L. Boulier, Tejwant S. Datta, and Robert S. Goldfarb (2007) "Vaccination Externalities," *The B.E. Journal of Economic Analysis & Policy*: Vol. 7: Iss. 1 (Contributions), Article 23. Available at: <http://www.bepress.com/bejeap/vol7/iss1/art23>, with permission from publisher.

They calibrate their model for influenza vaccination. As shown in Figure 25-2, from the first vaccination, the initial marginal private benefit is a little less than 0.6 cases of influenza prevented. The marginal external benefit is a little more than one additional case prevented, giving a marginal social benefit of 1.6 cases prevented. The marginal external benefit curve “balloons out” to be as high as 1.58, equaling  $(1.77 - 0.19)$  additional cases prevented, before falling toward zero, as the number vaccinated increases, and the number who could catch the disease correspondingly falls.

The authors’ analysis suggests that while vaccination is likely to yield positive externalities, analyses must be disease-specific. They emphasize that many analyses use parallel shifts of demand curves, or simple multiples, to characterize the social benefits from vaccinations, but that the simple characterizations provide misleading results when used in measuring the benefits accruing to vaccination.

## ECONOMIC EPIDEMIOLOGY

Economists have defined a new area of study described as “economic epidemiology.” We present these theoretical models and related empirical results. Then, we will examine how we might understand the world’s worst present-day epidemic, the case of AIDS in Africa, in light of economics. We seek to find how it came to be, and how its spread can be reduced.

### Rational Epidemics

What would rational individuals (both those infected and uninfected) do in the face of an epidemic? Geoffard and Philipson (1996) incorporate the assumption of rational economic behavior into epidemiological models. Suppose both the uninfected and the infected seek to maximize their utilities. Realizing that one’s customary behaviors entail a risk of infection, susceptible individuals will

demand at least some preventive care. Because preventive care entails costs, they will rationally choose less than the maximal amount of everyone's getting preventive care.

At the same time, however, individuals may also demand risky experiences. For example, where the disease is transmitted sexually, the individuals' demand for sexual intimacy becomes a demand for risky experiences, a risk to which the individuals may or may not be fully informed. In another context, suppose that a young grade-schooler wants very much to go to school, but her parents know that a measles epidemic has affected the school. The child's desire to go to school is a demand for a risky experience.

If fully informed, rational people seek to maximize utility, considering both costs and gains; people's behaviors may still lead to an epidemic. One special focus for economists is the responsiveness of prevention demand to new information regarding the prevalence of the disease.

### The Prevalence Elasticity of Demand for Prevention

Assume that the susceptible population responds to information regarding an increased prevalence of the disease by demanding more preventive care. Further, define:

$$E_p = \text{Prevalence Elasticity of Demand for Preventive Care}$$

$$E_p = \frac{\% \text{ Change in Prevention Demanded}}{\% \text{ Change in Disease Prevalence}} \quad (25.1)$$

If  $E_p$  is low (zero, or close to it), people will demand only a small amount of prevention, resulting in higher future prevalence. In contrast, if  $E_p$  is large (much greater than zero), then they will demand a large amount of prevention (for example, vaccinations), resulting in lower future prevalence.

These demands for preventive care will, in turn, alter the prevalence rate of the disease and could dampen or reverse any further growth of prevalence. Philipson (2000) shows theoretically that a sufficiently elastic response (large  $E_p$ ) leads to a decline in the prevalence rate, thus stemming the progress of the epidemic.

Do susceptible people respond in practice to the threat implied by an increased prevalence by demanding more prevention? Two studies examined data for a measles epidemic (Philipson, 1996; Goldstein et al., 1996). Both studies found evidence of a significant prevalence elasticity of demand for prevention. One found that the rate at which people got themselves "up-to-date" with the measles vaccine protection increased from 56 to 58 percent before a measles epidemic began to 70 percent during the epidemic. Researchers found the same sort of demand behavior in a cross-state study of AIDS (Ahituv, Hotz, and Philipson, 1996). There, the use of condoms was quite responsive to the prevalence of AIDS. Thus, these data support the economic model, and imply that incorporating an economic response into the epidemiology model will better describe the course of epidemics.

Because people may respond to the threat, some epidemics will be self-limiting. Does this imply that we need not worry about rampant epidemics, such as the case of AIDS? Not entirely. First, in some cases, the model predicts that an epidemic will continue to grow if there is no intervention. Second, the hoped-for behaviors can go awry. For example, given knowledge of high prevalence, people who customarily engage in high-risk activities might plausibly become more fatalistic (assuming that they will die anyhow), leading them to engage in yet more high-risk behaviors.

### The Economic Consequences of Epidemics

An epidemic is potentially costly in several ways. Persons contracting the illness face direct costs, as do those who have yet avoided it. Second, diseases potentially interact with the production of goods and service as well as the growth of the economy and its income per capita. We begin here with the costs to those current and potential sufferers. If an epidemic comes, then what is its true cost to society? A standard approach estimates the cost of illness (COI) as the product of prevalence times cost per illness

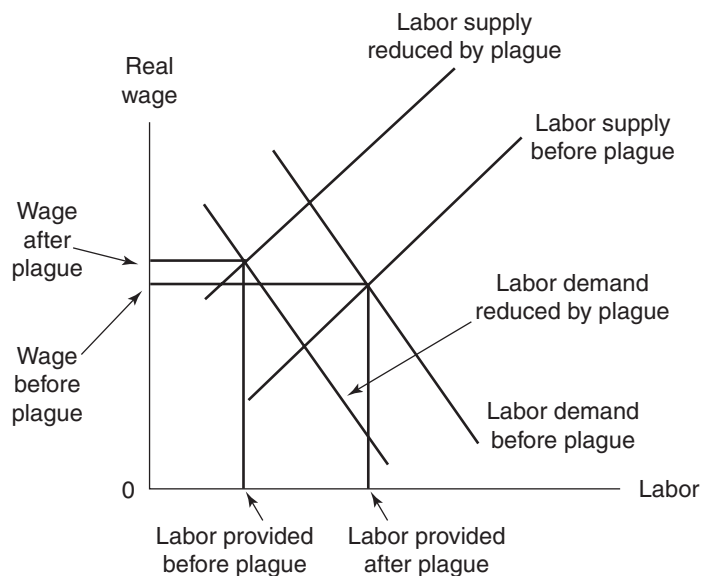
(both treatment and forgone earnings), based on severity of the disease. Philipson (1995), however, shows that this approach neglects the cost of patients' disease-avoidance behaviors. We spend money, time, and effort in reducing the risk that we face from the disease. We may choose to analyze a disease like we avoid a tax because the disease causes a loss of utility and naturally leads to efforts to avoid the disease. Treating an illness this way implies that the efforts at illness-avoidance are costs, and these avoidance costs should become part of the cost of illness calculation. Tax-avoidance costs are already treated as an economic loss, so why not illness-avoidance costs?

The relationships of epidemic illness to poverty, growth in GDP, and income per capita can be illustrated dramatically with the experience of AIDS in Africa. It is clear from available data of world experience that countries with lower income per capita experience greater HIV infection prevalence.

Thus, the experience of poverty and the related experience of AIDS are each more complex than is suggested by the world data. Poverty cuts two ways. The poor may have less information about the dangers of HIV infection and how to avoid it. Also, when AIDS occurs, the poor experience a much greater proportional burden in caring for the ill person (Bloom and Sevilla, 2002). Whether AIDS epidemics have actually slowed the growth of real wages is not nearly as clear. Work by Bloom and Mahal (1997) finds that there is no significant empirical relationship between the prevalence of AIDS across 50 countries studied and the growth rate in real income per capita.

Consider an analogy to the experience of Europeans with the Black Death, thought to have begun in 1348 and continuing for the next 200 years. The Black Death or "bubonic plague" spread a virus carried by rats; the infection was usually lethal. Consider Figure 25-3 with an initial demand and supply curve for labor. Before the plague, the laborers of that time produced agricultural products and crafted other goods. The heightened mortality eliminated portions of the general population, including a share of those laborers. The plague's effect was to reduce the supply of labor and shift that curve to the left. The general population reduction, however, also reduced the demand for the products of labor, and the derived demand for labor, therefore, was reduced (shifted leftward) as well.

The result of the plague is that the equilibrium quantity of labor employed, and the accompanying *output*, necessarily falls. However, we cannot determine from this analysis whether the *real wage* in equilibrium would necessarily rise or fall. The real wage here is analogous to the real income per capita, while the real output here is analogous to a country's real GDP. We have depicted



**FIGURE 25-3** The Effect of Plague on Real Wage

the real wage as increasing, and the standard result for studies of the Black Death corresponds to this picture. We note simply that neither a rise nor fall in the wage necessarily results from the plague, but that policymakers who argue that real wage has not changed as a result of a plague (as did some leaders in South Africa) are surely incorrect in their evaluation of the costs (in forgone output) of the plague.

One does not have to go back to the fourteenth century to look at the costs of plague. Some survivors of the “Great Influenza” of 1918–1919 (for an excellent account of this epidemic, and of medical practices at the time, see Barry, 2005) are still alive. Starting in Kansas, the disease spread in part by the mass movement of soldiers of the warring countries in World War I and by the unsanitary conditions in military camps. Experts believe that between 20 million and 100 million people (in a world population of 2 billion people) died in this outbreak.

There is no doubt that an epidemic can and does retard a country’s growth rate of GDP. While quandaries arise regarding the relation of total population to income per capita, it is clear that the massive loss of productive resources reduces total income. Similarly, there is no doubt that, should per capita income rise during and after a plague, it does so at a huge and irretrievable expense in the suffering and loss of life that occurs.

### **The Difficulty of Eradicating Diseases**

The excitement following the first report of the world eradication of smallpox in the 1980s was tempered by reports of the difficulty of achieving the same for other infectious diseases. Of about 40 vaccines available, only smallpox can claim this much success. It may surprise many that polio, for which major vaccination advances were introduced in the 1950s and 1960s, is still active in 2011, causing paralysis and death in wide swaths of Central Africa and the Indian subcontinent of Asia (Aylward and Yamada, 2011).

Complete eradication may often be an unwise goal. We earlier proposed that demand for prevention depends on disease prevalence. If so, the widespread application of an effective vaccine will reduce the prevalence of the disease, an essential step in the disease eradication process. However, when prevalence becomes sufficiently small, the demand for prevention may disappear. Why could this occur? What happens, in this theory, is that the benefits to vaccination can drop below any possible price of vaccination; the success of vaccination, somewhat ironically, will cause further vaccinations to lack sufficient benefits to cover the costs. The ensuing expansion of unvaccinated numbers, in turn, may make complete eradication unachievable.

### **Information**

Controlling epidemics, such as AIDS, requires screening, tracing of partners, and education about the nature and transmission of the disease. The problem is one of incomplete information. Susceptible people will improve their chances when learning of effective ways to avoid infection. Knowledge of the available technologies improves the chances people will seek the appropriate care. In some cases, the most effective technology may be among the least costly, with lack of information itself being the key bottleneck. In other cases, as noted in Box 25-2 people may act on the basis of inaccurate information.

Several interventions take their effect by shortening the incubation period of the infection. Incubation period begins at the onset of the disease and lasts until the disease generates observable symptoms. Information provided through accurate testing, for example, provides an early warning and shortens the incubation period. This generates beneficial results by speeding the process by which both the infected and the susceptible populations adopt protective and preventive behaviors.

The transmission of information can entail clashes of cultures; a medical culture, for example, based more or less on scientific practices and a population culture, based on different and even questionable modes of choosing behaviors. The choice of whether to urge intervention may entail the question of whether imposing one’s own culture is arrogant and interventionist, with the knowledge that suffering will continue, perhaps on a large scale, without such intervention.



**BOX 25-2****Study Linking Childhood Vaccine and Autism Was Fraudulent**

In the 1998 *Lancet*, Dr. Andrew Wakefield and colleagues reported on 12 children, eight of whom supposedly developed gastrointestinal trouble and “regressive autism,” a form of the disorder that strikes later in childhood, after getting a combination vaccine against measles mumps and rubella (MMR). No other researchers were able to replicate the study, and the *Lancet* retracted the study in 2010. Nonetheless, the publication, in a reputable journal, fueled suspicions by some parents that MMR vaccine was harmful and a 2008 *Pediatrics* study found that the rate of “selective MMR nonreceipt” had nearly tripled from 0.77% in 1995 to 2.10% in 2000.

A January 2011 National Public Radio (NPR) report by Scott Hensley looked at an article in the *BMJ*, the *British Medical Journal*, where investigative reporter Brian Deer argues that the *Lancet* study wasn’t just wrong—but *fraudulent* because key facts were altered to support the autism link

- Only one of nine children said to have regressive autism clearly had it. Three had no form of autism.
- Contrary to the paper’s assertion that all the children were normal before vaccination, five had some sort of pre-existing developmental problems.
- Behavioral problems the paper said occurred days after vaccination didn’t actually appear for months in some of the children, a fact that undercuts the causality of vaccination.

Moreover, Wakefield had not disclosed conflicts of interest to the *Lancet*, including that he received research funding from a lawyer representing parents who believed the MMR shot had harmed their children.

NPR reported that while more revelations about Wakefield are not likely to make the fear of vaccines go away, David Ropeik, a risk analyst affiliated with Harvard University, says something else eventually will “. . . as more and more people get measles and kids die, which is happening around the world. Eventually the threat of the disease will come back and surmount our fear of the vaccine.”

*Sources:* Smith et al (2008), Wakefield, Andrew J., et al (1998), Deer (2011).

**The Role of Government in Battling Epidemics**

The standard economic analysis proposes that collective action, generally involvement by government, may be justified when there is market failure or when issues of equity arise. Rationales of these kinds reasonably arise in the case of many epidemic diseases. First, where people can obtain immunity to the disease through vaccination or other means, and where the disease is contagious, economists commonly recommend that vaccinations (for example) be encouraged either through provision or subsidies. The reason is that private pay vaccination will tend to underestimate the external benefit of immunization; those immunized gain not only for themselves but also benefit others. By not becoming disease carriers, they reduce the risk to others. It is the classic example of an externality rationale for collective action.

Second, the optimality observed in models of competitive markets requires that information be adequate and symmetrically available to the market participants. Infectious disease commonly fails to fit this depiction. Even in sexually transmitted diseases (STD), partners may not recognize the risks that they face, either fully or at all. While many examples in life exist where people take on a risks knowingly, making private rational decisions, this picture is not as clear for AIDS and similar types of infections.

Third, one could make an equity case for some epidemic diseases. Equity may be violated, for example, should some segments of the populace suffer the disease more than others. It might also apply if observers believe that the disease has been forced upon the sufferer in some sense, partly by public actions, or inactions.

Fourth, Hilsenrath (1999) argues that the role of government in monitoring and controlling the overall economy forms an analogy to a potential role for government in epidemics like AIDS. Here, the government’s role to encourage growth and stability takes on a public health aspect inasmuch as raging epidemics have macroeconomic consequences.

Fifth, as we saw in Chapter 13, health economists have pointed out the market failures entailed in the provision of charity. Millions of Americans may experience the charitable motivation to send money

voluntarily to help halt AIDS in Africa, but the “free-rider effect” suggests that many of the potentially charitable people may choose not to donate, assuming that others will. The standard analysis here is that the potential donors gain from the healthful program in Africa even if they do not personally contribute.

The previous accounts explain economic rationales for government intervention on several grounds. They do not imply, however, that actual government interventions will necessarily be economically justified. For example, Philipson and Posner (1993) argue that the behaviors of epidemic sufferers matter substantially to the progress of an epidemic and to the wisdom of particular interventions. Should government encourage and provide AIDS testing? Intuition can run afoul with economic behavior here. Those testing negative may increase their risky behavior, or people testing positive may not care about infecting others. The examples add an extra condition to justifying government involvement—it must have a reasonable chance of being effective.

Philipson and Posner also argue that much of AIDS infection and contagion occurs through personal choice, plausibly even rational choice. To understand, predict, and control the epidemic may require the understanding and manipulation of “prices” and benefits that those involved face when making their choices. We may find means to encourage the dampening of prevalence growth by understanding the role of this economic behavior.

---

## CASE STUDY HIV/AIDS in Africa

### HIV/AIDS<sup>1</sup>

The term AIDS (auto-immune deficiency syndrome) first appeared in the *Morbidity and Mortality Weekly Report* (MMWR) of the Centers for Disease Control (CDC) in 1982 to describe “a disease, at least moderately predictive of a defect in cell-mediated immunity, occurring with no known cause for diminished resistance to that disease.” AIDS is caused by infection with the human retroviruses HIV-1 or -2. These viruses are passed through sexual contact; through contact with blood, blood products, or other bodily fluids (as in drug abusers who share contaminated intravenous needles); intrapartum or perinatally from mother to infant; or via breast milk. There is no evidence that the virus can be passed through casual or family contact or by insects, such as mosquitoes.

By 2008, as noted in Figure 25-4 almost 1.2 million people were living with AIDS in the United States. The annual number of deaths peaked at almost 50,000 in 1995, and has since leveled off at less than 20,000. As of the end of 2008, the cumulative number of deaths was over 590,000.

The data in Table 25-1 show that males make up about three-quarters of the current cases. The groups with the highest numbers and the highest rates are those between the ages of 25 and 34, and 35 and 44. African Americans have HIV/AIDS at a rate of 1,819 per 100,000 people, or over 7.6 times as high a rate as whites.

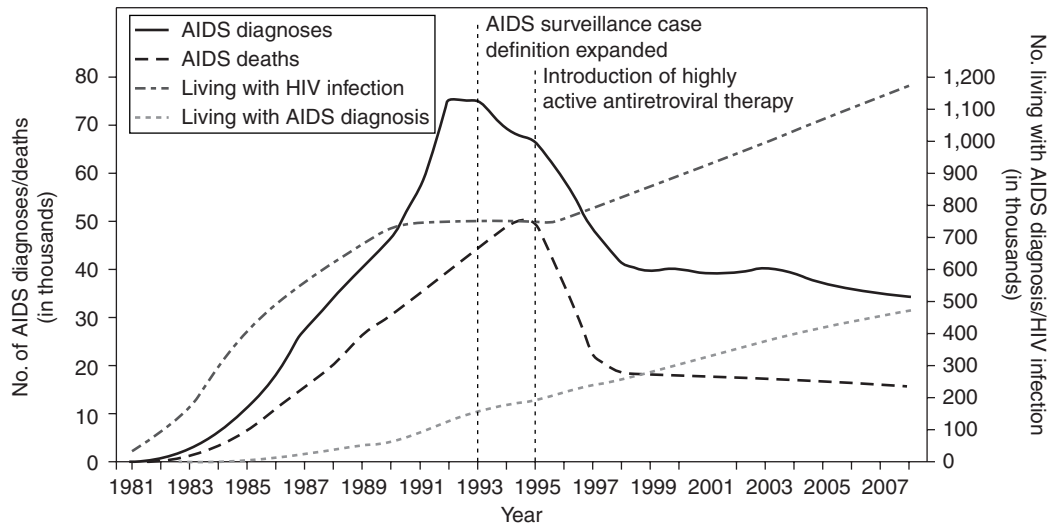
Globally, as shown in Table 25-2, HIV/AIDS is a pandemic, especially in developing countries. The current estimate of the number of cases of HIV infection worldwide is 33.3 million, with almost 22 million in Africa and another 6.1 million in Asia. Of the worldwide cases, almost half (47.7 percent) are women.

Infections that take advantage of a weakened immune system include:

- Pneumocystis carinii pneumonia
- Toxoplasmosis
- Tuberculosis

---

<sup>1</sup> Material in this section comes from Fauci, Anthony S. and H. Clifford Lane, Human Immunodeficiency Viral Disease, AIDS and Related Disorders, *Harrison's Principles of Internal Medicine*, 16th ed., 2005, [www.scribd.com/doc/509117/Harrisons-Principles-of-Internal-Medicine-16th-Edn?query2=internal%20medicine%20manual%20free%20download](http://www.scribd.com/doc/509117/Harrisons-Principles-of-Internal-Medicine-16th-Edn?query2=internal%20medicine%20manual%20free%20download), accessed July 12, 2008.



**FIGURE 25-4** Annual and Cumulative Rates of HIV/AIDS infections up to 2008

Source: Centers for Disease Control, "HIV Surveillance – United States, 1981–2008", Weekly June 3, 2011 / 60(21): 689-693, Figure, <http://www.cdc.gov/mmwr/preview/mmwrhtml/mm6021a2.htm>, accessed July 7, 2011.

- Extreme weight loss and wasting, exacerbated by diarrhea that is experienced in up to 90 percent of HIV patients worldwide
- Meningitis and other brain infections
- Fungal infections
- Syphilis
- Malignancies, such as lymphoma and cervical cancer
- Kaposi's sarcoma, which affects the skin and mouth and may spread to the lungs

Many of the economic theories just described take on a great reality in the context of the current African AIDS epidemic. While each continent has been infected, we will see that the epidemic in Africa has been extraordinarily bad, making it a reasonable focus of policy by those intent on stopping the worldwide AIDS epidemic.

With well over 20 million dead, projections of future human and material loss are disturbing and continue to worsen. Drugs and drug cocktails that stem the progress of the disease now exist. Human behavioral change, which might follow substantial improvements in the transmission of information, offers hope as well. The situation in Africa is so dire, however, that many propose that the response from the West must be stronger and sustained.

### Costs of AIDS

Another way of assessing both the prevalence of AIDS and its consequences for human life is to assess its costs in dollars lost. There are numerous ways to measure valuations of life. Some look at foregone productivity, which produces a very conservative value, because, for example, retired people would be valued at 0. More generous versions of valuation of life for the United State (Viscusi and Aldy, 2003) range from \$3 million to \$7 million dollars. Deaton and colleagues (2008) discuss the difficulties in measuring such numbers for sub-Saharan Africa.

Let us assign a (very low) value of \$100,000 per full statistical life lost to HIV/AIDS, and assume that it declines with age, so the average value is \$50,000. As noted in Table 25-2, South Africa (with a GDP of \$500 billion in 2009) has had 310,000 deaths annually. Evaluated at \$50,000 per year, this gives a loss of \$15.5 billion per year, or approximately 3 percent of the GDP.

**TABLE 25-1** Estimated number and rate of persons aged  $\geq$ 13 years or older living with HIV infection, and number and percentage whose HIV infection was undiagnosed, by selected characteristics—United States, 2008\*

Characteristic	Total persons living with HIV infection				Persons whose HIV infection was undiagnosed		
	No.	(95% CI)	Rate <sup>†</sup>	(95% CI)	No.	(95% CI)	%
<b>Total</b>	<b>1,178,350</b>	<b>(1,128,350–1,228,500)</b>	<b>469.4</b>	<b>(449.5–489.4)</b>	<b>236,400</b>	<b>(224,900–247,900)</b>	<b>20.1</b>
<b>Sex</b>							
Male	883,450	(841,450–925,450)	719.5	(685.3–753.7)	182,450	(172,450–192,450)	20.6
Female	294,900	(269,900–319,900)	230.0	(210.5–249.5)	53,950	(47,950–59,950)	18.3
<b>Age group (yrs)</b>							
13–24	68,600	(56,000–80,600)	134.1	(109.5–157.6)	40,400	(35,400–45,400)	58.9
25–34	180,600	(160,600–200,600)	440.9	(392.1–489.8)	56,800	(51,300–62,300)	31.5
35–44	357,500	(327,500–387,500)	846.3	(775.3–917.4)	64,300	(58,300–70,300)	18.0
45–54	385,400	(353,400–417,400)	871.3	(798.9–943.6)	53,200	(48,200–58,200)	13.8
55–64	147,700	(132,770–162,770)	439.3	(394.9–484.1)	17,600	(15,600–19,600)	11.9
$\geq$ 65	38,400	(34,400–42,400)	99.0	(88.7–109.3)	4,100	(3,600–4,600)	10.7
<b>Race</b>							
American Indian/Alaska Native	5,000	(3,500–6,500)	268.8	(188.2–349.4)	1,250	(650–1,850)	25.0
Asian/Pacific Islander	16,750	(14,250–19,250)	147.0	(125.0–168.9)	4,350	(2,850–4,850)	26.0
Black/African American	545,000	(513,000–577,000)	1819.0	(1,712.2–1,925.8)	116,750	(108,650–124,850)	21.4
White	406,000	(378,000–434,000)	238.4	(221.9–254.8)	75,200	(70,700–81,700)	18.5

Source: Centers for Disease Control, “HIV Surveillance—United States, 1981–2008”, Weekly June 3, 2011 / 60(21): 689-693, Table, <http://www.cdc.gov/mmwr/preview/mmwrhtml/mm6021a2.htm>, accessed July 7, 2011.

<sup>†</sup>Per 100,000 population.

**TABLE 25-2 Largest Losses Due to HIV/AIDS, 2009**

	Number with HIV/AIDS	Annual Deaths	% of Total Deaths	Adult (15–49) prevalence %	Loss @ 50,000/death (in \$ millions)
South Africa	5,600,000	310,000	18.3	17.8	15,500
Nigeria	3,300,000	220,000	13.0	3.6	11,000
India	2,400,000	170,000	10.0	0.3	8,500
Tanzania	1,400,000	86,000	5.1	5.6	4,300
Zimbabwe	1,200,000	83,000	4.9	14.3	4,150
Kenya	1,500,000	80,000	4.7	6.3	4,000
Mozambique	1,400,000	74,000	4.4	11.5	3,700
Uganda	1,200,000	64,000	3.8	6.5	3,200
Malawi	920,000	51,000	3.0	11.0	2,550
Zambia	980,000	45,000	2.7	13.5	2,250
Cameroon	610,000	37,000	2.2	5.3	1,850
Cote d'Ivoire	450,000	36,000	2.1	3.4	1,800
Thailand	530,000	28,000	1.7	1.3	1,400
China	740,000	26,000	1.5	0.1	1,300
Ukraine	350,000	24,000	1.4	1.1	1,200
Ghana	260,000	18,000	1.1	1.8	900
Burma	240,000	18,000	1.1	0.6	900
United States	1,200,000	17,000	1.0	0.6	850
Russia	980,000	16,660 *	1.0	1.0	833
Burundi	180,000	15,000	0.9	3.3	750
Lesotho	290,000	14,000	0.8	23.6	700
Vietnam	280,000	14,000	0.8	0.4	700
Colombia	160,000	14,000	0.8	0.5	700
Sudan	260,000	12,000	0.7	1.1	600
Mexico	220,000	12,000 *	0.7	0.3	600
Chad	210,000	11,000	0.6	3.4	550
Angola	200,000	11,000	0.6	2.0	550
Central African Republic	130,000	11,000	0.6	4.7	550
Indonesia	310,000	8,300	0.5	0.2	415
Togo	120,000	7,700	0.5	3.2	385
Haiti	120,000	7,100	0.4	1.9	355
Burkina Faso	110,000	7,100	0.4	1.2	355
Swaziland	180,000	7,000	0.4	25.9	350
Namibia	180,000	6,700	0.4	13.1	335
Botswana	320,000	5,800	0.3	24.8	290
Rwanda	170,000	4,100	0.2	2.9	205
Argentina	110,000	2,900	0.2	0.5	145
France	150,000	1,700	0.1	0.4	85
Spain	130,000	1,600	0.1	0.4	80
Italy	140,000	1,000	0.1	0.3	50
Total for Listed Countries	29,230,000	1,578,660			78,933
Total for Remaining Countries	4,070,000	116,017			5,801
Total	33,300,000	1,694,677			84,734

Source: CIA - *The World Factbook*, 2009, various tables; <https://www.cia.gov/library/publications/the-world-factbook/>, accessed July 7, 2011.

\* Estimated

Zimbabwe with a GDP of \$5 billion has had 83,000 deaths per year, with a loss of \$4.3 billion, which equals over 85 percent of its GDP. Summing these losses for the countries around the globe gives an aggregate loss of \$84.7 billion. This is almost certainly a severe underestimate because it values, for example, the 17,000 lives lost in the United States at only \$50,000 apiece.

### Fighting AIDS

Economic analysis can help to identify the reasons why halting this epidemic will benefit from strong responses from the private, the nonprofit, and the public sectors of these economies. However it especially explains why private and voluntary market interactions may not stop AIDS in Africa, and why it has often made things much worse.

Sexual practices that were common and acceptable in the culture became deadly with the introduction of the AIDS virus. Similarly, common economic practices, such as traveling long distances to look for work or as a part of one's work, became ready means for the virus to travel. When local jobs disappeared, sometimes due to side effects of the crisis, then the geographic search for work outside of the local area probably made the epidemic even worse. Note the similarity in terms of travel to the Spanish flu during World War I where traveling soldiers transmitted the virus.

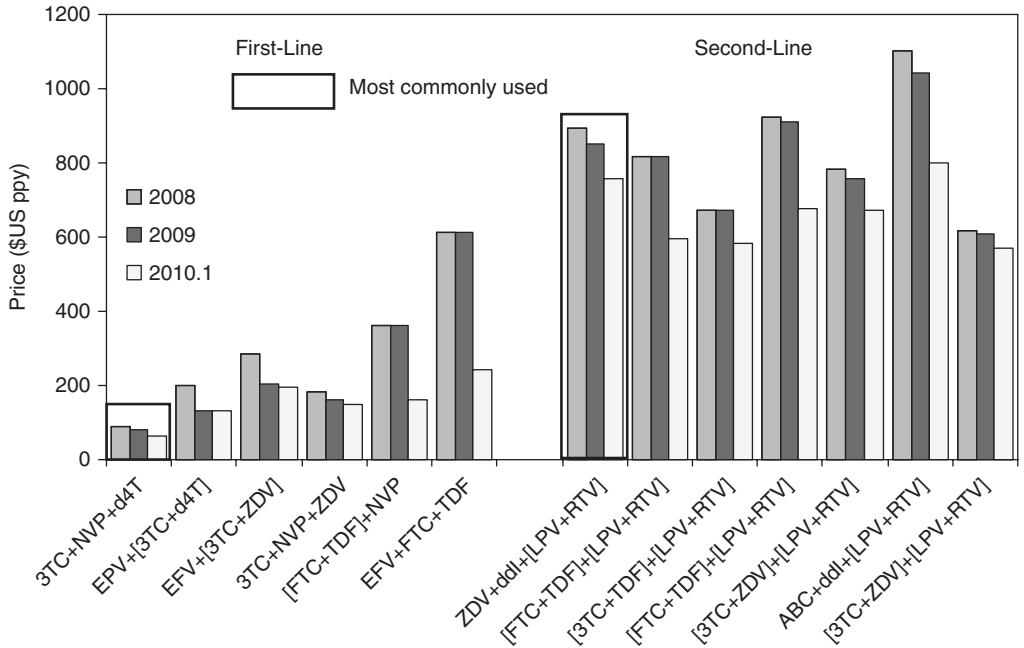
A two-pronged attack on the problem can bring substantial success. Before 2001 drug treatments for HIV/AIDS were effective although extremely costly, with the annual cost of a three-drug combination antiretroviral (ART) regimen for a patient in a poor country approximately \$10,000 to \$25,000. At the time, noted Over (2004), only a few of the richest developing countries, such as Brazil and Thailand, could attempt to finance ART for their AIDS patients. In most developing countries, the only patients receiving ART were the very rich or those who had access to rationed, low-price supplies through a variety of pilot or research projects.

Since 2003, there has been a push to offer antiretroviral treatment to people living with HIV/AIDS in resource-poor settings. Today, the World Health Organization (WHO) estimates that over 1 million people are on ART in the developing world. Treatments are categorized as the less expensive or "first-line" regimens, and the more expensive "second-line" regimens. The median price of medicines for adult major first-line regimens continued to decrease in low-income countries (LIC) between 2008 and March 2010. Within that period, the median price of the most commonly prescribed regimen for adults use: 3TC+NVP+d4T (150+200+30 mg), dropped from (US)\$88 in 2008 to \$64 (having dropped from \$153 in 2004 to \$92 in 2007), as noted in Figure 25-5. Other first-line regimens have also fallen in price.

However, experience from countries where ART has long been widely available, such as Brazil, the United States, or countries in Europe, shows that after a few years, the first-line of ART no longer works for many patients, who must switch onto a second-line regimen. Between 2008 and March 2010, the median price of the most commonly prescribed second-line regimen for adult use—ZDV+ddI+[LPV+RTV], 300mg+400mg+[200+50]mg—dropped from \$892 to \$759, but it is still almost 12 times as high as the most commonly used first-line treatment. For countries with limited budgets, determining how many, if any, patients get the second-line treatment will remain an important policy decision.

Information forms the other major branch in the attack on AIDS in Africa. Low-cost prophylactic measures were not commonly used when the epidemic hit, nor were the causes of the epidemic widely known or understood.

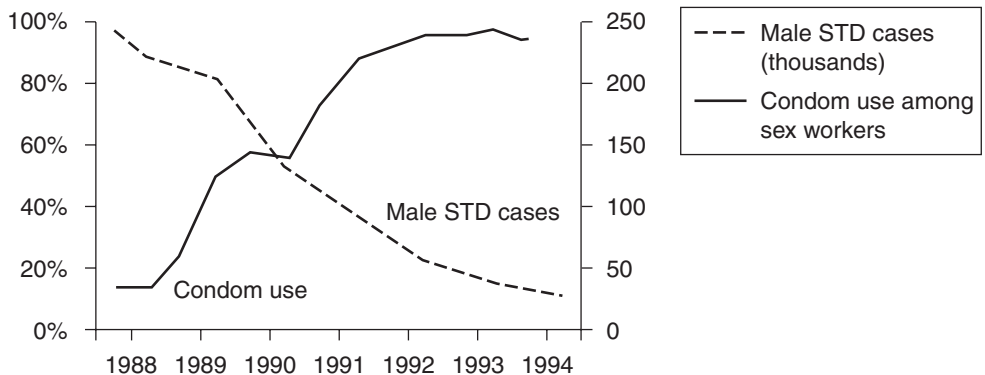
Figure 25-6 demonstrates one of the key production processes that apply to stopping AIDS: the distribution and use of latex condoms. Individuals participate in the production of their own health. Many of the material inputs are simple and inexpensive, but it is the information that makes the health gain possible. The figure shows that increased condom usage in Thailand accompanied a decline in the incidence of STDs. The striking effectiveness of condom use in this case is a cause for optimism. However, Westerners attempting to help must confront local cultures and beliefs that may stand in the way of prevention. It is legitimate to ask if we can justify Westerners attacking "superstitions" in the means they use to attack the spreading virus.



**FIGURE 25-5** ART Drug Prices: First- and Second-Line  
 Source: Global Price Reporting Mechanism, "Transaction prices for Antiretroviral Medicines and HIV Diagnostics from 2008 to March 2010," World Health Organization. Geneva 2010.

The approach of providing scientifically valid information, however, has been applied in practice in Africa and with indigenous support. These two branches of the attack on AIDS in Africa suggest the need for collective action through government because information is substantially a public good, in which its marginal cost of transmission is extremely low once the infrastructure has been established. Furthermore, it is hard to sell information in a private market because it is economically infeasible to limit the information to those who pay.

A similarly important public role follows from an economic analysis of both the detrimental and the beneficial externalities involved. Risky sexual behaviors involve risks to more people than just the individual making the risky choice. Economic logic implies that the effective price of risky behaviors must go up, and government may be instrumental in raising it. The government can intervene to



**FIGURE 25-6** The Rising Impact of Condom Use in Thailand  
 Source: Confronting AIDS: Public Priorities in a Global Epidemic, World Bank, 1997. Source: Based on information from confronting AIDS: Public Priorities in a Global Epidemic, World Bank, 1997.

reduce the effective price of safe sexual behavior, for example, by providing condoms freely yet with social discretion.

A beneficial externality occurs to the millions of people worldwide who feel a charitable gain in the gains to Africans who keep or regain health. This charitable externality can in principle occur through voluntary charities, and we can hope that this proves an important source. Collective provision of the aid to Africa can also reduce free-riding behaviors.

### Economic Theory and African Reality

Two strands of health economic theory bear most critically on the African experience. The first among these must be the role of information. Affordable technologies, such as the use of condoms, are available in principle but they must be adopted to become effective. In many cases, the local cultural patterns in some African countries involve either a nonscientific understanding of AIDS or cultural relationships between the sexes that have minor effects during times when the risk of infection is low, but have potentially disastrous effects during an ongoing epidemic. To intervene with reliable, scientific information may be costly when it disturbs the local culture. In this sense, it could bear a cost similar to the costly AIDS drugs that must also play an important role. However, information has proved to be an effective weapon against AIDS in some cases where it is vigorously applied.

Secondly, the concept of prevalence elasticity of demand for preventive goods and services may also be very important in Africa. We have seen that a sufficiently large elasticity of demand for preventive care can result in a prevention response so large that growing prevalence of disease can result in a lower rate of new infections. Africans have not always responded to growing evidence of the AIDS epidemic with behaviors that embrace preventive care. If, as this suggests, the prevalence elasticity of the demand for preventive care is very low, it portends a growing epidemic. Information could play a double role by encouraging the responsiveness of preventive demand to further evidence of a growing prevalence of the disease.

Third, African culture is often not hospitable to Western medical treatments and practices. The Republic of South Africa, one of the wealthiest countries on the continent, has lost between 300,000 and 400,000 people to HIV/AIDS each year for several years. Yet as noted in Box 25-3, effective treatment often loses out to cultural obstacles.

#### BOX 25-3

#### HIV/AIDS in South Africa: Treatment Meets Culture

On March 13, 2008, *The Economist* reported on a book by Jonny Steinberg, *Sizwe's Test: A Young Man's Journey Through Africa's AIDS Epidemic*. Steinberg spent months trying to understand what prevents millions of South Africans from even testing for their status and, if necessary, seeking treatment.

Sizwe, a 29-year-old from the rural district of Lusikisiki in the Eastern Cape, lives close to one of the country's most successful treatment programs but he refuses to test, in spite of losing his best friend to AIDS, having access to medication, and watching those on treatment recover their health. "Some people have maybe sent a demon to have sex with me: a demon with HIV. That is why I am scared to test," he says. His distrust of white doctors also feeds the fear of testing. When an internationally acclaimed group, Doctors Without Borders, arrived in the area, explains a volunteer, people believed that the white doctor had come to kill people with his needle and blood test. "They believed AIDS was caused by politics, by white people."

Then-President Thabo Mbeki's dissident views on AIDS and antiretroviral drugs reflected similar fears. A local government official told Mr. Steinberg that he did not trust the drugs because they are not African. "Mbeki did not fabricate the old man's paranoia," remarks Steinberg, "but he did draw it to the surface of South Africa's political culture."

For Sizwe, who was about to marry the mother of his son, testing positive would also destroy a future constructed on the remains of a traditional culture that commands marriage and heirs. The young man argues that knowing that his blood is "dirty," with a virus that can be tamed but not killed, would push him to a fast death.

Sources: [www.economist.com/books/displaystory.cfm?story\\_id=10843131](http://www.economist.com/books/displaystory.cfm?story_id=10843131), accessed July 29, 2008. Also, Steinberg, Jonny, *Sizwe's Test: A Young Man's Journey Through Africa's AIDS Epidemic*, Simon & Schuster, 2008.



**TABLE 25-3 Cost Effectiveness of HIV/AIDS Interventions**

	Change Transmission Rates Through STI Treatment <sup>a</sup>	Change Behavior Through Education <sup>b</sup>
Cost	\$1,068,326,724	\$2,800,263,630
Life-years	291,338,142	166,476,290
DALYs	311,746,543	176,081,992
Infections averted	13,655,027	6,409,195
Cost per LY	\$3.67	\$16.82
Cost per DALY	\$3.43	\$15.90
Cost per infection	\$78.24	\$436.91

All costs include both delivery costs and drug costs.

<sup>a</sup> Treatment of bacterial STIs, including syphilis, gonorrhea, chancroid, and others.

<sup>b</sup> Intervention to decrease number of sexual partners.

Source: Figures from Oster (2005).

We finish this section by looking at some provocative work by Emily Oster who in a 2005 article addresses the question of why HIV affected Africa so heavily, and what explains the variation within Africa. Her model decomposes epidemic level into differences in sexual behavior and differences in viral transmission rates. She argues, using evidence drawn from the existing medical literature, that Africa has very high HIV transmission rates, likely due to high rates of other untreated sexually transmitted infections, while transmission rates in the United States are low. The difference in transmission rates is large enough to explain the observed difference in prevalence between the United States and sub-Saharan Africa. The model also provides a good fit to cross-country data within Africa and suggests that differences within that continent can be attributed to differences in sexual behavior and epidemic timing.

Table 25-3, taken from her research, compares the cost effectiveness of two possible interventions. The first, following a pilot study in Tanzania, is designed to decrease transmission rates by treating other untreated (bacterial) sexually transmitted infections. The second considers a scaled-up version of the Ugandan experience in which, through advertising and educational campaigns, Uganda appears to have decreased most aspects of risky sexual behavior; estimates suggest that the HIV prevalence there has gone down substantially.

Oster examines these interventions in the model for Africa overall, assuming that both interventions achieve their entire effect in the first year. She begins by simulating the model under the case of no intervention and then for each intervention separately. Using information on current population at each age, she calculates measures of effectiveness from the model output. These include life-years saved, disability-adjusted life-years (DALYs) saved, and overall infections averted.

The table reports effectiveness and cost effectiveness. The transmission rate intervention would save 291 million life-years with 13.6 million infections averted (around 25 percent of the total infections over the next decade), at a cost of \$3.67 per life-year and \$78.24 per infection. The sexual behavior intervention is less effective, preventing 6 million infections at a cost of \$16.82 per life-year and \$436.91 per infection. She argues further that in addition to greater cost effectiveness, an intervention designed to treat other STIs is more readily generalized from one country to another than one designed to change sexual behavior. She believes it reasonable to infer the effect of treating STIs in one country based on the experience of another, but she argues that changing sexual behavior may be more culturally specific, and therefore the experience of one country may not be transferable to others.

## CONCLUSIONS

Economic epidemiology provides new theoretical ideas, speculative applications, and studies spurred on by a new urgency caused by the AIDS epidemic in progress across the world. Economic theory helps inform epidemiological theory by developing the dynamic role of the elasticity of the demand for preventive care as it responds to a growing prevalence of the disease.

Epidemics have economic consequences most often found in the reduction of the growth rate of GDP, especially through the loss of manpower and the diversion of economic activity away from more productive efforts. It is unclear, however, whether an epidemic will tend to lower per capita income.

Information theory plays a critical role in epidemics in that the nature of the disease in scientific terms must become clear in order for the demand for prevention to overcome the progress of the disease. The chief technology is the vaccine. Where available, would-be users must see that the vaccine's benefits overcome its costs.

Africa provides a sobering case study on the effects of the AIDS epidemic. Issues of information and the need to induce prevention responses among susceptible people are foci of attention, as is the treatment of AIDS sufferers. Changes in personal behaviors, such as the use of condoms, can involve small capital costs but have substantial effects. The epidemic meets the economic rationales for the involvement of collective action, and those speaking up most often about the problems of Africa usually emphasize the need for all three—government, private markets, and nonprofits—to play a part.

