U.S. Medical Care: Crisis or Conundrum

The concern over the future of health care revolves around three broad issues: quality, access, and affordability. As private health insurance coverage gradually declines, and the number of uninsured steadily rises, emerging public consensus indicates a system in need of reform. Gaps in coverage, combined with the upward trend in medical care spending over the past several decades, add to the commonly held belief that the United States has a health care crisis. What exactly is the nature of this health care crisis? Is it a crisis in spending? Or is it the way we pay for medical care? Many are concerned over access to care for the uninsured and the prospects for continued access for those who currently have insurance. An additional concern is whether quality may deteriorate as medical care is increasingly provided in a managed care environment.

Recent polls indicate that most insured Americans are satisfied with the quality of their own health insurance plan. In fact, over 80 percent of Americans said that the medical care that they or their family received during the past year was good to excellent. Access, on the other hand, is a real concern. About half are concerned that if they become seriously ill, they will not be able to afford the medical care they need (Donelan et al., 1999).

Those without insurance coverage who lack the resources to pay out-of-pocket must rely on public assistance and private charity for the care they receive. Even those with health insurance lack the assurance of continued coverage. Because most workers receive their health insurance as an employee benefit, losing a job can mean losing access to medical care. Thus, it is not a contradiction for a survey respondent to be satisfied with the health care received personally, and at the same time believe that the system is flawed and needs to be rebuilt.

The “experts” have a completely different perspective of the health care crisis, including the problem, its cause, and preferred solutions (Blendon et al., 1993). They see the health care crisis as one of rising aggregate spending and the government’s inability to sustain the expansion of its two major programs, Medicare and Medicaid. Experts themselves are sharply divided on the cause of the crisis. Some think the problem is the unrestrained use of medical technology. Others believe the culprits are the increased use of health insurance and tax subsidies that encourage individuals to overinsure. Proposed solutions address the perceived causes. If technology is being overused, restrict patient access to expensive procedures. If subsidized insurance is the problem, limit the subsidy.

Although the uninsured receive fewer services and less coordinated care, they do have access to high-quality medical care through public clinics and hospital...
emergency departments. Traditionally, indigent care has been financed directly by taxpayers and private charities and indirectly by shifting costs to those with insurance coverage. In 2001, for example, estimates of this cost shifting were as high as $34.5 billion, amounting to approximately $1,000 for every uninsured person (Hadley and Holahan, 2003). This kind of availability may be undesirable and inefficient, but it is access. When surveyed, 25 percent of the uninsured stated that they delayed care in the past year due to cost, but only half that number stated that they had an unmet medical need because of the cost of care (Kaiser, 2007). Is it accurate to refer to the circumstances surrounding the delivery and financing of health care in the United States as a crisis, or is it merely a conundrum—a problem that the wealthiest country in the world ought to address?

**HISTORICAL DEVELOPMENTS IN THE DELIVERY OF MEDICAL CARE**

No matter where a health care discussion begins, the topic of conversation soon turns to the issue of affordability. Employees and employers complain about high premiums, patients and providers note high treatment costs, and policymakers lament high and rising spending. Each perspective presents a different aspect of the same problem. In 2007, the average cost of a health insurance policy was $12,106 for a family and $4,479 for an individual (Kaiser, 2007). The average cost per hospital stay was over $8,000, and Americans spent almost $2 trillion on health care—16 percent of the gross domestic product (GDP).

From 1971 to 2006, the annual growth in nominal health care spending ranged between 4.4 percent and 15.6 percent, increasing at an annual compound rate of almost 10 percent for that 35-year period. Over that same period, the Consumer Price Index (CPI), a popular measure of the rate of inflation, increased an average of 4.7 percent per year. With nominal health care spending increasing at twice the overall rate of inflation, real spending (adjusted for inflation) grew approximately 5 percent per year. Figure 1.1 depicts the relative growth rates in nominal and real spending from 1971 to 2006.

Many commentators are encouraged by the dramatic slowing of the real rate of growth in health care spending during the 1990s, falling from 6.1 percent in

---

**cost shifting**
The practice of charging higher prices to one group of patients, usually those with health insurance, in order to provide free care to the uninsured or discounted care to those served by Medicare and Medicaid.

**premium**
A periodic payment required to purchase an insurance policy.

**gross domestic product (GDP)**
The monetary value of the goods and services produced in a country during a given time period, usually a year.

**POLICY ISSUE**
How many years does it take to constitute a trend?

Since 1985, the Emergency Medical Treatment and Active Labor Act (EMTALA) has made it illegal for hospital emergency departments to deny care to anyone requesting care. Turning away patients because of lack of health insurance is not an option.
1990 to 2.1 percent in 1996. History, however, warns against using such short trends as tools for policymaking. The 1972 to 1974 time period saw real growth rates fall from 8.2 percent to 2.1 percent, only to rise again to 8.6 percent by 1976. Beginning in that year, real growth rates started falling again, to less than 2 percent by 1980. This was followed by a steady upward march until 1990. By 2002, growth in real spending had reached 7.3 percent. Since then, real rates have fallen back to around 3 percent.

The major concern over health care spending is not that it is high; the concern is that the steady upward spiral does not seem to have an end to it. Government projections predict that medical care spending will rise to $3.4 trillion by 2016, almost 20 percent of GDP (Center for Medicare and Medicaid Services, CMS, 2007). Although economic theory has yet to determine what the optimal percentage ought to be, the United States spends more on medical care by virtually every measure than any other country in the world. If the optimal percentage is not known, what does it mean to spend 8 percent or 10 percent or 16 percent of a country’s GDP on medical care? And, more importantly, should the amount spent on medical care be a concern to policy makers?

Postwar Experience

Medical care spending in the United States over the post-World War II period is summarized in Table 1.1. The four summary measures provide evidence that medical care spending is high and growing. During the decade of the 1950s, total spending increased at a rate of 8 percent per year. Total spending at the beginning

<table>
<thead>
<tr>
<th>Year</th>
<th>Total Spending (in billions)</th>
<th>Percent Change(^1)</th>
<th>Percent of GDP</th>
<th>Per Capita Spending</th>
</tr>
</thead>
<tbody>
<tr>
<td>1950</td>
<td>$12.7</td>
<td>—</td>
<td>4.5</td>
<td>$82</td>
</tr>
<tr>
<td>1960</td>
<td>27.5</td>
<td>8.0</td>
<td>5.2</td>
<td>148</td>
</tr>
<tr>
<td>1970</td>
<td>74.9</td>
<td>10.5</td>
<td>7.2</td>
<td>356</td>
</tr>
<tr>
<td>1975</td>
<td>133.1</td>
<td>12.2</td>
<td>8.1</td>
<td>605</td>
</tr>
<tr>
<td>1980</td>
<td>253.9</td>
<td>13.7</td>
<td>9.1</td>
<td>1,102</td>
</tr>
<tr>
<td>1985</td>
<td>439.9</td>
<td>11.6</td>
<td>10.4</td>
<td>1,820</td>
</tr>
<tr>
<td>1990</td>
<td>714.0</td>
<td>10.2</td>
<td>12.3</td>
<td>2,813</td>
</tr>
<tr>
<td>1995</td>
<td>1,016.5</td>
<td>7.3</td>
<td>13.6</td>
<td>3,783</td>
</tr>
<tr>
<td>2000</td>
<td>1,353.6</td>
<td>5.9</td>
<td>13.8</td>
<td>4,790</td>
</tr>
<tr>
<td>2001</td>
<td>1,469.6</td>
<td>8.6</td>
<td>14.5</td>
<td>5,148</td>
</tr>
<tr>
<td>2002</td>
<td>1,603.4</td>
<td>9.1</td>
<td>15.3</td>
<td>5,559</td>
</tr>
<tr>
<td>2003</td>
<td>1,732.4</td>
<td>8.0</td>
<td>15.8</td>
<td>5,952</td>
</tr>
<tr>
<td>2004</td>
<td>1,852.3</td>
<td>6.9</td>
<td>15.9</td>
<td>6,301</td>
</tr>
<tr>
<td>2005</td>
<td>1,973.3</td>
<td>6.5</td>
<td>15.9</td>
<td>6,649</td>
</tr>
<tr>
<td>2006</td>
<td>2,105.5</td>
<td>6.7</td>
<td>16.0</td>
<td>7,026</td>
</tr>
</tbody>
</table>


\(^1\) Annual rate of change from the previous year listed.

POLICY ISSUE

What is the optimal percentage of GDP that a country should spend on health care? Is a continuously growing percentage affordable?
of the decade was $12.7 billion, doubling by its end. Medical care spending as a percent of GDP increased from 4.5 to 5.2 percent, and per capita medical care spending increased from $82 in 1950 to $148 ten years later.

The 1960s was the first of three decades characterized by rapid growth in medical care spending. The annual compound rate of growth in medical care spending was 11.5 percent between 1960 and 1990. At the beginning of that 30-year period, medical care spending was $27.5 billion, 5.2 percent of GDP, and $148 per capita. By 1990, it stood at $714.0 billion, 12.3 percent of GDP, and $2,738 per capita. Contributing factors included increased federal government involvement in the payment for medical care services for specific groups—Medicare for the elderly and Medicaid for the indigent—and cost shifting by providers to subsidize care for those without insurance.

Rapid advancement in medical technology and the subsequent cost-containment strategies that emphasized regulation and planning characterized the 1970s. The federal government became a major force in biomedical research and development with the expansion of the National Institutes of Health. Technological advances that included open-heart surgery, organ transplantation, various types of imaging, and the ability to preserve and prolong life in the intensive care unit increased public awareness of medicine and served as a major cost driver. While it all seemed justifiable, this emphasis on advanced technologies precipitated a growing concern over cost issues.

Federal legislation, specifically the National Health Planning Act of 1974, created a network of government planning agencies to control medical care costs. In addition, states passed certificate-of-need (CON) laws to limit the growth in hospital investment in capital improvements and technology. Even a brief national experiment with wage and price controls during the Nixon presidency did little to curb the growth in medical care costs and spending.

Possibly the most significant piece of legislation affecting health care was not viewed as particularly significant at the time. The Employee Retirement Income Security Act (ERISA) of 1974 was passed to regulate the corporate use of pension funds. One provision of the act exempted self-insured health plans from state-level health insurance regulations. The passage of ERISA provided an incentive for employers to switch to self-insurance. Today, more than half of all workers who participate in group health insurance plans are employed by companies who self-insure.

The 1980s ushered in a change in direction in health care policy, resulting in a shift away from regulation and planning and toward a greater reliance on market forces. A president who wanted to lower taxes and a Congress that refused to cut spending characterized the era. Federal budget deficits grew dramatically. By the end of the decade, those areas of the budget in which spending was mandated—the entitlement programs including Medicare and Medicaid—grew seemingly without limit and came under intense pressure to reduce their rate of growth. During this period, the introduction of alternative payment schemes and delivery systems was significant. Prospective payment, capitation, the use of diagnosis-related groups to pay hospitals, and the introduction of a relative-value scale to pay physicians were all examples of these changes. Health maintenance organizations, preferred provider organizations, and other systems of managed care became more common.

The 1990s saw a moderation in the growth in spending. Most experts attribute at least part of the slowdown to the movement of patients into managed care. The annual percentage increase in nominal spending fell from 15.9 percent in 1981 to around 5 percent in the mid 1990s. A steady increase in growth rates resulted in an
annual change of 9.1 percent in 2002, settling to 6.7 percent in 2006. The expansion of medical care spending as a percentage of GDP remained between 13.0 and 14.0 percent until 2001, when it nudged above 14 percent for the first time.

The federal government has taken more of an activist role in health care policy in the past decade. Although an attempt to completely restructure the health care system failed in 1994, important legislation has been enacted that is expected to improve access to care.\(^2\)

**The Question of High and Rising Spending**

Why do Americans spend so much on medical care? Various researchers have explored this issue, and several explanations have been offered. The reason cited most often for the rise in spending include the expansion of the third-party payment system, an aging population, the rise in medical malpractice litigation, and the increased use of medical technology (Aaron, 1991).

Insurance use has increased dramatically over the past four decades. Out-of-pocket spending has fallen from about half of total spending to 12.5 percent. At the same time, per capita spending on health care has increased more than 48 times. Common sense tells us that the demand for medical care increases as the cost to the individual declines. In all societies, the elderly demand more medical care than the nonelderly. As the U.S. population ages, more medical care is demanded. This trend will likely continue as the baby-boom generation reaches retirement age beginning in 2010.

Medical malpractice premiums represent costs that are passed on to patients in the form of higher fees. Those costs alone reached $29.4 billion in 2005 (Tillinghast-Towers Perrin, 2006). Additionally, fear of malpractice litigation changes the way physicians carry out the practice of medicine, resulting in the increased use of *defensive medicine*, practices and procedures whose primary purpose is to reduce the risk of a lawsuit. Estimates of the cost of this practice range from $60 billion to $100 billion per year (American College of Obstetricians and Gynecologists, 2003). The use of expensive technology has become an imperative in most medical practices. If withheld, even in situations where the probability of success is extremely low, the physician will be accused of negligence and likely sued for malpractice.

Many other factors contribute to the increased spending. The imbalance of information between provider and patient is seen by some as a major factor contributing to the increase in spending. The patient is typically uninformed and finds it difficult to get informed. Because the medical practitioner serves as both adviser and provider of care, he has the potential to recommend procedures and services with little expected benefit to the patient—a practice known as *physician-induced demand*.

Others have focused on certain aspects of the U.S. institutional setting as a major reason for the rise in expenditures. In particular, some see the restrictions on competition due to licensing requirements as important. The 125 medical schools in the United States enroll approximately 67,000 students. Not one new school has been accredited since 1985. In contrast, the original 15 countries that comprise the European Union have a combined population that is approximately

---

\(^2\) At the federal level, Congress established the Health Insurance Portability and Accountability Act (HIPPA) of 1996 providing insurance portability to individuals with health insurance. In 1997, Congress passed the Children's Health Insurance Program (CHIP), the largest expansion of a federal medical program since its original enactment. In late 2003, Congress voted to expand the coverage for outpatient prescription drugs within the Medicare program.
1.5 times that of the United States. Their 220 medical schools enroll almost 40,000 new students each year, or 2.3 times the 17,000 new students enrolled in U.S. medical schools annually (Curtoni and Sutnick, 1995). Furthermore, supply constraints are exacerbated by the increased specialization of U.S. physicians. The ratio of specialists to generalists is 70:30, roughly the reverse of that in most other developed countries.

### ISSUES IN MEDICAL CARE DELIVERY

**Spending Somebody Else’s Money**

A *Wall Street Journal* article provides an interesting example of how spending someone else’s money distorts the decision-making process. A 70-year-old man suffering from a ruptured abdominal aortic aneurysm was brought to the hospital. After several weeks in the intensive care unit — with all the modern technology that goes with it — and a three-month stay in the hospital, the bill approached $275,000, none of which would be paid out-of-pocket by the patient. The man’s physician determined that his poor eating habits, caused by poorly fitting dentures, were contributing to his slow recovery. He requested that the hospital dentist perform the necessary adjustments. Later, the doctor discovered that the man had not allowed the dentist to adjust the dentures. When asked the reason, the man replied, “$75 is a lot of money.” It seems that Medicare would not pay for the adjustment, so it would have been an out-of-pocket expenditure for the patient.

When you’re spending somebody else’s money, $275,000 does not seem like a lot. But when you are spending your own money, $75 is a lot. Our reliance on a third-party payment system is the major institutional feature that contributes to rising costs and increased spending. Cost-conscious consumers have little or no role in a system dominated by third-party payers.


Economic theory suggests that an increase in the number of physicians would decrease the prices they charge for their services. For many years, economists believed that the American Medical Association (AMA), through its control over medical schools via the accreditation process, controlled the number of licensed physicians and thus kept prices high (Friedman and Kuznets, 1945). The evidence presented in Table 1.2 makes it difficult to present a strong case that restrictions in supply have caused prices to increase. Over the past 20 years, the number of active physicians has increased by two-thirds, but the population has increased less than 20 percent. As a result, the physician–population ratio has actually increased by over 38 percent, from 18.3 physicians per 10,000 population in 1980 to 26.9 physicians per 10,000 population in 2005. Comparable numbers in the European Union place the number of practicing physicians at 1.3 million, or 33.3 per 10,000 population.

---

3 The 15 countries in the European Union include Austria, Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Luxembourg, The Netherlands, Portugal, Spain, Sweden, and the United Kingdom.
The predominance of not-for-profit providers has also been cited as a major factor in the spending increase (Goodman and Musgrave, 1992). Physicians are trained in not-for-profit medical schools, they work in predominantly not-for-profit hospitals, and the majority of their fees are paid by not-for-profit insurance companies (e.g., Blue Cross and Blue Shield) or the government. This reliance on not-for-profit institutions creates a dynamic environment in which innovation in general is encouraged, but where cost-saving innovations are seldom encountered.

Changes in Medical Care Delivery
The last 30 years have witnessed major changes that have affected medical care delivery and costs. The shift from private to public sector financing, the shift from out-of-pocket spending to third-party payment, the changes in hospital usage and pricing, deregulation, and the growth in managed care have all had profound effects on medical care delivery and pricing.

Shift From Private to Public Financing
Quite possibly, the single most important change affecting medical care delivery has been the shift from private to public sector financing. Referring to Table 1.3, the private sector was responsible for $3 of every $4 spent in the industry in 1960. The government role in financing was modest, standing at less than 25 cents out of every medical care dollar. The introduction of Medicare and Medicaid in the mid-1960s resulted in an increase in the government’s share of spending, to almost 40 percent by the end of the decade. Even though the government’s total share has remained at about half of total spending, the federal share has nearly tripled, from 10.6 percent in 1960 to 33.5 percent in 2006. This translates into a federal budgetary obligation that has grown from $2.9 billion to $704.9 billion in four and a half decades. Even as the federal share has exploded, the share of state and local governments has remained relatively stable at around 13 percent.

Shift to Third-Party Payment
Even as the private share of total spending has fallen, the role of private insurance has expanded. Private insurance paid a little more than 20 percent of the total cost of medical care in 1960, with that share rising

Table 1.2: Physicians and Physician–Population Ratio United States, Various Years

<table>
<thead>
<tr>
<th>Year</th>
<th>Number of active physicians</th>
<th>Active physicians per 10,000 population</th>
</tr>
</thead>
<tbody>
<tr>
<td>1975</td>
<td>340,280</td>
<td>—</td>
</tr>
<tr>
<td>1980</td>
<td>414,916</td>
<td>18.3</td>
</tr>
<tr>
<td>1985</td>
<td>497,140</td>
<td>20.1</td>
</tr>
<tr>
<td>1990</td>
<td>547,310</td>
<td>22.0</td>
</tr>
<tr>
<td>1995</td>
<td>625,443</td>
<td>23.8</td>
</tr>
<tr>
<td>2000</td>
<td>692,368</td>
<td>25.2</td>
</tr>
<tr>
<td>2001</td>
<td>713,375</td>
<td>25.0</td>
</tr>
<tr>
<td>2002</td>
<td>719,431</td>
<td>25.0</td>
</tr>
<tr>
<td>2003</td>
<td>736,211</td>
<td>25.3</td>
</tr>
<tr>
<td>2004</td>
<td>744,143</td>
<td>25.3</td>
</tr>
<tr>
<td>2005</td>
<td>762,438</td>
<td>26.9</td>
</tr>
</tbody>
</table>

Table 1.3: Financing of Health Care Expenditures, Various Years
In Billions of Dollars and Percentage of Total Personal Spending

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$</td>
<td>%</td>
<td>$</td>
<td>%</td>
<td>$</td>
<td>%</td>
<td>$</td>
<td>%</td>
</tr>
<tr>
<td>Private Funds</td>
<td>20.7</td>
<td>75.3</td>
<td>46.8</td>
<td>62.5</td>
<td>147.6</td>
<td>58.1</td>
<td>427.3</td>
<td>59.9</td>
</tr>
<tr>
<td>Out-of-Pocket Payments</td>
<td>12.9</td>
<td>46.9</td>
<td>24.9</td>
<td>33.2</td>
<td>58.6</td>
<td>23.1</td>
<td>136.1</td>
<td>19.1</td>
</tr>
<tr>
<td>Private Health Insurance</td>
<td>5.9</td>
<td>21.5</td>
<td>15.5</td>
<td>20.7</td>
<td>68.9</td>
<td>27.1</td>
<td>233.7</td>
<td>32.7</td>
</tr>
<tr>
<td>Other Private</td>
<td>2.0</td>
<td>7.3</td>
<td>6.4</td>
<td>8.5</td>
<td>20.1</td>
<td>7.9</td>
<td>57.5</td>
<td>8.0</td>
</tr>
<tr>
<td>Public Funds</td>
<td>6.8</td>
<td>24.7</td>
<td>28.1</td>
<td>37.5</td>
<td>106.3</td>
<td>41.9</td>
<td>286.7</td>
<td>40.1</td>
</tr>
<tr>
<td>Federal</td>
<td>2.9</td>
<td>10.6</td>
<td>17.7</td>
<td>23.6</td>
<td>71.6</td>
<td>28.2</td>
<td>193.9</td>
<td>27.2</td>
</tr>
<tr>
<td>State and Local</td>
<td>3.9</td>
<td>14.2</td>
<td>10.4</td>
<td>13.9</td>
<td>34.8</td>
<td>13.7</td>
<td>92.4</td>
<td>13.0</td>
</tr>
<tr>
<td>Medicare¹</td>
<td>—</td>
<td>—</td>
<td>7.7</td>
<td>10.3</td>
<td>37.2</td>
<td>14.7</td>
<td>102.5</td>
<td>15.3</td>
</tr>
<tr>
<td>Medicaid and SCHIP²</td>
<td>—</td>
<td>—</td>
<td>2.8</td>
<td>3.7</td>
<td>26.0</td>
<td>10.2</td>
<td>73.7</td>
<td>10.3</td>
</tr>
<tr>
<td>Total Health Care Spending</td>
<td>27.5</td>
<td>100.0</td>
<td>74.9</td>
<td>100.0</td>
<td>253.9</td>
<td>100.0</td>
<td>714</td>
<td>100.0</td>
</tr>
</tbody>
</table>


¹ Included in federal spending.
² Included in federal, state, and local spending.
³ Columns do not add to total due to double-counting for Medicare, Medicaid, and SCHIP.
to about one-third by 1990, where it has remained since that time. The major change in private spending has been the dramatic decline in private, out-of-pocket spending. Approximately half of total health care expenditures were classified as out-of-pocket spending in 1960. By 2006, that total had fallen to 12.2 percent. With the increased importance of third-party payers such as government and private insurers, the insured patient has relatively little out-of-pocket spending at the point of purchase.

Payment by third parties provides little incentive on the part of provider or patient to control spending. As long as insurance companies are willing to pay the bills, physicians will continue to provide all the care that patients request. Patients have no incentive to limit their utilization. Even when the expected benefit of a procedure is small, in most cases it will be demanded, because the patient’s share of the cost is small.

It should come as no surprise that the cost of services covered by insurance—public and private—has risen at a faster rate than the cost of services that are not covered. Why? When consumers purchase goods and services at discount prices, they tend to purchase more than if they paid the full price. What other reasonable explanation would explain the crowds that flock to clearance sales, and enthusiastic consumer acceptance of discount malls? Health economists refer to this phenomenon as **moral hazard**. Between 1970 and 2006, hospital spending for services usually covered by insurance increased 20 times over, whereas spending on eyeglasses—something typically not covered by insurance—increased only 10 times over. Insulating patients from the full cost of medical care has had the effect of making patients insensitive to the prices that are being charged, and at the same time has encouraged greater utilization.