---

## Summary

1. Incidence is the number of new cases of the disease relative to the population. The hazard rate is the instantaneous rate of infection. Prevalence is the total number of cases per the population.
2. The prevalence elasticity of demand for preventive care is the percentage increase in the quantity of preventive care demanded per a percentage increase in the prevalence of the disease.
3. The standard epidemiological model predicts that an increase in the prevalence of an infectious disease will tend to result in an increase in the incidence rate of the disease.
4. In the economic model, a sufficiently high prevalence elasticity of demand for preventive care can lead to a decrease in the incidence of the disease.
5. Contracting an infectious disease is analogous to being affected by a tax in that it leads to costly responses to avoid the harm. These avoidance costs are logically part of the cost of illness.
6. The difficulty in eradicating an infectious disease can be explained by an economic model. When a sufficiently lowered prevalence of the disease leads people to demand zero preventive care, its benefits no longer outweigh its costs.
7. AIDS in Africa has resulted in large losses of human resources in many countries. In countries such as Zimbabwe, annual losses due to AIDS equal over 85 percent of the country's GDP.
8. Methods to curb AIDS in Africa have included new drugs and efforts to get reduced prices for these drugs. They have also included effectively providing simple and inexpensive technologies such as the distribution of condoms.
9. Prices of HIV/AIDS drug treatments have plummeted in the past 15 years with the least expensive "first-line" treatments now costing less than \$100. Budgetary constraints continue to limit the utilization of these and more expensive "second-line" treatments.
10. Both the economic culture and the social culture revealed serious problems in dealing with the new epidemic.
11. Problems of information are central to most approaches toward stopping the AIDS epidemic in Africa.

## Discussion Questions

1. Identify several examples of preventive goods and services that are relevant to reducing the risk of infection from a disease that you name.
2. Newspaper accounts are available of the “Great Influenza” of 1918–1919. Using archival sources, try to delineate the economic costs of influenza for specific U.S. cities, in terms of treatment, and lives lost.
3. Philipson’s model describes the role of the prevalence elasticity of demand for preventive care. Does this imply that the epidemic is self-limiting, so that we need do nothing about it?
4. How might information regarding AIDS, its method of transmission, and the means to stop the epidemic be conveyed to people in Africa?
5. Is the complete eradication of all infectious diseases from the globe a wise objective? Is it a feasible objective?
6. How should drug companies price AIDS drugs for Africans? What the market will bear? The cost of producing the drugs? No charge whatsoever?

## Exercises

1. Suppose that the prevalence elasticity of demand for prevention was 0.1. What would be the impact of a 10 percent increase in prevalence of a contagious disease? Why?
2. Provide an explanation for why increasing income inequality is associated with greater prevalence of HIV infection in Africa.
3. Suppose that the incidence of AIDS,  $I$ , is determined by the prevalence of AIDS,  $A$ , and the demand for preventive care,  $P$ , as in the equation:

$$I = kA^{0.03}P^{-0.01}$$

where  $k$  is a constant. Let preventive care demand, in turn, be a function of the prevalence of AIDS, so that  $P = gA^{0.4}$ . By substitution, find the incidence solely in terms of  $A$ , and determine from this whether incidence rises or falls with prevalence  $A$ .

4. If asked to determine where to spend the next million dollars on curbing AIDS in Africa should go, would you recommend high-tech drug treatments or relatively low-tech behavioral practices? Sketch out the necessary categories of costs per case to assess the problem in a cost-effectiveness format.
5. We have seen that there was a major decline in the price of drugs for HIV/AIDS between 2008 and 2010.
  - (a) Check various Web sources to find what has happened to the prices since 2010.
  - (b) With the fall in the price of drugs, enumerate the other costs that stand in the way of larger-scale treatment for HIV/AIDS in Africa.
6. Table 25-2 calculates costs of HIV/AIDS in terms of value of life. Picking an individual country, determine what other costs, including treatment, and reduced productivity to add for a full measure of country-specific costs. Use secondary sources to try to estimate the full cost of HIV/AIDS.

# GLOSSARY

**Actuarially Fair** Insurance under which expected payouts equal the premiums paid by beneficiaries.

**Adverse Selection** A situation often resulting from asymmetric information in which individuals are able to purchase insurance at rates that are below actuarially fair rates plus loading costs.

**Agency Relationship** A situation in which one person (agent) makes decisions on behalf of another person (principal).

**Alternative Delivery System (ADS)** Insurance and organizational arrangements for health care delivery that are alternatives to traditional fee-for-service (FFS) arrangements.

**Asymmetric Information** Situations in which the parties on the opposite sides of a transaction have differing amounts of relevant information.

**Average Cost** Total cost represents the sum of all fixed costs and variable costs in the short run. Average cost equals total cost divided by the quantity of output and also equals the sum of average variable cost (AVC) and average fixed cost (AFC). In the long run, average total cost represents the minimum possible cost per unit of producing any given level of output when there are no fixed costs.

**Balance Billing** The practice of collecting the difference between the charge and the insurance reimbursement from the patient.

**Barriers to Entry** Impediments to the unrestricted flow of factors into or out of an industry or occupation (e.g., control over natural resources, licensure, patents).

**Body Mass Index (BMI)** A measurement of tendency toward obesity,  $BMI = \frac{\text{Weight in kilograms}}{(\text{height in meters})^2}$ . Current usage defines a BMI over 25 as overweight, with BMI greater than 30 as obese.

**Budget Constraint** The line representing combinations of goods that the consumer is just able to afford.

**Capitation** A method of reimbursement in managed care plans in which a provider is paid a fixed amount per person over a given period regardless of the amount of services rendered.

**Cardinal Utility** A quantitative measure of the value of a good in terms of metrically measurable utility. It is used in the study of risk and insurance.

**Case-Mix Index** A numerical measure of the assortment of patient cases treated by a given hospital, so that a higher value indicates a greater average degree of complexity of the cases.

**CAT (Computerized Axial Tomography)** Diagnostic equipment that produces cross-sectional images of the head and/or body.

**Centers for Medicare & Medicaid Services (CMS)** An agency of the U.S. Department of Health and Human Services

(DHHS). CMS is responsible for administering the financing and quality assurance programs for Medicare and the federal participation in Medicaid; formerly the Health Care Financing Administration (HCFA).

**Certificate-of-Need (CON)** Regulations that require health care providers to obtain approval from state planning agencies for capital expenditures that exceed various threshold levels (e.g., \$500,000).

**Ceteris Paribus** *Latin*, Other things being held constant.

**CHIP** A program administered by CMS that provides matching funds to states for health insurances for families with children. The program was initially designed to cover uninsured children in families with incomes that are modest, but too high to qualify for Medicaid. Formerly known as SCHIP.

**Coefficient of Variation** A measure of dispersion equal to the standard deviation divided by the mean (and sometimes multiplied by 100).

**Coinsurance (Rate)** The share of costs paid by the beneficiary of a health policy (often after some deductible).

**Community Rating** The practice of setting insurance premiums based on the utilization pattern of a broad population in a region. This approach to rate setting contrasts with experience rating.

**Comparative Statics** The analysis that calculates the level of a new equilibrium given changed values of one or more economic parameters, such as prices or income.

**Competition** (*See* Perfect Competition)

**Concentration Ratio** The share of the market sales or production accounted for by a certain number of the largest firms. Often the four-firm ratio is used.

**Consumer-Directed Health Plan (CDHP)** A high-deductible health plan (HDAP) coupled with a tax-advantaged health spending account (HSA or HRA). Consumers are provided with information and tools to help with health care service and financing decisions.

**Copayment** An amount paid out-of-pocket by the insurance beneficiary as a result of coinsurance and deductibles.

**Cost-Benefit Analysis (CBA)** A method of comparing the monetary value of all benefits of a social project with all costs of that project.

**Cost-Effectiveness Analysis (CEA)** A method that tries to find the least-cost method of achieving a desired objective(s) associated with a social project.

**Cost Sharing** (*See also* Copayment) Methods of financing health care that require some direct payments for services by patients.

**Cost Shifting** The practice by suppliers of increasing charges from some payers to offset uncompensated care costs and lower net payments from other payers.

**Cost-Utility Analysis (CUA)** A form of cost-effectiveness analysis in which outcomes, such as quality-adjusted life-years (QALYs), reflect the quantity and quality of life.

**Cross (Price) Elasticity of Demand** (*See also* Elasticity) The percentage change in the quantity demanded of one good resulting from a 1 percent change in the price of another good.

**Crowd Out** A response to the introduction or enhancement of public insurance, referring to the extent that those who have previously used private coverage now use public coverage.

**Deductible** The amount of health care charges for which a beneficiary is responsible before the insurer begins payment.

**Demand Function** The relationship between quantity demanded and price (and other independent variables, such as income and tastes). One could study individual demand as well as market demand.

**Depreciation** The change in the value of a good over time, due to deteriorating physical characteristics or technical obsolescence.

**Diagnosis Related Groups (DRGs)** A set of case types established under the prospective payment system (PPS) identifying patients with similar conditions and processes of care. In 2007–2008 CMS adopted a new set of 745 Medicare Severity Long-Term Care Diagnostic Related Groups (MS-DRGs) that replaced the existing 538 DRGs with ones that better recognized illness severity.

**Discount Rate** The interest rate used when converting sums to be received at a future date to a present value.

**Discounting** The process of converting sums to be received at a future date to a present value.

**Economic Profit** (*See also* Monopoly Profit) The return over and above that which is necessary to keep the firm from exiting the market over the long run. These profits are also called above-normal profits, excess profits, and supranormal profits.

**Economies of Scale** Situation in which the long-run average costs of a firm decline as output increases.

**Economies of Scope** Situation in which a firm can jointly produce two or more goods more cheaply than under separate production of the goods.

**Edgeworth Box (in Consumption)** A diagram that shows all possible allocations of fixed amounts of goods and services between two people.

**Efficiency** (*See also* Pareto Efficiency) Technical efficiency occurs when the firm produces the maximum possible sustained output from a given set of inputs. This idea is distinguished from allocative efficiency—situations in which either inputs or outputs are put to their best possible uses in the economy so that no further gains in output or welfare are possible. Both allocative and technical efficiency are prerequisites for Pareto efficiency.

**Elasticity** The percentage change in a dependent variable (e.g., quantity demanded) resulting from a 1 percent change in an independent variable (e.g., price). Elasticities that exceed 1 in absolute value are considered elastic; elasticities less than 1 are inelastic.

**Elasticity of Substitution** (*See also* Elasticity) The percentage change in the capital-labor ratio resulting from a 1 percent change in relative factor prices.

**Equilibrium Price (Quantity)** The price (quantity) at which the quantity demanded and the quantity supplied are equal.

**Expected Value** A measure used with a probability distribution of returns. The expected value is the sum of each probability multiplied by its corresponding return.

**Experience Good** A good for which evaluation is difficult prior to experience or purchase.

**Experience Rating** The practice of setting insurance premiums for an individual or group based on historical experience or risk associated with the individual or group.

**Externality** A case in which a consumer (producer) affects the utility (costs) of another consumer (producer) through actions that lie outside the price system.

**Fee-for-Service (FFS)** A method of payment under which the provider is paid for each procedure or service that is provided to a patient.

**Fee Schedule** A listing of fees by third-party payers showing the maximum amounts they will reimburse for specific services or procedures.

**Firm** Any entity that transforms inputs to some product or service that is sold in the marketplace.

**First Fundamental Theorem of Welfare Economics** The proposition that under specified conditions, competitive markets lead to Pareto efficient results.

**Fixed Costs (TFC and AFC)** Costs that do not vary with output. They are expressed either as total fixed cost (TFC) or average fixed cost (AFC).

**Formulary** A list of drugs developed by a managed care plan. Under a positive formulary, prescriptions on the list are covered. Under a negative formulary, prescriptions on the list are not covered.

**Frontier Analysis** A statistical analysis of firm efficiency that attempts to identify the best possible production practice and interprets firm inefficiency as a departure from the best possible production practice or frontier.

**Game Theory** A model that analyzes economic behavior as a series of strategic moves and countermoves by rival agents.

**Gatekeeper** The primary care provider who is responsible for coordinating a patient's care in a managed care plan. Often the gatekeeper must authorize referrals to specialists and nonemergency hospital admissions.

**Gross Domestic Product (GDP)** The market value of final goods and services produced within the borders of a country over a period of one year.

**Group Insurance** An insurance contract in which employees or members of a group are covered by a policy issued by their employer or group.

**Hazard** In epidemiology, the instantaneous rate of infection.

**Health and Health Status** (*See also* Morbidity Rate and Mortality Rate) The measures of the physical and emotional well-being of an individual or a defined population. Mortality and morbidity rates are often used to measure health status.

**Health Care** Goods and services used as inputs to produce health. Some analyses consider people's own time and knowledge used to maintain and promote health, in addition to conventional health care inputs.

**Health Insurance Portability and Accountability Act (HIPAA)** Federal legislation enacted in 1996 to protect the portability and continuity of health insurance coverage for workers who change or lose their jobs. The act requires hospitals, doctors, and insurance companies to share patient medical records and personal information on a wider basis to combat waste and fraud. The act also contains privacy provisions to protect the confidentiality of identifiable health data.

**Health Maintenance Organization (HMO)** A managed care plan that integrates financing and delivery of a comprehensive set of health care services to an enrolled population. HMOs may contract with or directly employ health care providers.

**Health Plan Employer Data and Information Set (HEDIS)** A set of standardized measures to evaluate health plan performance. HEDIS is used by the National Committee on Quality Assurance (NCQA) to accredit HMOs.

**Health Reimbursement Account (HRA)** The HRA is similar to a health savings account (HSA), but it is completely controlled by the employer and does not have to be linked to a high-deductible health plan. It is used to pay for qualified medical expenses and can also be used to purchase health insurance.

**Health Savings Account (HSA)** Introduced in 2003 as part of the Medicare prescription drug benefit legislation, the HSA is a less-restrictive medical savings account (MSA), owned by the employee, and open to anyone enrolled in a high-deductible health plan (HDHP) and not already covered by public or private insurance.

**Herfindahl-Hirschman Index (HHI)** A measure of market concentration that incorporates the size distribution of firms. It is found by summing the squares of the market shares of each firm and varies from 0 to 1, or (when shares are in percent terms) from 0 to 10,000.

**High-Deductible Health Plan (HDHP)** A health insurance plan with a much higher deductible and a lower insurance premium than a traditional plan. An individual must have an HDHP to open a health savings account (HSA).

**Human Capital** A form of intangible capital that includes the skills and other knowledge that workers possess, or acquire through education, training, and health care that yields valuable productive services over time.

**Incidence** (*See also* Prevalence) In epidemiology, the number of new cases per a defined population.

**Income Effect** (*See also* Substitution Effect) The effect on quantity demanded that results from the change in real income associated with a relative change in the price of the good or service under study.

**Income Elasticity of Demand** (*See also* Elasticity) The percentage change in quantity demanded resulting from a 1 percent change in income.

**Indemnity Insurance** Traditional health insurance often associated with fee-for-service payments, fee schedules, and which, unlike managed care, places few restrictions on choice of providers.

**Independent Practice Association (IPA)** Physicians in independent, solo, or small group practices who contract with a health maintenance organization to provide services to its members.

**Indifference Curve** Geometric construction showing all combinations of goods that provide a constant level of satisfaction (utility) to the individual under study.

**Inefficiency** (*See* Efficiency and Pareto Efficiency)

**Infant Mortality Rate** The ratio of the number of deaths in infants age one year or less during a year divided by the number of live births during the year.

**Inferior Good** (*See also* Normal Good) A good or service for which demand decreases as income increases.

**Internal Rate of Return** The discount rate that will equate the time streams of costs and returns of an investment. It is a measure of the profitability of an investment.

**Isoquant (Isoproduct Curve)** All combinations of factors of production yielding a constant level of output.

**Labor–Leisure Trade-off** The line that is the collection of points representing the combinations of leisure time and earnings from work that are possible for a person.

**Law of Demand** A statement of the inverse relationship between price and quantity demanded, *ceteris paribus*.

**Law of Diminishing Returns** After some point, the marginal product of a variable input must diminish.

**Loading Costs** The administrative and other costs associated with underwriting an insurance policy.

**Long Run** (*See also* Short Run) A period of time sufficient to permit a firm to vary all factors of production.

**Long-Term Care** The ongoing health and social services provided for individuals who need assistance on a continuing basis because of physical or mental disability. Services can be provided in an institution, the home, or the community, and

include informal services provided by family or friends as well as formal services provided by professionals or agencies.

**Luxury Good** A good that richer people tend to buy in greater proportions so that its income elasticity exceeds 1.

**Managed Care** Any payment or delivery arrangement used by a health plan or provider to control or to coordinate use of health services to contain health expenditures, improve quality, or both.

**Managed Care Organization (MCO)** A general term referring to the various health plans that use managed care arrangements and have a defined system of selected providers that contract with them. The most common are health maintenance organizations (HMOs), preferred provider organizations (PPOs), and point-of-service (POS) plans.

**Mandated Benefits** The coverage in health insurance policies for services that are mandated by state insurance statutes.

**Marginal Cost** The increase in total cost resulting from a one-unit increase in output.

**Marginal Labor (Factor) Cost** The addition to total labor (factor) costs associated with an additional unit of labor (factor of production).

**Marginal Product** The addition to total output resulting from an additional unit of the variable input.

**Marginal Rate of Substitution** The amount of one commodity given up per unit increase in another commodity, while maintaining the same level of production or satisfaction (for consumers).

**Marginal Rate of Technical Substitution** The amount of one factor of production given up per unit increase in another factor of production, while maintaining the same level of output.

**Marginal Rate of Transformation** The slope of the production possibilities curve, and the rate at which society can transform one good into another.

**Marginal Revenue** The addition to total revenue associated with a one-unit increase in output.

**Marginal Revenue Product** The addition to a firm's total revenue associated with employing one more unit of a variable input.

**Marginal Utility** The extra utility gained from consuming one more unit of a good, holding others constant. Utility is a measure of the satisfaction from consuming goods.

**Market Demand** The total demand for a good by all consumers in the market.

**Market Structure** The organization of an industry in terms of the number and distribution of firms and how firms compete among themselves.

**Medicaid** The health insurance programs administered by the states for qualifying low-income beneficiaries. The federal government establishes minimum standards and provides matching grants. The program became law in 1965.

**Medical Savings Account (MSA)** A limited health spending account (HSA) that was introduced on an experimental basis in the 1990s. MSAs were aimed at small businesses and individuals.

**Medicare** The federal health insurance program established in 1965 for the elderly and other selected groups.

**Medicare—Part A** The Medicare Hospital Insurance that covers beneficiaries for inpatient hospital, home health, hospice, and limited skilled nursing facility services. Beneficiaries are responsible for deductibles and copayments.

**Medicare—Part B** The Medicare Supplementary Medical Insurance that covers Medicare beneficiaries for physician services, medical supplies, and other outpatient treatment. Beneficiaries are responsible for monthly premiums, copayments, deductibles, and balance billing.

**Medicare—Part C** (Medicare Advantage) An expanded set of options for the delivery of health care under Medicare. While all Medicare beneficiaries can receive their benefits through the original fee-for-service program, most beneficiaries enrolled in both Part A and Part B can choose to participate in a Medicare Advantage plan instead. Managed care or fee-for-service providers that seek to contract as Medicare Advantage plans, must meet specific organizational, financial, and other requirements.

**Medicare—Part D** Part D provides subsidized access to prescription drug insurance coverage on a voluntary basis, upon payment of a premium, to individuals entitled to Part A or enrolled in Part B, with premium and cost-sharing subsidies for low-income enrollees. Beneficiaries may enroll in either a stand-alone prescription drug plan (PDP) or an integrated Medicare Advantage plan that offers Part D coverage.

**Medigap Policy** A privately purchased insurance policy that supplements Medicare coverage and meets specified requirements set by federal statute and the National Association of Insurance Commissioners.

**Monopoly** Situations in which a firm faces a negatively sloped demand curve. In a pure monopoly, no other firm produces a close substitute for the firm's product. The demand curve facing the monopolist is the market demand curve.

**Monopoly Profit (Rent)** The return over and above a normal profit resulting from monopoly.

**Monopsony** Situations in which a firm faces a positively sloped supply curve in the product or factor market because it is the only buyer. The supply curve facing the monopsonist is the market supply curve.

**Moral Hazard** A term that represents the disincentives created by insurance (more generally, any contractual arrangement) for individuals to take measures that would reduce the amount of care demanded. In the health services literature, it more commonly describes the additional quantity of health care demanded, due to a decrease in the net price of care attributable to insurance.

**Morbidity Rate** The rate of incidence of disease in a particular population.

**Mortality Rate** The death rate for a particular population. The crude death rate is the ratio of deaths during a year divided by midyear population. Because age is so important, the age-adjusted mortality rate is a measure that takes into account a population's age distribution.

**Necessity** A good whose consumption does not vary greatly with changes in peoples' incomes. More generally, a good with an income elasticity less than 1.

**Nominal Value** (*See also* Real Value) The money value measured in current dollars.

**Nonprofit Firm** A firm that is constrained by law from distributing any residual of income over costs to any party.

**Normal Good** (*See also* Inferior Good) A good or service for which demand increases as income increases.

**Normal Return (Normal Profit)** The return just sufficient to retain factors of production in an industry or an occupation in the long run. The return equal to the opportunity cost of a factor of production.

**Nursing Facility** An institution that provides skilled nursing care and rehabilitation services to injured, functionally disabled, or sick persons.

**Offer Curve** A set of points summarizing the amount of a good that an individual will offer for trade, given his or her preferences, endowment, and the prices of other goods.

**Opportunity Cost** The value of the best alternative that is forgone in order to get or produce more of the commodity under consideration.

**Ordinal Utility** Utility as evaluated through relative levels of satisfaction, when the particular unit of utility is not essential. Examples of ordinal numbers are first, second, and third.

**Pareto Efficiency (Optimum)** (*See also* Efficiency) An allocation in which it is impossible to improve the level of welfare of one party without hurting the welfare level of another party. Circumstances in which the level of welfare of one or more parties can be improved without hurting any other party are Pareto improvements.

**Participation/Assignment** A situation in which a provider agrees to accept the third-party payer's payment in full, thereby relieving the patient of any balance (except for applicable patient copayments).

**Patient Protection and Affordable Care Act (PPACA)** Signed into law by President Barack Obama in 2010. PPACA, through the existing private health insurance industry and public health insurance programs, in steps incremented through 2018, increases insurance coverage of pre-existing conditions, and expands access to insurance to over 30 million Americans.

**Pay-for-Performance (P4P)** The effort by managed care organizations and other payers to reward providers who improve the quality of their care by meeting certain performance standards. Providers typically receive bonus payments for meeting the goals.

**Peer Review Organization (PRO)** An organization that contracts with the Centers for Medicare & Medicaid Services to investigate the quality of health care furnished to Medicare beneficiaries and to educate beneficiaries and providers. PROs also conduct limited review of medical records and claims to evaluate the appropriateness of care provided.

**Perfect Competition** A market structure with (1) numerous buyers and sellers, (2) perfect information, (3) free entry and exit, and (4) a homogeneous product.

**Play or Pay** Insurance proposal that would require employers to either provide minimal levels of health insurance to their employees ("play") or *pay* into a fund that would be used to provide coverage.

**Point-of-Service (POS) Plan** A managed care plan that encourages patients to select a provider in a network. Members can select non-network providers but will incur higher out-of-pocket costs.

**Preferred Provider Organization (PPO)** An arrangement under which an enrollee is given financial incentives (e.g., zero copayments) to seek care from selected physicians and hospitals with which the payer has contracted.

**Prepaid Group Practice (PGP)** A prepayment arrangement under which participating providers agree to provide services to eligible enrollees in return for a fixed capitated payment.

**Prepayment (Prepaid Plans)** Health insurance proposal that provides unlimited amounts of covered services in return for a fixed predetermined premium.

**Present (Discounted) Value (PV or PDV)** (*See also* Discounting) The value of a stream of returns to be received at future dates, which is discounted to the equivalent of present dollars.

**Prevalence** (*See also* Incidence) In epidemiology, the fraction of the population that is currently infected. Incidence adds new cases to the total pool describing the prevalence of present cases.

**Price Discrimination** The sale of goods or services to different individuals at different prices.

**Price Elasticity of Demand** (*See also* Elasticity) The percentage change in quantity demanded resulting from a 1 percent change in price.

**Price Elasticity of Supply** (*See also* Elasticity) The percentage change in quantity supplied resulting from a 1 percent change in price.

**Price Index** An index that expresses the current prices of a group of goods relative to the prices of these goods in a base year. A price index, often used to convert nominal values to real values, shows how much prices of those goods have changed since the base year.

**Production Function** The relationship between the maximum output that can be produced corresponding to any combination of factor inputs.

**Production Possibilities Curve (Transformation Curve)** A relationship showing all combinations of goods that an economy



can produce with given amounts of input factors and the existing technology. The slope of the curve is the marginal rate of transformation showing the amount of one good that must be given up for a one-unit increase in the other good.

**Prospective Payment System (PPS)** The method of hospital reimbursement phased in by Medicare beginning in 1983 under which hospitals were reimbursed a fixed amount determined by the diagnostic related groups of their admissions.

**Public Good (Pure)** A good (e.g., national defense) that no one can be prevented from consuming, (i.e., nonexcludable) and that can be consumed by one person without depleting it for another (i.e., nonrival). The marginal cost of providing the good to another consumer is zero.

**Quality-Adjusted Life-Year (QALY)** A measure of health outcome that incorporates quantity and quality of life. It uses a weighting system that assigns a value ranging from 1 (perfect health) to 0 (health state equivalent to death).

**Rate Regulation (RR)** Regulations administered largely at the state level that define the terms under which hospitals are reimbursed.

**Real Value** (*See also* Nominal Value) Monetary value that is adjusted for changes in the general level of prices relative to some arbitrarily selected base year.

**Regression Analysis** Statistical analysis that posits a linear relationship between a variable to be explained and one or more (in multiple regression) explanatory variables, in the form  $y = a + bx$ .

**Reinforcement** A characteristic of addictive behavior indicating that greater past consumption of addictive goods increases the desire for present consumption.

**Relative Value Scale** Systems that assign relative weights to health care procedures or services. These are sometimes used as guides by providers to price their services and by third-party payers to determine reimbursement levels.

**Rent (Economic Rent)** The remuneration to a factor of production, over and above the amount that is necessary to induce its supply in the market.

**Reputation Good** A good for which consumers rely on information provided by friends, neighbors, and others.

**Resource-Based Relative Value Scale (RBRVS)** A relative value scale used in Medicare reimbursement that is based on the resources involved in providing a service.

**Risk Aversion** The degree to which a certain income is preferred to a risky alternative with the same expected income.

**Risk Selection** The enrollment choices made by health plans or enrollees on the basis of perceived risk relative to the premium to be paid.

**Search Good** A good whose characteristics can be fully evaluated upon inspection.

**Second Fundamental Theorem of Welfare Economics** The proposition that any Pareto efficient outcome can, in principle, be achieved by competitive markets, given an appropriate initial endowment.

**Selective Contracting** The practice of a managed care organization (MCO) by which the MCO enters into participation agreements only with certain providers (and not with all providers who qualify) to provide health care services to health plan participants as members of the MCO's provider panel.

**Short Run** (*See also* Long Run) Situations in which the firm is not able to vary all its inputs. There is at least one factor of production that is fixed.

**Single-payer System** Payment to health care providers administered by a single entity or organization, usually, but not necessarily, the government. Canada is often cited as an example of a single-payer system.

**Small Area Variations** The large variations in the per capita rates of utilization across small, homogeneous areas for many medical and surgical procedures.

**Social Insurance** Government insurance programs in which eligibility and premiums are not determined by the practices common to private insurance contracts. Premiums are often subsidized and there are typically redistributions from some segments of the population to others.

**Social Welfare Function** A decision rule under which society ranks all possible distributions of goods and services.

**Staff Model HMO** A health maintenance organization (HMO) in which physicians are directly employed by the HMO.

**Substitutes** The substitutes in consumption are goods that satisfy the same wants (e.g., beef and chicken) so that an increase in the price of one will increase the demand for the other. Substitutes in production are alternative goods that a firm can produce (e.g., corn and soybeans for a farmer) so that an increase in the price of one will lead to a decrease in the supply of another.

**Substitution Effect** (*See also* Income Effect) The change in quantity demanded resulting from a relative change in commodity prices, holding real income constant.

**Supplier-Induced Demand (SID)** The change in demand associated with the discretionary influence of providers, especially physicians, over their patients. Demand provided for the self-interests of providers rather than solely for patient interests.

**Take Up** A response to the introduction or enhancement of public insurance, referring to the extent that those who have previously been uninsured now use public coverage.

**Target Income Hypothesis** A model under which providers are thought to select a specified income level, and to adjust their amount of services provided or fees in order to reach this level.

**Technical Efficiency** (*See* Efficiency)

**Technological Change** A change in the process by which factors of production combine to produce outputs.

**Theorem of the Second Best** The economic theorem stating that the correction of some but not all market imperfections, in cases where there is more than one imperfection, may not necessarily improve society's welfare.

**Time Costs** The money value of the time lost through travel or waiting when consuming a product or service.

**Tolerance** A characteristic of addictive behavior indicating that the incremental utility from a given amount of consumption of the addictive good is lower when past consumption is greater. This suggests that the marginal future impacts of smoking or drinking or ingesting drugs decrease when we consume more at the present time.

**Uncompensated Care** The care rendered by hospitals or other providers without payment from the patient or a government-sponsored private insurance program. It includes both charity care, provided without expectation of payment, and bad debts, for which the provider has made an unsuccessful effort to collect payment due from the patient.

**Usual, Customary, and Reasonable (UCR) Reimbursement** A system of insurance reimbursement under which the third-party payer restricts payment to the maximum of several limits (screens) that are established.

**Utility and Utility Function** Utility represents satisfaction or the level of welfare of an individual, measured in cardinal or

ordinal utility terms. The utility function expresses the person's utility as a function of all possible combinations of goods and services.

**Utilization Review (UR)** The programs that attempt to determine whether specific services are medically necessary and delivered at an appropriate level and cost.

**Variable Costs (TVC and AVC)** Costs associated with variable factor(s) of production, often expressed as total variable cost (TVC) or average variable cost (AVC).

**Welfare** A measure of an individual's or a society's level of well-being.

**Welfare Loss or Deadweight Loss** A measure of the net loss of society's welfare resulting from a misallocation of resources, usually situations in which the marginal benefits of a good do not equal marginal costs.

**Yardstick Competition** A regulatory pricing policy in which an average of the marginal costs of all competing firms is used as a standard of payment to induce the firm to engage in cost-cutting innovation.

# REFERENCES

- Aaberge, Rolf et al., "The Distributional Impact of Public Services When Needs Differ," *Journal of Public Economics* 94 (2010): 549–562.
- Aaron, Henry J., "The Costs of Health Care Administration in the United States and Canada—Questionable Answers to a Questionable Question," *New England Journal of Medicine* 349 (2003): 801–803.
- Abbott, Thomas A., "Price Regulation in the Pharmaceutical Industry: Prescription or Placebo?" *Journal of Health Economics* 14 (1995): 551–565.
- Abso, Ignacio and Aki Tsuchiya, "Exploring Social Welfare Functions, and Violations of Monotonicity: An Example from Inequalities in Health," *Journal of Health Economics* 23 (2004): 313–329.
- Acemoglu, Daron, and Amy Finkelstein, "Input and Technology Choices in Regulated Industries: Evidence from the Health Care Sector," *NBER Working Paper w12254*, 2006.
- Acton, Jan P., "Demand for Health Care Among the Urban Poor with Special Emphasis on the Role of Time," in *The Role of Health Insurance in the Health Services Sector*, Richard Rosett (ed.) (New York: Neal Watson, 1976).
- Acton, Jan P., "Nonmonetary Factors in the Demand for Medical Services: Some Empirical Evidence," *Journal of Political Economy* 83 (1975): 595–614.
- Adamache, Killard W., and Louis F. Rossiter, "The Entry of HMOs into the Medicare Market: Implications for TEFRA's Mandate," *Inquiry* 23 (1986): 349–364.
- Adams, Scott J., "Health Insurance Market Reform and Employee Compensation: The Case of Pure Community Rating in New York," *Journal of Public Economics* 91 (2007): 1119–1133.
- Agency for Healthcare Research and Quality, *Annual Quality Report, 2009*, Washington 2010a.
- Agency for Health Care Research and Quality, *2009 Healthcare Disparities Report*, Washington, 2010b.
- Ahituv, Avner, V. Joseph Hotz, and Tomas Philipson, "The Responsiveness of the Demand for Condoms to the Local Prevalence of AIDS," *Journal of Human Resources* 31 (1996): 869–897.
- Aigner, Dennis J., C.A. Knox Lovell, and Peter Schmidt, "Formulation and Estimation of Stochastic Frontier Production Function Models," *Journal of Econometrics* 6 (1977): 21–37.
- Aizer, Anna, Janet Currie, and Enrico Moretti, "Does Managed Care Hurt Health? Evidence from Medicaid Mothers," *Review of Economics and Statistics* 89 (2007): 385–399.
- Akerlof, George A., "The Market for 'Lemons': Qualitative Uncertainty and the Market Mechanism," *Quarterly Journal of Economics* 84 (1970): 488–500.
- Albion, Mark S. and Paul W. Farris. *The Advertising Controversy: Evidence on the Economic Effects of Advertising*. (Boston: Auburn House, 1981).
- Alchian, Armen A. and Harold Demsetz, "Production, Information Costs, and Economic Organization," *American Economic Review* 62 (1972): 777–795.
- Alexander, Jeffrey A. et al., "Do Market-Level Hospital and Physician Resources Affect Small Area Variation in Hospital Use?" *Medical Care Research and Review* 56 (1999): 94–117.
- Almond, Douglas et al, "Estimating Marginal Returns to Medical Care: Evidence from at-risk Newborns," *Quarterly Journal of Economics* 125 (2010): 591–634.
- Altman, Stuart H., David Shactman, and Eilat Efrat, "Could U.S. Hospitals Go the Way of U.S. Airlines?" *Health Affairs* 25 (2006): 11–21.
- Alvarez-Dardet, Carlos and Maria Teresa Ruiz, "Thomas McKeown and Archibald Cochrane: A Journey Through the Diffusion of Their Ideas," *British Medical Journal* 306 (May 8, 1993): 1252–1254.
- American Hospital Association and the Lewin Group, *Trendwatch Chartbook 2006*, March 2006, Table 4-4.
- American Medical Association, *Competition in Health Insurance: A Comprehensive Study of U.S. Markets 2010 Update* (AMA: Chicago, 2010).
- Andermann, Anne Adina Judith, "Physicians, Fads, and Pharmaceuticals: A History of Aspirin," *McGill Journal of Medicine* 2 (1996), <http://www.medicinemcgill.ca/mjmv02n02/aspirin.html>, accessed April 2, 2009.
- Anderson, Gerard F. and Jean-Pierre Poullier, "Health Spending, Access, and Outcomes, Trends in Industrialized Countries," *Health Affairs* 18 (1999): 178–192.
- Anderson, Gerard F., Peter S. Hussey, Bianca K. Frogner, and Hugh R. Waters "Health Spending in the United States and the Rest of the Industrialized World," *Health Affairs* 24 (2005): 903–914.
- Anderson, Gerard F., Uwe E. Reinhardt, Peter S. Hussey, and Varduhi Petrosyan, "It's The Prices, Stupid: Why the United States Is So Different from Other Countries," *Health Affairs* 22 (2003): 89–105.
- Anderson, Odin W., *Health Services as a Growth Enterprise in the United States Since 1875*, 2nd ed. (Ann Arbor, MI: Health Administration Press, 1990).
- Anderson, Patricia M., Kristin F. Butcher, and Phillip B. Levine, "Maternal Employment and Overweight Children," *Journal of Health Economics* 22 (2003): 477–504.
- Andreoni, James, "Impure Altruism and Donations to Public Goods: A Theory of Warm-Glow Giving," *The Economic Journal* 100 (1990): 464–477.

- Archibald, Robert, Clyde Haulman, and Carlisle Moody, "Quality, Price, Advertising, and Published Quality Ratings," *Journal of Consumer Research* 9 (1983): 347–356.
- Arnould, Richard J., Lawrence W. Debrock, and John W. Pollard, "Do HMOs Produce Specific Services More Efficiently?" *Inquiry* 21 (1984): 243–253.
- Arrow, Kenneth J., *Social Choice and Individual Values*. (New York: Wiley, 1975).
- Arrow, Kenneth J., "Some Ordinalist-Utilitarian Notes on Rawls's Theory of Justice," *Journal of Philosophy* 70 (1973): 245–262.
- Arrow, Kenneth J., "Uncertainty and the Welfare Economics of Medical Care," *American Economic Review* 53 (1963): 941–973.
- Ash, Timothy Garton, *The Magic Lantern*. (New York: Vintage Books, 1993).
- Associated Press, "Hawaii Ends Universal Health Care for Kids," [www.cnn.com/2008/HEALTH/10/17/hawaii.health.ap/](http://www.cnn.com/2008/HEALTH/10/17/hawaii.health.ap/), accessed November 19, 2008.
- Atkinson, Scott E. and Thomas Crocker, "Econometric Health Production Functions: Relative Bias from Omitted Variables and Measurement Error," *Journal of Environmental Economics and Management* 22 (1992): 12–24.
- Auld, M. Christopher and Nirmal Sidhu, "Schooling, Cognitive Ability and Health," *Health Economics* 14 (2005): 1019–1034.
- Auster, Richard D., Irving Leveson, and Deborah Sarachek, "The Production of Health, an Exploratory Study," *Journal of Human Resources* 4 (1969): 411–436.
- Austin, D. Andrew, and Thomas L. Hungerford, "The Market Structure of the Health Insurance Industry," Congressional Research Service, 7-5700, December 17, 2010.
- Aylward, Bruce, and Tadataka Yamada, "The Polio Endgame," *New England Journal of Medicine* 364 (June 15, 2011): 2273–2275.
- Bae, Jay P., "Drug Patent Expiration and the Speed of Generic Entry," *Health Services Research* 22 (1997): 87–101.
- Bain, Joe S., *Barriers to New Competition* (Cambridge, MA: Harvard, 1956).
- Baker, Laurence C., "Managed Care and Technology Adoption in Health Care: Evidence from Magnetic Resonance Imaging," *Journal of Health Economics* 20 (2001): 395–421.
- Baker, Laurence C. and Kenneth S. Corts, "The Effects of HMOs on Conventional Insurance Premiums: Theory and Evidence," *NBER Working Paper* 5356, 1995.
- Baker, Laurence C. and Kenneth S. Corts, "HMO Market Penetration and the Cost of Health Care: Market Discipline or Market Segmentation," *American Economic Review* 86 (1996): 389–394.
- Baker, Lawrence, and Ciaran Phibbs, "Managed Care, Technology Adoption, and Health Care: The Adoption of Neonatal Intensive Care," *NBER Working Paper* w7883, 2000.
- Baker, Laurence C. and Joanne Spetz, "Managed Care and Medical Technology Growth," in *Frontiers in Health Policy Research, Volume 2*, Alan Garber (ed.) (Cambridge, MA: MIT Press, 2000).
- Bala, Mohan V. and Gary A. Zarkin, "Are QALYs an Appropriate Measure for Valuing Morbidity in Acute Diseases?" *Health Economics* 9 (2000): 177–180.
- Balcombe, Kelvin, Ian Fraser, and Salvatore Di Falco, "Traffic Lights and Food Choice: A Choice Experiment Examining the Relationship Between Nutritional Food Labels and Price," *Food Policy* 35 (2010): 211–220.
- Balia, Silvia and Andrew Jones, "Mortality, Lifestyle and Socio-economic Status," *Journal of Health Economics* 27 (2008): 1–26.
- Ballou, Jeffrey P., "Do the Nonprofit and Government Homes Enter Unprofitable Markets?" *Economic Inquiry* 46 (2008): 241–260.
- Ballou, Jeffrey P. and Burton A. Weisbrod, "Managerial Rewards and the Behavior of For-Profit, Governmental, and Nonprofit Organizations: Evidence from the Hospital Industry," *Journal of Public Economics* 87 (2003): 1895–1920.
- Balsa, Ana I., Zhun Cau, and Thomas G. McGuire, "Does Managed Care Reduce Health Care Disparities between Minorities and Whites?" *Journal of Health Economics* 26 (2007): 101–121.
- Baltagi, Badi H. and James M. Griffin, "The Econometrics of Rational Addiction: The Case of Cigarettes," *Journal of Business and Economic Statistics* 19 (2001): 449–454.
- Baltagi, Badi H. and Dan Levin, "Estimating Dynamic Demand for Cigarettes Using Panel Data: The Effects of Bootlegging, Taxation, and Advertising Reconsidered," *Review of Economics and Statistics* 68 (1986): 148–155.
- Barbetta, Gian Paolo, Gilberto Turati, and Angelo M. Zago, "Behavioral Differences Between Public and Private Not-for-Profit Hospitals in the Italian National Health Service," *Health Economics* 16: 75–96 (2007).
- Barry, John M., *The Great Influenza: The Epic Story of the Deadliest Plague in History* (revised ed.) (New York: Penguin Paperback, 2005).
- Barry, Michael J. et al., "Patient Reactions to a Program Designed to Facilitate Patient Participation in Treatment Decisions for Benign Prostatic Hyperplasia," *Medical Care* 33 (1995): 771–782.
- Barsky, Robert B. et al., "Preference Parameters and Behavioral Heterogeneity: An Experimental Approach in the Health and Retirement Study," *Quarterly Journal of Economics* 112 (1997): 537–579.
- Bates, Laurie J., and Rexford E. Santerre, "Do Health Insurers Possess Monopsony Power in the Hospital Service Industry?," *International Journal of Health Care Finance and Economics* 8 (2008): 1–11.
- Bator, Francis M., "The Simple Analytics of Welfare Maximization," *American Economic Review* 47 (1957): 22–59.