**Change in Hospital Usage and Pricing**  Hospital usage has also changed dramatically. As seen in Table 1.4, almost every measure of inpatient hospital usage has fallen in the past 30 years, in some cases quite dramatically. The number of hospital beds is down, admissions are down, the average length of stay is down, and occupancy rates have fallen significantly. Some would go so far as to say that hospitals have gone from overcrowded to underused. Another important trend is the shift from inpatient to outpatient care. The number of per capita outpatient visits has tripled since 1970, and outpatient visits per hospital admission are also three times higher.

---

**Table 1.4: Short-Stay Community Hospital Characteristics, United States**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Beds (per 1,000 population)</td>
<td>4.17</td>
<td>4.38</td>
<td>3.73</td>
<td>3.32</td>
<td>2.92</td>
<td>2.79</td>
<td>2.71</td>
</tr>
<tr>
<td>Admissions (per 1,000 population)</td>
<td>144.0</td>
<td>159.6</td>
<td>125.4</td>
<td>117.9</td>
<td>117.1</td>
<td>119.4</td>
<td>118.9</td>
</tr>
<tr>
<td>Average length of stay (days)</td>
<td>7.7</td>
<td>7.6</td>
<td>7.2</td>
<td>6.5</td>
<td>5.8</td>
<td>5.7</td>
<td>5.6</td>
</tr>
<tr>
<td>Outpatient visits (per 1,000 population)</td>
<td>657.2</td>
<td>893.2</td>
<td>1,211.6</td>
<td>1,578.5</td>
<td>1,845.7</td>
<td>1,934.0</td>
<td>1,971.7</td>
</tr>
<tr>
<td>Outpatient visits per admission</td>
<td>4.6</td>
<td>5.6</td>
<td>9.7</td>
<td>13.4</td>
<td>15.8</td>
<td>16.2</td>
<td>16.6</td>
</tr>
<tr>
<td>Percent occupancy</td>
<td>78.0</td>
<td>75.6</td>
<td>66.8</td>
<td>62.8</td>
<td>63.9</td>
<td>66.2</td>
<td>67.3</td>
</tr>
</tbody>
</table>

*Source: Health United States, various years.*
Cost Plus was the standard approach for hospital pricing from the inception of Medicare until 1983, when pricing shifted to prospective payment using diagnosis-related groups (DRGs). Under DRG pricing, payment is fixed in advance and based on the principal diagnosis at the time of hospital admission. In contrast, private insurance pays hospitals negotiated prices based on discounts from billed charges. As a result, the financial risk of treating patients has shifted from the payer to the provider, creating an incentive for providers to limit access to care. Many providers are affiliated with networks of providers that offer discounts to group members. Because all must abide by the fee limits placed on them by Medicare and Medicaid, actual transaction prices are deeply discounted from the invoice prices that show up on their bills.

**Deregulation and the Growth in Managed Care**  
Deregulation has resulted in an explosion of facilities and practices previously considered unthinkable. The use of ambulatory surgery centers has risen, as has the construction of physician-owned clinics and hospitals. More physicians are advertising, more practices offer evening and weekend hours, and some physicians are even making house calls.

The managed care approach is the prevailing form of insurance in the U.S. market. By 1999, nine out of ten employees covered by employer-based group insurance were enrolled in a managed care plan (a health maintenance organization, a preferred provider organization, or a point-of-service plan). The rest were still in a traditional indemnity insurance plan. The increased popularity of managed care has begun to change the incentive structure within the industry, forcing providers to consider costs more carefully. No longer are physicians’ fees constrained by a pricing model that limits fees to usual, customary, and reasonable (UCR) levels.

In 1986, the federal government established a pricing model for Medicare based on a relative-value scale (RVS). The Medicare RVS is an index of resource use for every medical procedure across all specialty areas. It translates into a fee schedule by adjusting resource use by a monetary conversion factor. Most fees charged by physicians are in some way tied to this index.

Many physicians participate in at least one risk-sharing contract with a health plan, in which they receive payment under a capitation arrangement. **Capitation** is defined as a fixed fee, paid in advance, for all necessary care provided to a well-defined group. Providing care for a fixed fee changes the nature of the physician–patient relationship. With cost increasingly an issue, the provider has a stake in eliminating all unnecessary care, which increases the risk that potentially beneficial care will be denied in the name of cost savings.

## The Nature of Medical Care as a Commodity

Before undertaking the study of medical care using economics, it is important to understand the differences between medical care and other commodities. If medical care were just like any other commodity, the use of economics to explain pricing and allocation decisions would not be questioned. But if it is substantially different, strict reliance on economic models may lead to inaccurate predictions and, ultimately, to serious policy mistakes.

Just how different is medical care from other commodities? Using the pioneering work of Kenneth Arrow (1963) as a guide, we can identify a number of distinguishing characteristics that contribute to the uniqueness of medical care as a commodity. First, unlike other commodities, the demand for medical care is irregular. Except for the small percentage of care that may be defined as preventive,
medical care demand follows an accidental injury or the onset of an illness. As a result, medical care is commonly associated with discomfort, pain, and suffering. It may even be an issue of life or death, depending on the nature of the accident or illness. Thus, access to medical care often has implications on the patient’s ability to return to a state of normal functioning.

Second, the medical care transaction is characterized by information problems that disproportionately affect patients. All consumers are frequently confronted with difficulties in collecting information about a product, but the problem is particularly acute for medical care consumers due to the complexity of medical knowledge. The typical consumer of medical care is poorly informed and finds it difficult to become well informed. Because of this information imbalance, patients rely on their physicians to diagnose their illnesses and prescribe treatments, and they expect the physician to proceed without consideration for his or her own personal gain. Thus, the medical transaction carries with it ethical overtones unlike any other transaction. To protect the interests of the uninformed public, government has established licensing requirements and educational standards to ensure a minimum level of quality among providers, and provider organizations have adopted codes of conduct to guard against unethical behavior.

In addition to the information problem, the medical transaction is characterized by widespread uncertainty. An individual can rarely predict the onset of an illness and usually cannot predict his or her demand for medical care. Physicians are confronted with uncertainty in diagnosis and treatment. Any given medical condition can be taken care of using a number of different treatment alternatives. One physician may recommend surgery; another may take a wait-and-see attitude. Both decisions are based on the interpretation of diagnostic tests and the physician’s best judgment. And treatment is not always clearly linked to the outcome. Thus, medicine is an art as much as it is a science.

Another interesting feature of the market for medical care is the widespread reliance on not-for-profit providers, especially in the provision of hospital services. Because trust plays such a big role in the patient–provider relationship, restraining the profit motive may be desirable. The conventional wisdom would have us believe that the absence of the profit motive will mean decision making without the influence of self-interest on the part of providers. Even with over 85 percent of the nation’s hospitals either government owned or otherwise not-for-profit, the profit motive has not been totally eliminated from the medical care sector. Most physicians’ practices are for profit, as are virtually all pharmaceutical companies, retail drug stores, and long-term care facilities.

Although it is difficult to predict the onset of illness for any one individual, it is possible to predict the number of people who will suffer from a particular medical condition within a large group of individuals. In order to spread the risk of financial loss due to an illness, the individual is willing to purchase insurance. Because the probability of a loss is predictable for large groups, insurance companies emerge to underwrite that risk and sell insurance policies. As a result, insurance has become the primary means of payment for medical care. With third parties financing most of the costs of medical care, individuals are insulated from the full cost of the care they receive. Those with insurance will demand more medical care than equally healthy individuals who are uninsured. Providers will adjust treatment recommendations

---

**uncertainty**
A state in which multiple outcomes are possible, but the likelihood of any one outcome is not known.

**self-interest**
A behavioral assumption of neoclassical economics that individuals are motivated to promote their own interests.

---

Dr. Mike Magee’s “Health Commentary” offers weekly discussions on important health issues. Popular topics on his blog include Men’s and Women’s health, aging, and health care reform. Check it out at [http://healthcommentary.org](http://healthcommentary.org)

---

4 Notable exceptions to this observation include the treatment of certain chronic conditions that may be postponed with little risk and the provision of elective procedures, such as cosmetic surgery, pregnancy, and corrective orthopedics.
Health economics emerged as a subdiscipline of economics in the 1960s, with the publication of two important papers by Kenneth J. Arrow (1963) and Mark V. Pauly (1968), both published in the American Economic Review. Arrow’s paper is considered by many to be the seminal contribution to the field of health economics and health policy. Recognizing its importance, the Journal of Health Politics, Policy, and Law (Peterson, 2001) devoted a special issue to the paper’s important contributions, including a forward written by Pauly.

Health economists examine a wide range of issues, extending from the nature and production of health to the market for health and medical care to the microeconomic evaluation of health care interventions and strategies. Figure 1.2 provides a diagrammatic overview of the structure of health economics. Beginning with the box labeled “Nature of Health,” we can ask ourselves a number of questions: What does it mean to be healthy? How do we measure health? What is the best possible way to measure quality of life? Because of the nature of the questions being asked, research on this topic is interdisciplinary. Even though economists are not the only ones studying these questions, their contributions have been significant. The development of the quality of life measure, called the quality-adjusted life year (QALY), was in part a result of the participation of economists.

Grossman (1972) developed an economic framework for the study of medical care demand in which medical care is simply one of many factors used to produce good health. In this framework, “Production of Health” looks at the determinants of health, including income, wealth, education, genetics, and public health. Our ability to maintain a desired level of health depends to a great extent on the lifestyle choices we make. The topic “Confounding Factors” develops the influence of, for example, tobacco, alcohol, drugs, obesity, and sexually transmitted diseases on our ability to produce good health for a given level of medical care spending. The aging population and the introduction of new technology affect the ability of the market to allocate resources in such a way as to effectively satisfy consumer demand.

The principle activity of health economists outside the United States is microeconomic evaluation, or the evaluation of alternative ways to treat a specific medical condition. Policy makers within fixed-budget systems find it necessary to conduct studies comparing the costs and consequences of diagnosis and treatment options in order to make informed decisions on the optimal allocation of scarce resources. Cost-benefit analysis, with its welfare economics framework, provides the foundation for most of the research in economic evaluation, and health economists have adapted that framework in developing cost-effectiveness analysis, the evaluation method of choice in medical care decision making.

John M. Keynes, author of The General Theory of Employment, Interest, and Money, wrote that “practical men, who believe themselves to be quite exempt from any intellectual influences, are usually the slaves of some defunct economist. Madmen in authority, who hear voices in the air, are distilling their frenzy from some academic scribbler of a few years back.” The Dead Economists Society is dedicated to the preservation of the insights of classical liberal economists, such as Adam Smith, Friedrich Hayek, Ludwig von Mises, Bookers T. Washington, and Benjamin Franklin. Check it out at http://www.personal.psu.edu/faculty/j/d/djm114/oldindex.html

Familiarize yourself with economic concepts and issues by staying abreast of recent developments in the world of business. A popular business daily newspaper is the Wall Street Journal, also available in an online, interactive edition at http://online.wsj.com/public/us

Differences of opinion among economists have been a constant source of humor. Jokes about economists and economics that even Adam Smith would enjoy can be found at http://netec.wustl.edu/jok&c.html
The primary focus of U.S. health economists is the market for health care. The boxes in Figure 1.2 numbered 5 through 7, and the topics covered in them, summarize this emphasis. The “Demand for Health Care” is affected by the elements discussed in boxes 1 and 2, the nature and production of health. The early contribution of economics to the study of health care demand considered improving health to be one way to increase future productivity (Mushkin, 1962). Thus, the demand for health care is not only influenced by a desire to feel better when ill, it is also viewed as investment in human capital. Factors affecting the demand for medical care include the socioeconomic characteristics of the population, patient demographics, access barriers (including cost-sharing arrangements), and the role of providers in determining the type and level of care prescribed.

The “Supply of Health Care” encompasses a broad spectrum of economics on such topics as production theory, input markets, and industrial organization. Specific issues examined include the cost of production, input substitution, and the nature and role of incentives. Demand and supply interact with one another to establish “Market Equilibrium.” Markets are able to effectively allocate scarce resources where they are most productive by establishing a price for everything.

Analysis of the overall goals and objectives of the health care system is the subject of “Macroeconomic Evaluation.” How well is the system performing? Is it accessible? Is it affordable? Is quality at the desired level? It is here where
national and international comparisons are made. How does our system compare to those of our neighbors? Finally, “Health Policy and Planning” involves the interaction of private sector, government, and nongovernmental organizations (NGOs) in setting national goals, determining the strategies for reaching those goals, and establishing the rules of the game that regulate how medical care markets work.

Health care systems are constantly changing. Policymakers and planners are always looking for better ways to produce, deliver, and pay for a growing menu of medical care services demanded by an insatiable public. The goal of this book is to provide you with the tools to better understand the role of economics in this important task.

| TEN KEY ECONOMIC CONCEPTS |

Given the complexity of economic theory, it may come as a surprise that economic thought is guided by a relatively small number of key concepts. These concepts will serve as unifying themes throughout the book.

1. **Scarcity and choice** address the problem of limited resources and the need to economize. Not enough resources are available to meet all the desires of all the people, making rationing in some form unavoidable. We are forced to make choices among competing objectives—an inescapable result of scarcity.

2. **Opportunity cost** recognizes that everything and everyone has alternatives. Time and resources used to satisfy one set of desires cannot be used to satisfy another set. The cost of any decision or action is measured in terms of the value placed on the opportunity foregone.

3. **Marginal analysis** is the economic way of thinking about the optimal allocation of resources. Choices are seldom made on an all-or-nothing basis—they are made at “the margin.” Decision makers weigh the trade-offs, a little more of one thing and a little less of another. In this environment, consideration is given to the incremental benefits and incremental costs of a decision.

4. **Self-interest** is the primary motivator of economic decision makers. Driven by the power of self-interest, people are motivated to pursue efficiency in the production and consumption decisions they make. According to the well-known eighteenth-century economist Adam Smith, this pursuit of self-interest, moderated by market competition, causes each individual to pursue a course of action that promotes the general goals of society.

5. **Markets and pricing** serve as the most efficient way to allocate scarce resources. The market accomplishes its tasks through a system of prices, what Adam Smith called the “invisible hand.” The invisible hand can allocate resources because everyone and everything has a price. Prices increase if more is desired and decrease if less is desired. Firms base their production decisions on relative prices and relative price movements. The price mechanism becomes a way to bring a firm’s output decisions into balance with consumer desires—something that we refer to as equilibrium.

6. **Supply and demand** serve as the foundation for all economic analysis. Pricing and output decisions are based on the forces underlying these two economic concepts. Goods and services are allocated among competing uses by striking
a balance, or attaining an equilibrium, between consumers’ willingness to pay and suppliers’ willingness to provide. This is rationing via prices.

7. Competition forces resource owners to use their resources to promote the highest possible satisfaction of society, including consumers, producers, and investors. If resource owners do this well, they are rewarded. If they are inept or inefficient, they are penalized. Competition takes production out of the hands of the less competent and places it into the hands of the more efficient to constantly promote more efficient methods of production.

8. Efficiency in economics measures how well resources are being used to promote social welfare. Inefficient outcomes waste resources, but the efficient use of scarce resources enhances social welfare. The fascinating aspect of competitive markets is how the more-or-less independent behavior on the part of thousands of decision makers serves to promote social welfare. Consumers attempt to make themselves better off by allocating limited budgets. Producers seek maximum profits by using cost-minimizing methods.

9. Market failure arises when the free market fails to promote the efficient use of resources by either producing more or less than the optimal level of output. Sources of market failure include natural monopoly, externalities in production and consumption, and public goods. Other market imperfections, such as incomplete information and immobile resources, also contribute to this problem.

10. Comparative advantage explains how people benefit from voluntary exchange when production decisions are based on opportunity cost. The individual or entity that has the lowest opportunity cost of production is said to have a comparative advantage.

SUMMARY AND CONCLUSIONS

The medical care industry in the United States is large and growing in relative size. Medical care is one of the largest industries in the vast U.S. economy. At more than $2.1 trillion, it was five times larger than the domestic auto industry and four times larger than the total defense budget in 2005. In addition, medical care employed more people and exported more goods and services than either defense or automobiles. It may be difficult to imagine, but the economic output of the U.S. medical care industry was almost 20 percent larger than the entire French economy.

As shown in Figure 1.3, a potpourri of public and private sources finances U.S. medical care. The public sector directly finances 46.1 percent of total spending. Private health insurance and private philanthropy finance 41.7 percent, leaving 12.2 percent to come from direct, out-of-pocket payments from individuals.

Most of the money Americans spend on medical care covers either hospital or physicians’ services (see Figure 1.4). The percentage of total spending in these two areas has remained at around 50 to 55 percent. Other professional services, pharmaceuticals, and nursing home care combine for approximately one-fourth of the total spending. The other five percent comprises home health care and other medical products and services. Even though it represents only 10.3 percent of total spending, pharmaceutical spending is the fastest growing portion of expenditures with an 80 percent increase since 2000.

The U.S. system of medical care delivery is far from perfect. Its weaknesses are easily identified. Critics claim there are too few primary care physicians and too many specialists, leading to greater reliance on acute and specialty care and underutilization of primary and preventive care. The gaps in health insurance coverage limit reliable access for many low-skilled workers and their families. Only recently has federal legislation introduced a modest measure of portability in the market for group health.
insurance. Even with changes in the law, many people are still considered uninsurable because of preexisting conditions.

The system also has its strengths, and its defenders argue that quality is unquestionably high. Citing evidence from polls, they note that around 85 percent of Americans are happy with the quality of their own medical care arrangements. It should be noted that the same polls show that one-third feel the system has so much wrong with it that it needs to be completely rebuilt (Donelon et al., 1999). The U.S. system has progressed much faster than its European counterparts in developing quality assessment and output measures. The United States is still the world leader in innovation, research, and the development of state-of-the-art technology.

The growth in medical care spending has moderated somewhat since 1990. It could be that the aggressive action by employers and state governments to reverse the escalation in spending is finally paying off, or possibly that the threat of government intervention at the federal level has served to intimidate providers, who now fear public backlash and political reprisals. Whatever the reason, spending growth has moderated without significant legislative action.

In general, spending growth in the public sector has outpaced spending growth in the private sector. Since
QUESTIONS AND PROBLEMS

1. Thomas Sowell, a senior fellow at the Hoover Institution, has stated that we “have difficulty understanding the strange way words are used by politicians and the media.” We often think of a crisis in terms of an emergency, a situation of utmost urgency, maybe even life or death. According to Sowell, politicians use the term differently. They define a crisis as any situation they want to change. How do you define the term crisis? Does the United States have a health care crisis?

2. Discuss the magnitude of the financing problem in medical care. What are the major reasons that medical spending is absorbing an increasing share of national output?

3. How important is cost containment in establishing a national health care policy? In addition to controlling costs, what are the alternative goals for a national medical care system?

4. What do economists mean by scarcity? Why is the concept so important in economic analysis?

REFERENCES


OASDI and Medicare Boards of Trustees, 1999 *Annual Reports of the OASDI and Medicare Boards of Trustees to Congress*, based on data from the Office of the Actuary, Health Care Financing Administration, 1999.


The Medical Care Price Index

The conventional wisdom in many policy circles embraces the notion that medical care inflation is out of control. How much of the increase in medical spending is due to inflation, and how much is due to improved services and changing demographic patterns? The way we answer this question will ultimately determine the type of medical care reform we will get. It is important, therefore, to understand how price indexes are used to measure medical care price inflation.

**Measuring Price Changes with Index Numbers**

The principal measure of inflation used by business and government policy makers is the year-to-year change in the consumer price index (CPI). The index plays an important role in determining cost of living adjustments (COLAs) for everything from union wages to social security and pension benefits to federal income tax brackets. The CPI is a fixed-weight or Laspeyres index that measures price changes for a market basket of items defined for a base time period. In other indexes, such as the GDP price deflator, the composition of the market basket changes every year to reflect different spending patterns.

A fixed-weight index has become the index of choice used to measure inflation. Because the weights do not change, movements in a fixed-weight index are due solely to changes in the prices of the goods included in the market basket. In contrast, a movement in a deflator reflects changes in prices of goods and the composition of the market basket. In reality, consumers adjust their spending away from goods whose prices increase, making it necessary to change the composition of the fixed-weight market basket periodically to better reflect consumer spending patterns. The weights for the CPI are based on a survey of consumer spending patterns and are changed approximately every ten years. The current CPI weighting scheme was revised in 1987 based on results from the 1982–1984 Consumer Expenditure Survey.

Table A1.1 presents data for the consumer price index from 1970 through 2007. Overall, the index is broken down into seven major spending categories: food (18 percent), housing (42 percent), apparel (6 percent), transportation (18 percent), medical care (6 percent), entertainment (4 percent), and other (6 percent). The index in each case equals 100 for the 1982 to 1984 time period. When interpreting these indexes, note that the inflation rate from one time period to the next can be calculated by dividing the change in the index by its previous value. For example, the CPI changed from 144.5 to 148.2 between 1993 and 1994. This change of 3.7 percentage points divided by 144.5 results in an estimated annual inflation rate of 2.56 percent.

Of the time period shown, the medical care component increased at a faster rate than any other component of the CPI—over 700 percent from 1970 to 2002.

**Medical Care Price Index**

The major index of medical care prices, the Medical Care Price Index (MCPI), is shown in Table A1.2. Medical care is divided into commodities and services. Medical commodities are subdivided into seven categories: prescription drugs, nonprescription drugs, first aid and dressings, general medical equipment, convalescent equipment, hearing aids, and unpriced items. Medical services are divided into nine categories: physician, dental, optometry, other professional, hospital room, other inpatient, outpatient, nursing home, and unpriced. Health insurance is priced using a separate category.

Typically cited as the measure of medical care inflation, the MCPI has steadily increased since 1950. Interpreting the index as a measure of inflation suggests that medical care prices have risen at a compounded rate of over 6.35 percent since 1980, over two-thirds faster than prices in general. If this is true, we have a real problem on our hands. But can we believe what the statistics seem to tell us? Is the MCPI a good measure of medical care price inflation?
### Table A1.1: Consumer Price Indexes for Major Expenditure Classes
Select Years, 1960 to 2007 (1982 to 1984 = 100)

<table>
<thead>
<tr>
<th>Year</th>
<th>All Items (CPI-U)</th>
<th>All Services</th>
<th>Food</th>
<th>Housing</th>
<th>Apparel</th>
<th>Energy</th>
<th>Medical Care</th>
<th>All Items Excluding Medical Care</th>
</tr>
</thead>
<tbody>
<tr>
<td>1960</td>
<td>29.6</td>
<td>24.1</td>
<td>30.0</td>
<td>—</td>
<td>45.7</td>
<td>22.4</td>
<td>22.3</td>
<td>30.2</td>
</tr>
<tr>
<td>1970</td>
<td>38.8</td>
<td>35.0</td>
<td>39.2</td>
<td>36.4</td>
<td>59.2</td>
<td>25.5</td>
<td>34.0</td>
<td>39.2</td>
</tr>
<tr>
<td>1980</td>
<td>82.4</td>
<td>77.9</td>
<td>86.8</td>
<td>81.1</td>
<td>90.9</td>
<td>86.0</td>
<td>74.9</td>
<td>82.8</td>
</tr>
<tr>
<td>1990</td>
<td>130.7</td>
<td>139.2</td>
<td>132.4</td>
<td>128.5</td>
<td>124.1</td>
<td>102.1</td>
<td>162.8</td>
<td>128.8</td>
</tr>
<tr>
<td>1995</td>
<td>152.4</td>
<td>168.7</td>
<td>148.4</td>
<td>148.5</td>
<td>132.0</td>
<td>105.2</td>
<td>220.5</td>
<td>148.6</td>
</tr>
<tr>
<td>2000</td>
<td>172.2</td>
<td>195.3</td>
<td>167.8</td>
<td>169.6</td>
<td>129.6</td>
<td>124.6</td>
<td>260.8</td>
<td>167.3</td>
</tr>
<tr>
<td>2001</td>
<td>177.1</td>
<td>203.4</td>
<td>173.1</td>
<td>176.4</td>
<td>127.3</td>
<td>129.3</td>
<td>272.8</td>
<td>171.9</td>
</tr>
<tr>
<td>2002</td>
<td>179.9</td>
<td>209.8</td>
<td>176.2</td>
<td>180.3</td>
<td>124.0</td>
<td>121.7</td>
<td>285.6</td>
<td>174.3</td>
</tr>
<tr>
<td>2003</td>
<td>184.0</td>
<td>216.5</td>
<td>180.0</td>
<td>184.8</td>
<td>120.9</td>
<td>136.5</td>
<td>295.1</td>
<td>178.1</td>
</tr>
<tr>
<td>2004</td>
<td>188.9</td>
<td>222.8</td>
<td>186.2</td>
<td>189.5</td>
<td>120.4</td>
<td>151.4</td>
<td>310.1</td>
<td>182.7</td>
</tr>
<tr>
<td>2005</td>
<td>195.3</td>
<td>230.1</td>
<td>190.7</td>
<td>195.7</td>
<td>119.5</td>
<td>177.1</td>
<td>323.2</td>
<td>188.7</td>
</tr>
<tr>
<td>2006</td>
<td>201.6</td>
<td>238.9</td>
<td>195.2</td>
<td>203.2</td>
<td>119.5</td>
<td>196.9</td>
<td>336.2</td>
<td>194.7</td>
</tr>
<tr>
<td>2007</td>
<td>207.3</td>
<td>246.8</td>
<td>202.9</td>
<td>209.6</td>
<td>119.0</td>
<td>207.7</td>
<td>351.1</td>
<td>200.1</td>
</tr>
</tbody>
</table>

Source: *Health United States, various years*.

### Table A1.2: The Medical Care Price Index and Its Major Components
Select Years 1950 to 2005 (1982 to 1984 = 100)

<table>
<thead>
<tr>
<th>Year</th>
<th>Total Medical Care</th>
<th>Compound Rate of Change from Previous Year Listed</th>
<th>Medical Care Commodities</th>
<th>Medical Care Services</th>
</tr>
</thead>
<tbody>
<tr>
<td>1950</td>
<td>15.1</td>
<td>—</td>
<td>39.7</td>
<td>12.8</td>
</tr>
<tr>
<td>1960</td>
<td>22.3</td>
<td>4.0</td>
<td>46.9</td>
<td>19.5</td>
</tr>
<tr>
<td>1970</td>
<td>34.0</td>
<td>4.3</td>
<td>46.5</td>
<td>32.3</td>
</tr>
<tr>
<td>1980</td>
<td>47.5</td>
<td>8.2</td>
<td>75.4</td>
<td>74.8</td>
</tr>
<tr>
<td>1990</td>
<td>162.8</td>
<td>7.4</td>
<td>163.4</td>
<td>162.7</td>
</tr>
<tr>
<td>1995</td>
<td>220.5</td>
<td>6.3</td>
<td>204.5</td>
<td>224.2</td>
</tr>
<tr>
<td>2000</td>
<td>260.8</td>
<td>3.4</td>
<td>238.1</td>
<td>266.0</td>
</tr>
<tr>
<td>2001</td>
<td>272.8</td>
<td>4.6</td>
<td>247.6</td>
<td>278.8</td>
</tr>
<tr>
<td>2002</td>
<td>285.6</td>
<td>4.7</td>
<td>256.4</td>
<td>292.9</td>
</tr>
<tr>
<td>2003</td>
<td>297.1</td>
<td>4.0</td>
<td>262.8</td>
<td>306.0</td>
</tr>
<tr>
<td>2004</td>
<td>310.1</td>
<td>4.4</td>
<td>269.3</td>
<td>321.3</td>
</tr>
<tr>
<td>2005</td>
<td>323.2</td>
<td>4.2</td>
<td>276.0</td>
<td>336.7</td>
</tr>
<tr>
<td>2006</td>
<td>336.2</td>
<td>4.0</td>
<td>285.9</td>
<td>350.6</td>
</tr>
<tr>
<td>2007</td>
<td>351.1</td>
<td>4.4</td>
<td>290.0</td>
<td>369.3</td>
</tr>
</tbody>
</table>

Source: *Health United States, various years*. 
PROBLEMS WITH USING A FIXED-WEIGHT INDEX AS A MEASURE OF INFLATION

In reality, changes in a fixed-weight index do not accurately reflect changes in the cost of living. Using a fixed-weight index, such as the MCPI, to measure medical care price inflation introduces a substantial upward bias to the estimate. It is important that we understand the problems associated with using indexes to measure inflation and take appropriate steps when interpreting indexes to minimize the bias.

Measuring Inputs Instead of Outcomes

The MCPI measures the wrong thing. The price index measures the cost of inputs: an office visit, a day in the hospital, a surgical procedure, or a prescription drug. Patients who are ill do not desire the inputs; they are interested in the restoration of their health. But, as we will see in Chapter 5, health is difficult to define, let alone measure.

Given the difficulty in measuring health, one possible solution would be to measure the cost of curing a particular illness. For example, the average length of stay in the hospital has steadily fallen over the course of the past several decades. Reduced stays have dampened the hospital-cost escalation measured in terms of average cost per day (what the CPI measures). Even more dramatic has been the increased use of outpatient procedures to treat illnesses that formerly required extensive hospital stays. Repair of an inguinal hernia, one of the most common surgical procedures, formerly required several days in the hospital and several months of limited activity. Today the procedure is performed on an outpatient basis and requires only a few hours in the surgicenter and minimal rehabilitation time. In fact, most patients are encouraged to resume their normal daily activities as soon as possible.

The shift to outpatient surgery has greatly reduced the cost of treating many common problems, but the cost savings has largely been lost on the MCPI. As outpatient procedures grow in popularity, two things happen: First, patients who continue to be treated in the hospital are, on average, sicker than before. They require more resources on average and thus drive up the average cost of their hospital stays. Second, when an outpatient procedure actually replaces a conventional hospital procedure, as is the case with cataract surgery and lens replacement and many orthopedic surgeries, it drops out of the hospital component of the price index and is picked up later in the outpatient component. The end result of both of these factors is an increase in the MCPI, even though the cost of treating the illness has decreased.

Measuring Quality Changes

Technological progress typically results in improvements in the products and services available to consumers. Price increases due to quality improvements are mistakenly identified as inflation in a fixed-weight index. This is not a severe problem in industries in which innovation takes place slowly, but technological progress takes place at different rates in different industries. This is especially true in the medical industry, in which quality of care has improved dramatically over the past 50 years. Treatments for once-untreatable diseases offer new hope. Inexpensive prevention of diseases such as polio and smallpox has led to near eradication of these once-costly illnesses, and improved surgical techniques allow patients to leave the hospital sooner and recover faster.

If price indexes are to be an accurate measure of changes in the cost of living, price changes due to quality improvements must have no impact on the value of the index. The Bureau of Labor Statistics (BLS) attempts to factor in quality improvements, but once again, infrequent changes in the composition of the index fail to keep up with the rapid advance of technology. As a result, quality improvements are mistakenly interpreted as pure price movements.

Accounting for New Products

The CPI, as a fixed-weight index, relies on the assumption that the product and service mix of the market basket remains unchanged. The use of this assumption makes it difficult to incorporate new products into the calculation. In some industries, this poses only minor problems. For gasoline and other components of the energy price index, this assumption works reasonably well. The same cannot be said for the medical care industry. The rapid introduction of new medicines and new technologies over the past several decades poses problems for the fixed-weight MCPI.

Infrequent revisions in the index mean that the price index fails to account for significant reductions in the price of newly discovered products. Penicillin, for example, did not enter into the index until its price had fallen to about 1 percent of its original level. A more common problem deals with the introduction of generic drugs. Generics are chemically identical to their name-brand alternatives and usually much cheaper. They do not enter into the calculation of...
the index until weights are periodically revised, and only then as an entirely new product. By that time, they may have captured a significant portion of the market and lowered costs to users substantially. Their addition to the index, however, does not reflect the price decline.

The introduction of the laparoscope has revolutionized many forms of surgery, from knee reconstruction for damaged ligaments to the removal of the gall bladder. In most cases, the new surgical method costs considerably less than the traditional alternative because of shorter hospital stays. Gall bladder removal using laparoscopic techniques requires a 1- to 2-day hospital stay compared with 3 to 7 days using traditional surgical techniques. Repairing a damaged anterior cruciate ligament using the new technique costs 75 percent less for the same medical result.

The BLS incorporates new products and procedures into the index by price linking, replacing the old product with a new one at some arbitrary point in time. This adjustment is made in such a way that the price index remains unchanged; price increases are considered an improvement in quality, but price decreases are simply lost to the index.

Other Problems
In addition to the problems already addressed, several other factors play an important role in creating biased indexes. These include statistical sampling problems, a substitution bias, and the use of list prices instead of transaction prices.

Use of List Prices All published indexes from the BLS use list prices in their calculations rather than transaction prices. The list price is the price paid by a full-paying patient. Information on list prices is easier to collect but may bear little resemblance to the payments that providers actually receive. As more and more providers, physicians, and hospitals enter into agreements with managed care networks and other insurers, actual transaction prices represent discounts from normal list prices. In practice, very few patients actually pay list prices for services.

Suppose a hospital that normally charges $2,500 for a hospital stay agrees to accept $2,000 from a private insurer as payment in full. In this case, $2,000 should be the price that enters into the index. But more often than not, the discounted price differs across payers and is more difficult to determine, so the list price of $2,500 is used.

If list prices and transaction prices change at roughly the same rate, the use of list prices is not particularly glaring. Medical discounting, however, has become an increasingly important phenomenon in recent years, so the use of list prices produces an upward bias on the medical care price index. In fact, the Centers for Medicare and Medicaid Services (CMS) have developed a transaction price index for hospital services. Since 1978, the transaction price index has increased about 70 percent as fast as the hospital index based on list prices (Tregarthen, 1993).

Sampling The high cost of collecting price data dictates that only a limited number of transactions are included in the price index. Sampling can introduce several types of biases into the price index. Because of routine discounts, list prices on the day the data are collected may not be totally representative of the prices that consumers actually pay. Prices paid in the sampled locales may not represent prices paid by most consumers. Discounts for bulk purchases and the increased popularity of generic and store brands are also lost in the sampling procedure used.

Substitution Bias Economists have observed that when the price of a good increases relative to other goods, consumers tend to buy less of it. So as the prices of goods change relative to one another, spending patterns change. Consumers substitute lower-priced items for higher-priced items. This changing pattern of spending, called the substitution effect, is missed completely by fixed-weight indexes like the CPI. As long as the prices of all items in the index rise at roughly the same rate, this phenomenon causes few measurement problems. Over time, however, small differences can add up and result in the statistical phenomenon called substitution bias. This bias does not pose a problem with a deflator, because the market basket changes annually to reflect changing spending patterns. In a fixed-weight index, the weights are changed infrequently (every ten years or so with the CPI), placing too much emphasis on goods whose prices rise the fastest.

**ALTERNATIVE METHODS TO MEASURE MEDICAL CARE INFLATION**

Researchers have suggested alternative measures that might better reflect changes in the price of medical care. Wilensky and Rossiter (1986) advance the case that a change in the measure of medical output would result in more accurate estimates of price changes in medical care. The most commonly
used measure of output is the procedure (e.g., one dose of chemotherapy for the treatment of cancer). Alternatively, output could be defined by the case, such as treatment of cancer from diagnosis to final outcome; the episode, using a particular phase of the illness; or on a per capita basis, measuring the total cost per patient for all medical care.

Another suggested method involves defining a good by a set of characteristics demanded by consumers. This so-called hedonic approach prices those individual characteristics and recombines them to determine the quality-constant price of the good. Trajtenberg (1990) used the hedonic approach to estimate the change in the cost of computerized tomographic X-rays, or CT scans. Defining a CT scan as a set of characteristics, the hedonic index actually declined from 100 to 27.3 from 1973 to 1982. In contrast, the standard index with no quality adjustment showed an increase from 100 to 259.4.

The use of these alternative approaches, though promising in some cases, is not appropriate in others. Even when appropriate, the cost of data collection rises dramatically. Unfortunately, data collection does not seem to be very high on the list of government priorities.

**SUMMARY AND CONCLUSIONS**

Measuring price changes with the indexes we have available is somewhat problematic. Outputs are difficult to measure, new products are included arbitrarily, and the methods for dealing with quality improvements are inadequate at best. Depending on how we interpret the evidence, medical care may be the fastest-rising component of the consumer price index or, using a quality-adjusted notion, medical care prices may be actually falling.

**REFERENCES**


Understanding what economics can and cannot do is the first and possibly most important step in using economics as a tool of public policy. Economics can offer a framework to study the implications of individual decision making, and it can help define the alternative mechanisms available to improve resource allocation. It cannot provide solutions to all the problems of medical care access and delivery. When using economics to study medical care, it is important to avoid extremes. Arguing that economics does not matter, or at least should not matter, when it comes to medical care issues is as ill advised as arguing that economics is all that matters. We cannot avoid the economic implications of our actions any more than we can avoid their moral implications.

Sound policy making is based on sound economic principles applied in a compassionate and consistent manner. The premise of this book is that policy making based on sound economics is better than policy making in an economic vacuum. Basic economics teaches us many lessons: about human behavior and the way individuals make decisions and respond to incentives, about the way people interact with each other, and about the efficient allocation of scarce resources. Economists do not claim to have the final word about how to organize and run a health care system, but they do have something relevant to add to the discussion.

The goals of this chapter are somewhat ambitious. Those of you who have been exposed to an economics course may be tempted to skip this chapter completely: Avoid that temptation. At a minimum, use the chapter to refresh your memory of the important concepts that will come into play in analyzing medical markets and the policies that affect them. Those of you who have never had the privilege of taking a course in economics will find this chapter useful in setting the tone for the rest of the book. The principal focus here will be the examination of the basic principles of supply and demand.

Rhetoric in Economics

An important element of economics is conversation, so economists must be persuasive communicators. Economics has its own rhetoric, and those unfamiliar with it have a difficult time understanding it. Economists use mathematical and statistical tests to make arguments, but when you listen closely to their conversations, you hear many literary
The Relevance of Economics in Health Care

Economics is a way of organizing our thinking about problems that confront us in our daily lives. To think like an economist requires a disciplined approach to problem solving, and sound reasoning within a systematic framework is essential. The value of economics stems from its usefulness in making sense out of complex economic and social issues, including issues in medical care delivery. Future health care decision makers will need training and knowledge in many areas: not only biology and chemistry but also statistics, epidemiology, behavioral science, ethics, decision analysis, and, of course, economics.

Economics is one of several social sciences that attempt to explain and predict human behavior. It is unique among the social sciences in establishing a context of scarcity and uncertainty. More specifically, economics is concerned with the way scarce resources are allocated among alternative uses to satisfy unlimited human wants.

The quest for economic efficiency stems from the fact that there are never enough resources to provide all the goods and services desired by a society. Economists call this concept scarcity. Using resources in one activity precludes the use of those same resources in a different activity. When resources are used in medical care delivery, those same resources are not available for use in other beneficial activities; for example, food distribution, education, housing, and national defense.

The economic concept of cost stems from the notion that resources have alternative uses. The term opportunity cost is defined as the potential benefit that could have been received if the resources had been used in their next-best alternative. Tax dollars used to purchase medical care for the elderly cannot be used to buy education for the young. Money spent in a rehabilitation program for drug addicts is not available to spend on prenatal care for indigent women. Adopting the
concept of economic efficiency implies that choices should be made in a way that maximizes the total benefit from the available resources. In the practice of medical care delivery, this involves the evaluation of health care alternatives by calculating the benefits and costs of each and allocating resources in a way that maximizes the net benefits to the community.

**Critical Assumptions in Economics**

All scientific models start with assumptions. Economic models start by assuming rational behavior on the part of decision makers, meaning everyone involved in a decision behaves in a purposeful manner.⁠¹ Economics is different from other social sciences in its emphasis on rational decision making under conditions of scarcity.

In microeconomics, the assumption of rational behavior establishes a consistent framework for individual decision making. We assume that individuals, in an attempt to reach certain objectives, must choose among competing alternatives. The problem becomes one of allocating scarce resources among these competing ends. In other words, we cannot satisfy every desire we have; we must make choices.

Decision makers, motivated by self-interest, respond to incentives. In fact, decision making is dominated by the pursuit of self-interest. Individuals use their resources to advance their own economic well-being. When confronted with alternative actions, they choose the one that makes them better off.²

People look for the best way to achieve their goals. This does not rule out impulsive behavior or mistakes. In fact, because information is costly to gather and process, decision makers often practice rational ignorance: They decide between alternative actions with incomplete information. From the decision maker’s perspective, the information left to be gathered costs more to gather than it is worth.

Scarcity is the reason we study economics. In a world of superabundance, there would be no compelling reason to make choices. All people could have all that they wanted without concern for alternative uses. Or, if all individuals had the divine nature of saints, then our attitude would be one of relative indifference toward material goals, and scarcity would not be an issue. But we do not live in a world of superabundance, and the world is not populated by saints, so decision making must take into consideration forgone opportunities.

**The Scientific Method**

The challenge at hand is to understand economic relationships without the luxury of controlled experiments. Economic inquiry utilizes the scientific method in much the same way that physics and chemistry do. There are five basic steps in the scientific method:

1. Every scientist starts with a premise, or postulate, that serves as a foundation for the inquiry. Some may call it an ideology or even a vision. Either way it represents the scientist’s understanding of the way the world works. The culture around us, the way our parents raised us, and years of scientific training and inquiry all affect the way we view the world around us. Even the

---

¹ Note that it is possible to study human behavior without assuming rationality, but that would not be economics.

² Altruistic behavior is not ruled out; it is merely interpreted as self-interested behavior.
most unbiased among us are affected by some bias; at minimum, our biases affect the nature of our inquiry.

2. The world arouses our curiosity. Scientists are careful observers of real-world phenomena and events. These observations concerning the real world are organized and catalogued.

3. A theory is developed to explain the observed behavior or predict future behavior. Model building captures the essential features of the observed behavior. It is a meaningful abstraction, decomposing the problem into its elemental parts.

4. The scientist then formulates a hypothesis to test the predictions of the theory. This requires gathering of facts and data.

5. In the final step, hypothesis testing, we use quantitative techniques to improve our understanding of the issue and promote more accurate predictions.

In practice, an economist might approach a problem using the scientific method as follows: One vision of the way the world works might be that people who are truly motivated by self-interest will respond in measurable ways to changes in incentives. From this vision, a theory is developed that people will respond to higher out-of-pocket payments for health care by demanding fewer elective procedures. The RAND insurance experiment conducted controlled trials that randomly placed individuals into different types of health plans (Manning et al., 1987). By varying the out-of-pocket payments required of individuals, their demand for medical care was analyzed. Empirical results supported the hypothesis that higher out-of-pocket payments would lead to lower utilization, measured as fewer physician visits. The RAND experiment has spawned many studies, testing numerous different hypotheses. The way we think about health insurance pricing and payment policies has been significantly affected by this important research.

These are the steps involved in the scientific method: an ideological base; observation of events; development of a theory; hypothesis testing; and, finally, rethinking. Empirical results that run counter to expectations may cause the scientist to rethink the theory or develop a different hypothesis.

Model Building

One of the main goals of economics is to understand, explain, and predict the behavior of decision makers. To this end, economists find it necessary to simplify that behavior, this simplification is accomplished through generalization, often through the construction of models.

A model is nothing more than a way of organizing knowledge on a particular issue so that it becomes more than a set of random observations. An economic model explains how the economy, or part of the economy, works. The terms model and theory are often used interchangeably. By their very nature, models are simplifications of the real-world phenomena they attempt to explain, and model building is an exercise in abstract thinking.

Microeconomic models examine the behavior of individual decision makers—individuals, households, firms, and government agents—and the behavior of specific markets. We use microeconomic models to study how a patient’s demand for a particular diagnostic test varies, depending on the out-of-pocket cost of the test. We can examine how a shortage of qualified nurses affects nurses’ salaries, or how the relative income of specialists affects the demand for residency-training positions in all specialties.
Problem Solving

Economics emerged as a science in the late eighteenth century with the publication of Adam Smith’s *The Wealth of Nations*. Since that time, a wealth of theory has accumulated to help us understand and describe *economizing behavior*. Most microeconomic theory can be classified under the framework of *neoclassical economics*. Relying heavily on the rationality assumption, the neoclassical framework classifies all decision makers as optimizers—those who attempt to maximize their well-being. *Optimizing behavior*, or optimization, is nothing more than a decision maker seeking to accomplish certain objectives: maximize sales or profit, minimize cost, or maximize income. Economists often talk of decision-making calculus, which refers to the notion that individuals make mental calculations before arriving upon a decision. Optimization fits the calculus model well in that it evaluates a mathematical function for its maximum or minimum value.

**ECONOMIC OPTIMIZATION**

When more than one alternative is available, the optimal choice produces an outcome that is most consistent with the decision maker’s stated objectives. Optimization is nothing more than discovering the best course of action given the decision maker’s goals and objectives. Constrained optimization takes into consideration the cost and availability of resources. Would it be better for the

---

**ISSUES IN MEDICAL CARE DELIVERY**

**Positive and Normative Analysis**

To a great extent, we will mix positive and normative analysis in our discussions. Positive analysis is the testing of hypotheses against facts; it examines the way things are. Normative analysis prescribes policies and actions to achieve certain goals; it purports to examine the way things ought to be.

The differences between positive and normative statements are easy to spot: “The United States spends more money per capita on medical care than any other country in the world” is an example of a positive statement. “Congress should guarantee universal insurance coverage by requiring all employers to provide health insurance to their workers” is a normative statement.

Positive statements are either true or false. It is the task of science to determine which they are. Normative statements are matters of opinion, so science is of little help in determining their legitimacy. Fuchs (1996), in a survey of 90 economists concerning issues in health economics and health policy, found that over 90 percent disagreed with the positive statement, “In the long run, employers bear the primary burden of their contributions to employees’ health insurance.” In contrast, opinion was divided almost equally on the normative statement: “National standardized health insurance benefit packages should be established.” Disputes over factual information can be settled through careful observation and analysis. Settling disputes over differences of opinion, on the other hand, is almost never easy. In fact, disagreements among economists are typically disputes over normative issues, and these disagreements represent differences of opinion based on differences in ideology.
hospital to enter into a contract for housekeeping services with an outside firm, or should this activity be performed in-house? Following an increase in patient volume, should physicians in a small group practice hire an office manager, an additional nurse, or both?

Choices in health care delivery must be made at two levels: individual physicians must decide on a particular course of treatment for a particular patient, and policy makers must decide on a course of action in planning the availability of health services for an entire community. The delivery of health care in any form must cover the following areas: whom to treat, when to begin treatment, where to treat, and how much treatment to offer. Of the many ways to go about choosing the best alternatives, economic efficiency will be the criterion examined in this section.

In a sense this decision making is nothing more than the classic “economic problem.” Resource allocation demands that we answer three basic questions: 1) What do we produce? 2) How do we produce it? 3) Who gets it? Regardless of our perspective, whether we are examining economic systems, health care systems, business firms, individuals, or decision makers of any kind, something must drive the system to produce and distribute what people want. Just remember: what, how, and for whom? This is the economic problem that must be solved to promote growth and welfare in any modern society.