- Baumol, William J., John C. Panzar, and Robert Willig, *Contestable Markets and the Theory of Industry* (San Diego: Harcourt Brace Jovanovich, 1982).
- Bays, Carson W., "Why Most Private Hospitals Are Nonprofit," *Journal of Policy Analysis and Management* 2 (1983): 366–385.
- Beaulieu, Nancy D., "Quality Information and Consumer Health Plan Choices," *Journal of Health Economics* 21 (2002): 43–63.
- Becker, Gary S., Michael Grossman, and Kevin M. Murphy, "Rational Addiction and the Price on Consumption," *American Economic Review* 81 (1991): 237–241.
- Becker, Gary S. and Kevin M. Murphy, "A Simple Theory of Advertising as a Good or Bad," *Quarterly Journal of Economics* 108 (1993): 941–964.
- Becker, Gary S. and Kevin M. Murphy, "A Theory of Rational Addiction," *Journal of Political Economy* 96 (1988): 675–700.
- Becker, Gary S., Kevin M. Murphy, and Michael Grossman, "The Economic Theory of Illegal Goods: The Case of Drugs," *NBER Working Paper* 10976, 2004.
- Becker, Gary S., Kevin M. Murphy, and Tomas Philipson, "The Value of Life Near Its End and Terminal Care," *NBER Working Paper* 13333, 2007.
- Benham, Lee, "The Effects of Advertising on the Price of Eyeglasses," *Journal of Law and Economics* 15 (1972): 421–477.
- Besser, Richard, "Doctor's Notebook", March 8, 2011, <http://abcnews.go.com/Health/GlobalHealth/dr-bessers-notebook-shot-save-hearing/story?id=13080818>, accessed June 3, 2011.
- Bhagwati, Jagdish and Sandan Madan, "We Need Free Trade in Health Care," *Wall Street Journal*, May 27, 2008, p. A19.
- Bhattacharya, Jayanta, "Specialty Selection and Lifetime Returns to Specialization Within Medicine," *Journal of Human Resources* 60 (2005): 115–143.
- Bhattacharya, Jay and Darius N. Lakdawalla, "The Labor Market Value of Health Improvements," *Forum for Health Economics & Policy* 9 (2006): (Biomedical Research and the Economy), Article 2. [www.bepress.com/fhep/biomedical\\_research/2](http://www.bepress.com/fhep/biomedical_research/2), accessed May 20, 2008.
- Bhattacharya, Jay and M. Kate Bundorf, "The Incidence of the Healthcare Costs of Obesity," *NBER Working Paper* 11303, 2005.
- Bilodeau, David, Pierre-Yves Cremieux, and Pierre Oullette, "Hospital Technology in a Nonmarket Health Care System," *Southern Economic Journal* 68 (2002): 511–529.
- Bishop, Christine E., "Where Are the Missing Elders? The Decline in Nursing Home Use, 1985 and 1995," *Health Affairs* 18 (1999): 146–155.
- Blank, Rebecca M., "Can Equity and Efficiency Complement Each Other?" *Labour Economics* 9 (2002): 451–468.
- Bleichrodt, Han and John Quiggin, "QALYs and Consumer Demand for Health Care," *Journal of Health Economics* 18 (1999): 681–708.
- Blendon, Robert J., Mollyann Brodie, and John Benson, "What Happened to Support for the Clinton Plan?" *Health Affairs* 14 (1995): 7–23.
- Blendon, Robert et al., "Common Concerns amid Diverse Systems: Health Care Experiences in Five Countries," *Health Affairs* 22 (2003): 106–121.
- Blendon, Robert J. et al., "Understanding the Managed Care Backlash," *Health Affairs* 17 (1998): 80–94.
- Blomquist, Ake, "QALYs, Standard Gambles, and the Expected Budget Constraint," *Journal of Health Economics* 21 (2002): 181–195.
- Blomquist, Ake G. and R.A.L. Carter, "Is Health Care Really a Luxury?" *Journal of Health Economics* 16 (1997): 207–230.
- Blomquist, Ake and Pierre Thomas Leger, "Information Asymmetry, Insurance, and the Decision to Hospitalize," *Journal of Health Economics* 24 (2005): 775–793.
- Bloom, David E. and Ajay Mahal, "Does the AIDS Epidemic Threaten Economic Growth?" *Journal of Econometrics* 77 (1997): 105–124.
- Bloom, David E. and Jaypee Sevilla, "Health, Wealth, AIDS, and Poverty," mimeo, 2002.
- Blundell, R. and F. Windmeijer, "Identifying Demand for Health Resources Using Waiting Time Information," *Health Economics* 9 (2000): 465–474.
- The Boards of Trustees, Federal Hospital Insurance and Federal Supplementary Medical Insurance Trust Funds, *2011 Annual Report of the Boards of Trustees of the Federal Hospital Insurance and Federal Supplementary Medical Insurance Trusts Funds, Washington DC, May 13, 2011* <https://www.cms.gov/ReportsTrustFunds/downloads/tr2011.pdf>, accessed June 7, 2011.
- Bolin, Kristian, Lena Jacobson, and Bjorn Lindgren, "The Family as the Health Producer—When Spouses Act Strategically," *Journal of Health Economics* 21 (2002): 475–495.
- Bond, Ronald et al., *Staff Report on Effects of Restrictions on Advertising and Commercial Practice in the Professions: The Case of Optometry* (Federal Trade Commission, Washington, DC: Government Printing Office, 1980).
- Bonsang, Eric, "Does Informal Care from Children to their Elderly Parents Substitute for Formal Care in Europe?," *Journal of Health Economics* 28 (2009): 143–154.
- Borisova, Natalia and Allen C. Goodman, "The Effects of Time and Money Prices on Treatment Attendance for Methadone Maintenance Clients," *Journal of Substance Abuse Treatment* 26 (2004): 345–352.
- Borjas, George J., H. E. Frech III, and Paul B. Ginsburg, "Property Rights and Wages: The Case of Nursing Homes," *Journal of Human Resources* 17 (1983): 232–246.

- Bosmans, Nathalie and Fabienne Fecher, "Performance of Belgian Hospitals: A Frontier Approach," *Health Economics* 4 (1995): 389–397.
- Bosworth, Barry and Gary Burtless, "Effects of Tax Reform on Labor Supply, Investment, and Saving," *Journal of Economic Perspectives* 6 (1992): 3–25.
- Boulier, Bryan L., Tejwant Singh Datta, and Robert S. Goldfarb, "Vaccination Externalities," *The B.E. Journal of Economic Analysis & Policy* 7. (Contributions), Article 23, <http://www.bepress.com/bejeap/vol7/iss1/art23>.
- Boyle, Sean, "United Kingdom (England): Health System Review," *Health Systems in Transition* 13 (2011): 1–486.
- Braden, Bradley R. et al., "National Health Expenditures, 1997," *Health Care Financing Review* 20 (1998): 83–110.
- Bradford, W David et al., "The Effect of Direct to Consumer Television Advertising on the Timing of Treatment" *Economic Inquiry* 48 (2010): 306–322.
- Braunwald, Eugene et al., "HIV Infection and AIDS," *Harrison's Manual of Medicine*, Chapter 86 (2001), [www.harrisonsmed.com/downloads/manual\\_ch86](http://www.harrisonsmed.com/downloads/manual_ch86), accessed January 16, 2003.
- Brekke, Kurt R., Luigi Siciliani, Odd Rune Straume, "Competition and Waiting Times in Hospital Markets," *Journal of Public Economics* 92 (2008): 1607–1628.
- Brekke, Kurt R., et al., "Competition and Quality in Health Care Markets: A Differential-Game Approach," *Journal of Health Economics* 29 (2010): 508–523.
- Brennan, Troyen A. et al., "Incidence of Adverse Events and Negligence in Hospitalized Patients," *New England Journal of Medicine* 324 (1991): 370–376.
- Brickley, James A. and R. Lawrence Van Horn, "Managerial Incentives in Nonprofit Organizations: Evidence from Hospitals," *Journal of Law and Economics* 45 (2002): 227–249.
- Brody, Howard, "Medicine's Ethical Responsibilities for Health Care Reform—The Top Five List," *New England Journal of Medicine* 362 (2010): 283–285.
- Brook, Robert H. et al., "Appropriateness of Acute Medical Care for the Elderly: An Analysis of the Literature," *Health Policy* 14 (1990): 225–242.
- Brook, Robert H. et al., "Does Free Care Improve Adults' Health? Results from a Randomized Trial," *New England Journal of Medicine* 309 (1983): 1426–1434.
- Brouwer, Werner B. F., "Need for Differential Discounting of Costs and Health Effects in Cost Effectiveness Analysis," *British Medical Journal* 331 (2005): 446–448.
- Brouwer, Werner B. F. et al., "Welfarism vs. Extra-welfarism," *Journal of Health Economics* 27 (2008): 325–338.
- Brown, Douglas M., "Do Physicians Underutilize Aides?" *Journal of Human Resources* 23 (1988): 342–355.
- Brown, Lawrence D., "The National Politics of Oregon's Rationing Plan," *Health Affairs* 10 (1991): 28–51.
- Browne, Mark J. and Helen I. Doeringhaus, "Information Asymmetries and Adverse Selection in the Market for Individual Medical Expense Insurance," *Journal of Risk and Insurance* 60 (1993): 300–312.
- Brown, Timothy T. et al., "The Empirical Relationship Between Community Social Capital and the Demand for Cigarettes," *Health Economics* 15 (2006): 1159–1172.
- Bryan, Stirling and Sue Jowett, "Hypothetical versus Real Preferences: Results from an Opportunistic Field Experiment," *Health Economics*, 19 (2010): 1502–1509.
- Buchanan, James, *The Limits of Liberty*. (Chicago: University of Chicago Press, 1975).
- Bunce, Victoria Craig and J. P. Wieske, "Health Insurance Mandates in the States 2010," [http://www.cahi.org/cahi\\_contents/resources/pdf/MandatesintheStates2010.pdf](http://www.cahi.org/cahi_contents/resources/pdf/MandatesintheStates2010.pdf), accessed May 24, 2011.
- Bundorf, M. Kate, et al., "Do Markets Respond to Quality Information? The Case of Fertility Clinics," *Journal of Health Economics* 28 (2009): 718–727.
- Burgess, James F. Jr. and Paul W. Wilson, "Variation in Inefficiency in U.S. Hospitals," *INFOR* 36 (1998): 84–102.
- Burke, Mary, Gary Fournier, and Kislaya Prasad, "Symposium: Third Annual Southwestern Health Economics Study Group Conference: The Diffusion of a Medical Innovation: Is Success in the Stars?" *Southern Economic Journal* 73 (2007): 588–603.
- Burstein, Philip L. and Jerry Cromwell, "Relative Incomes and Rates of Return for U.S. Physicians," *Journal of Health Economics* 4 (1985): 63–78.
- Cady, John F., *Drugs on the Market* (Lexington, MA: Lexington, 1975).
- Calem, Paul S. and John A. Rizzo, "Competition and Specialization in the Hospital Industry: An Application of Hotelling's Location Model," *Southern Economic Journal* 61 (1995): 1182–1198.
- Cameron, Sam, "Are Greek Smokers Rational Addicts?" *Applied Economics Letters* 4 (1997): 401–402.
- Campbell, Maude, "Looking Down the Pharmaceutical Pipeline," *American Health & Drug Benefits* (June 2008): 37–38.
- Cannon, Michael, "Large Health Savings Accounts: A Step Toward Tax Neutrality for Health Care," *Forum for Health Economics & Policy* 11 (2008): Article 3.
- Capps, Cory, David Dranove, and Richard C. Lindrooth, "Hospital Closure and Economic Efficiency," *Journal of Health Economics* 29 (2010): 87–109.
- Card, David, Carlos Dobkin, and Nicole Maestas. "Does Medicare Save Lives?," *Quarterly Journal of Economics*, 124 (2009): 597–636.
- Cardon, James H. and Igal Hendel, "Asymmetric Information in Health Insurance: Evidence from the National Medical Expenditure Survey," *RAND Journal of Economics* 23 (2001): 408–427.

- Carey, Kathleen and James F. Burgess Jr., "On Measuring the Hospital Cost/Quality Trade Off," *Health Economics* 8 (1999): 509–520.
- Carey, Kathleen and Theodore Stefos "Controlling for Quality in the Hospital Cost Function," *Health Care Management Science* 14 (2011): 125–134.
- Carlsen, Fredrik and Jostein Grytten, "Consumer Satisfaction and Supplier Induced Demand," *Journal of Health Economics* 19 (2000): 731–783.
- Carlson, Rick, *The End of Medicine*. (New York: Wiley Interscience, 1975).
- Caudill, Stephen B., John M. Ford, and David L. Kaserman, "Certificate of Need Regulation and the Diffusion of Innovations," *Journal of Applied Econometrics* 10 (1995): 73–78.
- Cawley, John, "The Impact of Obesity on Wages," *Journal of Human Resources* 39 (2004): 451–474.
- Centers for Disease Control and Prevention, "Health Care: See Why Being Insured Matters," <http://www.cdc.gov/Features/VitalSigns/HealthcareAccess/>, November 2010, accessed May 24, 2011.
- Centers for Disease Control, "HIV Surveillance – United States, 1981 – 2008", Weekly June 3, 2011 / 60(21): 689–693 <http://www.cdc.gov/mmwr/preview/mmwrhtml/mm6021a2.htm>, accessed July 7, 2011.
- The Center for Evaluative Clinical Studies, *Dartmouth Atlas of Health Care* (Dartmouth Medical School, American Hospital Publications, Inc., 1996).
- Central Intelligence Agency, *CIA World Factbook*, various years.
- Chakravarty, Sujoy et al., "Does the Profit Motive Make Jack Nimble? Ownership Form and the Evolution of the U.S. Hospital Industry," *NBER Working Paper* 11705, 2005.
- Chaloupka, Frank, "The Economics of Smoking," in *Handbook of Health Economics*, Anthony J. Culyer and Joseph P. Newhouse (eds.) (Amsterdam, Elsevier Science B.V., 2000).
- Chaloupka, Frank J., "Rational Addictive Behavior and Cigarette Smoking," *Journal of Political Economy* 99 (1991): 722–742.
- Chaloupka, Frank J. and Rosalie Liccardo Pacula, "An Examination of Gender and Race Differences in Youth Smoking Responsiveness to Price and Tobacco Control Policies," *NBER Working Paper* 6541, 1998.
- Chaloupka, Frank J. and Henry Saffer, "Clean Indoor Air Laws and the Demand for Cigarettes," *Contemporary Policy Issues* 10 (1992): 72–83.
- Chaloupka, Frank J. and Kenneth E. Warner, "The Economics of Smoking," *NBER Working Paper* 7047, 1999.
- Chassin Mark R., and Sally M. McCue. "A Randomized Trial of Medical Quality Assurance. Improving Physicians' Use of Pelvimetry." *Journal of the American Medical Association* 256 (1986):1012–1016.
- Chatterji, Pinka and Jeff DeSimone, "Adolescent Drinking and High School Dropout," *NBER Working Paper* 11337, 2005.
- Chen, Lin Arjan de Haan, Zhang Xiulan, and Ward Warmerdam, "Addressing Vulnerability in an Emerging Economy: China's New Cooperative Medical Scheme (NCMS)," Institute of Social Studies, Erasmus University, 2011.
- Cheney, Catherine, "Hungary Introduces 'Fat Tax'," *Spiegel Online*, <http://www.spiegel.de/international/europe/0,1518,783862,00.html>, accessed November 8, 2011.
- Chernew, Michael and Dennis P. Scanlon, "Health Plan Report Cards and Insurance Choice," *Inquiry* 35 (1998): 9–22.
- Chernew, Michael C. and Joseph P. Newhouse, "What Does the RAND Health Insurance Experiment Tell Us About the Impact of Cost Sharing on Health Outcomes?" *American Journal of Managed Care* 14 (2008): 412–414.
- Chetty, V. K., "Stochastic Technology, Production Organization and Costs," *Journal of Health Economics* 17 (1998): 187–210.
- Chilingerian, Jon, "Evaluating Physician Efficiency in Hospitals: A Multivariate Analysis of Best Practices," *European Journal of Operations Research* 80 (1995): 548–574.
- Chiswick, Barry R., "The Demand for Nursing Home Care: An Analysis of the Substitution Between Institutional and Noninstitutional Care," *Journal of Human Resources* 11 (1976): 295–316.
- Christianson, Jon B. et al., "The HMO Industry: Evolution in Population Demographics and Market Structures," *Medical Care Review* 48 (1991): 3–46.
- Christensen, Eric W. and Richard J. Arnould, "The Impact of Asymmetric Information and Ownership on Nursing Home Access," *International Journal of Health Care Finances and Economics* 5 (2005): 273–297.
- Chou, Shin-Yi, "Asymmetric Information, Ownership and Quality of Care: An Empirical Analysis of Nursing Homes," *Journal of Health Economics* 21 (2002): 293–311.
- Ciliberto, Frederico and Richard Lindrooth, "Exit from the Hospital Industry," *Economic Inquiry* 45 (2007): 71–81.
- Civan, Abdulkadir, and Bülent Köksal, "The Effect of Newer Drugs on Health Spending: Do They Really Increase the Costs?," *Health Economics* 19 (2010): 581–595.
- Clancy, Carolyn and Anne K. Gauthier, (eds.), "Consumer Driven Health Care: Beyond Rhetoric with Research and Experience," *Health Services Research* 30 Suppl. (2004): 1049–1233.
- Clark, Andrew E. and Youenn Loheac. " 'It Wasn't Me, It Was Them!' Social Influence in Risky Behavior by Adolescents," *Journal of Health Economics* 26 (2007): 763–784.
- Clark, Robert Charles, "Does the Nonprofit Form Fit the Hospital Industry?" *Harvard Law Review* 93 (1980): 1417–1489.

- Clarkson, Kenneth W., "The Effects of Research and Promotion on Rates of Return," in *Competitive Strategies in the Pharmaceutical Industry*, Robert B. Helms (ed.) (Washington, DC: American Enterprise Institute, 1996) 238–268.
- Claxton, Karl et al., "Discounting and Cost-effectiveness in NICE—Stepping Back to Sort Out a Confusion," *Health Economics* 15 (2006): 1–4.
- Cleverly, William O., "More Efficient Hospitals Are Closing," *Healthcare Financial Management* 47 (1993): 82–85.
- CNN, "Hawaii Ends Universal Health Care for Kids," [www.cnn.com/2008/HEALTH/10/17/hawaii.health.ap](http://www.cnn.com/2008/HEALTH/10/17/hawaii.health.ap), accessed November 19, 2008.
- Cochrane, Archibald L., *Effectiveness and Efficiency: Random Reflections on Health Services* (London: Nuffield Provincial Hospital Trust, 1972).
- Coddington, Dean C. et al., *The Crisis in Health Care: Costs, Choices, and Strategies*. (San Francisco: Jossey-Bass, 1990).
- Coffey, Rosanna M., "The Effect of Time Price on the Demand for Medical Care Services," *Journal of Human Resources* 18 (1983): 407–424.
- Cogan, John F., R. Glenn Hubbard, and Daniel P. Kessler, *Healthy, Wealthy, & Wise: Five Steps to a Better Health Care System*. (Washington, DC: American Enterprise Institute and Stanford, CA: Hoover Institution, 2005).
- Cogan, John F., R. Glenn Hubbard, and Daniel P. Kessler, "Cost Shifting From the Uninsured: Assessing the Evidence," Stanford University, 2011, unpublished manuscript.
- Cohen, Harold A., "Variations in Cost Among Hospitals of Different Sizes," *Southern Economic Journal* 33 (1967): 355–366.
- Cohen, Joel W. and William D. Spector, "The Effect of Medicaid Reimbursement on Quality of Care in Nursing Homes," *Journal of Health Economics* 15 (1996): 23–48.
- Cohen, Joshua T., Peter J. Neumann, and Milton C. Weinstein, "Dose Preventive Care Save Money? Health Economics and the Presidential Candidates," *New England Journal of Medicine* 358 (2008): 661–663.
- Collins, Robert, and Gregory A. Baker, "Of Junk Food and Junk Science," *International Food and Agribusiness Management Review* 12 (2009): 111–125.
- Collins, Sara R. et al., "Losing Ground: How the Loss of Adequate Health Insurance is Burdening Working Families," Findings from the Commonwealth Fund Biennial Health Insurance Surveys, 2001–2007, The Commonwealth Fund, August 2008a.
- Collins, Sara R. et al., "The 2008 Presidential Candidates' Health Reform Proposals: Choices for America," The Commonwealth Fund, October 2008b.
- Comanor, William S. and Thomas A. Wilson, *Advertising and Market Power*. (Cambridge, MA: Harvard, 1974).
- Comanor, William S. and Thomas A. Wilson, "The Effect of Advertising on Competition: A Survey," *Journal of Economic Literature* 17 (1979): 453–476.
- Comanor, William, H. E. Frech III, and Richard Miller, "Is the United States an Outlier in Health Care and Health Outcomes? A Preliminary Analysis," *International Journal of Health Care Finance and Economics* 6 (2006): 3–33.
- Commission for Healthcare Audit and Inspection, "State of Healthcare 2005," [www.healthcarecommission.org.uk/healthcareproviders/nationalfindings/publications/publications2004-05](http://www.healthcarecommission.org.uk/healthcareproviders/nationalfindings/publications/publications2004-05), accessed August 2, 2008.
- Commonwealth Fund Commission on a High Performance Health System, *Why Not the Best? Results from the National Scorecard on U.S. Health System Performance, 2008*, July 27, 2008. Available at [www.commonwealthfund.org/publications/publications\\_show.htm?doc\\_id=692682](http://www.commonwealthfund.org/publications/publications_show.htm?doc_id=692682).
- Conover, Christopher J. and Frank A. Sloan, "Does Removing Certificate of Need Regulations Lead to a Surge in Health Care Spending?" *Journal of Health Politics, Policy and Law* 23 (1998): 455–481.
- Conrad, Robert F. and Robert P. Strauss, "A Multiple-Output Multiple-Input Model of the Hospital Industry in North Carolina," *Applied Economics* 15 (1983): 341–352.
- Contoyannis, Paul and Andrew Jones, "Socio-economic Status, Health and Lifestyle," *Journal of Health Economics* 23 (2004): 965–995.
- Conway, Karen S. and Andrea Kutinova, "Maternal Health: Does Prenatal Care Make a Difference?" *Health Economics* 15 (2006): 461–488.
- Cook, Philip J. and Christopher Carpenter, "Cigarette Taxes and Youth Smoking: New Evidence from National, State and Local Youth Behavior Surveys," *Journal of Health Economics* 27 (2008): 287–299.
- Cook, Philip J. and Michael J. Moore, "Alcohol," in *Handbook of Health Economics*, Anthony J. Culyer and Joseph P. Newhouse (eds.) (Amsterdam, Elsevier Science B.V., 2000).
- Cook, Philip J. and Michael J. Moore, "Drinking and Schooling," *Journal of Health Economics* 12 (1993): 411–429.
- Cook, Phillip J., Jan Orterman, and Frank A. Sloan, "Are Alcohol Excise Taxes Good for Us? Short and Long Term Effects on Mortality Rates," *NBER Working Paper* 11139, 2005.
- Cookson, Richard, "QALYs and the Capability Approach," *Health Economics* 14 (2005): 817–830.
- Cooper, Philip F. and Alan C. Monheit, "Does Employment-Related Health Insurance Inhibit Job Mobility?" *Inquiry* 30 (1993): 400–416.
- Cooper, Philip F., Kosali I. Simon, and Jessica Vistnes, "A Closer Look at the Managed Care Backlash," *Medical Care* 44 (2006): 14–111.
- Cooper, Richard A. et al., "Economic and Demographic Trends Signal an Impending Physician Shortage," *Health Affairs* 21 (2002): 140–154.
- Corman, Hope and Michael Grossman, "Determinants of Neonatal Mortality Rate in the U.S.: A Reduced Form Model," *Journal of Health Economics* 4 (1985): 213–236.



- Corman, Hope, Theodore J. Joyce, and Michael Grossman, "Birth Outcome Production Function in the United States," *Journal of Human Resources* 22 (1987): 339–360.
- Cortemanche, Charles, and Michael Plotzke, "Does Competition from Ambulatory Surgical Centers Affect Hospital Surgical Output?," *Journal of Health Economics* 29 (2010): 765–773.
- Cost-Effectiveness Analysis Registry (2008), <https://research.tufts-nemc.org/cear/default.aspx>.
- Cowing, Thomas G. and Alphonse G. Holtmann, "Multiproduct Short-Run Hospital Cost Functions: Empirical Evidence and Policy Implications from Cross-Section Data," *Southern Economic Journal* 49 (1983): 637–653.
- Cowing, Thomas G., Alphonse G. Holtmann, and Susan Powers, "Hospital Cost Analysis: A Survey and Evaluation of Recent Studies," *Advances in Health Economics and Health Services Research* 4 (1983).
- Coyle, Douglas, Martin J. Buxton, and Bernie J. O'Brien, "Stratified Cost-Effectiveness Analysis: A Framework for Establishing Efficient Limited Criteria," *Health Economics* 12 (2003): 421–427.
- Coyte, Peter et al., "The Relationship between Variations in Knee Replacement Utilization Rates and the Reported Prevalence of Arthritis in Ontario, Canada," *Journal of Rheumatology* 24 (1997): 2403–2412.
- Creemieux, Pierre-Yves and Pierre Ouellette, "Omitted Variable Bias and Hospital Costs," *Journal of Health Economics* 20 (2001): 271–282.
- Creemieux, Pierre-Yves, Pierre Ouellette, and Caroline Pilon, "Health Care Spending as Determinants of Health Outcomes," *Health Economics* 8 (November 1999): 627–639.
- Cropper, Maureen L. et al., "The Health Benefit of Air Pollution Control in New Delhi," *American Journal of Agricultural Economics* 79 (1997): 1625–1629.
- Cropper, Maureen, James K. Hammit, and Lisa A. Robinson, "Valuing Mortality Risk Reductions: Progress and Challenges," Discussion Paper, Washington, DC, Resources for the Future, (2011).
- Culyer, Anthony J. "The Normative Economics of Health Care Finance and Provision," *Oxford Review of Economic Policy* 5 (1989): 34–58.
- Culyer, Anthony J., "The United Kingdom: Effective, Efficient, Equitable?," *Journal of Health Politics, Policy and Law* 17 (1992): 667–688.
- Culyer, Anthony J. and Adam Wagstaff, "Equity and Equality in Health and Health Care," *Journal of Health Economics* 12 (1993): 431–457.
- Culyer, Anthony J., Eddy van Doorslaer, and Adam Wagstaff, "Access, Utilisation and Equity: A Further Comment," *Journal of Health Economics* 11 (1992a): 207–210.
- Culyer, Anthony J., Eddy van Doorslaer, and Adam Wagstaff, "Utilisation as a Measure of Equity by Mooney, Hall, Donaldson and Gerard," *Journal of Health Economics* 11 (1992b): 93–98.
- Currie, Janet. "The Take-up of Social Benefits," in *Poverty, the Distribution of Income, and Public Policy*, Alan Auerbach, David Card, and John M. Quigley, eds., (New York: Russell Sage, 2006): 80–148.
- Currie, Janet and Nancy Cole, "Welfare and Child Health: The Link between AFDC Participation and Birth Weight," *American Economic Review* 83 (1993): 971–985.
- Currie, Janet and Jonathan Gruber, "Saving Babies: The Efficacy and Cost of Recent Changes in the Medicaid Eligibility of Pregnant Women," *Journal of Political Economy* 104 (1996): 1263–1296.
- Currie, Janet and Patricia B. Reagan, "Distance to Hospital and Children's Use of Preventive Care: Is Being Closer Better, and for Whom?," *Economic Inquiry* 41 (2005): 378–391.
- Custer, W. S. and Richard J. Willke, "Teaching Hospital Costs: The Effect of Medical Staff Characteristics," *Health Services Research* 25 (1991): 831–857.
- Cutler, David M., "The Lifetime Costs and Benefits of Medical Technology," *Journal of Health Economics* 26 (2007): 1081–1100.
- Cutler, David M., *Your Money or Your Life: Strong Medicine for America's Health Care System* (Oxford: Oxford University Press, 2004).
- Cutler, David M. and Adriana Lleras-Muney, "Education and Health Evaluating Theories and Evidence," *NBER Working Paper* 12352, 2006.
- Cutler, David M. and Dan P. Ly, "The (Paper)Work of Medicine: Understanding International Medical Costs," *Journal of Economic Perspectives* 25 (2011): 3–25.
- Cutler, David M. and Grant Miller, "The Role of Public Health Improvements in Health Advances: The Twentieth-Century United States," *Demography* 42 (2005): 1–22.
- Cutler, David M. and Jonathan Gruber, "Does Public Insurance Crowd Out Private Insurance?," *Quarterly Journal of Economics* 111 (1996): 391–430.
- Cutler, David M. and Jonathan Gruber, "Health Policy in the Clinton Era: Once Bitten, Twice Shy," in *American Economic Policy in the 1990s*, Peter R. Orszag and Jeffrey A. Frankel (eds.) (MIT Press: Cambridge, MA, 2002).
- Cutler, David M. and Jill R. Horwitz, "Converting Hospitals from Not-for-Profit to For-Profit Status: Why and What Effects?," *NBER Working Paper* 6672, 1998.
- Cutler, David M. and Ellen Meara, "The Medical Costs of the Young and Old: A Forty-Year Perspective," *NBER Working Paper* 6114, 1997.
- Cutler, David M., Edward L. Glaeser and Jesse M. Shapiro, "Why Have Americans Become More Obese?," *Journal of Economic Perspectives* 17 (2003): 93–118.

- Cutler, David M., Mark McClellan, and Joseph P. Newhouse, "How Does Managed Care Do It?" *RAND Journal of Economics* 31 (2000): 526–548.
- Cutler, David M., Mark McClellan, Joseph P. Newhouse, and Dahlia Remler, "Pricing Heart Attack Treatments," in Cutler, David M. and Ernst R. Berndt (eds.), *Medical Care Output and Productivity* (Chicago: University of Chicago Press, 2001).
- Cutler, David M. and Sarah J. Reber, "The Trade-Off Between Competition and Adverse Selection," *Quarterly Journal of Economics* 113 (1998): 433–466.
- Dalton, Kathleen and Hilda A. Howard, "Market Entry and Exit in Long-Term Care: 1985–2000," *Health Care Financing Review* 24 (2002): 17–32.
- Daniels, Norman, "Equity of Access to Health Care: Some Conceptual and Ethical Issues," *Milbank Memorial Fund Quarterly/Health and Society* 60 (1982): 51–81.
- Daniels, Norman, "Health Care Needs and Distributive Justice," *Philosophy and Public Affairs* 10 (1981): 146–179.
- Daniels, Norman, *Just Health Care*. (Cambridge: Cambridge University Press, 1985).
- David, Guy, and Scott Harrington, "Population Density and Racial Differences in the Performance of Emergency Medical Services," *Journal of Health Economics* 29 (2010): 603–615.
- Davis, Karen and Roger Reynolds, "The Impact of Medicare and Medicaid on Access to Medical Care," in *The Role of Health Insurance in the United States Health Services Sector*, Richard Rosett (ed.) (New York: Neal Watson, 1976).
- Davis, Mark A., "On Nursing Home Quality: A Review and Analysis," *Medical Care Review* 48 (1991): 129–166.
- Danzon, Patricia M., Andrew Epstein, and Sean Nicholson, "Mergers and Acquisitions in the Pharmaceutical and Biotech Industries," *Managerial and Decision Economics* 28 (2007): 307–328.
- Danzon, Patricia M., Y. Richard Wang, and Liang Wang, "The Impact of Price Regulation on the Launch Delay of New Drugs—Evidence from Twenty-Five Major Markets in the 1990s," *Health Economics* 14 (2005): 269–292.
- Deaton, Angus and Christina Paxson, "Mortality, Income, and Income Inequality over Time in Britain and the United States," *NBER Working Paper* 8534, 2001.
- Deaton, Angus, Jane Fortson, and Robert Tortora, "Life (Evaluation), HIV/AIDS, and Death in Africa," in Ed Diener, John Helliwell, and Daniel Kahneman Eds., *International Differences in Well-being*, (Oxford: Oxford University Press, 2010.)
- De Brantes, Francois, Guy D'Andrea, Meredith B. Rosenthal, "Should Health Care Come with a Warranty?" *Health Affairs* Web Exclusive (2009): w678–w687.
- Deb, Partha M. and Ann M. Holmes, "Substitution of Physicians and Other Providers in Outpatient Mental Health Care," *Health Economics* 7 (1998): 347–361.
- Deb, Partha et al., "The Effect of Managed Care on Use of Health Care Services: Results from Two Contemporaneous Household Surveys," *Health Economics* 15 (2006): 743–760.
- Debebe, Zelalem Yilma and Luuk van Kempen, "Ex-ante Moral Hazard in Health Insurance: Effect of the Ghanaian National Health Insurance on Malaria Prevention," presented at Institute of Social Studies, The Hague, Netherlands, February 2011, <http://www.iss.nl/News/Events/Ex-ante-Moral-Hazard-in-Health-Insurance-Effect-of-the-Ghanaian-National-Health-Insurance-on-Malaria-Prevention>, accessed April 21, 2011.
- Decker, Frederic H., "Nursing Home Performance in Resident Care in the United States: Is it Only a Matter of For-Profit versus Not-For-Profit?," *Health Economics, Policy and Law* 3 (2008): 115–140.
- Deer, Brian, "How the Case Against the MMR Vaccine was Fixed," *British Medical Journal* 342 (2011): 77–82.
- Defelice, Lisa C. and W. David Bradford, "Relative Inefficiencies in Production Between Solo and Group Practice Physicians," *Health Economics* 6 (1997): 455–465.
- Deily, Mary and Niccic McKay, "Cost Inefficiency and Mortality Rates in Florida Hospitals," *Health Economics* 115 (2006): 419–431.
- Dekker, T., et al. "The Effect of Risk Context on the Value of a Statistical Life: A Bayesian Meta-model," *Environmental and Resource Economics*, forthcoming (2011).
- Despotakis, Kostas A., "Economic Performance of Flexible Functional Forms," *European Economic Review* 30 (1986): 1107–1143.
- Deutsch, Larry, L., "Pharmaceuticals: The Critical Role of Innovation," in *Industry Studies*, 2nd ed., Larry L. Deutsch (ed.) (Armonk, NY: M.E. Sharpe, 1998).
- D'Hombres, Beatrice et al, "Does Social Capital Determine Health? Evidence From Eight Transition Countries," *Health Economics* 19 (2010): 56–74.
- DiCicca, Phillip et al., "Youth Smoking, Cigarette Prices and Anti-smoking Sentiment," *NBER Working Paper* 12458, 2006.
- Diehr, Paula et al., "What Is Too Much Variation: The Null Hypothesis in Small-Area Analysis," *Health Services Research* 24 (1990): 741–771.
- DiFranza, Joseph R. et al., "RJR Nabisco's Cartoon Camel Promotes Camel Cigarettes to Children," *Journal of the American Medical Association* 266 (December 11, 1991): 3149–3153.
- Dills, Angela K., Mirelle Jacobson, and Jeffrey A. Miron, "The Effect of Alcohol Prohibition on Alcohol Consumption: Evidence from Drunkenness Arrests," *Economics Letters* 86 (2005): 279–284.
- DiMasi, Joseph A., "New Drug Development in the United States from 1963 to 1999," *Clinical Pharmacology & Therapeutics* 16 (2001): 286–296.

- DiMasi, Joseph A., Ronald W. Hansen, and Henry G. Grabowski, "The Price of Innovation: New Estimates of Drug Development Costs," *Journal of Health Economics* 22 (2003): 151–185.
- DiMasi, Joseph A., Ronald W. Hansen, and Henry G. Grabowski, "Reply: Extraordinary Claims Require Extraordinary Evidence," *Journal of Health Economics* 24 (2005): 1034–1044.
- DiMasi, Joseph A. et al., "Cost of Innovation in the Pharmaceutical Industry," *Journal of Health Economics* 10 (1991): 107–142.
- Dixon, Anna, Jessica Greene, and Judith Hibbard, "Do Consumer-Directed Health Plans Drive Change in Enrollees' Health Care Behavior?" *Health Affairs* 27 (2008): 1120–1131.
- Dobkin, Carlos and Steven L. Puller, "The Effects of Government Transfers on Monthly Cycles in Drug Abuse, Hospitalization and Mortality," *Journal of Public Economics* 91 (2007): 2137–2157.
- Dobson, Allen et al., "An Evaluation of Illinois' Certificate of Need Program," Prepared for State of Illinois Commission on Government Forecasting and Accountability, The Lewin Group, February 15, 2007, [www.ilga.gov/commission/cgfa2006/Upload/LewinGroupEvalCertOfNeedPresentation.pdf](http://www.ilga.gov/commission/cgfa2006/Upload/LewinGroupEvalCertOfNeedPresentation.pdf), accessed July 22, 2008.
- Dolan, Paul et al. "QALY Maximization and People's Preferences: A Methodological Review of the Literature," *Health Economics* 14 (2005): 197–208.
- Donabedian, Avedis, *The Definition of Quality and Approaches to Its Assessment*. (Ann Arbor, MI: Health Administration Press, 1980).
- Donabedian, Avedis, "Evaluating the Quality of Medical Care," *Milbank Memorial Fund Quarterly/Health and Society* 44 (1966): 166–203.
- Dong, Frederick et al., "Cost Effectiveness of ACE Inhibitor Treatment for Patients with Type 1 Diabetes Mellitus," *Pharmacoeconomics* 22 (2004): 1015–1027.
- Donohue, Julie, "A History of Drug Advertising: The Evolving Role of Consumers and Consumer Protection," *The Milbank Quarterly* 84 (2006): 659–699.
- Dor, Avi, "Non-Minimum Cost Functions and the Stochastic Frontier: On Applications to Health Care Providers," *Journal of Health Economics* 13 (1994): 329–334.
- Dorfman, Robert and Peter O. Steiner, "Optimal Advertising and Optimal Quality," *American Economic Review* 44 (1954): 826–836.
- Dorn, Stan, Bowen Garrett, John Holahan, and Aimee Williams, "Medicaid, SCHIP and Economic Downturn: Policy Challenges and Policy Responses," Urban Institute, February 15, 2008, [www.allhealth.org/briefing\\_detail.asp?bi=121](http://www.allhealth.org/briefing_detail.asp?bi=121).
- Dougherty, Conor, "Slowdown's Side Effect: More Nurses," *Wall Street Journal*, May 7, 2008, p. D1.
- Douglas, Stratford and Govind Hariharan, "The Hazard of Starting Smoking: Estimates from a Split Population Duration Model," *Journal of Health Economics* 13 (1994): 213–230.
- Doyle, Joseph, "Health Insurance, Treatment and Outcomes: Using Auto Accidents As Health Shocks," *The Review of Economics and Statistics* 87 (2005): 256–270.
- Doyle, Joseph J., Jr., Steven M. Ewer, and Todd H. Wagner, "Returns to Physician Human Capital: Evidence from Patients Randomized to Physician Teams," *Journal of Health Economics* 29 (2010): 866–882.
- Dranove, David, "Demand Inducement and the Physician/Patient Relationship," *Economic Inquiry* 26 (1988): 281–298.
- Dranove, David, *The Economic Evolution of American Health Care*. (Princeton, NJ: Princeton University Press, 2000).
- Dranove, David, "A Problem with Consumer Surplus Measures of the Cost of Practice Variations," *Journal of Health Economics* 14 (1995): 243–251.
- Dranove, David and Kenneth Cone, "Do State Rate Setting Regulations Really Lower Hospital Expenses?" *Journal of Health Economics* 4 (1985): 159–165.
- Dranove, David and Richard Lindrooth, "Hospital Consolidation and Costs: Another Look at the Evidence," *Journal of Health Economics* 22 (2003): 983–997.
- Dranove, David and David Meltzer, "Do Important Drugs Reach the Market Sooner?" *RAND Journal of Economics* 25 (1994): 402–423.
- Dranove, David, Carol J. Simon, and William D. White, "Determinants of Managed Care Penetration," *Journal of Health Economics* 17 (1998): 729–745.
- Dranove, David, Carol J. Simon, and William D. White, "Is Managed Care Leading to Consolidation in Health-care Markets?" *Health Services Research* 37 (2002): 573–594.
- Dranove, David and Paul Wehner, "Physician Induced Demand for Childbirths," *Journal of Health Economics* 13 (1994): 61–73.
- Dranove, David and William D. White, "Agency and the Organization of Health Care Delivery," *Inquiry* 24 (1987): 405–415.
- Dranove, David et al., "Is the Impact of Managed Care on Hospital Prices Decreasing?" *Journal of Health Economics* 27 (2008): 362–376.
- Drummond, Michael F. et al., *Methods for the Economic Evaluation of Health Care Programmes*, 2nd ed. (New York: Oxford University Press, 1997).
- Dubos, Rene, *The Mirage of Health*. (London: George Allen and Unwin, Ltd., 1960).
- Duclos, Jean-Yves and Damien Échevin, "Health and Income: A Robust Comparison of Canada and the US," *Journal of Health Economics* 30 (2011): 293–302.
- Dunn, Daniel L. et al., "Economies of Scope in Physicians' Work: The Performance of Multiple Surgery," *Inquiry* 32 (1995): 87–101.