To resolve the problem, firms attempt to maximize profit, given production technology and the cost of available resources; consumers attempt to maximize satisfaction, subject to limited money income and the prices of goods consumed; and workers supply labor services in an attempt to maximize satisfaction derived from goods and services consumed and leisure time available subject to current wages. Together, this more or less independent behavior results in markets that tend toward equilibrium as represented by the familiar, or soon to be familiar, supply and demand framework.

Within this framework, what does optimal mean? Using the rhetoric of economics, it means that individuals will continue to purchase a good or service as long as the marginal benefits from consumption (MB) exceed the marginal costs (MC). Given that marginal benefits are declining and marginal costs are increasing as more of the good is consumed, eventually the two will be equal. As soon as \( MB = MC \), equilibrium is reached, and the individual will consume no more. In Figure 2.1, the total benefits (TB) received from a medical procedure increase as more care is provided, but at a decreasing rate. For reasons both ethical and practical, medical practitioners tend to provide additional care as long as the treatment results in positive benefits. Beyond point A, additional medical care is considered equivocal or wasteful—the marginal benefits are not worth the medical risk.

From the perspective of economics, exhausting all possible medical benefits wastes scarce resources. In fact, any care provided beyond point B is wasteful, because the marginal benefits received from the additional care fall short of the marginal costs. The resources used in providing the excess care could be put to better use somewhere else. Money wasted in the provision of unnecessary care cannot be used to further other important goals, such as improving education, repairing the interstate highway system, or cleaning up the environment.

When consumption is being subsidized, the cost to the consumer is less than the total resource cost, as in the case of medical care purchased with insurance.

---

3 In this discussion cost is measured in terms of total resource cost, the actual opportunity cost of the resources consumed in the production of medical care, not merely the out-of-pocket cost to the consumer.
In the case of the insurance subsidy for medical care, the cost of an extra unit of care to the individual is close to zero, providing an incentive to consume medical care with low marginal benefits. When the marginal cost to the consumer is artificially low, resources are treated as if they had little or no value—a prescription for overconsumption. This tendency to overconsume means that medical care consumption is likely to be closer to point A, where the marginal benefit is close to zero, than point B, where the marginal benefit is equal to marginal cost. This phenomenon is called flat-of-the-curve medicine.4

**Key Concept 9**

**Supply and Demand**

Many consider supply and demand the two most useful concepts in economics. Regardless of the issue being studied, the analysis often hinges on some aspect of supply and demand. The theory of supply and demand is also a powerful tool in predicting future behavior. How does a change in price affect the consumer’s willingness or ability to purchase a commodity? How does a change in the price of a key input affect the producer’s decision about the optimal input combination to use in the production process?

In modeling behavior, economists attempt to simplify relationships. The amount of a particular commodity that a consumer plans to purchase depends on several factors. Instead of looking at the large number of variables that would affect

---

4 The phrase “flat of the curve” is attributed to Alain Enthoven (1980).
demand, we focus on the most important ones: the price of the commodity; the price of related commodities; the number of people desiring the commodity; and consumer income, preferences, and expectations.

**The Law of Demand**

The theory of demand occupies such an important place in economic analysis that it has been given the status of a law. The law of demand states:

*There is an inverse relationship between the amount of a commodity that a person will purchase and the sacrifice that must be made to obtain it.*

The higher the price of an item, the less it is purchased, and the lower the price, the more it is purchased. It is important to understand that this inverse, or negative, relationship holds as long as the circumstances of the consumer do not change materially. Remember, other things affect the demand relationship: prices of related items, the consumer’s income, and preferences. As long as there are no changes in these other factors, the inverse relationship holds. When prices rise, less is desired. When prices fall, more is desired.

Changes in price affect the demand relationship in two very important ways: First, consumers have alternative ways to spend their money. If the price of a name-brand drug goes up, an alternative drug or even a generic can be substituted for the name brand. Or if money is tight and no insurance coverage is available, the patient can choose to skip the treatment and let the disease run its course. In any case, when price rises, the quantity demanded goes down. Economists refer to this phenomenon as the substitution effect.

A change in price affects the consumer in another important way. Paying higher prices for a desired commodity reduces the consumer’s overall level of satisfaction. Spending more for one item leaves less to spend on everything else. With less money to spend, the consumer is unable to buy as much of everything else as before and thus feels worse off. This aspect of a price change on quantity demanded is called the income effect.

Part (a) of Figure 2.2 illustrates how an increase in price affects demand. Suppose that the demand for a particular commodity is represented by the demand
Part 1 >> The Relevance of Economics in Health and Medical Care

curve $D_1$. Assuming no other changes, an increase in the price from $P_0$ to $P_1$ will reduce the amount demanded from $Q_0$ to $Q_1$. This is depicted by a movement along the stationary demand curve from point $A$ to point $B$. A change in price, holding everything else constant, changes the quantity demanded.

There are many factors other than price that influence our purchasing decisions. These other factors are sometimes referred to as ceteris paribus conditions (remember, economics has a language of its own). These conditions are factors that are held constant when examining the relationship between price and quantity demanded. They include:

- The price of related commodities
- The number and type of people desiring the commodity
- Consumer income
- Consumer preferences
- Consumer expectations about future prices and product availability

A change in the price of a related commodity changes the demand for the commodity in question. Related commodities are either substitutes or complements. An increase in the price of a substitute increases the demand for a commodity. Coronary artery bypass graft surgery (CABGS) and cardiac angioplasty are two procedures used to accomplish the same outcome. If the price of CABGS increases, heart patients—or rather whomever is paying for the procedure—will view cardiac angioplasty as a more viable alternative. The demand for cardiac angioplasty will increase.

When the price of a complement goes down, demand goes up, because complementary goods are consumed together. Dentists often recommend that full-mouth X-rays accompany the annual dental exam; X-rays complement the annual exam. If the price of the X-ray goes down, more patients will make appointments for dental exams.

An increase in the size of the population or its composition affects demand. More people means a higher demand for all goods and services, including medical care. The addition of an infant to a family increases the demand for visits to the pediatrician. An increase in the birth rate raises the demand for disposable diapers, even if the average baby still uses the same number of diapers per day. An older population has a higher demand for treatments for chronic illnesses, such as arthritis and emphysema.

A change in income affects the consumer's ability to purchase goods and services. In situations where higher income leads to increased demand, the good in question is referred to as a normal good. In some cases, an increase in income leads to a decrease in demand. In those situations, the good is called an inferior good. Medical care is usually considered a normal good. For individuals with comparable levels of health, higher income means a higher demand for medical care. Good health improves a person's ability to earn income. Higher income in turn increases the return to good health and increases the demand for medical care.

Consumer preferences play a key role in determining an individual's demand for goods and services. Some flu sufferers will consider a visit to the physician

---

5 The examples point out the importance of distinguishing between the individual demand and market demand. Clearly, the market demand curve is determined by combining the demand curves of all the individuals actively participating in the market.
only as a last resort. They prefer to treat their ailment with over-the-counter medications. Some people hold religious beliefs (e.g., Christian Scientists) that strongly discourage the use of medical care. Others are convinced of the efficacy of chiropractors, herbalists, acupuncturists, midwives, and other alternative providers. They prefer these alternatives to the more traditional health care providers, and this shift in preferences can have a powerful impact on demand.

Consumer expectations play a key role in determining the level of demand. If consumers expect prices to change steeply and suddenly, or if they are afraid the product will be difficult to obtain in the near future, demand will rise sharply.

Finally, it is important to note that the demand for resources is a derived demand. Whenever a resource is used to produce a final product, the demand for that resource is ultimately determined by the demand for the final product. If medical care is considered an essential element in promoting the health of an individual or a group of people, an increase in the demand for health will increase the demand for medical care.

A change in one of these other factors changes the level of demand and causes a shift in the demand curve. Refer once again to Figure 2.2. Part (b) depicts a change that increases the level of demand caused by an increase in the price of a substitute commodity, a decrease in the price of a complement, an increase in consumer income, a positive shift in preferences, the expectation of a price increase, or a decline in availability in the future. Suppose the level of demand is originally $D_1$ in part (b). At the price $P_0$, the quantity demanded is $Q_0$. With the price held constant, an increase in consumer income will cause a rightward shift in the demand curve to $D_2$. This shift in the demand curve depicts an increased demand for the commodity. The consumer will now desire $Q_1$ at the price $P_0$.

To summarize, a change in the price of a commodity or service, holding everything else constant, will result in a change in quantity demanded, shown as a movement along a stationary demand curve. A change in any of the factors that affect the level of demand results in a shift in the demand curve—more or less of the commodity or service is demanded at every price level.

**How to Survive Supply and Demand**

Succeeding in any economics course, especially a course in medical economics, depends on your mastery of the twin concepts of supply and demand. Listen carefully to economic commentators when they are queried on a complex issue in economic theory or policy, and their answer is frequently preceded by “It’s all because of supply and demand.” The introduction of supply and demand into the economics vocabulary is soon followed by adding supply and demand curves to the lexicon. In this hostile environment, survival depends on your ability to keep your wits about you while others around you fail. To ensure your success, follow these simple rules of survival:

- **Use common sense.** Most students already know a great deal about supply and demand. The key is to use what you know. Remember, economics is a way of thinking. For the most part, it is intuitive.
Think about the market for oatmeal. Scientific evidence has suggested that consuming large quantities of oat products every day reduces the level of cholesterol in the bloodstream and thus the risk of heart attack. What do you suppose happened to the demand for oatmeal, and its price, immediately after this information was made public? If you said that demand for oatmeal increased and its price also went up, then you already have some intuitive notion of the workings of supply and demand.

- **Learn the language.** After a few weeks in Econ 101, many students feel they are taking a foreign language. Mastery of economics requires that you learn the language of economists. When it comes to supply and demand, economists speak in graphs. Understand graphs and you understand supply and demand. If freshman literature were taught in Greek, it would be extremely difficult for the typical student. Not that the subject matter is so hard, it's the language. Introductory economics is taught in graphs. Learning to use graphs makes learning economics much easier.

- **Practice, practice, practice.** The rules of graphing are simple. Unlike a foreign language, there are no irregular verbs. But like a foreign language, it takes practice to master the subject matter. Practice whenever you can; economics is not a spectator sport. Watching your professor manipulate graphs is not enough: You have to do it yourself. Remember, demand curves are downward sloping, and supply curves are upward sloping. Economists place price on the vertical axis and quantity on the horizontal axis. Equilibrium price and quantity are determined by the intersection of the supply and demand curves.

- **Shift the appropriate curve.** The discovery that oat products have health benefits affected the market for oats. Did it affect supply or demand or both? Remember what causes shifts in the two curves. For the supply curve to shift, a change in the cost or profitability of making a product available to the market is needed. A shift in the demand curve is precipitated by anything that changes the willingness or ability of consumers to buy something. The discovery that oatmeal works like Roto-Rooter to clean out your arteries affected consumers' willingness to buy the product. So the demand curve shifted. Did it shift to the right or to the left? If in doubt at this point, go back to rule number one: An increase in demand will increase price. The only way to get this result is to shift the demand curve to the right. Shifting the demand curve to the left, or shifting the supply curve, is counterintuitive.

It is now time to test your mastery of supply and demand. Consider the market for hospital services. Use a graph similar to the one in Figure 2.5 and label the vertical axis “Price of Hospital Services” and the horizontal axis “Quantity of Hospital Services.” Draw the supply and demand curves and identify the equilibrium price and the quantity of hospital services. Now suppose that due to a nursing shortage, the average salary paid to nurses increases 10 percent. What affect will this increased cost have on the market for hospital services?

Price Elasticity of Demand

An important corollary to the law of demand is the concept of price elasticity of demand. The law of demand is used to answer the question, when price changes, what is the effect on the quantity demanded? Taking this notion one step further, price elasticity of demand is a technical concept used to answer the question, when price changes, how much does quantity demanded change? The inverse relationship between price and quantity is relatively easy to comprehend. In most cases, it is important to include not only the direction of the change but the magnitude of the change.

Price elasticity of demand measures consumer responsiveness to a change in price, holding the other variables that affect demand constant. Slope also measures the relationship between quantity demanded and price, but slope is not elasticity; slope measures the change in quantity demanded that results from a price change in absolute terms. Elasticity measures the change in relative terms.

Price elasticity of demand is defined as the percentage change in quantity demanded divided by the percentage change in price. Formally, price elasticity ($e_p$) is calculated as

$$e_p = \frac{\text{percentage change in } Q}{\text{percentage change in } P}$$

where $Q$ is quantity demanded, and $P$ is the unit price.

If consumer demand increases 10 percent because of a 5 percent price decrease, price elasticity of demand is 10 percent divided by 5 percent, or 2.0.\(^6\) Values for the elasticity coefficient range from zero (0) to infinity ($\infty$).

A summary of all possible values for the price elasticity coefficient is provided in Table 2.1. In the case in which price elasticity equals zero, consumers are completely unresponsive to changes in price. Their consumption patterns are fixed, and a higher price does not affect quantity demanded. Under these circumstances, demand is said to be perfectly inelastic, or totally unresponsive. The demand for addictive substances may come about as close to perfectly inelastic demand as anything. The demand for life-saving procedures, such as kidney dialysis and organ transplants, may also fall into this category.

### Table 2.1: Price Elasticity of Demand

<table>
<thead>
<tr>
<th>Coefficient Value</th>
<th>Nature of Demand</th>
<th>Impact of a 10 Percent Price Increase on Quantity Demanded</th>
<th>Impact of a 10 Percent Price Increase on Total Expenditures</th>
</tr>
</thead>
<tbody>
<tr>
<td>$</td>
<td>e</td>
<td>= \infty$</td>
<td>Perfectly elastic</td>
</tr>
<tr>
<td>$1 &lt;</td>
<td>e</td>
<td>&lt; \infty$</td>
<td>Elastic</td>
</tr>
<tr>
<td>$</td>
<td>e</td>
<td>= 1$</td>
<td>Unit elastic</td>
</tr>
<tr>
<td>$0 &lt;</td>
<td>e</td>
<td>&lt; 1$</td>
<td>Inelastic</td>
</tr>
<tr>
<td>$</td>
<td>e</td>
<td>= 0$</td>
<td>Perfectly inelastic</td>
</tr>
</tbody>
</table>

\(^6\) The actual calculation is $\left[\frac{(+0.10)}{(-0.05)}\right] = -2.0$. While the price elasticity coefficient is always negative, for simplicity we usually ignore the negative sign, or more precisely, we consider its absolute value.
A more likely scenario would be the case in which a price change has an impact on quantity demanded, but the consumer response is less than proportional. In other words, we consider consumer demand somewhat unresponsive when the percentage change in quantity demanded is less than the percentage change in price. In this case, the elasticity coefficient is less than one, and demand is inelastic. Even addicts and terminally ill patients have their limits on how much they are willing or able to pay for a desired commodity.

An elasticity that is greater than one represents a change in quantity demanded that is proportionately greater than the change in price. Consumers are said to be relatively responsive, and in this case demand is elastic. In the rare case where the elasticity coefficient is equal to infinity, demand is perfectly elastic; consumers are intolerant of even small changes in price and refuse to buy the item if its price goes up at all.

An important use of the concept of price elasticity is illustrated in the right-hand column of Table 2.1. When price changes, it is important to know how much quantity demanded changes. It is also important to realize that this same information enables us to predict what will happen to consumer expenditures. With perfectly elastic demand, any price increase causes quantity demanded to fall to zero. In this case, it may be obvious that consumer expenditures also fall to zero. The case of unit elasticity may not be so obvious. When price elasticity equals one, a 10 percent price increase causes quantity demanded to fall by 10 percent, and consumer expenditures do not change. Likewise, price increases cause consumer expenditures to fall when demand is elastic and to increase when demand is inelastic.

What determines the price elasticity of demand? Why are consumers more tolerant of price changes for some items but not others? Price elasticity depends primarily on the consumer's ability to find suitable substitutes for a good or service. The easier it is to substitute, the more elastic the consumer's demand. If the consumer perceives a number of good alternatives to the item, demand is likely to be more responsive to changes in price. Patients with no established preference for a general practitioner (GP) might view a 20 percent increase in the price of an office visit as intolerable in light of the number of suitable alternative GPs in practice. However, those individuals who have an established relationship with a GP may be willing to remain a loyal patient in spite of the price increase. In this case, the GP will lose some business but not all of it.

Key Concept 6
Supply and Demand

ISSUES IN MEDICAL CARE DELIVERY

Is “Safe” Sex Really Safe?

One of the costs of risky sexual practices is an increased likelihood of contracting a sexually transmitted disease (STD) such as syphilis and gonorrhea or even AIDS. As with any activity involving human choice, the higher the perceived cost of engaging in risky sex, the less the demand for it. This suggests that by making sex “safer” through free condom distribution — in effect lowering the cost of risky behavior — public health officials may be increasing the incidence of that behavior and actually increasing the incidence of STDs.

The logic of this possibility is based on the fact that there is a demand curve for sex. It is difficult to know its exact shape, but most economists
would agree that it is downward sloping. As the perceived cost of a sexual encounter (the risk of contracting an STD) falls, the number of sexual encounters will increase. The size of the increase is determined by the "risk elasticity of demand for sex."

The risk elasticity of demand for sex is defined as the percentage change in the number of sexual encounters divided by the percentage change in the risk of each encounter. If the risk elasticity is less than one, then free condom distribution will reduce the incidence of disease. If it is greater than one, the incidence of disease will increase.

Consider a closed community, where condoms must be purchased and no one uses them. According to research (Rosenberg et al., 1992), the risk of contracting three common STDs, gonorrhea, trichomoniasis, and chlamydia, during unprotected sexual activity is 23.4 in 100. If the number of risky sexual encounters is 250 per week, there will be 58 new infections every week. Assume that condoms are now distributed free of charge, and their use is widely encouraged through a sex education program. The use of condoms will result in a reduction in the incidence of STD to 18.8 per 100 risky sexual encounters — a 20 percent reduction.

If the demand for sex is inelastic, and the risk elasticity of demand is \( -0.50 \), the incidence of sexual intercourse will increase from 250 per week to 275, a 10 percent increase. In that case, there will be 52 new cases of STD every week, a 10 percent decrease. On the other hand, if the demand for sex is elastic, and the risk elasticity of demand for sex is \( -1.5 \), sexual intercourse increases from 250 incidents per week to 325 — a 30 percent increase. In that case, there will be 61 new cases of STD reported every week, a 5 percent increase.

Does the policy of making condoms available increase or decrease the number of cases of STD? While the value of risk elasticity of demand for sex is an empirical matter, there is some evidence that sexual activity is higher in situations in which condoms are widely available. According to Planned Parenthood, in schools with formal sex education programs and free condom distribution, the percentage of males engaging in sex increased from 60 to 84 percent, and the use of condoms actually decreased (Family Planning Perspectives, 1994). Kasun’s review (1994) of seven sex education programs with easy access to condoms revealed that six resulted in an increase in sexual activity.

Any attempt by policy makers to make sex safer could actually exacerbate the problem by encouraging sexual activity. The risk elasticity of demand for sex determines whether the incidence of STD infection increases or decreases.


Other factors that influence the degree of consumer responsiveness are the proportion of a person’s income spent on the item and the urgency of the purchase. If the cost of the item comprises a substantial portion of a consumer’s total income, demand will likely be elastic. Consumers are more sensitive to a price change on the purchase of big-ticket items. Insulin-dependent diabetics are more sensitive to
a change in the price of syringes than the typical nondiabetic patient. The diabetic patient buys a lot more syringes per year than the nondiabetic. Finally, demand for nonurgent procedures will be more elastic than demand for emergency procedures. The more time a patient has to make a decision, the more price sensitive he or she will likely be. A patient entering the emergency room with a compound fracture does not have much time to shop around for an orthopedic surgeon. Patients desiring elective rhinoplasty, however, have the opportunity and the luxury to shop around for the best plastic surgeon, the best price, the best financing, or whatever else they consider important. A patient who shops around is more likely to find suitable alternatives.

Demand curves are typically drawn as straight lines for the sake of simplicity. There are three possibilities, as shown in Figure 2.3. Perfectly inelastic demand curves are drawn as vertical lines indicating zero response, and perfectly elastic demand curves are depicted by horizontal lines. The typical downward-sloping demand curve is shown at the right. Although slope is the same at every point, elasticity is not. The relationship between slope and elasticity at any point on the demand curve can be shown to be

\[ e = \frac{\Delta Q/Q}{\Delta P/P} = \frac{P\Delta Q}{Q\Delta P} = \frac{P}{Q} \frac{\Delta Q}{\Delta P} \]

where \( Q \) is the quantity demanded, \( P \) is the unit price, and \( \Delta \) is used to represent a change in the variable.

A demand curve with a given slope has a constantly declining elasticity. Moving from the upper left to the lower right on a downward-sloping demand curve, the \( P/Q \) ratio is declining: as price falls, quantity demanded increases. It follows that the demand curve goes from elastic to inelastic as you move down a straight-line demand curve.\(^7\)

### The Law of Supply

The theory of supply assumes that decision makers, producers in this case, are faced with scarce resources and must choose among alternative uses. Supply

---

\(^7\) Economists sometimes refer to an entire demand curve as inelastic if it is generally steep and elastic if it is generally flat. While technically incorrect, as a matter of convenience, we often think in these terms.
decisions involve the valuation of resources among competing uses. The law of supply states:

*There is a direct relationship between the amount of a commodity that a producer will make available and the reward that is received.*

The higher the price of an item, the greater its availability. At lower prices, less will be available. Suppliers practice economizing behavior much as consumers do. The market rewards efficiency and punishes wastefulness.

Producers are concerned with cost. This concern is more than an accounting of the value of inputs; it involves establishing the opportunity cost of those inputs. In economics, cost reflects the value of resources in their next-best alternative use. In other words, forgone opportunities are an important element in determining value. Resources used in the production of one commodity are not available to produce another. Economizing behavior guarantees that resources will be used where they have the highest value. Therefore cost is determined by the value of what is being given up to produce any item.

Part (a) of Figure 2.4 illustrates how a change in the price of a commodity affects quantity supplied. Suppose that supply is depicted by the curve $S_1$. Assuming no other changes, an increase in price from $P_0$ to $P_1$ will increase the quantity supplied from $Q_0$ to $Q_1$. At higher prices, suppliers will transfer resources to the production of the higher-priced commodity, making more of it available to the market. A change in price, holding everything else constant, results in a change in quantity supplied and is depicted by a movement along the stationary supply curve.

Many other factors affect the availability of goods and services in a market. A change in any one of these factors, the *ceteris paribus* conditions, will change the level of supply. These other factors that affect the level of supply include:

- The prices of resources used to produce the commodity
- The number of firms supplying the commodity
- The state of technology
- Producer expectations about future prices and availability

![Figure 2.4: A Change in Quantity Supplied and a Change in the Level of Supply](image)
In general, anything that changes the costs of producing a commodity will affect the level of supply. Resources have alternative uses. In order to use resources to produce a particular commodity, producers must bid them away from their next-best alternative use. An increase in the price of a resource decreases the supply of the commodity that uses the resource as an input in the production process, and it raises its price. Technicians trained to operate the new magnetic resonance imaging (MRI) machines are in short supply. As competition bids up their wages, the cost of providing MRI services increases, shifting the supply curve for MRIs to the left and raising the price of the service in the market.

An increase in the number of suppliers increases access to a product or service. More suppliers means that consumers have more choices. The construction of a new 250-bed hospital in a community will increase the availability of inpatient hospital services to local residents. At any given price per day, there are now more beds available to serve the patient population.

New technology that reduces the cost of producing a commodity or service increases the level of supply. In the case of medical technology, certain analytical problems make it difficult to evaluate the different supply responses of cost-reducing and quality-enhancing technology. Arthroscopic surgery provides a clear example of a technological advance that represents both a cost-reducing and quality-enhancing change. The repair of a damaged anterior cruciate ligament was once a major ordeal for both surgeon and patient. Before the introduction of the laparoscope, an athlete who suffered this knee injury was faced with a four-hour surgery requiring a six-inch incision, several days in the hospital, and six weeks on crutches. Today, the same procedure can be performed as outpatient surgery. It requires three small incisions and a much shorter rehabilitation.

If suppliers expect the price of a commodity or service to fall in the future, they have an incentive to make it immediately available. If for some reason suppliers expect an increase in future availability, current supply will increase. As the medical marketplace moves systematically toward the managed care model, physicians scramble to join provider networks. Expectations create powerful incentives. As more physicians join networks, fueling expectations, others feel an urgency to join them, too.

An increase in the level of supply is illustrated graphically in part (b) of Figure 2.4. Anything that enhances a producer’s ability to bring a product to the market increases the level of supply and results in a rightward shift in the supply curve. A decrease in resource costs, an increase in the number of providers, a technological advance that increases production efficiency, and the expectation of downward-price movements all increase the level of supply and cause the supply curve to shift to the right. Suppose that the supply curve shifts from $S_1$ to $S_2$. At any given price level, say $P_0$, providers will be willing to increase the amount supplied from $Q_0$ to $Q_1$.

To summarize, a change in the price of a commodity or service, holding everything else constant, will result in a change in the quantity supplied. This change is shown as a movement along a stationary supply curve. A change in any of the factors that affect the level of supply results in a shift in the supply curve and a change in the availability of the commodity or service at any given price.

**Equilibrium**

Price changes affect buyers and sellers differently. An increase in price reduces the consumer’s willingness to buy and at the same time increases the producer’s willingness to provide. The most fascinating aspect of the marketplace is how the more or less independent behavior of buyers and sellers results in an allocation of resources that guarantees that all consumers willing to pay the market price will find willing sellers, and all sellers willing to accept the price will find buyers.
Adam Smith observed that it is as if an “invisible hand” were responsible for the price adjustments that promote the best use of resources.

We define the equilibrium price as the market price that exists when the quantity demanded equals the quantity supplied. Suppose that the price of the commodity depicted in Figure 2.5 is $P_1$. At that price, producers would like to sell more than consumers are willing to buy. There is a surplus, because the quantity supplied is greater than the quantity demanded. When prices are too high in the medical marketplace, hospitals, for example, will have unused capacity. This excess capacity takes the form of idle resources, empty beds, and unused operating rooms. Physicians find their appointment books unfilled and their waiting rooms empty. A surplus serves to increase competition among providers. The competition may manifest itself in many ways, but one sure way to eliminate the surplus and increase quantity demanded is to lower prices.

At the price $P_2$, quantity demanded exceeds quantity supplied, resulting in a shortage. Patients experience significant delays in getting appointments. When they do get an appointment, the waiting room is crowded and delays are frequent. Nonemergency surgeries have to be scheduled far in advance. Access to diagnostic imaging equipment is limited. Under these conditions, prices have a tendency to adjust upwards. Competition among consumers bids prices up and reduces quantity demanded. Coupled with an increase in quantity supplied, the shortage is eliminated. Only one price does not result in either a surplus or a shortage. That price, $P_0$, the equilibrium price, clears the market. At $P_0$, the behavior of buyers and sellers coincides. Buyers are willing to pay the price that providers are willing to accept. Everyone who wants to buy at $P_0$ is able to buy, and everyone who wants to sell at that price is able to sell. In a market economy, people are free to make transactions: they are free to bid for goods and services at any price and free to offer those same goods and services at any price. When buyers seek the lowest price that producers are willing to accept, and sellers seek the highest price that consumers are willing to pay, the transaction price that clears the market is the equilibrium price.

**THE COMPETITIVE MODEL**

Free markets play a crucial role in the free enterprise system. The market system is grounded in the concept of consumer sovereignty: what is produced is determined
Part 1 >> The Relevance of Economics in Health and Medical Care

by what people want and are able to buy. No one individual or group dictates what must be produced or purchased. No one limits the range of choice.

The market accomplishes its task of resource allocation through a system of prices, what Adam Smith called the “invisible hand.” In a market system, resources can be allocated by this invisible hand because everyone and everything has a price. There is a tendency for prices to increase if more is desired and to decrease if less is desired.

Firms base their production decisions on relative prices and relative price movements. The price mechanism becomes a way of bringing a firm’s output decisions into balance with consumer desires, something that we refer to as equilibrium.

Prices serve not only as a signal to producers but as a means of rewarding popular decisions. Producers who invest in appropriate technology are able to produce goods and services desired by consumers. Their rewards come in the form of profits. Poor decisions are in turn punished by the market, and the producer suffers losses. This market discipline, accompanied by the freedom to compete within a system that allows private property ownership, is largely responsible for the efficient use of resources.

**Key Concept 7**

**Competition**

The Theory of Firm Behavior

One desirable outcome of a perfectly competitive marketplace is the efficient use of resources. The characteristics of the model of perfect competition are many buyers and sellers, a standardized product, mobile resources, and perfect information. These four characteristics guarantee that risk-adjusted rates of return will be equal to the normal rate of return for the economy, that prices are equal to minimum average cost of production, and that all transactions beneficial to both buyer and seller will take place.

Every firm must decide how much to produce and what price to charge. The choice of an output level and a pricing strategy are ultimately determined by the firm’s costs. In a perfectly competitive market, the pricing decision is easy, because the product is standardized and firms must follow the dictates of the market. Firms that charge more than the market price lose customers. At the other extreme, firms have no incentive to charge a lower price, because they find willing customers at the market price. Firms are called price takers.

Figure 2.6 provides an illustration of the perfectly competitive market. Market price is determined by the interaction of supply and demand in part (a). At the price $P_0$, the representative firm can sell all it can produce. A profit maximizer will
produce every unit of output when the selling price is greater than the marginal cost of production—as long as $P_0$ is greater than $MC$. Because the competitive firm is a price taker, its demand curve is perfectly elastic at the market-determined price. In the case of a horizontal demand curve, the firm’s marginal revenue ($MR$) curve is equal to price. Profit is maximized where $MR = MC$, or at $q_0$ units of output.

Competitive forces will lead to prices equilibrating at minimum average costs. At a price above $P_0$, price is greater than the average cost of production. Firms enjoy excess profits, or higher than normal rates of return, which encourages the entry of new firms into the market. As these new entrants establish their presence, supply increases and prices fall, until all excess profits are eliminated.

**Price Ceilings and Price Floors**

In their zeal to control rising prices, policy makers are sometimes tempted to pursue a price-fixing strategy. If prices are currently too high, why not roll them back to lower levels? Simply legislate a price that is below the current equilibrium price and make the product more affordable. In Figure 2.7, suppose the legislature sets a maximum price of $P_c$ below the equilibrium price $P_e$. This price ceiling does two things: it reduces the availability of medical care from $Q_e$ to $Q_s$, and it increases the amount requested to $Q_d$. The difference between $Q_d$ and $Q_s$ represents a shortage in the medical market. The shortage manifests itself in terms of longer delays in getting appointments, longer waits at physicians’ offices, reduced access to high-tech surgical and diagnostic equipment, and lower quality of care.

Suppose the market for unskilled labor is depicted in Figure 2.8. Without government intervention, firms pay workers $W_e$ and employ $Q_e$. If the government raises the cost of hiring workers by mandating that all firms provide health insurance for their employees, the cost of this new benefit raises the effective wage to $W_f$. This price floor reduces quantity demanded and increases quantity supplied. The job losers, when added to the new entrants, add to the number of unemployed workers in the labor market. Workers who keep their jobs are better off, but those who lose their jobs because of the mandate are noticeably worse off.

Policy makers are desperate to control medical care spending. Many feel that desperate times call for desperate measures. Some even think that their ability to

---

**Key Concept 8**

**Efficiency**

---

**Figure 2.7: Price Ceiling**

![Price Ceiling Diagram](Image)
write laws also applies to the laws of supply and demand. Governments have been trying for centuries to rewrite those laws, and have all failed miserably.\textsuperscript{8}

\textbf{The Impact of an Excise Tax}

The excise tax is becoming an increasingly popular way of imposing user fees on the consumption of specific items, such as gasoline, tobacco, and alcohol. Excise taxes may be set at a fixed dollar amount or at a percentage of selling price, called either a \textit{specific tax} or an \textit{ad valorem tax}.

In a competitive market, depicted in Figure 2.9, price and output are determined by the interaction of supply and demand. The commodity will sell for the price $P_e$, and $Q_e$ will be purchased. An excise tax of a fixed amount will raise the cost of providing the commodity to the market and shift the supply curve leftward to the curve labeled $S + \text{tax}$. The dollar magnitude of the shift, measured by the vertical distance between the two supply curves, will be exactly equal to the specific tax.

The new equilibrium price will be $P_c$. Because producers are legally responsible for paying the tax, they only net $P_s$ from the transaction. The difference between the price consumers pay and the price producers receive is the amount of the excise tax. At the higher price, consumers buy less of the commodity, or $Q_t$ instead of $Q_e$. The excise tax generates revenues for the government of $P_sP_cBC$. The higher price and lower output causes a loss in surplus value—a deadweight loss from the tax of $ABC$.

The impact of this loss is minimized when the lost output is small; that is, when the demand curve is inelastic. It should come as no surprise that excise taxes on cigarettes, alcohol, health insurance, and hospital stays have been proposed as financing alternatives for the various health care reform options. Whenever taxes on alcohol and cigarettes are discussed, the tax is often called a \textit{sin tax}.

\begin{itemize}
\end{itemize}
Using Game Theory to Study Economic Behavior

Game theory is a branch of applied mathematics used by economists to study strategic behavior. As individuals we interact with parents, children, siblings, spouses, friends, rivals, and colleagues, and we often find it useful to behave strategically. Strategic behavior is practiced in business, policy making, international diplomacy, and anywhere else interactive decision making takes place. The study of game theory attempts to build on strategic ability to develop a systematic approach to strategic behavior and improve strategic skills. Game theory is not a game. It involves more important issues in economics, adding another dimension to the foundational assumption of rational behavior — the interaction of two or more rational decision makers.

When considering strategic games, we frequently think of head-to-head interaction between two rivals. The prevalent view in economics is that competition improves all outcomes. Competitive markets are more efficient, prices are lower, and everyone is better off. Game theory goes beyond the simple interaction of supply and demand in the standard competitive model. No longer are we dealing with the impersonal market but with interpersonal strategic interaction between two decision makers.

Interaction can be either sequential or simultaneous. Players can take turns, each waiting to see what the other does before responding, or they can choose without prior knowledge of the other’s decisions. Gambling is a zero-sum game; one person’s winnings are the other person’s losses. International trade is not zero-sum, because both nations generally benefit from increased economic activity. Some games are played one time, some are repeated. Sometimes information is equally available to all players, continued
often it is asymmetrically distributed. Game theory is used to explain past events, predict future events, and advise players on the appropriate strategies under different circumstances.

The classic case of the simultaneous game is the prisoner’s dilemma. The payoff structure of the prisoner’s dilemma is important, because it arises in many strategic situations and thus has a wide range of applicability. The payoff matrix below depicts the predicament that two bank robbers, Bonnie and Clyde, find themselves in once captured. They are placed in separate interrogation rooms and given an opportunity to provide evidence against the other for a reduced prison sentence.

<table>
<thead>
<tr>
<th></th>
<th>Confess</th>
<th>Deny</th>
</tr>
</thead>
<tbody>
<tr>
<td>Confess</td>
<td>20, 20</td>
<td>1, 30</td>
</tr>
<tr>
<td>Deny</td>
<td>30, 1</td>
<td>5, 5</td>
</tr>
</tbody>
</table>

First, examine the situation from Bonnie’s perspective. The payoff matrix represents the length of her prison sentence if both confess (20 years), if Clyde confesses and she does not (30 years), if she confesses and Clyde does not (1 year), and if they both choose not to confess (5 years). Even if they agreed prior to their arrest to never confess their crimes, what should she do now that they are both confronted with the opportunity to limit their sentences by confessing? Does she really trust Clyde not to confess when confronted with the same payoffs?

The prudent strategy in this situation, and Bonnie’s best response, is to base her decision on what is best for her regardless of Clyde’s choice. If Clyde confesses, Bonnie will spend 20 years in prison if she confesses, 30 if she does not. It is better to confess. If Clyde does not confess, Bonnie will spend 1 year in prison if she confesses and 5 if she does not. It’s better to confess. Bonnie is said to have a dominant strategy; regardless of Clyde’s decision, she spends less time in prison if she confesses. With this payoff structure, Clyde is faced with the same situation, so his dominant strategy is to confess. When both follow their dominant strategy, we reach a Nash equilibrium in which both confess and go to prison for 20 years.*

Regardless of the circumstances, the pursuit of the dominant strategy in a prisoner’s dilemma results in lower payoff. Even though cooperative behavior would result in a higher payoff, the consequences of the other’s defection are too great to take the risk. How do you avoid the consequences of opportunistic behavior? What can you do to guarantee a better outcome?

*John Nash won a Nobel Prize in Economics in 1994 for his contribution to economics in game theory.

Welfare Implications

Consider another way to look at demand and supply curves. Instead of viewing the demand curve as the amount demanded at various prices, it can be interpreted as the maximum price that consumers are willing to pay for each unit of a product. Likewise, the supply curve can be interpreted as the minimum price that providers are willing to accept for each unit of a product. From this perspective, demand curves may be viewed as “willingness-to-pay” curves and supply curves as “willingness-to-provide” curves.
**Consumer Surplus** Value depends on the consumer’s willingness to pay. Items are valued for the utility they provide when purchased and consumed. In free markets, consumers do not pay more for a good than the subjective value they place on it. In fact, much of the time the value placed on an item exceeds its price. In those instances in which value exceeds price, consumers enjoy surplus value, or what is called *consumer surplus*.

In Figure 2.10, the demand curve $DD'$ represents the maximum price that consumers are willing to pay to obtain a good, which is its subjective value. At the equilibrium price $P_0$, consumer surplus is depicted as the difference between the value consumers place on the good, shown by the demand curve itself, and the price they must pay ($P_0$). All $Q_0$ units of output sold have surplus value. The triangular area between the demand curve and the price, $P_0 \triangleq AD$, shows total consumer surplus.

**Producer Surplus** In the case of voluntary exchange, surplus value is created for both consumers and producers. A producer’s willingness to provide goods and services is determined to a great extent by the opportunity cost of the resources used in production. Supply curves reflect these forgone opportunities. Producer surplus is defined as the difference between the price that is received and the minimum price that producers are willing to accept. Graphically, producer surplus is the area below the equilibrium price ($P_0$) and above the supply curve ($SS'$). Total producer surplus is the triangular area $P_0 \triangleq AS$. Any output level other than $P_0$ results in a loss of surplus value and represents lost social welfare. In other words, given the demand and supply curves, $DD'$ and $SS'$, any price other than the perfectly competitive equilibrium price $P_0$ represents an inefficient outcome.

**Imperfect Competition** In the case of the medical marketplace, violations of the assumptions of perfect competition are common. Although the incidence of monopoly is rare, the number of providers often falls far short of the perfectly competitive ideal. For example, many communities around the United States are served by a single hospital. Many factors determine the strength of this monopoly status; among them are the relative ease of access to other hospitals and the urgency of the services provided. Monopoly power

---

**Figure 2.10: Consumer and Producer Surplus**

*Remember, the supply curve represents the subjective value providers place on the resources used to produce the good or service—it's opportunity cost.*
leads to monopoly returns, or excess payments. In the hospital industry, these extra payments are used to cross-subsidize care for the indigent population.

Other violations of the assumptions of the perfectly competitive model include entry restrictions that limit the number of providers that can practice in a particular area. These restrictions come in the form of certification requirements, such as compulsory licensure for physicians, and by limiting hospital privileges to certain providers. Information costs—in particular, unequal distribution of information between patient and provider—also present impediments to the market.

Supply-Side Imperfections

Imperfections on the supply side of the market allow providers to enjoy monopoly returns. These imperfections usually deal with the nature of the rivalry, or the lack of rivalry, among firms. Too few firms, a nonstandardized product, barriers to entry, and information problems manifest themselves in the medical marketplace.

The presence of a single firm in a market is referred to as monopoly. As the sole provider in a market, monopolists have market power—the ability to set a price. This market power is inversely related to the elasticity of demand for whatever the monopolist is selling. The more inelastic the demand, the greater the market power.

Monopolists enjoy their special position in the market because, for various reasons, rivals are prevented from competing effectively. Barriers to entry may be the result of cost advantages due to size, something economists call economies of scale. Barriers may exist because of the sole ownership of an essential input in the production process or the franchise rights to a particular geographic region. These barriers can arise naturally or can result from legal restrictions on competitors. Whatever the source of the monopoly power, the result is a single provider serving a given market.

Monopoly is really quite rare in the U.S. economy, even in the medical marketplace. A more likely scenario is oligopoly, or the presence of a few firms in a market. The most important aspect of oligopolistic markets is the nature of the rivalry among firms. The pricing and output decisions of one firm depend on those of its rivals. The recent wave of consolidations in the managed care industry is bringing this form of market organization into the spotlight.

A single firm, or even a small number of firms, does not dominate many local markets, especially those that deal in services. Often many small firms attempt to differentiate themselves from their competitors by serving these markets by various means. Successful differentiation leads to market power. The degree of market power depends on how different the product is from its alternatives. A market with a large number of suppliers selling a variety of similar products is classified as monopolistic competition.

In all cases of imperfect competition, the firms share a common characteristic: they face downward-sloping demand curves. Firms in perfectly competitive markets, facing horizontal demand curves, have no market power: they are price takers. Whenever a demand curve is downward sloping, the pricing strategy changes. Market power allows firms to set a higher price, one that increases profit. Firms that find themselves in this situation are called price searchers.

Figure 2.11 illustrates the pricing and output strategy of a price searcher.10 Faced with a downward-sloping demand curve, the firm must choose the profit-maximizing price and quantity. The price searcher is confronted with a marginal revenue curve that is situated below the downward-sloping demand curve. When the demand curve is downward sloping, the firm must lower the price to sell more

---

10 The model discussed here is that of the single-price monopolist, one that sells to each customer at the same price. Other pricing strategies include price discrimination, in which different consumers are charged different prices depending on their price elasticity of demand.
of the product. As a result, the extra revenue from the sale of one more unit of output is less than its price. To sell the extra unit of output, the provider must lower the price on all the output that could have been sold at a higher price. In other words, the marginal revenue curve is below the demand curve. It has the same intercept on the price axis and twice the slope.\textsuperscript{11} Although the rule of thumb for profit maximization is the same, \( MR = MC \), the intersection takes place below the demand curve. So the profit-maximizing output is lower than in the case of perfect competition, and the resulting price is higher.

Whether the price searcher makes a profit depends a great deal on the nature of the entry barriers. A monopolist can expect to maintain profits as long as the level of demand is maintained. In contrast, firms in monopolistic competition will see profits eliminated, because profits attract competitors, and competition for market share results in lower prices, higher costs, and lower profits.

**Demand-Side Imperfections** On the demand side of the market, imperfections manifest themselves in a number of ways: a limited number of buyers and imperfect information are two possibilities. The classic case of demand-side imperfections is called monopsony, or a single buyer. This situation emerges in medical care when consumers form into groups to consolidate their purchasing power and get lower prices from insurers and providers. The Canadian single-payer system is an example of a monopsony.

As sole purchaser in the market, the monopsonist faces an upward-sloping supply curve and a marginal cost curve that is above the supply curve. Figure 2.12 illustrates the operation of a market with a single buyer. Faced with an upward-sloping supply curve, the monopsonist must pay increasingly higher prices to obtain more output, even on those items that could have been purchased at lower prices if less had been bought. The relevant purchasing decision takes into consideration the marginal cost of purchasing one more unit of output, not the opportunity cost of that last unit of output. Instead of equilibrium occurring where supply and demand are equal, the monopsonist equates marginal cost with demand.

\textsuperscript{11} A mathematical proof of this proposition follows:

\begin{align*}
\text{Demand curve:} & \quad P = a + bQ \\
\text{Total revenue:} & \quad TR = P \times Q = (a + bQ) \times Q = aQ + bQ^2 \\
\text{Marginal revenue:} & \quad MR = \frac{dTR}{dQ} = a + 2bQ
\end{align*}
Economists seldom hesitate in applying economic tools in a variety of circumstances to evaluate individual choice and behavior. This tendency should not be misinterpreted. Few members of the economics profession believe that economics provides all the answers. As you progress through the book, it will become obvious that the health care marketplace fails to achieve its theoretical optimum in many cases, making the strict application of the neoclassical model inappropriate. The goal of this book, however, is to show that economics can provide insights into the study of human decision making that few other disciplines offer.

The central message of economics presented in this chapter can be stated briefly:

- **Key Concept 1**
  *Scarcity and Choice*

- **Key Concept 2**
  *Opportunity Cost*

- **Key Concept 3**
  *Marginal Analysis*

Monopsony equilibrium occurs at a lower level of output and a lower price than in the case of perfect competition. Society is worse off because fewer services are provided. At the lower price, quantity demanded \( (Q_D) \) exceeds quantity supplied \( (Q_s) \). The monopsonist exercises market power and creates a shortage that is not eliminated by competition with other purchasers, because none exists.

**Key Concept 1**

**Scarcity and Choice**

**Key Concept 2**

**Opportunity Cost**

**Key Concept 3**

**Marginal Analysis**

**SUMMARY AND CONCLUSIONS**

- It is important to strike a balance between incremental benefits and incremental costs. Most choices in medical care involve determining the level of an activity, not its very existence. The issue is not whether it is beneficial to perform widespread screenings for colon cancer, but whether it is cost effective to perform a sixth test, when five have already been done (Neuhauser and Lewicki, 1975). Decision making is seldom based on an all-or-nothing proposition. It usually involves a trade-off. If we are to spend a little more on one thing, we must be willing to spend a little less on something else.

- Human behavior is responsive to incentives and constraints. If you want people to practice economizing behavior, they must benefit individually from their own economizing. People spending other people's money show little concern for how it is spent. People spending their own money spend it more wisely.
As concern over escalating costs grows, economics takes on an increasingly important role in the study of medical issues. Future clinicians must be well-grounded in economic theory. Only then can they help shape the debate on the future direction of medical care delivery.

1. What are the likely consequences on the U.S. market for tobacco products for each of the events listed below? Would the supply curve or the demand curve shift? In which direction? State whether the equilibrium price and quantity would increase, decrease, or stay the same. Show the changes using a standard diagram with an upward-sloping supply curve and a downward-sloping demand curve.
   a. The Food and Drug Administration classifies tobacco an “addictive substance.”
   b. The Congress votes to raise the excise tax on all tobacco products.
   c. Hurricane Fran dumps 15 inches of rain on North Carolina and destroys 80 percent of that state’s tobacco crop.
   d. Sixteen states sue the major tobacco companies for billions of dollars because of tobacco-related costs in their Medicaid programs.
   e. Medical evidence that more than two cups of coffee a day, considered by many to be a substitute for smoking, greatly increases the risk of stomach cancer.

2. What is the proper role of economics in the study of health and medical care? What does economics have to offer? What are its limitations?

3. “The laws of supply and demand are immutable. No one, including government, can affect a commodity’s demand curve or supply curve.” True or false. Comment.

4. Indicate whether the following statements are positive or normative.
   a. Smokers should pay higher health insurance premiums than nonsmokers.
   b. The United States should enact a comprehensive health care plan that provides universal coverage for all Americans regardless of their ability to pay.
   c. The primary reason for the escalation in health care spending over the past 30 years has been the rapid development of expensive medical technology.
   d. The high cost of providing health care for employees is a major reason U.S. firms are not competitive with their foreign counterparts.
   e. Individuals born with certain genetic defects that predispose them to higher medical care spending over their lifetimes should be charged higher health insurance premiums than people without those defects.