- Dyck, Frank J., et al., "Effect of Surveillance on the Number of Hysterectomies in the Province of Saskatchewan," *New England Journal of Medicine* 296 (1977): 1326–1328.
- Eakin, B. Kelly and Thomas J. Kniesner, "Estimating a Non-Minimum Cost Function for Hospitals," *Southern Economic Journal* 54 (1988): 583–597.
- Easley, David and Maureen O'Hara, "The Economic Role of the Nonprofit Firm," *Bell Journal of Economics* 14 (1983): 531–538.
- Economic Report of the President, transmitted to the Congress, February 1985. (Washington, DC: United States Government Printing Office, 1985).
- Editorial Staff, *New England Journal of Medicine*, "PPACA Can Help Contain Growth in Healthcare Costs", May 12, 2010, [http://www.healthimaging.com/index.php?option=com\\_articles&article=22178](http://www.healthimaging.com/index.php?option=com_articles&article=22178), accessed June 24, 2011.
- Eddy, David M., "Clinical Decision Making: From Theory to Practice," *Journal of the American Medical Association* 263 (January 19, 1990): 441–443.
- Efron, Bradley and Robert Tibshirani, *An Introduction to the Bootstrap* (London: Chapman and Hall, 1993).
- Eggleston, Karen et al., "Health Service Delivery in China: A Literature Review", *Health Economics* 17 (2008a): 149–165.
- Eggleston, Karen et al., "Hospital Ownership and Quality Of Care: What Explains The Different Results In The Literature?" *Health Economics* 17 (2008b): 1345–1362.
- Eisenberg, J. M., *Doctors' Decisions and the Cost of Medical Care* (Ann Arbor, MI: Health Administration Press, 1986).
- Elixhauser, Anne (ed.), "Health Care Cost-Benefit and Cost Effectiveness Analysis (CBA/CEA)," *Medical Care* 31 (1993): JS1–JS150.
- Elixhauser, Anne et al., "Health Care CBA and CEA from 1991 to 1996: An Updated Bibliography," *Medical Care* 36 (1998): MS1–MS9.
- Ellis, Randall P., "Creaming, Skimping and Dumping: Provider Competition on the Intensive and Extensive Margins," *Journal of Health Economics* 17 (1998): 537–555.
- Ellis, Randall P., "Hospital Cost Function Estimation When Firms May Not Try to Minimize Total Costs," mimeo. (Boston: Boston University, 1993).
- Ellis, Randall P. and Thomas G. McGuire, "Provider Behavior Under Prospective Reimbursement," *Journal of Health Economics* 5 (1986): 129–151.
- Elmendorf, Douglas W. "Estimate of the Direct Spending and Revenue Effects of an Amendment in the Nature of a Substitute to H.R. 4872, the Reconciliation Act of 2010," Congressional Budget Office letter to Speaker Nancy Pelosi, March 20, 2010.
- Elzinga, Kenneth G. and T. F. Hogarty, "The Problem of Geographic Market Definition Revisited: The Case of Coal," *Antitrust Bulletin* 23 (1978): 1–18.
- Emanuel, Ezekiel J., and Victor R. Fuchs, "Who Really Pays for Health Care? The Myth of 'Shared Responsibility'" *Journal of the American Medical Association* 299 (2008): 1057–1059.
- Emmons, David W., José R. Guardado, and Carol K. Kane, *Competition in Health Insurance*. (Chicago: American Medical Association, 2010.)
- Enthoven, Alain C., "National Health Service: Some Reforms That Might Be Politically Feasible," *The Economist* 22 (1985): 61–64.
- Escarce, José J., "Explaining the Association Between Surgeon Supply and Utilization," *Inquiry* 29 (1992): 403–415.
- Escarce, José J., "Externalities in Hospitals and Physicians Adoption of a New Surgical Technology: An Exploratory Analysis," *Journal of Health Economics* 15 (1996): 715–734.
- Escarce, José J., "Would Eliminating Differences in Physician Practice Style Reduce Geographic Variations in Cataract Surgery Rates?" *Medical Care* 12 (1993): 1106–1118.
- Escarce, José J. and Mark V. Pauly, "Physician Opportunity Costs in Physician Practice Cost Functions," *Journal of Health Economics* 17 (1998): 129–151.
- Escarce, José J. et al., "Diffusion of Laparoscopic Cholecystectomy Among General Surgeons in the United States," *Medical Care* 33 (1995): 256–271.
- Eskin, Gerald J., "A Case for Test Marketing Experiments," *Journal of Advertising Research* 15 (1975): 27–33.
- Eskin, Gerald J. and Penny H. Barron, "Effect of Price and Advertising in Test-Market Experiments," *Journal of Marketing Research* 14 (1977): 499–508.
- Evans, Robert G., "Behavioral Cost Functions for Hospitals," *Canadian Journal of Economics* 4 (1971): 198–215.
- Evans, Robert G., "Supplier-Induced Demand; Some Empirical Evidence and Implications," in *The Economics of Health and Medical Care*, Mark Perlman (ed.) (London: Macmillan, 1974).
- Evans, William N. and Diana S. Lien, "The Benefits of Prenatal Care: Evidence from the PAT Bus Strike," *Journal of Econometrics* 125 (March 2005): 207–239.
- Evans, William N. and Jeanne S. Ringel, "Can Higher Cigarette Taxes Improve Birth Outcome?" *Journal of Public Economics* 72 (1999): 135–154.
- Fang, Hai, Hong Liu, and John A. Rizzo, "Has the Use of Physician Gatekeepers Declined among HMOs? Evidence from the United States," *International Journal of Health Care Finance and Economics* 9 (2009): 183–195.
- Fang, Hai, and John A. Rizzo, "Has the Influence of Managed Care Waned? Evidence from the Market for Physician Services," *International Journal of Health Care Finance and Economics* 10 (2010): 85–103.
- Fang, Hai et al. "Demanding Customers: Consumerist Patients and Quality of Care," Harvard Working Paper, Cambridge, 2008.

- Fare, Rolf and C. A. Knox Lovell, "Measuring the Technical Efficiency of Production," *Journal of Economic Theory* 19 (1978): 150–162.
- Farmer, Paul and Edward Nardell, "Editorial: Nihilism and Pragmatism in Tuberculosis Control," *American Journal of Public Health* 88 (1998): 1014–1015.
- Farsi, Mehdi, "Changes in Hospital Quality After Conversion in Ownership Status," *International Journal of Health Care Finance and Economics* 4 (2004): 211–230.
- Farsi, Mehdi, "The Temporal Variation of Cost-Efficiency In Switzerland's Hospitals: An Application Of Mixed Models," *Journal of Productivity Analysis*, 30 (2008): 155–168.
- Farsi, Mehdi and Massimo Philippini, "Effects of Ownership, Subsidization and Teaching: Activities on Hospital Costs in Switzerland," *Health Economics* 17 (2008): 335–350.
- Fauci, Anthony S. and H. Clifford Lane, "Human Immunodeficiency Viral Disease: AIDS and Related Disorders," *Harrison's Principles of Internal Medicine*, 16th ed., 2005.
- Feinglass, Joe and James J. Holloway, "The Initial Impact of the Medicare Prospective Payment System on U.S. Health Care: A Review of the Literature," *Medical Care Review* 48 (1991): 91–115.
- Feinstein, Jonathon S., "The Relationship Between Socioeconomic Status and Health: A Review of the Literature," *Milbank Quarterly* 71 (1993): 279–322.
- Feldman, Roger and J. Begun, "Effects of Advertising: Lessons from Optometry," *Journal of Human Resources* 13 (Supplement 1978): 247–262.
- Feldman, Roger and Bryan Dowd, "Is There a Competitive Market for Hospital Services?" *Journal of Health Economics* 5 (1986): 277–292.
- Feldman, Roger and Bryan Dowd, "A New Estimate of the Welfare Loss of Excess Health Insurance," *American Economic Review* 81 (1991): 297–301.
- Feldman, Roger and Frank Sloan, "Competition Among Physicians Revisited," in *Competition in the Health Care Sector: Ten Years Later*, Warren Greenberg (ed.) (Durham, NC: Duke University Press, 1988).
- Feldman, Roger et al., "The Effect of Premiums on the Small Firm's Decision to Offer Health Insurance," *Journal of Human Resources* 32 (1998): 635–658.
- Feldman, Roger et al., "Effects of HMOs on the Creation of Competitive Markets for Hospital Services," *Journal of Health Economics* 9 (1990): 207–222.
- Feldman, Roger, Stephen T. Parente, and Jon B. Christianson, "Consumer-Directed Health Plans: Evidence on Spending and Utilization," *Inquiry* 44 (2007): 26–39.
- Feldstein, Martin S., *Economic Analysis for Health Service Efficiency* (Amsterdam: North-Holland, 1967).
- Feldstein, Martin S., "Hospital Cost Inflation: A Study of Nonprofit Price Dynamics," *American Economic Review* 61 (1971): 853–872.
- Feldstein, Martin S., "The Rising Price of Physicians' Services," *Review of Economics and Statistics* 52 (1970): 121–133.
- Feldstein, Martin S., "The Welfare Loss of Excess Health Insurance," *Journal of Political Economy* 81 (1973): 251–280.
- Feldstein, Paul J., "The Emergence of Market Competition in the U.S. Health Care System. Its Causes, Likely Structure, and Implications," *Health Policy* 6 (1986): 1–20.
- Finkelstein, Amy, "The Aggregate Effects of Health Insurance: Evidence from the Introduction of Medicare," *Quarterly Journal of Economics* 122 (2007): 1–37.
- Finkelstein, Amy and Robin McKnight, "What Did Medicare Do? The Initial Impact of Medicare on Mortality and Out of Pocket Medical Spending," *Journal of Public Economics* 92 (2008): 1644–1669.
- Finkelstein; Eric A. et al, "Annual Medical Spending Attributable To Obesity: Payer-And Service-Specific Estimates," *Health Affairs* 28 (2009): w822–w831.
- Finkler, Merton D. and David D. Wirtschafter, "Cost Effectiveness and Data Envelopment Analysis," *Health Care Management Review* 18 (1993): 81–88.
- Fletcher, Suzanne W. and Graham A. Colditz, "Failure of Estrogen Plus Progestin Therapy for Prevention," *Journal of the American Medical Association* 288 (2002): 366–368.
- Flexner, Abraham, *Medical Education in the United States and Canada, Bulletin No. 4* (New York: Carnegie Foundation for the Advancement of Teaching, 1910).
- Fogel, Robert W., "The Fourth Great Awakening & the Future of Egalitarianism" (Chicago: University of Chicago Press, 2000).
- Fogel, Robert W., *The Escape from Hunger and Premature Death, 1700–2100: Europe, America and the Third World* (Cambridge: Cambridge University Press, 2004).
- Folland, Sherman, "Does Community Social Capital Contribute to Population Health?" *Social Science & Medicine* 64 (2007): 2342–2354.
- Folland, Sherman, "An Economic Model of Social Capital and Health," *Health Economics, Policy and Law* 3 (2008): 333–348.
- Folland, Sherman, "Value of Life and Behavior Toward Health Risks: An Interpretation of Social Capital," *Health Economics* 15 (2006): 159–171.
- Folland, Sherman T., "A Critique of Pure Need: An Analysis of Norman Daniels's Concept of Health Care Need," *International Journal of Social Economics* 17 (1990): 36–45.
- Folland, Sherman T., "Health Care in Small Areas of Three Command Economies: What Do the Data Tell Us?" *Eastern European Economics* 43 (2005): 31–52.
- Folland, Sherman T., "The Quality of Mercy: Social Health Insurance in the Charitable Liberal State," *International Journal of Health Care Finance and Economics* 5 (2005): 23–46.
- Folland, Sherman T., "Value of Life and Behavior Toward Risk: An Interpretation of Social Capital," *Health Economics* 14 (2005): 159–171.

- Folland, Sherman T., "Health Care Need, Economics, and Social Justice," *International Journal of Social Economics* 13 (1986): 98–116.
- Folland, Sherman T. and Richard A. Hofler, "How Reliable Are Hospital Efficiency Estimates? Exploiting the Dual to Homothetic Production," *Health Economics* 10 (2001): 683–698.
- Folland, Sherman T. and Miron Stano, "Small Area Variations: A Critical Review of Propositions, Methods, and Evidence," *Medical Care Review* 47 (1990): 419–465.
- Folland, Sherman T. and Miron Stano, "Sources of Small Area Variations in the Use of Medical Care," *Journal of Health Economics* 8 (1989): 85–107.
- Folland, Sherman, M. Kamrul Islam, and Oddvar Kaarboe, "The Social Capital and Health Hypothesis: An Integrating Theory and New Empirics," Oakland University, 2011.
- Fontana, John P., "Hospital Marketing Is Here to Stay," *Hospital Topics* 62 (1984): 12–13.
- Forder, Julien, "Long-Term Care and Hospital Utilisation by Older People: An Analysis of Substitution Rates," *Health Economics* 18 (2009): 1332–1338.
- Fournier, Gary M. and Jean M. Mitchell, "Hospital Costs and Competition for Services: A Multi-Product Analysis," *Review of Economics and Statistics* 74 (1992): 627–634.
- Fox, William F. and Richard A. Hofler, "Using Homothetic Composed Error Frontiers to Measure Water Utility Efficiency," *Southern Economic Journal* 53 (1986): 461–477.
- Fox News, "Supreme Court Backs Oregon Assisted Suicide Law," [www.foxnews.com/story/0,2933,181881,00.html](http://www.foxnews.com/story/0,2933,181881,00.html), accessed July 7, 2008.
- Frank, Richard G. "Editorial: New Estimates of Drug Development Costs," *Journal of Health Economics* 22 (2003): 325–330.
- Frank, Richard G. and David S. Salkever, "Generic Entry and the Pricing of Pharmaceuticals," *Journal of Economics and Management Strategy* 6 (1997): 75–90.
- Frank, Richard G. and David S. Salkever, "Nonprofit Organizations in the Health Sector," *Journal of Economic Perspectives* 8 (1994): 129–144.
- Frech, H. E., III, *Competition & Monopoly in Medical Care*. (Washington, DC: AEI Press, 1996).
- Frech, H. E., III, *The United States, Advances in Health Economics and Health Services Research Supplement 1: Comparative Health Systems*. (Greenwich, CT: JAI Press, 1990): 43–75.
- Frech, H. E., III and Paul B. Ginsburg, "Competition Among Health Insurers, Revisited," *Journal of Health Politics, Policy and Law* 13 (1988): 279–291.
- Frech, H. E., III and Richard D. Miller Jr. *The Productivity of Health Care and Pharmaceuticals* (Washington, DC: American Enterprise Institute, 1999).
- Frech, H. E., III and Lee R. Mobley, "The Demsetz Hypothesis and Hospital Markets: An Empirical Analysis," Working Papers in Economics 5–96, Department of Economics, University of California, Santa Barbara, CA, 1996.
- Frederick, Shane, George Loewenstein, and Ted O'Donoghue, "Time Discounting and Time Preference: A Critical Review," *Journal of Economic Literature* 40 (2002): 351–401.
- Frenk, Julio, "Editorial: Medical Care and Health Improvement: The Critical Link," *Annals of Internal Medicine* 129 (1998): 419–420.
- Friedman, Bernard and Claudia Steiner, "Does Managed Care Affect the Supply and Use of ICU Services?" *Inquiry* 36 (1999): 68–77.
- Friedman, Milton and Simon Kuznets, *Income from Independent Professional Practice* (New York: National Bureau of Economic Research, 1945).
- Fries, James F., "Aging, Natural Death, and the Compression of Morbidity," *New England Journal of Medicine* 303 (1980): 130–136.
- Fuchs, Victor R., "The Clinton Plan: A Researcher Examines Reform," *Health Affairs* 13 (1994): 102–114.
- Fuchs, Victor R., "Economics, Health, and Post Industrial Society," *Milbank Memorial Fund Quarterly/Health and Society* 57 (1979): 153–182.
- Fuchs, Victor R., "Health Care for the Elderly: How Much? Who Will Pay for It?" *Health Affairs* 18 (1999): 11–21.
- Fuchs, Victor R., "Medicare Reform: The Larger Picture," *Journal of Economic Perspectives* 14 (2000): 57–70.
- Fuchs, Victor R., "The Supply of Surgeons and the Demand for Operations," *Journal of Human Resources* 13 (Supplement 1978): 35–56.
- Fuchs, Victor R., "Time Preference and Health: An Exploratory Study," in *Economic Aspects of Health*, Victor R. Fuchs (ed.) (Chicago: The University of Chicago Press, 1982): 93–120.
- Fuchs, Victor R., "The Future of Health Economics," *Journal of Health Economics* 19 (2000): 141–157.
- Fuchs, Victor R., *Who Shall Live? Health, Economics, and Social Choice, Expanded Edition*. Singapore: World Scientific Publishing Company, 1975.
- Fuchs, Victor R. and James S. Hahn, "How Does Canada Do It? A Comparison of Expenditures for Physicians' Services in the United States and Canada," *New England Journal of Medicine* 323 (1990): 884–890.
- Fuhrman, Vanessa, "Insurer Reveals What Doctors Really Charge," *Wall Street Journal*, August 18, 2005, p. D1.
- Gabel, Jon and Gail A. Jensen, "The Price of State Mandated Benefits," *Inquiry* 26 (1989): 419–432.
- Gagne, Robert and Pierre Thomas Leger, "Determinants of Physicians' Decisions to Specialize," *Health Economics* 14 (2005): 721–735.
- Galbraith, John K., *The Affluent Society*. (Boston: Houghton Mifflin, 1958).
- Galewitz, Phil, "Rising Health Care Costs: Workers Pay Bigger Share," *The Fiscal Times*, September 2, 2010, <http://www>

- .thefiscaltimes.com/Articles/2010/09/02/Workers-Pay-More-for-Health-Care.aspx, accessed May 9, 2011.
- Gallett, Craig A., "Advertising and Restrictions in the Cigarette Industry: Evidence of State-by-State Variation," *Contemporary Economic Policy* 21 (2003): 338–348.
- Garber, Alan, "Advances in Cost-Effectiveness Analysis of Health Interventions," in *Handbook of Health Economics, Volume 1*, A. J. Culyer and Joseph P. Newhouse (eds.) (Amsterdam, Elsevier Science B.V., 2000).
- Garber, Alan M. and Charles E. Phelps, "Economic Foundations of Cost-Effectiveness Analysis," *Journal of Health Economics* 16 (1997): 1–31.
- García-Frapolli, et al., "Beyond Fuelwood Savings: Valuing the Economic Benefits of Introducing Improved Biomass Cookstoves in the Purépecha Region of Mexico," *Ecological Economics* 69 (2010): 2508–2605.
- Gaumer, Gary L., "Regulating Health Professionals: A Review of the Empirical Literature," *Milbank Memorial Fund Quarterly/Health and Society* 62 (1984): 380–416.
- Gaynor, Martin, "What Do We Know About Competition and Quality in Health Care Markets?" *NBER Working Paper* 12301, 2006.
- Gaynor, Martin and Gerard F. Andersen, "Uncertain Demand, the Structure of Hospital Costs, and the Cost of Empty Beds," *NBER Working Paper* 4460, 1995.
- Gaynor, Martin and Mark V. Pauly, "Compensation and Productive Efficiency in Partnerships: Evidence from Medical Group Practice," *Journal of Political Economy* 98 (1990): 544–573.
- Gaynor, Martin and Solomon Polachek, "Measuring Information in the Market: An Application to Physician Services," *Southern Economic Journal* 60 (1994): 815–831.
- Gaynor, Martin and William B. Vogt, "Antitrust and Competition in Health Care Markets," in *Handbook of Health Economics*, Anthony J. Culyer and Joseph P. Newhouse (eds.) (Amsterdam, Elsevier Science B.V., 2000).
- Gaynor, Martin and William B. Vogt, "Competition Among Hospitals," *RAND Journal of Economics* 34 (2003): 764–785.
- Gaynor, Martin, Deborah Haas-Wilson, and William B. Vogt, "Are Invisible Hands Good Hands? Moral Hazard, Competition, and the Second Best in Health Care Markets," *Journal of Political Economy* 108 (2000): 992–1005.
- Goeffard, Pierre-Yves and Tomas Philipson, "Rational Epidemics and Their Public Control," *International Economic Review* 37 (3) (1996): 603–624.
- Gerdtham, Ulf-G and M. Johannesson, "New Estimates of the Demand for Health: Results Based on a Categorical Health Measure and Swedish Micro Data," *Social Science and Medicine* 49 (1999): 1325–1332.
- Gerdtham Ulf-G, and Bengt Jonsson, "International comparisons of health expenditure," in *Handbook of Health Economics, Volume 1A*, Anthony J. Culyer and Joseph P. Newhouse (eds.) (Amsterdam, Elsevier Science, B.V., 2000), 11–53.
- Gerdtham, Ulf-G. et al., "An Econometric Analysis of Health Care Expenditure: A Cross-Section Study of the OECD Countries," *Journal of Health Economics* 11 (1992): 63–84.
- Gertler, Paul J., "Subsidies, Quality, and the Regulation of Nursing Homes," *Journal of Public Economics* 38 (1989): 33–52.
- Getzen, Thomas E., "Health Care Is an Individual Necessity and a National Luxury," *Journal of Health Economics* 19 (2000): 259–270.
- Getzen, Thomas E. and Jean-Pierre Poullier, "International Health Spending Forecasts: Concepts and Evaluation," *Social Science and Medicine* 34 (1992): 1057–1068.
- Giaccotto, Carmelo, Rexford Santerre, and John A. Vernon, "Drug Prices and Research and Development Investment Behavior in the Pharmaceutical Industry," *Journal of Law and Economics* 48 (2005): 195–214.
- Gibson, Teresa R., Ronald J. Ozminowski, and Ron Z. Goetzl, "The Effects of Prescription Drug Cost Sharing: A Review of the Evidence," *American Journal of Managed Care* 11 (2005): 730–740.
- Gillum B. S., E. J. Graves, and E. Wood, *National Hospital Discharge Survey: Annual Summary, 1995*, National Center for Health Statistics, *Vital Health Statistics* 13 (1998).
- Ginsburg, Paul B. and Joy M. Grossman, "When The Price Isn't Right: How Inadvertent Payment Incentives Drive Medical Care," *Health Affairs* (2005): W5-376 – W5-384, <http://content.healthaffairs.org/content/early/2005/08/09/hlthaff.w5.376.citation>.
- Glaeser, Edward L., David Laibson, and Bruce Sacerdote, "An Economic Approach to Social Capital," *Economic Journal* 112 (2002): F437–458.
- Glick, Peter and David E. Sahn, "Health and Productivity in a Heterogeneous Urban Labor Market," *Applied Economics* 30 (1998): 203–216.
- Glied, Sherry, "Managed Care," in *Handbook of Health Economics, Volume 1A*, Anthony J. Culyer and Joseph P. Newhouse (eds.) (Amsterdam, Elsevier Science, B.V., 2000).
- Global Price Reporting Mechanism, "Transaction prices for Antiretroviral Medicines and HIV Diagnostics from 2008 to March 2010," World Health Organization. Geneva 2010.
- Goddeeris, John H., "Insurance and Incentives for Innovation in Medical Care," *Southern Economic Journal* 51 (1984): 530–539.
- Goddeeris, John H. and Burton A. Weisbrod, "Conversion from Nonprofit to For-Profit Legal Status: Why Does It Happen and Should Anyone Care?" *Journal of Policy Analysis and Management* 17 (1998): 215–233.
- Gold, Marthe R. et al. (eds). *Cost Effectiveness in Health and Medicine* (New York: Oxford University Press, 1996).

- Goldberg, Lawrence G. and Warren Greenberg, "The Effect of Physician-Controlled Health Insurance," *Journal of Health Politics, Policy and Law* 2 (1977): 48–78.
- Goldberg, Theodore et al., "Evaluation of Economic Effects of Drug Product Selection Legislation," *Medical Care* 17 (1979): 411–419.
- Goldman, Dana P. et al., "Pharmacy Benefits and the Use of Drugs by the Chronically Ill," *Journal of the American Medical Association* 291 (2004): 2344–2350.
- Goldman, Dana P., "Pressure from Rising Health-Care Costs: How Can Consumers Get Relief?" *Riverside Press Enterprise*, October 24, 2005.
- Goldman, Dana P. et al., "Redistributional Consequences of Community Rating," *Health Services Research* 32 (1997): 71–86.
- Goldstein, Karen P. et al., "The Effect of Epidemic Measles on Immunization Rates," *Journal of the American Medical Association* 276 (1996): 56–58.
- Golec, Joseph and John A. Vernon, "New Estimates of Pharmaceutical Research and Development Spending by US-based Firms from 1984 to 2003," *Managerial and Decision Economics* 28 (2007): 481–483.
- Goodman, Allen C., "Estimation of Offset and Income Effects on the Demand for Mental Health," *Inquiry* 26 (1989): 235–248.
- Goodman, Allen C. and Miron Stano, "HMOs and Health Externalities: A Local Public Good Perspective," *Public Finance Review* 28 (2000): 247–269.
- Goodman, Allen C., Miron Stano, and John M. Tilford, "Applications and Extensions of the Grossman Health Care Model," *Southern Economic Journal* 65 (1999): 791–806.
- Goodman, Clifford S., "Healthcare Technology Assessment: Methods, Framework, and Role in Policy Making," *American Journal of Managed Care* 4 (1998: Special Issue): SP200–SP213.
- Gordon, Margaret S., *Social Security Policies in Industrial Countries*. (Cambridge: Cambridge University Press, 1988).
- Grabowski, David C., "Medicare and Medicaid: Conflicting Incentives for Long-Term Care," *The Milbank Quarterly* 85 (2007): 579–610.
- Grabowski, David C. and Jonathan Gruber, "Moral Hazard in Nursing Home Use," *Journal of Health Economics* 26 (2007): 560–577.
- Grabowski, David C. and Richard Hirth, "Competitive Spillovers Across Nonprofit and For-profit Homes," *Journal of Health Economics* 22 (2003): 1–22.
- Grabowski, Henry G. and John M. Vernon, "Brand Loyalty, Entry, and Price Competition in Pharmaceuticals After the 1984 Drug Act," *Journal of Law and Economics* 35 (1992): 331–350.
- Grabowski, Henry G. and John M. Vernon, "Prospects for Returns to Pharmaceutical R&D Under Health Care Reform," in *Competitive Strategies in the Pharmaceutical Industry*, Robert B. Helms (ed.) (Washington, DC: American Enterprise Institute, 1996): 194–207.
- Grabowski, Henry G. and John M. Vernon, "Returns to R&D on New Drug Innovations in the 1980s," *Journal of Health Economics* 13 (1994): 383–406.
- Graddy, Elizabeth, "Interest Groups or the Public Interest—Why Do We Regulate Health Occupations?" *Journal of Health Politics, Policy and Law* 16 (1991): 25–49.
- Grady, Denise, "After Caesareans, Some See Higher Insurance Cost," *New York Times*, June 1, 2008, p. 1, [www.nytimes.com/2008/06/01/health/01insure.html?\\_r=1&hp&oref=slogin](http://www.nytimes.com/2008/06/01/health/01insure.html?_r=1&hp&oref=slogin), accessed June 1, 2008.
- Grannemann, Thomas W., Randall S. Brown, and Mark V. Pauly, "Estimating Hospital Costs: A Multiple-Output Analysis," *Journal of Health Economics* 5 (1986): 107–127.
- Grant, Darren, "Physician Financial Incentives and Cesarean Delivery: New Conclusions from the Healthcare Cost and Utilization Project," *Journal of Health Economics* 28 (2009): 244–250.
- Greene, Jessica et al., "The Impact of Consumer-Directed Health Plans on Prescription Drug Use," *Health Affairs* 27 (2008a): 1111–1119.
- Greene, Jessica et al., "Comprehension and Choice of a Consumer-Directed Health Plan: An Experimental Study," *American Journal of Managed Care* 14 (2008b): 369–376.
- Green, Lee A. and Mark P. Becker, "Physician Decision Making and Variation in Hospital Admission Rates for Suspected Acute Cardiac Ischemia," *Medical Care* 32 (1994): 1086–1097.
- Grosskopf, Shawna and Vivian Valdimanis, "Measuring Hospital Performance: A Non-Parametric Approach," *Journal of Health Economics* 6 (1987): 89–107.
- Grossman, Michael, *The Demand for Health: A Theoretical and Empirical Investigation*. [New York: Columbia (for the National Bureau of Economic Research), 1972a].
- Grossman, Michael, "On the Concept of Health Capital and the Demand for Health," *Journal of Political Economy* 80 (1972b): 223–255.
- Grossman, Michael and Theodore J. Joyce, "Unobservables, Pregnancy Resolutions, and Birth Weight Production Functions in New York City," *Journal of Political Economy* 98 (1990): 983–1007.
- Grossman, Michael, Frank J. Chaloupka, and Ismail Sirtalan, "An Empirical Analysis of Alcohol Addiction: Results from the Monitoring the Future Panels," *Economic Inquiry* 36 (1998): 39–48.
- Gruber, Jonathan, "Covering the Uninsured in the United States," *Journal of Economic Literature* 46 (2008): 571–606.
- Gruber, Jonathan, "The Incidence of Mandated Maternity Benefits," *American Economic Review* 84 (1994): 622–641.



- Gruber, Jonathan, "Medicaid," in *Means-Tested Transfers in the U.S.*, Robert A. Moffit (ed.) (Chicago: University of Chicago Press, 2002).
- Gruber, Jonathan and Daniel M. Hungerman, "Faith-based Charity and Crowd-out During the Great Depression," *Journal of Public Economics* 91 (2007): 1043–1069.
- Gruber Jonathan and Botond Köszegi, "Is Addiction 'Rational'? Theory and Evidence," *The Quarterly Journal of Economics* 116 (2001): 1261–1303.
- Gruber, Jonathan and Alan B. Krueger, "The Incidence of Mandated Employer Provided Health Insurance: Lessons from Workers' Compensation Insurance," in *Health Benefits and the Workforce*, David Bradford (ed.) (Washington, DC: U.S. Government Printing Office, 1992).
- Gruber, Jonathan and Brigitte C. Madrian, "Health Insurance, Labor Supply, and Job Mobility: A Critical Review of the Literature," *NBER Working Paper* 8817, March 2002.
- Gruber, Jonathan and Maria Owings, "Physician Financial Incentives and the Diffusion of Cesarean Section Delivery," *RAND Journal of Economics* 27 (1996): 99–123.
- Gruber, Jonathan and Jonathan Zinman, "Youth Smoking in the U.S.: Evidence and Implications," *NBER Working Paper* 8872, 2002.
- Grytten, Jostein, Irene Skau, and Rune Sørensen, "Do Expert Patients get Better Treatment than Others? Agency Discrimination and Statistical Discrimination in Obstetrics," *Journal of Health Economics* 30 (2011): 163–180.
- Grytten, Jostein and Rune Sørensen, "Type of Contract and Supplier-Induced Demand for Primary Physicians in Norway," *Journal of Health Economics* 20 (2001): 379–393.
- Grytten, Jostein and Rune Sørensen, "Primary Physician Services—List Size and Primary Physician Service Production," *Journal of Health Economics* 26 (2007): 742–762.
- Grytten, Jostein, and Rune Sørensen, "Patient Choice and Access to Physician Services in Norway," *Health Economics, Policy and Law* 4 (2008): 11–27.
- Gulley, O. David and Rexford E. Santerre, "Market Structure Elements: The Case of California Nursing Homes," *Journal of Health Care Finance* 33 (2007): 1–16.
- Gulley, O. David and Rexford E. Santerre, "The Effect of Tax Exemption on the Market Share of Nonprofit Hospitals," *National Tax Journal* 46 (1993): 477–486.
- Haas-Wilson, Deborah, "Consumer Information and Providers' Reputations: An Empirical Test in the Market for Psychotherapy," *Journal of Health Economics* 9 (1990): 321–333.
- Hadley, Jack and Stephen Zuckerman, "The Role of Efficiency in Hospital Rate Setting," *Journal of Health Economics* 13 (1994): 335–340.
- Hadley, Jack et al., "Covering the Uninsured in 2008: Current Costs, Sources of Payment, and Incremental Costs," *Health Affairs*, Web Exclusive, 2008: w399–w415.content.healthaffairs.org/cgi/reprint/27/5/w399, accessed January 9, 2009.
- Hakim, Danny, "State Watch for 2 Percent of Doctors," *New York Times*, May 7, 2008, www.nytimes.com/2008/05/07/myregion/07doctors.html.
- Hall, Mark A. and Christopher J. Conover, "The Impact of Blue Cross Conversions on Accessibility, Affordability, and the Public Interest," *The Milbank Quarterly* 81 (2003): 509–542.
- Hall, Thomas D. and Cotton M. Lindsay, "Medical Schools: Producers of What? Sellers to Whom?" *Journal of Law and Economics* 23 (1980): 55–80.
- Hamilton, James L., "The Demand for Cigarettes: Advertising, the Health Scare, and the Cigarette Advertising Ban," *Review of Economics and Statistics* 54 (1972):401–411.
- Hamilton, James L., "The Effect of Cigarette Advertising Bans on Cigarette Consumption," Proceedings of the Third World Conference on Smoking and Health (Washington, DC: U.S. DHEW, 1975).
- Hamilton, Vivian, "The Medicare Hospice Benefit: The Effectiveness of Price Incentives in Health Care Policy," *RAND Journal of Economics* 24 (1993): 605–624.
- Hanewinkel, Reiner, Christian Radden, and Tobias Rosenkranz, "Price Increase Causes Fewer Sales of Factory-made Cigarettes and Higher Sales of Cheaper Loose Tobacco in Germany," *Health Economics* 17 (2007): 683–693.
- Hanratty, Maria, "Canadian National Health Insurance and Infant Health," *American Economic Review* 86 (1996): 276–284.
- Hansen, W. Lee, "Shortages and Investment in Health Manpower," in *The Economics of Health and Medical Care: Proceedings of the Conference on the Economics of Health and Medical Care*, May 10–12, 1962. (Ann Arbor, MI: University of Michigan, 1964).
- Hansmann, Henry, Daniel Kessler, and Mark McClellan, "Ownership Form and Trapped Capital in the Hospital Industry," *NBER Working Paper* 8989, 2002.
- Hansmann, Henry B., "The Effect of Tax Exemption and Other Factors on the Market Share of Nonprofit Versus For-Profit Firms," *National Tax Journal* 40 (1987): 71–82.
- Hansmann, Henry B., "The Role of Nonprofit Enterprise," *Yale Law Journal* 89 (1980): 835–901.
- Harrington, Charlene, James H. Swan, and Leslie A. Grant, "Nursing Home Bed Capacity in the States, 1978–86," *Health Care Financing Review* 9 (1988): 81–97.
- Harris, Jeffrey E., "The Internal Organization of Hospitals: Some Economic Implications," *Bell Journal of Economics* 8 (1977): 467–482.
- Harris, Jeffrey E. and Sandra W. Chan, "The Continuum-of-Addiction: Cigarette Smoking in Relation to Price Among Americans," *Health Economics Letters* 2 (1998): 3–12.

- Harris, Jeffrey E. and Sandra W. Chan, "The Continuum-of-Addiction: Cigarette Smoking in Relation to Price Among Americans Aged 15–29," *Health Economics* 8 (1999): 81–86.
- Harris, Jennifer L. et al., "A Crisis in the Marketplace: How Food Marketing Contributes to Childhood Obesity and What Can be Done," *Annual Review of Public Health* 30 (2009): 211–225.
- Harrison, Teresa, "Consolidation and Closures: An Empirical Analysis of Exits from the Hospital Industry," *Health Economics* 16 (2007): 457–474.
- Hassan, Mahmud, Gerard Wedig, and Michael A. Morrissey, "Charity Care by Non-Profit Hospitals: The Price of Tax-Exempt Debt," *International Journal of the Economics of Business* 7 (2000): 47–62.
- Hay, Joel W. and Gerard Anderson, "The Hospital Services Market: A Disequilibrium Analysis," *Southern Economic Journal* 54 (1988): 656–665.
- Hayek, Friedrich, *The Constitution of Liberty* (Chicago: The University of Chicago Press, 1960).
- Headen, Alvin E., Jr., "Wages, Returns to Ownership, and Fee Responses to Physician Supply," *Review of Economics and Statistics* 72 (1990): 30–37.
- Helland, Eric and Mark Showalter, "The Impact of Liability on the Physician Labor Market" *Journal of Law and Economics* 52 (2009): 635–663.
- Henderson, Rebecca and Iain Cockburn, "Scale, Scope, and Spillovers: The Determinants of Research Productivity in Drug Discovery," *RAND Journal of Economics* 27 (1996): 32–59.
- Henry J. Kaiser Family Foundation, "Health Insurance Coverage in America—2006 Data Update," Tables 1, 9; [www.kff.org/uninsured/7451.cfm](http://www.kff.org/uninsured/7451.cfm), accessed June 3, 2008.
- Henry J. Kaiser Family Foundation, "CHIP Enrollment: December 2009 Data Snapshot" #7642-05, <http://www.kff.org/medicaid/upload/7642-05.pdf>, 2011, accessed June 8, 2011.
- Hethcote, Herbert W., "The Mathematics of Infectious Diseases," *SIAM Review* 42 (2000): 599–653.
- Hill, Steven C. and Barbara L. Wolfe, "Testing the HMO Competitive Strategy: An Analysis of Its Impact on Medical Care Resources," *Journal of Health Economics* 16 (1997): 261–286.
- Hilsenrath, Peter, "Fiscal Policy Dilemmas and Health Spending in South Africa," *Journal of Health and Human Services Administration* 21 (1999): 310–324.
- Hirsch, Barry T. and Edwin J. Schumaker, "Classic or New Monopsony? Searching for Evidence in Nursing Labor Markets," *Journal of Health Economics* 24 (2005): 969–989.
- Hirsch, Barry T. and Edward Schumacher, "Monopsony Power and Relative Wages in the Labor Market for Nurses," *Journal of Health Economics* 14 (1995): 443–476.
- Hirsch, Barry T. and Edward Schumacher, "Union Wages, Rents, and Skills in Health Care Labor Markets," *Journal of Labor Research* 19 (1998): 125–147.
- Hirshleifer, Jack and Amihai Glazer, *Price Theory and Applications*, 5th ed. (Upper Saddle River, NJ: Prentice Hall, 1992).
- Hirth, Richard A., "Consumer Information and Competition Between Nonprofit and For-Profit Nursing Homes," *Journal of Health Economics* 18 (1999): 219–240.
- Hirth, Richard A., Michael E. Chernew, and Sean M. Orzol, "Ownership, Competition, and the Adoption of New Technologies and Cost Saving Practices in a Fixed-Price Environment," *Inquiry* 37 (2000): 282–294.
- Ho, Katherine and Ariel Pakes, "Do Physician Incentives Affect Hospital Choice? A Progress Report," *International Journal of Industrial Organization* 29 (2011): 317–322.
- Hoadley, Jack et al. "The Medicare Part D Coverage Gap: Costs and Consequences in 2007," Kaiser Family Foundation, August 2007, <http://www.kff.org/medicare/upload/7811.pdf>, accessed June 7, 2011.
- Hodgson, Thomas A., "Cigarette Smoking and Lifetime Medical Expenditures," *Milbank Quarterly* 70 (1992): 81–125.
- Hoff, Timothy J., "Doing the Same and Earning Less: Male and Female Physicians in a New Medical Specialty," *Inquiry* 41 (2004): 301–315.
- Holahan, John and David Liska, "Variations in Medicaid Spending Among States," *Assessing New Federalism: Issues and Options for States*, A–3. (Washington, DC: Urban Institute, January 1997).
- Holahan, John and Arunabh Ghosh, "Understanding the Recent Growth in Medicaid Spending, 2000–2003," *Health Affairs*, Web Exclusive, Exhibit 6, DOI 10.1377/hlthaff.W5.52. Available at <http://content.healthaffairs.org/cgi/content/short/hlthaff.w5.52>, accessed April 25, 2009.
- Holahan, John, Joshua Wiener, and Susan Wallin, "Health Policy for the Low-Income Population: Major Findings from the Assessing the New Federalism Case Studies." Washington, DC: Urban Institute, November 1998.
- Hollingsworth, Bruce, "The Measurement Of Efficiency And Productivity Of Health Care Delivery," *Health Economics* 17 (2008): 1107–1128.
- Holtmann, Alphonse G., "Theories of Non-Profit Institutions," *Journal of Economic Surveys* 2 (1988): 29–45.
- Holtmann, Alphonse G., "A Theory of Non-Profit Firms," *Economica* 50 (1983): 439–449.
- Holtmann, Alphonse G. and Todd L. Idson, "Wage Determination of Registered Nurses in Proprietary and Nonprofit Nursing Homes," *Journal of Human Resources* 28 (1993): 55–79.
- Hornbrook, Mark C. and Alan C. Monheit, "The Contribution of Case-Mix Severity to the Hospital Cost-Output Relation," *Inquiry* 22 (1985): 259–271.

- Howard, David H., "Quality and Consumer Choice in Healthcare: Evidence from Kidney Transplantation," *Topics in Economic Analysis & Policy* 5 (2005): Article 24.
- Hsiao, William C., "Resource-Based Relative Value Scale: An Option for Physician Payment," *Inquiry* 24 (1987): 360–361.
- Hsiao, William C., Daniel L. Dunn, and Diana K. Verrilli, "Assessing the Implementation of Physician-Payment Reform," *New England Journal of Medicine* 328 (1993): 928–933.
- Hughes, David and Alistair McGuire, "Stochastic Demand, Production Responses and Hospital Costs," *Journal of Health Economics* 22 (2003): 999–1010.
- Hughes, Susan L. et al., "Impact of Home Care on Hospital Days: A Meta Analysis," *Health Services Research* 32 (1997): 415–432.
- Hughes, Bethan, "2007 FDA Drug Approvals: A Year in Flux," *Nature Reviews Drug Discovery* 7 (2008): 107–109.
- Hurd, Michael D. and Kathleen McGarry, "Medical Insurance and the Use of Health Care Services by the Elderly," *Journal of Health Economics* 16 (1997): 129–154.
- Hurley, Jeremiah, "An Overview of the Normative Economics of the Health Sector," in *Handbook of Health Economics*, Volume I, A. J. Culyer and J. P. Newhouse (eds.) (Amsterdam, Elsevier Science B.V., 2000).
- Hurley, Jeremiah et al., "Publicly Funded Medical Savings Accounts: Expenditures and Distributional Impacts in Ontario, Canada," *Health Economics* 17 (2008): 1129–1151.
- Hurwitz, Mark A. and Richard E. Caves, "Persuasion or Information? Promotion and the Shares of Brand Name and Generic Pharmaceuticals," *Journal of Law and Economics* 31 (1988): 299–320.
- Iglehart, John K., "An Interview with Leonard D. Schaefer," *Health Affairs* 14 (1995): 131–142.
- Illich, Ivan, *Medical Nemesis* (New York: Bantam Books, 1976).
- Illife, Steve and Sasha Shepperd, "What Do We Know About Hospital at Home: Lessons from International Experience," *Applied Health Economics and Health Policy* 1 (2002): 141–147.
- Institute of Medicine, *Crossing the Quality Chasm: A New Health System for the 21st Century* (Washington, DC: 2001).
- Institute of Medicine, *To Err Is Human: Building a Safer Health System* (Washington, DC, 1999).
- International Health, *How Australia Compares*, www.aihw.gov.au/publications/health/ihhac, 2002.
- Islam, Kamrul et al., "Social Externalities and Mortality in Sweden," *Economics and Human Biology* 6 (2008): 19–42.
- Islam, M. K. et al., "Social Capital and Health: Does Egalitarianism Matter? A Literature Survey," *International Journal for Equity in Health* 5 (2006): 1–28.
- Iversen, Tor (2004), "The Effects of a Patient Shortage on General Practitioners Future Income and Patient Lists," *Journal of Health Economics* 23 (2004): 673–694.
- Iwasaki, Natsuko, Carol H. Tremblay, and Victor J. Tremblay, "Advertising Restrictions and Cigarette Smoking: Evidence from Myopic and Rational Addiction Models," *Contemporary Economic Policy* 24 (2006): 370–381.
- Jacobson, Peter D., Louise E. Parker, and Ian D. Coulter, "Nurse Practitioners and Physician Assistants as Primary Care Providers in Institutional Settings," *Inquiry* 35 (1998/99): 432–446.
- Jaegher, Kris De and Marc Jegers, "A Model of the Physician with Demand Inducement," *Journal of Health Economics* 19 (2000): 231–258.
- Jantzen, Robert and Patricia R. Loubeau, "Hospital Selection by Managed Care Insurers," *Health Care Management Review* 25 (2000): 93–102.
- Jayachandran, Seema et al, Modern Medicine and the Twentieth Century Decline in Mortality: Evidence on the Impact of Sulfa Drugs," *American Economic Journal: Applied Economics* 2 (2010): 118–146.
- Jensen, Gail A. and Allen C. Goodman, "The Economics of Gynecology and Obstetrics," in *Practical Strategies in Obstetrics and Gynecology*, Scott B. Ransom (ed.) (New York: W. B. Saunders, 1999): 719–733.
- Jensen, Gail A. and Michael A. Morrisey, "Employer Sponsored Health Insurance and Mandated Benefit Laws," *Milbank Quarterly* 77 (1999): 425–459.
- Jensen, Gail A. and Michael A. Morrisey, "Endogenous Fringe Benefits, Compensating Wage Differentials and Older Workers," *International Journal of Health Care Finance and Economics* 1 (2001): 203–226.
- Jensen, Gail A. and Michael A. Morrisey, "Group Health Insurance: A Hedonic Price Approach," *Review of Economics and Statistics* 72 (1990): 38–44.
- Jensen, Gail A. and Michael A. Morrisey, "The Role of Physicians in Hospital Production," *Review of Economics and Statistics* 68 (1986): 432–442.
- Jha, Ashish K. et al., "Care in U.S. Hospitals—The Hospital Quality Alliance Program," *New England Journal of Medicine* 353 (2005): 265–274.
- Jiménez-Martin, Serji, José M. Labeaga, and Maite Martínez-Granado, "An Empirical Analysis of the Demand for Physician Services Across the European Union," *European Journal of Health Economics* 5 (2004): 150–165.
- Joesch, Jutta M., Thomas M. Wickizer, and Paul J. Feldstein, "Does Competition by Health Maintenance Organizations Affect the Adoption of Cost-Containment Measures by Fee-for-Service Plans?" *American Journal of Managed Care* 4 (1998): 832–838.
- Johnston, Jack, "Advertising and the Aggregate Demand for Cigarettes: A Comment," *European Economic Review* 14 (1980): 117–125.
- Jondrow, James et al., "On the Estimation of Technical Inefficiency in the Stochastic Frontier Production Function Model," *Journal of Econometrics* (1982): 233–238.