5. This problem is based on material discussed in Appendix 2B. The relationship between health care spending (E) and per capita national income (Y) was estimated using cross-section data from 24 developed countries. The resulting equation \( \hat{E} = 200 + 0.09 \hat{Y} \) relates spending and income in U.S. dollars.
   a. Interpret the coefficient on the national income variable.
   b. Complete the table.
<table>
<thead>
<tr>
<th>Income in $</th>
<th>Health Care Spending</th>
</tr>
</thead>
<tbody>
<tr>
<td>5,000</td>
<td>250</td>
</tr>
<tr>
<td>10,000</td>
<td>290</td>
</tr>
<tr>
<td>15,000</td>
<td>330</td>
</tr>
<tr>
<td>20,000</td>
<td>370</td>
</tr>
<tr>
<td>25,000</td>
<td>410</td>
</tr>
</tbody>
</table>
   c. Graph the relationship

---

**Professor Kenneth J. Arrow**

Kenneth J. Arrow, known primarily for his work on general equilibrium and welfare economics, wrote what is considered by many to be one of the classic articles in the field of health economics. “Uncertainty and the Welfare Economics of Medical Care” (*American Economic Review*, 1963)

continued
Part 1 >> The Relevance of Economics in Health and Medical Care

has had as much impact on economic thinking as any single paper written in the modern era. Members of the International Health Economics Association considered his contribution so important that they named their annual award for the outstanding published paper in health economics after him.

Born of immigrant parents in 1921, Arrow spent his early childhood in relatively comfortable surroundings. His father's business, however, fared poorly during the Great Depression, forcing Arrow to attend City College, which was free at that time to residents of New York. After graduating at the age of 19, and unable to get a job, he decided to pursue graduate studies in statistics at Columbia. Even though his interests were in mathematical statistics, he switched to economics to receive financial aid. He soon discovered his interest in economics surpassed his love for statistics.

Arrow's early work completely revolutionized the way economists think about general equilibrium and social choice. Winner of the 1972 Nobel Prize in Economics at the age of 51, he is widely considered one of the most important figures in general economic equilibrium theory and welfare theory.

In his own words, he describes his contribution to health economics as “not so much a specific and well-defined technical accomplishment as a point of view that has served to reorient economic theory” (Breit and Spencer, 1995). Arrow's work to integrate uncertainty into economic models led to his 1963 paper on the economics of medical care. In it he was able to show that the key element in insurance markets was the difference in information between the buyers and sellers of insurance. The very existence of health insurance causes individuals to spend more on medical care than they would otherwise. His emphasis on moral hazard and adverse selection served to focus research in health economics on these important issues.

Arrow joined the U.S. Air Force during the Second World War and served as a weather officer. His wartime contribution included important work on long-distance flight planning. At the time, the important theoretical work was all based on the assumption of a flat earth. Arrow's reformulation took into consideration the true nature of flight in a spherical world and helped determine optimal flight paths. After almost five years in the military, and still in his mid-twenties, he returned to Columbia University to finish his graduate studies. Before receiving his Ph.D., Arrow joined the Cowles Commission at the University of Chicago but soon moved to Stanford University, where he became a full professor at age 32. By the end of his first decade in academics, he was named president of the Econometric Society and winner of the John Bates Clark medal, given by the American Economic Association for the most distinguished work by an economist under the age of 40.

Most of his academic career has been spent at Stanford, except for 11 years at Harvard. He returned to Stanford in 1979, where he is currently emeritus Professor of Economics. In 1981, Arrow was named Senior Fellow at the Hoover Institution. In addition to his many honors and affiliations, he has been President of the American Economic Association, the Institute of Management Sciences, the Western Economic Association, the American Association for the Advancement of Science, and the International Economic Association. Often quoted and frequently criticized, his work has been so far reaching that we may never fully appreciate the extent of his contribution to economic and political thought.

REFERENCES


Graphing Data

Someone once said that a picture is worth a thousand words. Economists must take this axiom to heart. Seldom will an economist get far into a discussion without reaching for a pencil and paper. The picture often takes the form of a graph, one of several ways that economists use to convey ideas.

Some Basics of Graphing

Most graphs that we use in economics are two-variable graphs. The relationship between the two variables is illustrated by drawing two axes perpendicular to each other. The dependent variable is usually plotted on the vertical, or y axis; the independent variable on the horizontal, or x axis. Point a in Figure 2A.1 represents a combination of the variables x and y equal to x₀ and y₀, respectively. The x-y values for point a are called the coordinates of point a.

Graphs are used to describe relationships between variables. Scatter diagrams are often used for this purpose. The scatter diagram in Figure 2A.1 suggests that variable x and variable y are associated with one another; as the value of x increases, the corresponding values of y are also larger. Economists use scatter diagrams to get a feel for the relationship between two variables, looking for linkages, a correlation, or simply a random pattern.

When a relationship between variables is hypothesized, it is often depicted by a linear function or curve. Straight-line relationships can be expressed by the familiar equation $y = mx + b$, where m is the slope of the line and b is its y intercept. Graphically,
this relationship is shown in part (a) of Figure 2A.2. The slope of a straight line is calculated by dividing the change in the variable on the $y$ axis ($\Delta y$) by the change in the variable on the $x$ axis ($\Delta x$). The slope of the curve in part (b) below is determined by the slope of its tangent, a straight line that touches the curve at only one point.

The slope of a function or curve is a convenient way to describe the relationship between two variables. A slope of $1.3.0$ indicates that for every one unit increase in the variable measured on the $x$ axis, the variable on the $y$ axis increases by 3. The intercept represents the value of the variable measured on the axis $y$ when the variable on the $x$ axis has a value of zero.

### Functional Relationships

Graphs are an efficient means of expressing relationships between variables. Often the relationship between two variables is functional in nature, implying dependence or causation. A causal relationship has a dependent and an independent variable. The value of the dependent variable is determined by the value of the independent variable. Suppose that we want to examine the relationship between the amount of money spent on medical care and the health of a person or a group of people. Instead of spending one or two pages of valuable paper describing this relationship, I can simply use a graph to convey the main idea.

Figure 2A.3 indicates that there is a direct (positive) relationship between the level of health and the amount spent on medical care. The higher the level of spending, the healthier the person or population. The shape of the line indicates that there is a limit to how much health you can buy with increased medical care spending. Additional medical spending buys progressively smaller increments of health. There are other variables that affect the relationship between health and medical spending, such as genetics and lifestyle choices. Smokers as a group experience more respiratory and circulatory problems than nonsmokers. Figure 2A.4 depicts the relationship between the level of health and medical spending for smokers and nonsmokers. The graph indicates that at any given level of spending, nonsmokers are healthier than smokers on average.

Sometimes two variables are indirectly (negatively) related to one another. The relationship between infant mortality rates and birth weights is a good example of this phenomenon. Empirical data suggest that as birth weight increases, mortality rates decline. Figure 2A.5 illustrates the negative relationship between infant mortality and birth weight category. Some hypotheses question whether high mortality rates are due to low birth weights or some other factor, such as prematurity (Behrman, 1995). Those issues will be discussed later.
For now, focus your attention on the nature of the relationship and how to depict it graphically.

As we discussed earlier, one of the important concepts in economics is optimization. Efficient production techniques promote the goals of average cost minimization. Optimal pricing strategies enable firms to maximize profits. Graphs showing a minimum or a maximum are illustrated in Figure 2A.6.

Part (a) illustrates the hypothetical relationship between the average cost of services and the number of beds in a typical community hospital. This U-shaped relationship is typical of average costs in producing a product or service. As the size of the operation increases, average costs decrease. If the operation expands beyond a certain level, average costs begin to increase. The most efficient level of operation for the hospital, the optimal level, is $B_0$.

A functional relationship with a maximum is shown in part (b). Here the relationship between the total revenues of a physician’s practice and the number of patient visits is illustrated. To generate more patient visits, a physician must offer discount prices to some groups—a practice that is typical for physicians who participate in managed care networks. What is the optimal pricing policy? A physician trying to maximize total revenue will charge a price that will result in a volume of business equal to $V_0$.

**Time-Series Graphs**

On occasion it is important to examine how variables change over time. The use of longitudinal, or time-series, graphs often illustrates trends in a data series. Time-series graphs typically use daily, weekly, monthly, quarterly, or annual data to track changes.
in an economic variable. Figure 2A.7 graphs the changes in U.S. health care spending over three decades. Health care spending has shown a long-term upward trend since 1970. Starting at less than $100 billion, it has risen dramatically to almost 20 times that amount in just over three decades.

If we were interested in examining the relationship between health care spending and income, we could collect data on spending and income in a single country over a number of years. While a time series on two variables provides insight into the relationship, so many other factors change over time that we may not be sure of our results. Figure 2A.8 illustrates a time-series relationship between per capita health care spending and per capita gross domestic product (GDP) in the United States between 1970 and 2001.

### Cross-Section Graphs

Another approach to graphing the same relationship is the use of cross-section data. A cross-section graph provides a number of observations on two variables at a given point in time across different entities: individuals, firms, states, or countries. Figure 2A.9 illustrates the same relationship for the year 2000 using data from the Organization for Economic Cooperation and Development (OECD). The two graphs depict the relationship between income and spending. Each point on the time-series graph shows U.S. spending compared to income over a number of years. The cross-section graph shows the same two variables for 28 different countries during a single year. Each point represents income and spending (in U.S. dollars) for a given country.

**Figure 2A.8: Per Capita Health Care Spending and Per Capita GDP (United States, 1970–2001)**

**Figure 2A.9: Relationship between Health Care Spending and Income in OECD Countries (Per Capita Amounts, 2000)**

Statistical Tools

**DESCRIPTIVE STATISTICS**

Whenever we are confronted with a body of data, the challenge is how to summarize the relevant information to make it useful to the reader. Economic researchers are often confronted with large amounts of data, hundreds and sometimes thousands of observations on a number of variables. A useful way of summarizing large amounts of data is by way of a graph, sometimes called a **histogram**.

Figure 2B.1 shows the distribution of maternity patients by age at Hillcrest Baptist Memorial Hospital in Waco, Texas, for 1991. A simple viewing of the histogram tells us much about the ages of the 2,476 mothers who delivered that year. The youngest was 12 years old, the oldest 44—a spread of 32 years. The most frequent age was 25 years, the approximate center of the distribution.

Histograms can be summarized by statistical measures. These statistical measures help define the center of the distribution and the spread around the center. These concepts are formally called **central tendency** and **dispersion**.

**Measures of Central Tendency**

Measures of central tendency are often used to describe the typical value in a data set. The most commonly used measure of central tendency is the **mean**. Often referred to as the **average**, the mean of a distribution is the sum of the individual values divided by the total number of cases. Summing the ages for the maternity patients comes to 64,137 years. Dividing by the total number of patients (2,476) gives a mean value of 25.9 years.

Reporting the mean value as the typical value can be misleading, because it may place too much...
weight on extreme values. Suppose five infants were born on a given day, and their mothers were 42, 27, 25, 23, and 22 years old. The average age of these five women is

\[
\frac{42 + 27 + 25 + 23 + 23}{5} = 28 \text{ years}
\]

By weighting the observations equally, the 42-year-old causes the measure of central tendency, or mean in this case, to be inflated and not very typical of the rest of the data.

When dealing with data that has a relatively small number of unusually large or small numbers, many researchers use an alternative measure of central tendency known as the median. The median is a popular summary statistic for demographic data with extreme values or outliers. To calculate the median, the values of a group of numbers are ranked from largest to smallest. In the case of an odd number of observations, the median is the middle number. In the case of an even number of observations, the median is the average of the middle two values. Its position at the fiftieth percentile implies that exactly half of the distribution falls above the median and half falls below it. The median age of the five new mothers listed above is 25 years, a much better indication of the typical age of that sample of patients. The median for all 2,476 maternity patients is 26 years.

Another measure of central tendency is the mode. The mode is the value occurring most frequently in the distribution. The most common age of the five maternity patients listed above is 23. For the entire group it is 25. The mode is used primarily on those occasions where the distribution has more than one mode. Under these circumstances, care should be taken to understand what is truly typical of the data values. Confounding factors may cause measures of central tendency to convey quite different results concerning the overall data set. Without controlling for these confounding factors, reliance on a single measure of central tendency may produce spurious results.

**Measures of Dispersion**

Focusing on the central tendency can obscure other interesting features of a collection of numbers. Concentrating on averages would lead us to conclude that a person standing with one foot in a bucket of scalding hot water and the other foot in a bucket of ice water is, on average, comfortable. Instead of simply looking at the central tendency of the data, it is useful to examine the way the numbers spread out around the center or average. Deviations around the average are typically indexed by statistical measures termed the variance and the standard deviation.

The **variance** is a measure of the dispersion of the data around the mean (average) value. It is one way of describing how closely individual observations in a data set cluster around the mean. The sample variance, denoted \( s^2 \), is calculated as follows:

\[
s^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{X})^2}{n-1}
\]

where \( X_i \) is the “ith” observation of the variable \( X \), \( \bar{X} \) is the sample mean, and \( n \) is the number of observations in the sample. The deviations from the mean, \( x_i - \bar{X} \), are squared to take into consideration all values above or below the mean. Otherwise, deviations for values below the mean would enter the numerator as negative numbers and result in an artificially low measure of dispersion. Whenever the values of a variable are similar, the variance will be small. Variance, or the variability in the observed values, is a key concept in statistics and plays an important role in the calculation of many statistical tests and procedures. In fact, one of the goals in empirical research is to explain as much of the variance as is practicable.

A related measure of dispersion around the mean is the **standard deviation**. Even though the variance is computed in terms of squared values of the deviations, the standard deviation measures the average deviation; it is an estimate of how far on average the values are from the mean value. Mathematically, the standard deviation is the square root of the variance. This measure of deviation has more intuitive appeal, because it is measured in the same units as the original variable. If the variable being considered is years, variance is measured in square years and standard deviation in years. For our sample of maternity patients, the variance is 28.6 square years, and the standard deviation is 5.3 years.

Another common issue concerning a distribution is its shape. A distributions that is symmetrical is often called a **normal distribution**. A distribution that has a long tail is called a **skewed distribution** (see Figure 2B.2 above). A normal distribution is bell shaped and can be reconstructed rather well from its summary statistics, mean and standard deviation. For a normal distribution, roughly 70 percent of the...
observations fall within plus-or-minus one standard deviation of the mean, and about 95 percent fall within two. For our maternity patients, over 72 percent fall within one standard deviation (± 5.3 years) of the mean, 25.9 years. In other words, 1,787 of the 2,476 patients are between the ages of 20 and 31 years. Additionally, over 96 percent (2,386 out of 2,476) are between the ages of 15 and 36 years, or two standard deviations from the mean.

**CORRELATION**

Descriptive statistics are useful when dealing with one variable at a time. However, a study of the relationship between two or more variables is more interesting and requires other techniques. The scatter diagram described in Appendix 2A is one way of examining the relationship between two variables (see Figure 2A.1). Consider the points on a scatter diagram: A tight clustering around a straight line indicates a strong linear association between the two variables. A loose clustering indicates a weak linear association.

The strength of the association can be measured by a summary statistic commonly called the **correlation coefficient**. The correlation coefficient may be visualized as an expression of how two variables are “co-related.” It is calculated using the respective standard deviations and means of the variables. Practically speaking, a perfect correlation between two variables indicates that all the observations lie on a straight line that is either positively sloped or negatively sloped. In these two cases, the correlation coefficient will have the value of either +1 or −1. If the two variables show no tendency to increase or decrease together, the points on a scatter diagram will show no clustering. In such cases, the correlation coefficient will have the value of zero.

It is important to understand that a correlation coefficient indicates an association between two variables. Association, however, does not imply causation. Suppose researchers found a strong negative correlation between the number of cases of influenza and the amount of ice cream consumed. Could we say that eating ice cream reduces the incidence of influenza? As popular as this would be with the children of the world, we cannot honestly make the statement. If it were true, physicians would encourage the consumption of ice cream to reduce the chances of contracting an influenza virus.

Correlation may be telling us that there is a third factor at work in the influenza–ice cream connection: namely, the season of the year. Coincidentally, the flu is most prevalent during the winter months, when ice cream sales are low, and least prevalent during the summer months, when ice cream sales are high. Correlation says nothing about these confounding factors. If it were possible to control for all of these confounding factors, correlation would provide a much stronger argument for causation. What is needed is a way of controlling for these other factors.

**REGRESSION**

Simple measures of central tendency and dispersion reveal little about the way two or more variables are “co-related.” An empirical technique used to determine the nature of the **statistical relationship** among a dependent variable and one or more independent variables is called **regression analysis**. Regression analysis not only allows us to identify systematic relationships among variables, it provides estimates of the relative magnitude of the various relationships. The relationships may be discussed in terms of independent and dependent variables, stimuli and response, explanatory and explained variables, or cause and effect. Because it is one of the most frequently used empirical techniques in economic research, it is important to have a clear understanding of this powerful tool.
Least Squares Methodology

Regression analysis is used to identify a dependent relation of one variable or a set of variables to another. Most regression models use the least squares method for estimating parameters. The least squares method provides a means of fitting a curve to a set of data points. This technique is not without its methodological problems. Moving the line closer to some points moves it farther away from other points. Solving the problem is simple. First, find the average distance from the line to all points. Second, minimize the average distance. The least squares method uses this approach with one difference: instead of using the average distance, it uses the average of the squared distance. This approach avoids the problem of positive and negative differences canceling each other out, hence the name ordinary least squares.

Suppose we are interested in examining the causes of increased health care spending. The first step in our analysis is to specify the variables to include in the model. The variables that influence health care spending are numerous and may include income, age, and gender among other things. To simplify our discussion, we will specify a simple regression model with one dependent variable and one independent variable. The dependent variable is health care spending and the independent variable is income.

Step two in the analysis involves collecting reliable estimates for the two variables. Two approaches are possible: time series and cross section. A time-series approach would require the collection of data over time, locating data from a published source that looks at spending and income over time for a single entity, such as a state, region, or country. A cross-section approach requires data from a number of entities during a single time period.

Data for a cross-section analysis of the effect of income on spending is provided in Table 2B.1. The data come from the Organization of Economic Cooperation and Development for 28 developed nations. Income is defined as per capita GDP, and spending is defined in per capita terms. All values are translated into U.S. dollars using purchasing power parity exchange rates.

After collecting the data, the third step is to decide on the functional form of the relationship, or the regression equation. The two standard choices are linear and multiplicative. The linear form can be written as \( E = a + bY \), where \( E \) is per capita health care expenditures and is per capita GDP. The multiplicative form can be written \( E = aY^b \). Although the linear model is simpler, the multiplicative form has its advantages. One advantage is the coefficient \( b \) in a log transformation of the equation \( \log E = \log a + b \log y \) has a simple economic interpretation—it is an estimate of “income elasticity.”

<table>
<thead>
<tr>
<th>Country</th>
<th>Per Capita GDP</th>
<th>Per Capita Spending</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mexico</td>
<td>8,845</td>
<td>491</td>
</tr>
<tr>
<td>Slovak Republic</td>
<td>11,279</td>
<td>641</td>
</tr>
<tr>
<td>Hungary</td>
<td>12,204</td>
<td>817</td>
</tr>
<tr>
<td>Czech Republic</td>
<td>13,802</td>
<td>987</td>
</tr>
<tr>
<td>Korea</td>
<td>15,186</td>
<td>893</td>
</tr>
<tr>
<td>Greece</td>
<td>16,481</td>
<td>1,556</td>
</tr>
<tr>
<td>Portugal</td>
<td>16,857</td>
<td>1,519</td>
</tr>
<tr>
<td>Spain</td>
<td>20,080</td>
<td>1,497</td>
</tr>
<tr>
<td>New Zealand</td>
<td>20,214</td>
<td>1,611</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>24,933</td>
<td>1,813</td>
</tr>
<tr>
<td>Italy</td>
<td>25,245</td>
<td>2,060</td>
</tr>
<tr>
<td>Finland</td>
<td>25,414</td>
<td>1,699</td>
</tr>
<tr>
<td>France</td>
<td>25,594</td>
<td>2,387</td>
</tr>
<tr>
<td>Japan</td>
<td>26,003</td>
<td>1,984</td>
</tr>
<tr>
<td>Sweden</td>
<td>26,146</td>
<td>2,195</td>
</tr>
<tr>
<td>Belgium</td>
<td>26,239</td>
<td>2,293</td>
</tr>
<tr>
<td>Germany</td>
<td>26,269</td>
<td>2,780</td>
</tr>
<tr>
<td>Austria</td>
<td>26,473</td>
<td>2,350</td>
</tr>
<tr>
<td>Netherlands</td>
<td>27,183</td>
<td>2,348</td>
</tr>
<tr>
<td>Australia</td>
<td>28,046</td>
<td>2,233</td>
</tr>
<tr>
<td>Iceland</td>
<td>28,139</td>
<td>2,562</td>
</tr>
<tr>
<td>Canada</td>
<td>28,187</td>
<td>2,580</td>
</tr>
<tr>
<td>Ireland</td>
<td>28,200</td>
<td>1,793</td>
</tr>
<tr>
<td>Denmark</td>
<td>28,734</td>
<td>2,398</td>
</tr>
<tr>
<td>Switzerland</td>
<td>29,553</td>
<td>3,160</td>
</tr>
<tr>
<td>United States</td>
<td>34,602</td>
<td>4,540</td>
</tr>
<tr>
<td>Norway</td>
<td>36,248</td>
<td>2,787</td>
</tr>
<tr>
<td>Luxembourg</td>
<td>48,537</td>
<td>2,719</td>
</tr>
</tbody>
</table>

Choosing the simple linear model, the simple regression model that relates per capita health care spending to per capita gross domestic product for these 28 OECD countries is written

\[ E_i = a + bY_i + u_i \]

where \( i \) represents the subscript of each observation (countries numbered 1 through 28) and \( u_i \) represents the random elements in the relationship.

Figure 2B.3 plots the actual data on spending and income provided in Table 2B.1. The constant term represents the intersection of the regression line with the axis, and the coefficient on income represents its slope. Using the least squares technique, the regression estimate predicts that, on average, for every one-dollar increase in income health care, expenditures increase ten cents.

In social science and demographic research, often more than one causal variable is identified. The technique used in this situation is called multiple regression analysis. Researchers use multiple regression to control for confounding variables; that is, other variables associated with changes in the dependent variable. For example, health care spending may also depend on other factors, such as the percentage of population covered by insurance or the number of active physicians per capita. A multiple regression equation adding these two regressors would be written in linear form as

\[ E = a + bY + cI + dP \]

where \( I \) is the percentage of the population with health insurance coverage, and is the number of active physicians per 100 population.\(^{12}\) The coefficient on the income variable would now show the independent effect of income on expenditures, free from the influence of insurance coverage and the availability of providers.

**Measures of Significance**

Foremost on the minds of researchers is the reliability of the estimated coefficients. The accuracy of a regression equation can be determined by a number of significance tests. The standard error of the estimate (SEE) is the standard deviation of the dependent variable after controlling for the influence of all the independent variables. When data points are widely dispersed about the estimated regression line,

\[ E = -17.29 + 0.084Y \]

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Coefficient</th>
<th>St. Dev.</th>
<th>t-ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-17.29</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( Y )</td>
<td>0.084</td>
<td>0.0117</td>
<td>7.15</td>
</tr>
</tbody>
</table>

\[ F = 51.2 \]

\[ R^2 = 66.3\% \]

\(^{12}\) The multiplicative form of this same equation would be \( E = aY^pP^d \).
standard error is large. If all the data points were to fall on the regression line, the standard error would be zero.

One of the objectives of regression analysis is prediction. Standard error provides an estimate of the accuracy of a prediction based on a particular regression equation. Based on statistical probabilities, when there are roughly 30 or more observations, there is a 95 percent probability that the dependent variable will lie within two standard errors of its estimated value. The smaller the standard error, the greater the confidence in the accuracy of the estimate.

Often the standard error of the estimate is used to estimate confidence intervals around a given estimated equation. The 95 percent confidence interval has a range of roughly ±2 standard errors around the estimate.

A second measure of accuracy is the coefficient of determination, or \( R^2 \). The coefficient of determination is an estimate of the percentage of variation in the dependent variable explained by the independent variables, sometimes called goodness of fit. \( R^2 \) ranges between zero and one. The higher its value, the greater the overall explanatory power of the regression equation.

Standard error and \( R^2 \) are both important significance measures, but neither addresses the question of whether the independent variables as a whole explain a significant proportion of the variation of the dependent variable. The \( F \) statistic fills this void. Values of range from zero upward. At the extreme, when \( R^2 \) equals zero, equals zero.

Whether a particular value of the statistic indicates a significant set of regressors depends not only on its value, but also on the number of regressors and the number of observations on which the estimated equation is based. In general, the larger \( F \) is, the greater the likelihood that the set of independent variables explains a significant proportion of the variance in the dependent variable.

Critical values of \( F \) are provided in statistical tables that are readily available in almost all introductory statistics textbooks. Roughly speaking, with five or fewer independent variables and 25 or more observations, values of that are greater than 3 or 4 indicate a statistically significant proportion of the variance explained by the set of independent variables. Smaller sample sizes and a larger number of independent variables require larger values of for significance.

In addition to the significance of the overall equation, often the researcher is interested in the significance of each independent variable. The standard deviation, or standard error, of the coefficient for each independent variable provides a means of creating a test statistic expressly for this purpose. The most commonly used \( t \) statistic in regression analysis is calculated to determine if an individual coefficient is statistically different from zero. The \( t \) value is calculated by dividing the coefficient estimate by its standard error. Values of \( t \) greater than 2 are usually associated with coefficients that are statistically different from zero. The critical values of the statistic are found in tables in most introductory statistics textbooks.

**SUMMARY AND CONCLUSIONS**

With the development of the microcomputer, data analysis is no longer the exclusive purview of statisticians. A standard personal computer equipped with a statistical software package gives the user a powerful set of tools for analyzing information.

The analytical techniques discussed in this appendix are among the most commonly used in the social sciences. Many of the referenced articles use them extensively. A thorough understanding of these tools will go a long way in making the study of health economics more enjoyable and easier.
A compelling argument can be made that medical care delivery is far more complex and dynamic than is typically the case in the standard treatment of the market process. The trade-off between equity and efficiency is quite acute, calling for active government involvement to ensure that the process works. Critics of government involvement offer an equally compelling argument. Even well-meaning government policy has its unintended consequences. Oversight is costly and serves to impede growth and productivity in the private sector.

In this chapter, we will examine the competitive market model and its applicability to the medical market. After considering the breakdown in the traditional market model, we will then examine how governments intervene to address the problems that arise. A general discussion of the causes and consequences of market failure will be followed by a more specific examination of market failure in medical markets. Government intervention in the form of regulation, public provision, and licensing will then be discussed. Finally, the question of how to deal with government failure is addressed.

**The Medical Care Marketplace**

As costs have escalated, and the number of uninsured individuals has increased, calls for government intervention have mounted. Proponents of more government involvement in medical care claim that medical care is far too complicated to be left to market forces. Because medicine is difficult to understand, patients must rely on their physicians’ recommendations. Others add that medical care is a social good and too important to leave to the workings of the impersonal marketplace. Some argue that the externalities involved in medicine, particularly in the area of infectious diseases, require collective action to maximize the benefits to society. Many base their support for government intervention on ethical grounds, claiming that the provision of medical care based on the ability to pay is morally repugnant. Together these arguments are responsible, in varying degrees, for the development of government-financed medical care in most developed countries throughout the world.

Those who oppose more government involvement argue that the U.S. system has remained, for the most part, market based, which is in part evidence of the deep American distrust of federal government involvement in health care matters.1

---

1 Blendon and colleagues (1995) note that only 7 percent of Americans express a “great deal of confidence” in federal health care agencies, compared with 19 percent of Canadians and 41 percent of Germans.
Experience has taught that government-run programs are costly. For example, when originally proposed in the mid-1960s, Medicaid spending was projected to reach $9 billion in 1990; the actual cost in 1990 was $109 billion. The preamble to the original Medicare bill actually prohibited any federal “supervision or control over the practice of medicine or the manner in which medical services are provided.” Anyone familiar with medical care delivery is well aware of how the federal government has violated the original intent of this legislation.

Health Care Spending
One of the major factors driving the health care reform debate is spending, including total spending, spending per person, and spending as a share of total economic output. Referring to Table 3.1, national health expenditures rose to $2,105.5 billion in 2006, 16.0 percent of GDP. Of this amount, 83.7 percent, or $1,762.0 billion, was for personal health care expenditures. This category of spending includes the purchase of all goods and services associated with individual health care, such as hospital care, the services of physicians and dentists, prescription drugs, vision care, home health care, and nursing home care.

Hospital Care  Spending on hospital services increased to $648.2 billion in 2006. Hospital costs, valued as actual revenues received, experienced five years of accelerated growth between 1987 and 1991. For much of the decade of the 1990s, the growth in hospital spending moderated due primarily to aggressive cost-control efforts on the part of private payers. From 2000 to 2005, hospital spending grew at a compound rate of 7.62 percent, increasing concerns that spending would continue to accelerate. Hospital care accounted for 36.8 percent of personal health care spending, and patients paid for approximately 3 percent of hospital care out-of-pocket.

Physicians’ Services  Spending on physicians’ services amounted to 25.4 percent of the total spent on personal health care in 2006. The total of $447.6 billion tends to mask the importance of physicians in the health care sector. Even though only 25 cents of every medical care dollar flows directly to physicians, they are indirectly responsible for most of the rest. Physicians admit patients to hospitals, recommend surgeries, prescribe drugs and eyeglasses, and in general oversee the entire health care delivery system. Roughly 10 percent of physicians’ services are financed by patient out-of-pocket payments.

Prescription Drugs and Other Medical Products  Consumers spent $216.7 billion on pharmaceuticals and another $59.3 billion on other medical products in 2006. This category accounts for 15.7 percent of personal health care spending and is one of the fastest growing categories of spending. Patients pay only 22 percent of all prescription drugs out-of-pocket.

Other Personal Health Care Spending  Other spending includes payments for dentists’ services and other professional services, nursing home care, and home health services. When combined, these categories of care accounted for approximately 18.6 percent of all personal health care spending. Nursing home care amounted to $124.9 billion and 7.3 percent of total personal health care spending in 2006, making it the fourth largest spending category. Dental services accounted for $91.5 billion, and other professional services, $58.9 billion. Home health spending at $52.7 billion has increased four times since 1990.

Prospects for the Future  The average American consumed $7,026 in medical care in 2006. At this level, United States per capita spending on medical care is anywhere
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Hospital Care</td>
<td>$ 9.2</td>
<td>$27.6</td>
<td>$101.0</td>
<td>$251.6</td>
<td>$340.7</td>
<td>$417.1</td>
<td>$525.4</td>
<td>$564.4</td>
<td>$605.5</td>
<td>648.2</td>
</tr>
<tr>
<td>Physician and Clinical Services</td>
<td>5.4</td>
<td>14.0</td>
<td>47.1</td>
<td>157.5</td>
<td>220.5</td>
<td>288.6</td>
<td>366.7</td>
<td>393.6</td>
<td>422.6</td>
<td>447.6</td>
</tr>
<tr>
<td>Dental Services</td>
<td>2.0</td>
<td>4.7</td>
<td>13.3</td>
<td>31.5</td>
<td>44.5</td>
<td>62.0</td>
<td>76.9</td>
<td>81.5</td>
<td>86.6</td>
<td>91.5</td>
</tr>
<tr>
<td>Other Professional Services</td>
<td>0.4</td>
<td>0.7</td>
<td>3.6</td>
<td>18.2</td>
<td>28.5</td>
<td>39.1</td>
<td>49.0</td>
<td>52.4</td>
<td>56.2</td>
<td>58.9</td>
</tr>
<tr>
<td>Home Health Care</td>
<td>0.1</td>
<td>0.2</td>
<td>2.4</td>
<td>12.6</td>
<td>30.5</td>
<td>30.5</td>
<td>38.0</td>
<td>42.7</td>
<td>47.9</td>
<td>52.7</td>
</tr>
<tr>
<td>Nursing Home Care</td>
<td>0.8</td>
<td>4.0</td>
<td>19.0</td>
<td>52.6</td>
<td>74.1</td>
<td>95.3</td>
<td>110.5</td>
<td>115.2</td>
<td>120.7</td>
<td>124.9</td>
</tr>
<tr>
<td>Prescription Drugs</td>
<td>2.7</td>
<td>5.5</td>
<td>12.0</td>
<td>40.3</td>
<td>60.9</td>
<td>120.6</td>
<td>174.2</td>
<td>188.8</td>
<td>199.7</td>
<td>216.7</td>
</tr>
<tr>
<td>Other Medical Products</td>
<td>2.3</td>
<td>5.0</td>
<td>13.6</td>
<td>33.7</td>
<td>40.9</td>
<td>49.5</td>
<td>54.8</td>
<td>55.9</td>
<td>57.6</td>
<td>59.3</td>
</tr>
<tr>
<td>Other Personal Health Care</td>
<td>0.6</td>
<td>1.2</td>
<td>3.3</td>
<td>9.6</td>
<td>23.0</td>
<td>37.0</td>
<td>50.3</td>
<td>53.2</td>
<td>56.8</td>
<td>62.2</td>
</tr>
<tr>
<td>Personal Health Care</td>
<td>$23.3</td>
<td>$62.9</td>
<td>$215.3</td>
<td>$607.5</td>
<td>$863.7</td>
<td>$1,139.6</td>
<td>$1,445.9</td>
<td>$1,547.7</td>
<td>$1,653.7</td>
<td>$1,762.0</td>
</tr>
<tr>
<td>Government</td>
<td>1.2</td>
<td>2.8</td>
<td>12.2</td>
<td>39.2</td>
<td>58.1</td>
<td>81.8</td>
<td>121.0</td>
<td>129.0</td>
<td>133.6</td>
<td>145.4</td>
</tr>
<tr>
<td>Administration and Net Cost of Private Health Insurance</td>
<td>0.4</td>
<td>1.4</td>
<td>6.4</td>
<td>20.0</td>
<td>31.0</td>
<td>43.4</td>
<td>53.8</td>
<td>53.9</td>
<td>56.3</td>
<td>58.7</td>
</tr>
<tr>
<td>Public Health Activities</td>
<td>0.7</td>
<td>2.0</td>
<td>5.4</td>
<td>12.7</td>
<td>18.3</td>
<td>25.6</td>
<td>35.5</td>
<td>38.8</td>
<td>40.6</td>
<td>41.8</td>
</tr>
<tr>
<td>Noncommercial Biomedical Research</td>
<td>1.9</td>
<td>5.8</td>
<td>14.5</td>
<td>34.7</td>
<td>45.4</td>
<td>63.2</td>
<td>76.3</td>
<td>83.0</td>
<td>89.1</td>
<td>97.6</td>
</tr>
<tr>
<td>New Construction</td>
<td>$27.5</td>
<td>$74.9</td>
<td>$253.9</td>
<td>$714.0</td>
<td>$1,016.5</td>
<td>$1,353.6</td>
<td>$1,732.4</td>
<td>$1,852.3</td>
<td>$1,973.3</td>
<td>$2,105.5</td>
</tr>
<tr>
<td>National Health Expenditures</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Per Capita Personal Spending</td>
<td>$ 126</td>
<td>$ 301</td>
<td>$ 931</td>
<td>$ 2,397</td>
<td>$ 3,214</td>
<td>$ 4,033</td>
<td>$ 4,967</td>
<td>$ 5,264</td>
<td>$ 5,572</td>
<td>$ 5,879</td>
</tr>
<tr>
<td>Per Capita National Spending</td>
<td>$ 148</td>
<td>$ 356</td>
<td>$ 1,102</td>
<td>$ 2,813</td>
<td>$ 3,783</td>
<td>$ 4,790</td>
<td>$ 5,952</td>
<td>$ 6,301</td>
<td>$ 6,649</td>
<td>$ 7,026</td>
</tr>
<tr>
<td>National Spending as a Percent of GDP</td>
<td>5.2</td>
<td>7.2</td>
<td>9.1</td>
<td>12.3</td>
<td>13.7</td>
<td>13.8</td>
<td>15.8</td>
<td>15.9</td>
<td>16.0</td>
<td></td>
</tr>
</tbody>
</table>

http://www.cms.hhs.gov/NationalHealthExpendData/02_NationalHealthAccountsHistorical.asp#TopOfPage
from 40 to 300 percent higher than in other developed countries. Much of the
difference is predictable: countries with higher living standards, measured by per
capita income, spend more on promoting health.

Although high per capita spending paints a dramatic picture of spending
disparities, the share of output devoted to medical care is more reflective of shifts
in priorities. The percentage of GDP devoted to medical care spending has risen
dramatically in the United States since the late 1960s, from less than 6 percent to
16 percent. In comparison, in most developed countries worldwide, the percentage
ranges from 9 to 11 percent. Increasing health care expenditures as a percent
of GDP may reflect a conscious choice on the part of the consuming public to
spend more for health care. Or it may reflect an inefficient approach to health care
financing and a piecemeal attempt at reform that to date has been concentrated
on community hospital inpatient services, virtually ignoring every other aspect of
medical care delivery.

Clearly, the United States spends more on medical care, and devotes a larger
percentage of economic output to medical care, than any other country in the
world. Although interesting, these facts ignore three important questions: What is
a reasonable percentage of output to devote to medical care spending? How much
can we afford? Are we getting our money’s worth?

First of all, no one knows the ideal percentage of GDP that medical care
spending should consume. We do know, however, that as income increases, spend-
ing on services, including health care, tends to increase. Wealthy countries spend
proportionately more on medical care than poor countries. Since the United States
is among the leaders in per capita income in the industrialized world, it should come
as no surprise that U.S. medical care spending is the highest.

Second, a growing economy allows more resources to be devoted to those areas
of the service sector where productivity may lag, including medical care, education,
police protection, and the performing arts. In an economy where productivity
is growing in most sectors and declining in none, consumers can have more of
everything. It is merely a matter of devoting a different proportion of income to
the production of the various sectors (Baumol, 1993). This reappropriation is
accomplished by transferring resources from those sectors where productivity is
increasing to those where it is stagnant.

Baumol refers to the phenomenon of lagging productivity in the service sector
as the “cost disease of personal services.” Applying his reasoning to medical care,
the lag in productivity may be traced to two main factors: First, medical services
are hard to standardize, making it difficult to automate. Before you can cure
someone, it is necessary to diagnose the problem. Diagnosis and cure are done on
a case-by-case basis. Thus, efficiency and productivity tend to lag behind the rest
of the economy. Second, most people perceive that quality of care is positively
correlated with the amount of time the physician spends with the patient. Thus, it
is difficult to reduce the labor content of medical services. Physicians who speed
up the examination process are often accused of shortchanging their patients.
This same reasoning may also be applied to education, the performing arts, legal
services, and insurance.

Finally, empirical evidence indicates that the increase in health care spending
witnessed over the past 40 years provides substantial benefits to society that far
outweigh the associated costs. Lichtenberg’s (2002) analysis strongly supports the
hypothesis that medical innovation in the form of new drugs and overall health
care spending contributed positively to increased longevity between 1960 and 1997.
In fact, he concluded that the most cost-effective way to increase life expectancy
is through increased spending on new drug development. Cutler and McClellan
Access to Care

According to recent census estimates, approximately 47 million Americans were without health insurance in 2005, creating mounting pressure on policy makers to come up with a plan to ensure access to medical care for all Americans (DiNavas, Proctor, and Smith, 2007). It is interesting to note that over 40 percent of the uninsured are between the ages of 18 and 34, age categories that use relatively less medical care.

Having no health insurance is not the same thing as having no access to medical care. In fact, the uninsured in this country receive about 60 percent of the medical care per capita of those with insurance. Nonelderly Americans who were privately insured spent $2,484 per capita on medical care in 2001, compared to $1,587 for the uninsured. In contrast, per capita spending in Canada was $1,173, approximately three-fourths of U.S. spending on the uninsured. While uninsured Americans are not going without care, they do receive less care than insured Americans (Hadley and Holahan, 2003).

The ideological struggle surrounding medical care reform has focused on two competing visions of universality. One vision argues for universal coverage in a system that requires mandatory participation, and the other supports universal access in a voluntary system in which everyone can buy health insurance if they desire to do so. The debate has not progressed far beyond an argument over the percentage of the population that would have health insurance under the various alternatives. To truly advance the debate, we must address the critical issue of individual rights versus social responsibility. Is access to medical care an individual right, or is it a social responsibility to provide access to those who cannot afford to pay for it? How we choose to answer this question will go a long way toward determining the future of medical care delivery and finance.

Medical Outcomes

The third area of concern is the health of the population. Those critical of the U.S. delivery system cite the relatively poor health outcomes experienced in this country. The typical indicators used for comparisons are presented in Table 3.2. Male life expectancy at birth is the lowest among the six countries listed, at 75.2 years. Female life expectancy, also last among the six countries listed, is 80.4 years. Infant mortality rates are the highest in the United States, over two times the Japanese rate. Spending, both as a percentage of GDP and on a per capita basis, is much higher in the United States. In fact, per capita spending in Switzerland, ranked second behind the United States, is less than 65 percent of U.S. spending. Using these indicators, it appears that we may not be getting enough value for the money being spent. Is the U.S. system delivering an inferior product, or is there another way to look at the evidence?

The use of health indicators to praise or fault a delivery system ignores the contribution of the underlying demographic and social factors entirely. Health indicators reflect more than health care delivery. Life expectancy and infant mortality say a lot about environment, lifestyle choices, and social problems. The U.S. system must deal with a higher incidence of most of these problems than other industrialized countries—drug abuse, violence, reckless behavior, sexual promiscuity, and illegitimacy. These problems complicate the delivery of medical care and are, in part, responsible for the poor health indicators.

(2001) examine the benefits of technological change in five common conditions: heart attacks, low-birth weight infants, depression, breast cancer, and cataracts. They conclude that health care spending on these conditions is worth the cost of care.
Others argue that other indicators more accurately reflect the effectiveness of a health care system. In particular, how does the system treat people who are critically ill? The story is much different when disease-specific death rates are examined. Data from the OECD in Table 3.3 provide details for death rates per 100,000 for the top ten causes of death in the United States in 2002. Overall, the U.S. had the highest death rate that year. In the ten specific categories listed, the United States ranked in the top three in four of them. In the top two categories, the United States ranked high: third in heart disease and stroke and fifth in cancer.

In Table 3.4, Verdecchia and colleagues (2007) provide international comparisons of age-adjusted five-year survival rates for different types of cancer. Using data from European and U.S. cancer registries, they find that the United States has the highest survival rates for most types of cancer. For all malignancies, men in the United States have a 66.3 percent survival rate five years after diagnosis, and women have a survival rate of 62.9 percent. Whether Americans with cancer actually live longer is another issue. The higher five-year survival rates may be the result of earlier screening. No doubt when cancer is diagnosed earlier, there is a better chance that it can be controlled. So earlier screening is likely to lead to longer life expectancies.

As presented in Table 3.5, U.S. life expectancy at age 80 ranks second among males in the seven countries listed, behind only Japan, and third among females, behind Canada and Japan. The other three indicators shown in the table provide a measure of the efficiency of the system in delivering medical care. The U.S. has the lowest average acute inpatient length of stay at 5.6 days. Most government-run systems pay a fixed rate per hospital day, resulting in comparatively long average stays in the hospital. Over the course of the typical hospital stay, the later days are usually less costly than the earlier days. Keeping patients in the hospital longer provides the opportunity for hospitals to recover the higher costs of the first
### Table 3.3: Crude Death Rate per 100,000 population, 2002
Top Ten Causes of Death, United States

<table>
<thead>
<tr>
<th>Cause of Death</th>
<th>Canada(^1)</th>
<th>France</th>
<th>Germany</th>
<th>Japan</th>
<th>Switzerland</th>
<th>UK</th>
<th>USA</th>
</tr>
</thead>
<tbody>
<tr>
<td>All Causes</td>
<td>558.1</td>
<td>565.0</td>
<td>623.0</td>
<td>449.3</td>
<td>514.9</td>
<td>647.2</td>
<td>666.7</td>
</tr>
<tr>
<td>Diseases of the Circulatory System</td>
<td>175.5</td>
<td>149.7</td>
<td>268.1</td>
<td>129.8</td>
<td>175.6</td>
<td>240.1</td>
<td>235.0</td>
</tr>
<tr>
<td>Malignant Neoplasms</td>
<td>173.2</td>
<td>172.7</td>
<td>166.7</td>
<td>146.5</td>
<td>144.1</td>
<td>182.1</td>
<td>166.3</td>
</tr>
<tr>
<td>Accidents and Adverse Effects</td>
<td>38.3</td>
<td>51.5</td>
<td>32.7</td>
<td>42.5</td>
<td>37.9</td>
<td>27.5</td>
<td>52.3</td>
</tr>
<tr>
<td>Cerebrovascular Diseases</td>
<td>35.3</td>
<td>34.6</td>
<td>52.2</td>
<td>55.2</td>
<td>31.7</td>
<td>63.4</td>
<td>39.9</td>
</tr>
<tr>
<td>Diseases of the Nervous System</td>
<td>24.2</td>
<td>22.8</td>
<td>13.6</td>
<td>5.3</td>
<td>20.7</td>
<td>18.1</td>
<td>25.0</td>
</tr>
<tr>
<td>Diabetes Mellitus</td>
<td>19.6</td>
<td>11.4</td>
<td>16.6</td>
<td>5.9</td>
<td>12.9</td>
<td>7.5</td>
<td>20.9</td>
</tr>
<tr>
<td>Infectious and Parasitic Diseases</td>
<td>8.2</td>
<td>11.0</td>
<td>8.2</td>
<td>9.4</td>
<td>5.8</td>
<td>5.7</td>
<td>18.0</td>
</tr>
<tr>
<td>including AIDS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pneumonia and Influenza</td>
<td>10.2</td>
<td>10.0</td>
<td>12.7</td>
<td>34.5</td>
<td>11.5</td>
<td>31.4</td>
<td>15.3</td>
</tr>
<tr>
<td>Mental Disorders</td>
<td>14.1</td>
<td>16.7</td>
<td>6.7</td>
<td>1.7</td>
<td>20.3</td>
<td>15.8</td>
<td>12.5</td>
</tr>
<tr>
<td>Chronic Liver Disease and Cirrhosis</td>
<td>6.9</td>
<td>11.7</td>
<td>15.4</td>
<td>7.1</td>
<td>NA</td>
<td>10.1</td>
<td>9.5</td>
</tr>
</tbody>
</table>

\(^1\) 1997.