- Jones, Andrew M., "Smoking Cessation and Health: A Response," *Journal of Health Economics* 15 (1996): 755–759.
- Jones, Robert F. et al., "Review of U.S. Medical School Finances, 1996–1997," *Journal of the American Medical Association* 280 (1998): 813–818.
- Jorgenson, Dale W., "Capital Theory and Investment Behavior," *American Economic Review* 53 (1963): 247–259.
- Joskow, Paul L., "The Effects of Competition and Regulation on Hospital Bed Supply and the Reservation Quality of the Hospital," *Bell Journal of Economics* 11 (1980): 421–447.
- Joyce, Geoffrey F. et al., "Employer Drug Benefit Plans and Spending on Prescription Drugs," *Journal of the American Medical Association* 288 (2002): 1733–1739.
- Joyce, Theodore, "Impact of Augmented Prenatal Care on Birth Outcomes of Medicaid Recipients in New York City," *Journal of Health Economics* 18 (1999): 31–67.
- Joyce, Theodore, "What Can Massachusetts Teach Us About National Health Insurance Reform?" *Journal of Policy Analysis and Management*, 30 (2011): 177–195.
- Joyce, Theodore, Andrew D. Racine, and Naci Mocan, "The Consequences and Costs of Maternal Substance Abuse in New York City: A Pooled Time-Series, Cross-Section Analysis," *Journal of Health Economics* 11 (1992): 297–314.
- Jung, Ki-Taig, "Influence of the Introduction of a Per-Visit Copayment on Health Care Use and Expenditures: The Korean Experience," *Journal of Risk and Insurance* 65 (1998): 35–56.
- Kaestner, Robert, Theodore Joyce, and Hassan Wehbeh, "The Effect of Maternal Drug Use on Birth Weight: Measurement Error in Binary Variables," *Economic Inquiry* 34 (October 1996): 617–629.
- Kaiser Family Foundation, 2005, "Ten Myths About The Uninsured," <http://www.kff.org/uninsured/7307.cfm>, accessed May 24, 2011.
- Kaiser Family Foundation, *Trends and Indicators in the Changing Health Care Marketplace*, various years.
- Kaiser Family Foundation, "Summary of New Health Reform Law," April 15, 2011. Available at <http://www.kff.org/healthreform/upload/8061.pdf>, accessed June 9, 2011.
- Kalist, David E., "Registered Nurses and the Value of Bilingualism," *Industrial & Labor Relations Review* 59 (2005): 101–118.
- Kantarevic, Jasmin, Boris Kralj, and Darrel Weinkauff, "Enhanced Fee-for-Service Model and Physician Productivity: Evidence from Family Health Groups in Ontario," *Journal of Health Economics* 30 (2011): 99–111.
- Kaplan, Robert M. and John P. Anderson, "A General Health Policy Model: Update and Applications," *Health Services Research* 23 (1988): 203–235.
- Kapur, Kanika, Carole Roan Gresenz, and David M. Studdert, "Managing Care: Utilization Review in Action at Two Capitated Medical Groups Health Affairs," 10.1377/hlthaff.w3.275, posted June 18, 2003, accessed June 9, 2008.
- Katzman, Brett, Sara Markowitz, and Kerry Anne McGeary, "An Empirical Investigation of the Social Market for Cigarettes," *Health Economics* 16 (2007): 1025–1039.
- Keehan, Sean P., et al., "National Health Spending Projections Through 2020: Economic Recovery and Reform Drive Faster Spending Growth," *Health Affairs* 30 (2011): 1594–1605.
- Keeler, Emmett B. and John E. Rolph, "The Demand for Episodes of Treatment in the Health Insurance Experiment," *Journal of Health Economics* 7 (1988): 333–367.
- Keeler, Emmett B., Willard G. Manning, Jr., and Kenneth B. Wells, "The Demand for Episodes of Mental Health Services," *Journal of Health Economics* 7 (1988): 369–392.
- Keeler, Theodore E. et al., "Taxation, Regulation, and Addiction: A Demand Function for Cigarettes Based on Time-Series Evidence," *Journal of Health Economics* 12 (1993): 1–18.
- Keeler, Theodore E. et al., "The U.S. National Tobacco Settlement: The Effects of Advertising and Price Changes on Cigarette Consumption," *Applied Economics* 36 (2004): 1623–1629.
- Kenkel, Donald S., "The Demand for Preventive Medical Care," *Applied Economics* 26 (1994): 313–325.
- Kermack, William O. and Anderson G. McKendrick, "Contributions to the Mathematical Theory of Epidemics," *Proceedings of the Royal Society A* 115 (1927): 700–721.
- Kessel, Reuben, "Price Discrimination in Medicine," *Journal of Law and Economics* 1 (1958): 20–53.
- Kessides, Ioannis N., "Advertising, Sunk Costs, and Barriers to Entry," *Review of Economics and Statistics* 68 (1986): 84–95.
- Kessler, Daniel P. and Mark B. McClellan, "Is Hospital Competition Socially Wasteful?" *Quarterly Journal of Economics* 115 (2000): 577–615.
- Kim, Daniel, and Ichiro Kawachi, "Food Taxation and Pricing Strategies to "Thin Out" the Obesity Epidemic," *American Journal of Preventive Medicine* 30 (2006): 430–437.
- Kniesner, Thomas J. and John Leeth, "Compensating Wage Differentials for Fatal Injury Risk in Australia, Japan, and the United States," *Journal of Risk and Insurance* 4 (1991): 75–90.
- Kochi, I., B. Hubbell, and R. Kramer, "An Empirical Bayes Approach to Combining and Comparing Estimates of the Value of a Statistical Life for Environmental Policy Analysis," *Environmental and Resource Economics* 34 (2006): 385–406.
- Kolata, Gina, "Co-Payments Soar for Drugs With High Prices," *New York Times*, April 14, 2008, [http://www.nytimes.com/2008/04/14/us/14drug.html?\\_r=1&ref=health&oref=slogin](http://www.nytimes.com/2008/04/14/us/14drug.html?_r=1&ref=health&oref=slogin), accessed April 15, 2008.
- Konetzka, R. Tamara, and Ye Luo, "Explaining Lapse in Long-Term Care Insurance Markets," *Health Economics* 20 (2010): 1169–1183.

- Koolman, Xander and Eddy van Doorslaer, "On the Interpretation of a Concentration Index of Inequality," *Health Economics* 13 (2004): 649–656.
- Kooreman, Peter, "Data Envelopment Analysis and Parametric Frontier Estimation: Complementary Tools," *Journal of Health Economics* 13 (1994): 345–346.
- Kooreman, Peter, "Nursing Home Care in the Netherlands: A Nonparametric Efficiency Analysis," *Journal of Health Economics* 13 (1994): 301–316.
- Kronick, Richard, Louis Olsen, and Todd Gilmer, "The Response of Small Businesses to Variation in the Price of Insurance: Evidence from a Randomized Trial," *Medical Care Research and Review* 65 (2008): 187–206.
- Krueger, Alan B. and Uwe E. Reinhardt, "The Economics of Employer Versus Individual Mandates," *Health Affairs* 13 (1994): 34–53.
- Ku, L., "Medicaid: Improving Health, Saving Lives," Center on Budget and Policy Priorities Analysis of National Health Interview Survey data, [www.kff.org/medicaid/upload/2177-05.pdf](http://www.kff.org/medicaid/upload/2177-05.pdf), accessed August 2005.
- Kwoka, John E., Jr., "Advertising and the Price and Quality of Optometric Services," *American Economic Review* 74 (1984): 211–216.
- Labelle, Roberta, Gregg Stoddard, and Thomas Rice, "A Re-Examination of the Meaning and Importance of Supplier-Induced Demand," *Journal of Health Economics* 13 (1994): 347–368.
- Laine, Juha et al., "Measuring the Productive Efficiency and Clinical Quality of Institutional Long-term Care for the Elderly," *Health Economics* 14 (2005): 245–256.
- Lakdawalla, Darius and Tomas Philipson, "The Nonprofit Sector and Industry Performance," *Journal of Public Economics* 90 (2006): 1681–1698.
- Lalonde, Marc, *A New Perspective on the Health of Canadians*, Office of the Canadian Minister of National Health and Welfare, Ottawa, 1974.
- Lamberton, C. E., W. D. Ellington, and K. R. Spear, "Factors Determining the Demand for Nursing Home Services," *Quarterly Review of Economics and Business* 26 (1986): 74–90.
- Lambin, J. J., *Advertising, Competition and Market Conduct in Oligopoly over Time* (Amsterdam: North-Holland, 1976).
- Landon, Bruce E. et al., "Quality Monitoring and Management in Commercial Health Plans," *American Journal of Managed Care* 14 (2008): 377–386.
- Lanoie, Paul and Paul Leclair, "Taxation or Regulation: Looking for a Good Anti-Smoking Policy," *Economics Letters* 58 (1998): 85–89.
- Laugesen, Murray and Chris Meads, "Tobacco Advertising Restrictions, Price, Income, and Tobacco Consumption in OECD Countries, 1960–1986," *British Journal of Addiction* 86 (1991): 1343–1354.
- Law, Michael R., and Karen A. Grépin, "Is Newer Always Better? Re-Evaluating the Benefits of Newer Pharmaceuticals," *Journal of Health Economics* 29 (2010): 743–750.
- Le Grand, Julian, "Further Tales from the British National Health Service," *Health Affairs* 21 (2002): 116–128.
- Lee, Dwight R., "Why Workers Should Want Mandated Benefits to Lower Their Wages," *Economic Inquiry* 34 (1996): 401–407.
- Lee, Robert H. and Jack Hadley, "Physicians' Fees and Public Medical Care Programs," *Health Services Research* 16 (1981): 185–203.
- Lee, Roger I. and Lewis W. Jones, *The Fundamentals of Good Medical Care*. Publication of the Committee on the Costs of Medical Care. (Chicago: University of Chicago, 1933).
- Lee, Young Hoon and Peter Schmidt, "A Production Frontier Model with Flexible Temporal Variation in Technical Efficiency," in *Measurement of Production Efficiency*, Harold O. Fried, C. A. Knox Lovell et al. (eds.) (New York: Oxford University Press, 1993).
- Leffler, Keith B., "Persuasion or Information? The Economics of Prescription Drug Advertising," *Journal of Law and Economics* 24 (1981): 45–74.
- Leffler, Keith B., "Physician Licensure: Competition and Monopoly in American Medicine," *Journal of Law and Economics* 21 (1978): 165–186.
- Leffler, Keith B. and Cotton M. Lindsay, "Student Discount Rates, Consumption Loans, and Subsidies to Professional Training," *Journal of Human Resources* 16 (1981): 468–476.
- Lehner, Laura A. and James F. Burgess, "Teaching and Hospital Production: The Use of Regression Estimates," *Health Economics* 4 (1995): 113–125.
- Leibowitz, Arlene A., "The Demand for Health and Health Concerns after 30 Years," *Journal of Health Economics* 23 (2004): 663–671.
- Leibowitz, Arlene A., "An Economic Perspective on Work, Family and Well-Being," in S.M. Bianchi, L.M. Casper, and R.B. King, (eds.) *Work, Family, Health, and Well-Being* (Mahwah, New Jersey: Lawrence Erlbaum, 2005).
- Leonard, Kenneth L., "Is Patient Satisfaction Sensitive to Changes in the Quality of Care? An Exploitation of the Hawthorne Effect," *Journal of Health Economics* 27 (2008): 444–459.
- Leonard, Kenneth L. and Joshua G. Zivin, "Outcome Versus Service Based Payments in Health Care: Lessons from African Traditional Healers," *Health Economics* 14 (2005): 575–593.
- Lerner, Abba P., "The Concept of Monopoly and the Measurement of Monopoly Power," *Review of Economic Studies* 11 (1943): 157–175.
- Lerner, Arthur N., "Federal Trade Commission Anti-Trust Activities in the Health Care Services Field," *Antitrust Bulletin* 29 (1984): 205–224.

- Lessler, Daniel S. and Wickizer, Thomas M. "The Impact of Utilization Management on Hospital Readmissions for Patients with Cardiovascular Disease," *Health Services Research* 34 (2000): 1315–1330.
- Levit, Katherine et al., "Health Spending Rebound Continues in 2002," *Health Affairs* 23 (2004): 147–150.
- Levitt, Steven D. and Jack Ray Porter, "Estimating the Effect of Alcohol on Driver Risk Using Only Fatal Accident Statistics," *NBER Working Paper* 6944, 1999.
- Levy, Helen and David Weir. "Take-up of Medicare Part D: Results From the Health and Retirement Study," *Journals of Gerontology B Psychology Science Social Science* 65B (2010): 492–501.
- Lawrence, Timothy J. Eckels, and Linda B. Miller, "Setting the Record Straight: The Provision of Uncompensated Care by Not-for Profit Hospitals," *New England Journal of Medicine* 318 (1988): 1212–1215.
- Lewit, Eugene M. and Douglas Coate, "The Potential for Using Excise Taxes to Reduce Smoking," *Journal of Health Economics* 1 (1982): 121–145.
- Lewit, Eugene M., Douglas Coate, and Michael Grossman, "The Effects of Government Regulation on Teenage Smoking," *Journal of Law and Economics* 24 (1981): 545–569.
- Li, Tong and Robert Rosenman, "Estimating Hospital Costs With a Generalized Leontief Function," *Health Economics* 10 (2001): 523–538.
- Liang, J. Nellie and Jonathon D. Ogur, *Restrictions on Dental Auxiliaries: An Economic Policy Analysis* (Washington, DC: Federal Trade Commission, 1987).
- Lichtenberg, Frank R., "Pharmaceutical Innovation, Mortality Reduction, and Economic Growth," *NBER Working Paper* 6569, 2006.
- Lichtenberg, Frank R., "Importation and Innovation," *Economics of Innovation and New Technology* 16 (2007a): 403–417.
- Lichtenberg, Frank R., "Benefits and Costs of New Drugs: An Update," *Managerial and Decision Economics* 28 (2007b): 485–490.
- Lichtenberg, Frank R., "Have Newer Cardiovascular Drugs Reduced Hospitalization? Evidence from Longitudinal Country-Level Data on 20 OECD Countries, 1995–2003," *Health Economics* 18 (2009): 519–534.
- Lien, Hsien-Ming, et al., "Hospital Ownership and Performance: Evidence from Stroke and Cardiac Treatment in Taiwan," *Journal of Health Economics* 27 (2008): 1208–1223.
- Light, Donald W. and Rebecca N. Warburton, "Extraordinary Claims Require Extraordinary Evidence," *Journal of Health Economics* 24 (2005): 1030–1033.
- Light, Donald L., and Rebecca Warburton, "Demythologizing the High Costs of Pharmaceutical Research," *BioSocieties* 6 (2011): 34–50.
- Lindeboom, Maarten, Ana Llana-Nozai, and Bas van der Klaauw, "Parental Education and Child Health: Evidence from a Schooling Reform," Tinbergen Institute, Discussion Paper No. 2516 (2006).
- Lindhjem, H., S. Navrud, and N.A. Braathen, "Valuing Lives Saved from Environmental Transport and Health Policies: A Meta-Analysis of Stated Preference Studies, Prepared for the Organization for Economic Co-operation and Development," ENV/EPOC/WPNEP(2008)10/FINAL, (2010).
- Lindrooth, Richard C., Anthony T. Lo Sasso, and Gloria J. Bazzoli, "The Effect of Urban Hospital Closure on Markets," *Journal of Health Economics* 22 (2003): 691–712.
- Linnet, Martha S. et al., "Residential Magnetic Field Exposures and Childhood Acute Lymphoblastic Leukemia," *New England Journal of Medicine* 337 (July 3, 1997): 1–7.
- Linn, Margaret W., Lee Gurel, and Bernard S. Linn, "Patient Outcome as a Measure of Quality of Nursing Home Care," *American Journal of Public Health* 4 (1977): 337–344.
- Linna, Miika and Unto Häkkinen, "Reimbursing for Costs of Teaching and Research in Finnish Hospitals: A Stochastic Frontier Approach," *International Journal of Health Care Finance and Economics* 6 (2006): 83–97.
- Liu, Liqun, Andrew J. Rettenmaier, and Thomas R. Saving, "Longevity Bias in Cost-effectiveness Analysis," *Health Economics* 17 (2008): 523–534.
- Liu, Zhimei, William H. Dow, Edward C. Norton, "Effect of Drive-Through Delivery Laws on Postpartum Length of Stay and Hospital Charges," *Journal of Health Economics* 23 (2004): 129–155.
- Lleras-Muney, Adriana, "The Relationship between Education and Adult Mortality in the United States," *Review of Economic Studies* 72 (2005): 189–221.
- Lo Sasso, Anthony T., et al., "The \$16,819 Pay Gap for Newly Trained Physicians: The Unexplained Trend of Men Earning More than Women," *Health Affairs* 30 (2011): 193–200.
- Locke, John, *Epistola de Tolerantia*, *John Locke Bibliography*, [www.libraries.psu.edu/iasweb/locke/ch0b.htm](http://www.libraries.psu.edu/iasweb/locke/ch0b.htm), July 1, 2002.
- Locke, John, *On the Poor Law and Working Schools* (1697) [eco.utexas.edu/~hmcleave/350kPEELockePoorEdTable.pdf](http://eco.utexas.edu/~hmcleave/350kPEELockePoorEdTable.pdf), accessed May 14, 2006.
- Locke, John, *Two Treatises of Government* (1690; Cambridge: Cambridge University Press, 1960).
- Loftus, Peter, "Drug Makers Diversify as the U.S. Pharmaceutical Market Slows," *Wall Street Journal*, July 30, 2008, p. B4.
- LoGerfo, James P. et al., "Rates of Surgical Care in Prepaid Group Practices and the Independent Setting: What Are the Reasons for the Differences?" *Medical Care* 17 (1979): 1–10.
- Lohr, Steve, "Most Doctors Aren't Using Electronic Records," *New York Times*, June 19, 2008, [www.nytimes.com/2008/06/19/technology/19patient.html](http://www.nytimes.com/2008/06/19/technology/19patient.html).

- Long, Stephen H., Russell F. Settle, and Charles W. Wrightson Jr., "Employee Premiums, Availability of Alternative Plans, and HMO Disenrollment," *Medical Care* 26 (1988): 927–938.
- Lovell, C. A. Knox, and Shelton S. Schmidt (eds.), *The Measurement of Productive Efficiency*. (Oxford: Oxford University Press, 1993).
- Lu, John Z. and William S. Comanor, "Strategic Pricing of New Pharmaceuticals," *Review of Economics and Statistics* 80 (1998): 108–117.
- Luft, Harold S., "Health Maintenance Organizations and the Rationing of Medical Care," *Milbank Memorial Fund Quarterly/Health and Society* 60 (1982): 268–306.
- Luft, Harold S., *Health Maintenance Organizations: Dimensions of Performance*. (New York: John Wiley & Sons, 1981).
- Luft, Harold S., "How Do Health-Maintenance Organizations Achieve Their 'Savings'?" *New England Journal of Medicine* 298 (1978): 1336–1343.
- Lundborg, Petter and Bjorn Lindren, "Do They Know What They Are Doing? Risk Perception and Smoking Behavior Among Swedish Teenagers," *Journal of Risk and Uncertainty* 28 (2004): 261–286.
- Ma, Jin, Mingshan Lu, Hude Quan, "From A National, Centrally Planned Health System To A System Based On The Market: Lessons from China," *Health Affairs* 27 (2008): 937–948.
- MacDonald, Ziggy, "What Price Drug Use? The Contribution of Economics to an Evidence-Based Drugs Policy," *Journal of Economic Surveys* 18 (2004): 113–152.
- Madrian, Brigitte C., "Employment-Based Health Insurance and Job Mobility: Is There Evidence of Job-Lock?" *Quarterly Journal of Economics* 109 (1994): 27–54.
- Magnusson, Jon, "Efficiency Measurement and the Operationalization of Hospital Production," *Health Services Research* 31 (1996): 21–37.
- Mandon, Bruce E. et al., "Quality Monitoring and Management in Commercial Health Plans," *American Journal of Managed Care* 14 (2008): 377–386.
- Manning, Willard G. and M. Susan Marquis, "Health Insurance: The Tradeoff Between Risk Pooling and Moral Hazard," *Journal of Health Economics* 15 (1996): 609–639.
- Manning, Willard G. et al., "A Controlled Trial of the Effect of a Prepaid Group Practice on Use of Services," *New England Journal of Medicine* 310 (1984): 1505–1510.
- Manning, Willard G. et al., *The Costs of Poor Health Habits*. (Cambridge, MA: Harvard University Press, 1991).
- Manning, Willard G. et al., "Health Insurance and the Demand for Medical Care: Evidence from a Randomized Experiment," *American Economic Review* 77 (1987): 251–277.
- Mansfield, Edwin, "Patent Innovation: An Empirical Study," *Management Science* 32 (1986): 173–181.
- Mansfield, Edwin et al., *Managerial Economics: Theory, Applications, and Cases* (New York: Norton, 2002).
- Mark, Tami L., "Analysis of the Rationale for, and Consequences of, Nonprofit and For-Profit Ownership Conversions," *Health Services Research* 34 (1999): 83–101.
- Mark, Tami L., "Psychiatric Hospital Ownership and Performance: Do Nonprofit Organizations Offer Advantages in Markets Characterized by Asymmetric Information?" *Journal of Human Resources* 31 (1996): 631–649.
- Martinez, Barbara and Jacob Goldstein, "Big Pharma Faces Grim Prognosis," *Wall Street Journal*, December 6, 2007, p. 1.
- Marquis, M. Susan, Jeannette A. Rogowski, and Jose Escarcé, "The Managed Care Backlash: Did Consumers Vote with Their Feet?" *Inquiry* 41 (2004/2005): 376–390.
- Mason, Michael, "Bargaining Down That CT Scan Is Suddenly Possible," *New York Times*, February 27, 2007, [www.nytimes.com/2007/02/27/health/27cons.html?\\_r=1&oref=slogin](http://www.nytimes.com/2007/02/27/health/27cons.html?_r=1&oref=slogin), accessed July 24, 2008.
- Mather, Dell, B., "Incorporating Clinical Outcomes and Economic Consequences into Drug Formulary Decisions: A Practical Approach," *American Journal of Managed Care* 5 (1999): 277–285.
- Mayberry, Robert M and Mili, Fatima and Ofili, Elizabeth, "Racial and Ethnic Differences in Access to Medical Care," *Medical Care Research and Review* 57 Supplement (2000): 108–145.
- Maynard, Alan, "The United Kingdom," in *Advances in Health Economics and Health Services Research, Supplement 1: Comparative Health Systems*, Jean-Jacques Rosa (ed.) (Greenwich, CT: JAI Press, 1990), 1–26.
- Mays, Glen P., and Sharla A. Smith, "Evidence Links Increases in Public Health Spending to Declines in Preventable Deaths," *Health Affairs* 30 (2011): 1585–1593.
- McCain, John, "Straight Talk on Health System Reform," [www.johnmccain.com/Informing/Issues](http://www.johnmccain.com/Informing/Issues), accessed August 22, 2008.
- McCarthy, Patrick S., "The Effect of Higher Rural Interstate Speed Limits in Alcohol-Related Accidents," *Journal of Health Economics* 12 (1993): 281–299.
- McCarthy, Thomas R., "The Competitive Nature of the Primary-Care Physician Services Market," *Journal of Health Economics* 4 (1985): 93–117.
- McClellan, Mark and Douglas Staiger, "Comparing Hospital Quality at For-Profit and Not-For-Profit Hospitals," *NBER Working Paper* 7324, 1999.
- McClellan, Mark, "Hospital Reimbursement Incentives: An Empirical Analysis," *Journal of Economics and Management Strategy* 6 (1997): 91–128.
- McCullough, Jeffrey S., "The Adoption of Hospital Information Systems," *Health Economics* 17 (2008): 649–664.
- McGlynn, Elizabeth A. et al., "The Quality of Health Care Delivered to Adults in the United States," *New England Journal of Medicine* 348 (2003): 2635–2645.

- McGuinness, Tony and Keith Cowling, "Advertising and the Aggregate Demand for Cigarettes," *European Economic Review* 6 (1975): 311–328.
- McGuire, Thomas G., "Physician Agency," in *Handbook of Health Economics*, Volume 1A, Anthony J. Culyer and Joseph P. Newhouse (eds.) (Amsterdam, Elsevier Science B.V., 2000).
- McGuire, Thomas G. and Mark V. Pauly, "Physician Response to Fee Changes with Multiple Payers," *Journal of Health Economics* 10 (1991): 385–410.
- McKay, Niccye L., "The Prisoner's Dilemma: An Obstacle to Cooperation in Health Care Markets," *Medical Care Review* 51 (1994): 179–204.
- McKeown, Thomas, *The Modern Rise of Population* (New York: Academic Press, 1976).
- McKinlay, John B. and Sonja M. McKinlay, "The Questionable Contribution of Medical Measures to the Decline of Mortality in the United States in the Twentieth Century," *Milbank Memorial Fund Quarterly/Health and Society* 55 (1977): 405–428.
- McLaughlin, Catherine G., "Market Responses to HMOs: Price Competition or Rivalry?" *Inquiry* 25 (1988): 207–218.
- McLaughlin, Catherine G. and Wendy K. Zellers, "Small Business and Health Care Reform," The University of Michigan School of Public Health, 1994.
- McPherson, Klim et al., "Small-Area Variations in the Use of Common Surgical Procedures: An International Comparison of New England, England, and Norway," *New England Journal of Medicine* 307 (1982): 1310–1313.
- McQueen, M. P., "Paying Workers to Go Abroad for Health Care," *Wall Street Journal*, September 30, 2008, p. B1.
- Medical Study News*, "First Evaluation of California's Much-Watched Ratios Law," August 9, 2005.
- Medicare Payment Advisory Commission (MedPAC), *Report to Congress, Medicare Payment Policy*, March 1999: ch. 7, app. D.
- Medicare Payment Advisory Commission (MedPAC), *Aligning Incentives in Medicare*, Report to Congress, Ch. 4, June 2010.
- Medicare Payment Advisory Commission (MedPAC), *Healthcare Spending and the Medicare Program*, June 2010, <http://www.medpac.gov/documents/jun10datbookentirereport.pdf>.
- Mello, Michelle M., Amitabh Chandra, Atul A. Gawande and David M. Studdert. "National Costs of the Medical Liability System," *Health Affairs* 29 (2010): 1–9.
- Mellor Jennifer M. "Do Cigarette Taxes Affect Children's Body Mass Index? The Effect of Household Environment on Health" *Health Economics* 20 (2011): 417–31.
- Melnick, Glen A. et al., "The Effects of Market Structure and Bargaining Position on Hospital Prices," *Journal of Health Economics* 11 (1992): 217–233.
- Melnick, Glen and Emmett Keeler, "The Effects of Multi-Hospital Systems on Hospital Prices," *Journal of Health Economics* 26 (2007): 400–413.
- Meyerhoeffer, Chad D., and Samuel H. Zuvekas, "New Estimates of the Demand for Physical and Mental Health Treatment," *Health Economics* 19 (2010): 297–315.
- Miles, J. J., *Health Care and Antitrust Law*. (St. Paul, MN: West Group, 1998).
- Miller, Douglas L. et al., "Social Capital and Health in Indonesia," *World Development* 34 (2006): 1084–1098.
- Miller, Rebecca S., Harry S. Jonas, and Michael E. Whitcomb, "The Initial Employment Status of Physicians Completing Training in 1994," *Journal of the American Medical Association* 275 (1996): 708–712.
- Miller, Richard D. Jr. and H. E. Frech III, *Health Care Matters: Pharmaceuticals, Obesity, and the Quality of Life* (Washington, D.C.: American Enterprise Institute, 2004).
- Miller, Robert H. and Harold S. Luft, "Does Managed Care Lead to Better or Worse Quality of Care?" *Health Affairs* 16 (1997): 7–25.
- Miller, Robert H. and Harold S. Luft, "HMO Plan Performance Update: An Analysis of the Literature: 1997–2001," *Health Affairs* 21 (2002): 63–86.
- Miller, Robert H. and Harold S. Luft, "Managed Care Plan Performance Since 1980: A Literature Analysis," *Journal of the American Medical Association* 271 (May 18, 1994): 1512–1519.
- Miller, Theodore R., "Variation Between Countries in Values of Statistical Life," *Journal of Transport Economics and Safety* 34 (2000): 169–88.
- Ministry of Health. 2004. *Study on National Health Services in 2003*. Ministry of Health, Beijing, China (in Chinese).
- Miron, Jeffrey A., "The Effect of Alcohol Prohibition on Alcohol Consumption," *NBER Working Paper* 7130, 1999.
- Mitchell, Janet B. and Jerry Cromwell, "Impact of All-or-Nothing Assignment Requirement Under Medicare," *Health Care Financing Review* 4 (1983): 59–78.
- Mitchell, Janet B. and Jerry Cromwell, "Physician Behavior Under the Medicare Assignment Option," *Journal of Health Economics* 1 (1982): 245–264.
- Monroe, Kent B., "Buyers Subjective Perceptions of Price," *Journal of Marketing Research* 10 (1973): 70–80.
- Montgomery-Karp, Beverly, "Market Demands Call for New Treatment," *Advertising Age* 54 (September 26, 1983): M9–11.
- Mooney, Gavin et al., "Reweighing Heat: Response to Culyer, van Doorslaer, and Wagstaff," *Journal of Health Economics* 11 (1992): 199–205.
- Mooney, Gavin et al., "Utilisation as a Measure of Equity: Weighing Heat?" *Journal of Health Economics* 10 (1991): 475–480.



- Moore, Michael J., "Death and Tobacco Taxes," *NBER Working Paper* 5153, 1995.
- Morgan, Eleanor J., "Innovation and Merger Decisions in the Pharmaceutical Industry," *Review of Industrial Organization* 19 (2001): 181–197.
- Morrisey, Michael A., "Why Do Employers Do What They Do? Compensating Differentials" *International Journal of Health Care Finance and Economics* 1 (2001a): 195–201.
- Morrisey, Michael A., "Competition in Hospital and Health Insurance Markets: A Review and Research Agenda," *Health Services Research* 36 (2001b): 191–222.
- Morrisey, Michael A., "Movies and Myths: Hospital Cost Shifting," *Business Economics* 30 (1995): 22–25.
- Morrisey, Michael A., *Health Insurance*, Chapter 8 (Chicago: Health Administration Press, 2008).
- Morrisey, Michael and John Cawley, "U.S. Health Economists: Who We Are and What We Do," *Health Economics* 17 (2008): 535–543.
- Motheral, Brenda R., and Rochelle Henderson, "The Effect of a Copay Increase on Pharmaceutical Utilization, Expenditures, and Treatment Continuation," *American Journal of Managed Care* 5 (1999): 1383–1394.
- Mrozek, J.R., and L.O. Taylor, "What Determines the Value of Life? A Meta-Analysis," *Journal of Policy Analysis and Management* 21 (2002): 253–70.
- Mueller, Curt D. and Alan C. Monheit, "Insurance Coverage and the Demand for Dental Care," *Journal of Health Economics* 7 (1988): 59–72.
- Mukamel, Dana B. and William D. Spector, "The Competitive Nature of the Nursing Home Industry: Price Mark Ups and Demand Elasticities," *Applied Economics* 34 (2002): 413–420.
- Mullahy, John and Jody L. Sindelar, "Alcoholism, Work, and Income," *Journal of Labor Economics* 11 (1993): 494–520.
- Mullahy, John and Jody Sindelar, "Do Drinkers Know When to Say When? An Empirical Analysis of Drunk Driving," *Economic Inquiry* 22 (1994): 383–394.
- Mullen, Kathleen J., Richard G. Frank, and Meredith B. Rosenthal, "Can You Get What You Pay for? Pay-for-Performance and the Quality of Healthcare Provider," *RAND Journal of Economics* 41 (2010): 64–91.
- Murphy, Kevin M. and Robert H. Topel, "The Value of Health and Longevity," *NBER Working Paper* 11405, June 2005.
- Mushkin, Selma J., *Biomedical Research: Costs and Benefits*. (Cambridge, MA: Ballinger, 1979).
- Nassiri, Abdelhak and Lise Riochaix, "Revisiting Physician Financial Incentives in Quebec: A Panel System," *Health Economics* 15 (2006): 49–64.
- National Association of Home Care, "Basic Statistics About Hospice," [www.nahc.org/consumer/hpcstats](http://www.nahc.org/consumer/hpcstats), 1999.
- National Partnership for Women and Families, January 31, 2011; [http://www.nationalpartnership.org/site/News2?page=NewsArticle&id=27551&news\\_iv\\_ctrl=0&abbr=daily2\\_](http://www.nationalpartnership.org/site/News2?page=NewsArticle&id=27551&news_iv_ctrl=0&abbr=daily2_);
- Naylor, C. David, "Health Care in Canada: Incrementalism Under Fiscal Duress," *Health Affairs* 18 (1999): 9–26.
- Nelson, Phillip, "Advertising as Information," *Journal of Political Economy* 82 (1974): 729–754.
- Nelson, Phillip, "Information and Consumer Behavior," *Journal of Political Economy* 78 (1970): 311–329.
- New England Journal of Medicine Editorial Staff, "PPACA Can Help Contain Growth in Healthcare Costs," Published on-line May 12, 2010, [http://www.cardiovascular\\_business.com/index.php?option=com\\_articles&article=22178](http://www.cardiovascular_business.com/index.php?option=com_articles&article=22178), accessed June 24, 2011.
- Newbold, Paul, Carlson, William L., and Thorne, Betty *Statistics for Business and Economics*, 6th ed. (Upper Saddle River, NJ: Prentice Hall, 2007).
- Newhouse, Joseph P., *The Economics of Medical Care* (Reading, MA: Addison-Wesley, 1978a).
- Newhouse, Joseph P., "Frontier Estimation: How Useful a Tool in Health Economics?" *Journal of Health Economics* 13 (1994): 317–322.
- Newhouse, Joseph P., "Has the Erosion of the Medical Marketplace Ended?" *Journal of Health Politics, Policy and Law* 13 (1988): 263–277.
- Newhouse, Joseph P., "Medical Care Costs: How Much Welfare Loss?" *Journal of Economic Perspectives* 6 (1992): 3–21.
- Newhouse, Joseph P., "Medical-Care Expenditures: A Cross-National Survey," *Journal of Human Resources* 12 (1977): 115–125.
- Newhouse, Joseph P., "Medicare Policy in the 1990s," *NBER Working Paper* 8531, 2001.
- Newhouse, Joseph P., "The Structure of Health Insurance and the Erosion of Competition in the Medical Marketplace," in *Competition in the Health Care Sector: Past, Present and Future*, W. Greenberg (ed.) (Germantown, MD: Aspen Systems, 1978b).
- Newhouse, Joseph P., "Toward a Theory of Nonprofit Institutions: An Economic Model of a Hospital," *American Economic Review* 60 (1970): 64–74.
- Newhouse, Joseph P., "Why Is There a Quality Chasm?" *Health Affairs* 21 (2002): 13–25.
- Newhouse, Joseph P. and Charles E. Phelps, "New Estimates of Price and Income Elasticities of Medical Care Services," in *The Role of Health Insurance in the Health Services Sector*, Richard Rosett (ed.) (New York: Neal Watson, 1976).
- Newhouse, Joseph P., Geoffrey Anderson, and Leslie L. Roos, "Are Fee-for-Service Costs Increasing Faster Than HMO Costs?" *Medical Care* 23 (1985): 960–966.
- Newhouse, Joseph P., Geoffrey Anderson, and Leslie L. Roos, "Hospital Spending in the United States and Canada: A Comparison," *Health Affairs* 7 (1988): 6–16.
- Newhouse, Joseph P. et al., *Free for All? Lessons from the RAND Health Insurance Experiment*. (Cambridge MA: Harvard University Press, 1993).

- Nguyen, Nguyen X. and Frederick W. Derrick, "Physician Behavioral Response to a Medicare Price Reduction," *Health Services Research* 32 (1997): 283–298.
- Nicholson, Sean, "How Much Do Medical Students Know about Physician Income?" *Journal of Human Resources* 40 (2005): 100–114.
- Nord, Erik, "Health State Values from Multi-Attribute Utility Instruments Need Correction," *Annals of Medicine* 33 (2001): 371–374.
- Nordhaus, William, "Critical Assumptions in the Stern Review of Climate Change," *Science* 317 (2007): 201–202.
- Norton, Edward C., "Long-Term Care," in *Handbook of Health Economics*, Volume 1B, Anthony J. Culyer and Joseph P. Newhouse (eds.) (Amsterdam, Elsevier Science B.V., 2000).
- Norton, Edward C. and Douglas O. Staiger, "How Hospital Ownership Affects Access to Care for the Uninsured," *RAND Journal of Economics* 25 (1994): 171–185.
- Nozick, Robert, *Anarchy, State, and Utopia*. (New York: Basic Books, 1974).
- Nyman, John A., "The Economics of Moral Hazard Revisited," *Journal of Health Economics* 18 (1999): 811–824.
- Nyman, John A., "Testing for Excess Demand in Nursing Home Markets," *Medical Care* 31 (1993): 680–693.
- Nyman, John A. and Roland Maude-Griffin, "The Welfare Economics of Moral Hazard," *International Journal of Health Care Finance and Economics* 1 (2001): 23–42.
- Nyman, John A. et al., "Changing Physician Behavior: Does Medical Review of Part B Medicare Claims Make a Difference?" *Inquiry* 27 (1990): 127–137.
- Oakeshott, Isabel, "NHS Bars Woman after She Saw Private Doctor," London: *The Times*, April 18, 2010, [http://www.timesonline.co.uk/tol/life\\_and\\_style/health/article7100968.ece](http://www.timesonline.co.uk/tol/life_and_style/health/article7100968.ece), accessed June 27, 2011.
- O'Halloran, Patrick L. and David J. Bashaw, "U.S. Physician Board Certification and Labor Market Returns," *International Journal of Manpower* 27 (2006): 624–646.
- Organization for Economic Cooperation and Development (OECD), "Health Care Systems: Getting More Value for Money," *OECD Economics Department Policy Notes*, No. 2, 2010.
- Organization for Economic Cooperation and Development (OECD), *OECD Health Data*, various years various tables, 75775 Paris Cedex 16 France.
- Ohsfeldt, Robert L., Raymond G. Boyle, and Eli I. Capilouto, "Tobacco Taxes, Smoking Restrictions, and Tobacco Use," *NBER Working Paper* 6486, 1998.
- Okun, Arthur, *Efficiency and Equity: The Big Tradeoff* (Washington, DC: Brookings, 1975).
- Okunade, Ablert A. and Chutima Suraratdecha, "Cost Efficiency, Factor Interchange, and Technical Progress in U.S. Specialized Hospital Pharmacies," *Health Economics* 7 (1998): 363–371.
- Olesen, Ole B. and Niels C. Petersen, "The Use of Data Envelopment Analysis with Probabilistic Assurance Regions for Measuring Hospital Efficiency," *Journal of Productivity Analysis* 17 (2002): 83–109.
- Olivella, Pau, and Marcos Vera-Hernández, "Competition among Differentiated Health Plans Under Adverse Selection," *Journal of Health Economics* 26 (2007): 233–250.
- Olkin, Ingram and Jerome Sacks, "Cross-Disciplinary Research in the Statistical Sciences," Hayward, CA: Institute of Mathematical Statistics, September 1988.
- Olsen, Jan Abel, "But Health Can Still Be a Necessity. . . .," *Journal of Health Economics* 12 (1993): 187–191.
- Olsen, Jan Abel, "Theories of Justice and Their Implications for Priority Setting in Health Care," *Journal of Health Economics* 16 (1997): 625–639.
- Olsen, Jan Abel and Richard D. Smith, "Theory Versus Practice: A Review of 'Willingness-to-Pay' in Health and Health Care," *Health Economics* 10 (2001): 39–52.
- Olson, Mary K. "The Risk We Bear: The Effects of Review Speed and Industry User Fees on New Drug Safety," *Journal of Health Economics* 27 (2008): 175–200.
- O'Neill, June E. and Dave M. O'Neill, "Health Status, Health Care and Inequality: Canada vs. the U.S.," *Forum for Health Economics & Policy* 10 (2008) (Frontiers in Health Policy Research): Article 3. [www.bepress.com/fhep/10/1/3](http://www.bepress.com/fhep/10/1/3).
- Oregon Health Services Commission, State of Oregon, *Prioritization of Health Services: A Report to the Governor and Legislature*, May 1991.
- Oreopoulos, Philip, "Estimating Average and Local Average Treatment Effects of Education When Compulsory Schooling Laws Really Matter," *American Economic Review* 96 (2006): 125–175.
- Oster, Emily, "Sexually Transmitted Infections, Sexual Behavior, and the HIV/AIDS Epidemic," *Quarterly Journal of Economics* 120 (2005): 467–515.
- Over, Mead, "Impact of the HIV/AIDS Epidemic on the Health Sectors of Developing Countries," in Haacker, Markus, Ed., *The Macroeconomics of HIV/AIDS*, Washington, International Monetary Fund, November 2004.
- Owen, Gary M. "Seizing the Opportunity," *American Health & Drug Benefits* (April 2008): 52–55.
- Parente, Stephen and Roger Feldman, "Consumer Response to a National Marketplace for Individual Insurance," presented at the American Enterprise Institute, Washington, DC, July 31, 2008.
- Paris, Valérie, Marion Devaux, and Lihan Wei, "Health Systems Institutional Characteristics: A Survey of 29 OECD Countries," *OECD Health Working Paper No. 50*, OECD Publishing, 2010.
- Parker-Pope, Tara, "Antismoking Sentiment Flares in Europe's Smoke-Filled Cafes," *Wall Street Journal*, August 28, 1995, sec. B1.

- Parkin, David, Alistair McGuire, and Brian Yule, "Aggregate Health Care Expenditures and National Income: Is Health Care a Luxury Good?" *Journal of Health Economics* 6 (1987): 109–128.
- Paul, Chris, "Physician Licensure and the Quality of Medical Care," *Atlantic Economic Journal* 12 (1984): 18–30.
- Pauly, Mark V., *Doctors and Their Workshops* (Chicago: University of Chicago Press, 1980).
- Pauly, Mark V., "The Economics of Moral Hazard: Comment," *American Economic Review* 58 (1968): 531–537.
- Pauly, Mark V., "Editorial: A Re-Examination of the Meaning and Importance of Supplier-Induced Demand," *Journal of Health Economics* 13 (1994a): 369–372.
- Pauly, Mark V., *Health Benefits at Work: An Economic and Political Analysis of Employment-Based Health Insurance* (Ann Arbor, MI: The University of Michigan Press, 1997).
- Pauly, Mark V., "Is Medical Care Different?" in *Competition in the Health Care Sector*, Warren Greenberg (ed.) (Germantown, MD: Aspen Systems, 1978).
- Pauly, Mark V., "Is Medical Care Different? Old Questions, New Answers," *Journal of Health Politics, Policy and Law* 13 (1988b): 227–237.
- Pauly, Mark V., "Making a Case for Employer Enforced Individual Mandates," *Health Affairs* 13 (1994b): 21–33.
- Pauly, Mark V., *Medical Care at Public Expense* (New York: Praeger, 1971).
- Pauly, Mark V., "A Primer on Competition in Medical Markets," in *Health Care in America*, H. E. Frech III (ed.) (San Francisco: Pacific Research Institute for Public Policy, 1988a).
- Pauly, Mark V. and John C. Goodman, "Tax Credits for Health Insurance and Medical Savings Accounts," *Health Affairs* 14 (1995), 125–139.
- Pauly, Mark V. and Michael Redisch, "The Not-for-Profit Hospital as a Physician's Cooperative," *American Economic Review* 63 (1973): 87–100.
- Pauly, Mark V. and Uwe E. Reinhardt, "The Rise and Fall of Health Care Reform," in Institute of Medicine (1996), *Looking Back, Looking Forward, Staying Power Issues in Health Care Reform*, Washington DC: National Academies Press.
- Pauly, Mark V. and Mark A. Satterthwaite, "The Pricing of Primary Care Physicians' Services: A Test of the Role of Consumer Information," *Bell Journal of Economics* 12 (1981): 488–506.
- Payne, Abigail A., "Does the Government Crowd Out Private Donations? New Evidence from a Sample of Non-Profit Firms," *Journal of Public Economics* 69 (1998): 323–345.
- Pearson, Steven D. et al., "The Impact of Pay-for-Performance on Health Care Quality in Massachusetts, 2001–2003," *Health Affairs* 27 (2008): 1167–1176.
- Peden, Edgar A. and Mark S. Freeland, "Insurance Effects on U.S. Medical Spending (1960–1993)," *Health Economics* 7 (1998): 671–687.
- Peltzman, Sam, "Regulation of Pharmaceutical Innovation: The 1962 Amendments," American Enterprise Institute, Washington, DC, 1974.
- Peltzman, Sam, "Toward a More General Theory of Regulation," *Journal of Law and Economics* 19 (1976): 211–248.
- Pernick, Martin S., "The Cultural Ecology of Health: Science, Values, and Society in the History of American Medicine," manuscript of book in progress, 1991.
- Peterson, Mark A. (ed.), "Special Issue: Kenneth Arrow and the Changing Economics of Health Care," *Journal of Health Politics, Policy and Law* 26 (October 2001).
- Pharmaceutical Research and Manufacturers of America (PhRMA), *Pharmaceutical Industry Profile 2011*, Washington DC, 2011 (available at www.phrma.org).
- Pharmaceutical Research and Manufacturers of America (PhRMA), *New Drug Approvals in 2009*, Washington DC, 2010 (available at www.phrma.org).
- Pharmaceutical Research and Manufacturers of America (PhRMA), *The Facts About Pharmaceutical Marketing and Promotion*, Washington DC, 2008 (available at www.phrma.org).
- Phelps, Charles E., "Good Technologies Gone Bad: How and Why the Cost-Effectiveness of a Medical Intervention Changes for Different Populations," *Medical Decision Making* 17 (1997): 107–117.
- Phelps, Charles E., "Induced Demand—Can We Ever Know Its Extent?" *Journal of Health Economics* 5 (1986): 355–365.
- Phelps, Charles E., "Information Diffusion and Best Practice," in *Handbook of Health Economics*, Volume I, Anthony J. Culyer and Joseph P. Newhouse (eds.) (Amsterdam, Elsevier Science B.V., 2000).
- Phelps, Charles E., "Welfare Loss from Variations: Further Considerations," *Journal of Health Economics* 14 (1995): 253–260.
- Phelps, Charles E. and Cathleen Mooney, "Correction and Update on 'Priority Setting in Medical Technology Assessment,'" *Medical Care* 30 (1992): 744–751.
- Phelps, Charles E. and Stephen T. Parente, "Priority Setting in Medical Technology and Medical Practice Assessment," *Medical Care* 28 (1990): 703–723.
- Philipson, Tomas, "Economic Epidemiology and Infectious Diseases," in *Handbook for Health Economics*, Volume I, Anthony J. Culyer and Joseph P. Newhouse (eds.) (Amsterdam, Elsevier Science B.V., 2000).
- Philipson, Tomas, "Private Vaccination and Public Health: An Empirical Examination for U.S. Measles," *Journal of Human Resources* 31 (1996): 611–630.
- Philipson, Tomas, "The Welfare Loss of Disease and the Theory of Taxation," *Journal of Health Economics* 14 (1995): 387–395.

- Philipson, Tomas et al., "Cost-Benefit Analysis of the FDA: The Case of the Prescription Drug User Fee Acts," *Journal of Public Economics* 92 (2008): 1306–1325.
- Philipson, Tomas J., Essy Mozaffari, and John R. Maclean, "Pharmacy Cost Sharing, Antiplatelet Therapy Utilization, and Health Outcomes for Patients with Acute Coronary Syndrome," *American Journal of Managed Care* 16 (2010): 290–297.
- Philipson, Tomas J. and Richard Posner, *Private Choices and Public Health: The AIDS Epidemic in Economic Perspective* (Cambridge: Harvard University Press, 1993).
- Photius.com, "GDP – purchasing power parity 2008 Country Ranks," [www.photius.com/rankings/economy/gdp\\_purchasing\\_power\\_parity\\_2008\\_0.html](http://www.photius.com/rankings/economy/gdp_purchasing_power_parity_2008_0.html).
- Picone, Gabriel A., Martin Uribe, and R. Mark Wilson, "The Effect of Uncertainty on the Demand for Medical Care, Health Capital and Wealth," *Journal of Health Economics* 17 (1998): 171–185.
- Picone, Gabriel, Shin-Yi Chou, and Frank Sloan, "Are For-Profit Hospital Conversions Harmful to Patients and to Medicare?" *RAND Journal of Economics* 33 (2002): 507–523.
- Pindyck, Robert S. and Daniel L. Rubinfeld, *Microeconomics*, 4th ed. (Upper Saddle River, NJ: Prentice Hall, 1998).
- Pitt, Mark M., "Estimating the Determinants of Child Health When Fertility and Mortality Are Selective," *Journal of Human Resources* 32 (1997): 129–158.
- Plotzke, Michael Robert and Courtemanche, Charles, "Does Procedure Profitability Impact Whether An Outpatient Surgery Is Performed At An Ambulatory Surgery Center Or Hospital?" *Health Economics* 20 (2011): 817–830.
- Pollard, Michael R. and Robert F. Leibenluft, *Antitrust and the Health Professions*, Office of Policy and Planning, Federal Trade Commission, 1981.
- Polsky, Daniel et al., "Malpractice Premiums and the Supply of Obstetricians" *Inquiry* 47 (2010): 48–61.
- Polsky, Daniel and Sean Nicholson, "Why Are Managed Care Plans Less Expensive: Risk Selection, Utilization, or Reimbursement?" *Journal of Risk and Insurance* 71 (2004): 21–40.
- Porell, Frank et al., "A Longitudinal Analysis of Nursing Home Outcomes," *Health Services Research* 33 (1998): 835–865.
- Porter, Eduardo, "Do New Drugs Always Have to Cost So Much?" *New York Times*, November 14, 2004; Section 3, p. 5.
- Powell, Lisa M., and Frank T. Chaloupka, "Food Prices and Obesity: Evidence and Implications for Taxes and Subsidies," *The Milbank Quarterly* 87 (2009): 229–257.
- Pozen, Alexis, and David M. Cutler, "Medical Spending Differences in the United States and Canada: The Role of Prices, Procedures, and Administrative Expenses," *Inquiry* 47 (2010): 124–34.
- President of the United States, *The President's Comprehensive Health Reform Program*, February 6, 1992.
- Preya, Colin and George Pink, "Scale and Scope Efficiencies Through Hospital Consolidations," *Journal of Health Economics* 25 (2006): 1049–1068.
- Pritchett, Lant and Lawrence H. Summers, "Wealthier Is Healthier," *Journal of Human Resources* 31 (1996): 841–868.
- Prospective Payment Assessment Commission, *Annual Report to Congress*, April 11, 1997.
- Prospective Payment Review Commission, *Medicare Prospective Payment and the American Health Care System*, Report to Congress, June 24, 1991 (Chicago: Commerce Clearing House, Inc., 1991).
- Putnam, Robert D., *Bowling Alone* (New York: Simon & Schuster, 2000).
- Quast, Troy, David Sappington, and Elizabeth Shenkman, "Does the Quality of Care in Medicaid MCOs Vary with the Form of Physician Compensation?" *Health Economics* 17 (2008): 545–550.
- Quick, Jonathan D., M. R. Greenlick, and K. L. Roghmann, "Prenatal Care and Pregnancy Outcome in an HMO and General Population: A Multivariate Cohort Analysis," *American Journal of Public Health* 71 (1981): 381–390.
- Raffel, Marshall W., Norma K. Raffel, and Camille K. Barsukiewicz, *The U.S. Health System: Origins and Functions*, 5th ed. (Albany, New York: Delmar, 2002).
- Rashad, Inas and Eric Sarpong "Employer-provided Health Insurance and the Incidence of Job Lock: a Literature Review and Empirical Test," *Expert Review of Pharmacoeconomics & Outcomes Research* 8 (2008): 583–91.
- Rawls, John, *A Theory of Justice*. (Cambridge, MA: Belnap Press of the Harvard University Press, 1971).
- Register, Charles A. and Edward R. Bruning, "Profit Incentives and Technical Efficiency in the Production of Hospital Care," *Southern Economic Journal* 53 (1987): 899–914.
- Reinhardt, Uwe E., "The Pricing of U.S. Hospital Services: Chaos Behind a Veil of Secrecy," *Health Affairs* 25 (2006): 57–69.
- Reinhardt, Uwe E., "A Production Function for Physician Services," *Review of Economics and Statistics* 54 (1972): 55–65.
- Relman, Arnold S. and Marcia Angell, "America's Other Drug Problem," *The New Republic* (December 16, 2002): 27–41.
- Renn, Steven C. et al., "The Effects of Ownership and System Affiliation on the Economic Performance of Hospitals," *Inquiry* 22 (1985): 219–236.
- Research Data Assistance Center (ResDAC), "Calculating Hospital Specific DRG Adjusted Payments," University of Minnesota, [www.resdac.umn.edu/Tools/TBs/TN-004-2.pdf](http://www.resdac.umn.edu/Tools/TBs/TN-004-2.pdf), accessed July 31, 2008.
- Reyes, Jessica Wolpaw, "Do Female Physicians Capture Their Scarcity Value: The Case of OB/GYNs," *NBER Working Paper* 12528, 2006.

- Rhee, Sang-O et al., "USMGs Versus FMGs: Are There Performance Differences in the Ambulatory Setting?" *Medical Care* 24 (1986): 248–258.
- Rice, Dorothy, *Estimating the Cost of Illness*, U.S. Public Health Service Publication No. 947-6. (Washington, D.C.: U.S. Government Printing Office, May 1966).
- Rice, Thomas, *The Economics of Health Reconsidered* (Chicago: Health Administration Press, 1998).
- Rice, Thomas H. and Roberta J. Labelle, "Do Physicians Induce Demand for Medical Service?" *Journal of Health Politics, Policy and Law* 14 (1989): 587–600.
- Ried, Walter, "QALYs Versus HYE—What's Right and What's Wrong: A Review of the Controversy," *Journal of Health Economics* 17 (1998): 607–625.
- Rizzo, John A., "Are HMOs Bad for Health Maintenance?" *Health Economics* 13 (2005): 1117–1131.
- Rizzo, John A., "Physician Advertising Revisited," *Medical Care* 26 (1988): 1238–1244.
- Rizzo, John A. and David Blumenthal, "Is the Target Income Hypothesis an Economic Heresy?" *Medical Care Research and Review* 53 (1996): 243–266.
- Rizzo, John A. and Richard J. Zeckhauser, "Advertising and Entry: The Case of Physician Services," *Journal of Political Economy* 98 (1990): 476–500.
- Rizzo, John A. and Richard J. Zeckhauser, "Advertising and the Price, Quantity, and Quality of Primary Care Physician Services," *Journal of Human Resources* 27 (1992): 381–421.
- Rizzo, John A. and Richard J. Zeckhauser, "Reference Incomes, Loss Aversion and Physician Behavior," *Review of Economics and Statistics* 85 (2003): 909–922.
- Roberts, Brandon and Irving Hoch, "Malpractice Litigation and Medical Costs in the United States," *Health Economics* 18 (2009): 1394–1419.
- Roberts, Mark J. and Larry Samuelson, "An Empirical Analysis of Dynamic, Nonprice Competition in an Oligopolistic Industry," *RAND Journal of Economics* 19 (1988): 200–220.
- Robinson, James C., "The Commercial Health Insurance Industry in an Era of Eroding Employer Coverage," *Health Affairs* 25 (2006): 1475–1486.
- Rochaix, Lise, "Information Asymmetry and Search in the Market for Physicians' Services," *Journal of Health Economics* 8 (1989): 53–84.
- Roos, N. P. and L. L. Roos, Jr., "Surgical Rate Variations: Do They Reflect the Health or Socioeconomic Characteristics of the Population?" *Medical Care* 20 (1982): 945–958.
- Rosen, Harvey S., Chairman, Council of Economic Advisers, "Comments on Health Care at the Conference of the National Association of Business Economics," March 21, 2005, <http://georgewbush-whitehouse.archives.gov/cea/nabe-healthcare032105.pdf>.
- Rosen, Sherwin, "Theory of Equalizing Differences," in *Handbook of Labor Economics*, Volume 1, Orley Ashenfelter and Richard Layard (eds.) (New York: North-Holland, 1986, 641–692).
- Rosenthal, Meredith B., "Nonpayment for Performance? Medicare's New Reimbursement Rule," *New England Journal of Medicine* 357 (2007): 1573–1575.
- Rosenthal, Meredith B. et al., "Early Experience with Pay-for-Performance: From Concept to Practice," *Journal of the American Medical Association* 294 (2005): 1788–1793.
- Rosenthal, Meredith B. et al., "Pay for Performance in Commercial HMOs," *New England Journal of Medicine* 355 (2006): 1895–1902.
- Rosenzweig, Mark R. and T. Paul Schultz, "Estimating a Household Production Function: Heterogeneity, The Demand for Health Inputs, and Their Effects on Birth Weight," *Journal of Political Economy* 91 (1983): 723–746.
- Rosenzweig, Mark R. and Kenneth I. Wolpin, "Sisters, Siblings, and Mothers: The Effect of Teen-Age Childbearing on Birth Outcomes in a Dynamic Family Context," *Econometrica* 63 (1995): 303–326.
- Rosett, Richard N. and Lien-fu Huang, "The Effect of Health Insurance on the Demand for Medical Care," *Journal of Political Economy* 81 (1973): 281–305.
- Rosko, Michael D., "The Performance of U.S. Teaching Hospitals: A Panel Analysis of Cost Efficiency," *Health Care Management Science* 7 (2004): 7–16.
- Rotfeld, Herbert J. and Kim B. Rotzoll, "Advertising and Product Quality: Are Heavily Advertised Products Better?" *Journal of Consumer Affairs* 10 (1976): 37–47.
- Ruffin, Roy J. and Paul R. Gregory, *Principles of Economics*, 7th ed. (New York: Addison Wesley, 2001).
- Russell, Louise B., "The Diffusion of Hospital Technologies: Some Econometric Evidence," *Journal of Human Resources* 12 (1977): 482–502.
- Sack, Kevin, "For Governors, Medicaid Looks Ripe for Slashing," [http://www.nytimes.com/2011/01/29/us/politics/29medicaid.html?\\_r=1](http://www.nytimes.com/2011/01/29/us/politics/29medicaid.html?_r=1), both accessed June 17, 2011.
- Saffer, Henry, "Alcohol Advertising and Motor Vehicle Fatalities," *Review of Economics and Statistics* 79 (1997): 431–442.
- Saffer, Henry, "Alcohol Advertising Bans and Alcohol Abuse: An International Perspective," *Journal of Health Economics* 10 (1991): 65–79.
- Saffer, Henry, "Alcohol Consumption and Alcohol Advertising Bans," *NBER Working Paper* 7758, 2000.
- Saffer, Henry and Dhaval Dave, "Alcohol Advertising of Alcohol Consumption by Adolescents," *NBER Working Paper* 9676, 2003.
- Saffer, Henry and Frank Chaloupka, "Demographic Differentials in the Demand for Alcohol and Illicit Drugs," *NBER Working Paper* 6432, 1998.
- Saffer, Henry and Dhaval Dave, "Mental Illness and the Demand for Alcohol, Cocaine, and Cigarettes," *Economic Inquiry* 43 (2005): 229–246.

- Safriet, Barbara J., "Impediments to Progress in Health Care Workforce Policy: License and Practice Laws," *Inquiry* 31 (1994): 310–317.
- Salkever, David S., "Regulation of Prices and Investment in Hospitals in the United States," in *Handbook of Health Economics*, Volume 1B, Anthony J. Culyer and Joseph P. Newhouse (eds.) (Amsterdam, Elsevier Science B.V., 2000).
- Salvatore, Dominick, *Schaum's Outline Series: Microeconomic Theory*, 2nd ed. (New York: McGraw-Hill, 1983).
- Santerre, Rexford, "The Welfare Loss from Hospital Cost-Shifting Behavior: A Partial Equilibrium Analysis," *Health Economics* 14 (2005): 621–626.
- Santerre, Rexford E. and John A. Vernon, "The Consumer Welfare Implications of the Hospital Ownership Mix in the U.S.: An Exploratory Study," *Health Economics* 15 (2006): 1187–1199.
- Sanz de Galdeano, Anna, "Job-Lock and Public Policy: Evidence from Clinton's Second Mandate," *Industrial and Labor Relations Review* 59 (2006): 430–437.
- Sasser, Alicia C., "Gender Differences in Physician Pay: Tradeoffs Between Career and Family," *Journal of Human Resources* 60 (2005): 477–504.
- Satterthwaite, Mark A., "Consumer Information, Equilibrium Industry Price, and the Number of Sellers," *Bell Journal of Economics* 10 (1979): 483–502.
- Scanlon, Dennis P. et al., "Competition and Health Plan Performance: Evidence from Health Maintenance Organization Insurance Market," *Medical Care* 43 (2005): 338–346.
- Scanlon, Dennis P. et al., "The Impact of Health Plan Report Cards on Managed Care Enrollment," *Journal of Health Economics* 21 (2002): 19–41.
- Scanlon, William J., "A Theory of the Nursing Home Market," *Inquiry* 17 (1980): 25–41.
- Schaefer, Mark E., "Demand Versus Need for Medical Services in a General Cost Benefit Setting," *American Journal of Public Health* 65 (1975): 293–295.
- Schelling, Thomas, "Ergonomics: The Art of Self Management," *American Economic Review* 68 (1978): 290–294.
- Scherer, F. Michael., *Innovation and Growth: Schumpeterian Perspectives*. (Cambridge, MA: MIT, 1984).
- Schmalensee, Richard, *The Economics of Advertising* (Amsterdam: North-Holland, 1972).
- Schmalensee, Richard, "A Model of Advertising and Product Quality," *Journal of Political Economy* 86 (1978): 485–503.
- Schmidt, Peter and Robin Sickles, "Production Frontiers and Panel Data," *Journal of Business and Economic Statistics* 2 (1984): 367–374.
- Schnelle, John F. et al., "Relationship of Nursing Home Staffing to Quality of Care," *Health Services Research* 39 (2004): 225–250.
- Schneyer, Solomon, J. Steven Landefeld, and Frank H. Sandifer, "Biomedical Research and Illness: 1900–1979," *Milbank Memorial Fund Quarterly Health and Society* 59 (1981): 44–58.
- Schoen, Cathy, et al. "How Health Insurance Design Affects Access To Care And Costs By Income In Eleven Countries", *Health Affairs* 29 (2010): 2323–2334.
- Schwartz, Joel and Douglas W. Dockery, "Increased Mortality in Philadelphia Associated with Daily Air Pollution Concentrations," *American Review of Respiratory Diseases* 145 (1992): 600–604.
- Schwartz, Michael et al., "Small Area Variations in Hospitalization Rates: How Much You See Depends on How You Look," *Medical Care* 32 (1994): 189–201.
- Schwartz, William B., "In the Pipeline: A Wave of Valuable Medical Technology," *Health Affairs* 13 (1994): 70–79.
- Schwartz, William B. and Daniel N. Mendelson, "Eliminating Waste and Inefficiency Can Do Little to Contain Costs," *Health Affairs* 13 (1994): 224–238.
- Schweitzer, Stuart O., *Pharmaceutical Economics and Policy* (New York: Oxford University Press, 1996).
- Segal, Leonie and Jeff Richardson, "Economic Framework for Allocative Efficiency in the Health Sector," *Australian Economic Review* 106 (1994): 89–98.
- Seldon, Barry J. and Khowsrow Doroodian, "A Simultaneous Model of Cigarette Advertising: Effects on Demand and Industry Response to Public Policy," *Review of Economics and Statistics* 71 (1989): 673–677.
- Seldon, Barry J., Chulho Jung, and Roberto J. Cavazos, "Market Power Among Physicians in the U.S., 1983–1991," *Quarterly Review of Economics and Finance* 38 (1998): 799–824.
- Sen, Amartya, "Equality of What?" in McMurrin S. (ed.), *The Tanner Lectures on Human Values* (Cambridge: Cambridge University Press, 1980).
- Sen, Anindya and Tony S. Wirjanto "Estimating the Impacts of Taxes on the Initiation and Persistence of Youth Smoking: Empirical Evidence from Ontario, Canada," *Health Economics*, 19 (2010): 1264–1280.
- Sen, Bisakha and Christine M. Campbell, 2010. "Alcohol Prevalence, Alcohol Policies, And Child Fatal Injury Rates From Motor Vehicle Crashes," *Contemporary Economic Policy* 28 (2010): 392–405.
- Sendi, Pedram, Maiwenn J. Al, and Heinz Zimmerman, "A Risk-adjusted Approach to Comparing the Return on Investment in Health Care Programs," *International Journal of Health Care Finance and Economics* 4 (2004): 199–210.
- Shaw, James W., William C. Horrace, and Ronald J. Vogel, "The Determinants of Life Expectancy: An Analysis of the OECD Health Data," *Southern Economic Journal* 71 (2005): 768–783.
- Shen, Yu-chu, "The Effect of Financial Pressure on the Quality of Care in Hospitals," *Journal of Health Economics* 22 (2003): 243–269.

- Sherman, H. David, "Hospital Efficiency Measurement and Evaluation," *Medical Care* 22 (1984): 922–935.
- Sheu, Mei-ling et al., "The Effect of a Major Cigarette Price Change on Smoking Behavior in California: A Zero-Inflated Negative Binomial Model," *Health Economics* 13 (2004): 781–791.
- Shleifer, Andrei, "A Theory of Yardstick Competition," *RAND Journal of Economics* 16 (1985): 319–327.
- Shortell, Stephen M., *A Model of Physician Referral Behavior: A Test of Exchange Theory in Medical Practice* (Chicago: Center for Health Administration Studies, 1973).
- Shortell, Stephen M., Robin R. Gillies, and David A. Anderson, "The New World of Managed Care: Creating Organized Delivery Systems," *Health Affairs* 13 (1994): 46–64.
- Sickles, Robin C. and Abdo Yazbeck, "On the Dynamics of Demand for Leisure and the Production of Health," *Journal of Business and Economic Statistics* 16 (1998): 187–197.
- Silverman, Elaine and Jonathan Skinner, "Are For Profit Hospitals Really Different? Medicare Upcoding and Market Structure," *NBER Working Paper* 8133, 2001.
- Silverman, Elaine and Jonathan Skinner, "Medicare Upcoding and Hospital Ownership," *Journal of Health Economics* 23 (2004): 369–389.
- Simon, Carol J., David Dranove, and William White, "The Effect of Managed Care on the Incomes of Primary and Specialty Physicians," *Health Services Research* 33 (1998): 549–569.
- Simon, Kosali I., "Nation's Big Uninsurance Problem Revolves Around Small Firms," Economic Research Initiative on the Uninsured, [eriu.sph.umich.edu/forthemedia/interviews\\_simon.html](http://eriu.sph.umich.edu/forthemedia/interviews_simon.html), accessed September 1, 2008.
- Sindelar, Jody, "The Declining Price of Health Insurance," in *Health Care in America*, H. E. Frech III (ed.). (San Francisco: Pacific Research Institute for Public Policy, 1988).
- Sinjay, Ugur A. and Claudia R. Campbell, "Scope and Scale Economies in Merging Hospitals Prior to Merger," *Journal of Economics and Finance* 19 (1995): 107–123.
- Skinner, Jonathon, "What Do Stochastic Cost Functions Tell Us About Inefficiency?" *Journal of Health Economics* 13 (1994): 323–328.
- Skocpol, Theda, "The Rise and Resounding Demise of the Clinton Plan," *Health Affairs* 14 (1995): 66–87.
- Sloan, Frank A., et al., "Does Where You Are Admitted Make a Difference? An Analysis of Medicare Data," *NBER Working Paper* 6896, 1999.
- Sloan, Frank A. et al., "Hospital Ownership and Cost and Quality of Care: Is There a Dime's Worth of Difference?" *Journal of Health Economics* 20 (2001): 1–22.
- Sloan, Frank A., "Property Rights in the Hospital Industry," in *Health Care in America*, H. E. Frech III (ed.). (San Francisco: Pacific Research Institute for Public Policy, 1988).
- Sloan, Frank A., "Rate Regulation as a Strategy for Hospital Cost Control: Evidence from the Last Decade," *Milbank Memorial Fund Quarterly/Health and Society* 61 (1983): 195–221.
- Sloan, Frank A. and Edmund R. Becker, "Internal Organization of Hospitals and Hospital Costs," *Inquiry* 28 (1981): 224–239.
- Sloan, Frank A., Roger D. Feldman, and A. Bruce Steinwald, "Effects of Teaching on Hospital Costs," *Journal of Health Economics* 3 (1983): 1–28.
- Sloan, Frank A., Michael A. Morrissey, and Joseph Valvona, "Effects of the Medicare Prospective Payment System on Hospital Cost Containment: An Early Appraisal," *Milbank Quarterly* 66 (1988a): 191–219.
- Sloan, Frank A., Michael A. Morrissey, and Joseph Valvona, "Medicare Prospective Payment and the Use of Medical Technologies in Hospitals," *Medical Care* 26 (1988b): 837–853.
- Sloan, Frank A. et al., "The Demise of Hospital Philanthropy," *Economic Inquiry* 28 (1990): 725–743.
- Smedley, Brian D., Adrienne Y. Stith, and Alan R. Nelson, (eds.), *Unequal Treatment: Confronting Racial and Ethnic Disparities in Health Care*. Institute of Medicine, National Academy Press, Washington, DC, 2002.
- Smith, Adam, *An Inquiry into the Nature and Causes of the Wealth of Nations* (Chicago: University of Chicago Press, 1976).
- Smith, Francis B., *The Retreat of Tuberculosis 1850–1950*. (London: Croom Helm, 1988).
- Smith, James P., "Socioeconomic Status and Health," *American Economic Review* 88 (1998): 192–196.
- Smith, Michael J., Susan S. Ellenberg, Louis M. Bell and David M. Rubin, "Media Coverage of the Measles-Mumps-Rubella Vaccine and Autism Controversy and Its Relationship to MMR Immunization Rates in the United States," *Pediatrics* 121 (2008): e836–e843.
- Sonnenfeld, Sally T. et al., "Projections of National Health Expenditures Through the Year 2000," *Health Care Financing Review* 13 (1991): 1–27.
- Spence, A. Michael, "Monopoly, Quality, and Regulation," *Bell Journal of Economics* 6 (1973): 417–429.
- STAKES, [www.stakes.fi/english/facts/index.html](http://www.stakes.fi/english/facts/index.html), 2002.
- Stano, Miron, "An Analysis of the Evidence on Competition in the Physician Services Markets," *Journal of Health Economics* 4 (1985): 197–211.
- Stano, Miron, "A Further Analysis of the Physician Inducement Controversy," *Journal of Health Economics* 6 (1987): 227–238.
- Stano, Miron, "Individual Health Accounts: An Alternative Health Care Financing Approach," *Health Care Financing Review* 3 (1981): 117–125.

- Stano, Miron and Sherman Folland, "Variations in the Use of Physician Services by Medicare Beneficiaries," *Health Care Financing Review* 9 (1988): 51–58.
- Stano, Miron et al., "The Effects of Physician Availability on Fees and the Demand for Doctors' Services," *Atlantic Economic Journal* 13 (1985): 51–60.
- Starr, Paul, "Health Care for the Poor: The Past Twenty Years," in *Fighting Poverty: What Works and What Doesn't*, Sheldon Danziger and Daniel Weinberg (eds.). (Cambridge: Harvard, 1986).
- Starr, Paul, *The Social Transformation of American Medicine*. (New York: Basic Books, 1982).
- Steinberg, Jonny, *Sizwe's Test: A Young Man's Journey Through Africa's AIDS Epidemic* (New York: Simon & Schuster, 2008).
- Steinberg, Richard, "Charitable Giving as a Mixed Public/Private Good: Implications for Tax Policy," *Public Finance Quarterly* 14 (1986): 415–431.
- Steinberg, Richard, "Voluntary Donations and Public Expenditures in a Federalist System," *American Economic Review* 77 (1987): 24–36.
- Steinfelds, Peter, "Oregon Medicaid's Doctor-Assisted Suicide," *New York Times*, March 7, 1998, National/Metro Section.
- Steinmann, Lukas and Peter Zweifel, "On the [In]efficiency of Swiss Hospitals-English Summary," *Ifo Studien* 46 (2000): 197–217.
- Stern, Nicholas, *The Economics of Climate Change: The Stern Review* (Cambridge: Cambridge University Press, 2007).
- Stewart, Kathryn, "Seven Ways to Help Your Hospital Stay in Business," [www.aafp.org/fpm/20030500/27seve.html](http://www.aafp.org/fpm/20030500/27seve.html), accessed July 23, 2008.
- Stigler, George J., "The Economics of Information," *Journal of Political Economy* 69 (1961): 213–225.
- Stigler, George J., "The Theory of Economic Regulation," *Bell Journal of Economics* 2 (1971): 3–21.
- Stiglitz, Joseph E., *Economics of the Public Sector*, 2nd ed. (New York: Norton, 1988).
- Street, Andrew, "How Much Confidence Can We Have in Hospital Efficiency Estimates?" *Health Economics* 12 (2003): 896–907.
- Succi, Melissa J., Shoou-Yih Lee, and Jeffrey A. Alexander, "Effects of Market Position and Competition on Rural Hospital Closures," *Health Services Research* 31 (1997): 679–699.
- Sullivan, Daniel, "Monopsony Power in the Market for Nurses," *Journal of Law and Economics* 32 (1989): S135–S179.
- Surgeon General of the United States, *Reducing the Health Consequences of Smoking: 25 Years of Progress, Executive Summary*. U.S. Department of Health and Human Services, Public Health Service, Office on Smoking and Health, Rockville, MD, 1989.
- Svorny, Shirley, "Should We Reconsider Licensing Physicians?" *Contemporary Policy Issues* 10 (1992): 31–38.
- Svorny, Shirley, "Medical Licensing: An Obstacle to Affordable, Quality Care," *Policy Analysis No. 621*, Cato Institute, September 17, 2008.
- Tauras, John A. and Frank J. Chaloupka, "Price, Clean Indoor Air, and Cigarette Smoking: Evidence from the Longitudinal Data for Young Adults," *NBER Working Paper* 6937, 1999.
- Tay, Abigail, "Assessing Competition in Hospital Care Markets: The Importance of Accounting for Quality Difference," *RAND Journal of Economics* 34 (2003): 786–814.
- Telser, Lester G., "Advertising and Competition," *Journal of Political Economy* 72 (1964): 537–562.
- Temin, Peter, "An Economic History of American Hospitals," in *Health Care in America*, H. E. Frech III (ed.). (San Francisco: Pacific Research Institute for Public Policy, 1988).
- Tengs, Tammy O. et al., "Five Hundred Life-Saving Interventions and Their Cost-Effectiveness," *Risk Analysis* 15 (1995): 369–390.
- Teplensky, Jill D. et al., "Hospital Adoption of Medical Technology: An Empirical Test of Alternative Models," *Health Services Research* 30 (1995): 437–466.
- Thaler, Richard and Sherwin Rosen, "The Value of Saving a Life: Evidence from the Labor Market," in *Household Production and Consumption*, Nestor E. Terleckyi (ed.) (New York: National Bureau of Economic Research, 1975): 265–297.
- Thomas, Lewis, *The Lives of a Cell* (New York: Viking, 1974).
- Thornton, James, "Estimating a Health Production Function for the U.S.," *Applied Econometrics* 34 (2002): 59–62.
- Tilford, John M. et al., "Response to Health Insurance by Previously Uninsured Rural Children," *Health Services Research* 34 (1999): 761–775.
- Torrance, George W., "Preferences for Health Outcomes and Cost-Utility Analysis," *American Journal of Managed Care* 3 (1997 Supplement): S8–S20.
- Toxic Substances Board, *Health or Tobacco: An End to Tobacco Advertising and Promotion*, Department of Health, Wellington, New Zealand, May 1989.
- Tremblay, Carol Horton and Victor J. Tremblay, "The Impact of Cigarette Advertising on Consumer Surplus, Profit, and Social Welfare," *Contemporary Economic Policy* 13 (1995): 113–124.
- Trigg, Andrew B. and Nick Bosanquet, "Tax Harmonisation and the Reduction of European Smoking Rates," *Journal of Health Economics* 11 (1992): 329–346.
- Troyer, Jennifer L., "Examining Differences in Death Rates for Medicaid and Non-Medicaid Nursing Home Residents," *Medical Care* 42 (2004): 985–991.
- Truffer, Christopher J. et al, *2010 Actuarial Report on the Financial Outlook for Medicaid*, Washington DC,



- December 2010, Centers for Medicare & Medicaid Services United States Department of Health & Human Services, accessed June 7, 2011.
- Trude, Sally, Melanie Au, and Jon B. Christianson, "Health Plan Pay-for-Performance Strategies," *American Journal of Managed Care* 12 (2006): 537–542.
- Tsuchiya, Aki, "QALYs and Ageism: Philosophical Theories and Age Weighting," *Health Economics* 9 (2000): 57–68.
- Tumlinson, Anne et al., "Choosing a Health Plan: What Information Will Consumers Use?" *Health Affairs* 16 (1997): 229–238.
- Tye, Joe B., Kenneth E. Warner, and Stanton A. Glantz, "Tobacco Advertising and Consumption: Evidence of a Causal Relationship," *Journal of Public Health Policy* 8 (1987): 492–508.
- United Kingdom Commission for Healthcare Audit and Inspection, *State of Healthcare 2005*, www.healthcarecommission.org.uk/NationalFindings/StateOfHealthcare/fs/en?CONTENT\_ID=4006361&chk=sNVkqk, accessed November 15, 2005.
- United Nations, "World Population from Year 0 to Stabilization," www.gopher.undp.org:70/00/ungophers/popin/wdtrends/history, 1996.
- United Nations Department of Economic and Social Affairs Population Division, *Population and HIV/AIDS 2007* (United Nations Publication, Sales No. E.08.XIII.9), March 2008.
- U.S. Department of Commerce. *Statistical Abstract*. Washington, DC, various years.
- U.S. Department of Health and Human Services, National Institute of Allergies and Infectious Diseases, "The Relationship Between the Human Immunodeficiency Virus and the Acquired Immunodeficiency Syndrome," www.niaid.nih.gov/publications/hivaids/14.htm, accessed January 16, 2003.
- U.S. Department of Health and Human Services, Public Health Service. *Health United States*, various years.
- U.S. Department of Health and Human Services, *Healthy People 2010: Understanding and Improving Health* (2<sup>nd</sup> ed), US Government Printing Office, 2000.
- U.S. Department of Health and Human Services, Public Health Service. *Seventh Report to the President and Congress on the Status of Health Personnel in the United States*. DHHS Publication No. HRS-P-OD-90-1, March 1990.
- U.S. Department of Health and Human Services, *Report of the Expert Panel on Population Strategies for Blood Cholesterol Reduction*. NIH Publication No. 90-3046, November 1990.
- U.S. Department of Health and Human Services, *Understanding Estimates of the Uninsured: Putting the Differences in Context*, <http://aspe.os.dhhs.gov/health/reports/hiestimates.htm>, accessed March 17, 2003.
- U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, 2008, "HIV/AIDS Among Men Who Have Sex with Men," <http://www.cdc.gov/hiv/topics/msm/resources/factsheets/msm.htm>, accessed April 20, 2009.
- U.S. Department of Health and Human Services, Centers for Medicare and Medicaid Services, *Improving Medicare's Hospital Inpatient Prospective Payment System*, www.cms.hhs.gov/apps/media/press/factsheet.asp?Counter=2336, accessed July 24, 2008.
- U.S. Department of Health and Human Services, A Databook, *Healthcare and the Medicare Program*, MedPAC, June 2008, various charts and tables, www.medpac.gov/documents/Jun08DataBook\_Entire\_report.pdf, accessed October 12, 2008.
- U.S. Department of Justice, "Health Care Business Review Letters Issued," www.usdoj.gov/atr/public/health\_care/9950.htm, accessed July 23, 2008.
- U.S. Department of Labor, Bureau of Labor Statistics, *Occupational Outlook Handbook*, 1990–1991 edition, April 1990.
- Valdez, R. Burciaga et al., "Consequences of Cost Sharing for Children's Health," *Pediatrics* 75 (1985): 952–961.
- Valdmanis, Vivian G., "Ownership and Technical Efficiency of Hospitals," *Medical Care* 28 (1990): 552–561.
- van Doorslaer, Eddy, Xander Koolman, and Andrew M. Jones, "Explaining Income-related Inequalities in Doctor Utilization in Europe," *Health Economics* 13 (2004): 629–647.
- Van Houtven, Courtney and Edward C. Norton, "Informal Care and Health Use of Older Adults," *Journal of Health Economics* 23 (2004): 1159–1180.
- Van Houtven, Courtney H., and Edward C. Norton, "Informal Care and Medical Expenditures: Testing for Heterogeneous Treatment Effects," *Journal of Health Economics* 27 (2008): 134–156.
- van Lear, William and Lynette Fowler, "Efficiency and Service in the Group Home Industry," *Journal of Economic Issues* 21 (1997): 1039–1050.
- Varkevisser, Marco, Stéphanie A. van der Geest, and Frederik T. Schut, "Assessing Hospital Competition When Prices Don't Matter to Patients: The Use of Time-Elasticities," *International Journal of Health Care Finance and Economics* 10 (2010): 43–60.
- Vernon, John A., "Examining the Link Between Price Regulation and Pharmaceutical R&D Investment," *Health Economics* 14 (2005): 1–16.
- Viscusi, W. Kip, "Age Variations in Risk Perceptions and Smoking Decisions," *Review of Economics and Statistics* 73 (1991): 577–588.
- Viscusi, W. Kip, "Cigarette Smokers and the Social Consequence of Smoking," *Tax Policy and the Economy* 9 (1995): 51–101.

- Viscusi, W. Kip, "Promoting Smokers' Welfare with Responsible Taxation," *National Tax Journal* 57 (1994): 547–558.
- Viscusi, W. Kip, "Risk Equity," *Journal of Legal Studies* 29 (2000): 843–871.
- Viscusi, W. Kip, "The Value of Risks to Life and Health," *Journal of Economic Literature* 31 (1993): 1912–1946.
- Viscusi, W. Kip, and Joseph E. Aldy, "The Value of a Statistical Life: A Critical Review of Market Estimates Throughout the World," *Journal of Risk and Uncertainty* 27 (2003): 5–76.
- Viscusi, W. Kip, John M. Vernon, and Joseph E. Harrington, Jr., *Economics of Regulation and Antitrust*, 2nd ed. (Cambridge, MA: MIT Press, 1995).
- Vita, Michael G., "Exploring Hospital Production Relationships with Flexible Functional Forms," *Journal of Health Economics* 9 (1990): 1–21.
- Vitaliano, Donald F., "On the Estimation of Hospital Cost Functions," *Journal of Health Economics* 6 (1987): 305–318.
- Vitaliano, Donald F. and Mark Toren, "Cost Efficiency in Nursing Homes: A Stochastic Frontier Approach," *Journal of Health Economics* 13 (1994): 281–300.
- Vitaliano, Donald F. and Mark Toren, "Hospital Cost and Efficiency in a Regime of Stringent Regulation," *Eastern Economic Journal* 22 (1996): 161–175.
- Waber, Rebecca L. et al., "Commercial Features of Placebo and Therapeutic Efficacy," *Journal of the American Medical Association* 299 (2008): 1016–1017.
- Wade, Nicholas, "Couch Mouse to Mr. Mighty by Pills Alone," *New York Times*, August 1, 2008, [www.nytimes.com/2008/08/01/science/01muscle.html?\\_r=1&ref=health&oref=slogin](http://www.nytimes.com/2008/08/01/science/01muscle.html?_r=1&ref=health&oref=slogin), accessed November 26, 2008.
- Wagstaff, Adam, "Estimating Efficiency in the Hospital Sector: A Comparison of Three Statistical Cost Frontier Models," *Applied Economics* 21 (1989): 659–672.
- Wagstaff, Adam, "QALYs and the Equity-Efficiency Trade-Off," *Journal of Health Economics* 10 (1991): 21–41.
- Wagstaff, Adam and Guillem Lopez, "Hospital Costs in Catalonia: A Stochastic Frontier Analysis," *Applied Economics* 3 (1996): 471–474.
- Wakefield, Andrew J., et al., "RETRACTED: Ileal-lymphoid-nodular Hyperplasia, Non-specific Colitis, and Pervasive Developmental Disorder in Children," *The Lancet* 351 (1998): 637–641.
- Waldmann, Robert J., "Income Distribution and Infant Mortality," *Quarterly Journal of Economics* 107 (1992): 1283–1302.
- Wang, Jian, "Major Tasks for Five Priorities of Health Care Reform in 2011," China Center for Economic Research, manuscript, 2011.
- Wang, Justin, et al., "Do Bad Report Cards have Consequences?: Impacts of Publicly Reported Provider Quality Information on the CABG Market in Pennsylvania," *Journal of Health Economics* 30 (2011): 392–407.
- Warner, Kenneth E. and Linda M. Goldenhar, "The Cigarette Advertising Broadcast Ban and Magazine Coverage of Smoking and Health," *Journal of Public Health Policy* 10 (1989): 32–42.
- Warner, Kenneth E. et al., "Promotion of Tobacco Products: Issues and Policy Options," *Journal of Health Politics, Policy and Law* 11 (1986): 367–392.
- Watson, Diane, and Sara Allin, "The Canadian Health Care System, 2010," (New York: Commonwealth Fund, 2010).
- Wedig, Gerard J., "Health Status and the Demand for Health: Results on Price Elasticities," *Journal of Health Economics* 7 (1988): 151–163.
- Weiner, Jonathon P., "Forecasting the Effects of Health Reform on U.S. Physician Workforce Requirement," *Journal of the American Medical Association* 272 (1994): 222–230.
- Weinick, Robin M., Samuel H. Zuvekas, and Joel W. Cohen, "Racial and Ethnic Differences in Access to and Use of Health Care Services, 1977 to 1996," *Medical Care Research and Review* 57 (Supplement) (2000): 36–54.
- Weinstein, M. A., "The Sum of Values from a Normal and a Truncated Normal Distribution," *Technometrics* 6 (1964): 104–105.
- Weisbrod, Burton A., "Costs and Benefits of Medical Research: A Case Study of Poliomyelitis," *Journal of Political Economy* 79 (1971): 527–544.
- Weisbrod, Burton A., "The Health Care Quadrilemma: An Essay on Technological Change, Insurance, Quality of Care, and Cost Containment," *Journal of Economic Literature* 29 (1991): 523–552.
- Weisbrod, Burton A., *The Nonprofit Economy*. (Cambridge, MA: Harvard, 1988).
- Weisbrod, Burton A., "Toward a Theory of the Voluntary Non-Profit Sector in a Three-Sector Economy," in *Altruism, Morality, and Economic Theory*, Edmund S. Phelps (ed.) (New York: Russell Sage Foundation, 1975).
- Wells, William, John Burnett, and Sandra Moriarity, *Advertising: Principles and Practice*. (Upper Saddle River, NJ: Prentice Hall, 1989).
- Wennberg, John E., "Dealing with Medical Practice Variations: A Proposal for Action," *Health Affairs* 3 (1984): 6–32.
- Wennberg, John E., "Population Illness Rates Do Not Explain Population Hospitalization Rates," *Medical Care* 25 (1987): 354–359.
- Wennberg, John E., "Small Area Analysis and the Medical Care Outcome Problem," in *Research Methodology: Strengthening Causal Interpretation of Non-Experimental Data*, E. Perrin and J. Bunker (eds.) Rockville: Department of Health and Human Services, PHS 90-3454, 1990: 177–213.
- Wennberg, John E. and F. J. Fowler, "A Test of Consumer Contribution to Small Area Variations in Health Care

- Delivery,” *Journal of the Maine Medical Association* 68 (1977): 275–279.
- Wennberg, John E. and Alan Gittelsohn, “Small Area Variations in Health Care Delivery,” *Science* 182 (1973): 1102–1108.
- Wennberg, John E. and Alan Gittelsohn, “Variations in Medical Care Among Small Areas,” *Scientific American* 246 (1982): 120–134.
- Wennberg, John E., B. A. Barnes, and M. Zubkoff, “Professional Uncertainty and the Problem of Supplier-Induced-Demand,” *Social Science and Medicine* 16 (1982): 811–824.
- Wennberg, John E. et al., “Changes in Tonsillectomy Rates Associated with Feedback and Review,” *Pediatrics* 59 (1977): 821–826.
- Wennberg, John E. et al., “Hospital Use and Mortality Among Medicare Beneficiaries in Boston and New Haven,” *New England Journal of Medicine* 321 (1989): 1168–1173.
- Wheeler, Benedict W., “Health-Related Environmental Indices and Environmental Equity in England and Wales,” *Environment and Planning A* 36 (2004): 803–822.
- Whynes, David K., Emma J. Frew, and Jane L. Wolstenholme, “Willingness-to-Pay and Demand Curves: A Comparison of Results Obtained Using Different Elicitation Formats,” *International Journal of Health Care Finance and Economics* 5 (2005): 369–386.
- Wickizer, Thomas M. and Paul J. Feldstein, “The Impact of HMO Competition on Private Health Insurance Premiums, 1985–92,” *Inquiry* 32 (1995): 241–251.
- Wickizer, Thomas M., John R. C. Wheeler, and Paul J. Feldstein, “Have Hospital Inpatient Cost Containment Programs Contributed to the Growth in Outpatient Expenditures?” *Medical Care* 29 (1991): 442–451.
- Wiggins, Steven N. and Robert Mannes, “Price Competition in Pharmaceuticals: The Case of Anti-Infectives,” *Economic Inquiry* 42 (2004): 247–262.
- Williams, Alan, “Intergeneration Equity: An Exploration of the ‘Fair Innings’ Argument,” *Health Economics* 6 (1997): 117–132.
- Williams, Alan, “The ‘Fair Innings Argument’ Deserves a Fairer Hearing!,” *Health Economics* 10 (2001): 583–585.
- Williams, Alan and Richard Cookson, “Equity in Health,” *Handbook of Health Economics*, Volume I, A. J. Culyer and J. P. Newhouse (eds.) (Amsterdam, Elsevier Science, B.V., 2000).
- Williams, Jenny, Lisa M. Powell, and Henry Wechsler, “Does Alcohol Consumption Reduce Human Capital Accumulation? Evidence from a College Alcohol Study,” *Applied Economics* 10 (2003): 1227–1239.
- Williams, Scott C. et al., “Quality of Care in U.S. Hospitals as Reflected by Standardized Measures, 2002–2004,” *New England Journal of Medicine* 353 (2005): 255–264.
- Williamson, Oliver E., *The Economics of Discretionary Behavior: Managerial Objectives in the Theory of the Firm* (Upper Saddle River, NJ: Prentice Hall, 1964).
- Wittman, Donald, “Efficient Rules in Highway Safety and Sports Activity,” *American Economic Review* 72 (1982): 78–90.
- Wolfe, Barbara L. and Jere R. Behrman, “Women’s Schooling and Children’s Health: Are the Effects Robust with Adult Sibling Control for the Women’s Childhood Background?” *Journal of Health Economics* 6 (1987): 239–254.
- Wolfe, John R. and John H. Goddeeris, “Adverse Selection, Moral Hazard, and the Wealth Effects in the Medigap Insurance Market,” *Journal of Health Economics* 10 (1991): 433–459.
- Woods, Robert and P. R. Andrew Hinde, “Mortality in Victorian England: Models and Patterns,” *Journal of Interdisciplinary History* 18 (1987): 27–54.
- Woolhandler, Steffie, Terry Campbell, and David U. Himmelstein, “Costs of Health Care Administration in the United States and Canada,” *New England Journal of Medicine* 349 (2003): 768–775.
- World Bank Datasite, [www.worldbank.org/data/databytopic/class.htm](http://www.worldbank.org/data/databytopic/class.htm), accessed April 29, 2002.
- World Health Organization, *Vital Statistics*, various years.
- Writing Group for the Women’s Health Initiative Investigators, “Risks and Benefits of Estrogen Plus Progestin in Healthy Postmenopausal Women,” *Journal of the American Medical Association* 288 (2002): 321–333.
- Wu, Vivian Y., “Hospital Cost Shifting Revisited: New Evidence from the Balanced Budget Act of 1997,” *International Journal of Health Care Finance and Economics* 10 (2010): 61–83.
- Wu, Vivian Y., “Managed Care’s Price Bargaining with Hospitals,” *Journal of Health Economics* 28 (2009): 350–360.
- Yaisawarng, Suthathip, “Performance-based Budgeting in the Public Sector: An Illustration from the VA Health Care System,” *Health Economics* 15 (2006): 295–310.
- Yaniv, Gideon, Odelia Rosin, and Yossef Tobel, “Junk-food, Home Cooking, Physical Activity and Obesity: The Effect of the Fat Tax and the Thin Subsidy,” *Journal of Public Economics* 93 (2009): 823–830.
- Yip, Winnie C., “Physician Response to Medicare Fee Reductions: Changes in the Volume of Coronary Bypass Graft (CABG) Surgeries in the Medicare and Private Sectors,” *Journal of Health Economics* 17 (1998): 675–699.
- Yoo, Byung-Kwang et al., “Impacts of Informal Caregiver Availability on Long-term Care Expenditures in OECD Countries,” *Health Services Research* 39 (2004): 1971–1992.

- Zeckhauser, Richard J. and David V. P. Marks, "Sign-Posting Selected Product Information and Market Function," manuscript, John F. Kennedy School of Government, 1989.
- Zipkin, Amy, "The Concierge Doctor is Available (at a Price)," *New York Times*, July 31, 2005, [www.nytimes.com/2005/07/31/business/your\\_money/31doctor](http://www.nytimes.com/2005/07/31/business/your_money/31doctor).
- Zuckerman, Stephen, Jack Hadley, and Lisa Iezzoni, "Measuring Hospital Efficiency with Frontier Cost Functions," *Journal of Health Economics* 13 (1994): 255–280.
- Zwanziger, Jack and Adam Meirowitz, "Strategic Factors in Hospital Selection for HMO and PPO Networks," in *Managed Care and Changing Health Markets*, Michael A. Morrisey (ed.) (Washington, DC: AEI Press, 1998).
- Zwanziger, Jack and Glen A. Melnick, "The Effects of Hospital Competition and the Medicare PPS Program on Hospital Cost Behavior in California," *Journal of Health Economics* 7 (1988): 301–320.
- Zwanziger, Jack and Cathleen Mooney, "Has Competition Lowered Prices?" *Inquiry* 42 (2005): 73–85.

# NAME INDEX

## A

Aaberge, Rolf, 383  
Aaron, Henry J., 481  
Abbott, Thomas A., 357  
Absolo, Ignacio, 368  
Acemoglu, Daron, 109  
Acton, Jan P., 180  
Adams, Scott J., 217  
Ahituv, Avner, 535  
Aigner, Dennis J., 119  
Aizer, Anna, 97  
Akerlof, George A., 198, 199  
Al, Maiwenn J., 72  
Aldy, Joseph E., 74, 75, 540  
Alexander, Jeffrey A., 311  
Allin, Sara, 476, 477, 482  
Almond, Douglas, 95  
Altman, Stuart H., 292  
Alvarez-Dardet, Carlos, 87  
Andermann, Anne Adina Judith, 123  
Anderson, Geoffrey, 479  
Anderson, Gerard F., 140, 487, 489  
Anderson, John P., 139  
Anderson, Odin W., 401  
Andreoni, James, 267  
Angell, Marcia, 359  
Arnould, Richard J., 252, 269  
Arrow, Kenneth J., 13, 195, 267, 269, 386  
Ashcroft, John, 448  
Atkinson, Scott E., 94  
Auld, M. Christopher, 102  
Austin, D. Andrew, 224  
Aylward, Bruce, 537

## B

Bae, Jay P., 357  
Bain, Joe S., 521  
Baker, Laurence C., 126, 259, 397  
Bala, Mohan V., 79  
Balcombe, Kelvin, 397  
Balia, Silvia, 99  
Ball, James, 473  
Ballou, Jeffrey P., 277  
Balsa, Ana I., 189  
Baltagi, Badi H., 522, 526  
Barbetta, Gian Paolo, 120  
Barry, John M., 537  
Barry, Michael J., 316  
Barsukiewicz, Camille K., 283  
Bashaw, David J., 338  
Bates, Alan, 448  
Bates, Laurie J., 292  
Bays, Carson W., 268  
Beaulieu, Nancy D., 206  
Becker, Gary S., 73, 140, 517, 521, 527, 528  
Becker, Mark P., 313  
Bentham, Jeremy, 385  
Bergman, Richard N., 131  
Besser, Richard, 532–533  
Bhagwati, Jagdish, 501  
Bhattacharya, Jay, 141, 217, 340

Bilodeau, David, 114  
Bishop, Christine E., 298  
Blank, Rebecca, 372  
Bleichrodt, Han, 79  
Blendon, Robert J., 259  
Blomquist, Ake, 186  
Bloom, David E., 536  
Blumenthal, David, 307, 310  
Blundell, Richard, 180  
Bolin, Kristian, 100  
Bonsang, Eric, 298  
Borisova, Natalia, 141  
Boulier, Bryan L., 534  
Boyle, Sean, 469  
Bradford, W. David, 120, 316  
Brekke, Kurt R., 45, 246  
Brennan, Troyen A., 338  
Brewer, Jan, 463  
Brickley, James A., 277  
Brody, Howard, 328  
Brouwer, Werner B. F., 71, 379  
Brown, Douglas M., 108, 324–325  
Brown, Jerry, 463  
Brown, Randall S., 114  
Brown, Timothy T., 100, 527  
Browne, Mark J., 201  
Bryan, Stirling, 73  
Bunce, Victoria Craig, 230  
Bundorf, M. Kate, 207, 217  
Burgess, James F., Jr., 109, 120  
Burke, Mary, 125  
Burststein, Philip L., 337

## C

Cahill, Timothy, 408  
Calem, Paul S., 288  
Campbell, Christine M., 527  
Campbell, Denis, 473, 481  
Cannon, Michael, 500  
Capilouto, Eli I., 524  
Capps, Cory, 121, 292  
Card, David, 458  
Cardon, James H., 201  
Carey, Kathleen, 115  
Carlsen, Fredrik, 310  
Carlson, William L., 54  
Carpenter, Christopher, 525  
Carter, Richard A. L., 186  
Caudill, Stephen B., 125  
Cawley, John, 2, 143  
Chakvarty, Sujoy, 270, 278  
Chaloupka, Frank J., 397, 516, 520, 522, 523, 525, 527  
Chan, Sandra W., 523  
Chaoulli, Jacques, 482  
Chassin, Mark R., 313  
Chatterji, Pinka, 516  
Chen, Lin, 474  
Chen, Pauline W., 328  
Chernew, Michael C., 191, 206  
Chilingerian, Jon, 120  
Chiswick, Barry R., 184, 186  
Chou, Shin-Yi, 206, 280

Christensen, Eric W., 269  
Ciliberto, Frederico, 292  
Civan, Abdülkadir, 364  
Clancy, Carolyn, 500  
Clark, Andrew E., 527  
Clarkson, Kenneth W., 354  
Claxton, Karl, 71  
Cleverly, William O., 291  
Clinton, Hilary, 439  
Coate, Douglas, 35, 523, 525  
Cobb, Charles, 37  
Cochrane, Archibald L., 87  
Cockburn, Iain, 360  
Coffey, Rosanna M., 180  
Cogan, John F., 497, 500, 501  
Cohen, Joel W., 189, 270  
Cohen, Joshua T., 505  
Colditz, Graham A., 59  
Cole, Nancy, 101  
Collins, Sara R., 397  
Comanor, William S., 355, 521  
Commanor, William S., 100  
Cone, Kenneth, 411  
Conover, Christopher J., 411  
Conrad, Robert F., 114  
Contoyannis, Paul, 99  
Conway, Karen S., 97  
Cook, Phillip J., 516, 525, 527  
Cookson, Richard, 79, 368, 380–381  
Cooper, Philip F., 220, 260  
Corman, Hope, 96  
Corts, Kenneth S., 259  
Courtemanche, Charles, 292, 305  
Cowing, Thomas G., 114  
Cowling, Keith, 522  
Cremieux, Pierre-Yves, 94  
Crocker, Thomas, 94  
Cromwell, Jerry, 184, 337  
Cropper, Maureen L., 75, 101  
Cruz-Smith, Martin, 514  
Culyer, Anthony J., 368, 376, 377  
Currie, Janet, 96, 97, 98, 101, 191, 451, 459  
Custer, William S., 109  
Cutler, David M., 2, 74, 75, 76, 92, 95, 102, 144, 239, 249–250, 254, 280, 453, 456, 481, 497, 504

## D

Dalton, Kathleen, 450  
D' Andrea, Guy, 203  
Daniels, Norman, 80, 381  
Danzon, Patricia M., 360, 361  
Datta, Tejwant S., 534  
Dave, Dhaval, 523, 526  
David, Guy, 189  
Davis, Karen, 457  
Davis, Mark A., 270, 295  
Deaton, Angus, 101  
Deb, Partha M., 108, 254  
Debebe, Zelalem, 161  
De Brantes, Francois, 203  
Decker, Frederic H., 295

- Deer, Brian, 538  
Defelice, Lisa A., 120  
Deily, Mary, 120  
Dekker, Thijs, 75  
Derrick, Frederick W., 310  
DeSimone, Jeff, 516  
d'Hombres, Beatrice, 100  
DiCicca, Phillip, 527  
Dills, Angela K., 520  
DiMasi, Joseph A., 349, 359, 400  
Dixon, Anna, 500  
Dobkin, Carlos, 458  
Dobson, Allen, 411  
Dockery, Douglas W., 101  
Doerpinghaus, Helen I., 201  
Dolan, Paul, 79  
Donabedian, Avedis, 251  
Dong, Frederick, 78  
Donohue, Julie, 348  
Dorn, Stan, 462  
Doroodian, Khowsrow, 522  
Douglas, Paul, 37  
Dow, William H., 260  
Dowd, Bryan, 167, 185  
Doyle, Joseph, 98, 326  
Dranove, David, 116, 121, 238, 257, 289,  
309, 310, 314, 316, 350, 410, 411  
Drummond, Michael F., 66  
Dubos, Rene, 87  
Duclos, Jean-Yves, 484  
Dyck, Frank J., 313
- E**  
Easley, David, 269  
Échevin, Damien, 484  
Eddy, David M., 316  
Efron, Bradley, 53  
Eggleston, Karen, 277, 473, 475  
Elixhauser, Anne, 81  
Ellis, Randall P., 111, 248, 280  
Elzinga, Kenneth G., 430  
Emanuel, Ezekiel J., 217  
Emmons, David W., 429  
Epstein, Andrew, 312, 314  
Escarce, José J., 124, 313, 325  
Evans, Robert G., 114, 306  
Evans, Ronald M., 131  
Evans, William N., 96, 100, 526  
Ewer, Steven M., 326
- F**  
Fang, Hai, 261, 136  
Fare, Rolf, 118  
Farmer, Paul, 87  
Farsi, Mehadi, 114, 280  
Fauci, Anthony S., 539  
Feinglass, Joe, 422  
Feldman, Roger D., 167, 185, 229, 258, 306,  
310, 334, 500, 501  
Feldstein, Martin, 166, 167  
Feldstein, Paul J., 259  
Finkelstein, Amy, 109, 143, 456, 459  
Fletcher, Suzanne W., 59  
Flood, Colleen, 483  
Fogel, Robert W., 91, 92, 93, 533  
Folland, Sherman T., 99, 100, 120, 280, 312,  
313, 315, 386, 527  
Fontana, Patrick, 408
- Ford, Henry, 497  
Forder, Julien, 298  
Fournier, Gary M., 114  
Fowler, Floyd J., 313  
Fowler, Lynette, 120  
Frank, Richard G., 261, 357, 359  
Frech, H. E., III, 97, 100, 258, 268, 364  
Frederick, Shane, 70  
Freeland, Mark S., 189, 233  
Frenk, Julio, 87  
Frew, Emma J., 73  
Friedman, Bernard, 126  
Friedman, Milton, 11, 336  
Fuchs, Victor R., 17, 91, 98, 99, 102, 106,  
217, 327, 335, 382, 461, 478
- G**  
Gabel, Jon, 400  
Galewitz, Phil, 214  
Gallett, Craig A., 522  
Garavaglia, 280  
Garber, Alan M., 77, 79  
Garcia-Frapolli, 69  
Garrett, Bowen, 462  
Gaumer, Gary L., 338, 339  
Gauthier, Anne K., 500  
Gaynor, Martin, 45, 120, 185, 204, 277,  
427, 429  
Geoffard, Pierre-Yves, 534  
George, Lloyd, 437  
Gerdtham, Ulf-G., 140, 186, 475  
Gertler, Paul J., 270  
Getzen, Thomas E., 186  
Ghosh, Arunabh, 462  
Gibson, Teresa R., 362  
Gillum, Brenda S., 260  
Ginsburg, Paul B., 258, 416  
Glaeser, Edward L., 100, 144  
Glantz, Stanton A., 522  
Glied, Sherry, 254  
Goddeeris, John H., 201, 231, 280  
Gold, Marthe R., 66, 77, 81  
Goldberg, Lawrence G., 222  
Goldberg, Theodore, 363  
Goldfarb, Robert S., 534  
Goldman, Dana P., 3, 201, 362  
Goldstein, Karen P., 535  
Golec, Joseph, 357  
Gompers, Samuel, 438  
Goodman, Allen C., 56, 135, 141, 185,  
246, 260  
Goodman, Clifford S., 66  
Goodman, John C., 499  
Gordon, Margaret, 466  
Grabowski, David C., 270, 277, 297, 359, 449  
Grabowski, Henry G., 357, 359  
Graddy, Elizabeth, 338  
Grady, Denise, 218–219  
Grannemann, Thomas W., 114  
Grant, Darren, 202  
Graves, Edmund J., 260  
Green, Lee A., 313, 500  
Greenberg, Warren, 222  
Greene, Jessica, 500  
Gregory, Paul R., 35  
Grépin, Karen A., 364  
Griffin, James M., 526  
Grossman, Joy M., 416
- Grossman, Michael, 35, 94, 96, 102, 129,  
191, 517, 523, 527, 528  
Gruber, Jonathan, 96, 98, 101, 217, 219, 268,  
277, 297, 305, 310, 445, 452, 453, 459,  
462, 507, 525, 526  
Grytten, Jostein, 202, 310  
Guardado, José R., 429  
Gully, O. David, 280
- H**  
Haas-Wilson, Deborah, 205, 427  
Hadley, Jack, 94, 120, 185, 494–495  
Hahn, James S., 478  
Häkkinen, Unto, 114  
Hall, Thomas D., 336  
Hamilton, James L., 523  
Hamilton, Vivian, 298  
Hammit, James K., 75  
Hanratty, Maria, 98  
Hansen, Ronald W., 359  
Hansen, W. Lee, 330  
Hansmann, Henry B., 269, 270, 277, 280  
Harrington, Scott, 189  
Harris, Jennifer L., 397  
Harris, Gardiner, 328, 345  
Harris, Jeffrey E., 278–279, 523  
Harrison, Teresa, 292  
Hassan, Mahmud, 277  
Hatch, Elizabeth E., 54  
Hawley, Chris, 192  
Hayek, Friedrich, 386  
Headen, Alvin E., Jr., 337  
Helland, Eric, 315  
Hendel, Igal, 201  
Henderson, Rebecca, 360  
Henderson, Rochelle, 362  
Hensley, Scott, 538  
Hethcote, Herbert W., 533  
Hill, Steven C., 126  
Hilsenrath, Peter, 538  
Himmelstein, David U., 481  
Hirsch, Barry T., 331  
Hirth, Richard, 270, 277, 278  
Ho, Katherine, 305  
Hoadley, Jack, 443  
Hoch, Irving, 315  
Hoff, Timothy J., 340  
Hofler, Richard A., 120, 280  
Hogarty, Thomas F., 430  
Holahan, John, 458, 462  
Hollingsworth, Bruce, 120, 280  
Holloway, James J., 422  
Holmes, Ann M., 108  
Holtmann, Alphonse G., 114, 281  
Hornbrook, Mark C., 115  
Horrace, William C., 97  
Horwitz, Jill R., 280  
Hotz, Joseph, 535  
Howard, David H., 205  
Howard, Hilda A., 450  
Hsiao, William, 425  
Huang, Lien-fu, 186  
Hubbard, R. Glenn, 497, 500, 501  
Hughes, Susan L., 298  
Hungerford, Thomas L., 224  
Hungerman, Daniel M., 268, 277  
Hurd, Michael D., 188  
Hurley, Jeremiah, 368, 377, 500

- I**  
 Idson, Todd L., 281  
 Iezzoni, Lisa, 120  
 Iglehart, John K., 223  
 Ignagni, Karen, 164  
 Illich, Ivan, 87  
 Illife, Steve, 109  
 Islam, Kamrul, 100  
 Iversen, Tor, 305, 310  
 Iwasaki, Natsuko, 523
- J**  
 Jacobson, Lena, 100  
 Jacobson, Mirelle, 520  
 Jacobson, Peter D., 325  
 Jantzen, Robert, 292  
 Jayachandran, Seema, 95  
 Jenner, Edward, 533  
 Jensen, Gail A., 109, 217, 230, 260, 400  
 Jowett, Sue, 73  
 Jha, Ashish K., 293  
 Jiménez-Martin, Serji, 191  
 Joesch, Jutta M., 259  
 Johannesson, M., 140  
 Johnson, Lyndon B., 438  
 Jones, Andrew, 99, 383, 384  
 Jones, Andrew M., 99  
 Jones, Lewis W., 327  
 Jones, Robert F., 332  
 Jonsson, Bengt, 475  
 Joyce, Geoffrey F., 362  
 Joyce, Theodore J., 94, 96, 100, 507
- K**  
 Kaarboe, Oddvar, 100  
 Kane, Carol K., 429  
 Kantarevic, Jasmin, 326  
 Kaplan, Robert M., 139  
 Kapur, Kanika, 410  
 Katzman, Brett, 527  
 Kawachi, Ichiro, 397  
 Keehan, Sean P., 16, 402, 508  
 Keeler, Emmett B., 292  
 Keeler, Theodore E., 35, 523, 524, 525  
 Kefauver, Estes, 349  
 Kenkel, Donald S., 140  
 Kennedy, Anthony, 448  
 Kermack, William O., 533  
 Kessel, Reuben, 11, 244, 335, 336  
 Kessler, Daniel P., 270, 289, 497, 500, 501  
 Kim, Daniel, 397  
 Kleinerman, Ruth A., 54  
 Kochi, Ikuho, 75  
 Köksal, Bülent, 364  
 Kolata, Gina, 164  
 Konetzka, Tamara, 297  
 Koolman, Xander, 383, 384  
 Kooreman, Peter, 119, 120  
 Köszegi, Botond, 526  
 Kronick, Richard, 229  
 Krueger, Alan B., 217, 495, 496  
 Kutinova, Andrea, 97  
 Kuznets, Simon, 11, 336
- L**  
 Laine, Juha, 120  
 Lakdawalla, Darius N., 141, 279  
 Lalonde, Marc, 86  
 Lamberton, Charles E., 184  
 Landon, Bruce E., 254  
 Lane, H. Clifford, 539  
 Lanoie, Paul, 520  
 Laugesen, Murray, 523  
 Law, Michael R., 364  
 Leclair, Paul, 520  
 Lee, Dwight R., 207  
 Lee, Robert H., 185  
 Lee, Roger I., 327  
 Lee, Young Hoon, 119  
 Leeper, Ed, 53  
 Leffler, Keith B., 333, 337  
 Le Grand, Julian, 471–472  
 Lehner, Laura A., 109  
 Leibowitz, Arlene A., 140  
 Leonard, Kenneth L., 205, 316, 317  
 Lerner, Abba, 427  
 Lessler, Daniel S., 410  
 Levin, Dan, 522  
 Levitt, Steven D., 515  
 Levy, Helen, 443  
 Lewit, Eugene M., 35, 523, 525  
 Liang, J. Nellie, 108  
 Lichtenberg, Frank R., 361, 364  
 Lien, Diana S., 96, 277  
 Light, Donald W., 359  
 Lindeboom, Maarten, 102  
 Lindgren, Bjorn, 100, 519  
 Lindhjem, Henrik, 75  
 Lindrooth, Richard, 116, 121, 292  
 Lindsay, Cotton M., 333, 336  
 Linet, Martha S., 54  
 Lingle, Linda, 454  
 Linna, Miika, 114  
 Liska, David, 462  
 Liu, Hong, 261  
 Liu, Zhimei, 260  
 Lleras-Muney, Adriana, 95, 102  
 Locke, John, 386  
 Loewenstein, George, 70  
 Loheac, Youenn, 527  
 Lo Sasso, Anthony T., 341  
 Loubeau, Patricia R., 292  
 Lovell, C.A. Knox, 119  
 Lu, John Z., 355  
 Luft, Harold S., 252, 253, 256  
 Lundborg, Petter, 519  
 Luo, Ye, 297  
 Ly, Dan P., 497
- M**  
 Ma, Jin, 475  
 MacDonald, Ziggy, 517  
 Madan, Sandan, 501  
 Madrian, Brigitte C., 219, 220  
 Maestas, Nicole, 458  
 Magnusson, Jon, 120  
 Mahal, Ajay, 536  
 Mann, Thomas, 90  
 Mannes, Robert, 357  
 Manning, Willard G., 35, 60, 167, 184, 252, 253, 519  
 Mansfield, Edwin, 35, 358  
 Mark, Tami L., 120, 280  
 Markowitz, Sara, 527  
 Marquis, M. Susan, 167, 260  
 Mason, Michael, 408  
 Mather, Dell, 363  
 Maude-Griffin, 168  
 Mayberry, Robert M., 189  
 Maynard, Alan, 471  
 Mays, Glen P., 401  
 Mbeki, Thabo, 545  
 McCain, John, 439, 504  
 McCarthy, Patrick S., 520  
 McCarthy, Thomas R., 185  
 McClellan, Mark, 239, 254, 270, 277, 289, 413  
 McCue, Sally M., 313  
 McCullough, Jeffrey S., 125  
 McGarry, Kathleen, 188  
 McGeary, Kerry Anne, 527  
 McGlynn, Elizabeth A., 204  
 McGuinness, Tony, 522  
 McGuire, Thomas G., 189, 195, 303, 308, 309, 310  
 McKay, Niccie L., 120, 288  
 McKendrick, Anderson G., 533  
 McKeown, Thomas, 87, 88, 89, 91, 533  
 McKinlay, John B., 90, 533  
 McKinlay, Sonja M., 90, 533  
 McKnight, Robin, 459  
 McLaughlin, Catherine G., 228, 258  
 McQueen, M. P., 501  
 Meads, Chris, 523  
 Meara, Ellen, 456  
 Meiorowitz, Adam, 244  
 Mello, Michelle, 489  
 Mellor, Jennifer, 525  
 Melnick, Glen A., 258, 292, 421  
 Meltzer, David, 350  
 Meyerhoeffer, Chad D., 188  
 Miles, John J., 429  
 Mili, Fatima, 189  
 Mill, John Stuart, 386  
 Miller, Theodore R., 75  
 Miller, Douglas L., 100  
 Miller, Grant, 92  
 Miller, Richard, 97, 100, 364  
 Miller, Robert H., 252, 253  
 Miron, Jeffrey A., 520  
 Mitchell, Janet B., 184  
 Mitchell, Jean M., 114  
 Mocan, Naci, 100  
 Monheit, Alan C., 115, 180, 220  
 Mooney, Gavin, 289, 314  
 Moore, Michael J., 516, 526, 527  
 Moretti, Enrico, 97  
 Morrisey, Michael A., 2, 109, 217, 230, 239, 258, 291, 410, 422  
 Motheral, Brenda R., 362  
 Mrozek, Janusz R., 75  
 Mueller, Curt D., 180  
 Mukamel, Dana B., 185  
 Mullahy, John, 140, 520  
 Mullen, Kathleen, 261  
 Murphy, Kevin M., 73, 93, 95, 140, 517, 521, 527, 528  
 Mushkin, Selma, 93
- N**  
 Nardell, Edward, 87  
 Nash, John, 288, 416  
 Nassiri, Abdelhak, 305  
 Nauenberg, Eric, 482  
 Nelson, Alan R., 189, 520  
 Newbold, Paul, 54

Newhouse, Joseph P., 15, 60, 97, 98, 184,  
186, 188, 191, 233, 239, 254, 256, 271,  
275–276, 277, 333, 455, 479  
Nguyen, Nguyen X., 310  
Nicholson, Sean, 254, 312, 314, 339  
Nord, Eric, 80  
Nordhaus, William, 71  
Norton, Edward C., 260, 277, 294, 299  
Nozick, Robert, 385, 386–387  
Nyman, John A., 168, 170, 297

**O**

Oakeshott, Isabel, 471  
Obama, Barack, 365, 439, 504  
O'Donoghue, Ted, 70  
Ofili, Elizabeth, 189  
Ogur, Jonathon D., 108  
O'Halloran, Patrick L., 338  
O'Hara, Maureen, 269  
Ohsfeldt, Robert L., 524  
Okun, Arthur, 372  
Okunade, Albert, 108  
Olesen, Ole B., 114  
Olivella, Pau, 246  
Olkin, Ingram, 60  
Olsen, Jan Abel, 79  
O'Neill, Dave, 479, 483–484  
O'Neill, June, 479, 483–484  
Oreopoulos, Philip, 102  
Oster, Emily, 546  
Oullette, Pierre, 94  
Over, Mead, 543  
Owings, Maria, 305, 310  
Ozcan, Yasar A., 120

**P**

Pakes, Ariel, 305  
Parente, Stephen T., 313, 314, 501  
Pareto, Vilfredo, 368  
Paris, Valérie, 403  
Parkin, David, 186  
Paul, Chris, 337  
Pauly, Mark V., 109, 114, 115, 120, 153, 195,  
203, 215, 274, 275–276, 279, 302, 303,  
308, 309, 325, 376, 380, 387, 420, 496,  
498, 499, 502  
Paxson, Christina, 101  
Pearson, Steven D., 505  
Peden, Edgar A., 189, 233  
Peltzman, Sam, 337, 349, 403  
Petersen, Niels S., 114  
Peterson, Mark A., 13  
Phelps, Charles E., 77, 79, 184, 186, 306,  
311, 313, 314  
Phibbs, Ciaran, 126  
Philippini, Massimo, 114  
Philipson, Tomas, 73, 272, 279, 350, 362,  
534, 535, 536, 539  
Picone, Gabriel A., 191, 280  
Pilon, Caroline, 94  
Pink, George, 116  
Pisano, Susan, 218  
Plotzke, Michael Robert, 292, 305  
Polachek, Solomon, 204  
Polling, Red, 502  
Polsky, Daniel, 254, 315  
Porter, Jack Ray, 515  
Posner, Richard, 539  
Poullier, Jean-Pierre, 186

Powell, Lisa M., 295, 397, 516  
Pozen, Alexis, J. 481  
Preya, Colin, 116  
Pritchett, Lant, 101, 104

**Q**

Quast, Troy, 305  
Quiggin, John, 79

**R**

Racine, Andrew D., 100  
Raffel, Marshall W., 283, 294  
Raffel, Norma K., 283  
Rashad, Inas, 220  
Rawls, John, 385–386  
Reagan, Patricia B., 191  
Reber, Sarah, J., 249–250  
Redisch, Michael, 274, 275–276, 279, 420  
Reinhardt, Uwe E., 292, 306, 324, 380,  
495, 496  
Relman, Arnold S., 359  
Reynolds, Roger, 457  
Rhee, Sang-O., 335  
Ried, Walter, 77  
Ringel, Jeanne S., 100, 526  
Rizzo, John A., 125, 254, 261, 288, 305,  
307, 310  
Roberts, Brandon, 315  
Roberts, Mark J., 522  
Robertson, Peggy, 218–219  
Robinson, James C., 223  
Robinson, Lisa A., 75  
Rochaix, Lise, 305  
Rolph, John E., 12  
Romney, Mitt, 507  
Roos, Leslie L., 479  
Ropeik, David, 538  
Rosen, Harvey S., 3  
Rosenthal, Meredith B., 203, 261, 504  
Rosenzweig, Mark R., 94, 100, 515  
Rosett, Richard N., 186  
Rosin, Odelia, 142  
Ruffin, Roy J., 35  
Ruiz, Maria Teresa, 87  
Russell, Louise B., 81

**S**

Sacks, Jerome, 60  
Saffer, Henry, 520, 522, 523, 526, 527  
Safriet, Barbara J., 339  
Salkever, David S., 357, 410  
Samuelson, Larry, 522  
Santerre, Rexford E., 270, 280, 281,  
290, 292  
Sanz de Galdeano, Anna, 220  
Sappington, David, 305  
Sarpong, Eric, 220  
Sasser, Alicia C., 340  
Satterthwaite, Mark A., 203  
Scanlon, Dennis P., 206, 256  
Scanlon, William J., 296  
Schaeffer, Leonard, 223  
Schelling, Thomas, 516  
Schmidt, Peter, 119  
Schnelle, John F., 295  
Schoen, Cathy, 484, 486  
Schultz, T. Paul, 94, 100  
Schumacher, Edward, 331  
Schwartz, Joel, 101  
Schweitzer, Stuart O., 347, 358  
Segal, Leonie, 19, 81  
Seldon, Barry J., 337, 522  
Sen, Amartya, 79, 380  
Sen, Anindya, 525  
Sen, Bisakha, 527  
Sendi, Pedram, 72  
Sevilla, Jaypee, 536  
Shapiro, Jesse M., 144  
Shaw, James W., 97  
Shen, Yu-chu, 277  
Shen, Yu-chu, 277, 280  
Shenkman, Elizabeth, 305  
Shepperd, Sasha, 109  
Sheu, Mei-ling, 525  
Shleifer, Andrei, 413, 416  
Shortell, Stephen M., 238  
Showalter, Mark, 315  
Siciliani, Luigi, 45  
Sickles, Robin C., 94, 139  
Sidhu, Nirmal, 102  
Silverman, Elaine, 278, 419  
Simon, Carol J., 238, 257  
Simon, Kosali I., 229  
Sindelar, Jody, 140, 520  
Sisko, Andrea, 508  
Skinner, Jonathan, 278, 419  
Skocpol, Theda, 439  
Sloan, Frank A., 120, 125, 270, 276, 280,  
306, 310, 334, 411, 422  
Smedley, Brian D., 189  
Smith, Adam, 73, 386  
Smith, Francis B., 90  
Smith, James P., 101  
Smith, Richard D., 79  
Smith, Sharla A., 401  
Sorensen, Rune, 310  
Spector, William D., 185, 270  
Spence, A. Michael, 271  
Spetz, Joanne, 259  
Staiger, Douglas O., 277  
Stano, Miron, 135, 246, 312, 313  
Starr, Paul, 283, 335–336  
Stefos, Theodore, 115  
Steinberg, Jonny, 545  
Steinberg, Richard, 267  
Steiner, Claudia, 126  
Steinfels, Peter, 448  
Steinwald, A. Bruce, 334  
Stern, Nicholas, 71  
Stewart, Kathryn, 416–417  
Stigler, George J., 204, 337  
Stiglitz, Joseph, 72  
Stith, Adrienne Y., 189  
Straume, Odd Rune, 45  
Strauss, Robert P., 114  
Street, Andrew, 120  
Studdert, David M., 410  
Succi, Melissa J., 291  
Sullivan, Daniel, 331  
Summers, Lawrence H., 101, 104  
Suraratdecha, Chutima, 108  
Svorny, Shirley, 339

**T**

Tay, Abigail, 293  
Taylor, Laura O., 75  
Temin, Peter, 268, 283, 399  
Tengs, Tammy O., 68



## 594 Name Index

Teplensky, Jill D., 125  
Thorne, Betty, 54  
Tibshirani, Robert, 53  
Tilford, John M., 135, 184  
Tobel, Yossef, 142  
Topel, Robert H., 93, 95  
Toren, Mark, 120  
Torrance, George W., 77  
Tremblay, Carol Horton, 522, 523, 524  
Tremblay, Victor J., 522, 523, 524  
Troyer, Jennifer L., 295  
Trude, Sally, 261  
Truffer, Christopher J., 446  
Tsuchiya, Aki, 368  
Tumlinson, Anne, 206  
Turati, Gilberto, 120  
Tye, Joe B., 522

## U

Uribe, Martin, 191

## V

Valdez, R. Burciaga, 97  
Valdmanis, Vivian G., 120  
Valvona, Joseph, 422  
van Doorslaer, Eddy, 377, 383, 384  
Van Horn, R. Lawrence, 277  
Van Houten, Courtney H., 299  
van Kempen, Luuk, 161  
van Lear, William, 120  
Vara-Hernández, Marcos, 246  
Varkevisser, Marko, 180  
Vernon, John A., 270, 280, 281, 357,  
359, 360  
Viscusi, W. Kip, 67, 74, 75, 519, 540  
Vita, Michael G., 114  
Vitaliano, Donald F., 120

Vogel, Ronald J., 97  
Vogt, William B., 45, 185, 427, 429

## W

Waber, Rebecca L., 205  
Wade, Nicholas, 131  
Wagner, Todd H., 326  
Wagstaff, Adam, 376, 377  
Wakefield, Andrew J., 538  
Wallin, Susan, 458  
Wang, Jian, 476  
Warburton, Rebecca N., 359  
Warner, Kenneth E., 516, 522  
Watson, Diane, 476, 477, 482  
Wechsler, Henry, 516  
Wedig, Gerard J., 35, 184, 191  
Wehner, Paul, 309, 310  
Weinick, R. M., 189  
Weinstein, M.A., 81  
Weir, David, 443  
Weisbrod, Burton A., 230–231, 266, 268,  
277, 280  
Wells, Kenneth B., 12  
Wennberg, John E., 311, 312, 313  
Wertheimer, Nancy, 53  
Wheeler, Benedict W., 315  
White, William D., 202, 238, 257, 316  
Whitehead, Jenny, 471  
Whynes, David K., 73  
Wickizer, Thomas M., 259, 410  
Wiener, Joshua, 458  
Wiggins, Steven N., 357  
Wilde, Oscar, 74  
Williams, Aimee, 462  
Williams, Alan, 80, 368, 381  
Williams, Jenny, 516  
Williams, Scott C., 293

Willke, Richard J., 109  
Wilson, Duff, 345  
Wilson, Paul W., 120  
Wilson, R. Mark, 191  
Wilson, Thomas A., 521  
Windmeijer, Frank, 180  
Wirjanto, Tony S., 525  
Wittman, Donald, 387  
Wolfe, Barbara L., 126  
Wolfe, John R., 201  
Wolpin, Kenneth I., 100, 515  
Wolstenholme, Jane L., 73  
Wood, Elaine, 260  
Woolhandler, Steffie, 481  
Wu, Vivian Y., 254, 291

## Y

Yaisawarng, Suthathip, 121  
Yamada, Tadataka, 537  
Yaniv, Gideon, 142  
Yazbeck, Abdo, 94, 139  
Yip, Winnie C., 310  
Yoo, Byung-Kwang, 298

## Z

Zago, Angelo M., 120  
Zarkin, Gary A., 79  
Zeckhauser, Richard J., 125, 305, 307, 310  
Zeliotis, Georges, 482  
Zellers, Wendy K., 228  
Zhun Cau, 189  
Zimmerman, Heinz, 72  
Zinman, Jonathan, 525  
Zivin, Joshua Graff, 316, 317  
Zuckerman, Stephen, 120, 280  
Zuvekas, Samuel H., 188, 189  
Zwanziger, Jack, 244, 289, 421

# SUBJECT INDEX

- A**
- Access  
costs and, 9  
employer vs. individual mandates, 495–496  
ensuring, 495–498  
insurance separation from employment and, 496–497  
single payer vs. multiple insurers, 497–498
- ACE. *See* Angiotensin-converting enzyme inhibitor
- Addiction models  
application of, 518–519  
capital stock and, 518  
imperfectly rational, 516  
myopic, 516–517  
policy implications of, 519  
rational, 140, 517–518  
steady state and, 518
- Addictive stock, 518
- Adoption of technology, 124–126
- Adverse selection, 187, 218  
HMOs and, 249–251  
insurance and, 200–201
- Advertising  
alcohol consumption and, 523  
bads restrictions and, 520–523  
as barrier to entry, 521  
brand switching and, 522–523  
as complementary good, 521–522  
DTC, 316, 348  
as information, 520–521  
quackery and, 521  
theories, 520
- Africa case study, 539–546
- Age  
demand and, 191  
equilibrium and, 137–138
- Ageism, QALYs and, 80
- Agency  
defined, 202  
health care and, 202–203  
physician's practice and, 305–306
- Aggregate expenditures, insurance and, 189
- AIDS. *See* Auto-immune deficiency syndrome
- Alcohol consumption  
advertising and, 523  
bads and, 514, 515–516  
across countries, 514  
dangers of, 515–516  
excise taxes and, 523–528
- Allocative inefficiency, 117–118
- Alternative delivery systems  
development, 499
- Analysis, economic  
examples of, 11  
features of, 10
- Angiotensin-converting enzyme (ACE)  
inhibitor, 78
- Antitrust  
enforcement, 426  
exemptions, 426–427  
government regulation and, 426–431  
monopoly power measurement, 427–429  
procedures, 429–430
- Asian countries, 473
- Aspirin, 123
- Asymmetric information. *See also* Agency;  
Consumer information  
adverse selection and, 200–201  
experience rating and, 201  
imperfect information vs., 199  
information problems and, 196–197, 199  
lemons principle and, 198–201  
overview, 196
- Autism, 538
- Auto-immune deficiency syndrome (AIDS)  
annual and cumulative rates of, 539, 540  
case study, 539–546  
cost of, 540, 543  
diseases associated with, 539–540  
economic theory vs. reality, 545–546  
fighting, 543–545  
overview about, 539  
as pandemic, 539, 542  
in South Africa, 545
- B**
- Bads  
addiction models and, 516–519  
advertising restrictions and, 520–523  
alcohol consumption and, 514, 515–516  
cultures and behaviors and, 514–515  
excise taxes and, 523–528  
other interventions for, 520  
overview about, 513  
public intervention rationales for, 519–520  
smoking and, 515
- Basal Metabolic Rate (BMR), 142–143
- Benchmark model, supplier-induced demand, 308–309
- BMI. *See* Body Mass Index
- BMR. *See* Basal Metabolic Rate
- Body Mass Index (BMI), 141–142
- Brand switching, 522–523
- Budget constraint model, standard  
consumer's equilibrium and, 175–176  
demand shifters and, 176–177  
health status and demand and, 178  
overview, 174–175
- Budget constraints, 31
- Budgeting, performance-based, 121
- C**
- Cadillac tax, 439, 506
- Caesareans, 218–219
- Canada  
administrative costs in, 481  
background about, 476–478  
costs in, 479, 481  
GDP and, 469  
physician fees in, 478–479  
private insurance future in, 482  
resources, 469  
self-reported unmet need in, 484  
spending in, 468  
system origins, 477  
technology availability in, 480  
U.S. compared with, 477–478, 481–484
- Canada Health Transfer (CHT), 477
- Capital, 6–7
- Capital stock, 518
- CBA. *See* Cost-benefit analysis
- CDHPs. *See* Consumer-directed health plans
- CEA. *See* Cost-effectiveness analysis
- Certificate-of-need (CON), 411–412
- Chaoulli v. Quebec* lawsuit, 482
- Cheatham, J. W., LLC, 214
- Chicken Pox, 533
- Children, Medicaid and, 458
- Children's Health Insurance Program (CHIP), 450–451, 453–454, 462
- China  
conclusions about, 475–476  
emerging system in, 473–476  
GDP and, 469  
government policy initiatives in, 474  
nonprofit and for-profit in, 475  
organization of care in, 473–474  
performance of system in, 474–475  
private sector and, 474  
spending in, 468  
urban and rural areas in, 474
- CHIP. *See* Children's Health Insurance Program
- Choice, for patients/providers, 493, 508
- Cigarettes, 99–100. *See also* Smoking  
advertising and, 520–523  
excise taxes and, 523–528
- Clinical decision making, 316
- Coinurance, 149–150  
demand shifters and, 180–182  
impact of, 162–166  
market effects and, 181–182  
reduced, 180–181
- Commodity taxes and subsidies, 395–396
- Comparative health care systems. *See also*  
Contemporary health care systems  
public's evaluation, 484–487  
spending differences across countries, 487–489
- Comparative statics, 25–26
- Comparative utilization, HMOs and, 252
- Compensating differentials, 217–218
- Competition  
costs and, 287–291  
hospitals and, 257–258, 287–291  
insurance market, 258–259  
international, 502–504  
managed care and, 256–259  
model deviations, 373–374  
normal profits and, 157–159  
perfect, 41–43, 373–374  
pharmaceutical industry, 346–347, 357  
promotion of, 374–375  
restriction of, 15  
theoretical issues of, 256–257  
yardstick competition theory, 413–417

- Competitive strategies  
 alternative delivery systems  
   development, 499  
 CDHPs, 499–500  
 free trade, 501  
 other market, 500–501  
 reform, health system, 498–502  
 regulation vs., 498  
 representation of, 501–502  
 Complementary good, 521–522  
 CON. *See* Certificate-of-need  
 Consumer  
   DTC advertising, 316, 348  
   equilibrium, 32–33, 175–176  
   as health producer, 129–130  
 Consumer-directed health plans (CDHPs), 499–500  
 Consumer information  
   informed buyers and, 204  
   prices and, 203–204  
   quality and, 204–207  
 Consumer theory  
   budget constraints and, 31  
   consumer equilibrium and, 32–33  
   demand curve and, 28–33  
   described, 28  
   indifference curves and, 30–31  
   utility and, 29–30  
 Contemporary health care systems  
   Asian countries compared, 473  
   Canada, 476–484  
   China, 473–476  
   country resources compared, 469  
   spending by country, 468  
   technology availability by country, 480  
   typology of, 466–467  
   United Kingdom, 467–473, 480  
 Contract failure, 269–270  
 Cookstoves, 69  
 Copayment, 149–150  
   drugs and, 164  
   impact of, 162–166  
   pharmaceutical industry and, 361–362  
 Cost-benefit analysis (CBA)  
   background, 66  
   basic principles, 66–72  
   heart care treatment and, 75–76  
   human life valuing and, 72–76  
   marginal analysis in, 68–69  
   measuring costs, 67  
   overview about, 66–67  
   risk and, 67–68  
 Cost containment, 361–363, 493, 507–508  
 Cost-effectiveness analysis (CEA), 76  
   ACE inhibitor and, 78  
   advantages of, 77  
 Cost-function  
   deriving, 110–111  
   empirical studies, 114–116  
   structural vs. behavioral, 114  
   study difficulties, 114–115  
 Costs. *See also* Loading costs  
   access and, 9  
   of AIDS, 540, 543  
   Canada's, 479, 481  
   competition and, 287–291  
   containment, 361–363, 493, 507–508  
   cost curves, marginal and average, 40  
   demand for insurance and, 166  
   GDP and, 3–5  
   of health capital, 135–136  
   HMOs and, 252  
   hospitals, 286–291  
   inflation and, 8–9  
   managed care and, 255–256  
   measuring, 67  
   minimization, 39–40, 111–112  
   nursing homes, 287, 294  
   other countries and, 4–5  
   personal spending and, 5–6  
   preventive care and, 505  
   of prospective payment system, 421–423  
   quality and, 9  
   relevance of, 12–13  
   saving lives and, 67–68  
   shifting, 289–291  
   society and, 40  
   technological change and, 3, 121–124, 231  
   in theory and practice, 109–116  
   universal coverage and, 494–495  
 Cost-utility analysis (CUA), 77–80  
 Creaming, 248, 249  
 Crowd-out, 453  
 CUA. *See* Cost-utility analysis
- D**  
 DALYs. *See* Disability-adjusted life-years  
 Data envelopment analysis (DEA), 118–119  
 Decision making, rational, 10  
 Deductible, 150, 162  
 Demand. *See also* Supplier-induced demand;  
   Supply and demand  
   age, health status, and uncertainty and, 191  
   brand switching and, 522–523  
   consumer as health producer and, 129–130  
   derived, 28  
   education and, 190  
   elasticities, 184–188  
   ethnicity and gender and, 189–190  
   excess, 296–297, 328–329  
   functions, 27–28  
   for health, 129–131  
   for health capital, 136–137  
   health status and, 178  
   individual and market functions of,  
     182–183  
   for insurance, 153–157, 166  
   issues in measuring, 182–184  
   of labor, 319–324  
   labor shortages and, 328–329  
   moral hazard and, 159–161  
   nursing homes and, 296–297  
   other variables influencing, 189–191  
   prevention, 535  
   price of care and, 166  
   regression, 56–57  
   resource allocation and, 162–168  
   unmet, 266–267  
   urban vs. rural, 190  
 Demand curve  
   budget constraint model and, 176–177  
   consumer theory and, 28–33  
   demand shifters and, 23–24  
   economic efficiency and, 64  
 Demand shifters, 176–178  
   coinsurance and, 180–182  
   demand curve and, 23–24  
   time and, 178–180  
 Department of Defense (DOD), 400  
 Department of Veterans Affairs (VA), 400  
 Derived demand, 28  
 Diagnosis Related Groups (DRGs)  
   prospective payment and, 412–413  
   yardstick competition theory and, 413–417  
 Difference of means  
   distribution, variance of, and, 50–51  
   hypotheses and inferences and, 53–54  
   overview about, 49–50  
   standard error of mean and, 51–53  
   as statistical tool, 49–54  
 Diffusion of technology, 124–126  
 Direct-to-consumer (DTC) advertising,  
   316, 348  
 Disability-adjusted life-years (DALYs), 80  
 Discounting, 69–71  
 Discrimination  
   price, 355–356, 372  
   statistical, 202  
 Disease eradication difficulties, 537  
 Disparities, in health care, 189  
 Distribution, variance of, 50–51  
 Distributional adjustments, 72  
 Doctor-assisted suicide, 448  
 DOD. *See* Department of Defense  
 Donations, 267–268  
 Doughnut hole, 442, 443  
 DRGs. *See* Diagnosis Related Groups  
 Drug(s)  
   copayment, 164  
   DTC advertising and, 316  
   FDA and, 400  
   formularies, 363  
   insurance, 442–444  
   legalization, 527–528  
   new, 363–364  
   war on, 527–528  
 DTC. *See* Direct-to-consumer advertising  
 Dummy variables, 59–60  
 Dumping, 248–249
- E**  
 Economic efficiency. *See also* Cost-benefit  
   analysis  
   demand curve and, 64  
   efficient quantity and, 65–66  
   overview about, 63–64  
   supply curve and, 64–65  
 Economies of scale  
   described, 112  
   empirical cost-function studies and,  
     114–116  
   value of, 113  
 Economies of scope  
   described, 112–113  
   value of, 113  
 Education  
   control of, 335–336  
   demand and, 190  
   equilibrium and, 139  
   foreign graduates, 334–335  
   join production and, 333–334  
   labor issues and, 332–336  
   medical school revenue sources, 332  
   subsidy justification in, 332–333

- Efficiency. *See also* Economic efficiency  
 allocative inefficiency and, 117–118  
 competitive equilibrium and, 370  
 competitive markets and, 368–373  
 equity and, 372–373  
 for-profit/nonprofit firms, 280–281  
 frontier analysis and, 118–119  
 hospital, 119–121  
 hospital quality and, 120  
 Pareto, 368–369  
 social health insurance and, 375–376  
 technical inefficiency and, 116–117
- Efficient quantity, 65–66
- Elasticities, 177  
 across countries, 186–187  
 demand, 184–188  
 estimating, 57  
 income, 185–187  
 insurance, 187–188  
 as microeconomics tool, 33–35  
 prevalence, 535  
 price, 184–185  
 of substitution, 107–109
- Electromagnetic fields (EMFs), 53–54
- Elzinga and Hogarty criterion, 430–431
- EMFs. *See* Electromagnetic fields
- Employer-employee mandate, 493
- Employer mandate, 493, 495–496
- Employer provision  
 compensating differentials and, 217–218  
 history of, 212–213  
 of insurance, 212–219  
 labor market and, 213–215, 219–221  
 labor supply and, 219–221  
 other impacts of, 218–219  
 spousal coverage and, 215  
 tax system and, 215–217
- Employer sponsored managed care, 240–242
- Endowment, redistribution of, 371–372
- Environmental pollution, 101
- Epidemics  
 economic consequences of, 535–537  
 government role battling, 538–539  
 rational, 534–535
- Epidemiology  
 concepts from, 531–534  
 disease eradication difficulties, 537  
 epidemics economic consequences, 535–537  
 information and, 537  
 SIR model and, 533–534
- Epidemiology, economic  
 HIV/AIDS case study, 539–546  
 prevention demand prevalence elasticity and, 535  
 rational epidemics and, 534–535
- Equilibrium, 25  
 age and, 137–138  
 changes in, 137–139  
 competitive, 370  
 consumer, 32–33, 175–176  
 education and, 139  
 HMOs and, 249–251  
 wage rate and, 138–139
- Equity, 16  
 efficiency, 372–373  
 horizontal, 382–384
- Ethnicity, 189–190
- Exchanges, 506
- Excise taxes  
 bads and, 523–528  
 consumption reducing effects of, 524–527  
 illegal drugs/prohibition and, 527–528  
 influences on smoking, 526–527  
 in practice, 524–527  
 in theory, 524
- Exclusions, 150
- Exercise, 99–100, 131
- Expected loss, 155–156
- Expected value, 150–152
- Expenditure shares model, 487–489
- Experience rating, 201
- Externalities, 266–267, 394
- Externality model, of HMO, 247
- Extra-welfarism, 79, 379–381
- F**
- Factor substitution, labor and, 322, 324–325
- Fair innings proposal, 381
- FDA. *See* Food and Drug Administration
- Fee-for-service (FFS), 237  
 HMOs and, 248–249  
 managed care vs., 239
- FFS. *See* Fee-for-service
- Food and Drug Administration (FDA), 400
- For-profit firms  
 China and, 475  
 conversion into, 279–280  
 efficiency, 280–281  
 hospitals, 120, 276–280
- Free trade, 501
- Frontier analysis, 118–119
- Functions and curves  
 demand functions, 27–28  
 derived demand, 28  
 linear functions, 27  
 overview, 26–27
- G**
- Game theory, 288
- GDP. *See* Gross Domestic Product
- Gender  
 demand and, 189–190  
 physician income and, 340–341
- Generic entry, 357
- Generic substitutes, 362–363
- Germany, 485–486
- Global warming  
 cookstoves and, 69  
 discounting and, 69
- Goddeeris's model, 231–232
- Government  
 epidemics and, 538–539  
 initiatives of Chinese, 474  
 role of, 16
- Government intervention  
 externalities and, 394  
 failure, 402–404  
 forms of, 395–398  
 in health markets, 398–402  
 monopoly power and, 391–392  
 public goods and, 392–394  
 rationale for, 390–395
- Government regulation  
 antitrust and, 426–431  
 competitive strategies vs., 498  
 of hospital sector, 410–412  
 objectives of, 409–410  
 of physician payment, 423–425  
 prospective payment system and, 412–413, 417–423
- Gross Domestic Product (GDP)  
 by country, 469  
 health care's share of U.S., 3–5  
 NHE and, 8
- Grossman's model  
 health production and, 139–145  
 rational addiction and, 140
- H**
- Hazard rate, 531
- Health  
 demand for, 129–131  
 investment/consumption aspects of, 133–135  
 measuring, 94
- Health capital  
 consumer as health producer and, 129–130  
 cost of, 135–136  
 demand for, 136–137  
 exercise and, 131  
 obesity and, 141–145  
 time and, 130–131
- Health care. *See also* Comparative health care systems; Contemporary health care systems  
 agency and, 202–203  
 competition restriction and, 15  
 disparities in, 189  
 economics relevance to, 12–13  
 government role and, 16  
 insurance prominence and, 13–14  
 need and equity and, 16  
 NHE for, 8  
 nonprofit firms role in, 15  
 RHIE and, 97–98  
 social insurance and, 451–454  
 uncertainty and, 13  
 unique features of, 13–16  
 U.S. GDP and, 3–5  
 value of, 95
- Health economics  
 defined, 2  
 future regarding, 17  
 health care as unique, 13–16  
 relevance of, 3–9, 12–13
- Health economists  
 role of, 2
- Health economy  
 access and, 9  
 GDP and, 3–5  
 inflation and, 8–9  
 labor and capital in, 6–7  
 NHE and, 8  
 other countries and, 4–5  
 other health issues and, 9  
 personal spending and, 5–6  
 size and scope of, 3  
 time and, 7–8
- Health Insurance Portability and Accountability Act (HIPAA), 399

- Health maintenance organizations (HMOs), 241  
 comparative utilization and costs and, 252  
 empirical results on, 251–254  
 equilibrium and adverse selection and, 249–251  
 experience rating and, 201  
 externality model of, 247  
 FFS and, 248–249  
 methodological study issues and, 251  
 modeling individual, 246  
 practices of, 254  
 prediction framework illuminating, 248  
 RAND study on, 252  
 recent study evidence regarding, 252–254  
 types of care and, 247–248
- Health Savings Account (HSA), 499–500
- Heart care treatment, 75–76
- Hill-Burton Act, 399–400
- HIPAA. *See* Health Insurance Portability and Accountability Act
- HIV. *See* Human immunodeficiency virus
- HMOs. *See* Health maintenance organizations
- Home health, 298–299
- Horizontal equity, 382–384
- Hormone replacement therapy (HRT), 59
- Hospice, 298–299
- Hospitals  
 closing inefficient, 121  
 closures, mergers, and restructuring, 291–293  
 competition and, 257–258, 287–291  
 costs, 286–291  
 cost shifting and, 289–291  
 efficiency, 119–121  
 government regulation of, 410–412  
 history of, 284–285  
 managed care and, 243–244, 257–258  
 MAR and, 288  
 modeling, realism in, 278  
 nonprofit behavior models, 270–280  
 nonprofit compared with for-profit, 120, 276–279  
 nonprofit conversion into for-profit, 279–280  
 nonprofit efficiency and, 280–281  
 organization of, 285–286  
 overview, 283–284  
 as physicians' cooperative, 274  
 profit-deviating nonprofit, 272–274  
 prospective payment system and, 420–421  
 quality, 120, 293  
 regulation and accreditation, 286, 410–412  
 substitution and, 109  
 as two firms, 278–279
- HRT. *See* Hormone replacement therapy
- HSA. *See* Health Savings Account
- Human immunodeficiency virus (HIV). *See* *also* Auto-immune deficiency syndrome  
 case study, 539–546  
 as pandemic, 539, 542  
 statistics on, 539, 541, 542
- Human life, valuing  
 CBA and, 72–76  
 contingent valuation, 73  
 heart care treatment and, 75–76  
 last year of life, 73–75  
 VSL and, 74–75  
 willingness to pay and accept and, 73
- Hypothesis  
 difference of means and, 53–54  
 physician practice style, 312–313  
 target income, 306–308  
 testing, 49
- I**
- Illegal drugs, excise taxes and, 527–528
- Imperfect information, 199
- Imperfectly rational addiction models, 516
- Incidence rate, 531
- Income, 101  
 elasticities, 185–187  
 leisure *vs.*, 132–133  
 transfer effects of insurance, 168–170
- Indifference curves, 30–31
- Individual and market demands, 33
- Individual mandate, 439, 493, 505  
 employer *vs.*, 495–496
- Inferences  
 difference of means and, 53–54  
 in sciences and social sciences, 60
- Inflation, 8–9  
 marginal analysis and, 72  
 Medicaid and Medicare and, 455–456  
 technological change and, 122–124, 233
- Informal care, 298–299
- Information. *See also* Agency; Asymmetric information; Consumer information issues, 196  
 problems of, 14–15, 196–197
- Informed buyers, 204
- Innovation  
 pharmaceutical industry, 360–361
- Insecticide-treated bed net (ITN), 161
- Insurance. *See also* Coinsurance; Social insurance  
 adverse selection and, 200–201  
 aggregate expenditures and, 189  
 caesareans and, 218–219  
 CHIP, 450–451, 453–454, 462  
 demand for, 153–157, 166  
 elasticities, 187–188  
 expected loss changes and, 155–156  
 experience rating and, 201  
 explained, 148–149  
 future of Canadian private, 482  
 HIPAA and, 399  
 income transfer effects of, 168–170  
 labor market and, 213–215  
 lemons principle applied to, 199–201  
 mandated, 400  
 mobility and, 219–221  
 moral hazard and, 159–162  
 personal spending and, 14  
 pharmaceutical industry and, 352–353  
 premium changes and, 155  
 price of care and, 166  
 prominence of, 13–14  
 purchasing, 152–153  
 resource allocation and, 162–168  
 retirement and, 219  
 risk and, 150–153  
 secondary, 164–165  
 separation from employment, 496–497  
 single payer *vs.* multiple insurers, 497–498  
 substitution and, 352–353  
 supply of, 157–159  
 tax system and, 215–217  
 terminology, 149–150  
 wealth changes and, 156–157  
 welfare loss of excess, 167–168
- Insurance markets  
 compensating differentials and, 217–218  
 employer provision and, 212–219  
 labor market and, 213–215  
 labor supply and employer-based, 219–221  
 loading costs and, 210–212  
 managed care competition and, 258–259  
 past 30 years in, 223–224  
 practices in, 222–223  
 for private insurance, 221–222  
 tax system and, 215–217  
 technological change impacting, 230–233  
 uninsured and, 212, 224–230
- Isocost curves, 38–39
- Isoquants, 320–321
- ITN. *See* Insecticide-treated bed net
- J**
- Join production, 333–334
- Junk food tax, 397
- K**
- Keiki Care, 454
- L**
- Labor  
 demand of, 319–324, 328–329  
 education and, 332–336  
 employer provision and, 213–215, 219–221  
 factor substitution and, 322, 324–325  
 insurance and, 213–215  
 licensure and monopoly rents, 336–339  
 manpower availability, 325–332  
 marginal productivity of, 321–322  
 monopsony and, 330–332  
 overview of, 6–7  
 physician assistants, 324–325  
 physician issues of, 339–341  
 physician productivity measurement, 324  
 production functions and isoquants and, 320–321  
 shortages, 327–332  
 substitution among factors and, 324–325  
 supply of, 322–323
- Last year of life, 73–75
- Legalization, of drugs, 527–528
- Leisure  
 income *vs.*, 132–133  
 traded for wages, 132
- Lemons principle  
 asymmetric information and, 198–201  
 insurance and, 199–201  
 used-car market and, 198–199
- Liberalism, classical and modern, 386–387
- Licensure and monopoly rents  
 overview about, 336–339  
 public and self-interest theories, 337–338  
 quality and, 338–339
- Life saving, 67–68
- Limitations, 150
- Linear functions, 27
- Linearity *vs.* assumptions, 79–80
- Loading costs  
 impacts of, 211–212

- insurance firms and, 210–212  
uninsured and, 212
- Loading fees, 150
- Long-run average cost (LRAC), 40
- Long-term care (LTC)  
financing, 297–298  
hospice, home health, and informal care,  
298–299  
Medicare and Medicaid and, 449–450
- LRAC. *See* Long-run average cost
- LTC. *See* Long-term care
- M**
- Makena, 345
- Malpractice, 315
- Managed care. *See also* Health maintenance organizations  
background regarding, 236–237  
backlash, 259–262  
competitive effects, 256–259  
costs and, 255–256  
development and growth of, 244–246  
economic characteristics of, 238–239  
economics of, 245–246  
employer sponsored, 240–242  
features of, 238  
federal policy and, 244–245  
FFS vs., 239  
hospital contracts, 243–244  
hospital market competition, 257–258  
insurance market competition, 258–259  
Medicaid and Medicare, 242  
modeling, 246–251  
organizational structure within, 237–238  
physician contracts, 242–243  
plans emergence, 239–244  
P4P and, 261  
spending growth and, 255–256  
technological change and, 259  
technology diffusion and, 126  
theoretical competition issues, 256–257
- Mandated coverage, 229–230
- MAR. *See* Medical arms race
- Marginal analysis, 10  
in CBA, 68–69  
discounting and, 70–71  
distributional adjustments and, 72  
inflation and, 72  
risk adjustment and, 71–72
- Marginal efficiency of investment (MEI)  
decreasing, 136–137  
rate of return and, 136
- Marginal productivity, of labor, 321–322
- Marginal utility, of wealth, 151–152
- Marin County, 430
- Market failure  
externalities and, 266–267  
public goods and, 267–268
- Mean, standard error of, 51–53
- Measuring costs, 67
- Medicaid  
child population and, 458  
cost sharing impact, 448  
criticisms of, 462–463  
cuts in, 463  
doctor-assisted suicide and, 448  
effects of, 454–459  
eligibility, 446–448  
enrollment by group, 446  
establishment, 438–439  
health status improvement, 458–459  
improvement of, 452  
indirect impacts, 457  
inflation and, 455–456  
as managed care, 242  
Medicare relationship, 449–450  
overview, 445  
payment for, 449  
services scope and duration, 448
- Medical arms race (MAR), 288
- Medical Savings Account (MSA), 499
- Medicare  
beneficiary payment liabilities, 444  
case mix severity, 418–420  
criticisms of, 462–463  
effects of, 454–461  
enrollment numbers, 441  
establishment, 438–439  
expenditures, 460–461  
future prospects, 459–461  
health status improvement, 458–459  
indirect impacts, 457  
inflation and, 455–456  
length of stay and, 417–418  
as managed care, 242  
Medicaid relationship, 449–450  
overview, 440  
Part A, 440  
Part B, 440–441  
Part C, 441  
Part D, 442–444  
program financing, 444  
prospective payment system and, 417–423  
provider payments, 445  
recent changes, 459
- Medicine  
history and, 93  
mortality rate decline and, 89–92  
production of health and, 88–93  
rising population and, 88–89
- MEI. *See* Marginal efficiency of investment
- MenAfriVac, 533
- Mexican nursing homes, 192
- Microeconomics tools  
consumer theory, 28–33  
elasticities, 33–35  
functions and curves, 26–28  
individual and market demands, 33  
monopoly and, 43–45  
overview, 20  
perfect competition supply curve, 41–43  
production and market supply, 35–39  
scarcity and PPF, 21–23  
supply and demand analysis, 23–26
- Models, use of, 10
- Monopoly  
antitrust and, 427–429  
government intervention and, 391–392  
microeconomics tools and, 43–45  
pricing, 354–355, 356–357  
rents, 336–339
- Monopsony, 330–332
- Moral hazard  
coinsurance/deductibles effects and, 162  
demand for care and, 159–161  
insurance and, 159–162
- Mortality rate  
clean water and, 92  
decline, 89–92  
nutrition reduced, 91  
prenatal care and, 96–97  
public health reduced, 91–92
- MSA. *See* Medical Savings Account
- Myopic addiction models, 516–517
- N**
- National health expenditures (NHE), 8
- National Health Service (NHS)  
overview about, 469–470  
performance and reforms under, 471–473  
private care vs., 470–471  
United Kingdom's, 469–473  
woman excluded from, 471
- Need, 16  
CON, 411–412  
Daniels' concept of, 381  
distribution criticisms, 382  
distributions from, 376–382  
horizontal equity and, 382–384  
self-reported unmet, 484  
Netherlands, 485–486
- Net revenue per physician (NR/M), 274–275
- NHE. *See* National health expenditures
- NHS. *See* National Health Service
- Nonprofit firms  
behavior models, 270–280  
China and, 475  
as contract failure and, 269–270  
conversion into for-profit, 279–280  
donations and, 267–268  
efficiency vs. for-profits, 280–281  
financial matters and, 270  
for-profit hospitals and, 120, 276–280  
health care role of, 15  
market failure: externalities and, 266–267  
market failure: public goods and, 267–268  
NR/M maximization and, 274–275  
overview about, 265–266  
physicians' cooperative and, 274,  
275–276, 277  
prevalence of, reasons for, 266–270  
profit-deviating, 272–274  
quality-quantity theory, 271–272,  
275–276, 277  
relevance of, 268–269  
unmet demands and, 266–267
- NR/M. *See* Net revenue per physician
- Nurses, monopsony and, 330–332
- Nursing homes  
background about, 294  
costs, 287, 294  
excess demand and, 296–297  
Mexican, 192  
nonprofit efficiency and, 280–281  
quality of care, 295–296
- Nutrition reduced mortality, 91
- O**
- Obesity  
causes of increasing, 144–145  
economic effects, 143–144  
economic treatment of, 142–143  
health capital and, 141–145
- OLS. *See* Ordinary least squares

- Opportunity cost, 21  
 Ordinary least squares (OLS), 55–56  
 Oregon doctor-assisted suicide, 448  
 Output maximization, 39–40
- P**
- Pandemic, 539, 542  
 Pareto efficiency, 368–369  
 Patents, 345  
 Patient Protection and Affordable Care Act (PPACA), 1, 439, 446  
   administrative ease and, 508  
   cost containment and, 507–508  
   estimated fiscal effects of, 509  
   exchanges and, 506  
   key areas of focus, 505  
   patient/physician choice and, 508  
   predictions about, 508–510  
   provisions of, 505–506  
   safety net and, 507  
 Pay-for-performance (P4P), 261  
 Penicillin, 95  
 Perfect competition, 373–374  
   supply curve, 41–43  
 Performance-based budgeting, 121  
 Personal spending, 5–6  
   insurance and, 14  
 Pharmaceutical industry  
   competition, 346–347, 357  
   copayment and, 361–362  
   cost containment, 361–363  
   DTC advertising, 348  
   entry barriers in, 347–348  
   innovation, 360–361  
   insurance and, 352–353  
   least-cost-production and, 351–352  
   new drugs and, 363–364  
   overview, 344–346  
   patents and media attention and, 345  
   pricing and, 354–357  
   profits and, 354  
   R&D, 357–361  
   regulation of, 349–350  
   structure of, 346–348  
   substitutability and, 350–357  
   technological change and, 353–354  
 Physician assistants, efficient utilization of, 324–325  
 Physicians  
   availability of, 326–327  
   Canada and, 478–479  
   income by gender, 340–341  
   labor issues, 339–341  
   managed care and, 242–243  
   NR/M and, 274–275  
   PPACA and, 508  
   productivity measurement of, 324  
   recent trends regarding, 328  
   regulation of payment to, 423–425  
   specialization and, 339–340  
 Physicians' cooperative, 274, 275–276, 277  
 Physician's practice  
   agency and, 305–306  
   clinical decision making and, 316  
   DTC drug advertising, 316  
   financial incentive and, 305  
   issues and policy puzzles, 315–317  
   malpractice and, 315  
   model of, 303–305  
   overview, 302  
   paying for outcomes regarding, 316–317  
   SAV and, 311–315  
   style hypothesis and, 312–313  
   supplier-induced demand and, 305–311  
 Point of service (POS), 241  
 Pollution, environmental, 101  
 Population  
   health, 94–95  
   rising, 88–89  
 POS. *See* Point of service  
 P4P. *See* Pay-for-performance  
 PPACA. *See* Patient Protection and Affordable Care Act  
 PPF. *See* Production possibilities frontier  
 PPOs. *See* Preferred provider organizations  
 Pre-existing conditions, 150  
 Preferred provider organizations (PPOs), 241  
 Pregnancies, 310  
 Premium, 149  
 Premium, changes in, 155  
 Prenatal care, 96–97  
 Prescription drug insurance, Medicare, 442–444  
 Prevalence elasticity, 535  
 Prevalence rate, 531  
 Prevention demand prevalence elasticity, 535  
 Preventive care, 505  
 Price  
   consumer information and, 203–204  
   discrimination, 355–356, 372  
   dispersion, 204  
   elasticities, 184–185  
   insurance and, 166  
   monopoly pricing, 354–355, 356–357  
   relevance of, 12–13  
 Production  
   of health and home goods, 134–135  
   of healthy days, 133–134  
 Production and market supply  
   cost curves, marginal and average, 40  
   cost minimization or output maximization, 39–40  
   isocost curves and, 38–39  
   as microeconomics tools, 35–39  
 Production function  
   examples of, 36–38  
   explained, 35–36  
   and labor, 320–321  
   recent studies regarding, 326  
   substitution and, 38  
   total and marginal contributions and, 85–87  
 Production of health. *See also* Substitution  
   Grossman's model and, 139–145  
   issues, 93–94  
   medicine and, 88–93  
   modern day, 93–97  
   other health measures and, 97–101  
   overview about, 85  
   population health and, 94–95  
   prenatal care and, 96–97  
   production function and, 85–87  
   schooling role in, 101–102  
 Production possibilities frontier (PPF), 21–23  
 Profit-deviating nonprofit hospital, 272–274  
 Prohibition, excise taxes and, 527–528  
 Prospective payment system  
   cost effects of, 421–423  
   DRGs and, 412–413  
   government regulation, 412–413, 417–423  
   hospital's financial condition from, 420–421  
   Medicare's, 417–423  
 Public goods, 267–268  
 Public health reduced mortality, 91–92  
 Public provision, 396–398  
 Pure premiums, 150
- Q**
- QALY. *See* Quality-adjusted life-year  
 Quackery, 521  
 Quality  
   assurance, 238  
   consumer information and, 204–207  
   costs and, 9  
   hospitals, 120, 293  
   of licensure, 338–339  
   nursing homes, 295–296  
   other indicators of, 206–207  
   patient perceptions of, 205  
   reform and, 504–505  
 Quality-adjusted life-year (QALY)  
   ACE inhibitor and, 78  
   ageism and, 80  
   extra-welfarism and, 79  
   linearity vs. assumptions and, 79–80  
   overview about, 77–78  
   Sen's capabilities and, 79  
   standard welfare economics and, 78–79  
   technological change and, 123–124  
 Quality-quantity nonprofit theory, 271–272, 275–276, 277
- R**
- RAND Health Insurance Experiment (RHIE), 97–98  
 RAND HMO study, 252  
 Rate of return, 136  
 Rational addiction, 140, 517–518  
 Rational decision making, 10  
 Rational epidemics, 534–535  
 Rawlsian justice, 385–386  
 R&D. *See* Research and Development  
 Reform, health system  
   background about, 492  
   basic issues in, 493–494  
   competitive strategies, 498–502  
   employer vs. individual mandates, 495–496  
   ensuring access to care, 495–498  
   goals of, 492–495  
   insurance separation from employment, 496–497  
   international competitiveness, 502–504  
   under NHS, 471–473  
   other market, 500–501  
   PPACA and, 505–510  
   preventive care and, 505  
   quality of care, 504–505  
   single payer vs. multiple insurers, 497–498  
   universal coverage costs, 494–495

- Regression analysis  
 demand regression and, 56–57  
 dummy variables, 59–60  
 multiple, 58–60, 313  
 OLS and, 55–56  
 overview about, 54–55  
 regression coefficients interpretation, 58–59  
 SAV and, 313  
 as statistical tool, 54–57
- Regulation. *See* Government regulation
- Reinforcement, 517
- Relative rates of return, 330
- Report cards, 206–207
- Reputation goods, 203–204
- Research and Development (R&D), 357–361
- Resource allocation, 162–168
- Retail value scales (RVs), 424–425
- Retirement, 219
- Revascularization, 2
- RHIE. *See* RAND Health Insurance Experiment
- Risk, 67–68  
 adjustment, 71–72  
 aversion, 151–152  
 insurance and, 150–153
- RVs. *See* Retail value scales
- S**
- Safety net, 492–493, 507
- SAV. *See* Small area variations
- Scarcity  
 defined, 22  
 PPF and, 21–23
- Schooling, role of, 101–102
- Selective contracting, 238–239
- Sen's capabilities, 79
- SFA. *See* Stochastic frontier analysis
- Shortages  
 dynamic, 329–330  
 excess demand and, 328–329  
 labor, 327–332  
 monopsony and, 330–332  
 relative rates of return and, 330
- Single payer, *vs.* multiple insurers, 497–498
- SIR. *See* Susceptible-Infective-Removed
- Skimping, 248, 249
- Sleep, 99–100
- Small area variations (SAV)  
 contributions to, 311–312  
 diffusion of information and, 311–315  
 inappropriate utilization cost and, 313–314  
 multiple regression and, 313  
 other applications of, 315  
 physician practice style hypothesis and, 312–313
- Smoking. *See also* Cigarettes  
 across countries, 515  
 excise taxes influence on, 526–527
- Social capital, 100
- Social insurance. *See also* Medicaid; Medicare  
 CHIP, 450–451  
 criticisms of, 462–463  
 early U.S., 438  
 efficiency rationale for, 375–376  
 European beginnings, 437–438  
 health and, 451–454  
 historical roots, 437–439  
 policies and programs, 435–437  
 PPACA, 439  
*vs.* private, 149
- Social justice theories, 385–387
- Social welfare. *See* Welfare
- South Africa, 545
- Specialization, 339–340
- Spending  
 across countries, 468, 487–489  
 managed care, 255–256  
 personal, 5–6, 14
- Spousal coverage, 215
- Standard error  
 of estimate, 56  
 of mean, 51–53
- Statistical discrimination, 202
- Statistical tools  
 difference of means, 49–54  
 elasticities, estimating, 57  
 hypothesis testing, 49  
 inference, 60  
 multiple regression analysis, 58–60  
 overview about, 48–49  
 regression analysis, 54–57
- Steady state, 518
- Steering, 238
- Stochastic frontier analysis (SFA), 119
- Substitution  
 care professionals and, 108  
 degrees of, 107  
 elasticity of, 107–109  
 example explaining, 106–107  
 hospital care estimates of, 109  
 insurance and, 352–353  
 pharmaceutical industry and, 350–357  
 production function and, 38
- Suicide, doctor-assisted, 448
- Supplier-induced demand  
 benchmark model and, 308–309  
 fees and, 309–311  
 inducement/marketing parallel and, 309  
 modeling, 306–311  
 physician agency and, 305–306  
 pregnancies and, 310  
 target income hypothesis and, 306–308
- Supply  
 of insurance, 157–159  
 of labor, 322–323
- Supply and demand  
 analysis, 23–26  
 comparative statics and, 25–26  
 demand curve and shifters, 23–24  
 equilibrium and, 25  
 law applicability, 407–408  
 supply curve and shifters, 24–25
- Supply curve  
 economic efficiency and, 64–65  
 perfect competition, 41–43  
 shifters and, 24–25
- Susceptible-Infective-Removed (SIR), 533–534
- Switzerland, 485–486
- T**
- Target income hypothesis, 306–308
- Taxes, 215–217, 395–396, 397, 401
- Technical inefficiency, 116–117
- Technological change  
 cost and, 3, 121–124, 231  
 diffusion of, 124–126  
 Goddeeris's model and, 231–232  
 increasing/decreasing cost and, 121–122  
 inflation and, 122–124, 233  
 insurance markets and, 230–233  
 managed care and, 126, 259  
 pharmaceutical industry and, 353–354  
 QALYs and, 123–124
- Technology, availability by country, 480
- Thalidomide, 349
- The Magic Mountain* (Mann), 90
- Theorem of the Second Best, 374–375
- Time, 7–8  
 demand shifters and, 178–180  
 health capital and, 130–131  
 investment over, 135–136  
 trading leisure for wages, 132
- Tolerance, 517
- Tourism, medical, 501
- Transfer programs, 398
- Tuberculosis, 90
- U**
- UCR. *See* Usual, customary, and reasonable
- Uncertainty, 13, 14–15
- Uninsured  
 counting, 224  
 firm size and, 227  
 impact of, 227  
 loading costs and, 212  
 mandated coverage impacts and, 229–230  
 myths, 226  
 overview about, 224  
 statistics on, 224–226  
 working, 227–229
- United Kingdom, 467  
 GDP and, 469  
 NHS of, 469–473  
 resources, 469  
 spending in, 468  
 technology availability in, 480
- United States (U.S.)  
 Canada compared with, 477–478, 481–484  
 GDP and, 3–5  
 self-reported unmet need in, 484  
 social insurance history in, 438  
 technology availability in, 480
- Universal coverage, 494–495
- Used-car market, 198–199
- Usual, customary, and reasonable (UCR), 423
- Utilitarianism, 385
- Utility, 29–30. *See also* Cost-utility analysis; Marginal utility
- Utility-possibility frontier, 377–378
- Utilization review, 238



**V**

VA. *See* Department of Veterans Affairs  
Vaccination, 532–534, 538  
Value of statistical life (VSL), 74–75  
Valuing human life. *See* Human life,  
valuing  
Variables, dummy, 59–60  
VSL. *See* Value of statistical life

**W**

Wage rate, 138–139  
War on Drugs, 527–528  
Water, clean, 92  
Wealth changes, 156–157  
Welfare. *See also* Need  
extra-welfarism, 79, 379–381  
First Fundamental Theorem of, 371

insurance and, 167–168  
QALY and, 78–79  
Second Fundamental Theorem of, 372  
Willingness to pay and accept, 73

**Y**

Yardstick competition theory,  
413–417