### Table 3.4: Age-Adjusted 5-year Survival Rates
2000–2002 (In percentages)

<table>
<thead>
<tr>
<th>Country</th>
<th>Colorectal</th>
<th>Breast Cancer</th>
<th>Prostate Cancer</th>
<th>All types (Men)</th>
<th>All types (Women)</th>
</tr>
</thead>
<tbody>
<tr>
<td>France</td>
<td>59.9</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Germany</td>
<td>61.2</td>
<td>78.2</td>
<td>85.3</td>
<td>50.0</td>
<td>58.8</td>
</tr>
<tr>
<td>Italy</td>
<td>59.4</td>
<td>83.7</td>
<td>85.0</td>
<td>49.8</td>
<td>59.7</td>
</tr>
<tr>
<td>Sweden</td>
<td>59.8</td>
<td>86.3</td>
<td>82.5</td>
<td>60.3</td>
<td>61.7</td>
</tr>
<tr>
<td>Switzerland</td>
<td>63.8</td>
<td>84.5</td>
<td>87.3</td>
<td>54.6</td>
<td>61.1</td>
</tr>
<tr>
<td>United Kingdom(^1)</td>
<td>51.8</td>
<td>77.8</td>
<td>NA</td>
<td>44.8</td>
<td>52.7</td>
</tr>
<tr>
<td>Europe average</td>
<td>56.2</td>
<td>79.0</td>
<td>77.5</td>
<td>47.3</td>
<td>55.8</td>
</tr>
<tr>
<td>United States</td>
<td>65.5</td>
<td>90.1</td>
<td>99.3</td>
<td>66.3</td>
<td>62.9</td>
</tr>
</tbody>
</table>

Source: Verdecchia et al., 2007.
\(^1\) England only.
few days. Longer stays translate into a need for more hospital beds per capita, representing a waste of hospital resources. Predictably, the United States has the fewest number of hospital beds per 1,000 population.

The United States ranks third in terms of the number of physicians per 1,000 population. When physician payments are based on established fee schedules, physicians are able to compensate for low fees by requiring extensive follow-up visits. In France, with the second highest physician-to-population ratio, patients saw their physician an average of 6.6 times in 2004. In Japan, with the fewest physicians per capita, the average was 13.8. The typical American had 3.8 physician’s visits that year. In general, patients find it easier to schedule appointments in the United States, and they spend more time with their physicians during each appointment.

<table>
<thead>
<tr>
<th>Country</th>
<th>Life Expectancy at Age 801</th>
<th>Acute Care Inpatient Length of Hospital Stay</th>
<th>Hospital Beds per 1,000 Population</th>
<th>Physicians per 1,000 Population</th>
</tr>
</thead>
<tbody>
<tr>
<td>Canada</td>
<td>8.1</td>
<td>7.3</td>
<td>2.9</td>
<td>2.1</td>
</tr>
<tr>
<td>France</td>
<td>7.81</td>
<td>8.7</td>
<td>3.8</td>
<td>3.4</td>
</tr>
<tr>
<td>Germany</td>
<td>7.22</td>
<td>8.7</td>
<td>6.4</td>
<td>3.4</td>
</tr>
<tr>
<td>Japan</td>
<td>8.4</td>
<td>20.2</td>
<td>8.4</td>
<td>2.0</td>
</tr>
<tr>
<td>Switzerland</td>
<td>7.52</td>
<td>8.8</td>
<td>3.9</td>
<td>3.8</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>7.12</td>
<td>6.6</td>
<td>3.2</td>
<td>2.3</td>
</tr>
<tr>
<td>United States</td>
<td>8.2</td>
<td>5.6</td>
<td>2.8</td>
<td>2.4</td>
</tr>
</tbody>
</table>


1 In years.  2 2003.  3 2002.

Adam Smith asserted in his famous treatise, The Wealth of Nations, that individual decision making is motivated by self-interest. Guided by the “invisible hand” of the market, this self-serving behavior, in turn, serves to promote the interests of others. In other words, when markets exhibit certain ideal conditions, or perfectly competitive conditions, optimizing behavior on the part of individuals and firms leads to efficient outcomes.

Following the traditions established by Smith and the classical school of economics, modern-day economists evaluate markets according to the twin criteria of efficiency and equity. There are two aspects of efficiency—allocative efficiency and technical efficiency. Allocative efficiency may be viewed as efficiency in the final distribution of consumption. Consumers buy a good until the benefits received from the last unit purchased equals the price.2 Thus, everyone purchasing a good

---

2 Downward-sloping demand curves are implied from the law of diminishing returns, indicating that the last unit of a good purchased has a marginal value equal to its market price.
places a marginal value on the good at least equal to its market price. When everyone pays the same price for the good, there is no way to reallocate consumption from consumers to nonconsumers without lowering overall consumer welfare.

Technical efficiency may be thought of as efficiency in production, or cost efficiency. In perfectly competitive markets, producers must minimize costs to maximize profits. When all producers pay the same input prices, goods and services that are produced will have marginal valuations that are higher than goods and services that could have been produced with the same resources. In summary, perfect competition guarantees both allocative and technical efficiency.

Equity considerations are also important when evaluating economic systems. Even though the issue of equity is based on some standard of fairness, ideological differences dictate whether that standard is defined either in terms of outcomes or in terms of opportunities. For example, one economist might define equity in terms of final outcomes. In this case, any differences in infant mortality rates between, say, whites and African Americans would be viewed as inequitable and obviously the result of unequal access to the medical care system. How else could you possibly explain the large gulf between the 5.7 deaths per 1,000 live births among white Americans and 13.6 among African Americans (Matthews and MacDorman, 2007)?

Another economist might have a unique perspective on the same issue. Defining equity in terms of opportunities rather than outcomes, the same disparities in mortality rates would be interpreted another way. From this perspective, even in a world of equal opportunities, there will be varied outcomes. Blaming the differences on unequal access ignores demographic differences such as age, education, and marital status between the two population cohorts. Additionally, differences in lifestyle choices are also important, including the decision to smoke cigarettes, drink alcohol, or take drugs during pregnancy. Whether defined in terms of outcomes or opportunities, equity has become an important component in the evaluation of markets, especially medical markets.

Few people will argue against the importance of an equitable distribution of health care availability. But health care is like any other desirable commodity: It is subject to an equity–efficiency trade-off. Access to medical care differs according to individual circumstances, such as age, gender, income, geographic location, and insurance coverage. No matter how much we may desire equity, it comes at a price; mandating equity may be desirable, but it is costly.

The formal argument for competitive markets is based on the notions of economic efficiency and social equity, but some favor competition simply because it guards against the concentration of market power and promotes consumer sovereignty. Competition among providers and their desire to satisfy consumer preferences ensures against consumer exploitation. Consumers always have alternative sources of supply in competitive markets. Cost-conscious behavior on the part of consumers increases their sensitivity to price changes. Individual providers face perfectly elastic demand curves when cost-conscious consumers have alternative sources of supply. Consequently, prices of goods and services equal the marginal cost of production.

When markets work, prices reflect the valuation of forgone opportunities. As equilibrium is reached, marginal values and prices converge, and the value of the goods and services that are produced is greater than the value of the goods and services that could have been produced with the same resources. In other words, if individuals in society placed a higher value on the last dollar spent on medical care than on the last dollar spent on, say, education, then they would demand that more be spent on medical care and less on education, until the marginal valuations were equal.
**Market Failure**

According to Murphy's Law, if anything can go wrong, it will. Various imperfections in medical markets make the task of delivering a product equitably and efficiently more difficult. When the underlying assumptions of competitive markets are not met, markets fail to deliver the optimal output levels (Rice, 1998). Markets fail to allocate resources optimally when firms have market power, when there are externalities in consumption and production, and when the good produced is a public good.

**Market Power**

Any departure from perfect competition—whether it be monopoly, oligopoly, cartel, monopolistic competition, monopsony, or any other market structure imperfection—violates the optimality considerations discussed earlier. A profit-maximizing firm with market power sets prices at levels that exceed marginal costs. To maintain those prices, the firm must restrict output to levels that are less than optimum. Prices will be too high, costs will be too high, resources will be underutilized, and society will suffer an economic loss.

Market power is depicted graphically by any departure from perfectly elastic demand curves. Figure 3.1 points out the differences in pricing and output between firms in perfectly competitive markets and those with market power. When demand curves are perfectly elastic, they are drawn as horizontal lines. Profit maximizers set marginal revenue \( MR \) equal to marginal cost \( MC \). With price equal to marginal revenue, \( MR = MC \) at the same output level \( Q_0 \) where \( P_0 = MC \) (the condition for allocative efficiency).

Market power gives a firm some control over its pricing decisions. Raising price reduces quantity sold without the complete loss of customers. With a downward-sloping demand curve, the firm's marginal revenue is less than the price it charges. Setting \( MR = MC \) now results in a lower output level \( Q_1 \) and the ability to charge a higher price \( P_1 \). Higher prices, lower output, and underutilization of resources result in a loss in welfare as measured by the loss in consumer and producer surplus.

In spite of these problems, monopoly may still be the most effective way to organize production in a market. When production is subject to economies of scale, the long-run average cost curve declines continuously. Competition will

---

**Figure 3.1: The Consequences of Market Power on Price and Output**

- **Price**
  - \( P_1 \)
  - \( P_0 \)

- **Quantity**
  - \( Q_0 \)
  - \( Q_1 \)

- **Marginal Cost**
- **Marginal Revenue**

- **Demand under Perfect Competition**
- **Demand with Market Power**
result in the exit of all but one firm. That remaining firm, the natural monopoly, will not set price competitively; and since \( P > MC \), output is not provided at its optimal level. To correct this misallocation of resources, the most effective option may be regulation.³

Using Figure 3.2 to illustrate this point, suppose the firm has a long-run average cost curve that is downward-sloping as it crosses the market demand curve. Under these circumstances, a single firm can supply enough output to satisfy consumer demand and can do so at progressively lower unit costs.⁴ Shielded from competition from rival firms, the monopolist has no compelling reason to be efficient. Focusing solely on profit maximization, the firm will produce less than the optimal level of output \( (Q_0) \), and prices will be higher than if the market were competitive \( (P_0) \). To correct this problem, government price controllers often try to establish a maximum price the monopolist can charge that more closely approximates the perfectly competitive solution. Setting a price at \( P_1 \), for example, enables the firm to earn a normal return on its investment and produce at higher output levels \( (Q_1) \).

Market power in an input market also causes an inefficient allocation of resources. A monopsonist, as the sole buyer of a particular resource, faces an upward-sloping supply curve instead of a perfectly elastic supply curve. As a result, the firm has some discretion over the price it pays for the resource. If more is desired, then the firm must pay a higher price. If less is desired, then prices fall accordingly. The results are shown in Figure 3.3, where the monopsonist faces a situation in which the marginal cost of the resource is greater than the price of the resource. Instead of setting demand equal to supply and paying \( P_0 \) to employ \( Q_0 \) units of the resource, the monopsonist equates demand—its assessment of the marginal value of the resource used in production—with the marginal cost of the resource, and employs

---

³ Certain tax and subsidy schemes might actually be more efficient, but discussion of these alternatives is beyond the scope of this presentation.

⁴ Because price, represented by the demand curve, is above the average cost curve at every point, the firm can increase sales by lowering price and still make an economic profit.
Market power in the resource market enables firms to employ fewer resources and pay lower prices for their use than if the market were perfectly competitive. The result of this lost output is lost income to resource owners and fewer goods and services available to consumers. In summary, market power insulates a firm from the competitive forces that ensure allocative and technical efficiency, resulting in a loss to society.

**Monopsony: When Buyers Have Market Power**

Market power on the buyers’ side, called *monopsony*, gives buyers more leverage in determining the prices they pay for goods and services. If buyers have the ability to consolidate their demand under the control of a single collective, they may have market power. In this case, the monopsonist would function as a buyers’ union. In today’s language, this union would be called an *alliance* or a *cooperative*. The larger the cooperative, the more control the group can assert over the prices charged to its members. As already discussed, equilibrium for the monopsonist occurs at a price and output level that is below the level that would exist in perfectly competitive markets, representing lost economic welfare.

Even with the lost productivity, some still argue that monopsony provides a net benefit to society. Proponents of market power for buyers agree that the unilateral exercise of market power should be illegal on either side of the market. They contend, however, that providers in medical markets already exercise a significant degree of market power on the sellers’ side of the market. The use of power on the buyers’ side represents a countervailing force that encourages competitive behavior among sellers and promotes the efficient use of resources.

The formal explanation of this phenomenon is described in most intermediate microeconomics textbooks under the heading “bilateral monopoly.” A bilateral monopoly exists in a market when there is only one buyer seeking the output of a single seller. In other words, bilateral monopoly
monopoly is characterized by monopsony on the demand side and monopoly on the supply side. In the following graph, \( D \), \( MR \), and \( MC \) are the demand, marginal revenue, and marginal cost curves confronting the monopolist seller. Profit maximizing price and output, \( P_2 \) and \( Q_2 \), are determined by \( MC = MR \) at point \( A \).

A monopsonist with absolute control over demand could force the monopolist to behave like a firm in a perfectly competitive market. Under these conditions, \( MC \) is also the firm’s supply curve. Likewise, \( MC_B \) becomes the relevant marginal cost of buying an additional unit of the output. The monopsonist attempts to equate the marginal cost of buying with its own marginal valuation of the product (\( MV_p \)) at point \( B \). At the optimal level of output, \( Q_f \), the monopsonist pays the lowest price the provider is willing to accept and still cover marginal cost, \( P_1 \).

In terms of the final price, the negotiated outcome will fall somewhere between the two extremes, \( P_1 \) and \( P_2 \). The exact solution depends on the relative bargaining strengths of the two sides. The monopoly provider enters the negotiations wanting a higher price and lower output than the monopsony buyer. To avoid an impasse, the monopolist will likely offer somewhat lower prices and slightly more output. The monopsonist will agree to pay more than \( P_1 \) if the monopolist provides more than \( Q_2 \). As long as price does not fall below \( P_3 \), the final output level will fall between \( Q_3 \) and \( Q_2 \). Thus, for bilateral monopoly to benefit society, bargaining strengths of buyers and sellers must be approximately equal. If either side has a disproportionate share of the bargaining power, it will be able to tilt the balance in its favor to the detriment of society. (Technical note: Relative bargaining strengths and the final outcome will be different if the supply curve is so steeply sloped that \( Q_2 > Q_1 \). In this case, the monopolist wants to provide more output than the monopsonist wants to buy, weakening the monopolist’s bargaining position.)


**Externalities**

Sometimes the actions taken by individuals in the process of producing or consuming will have an effect on the welfare of others. An *externality* may be either positive or negative, depending on whether it benefits or harms other people. By maintaining
her property, a homeowner generates a positive externality for all her neighbors. Not only is it pleasing to look at a freshly painted house and well-kept garden, but the market values of surrounding properties are enhanced at the same time.

Examples of negative externalities abound. Anyone smoking a cigar in a crowded room imposes costs on everyone else in the room. Everyone has less fresh air to breathe than if the smoker were forced to internalize all the costs of his smoking. A factory that dumps toxic waste into a nearby river shifts some of the cost of production (i.e., waste disposal) onto those people who live downstream from the plant. The same can be said about acid rain, traffic congestion, and the many other examples of negative externalities that could be listed.

Externalities affect economic efficiency, and normal market mechanisms have no way of accounting for them. Decision makers are not required to absorb the costs of negative externalities and have no way to capture the benefits of positive externalities. The result is a level of output that is nonoptimal.

Externalities exist as by-products of the decision to produce and consume. Because formal markets do not exist for these by-products, they are produced in nonoptimal quantities. Take, for example, the case of automobile emissions in a crowded metropolitan area. By choosing to drive your own car to work, you impose costs on others in the form of carbon monoxide emissions from the exhaust. A large percentage of the costs of commuting are internalized. You pay for the car, the gasoline, and the insurance. But your fellow commuters pay the costs that cannot be internalized, namely the costs of the by-products of your commute: traffic congestion and air pollution.

Figure 3.4 illustrates the impact of an externality in a private market, the daily commute to work or school. Externalities arise because the driver does not internalize the full cost of the commute. Graphically, the vertical distance between the marginal social cost (MSC) curve and the marginal private cost (MPC) curve represents the external costs that the driver forces others to pay. Individual decision makers determine their own commuter miles by equating marginal benefit (MB) with MPC. Given the additional costs that society at large must pay, the number of commuter miles actually driven (Q_m) is greater than the optimal number (Q*)

To incorporate these externalities into individual decision making requires some form of collective action to force commuters to pay the full costs of their actions. For
example, through their elected representatives, voters may decide to reduce the number of commuter miles driven by private automobiles by erecting toll booths on all major freeways or simply forcing everyone who drives into the city to pay a commuter tax. In either case, the goal is to force private decision makers to take into account the external costs of their actions. By moving the MPC closer to the MSC, the number of commuter miles driven will approach its optimal level, \( Q^* \).

In the case of positive externalities, the competitive output rate will be too small if the decision maker cannot capture the externalities generated. The problem emerges because the marginal private benefit is less than the marginal social benefit. When marginal cost and marginal private benefit are equated, the resulting output is less than optimal.

**BACK-OF-THE-ENVELOPE**

**Optimal Output: Private versus Public Goods**

The market demand for a private good is derived by horizontally summing all the individual demands. In this case, total output is the sum of the amounts consumed by each individual in the market. When goods are rival goods, the amount consumed by one individual cannot be consumed by anyone else. In the diagram below, assume two consumers with demand curves \( D_1 \) and \( D_2 \). Equating market demand (\( \Sigma D \)) with supply results in a price of \( P_0 \) and an optimal output level of \( Q^* \). Given the market price, each consumer will demand a level of output where price is equal to the marginal cost of production.

In the case of a public good, the market demand curve is determined by the group’s willingness to pay for a given level of output. (In this case, the group consists of two people, 1 and 2.) Since the good is nonrival, the market demand curve is derived by summing the individual demand curves vertically, instead of horizontally, as was the case for a private (rival) good. At the optimal level of output (\( Q^* \)), the group is willing to pay \( P_3 \), the sum of \( P_1 \) and \( P_2 \). Remember that \( Q^* \) is the optimal level for the good, because
at that level, the marginal social benefit is equal to the marginal social cost of production.

In the society at large, identifying the marginal social benefit curve is problematic. No one is required to reveal his or her individual marginal valuations, so determining society's willingness to pay becomes a challenge. Some individuals will find it worthwhile to become free riders. Because of peer pressure, the free-rider problem may not be a big issue when there are only two people in the group. But in a large society, no one person places a high enough marginal value on the good to ensure its provision. In other words, the marginal costs are substantially higher than any one person's or small group's demand ($D_s$ in the diagram below). Under these circumstances, the market simply will not ensure the production of the good; its cost will simply be too high for anyone to absorb without collective action.

### Public Goods
Markets distribute goods efficiently when people spend their own money to enjoy the benefits of consumption. The market for Nike shoes works because those unwilling to pay the price for Nikes do not own Nikes. The market mechanism provides purchasers with the benefits of consumption and excludes nonpurchasers from receiving those benefits. Additionally, the benefits flow to specific individuals. Consumption of a crispy taco by one person does not satisfy the hunger of another.

In certain situations, these two characteristics do not hold. In fact, many important goods, such as national defense and air traffic control, do not exhibit them fully. Nonexcludable and nonrival goods are called public goods. Nonexcludability in the distribution of a good results when the costs of preventing nonpayers from consuming are high, making it difficult to impose prices on these individuals. Once a strategic national defense system is operational, there is no way to exclude individuals from its protective umbrella simply because they refuse to pay their share of the costs.

Nonrivalry in consumption means that more than one person can enjoy the benefits of consuming a commodity without affecting the enjoyment of the other. One person's consumption does not reduce the benefit received by someone else. In technical terms, the marginal cost of providing the good to additional consumers is zero. For example, after the Army Corps of Engineers builds a levy, any number of houses may be built in the flood plain without increasing the marginal cost of flood control. If an air traffic control system is in place, the marginal cost of monitoring the flight path of an additional aircraft is zero.

**free rider**
An individual who does not buy insurance, knowing that in the event of a serious illness, medical care will be provided free of charge.

**nonrival goods**
A good or service which does not, when consumed by one individual, limit the amount available to anyone else.
Serious efficiency problems arise when we attempt to provide **nonexcludable goods** through private markets. To understand the problem, note the difference between the provision of excludable and nonexcludable goods. Transactions involving private (excludable) goods take place in markets as long as the individual's marginal valuation of that good exceeds its price. Individuals have no incentive to lie about the marginal value placed on a good. Because of excludability, if you underestimate the marginal value you place on a good, you run the risk of not getting the good and losing out on the marginal benefits of consumption. If you have ever witnessed an auction of any kind, you are familiar with this concept. Marginal valuations are reflected in the prices individuals are willing to pay for items that are being auctioned. You must make those marginal valuations known, or you run the risk of finding yourself empty-handed at the end of the auction.

In contrast, when goods are nonexcludable, there is an incentive for individuals to understate their true marginal valuations. If I can enjoy all the benefits of consumption without paying for that privilege, why pay? Those individuals who refuse to pay for a good while still enjoying the benefits of consumption are called *free riders* (some might even call them *freeloaders*). Public television provides a good example of the free-rider problem. The number of people who watch public television far exceeds the number who subscribe. Of course, some ride free, but others have to pay, or no one rides at all. And that's the point. Private markets tend to undersupply nonexcludable goods.

The case of public goods is simply a special kind of positive externality. So to ensure its availability at optimal levels, public provision of the good may be required. Governments can require individuals to participate in paying for goods through the power to tax. Clearly, all goods publicly provided are not public goods. Whether the good is provided by a government entity is not the issue. Governments often engage in the provision of private goods, for example, by staging concerts in the park and collecting garbage. In both cases, nonpayers may be excluded from consumption at very little cost, eliminating the problem of the free rider.

Even strong defenders of the market admit that private markets do not always provide goods and services at efficient levels. But those critical of market outcomes must address the issue of whether the government can do a better job. Is government provision any more efficient than private provision? Does it result in a more equitable distribution of resources? Is a more equitable distribution of resources worth the cost? We will focus on this question later in the chapter.

**merit good**
A good whose benefits are not fully appreciated by the average consumer and thus should be provided collectively.

**ISSUES IN MEDICAL CARE DELIVERY**

**Medical Care as a “Merit Good”**

Economic models predicting consumer behavior usually assume, among other things, that individuals know what they want and are able to rank their preferences. But often people avoid what is good for them and choose items that are actually harmful. Recognizing this fact, Musgrave (1959) classified certain goods as *merit goods* to describe commodities that ought to be provided even if private demand is lacking. Since merit goods have benefits that are not fully appreciated by the average consumer, their consumption should be encouraged through collective action.

Many would place medical care in the merit-good category. Individuals lacking the ability to fully appreciate the importance of primary and preventive care will underconsume when it comes to this valuable commodity.
MARKET FAILURE IN MEDICAL MARKETS

The obvious starting point in analyzing market failure in medical markets begins with the three causes of market failure discussed above. How prevalent are monopolies in medical markets? Are there significant externalities in consumption and production? Is medical care a public good, nonexcludable in distribution and nonrival in consumption?

Whether this classification is merely a case of imposing preferences on society, or whether it is a genuine merit-good situation, is open to debate.

The usual arguments used to justify government involvement in medical care delivery and finance include market failure, information problems, third-party financing, and even merit goods. These arguments are often compelling, if not always convincing. But when using the merit-goods argument, we must be careful that we are not merely replacing a personal value judgment — that everyone is entitled to medical care — with formal terminology to justify our personal preferences (Baumol and Baumol, 1981).


POLSICY ISSUE

Should everyone be required to participate in an immunization program designed to protect the entire population against a communicable disease?

The Economics of Subsidizing Childhood Immunizations

Public health officials recommend that children receive a full round of vaccinations — including polio, measles, mumps, and whooping cough — before the age of two. To the extent that any children are not vaccinated, the entire childhood immunization program is undermined. The children who go unvaccinated are more likely to get sick, which lowers their welfare. They are also more likely to serve as carriers of the disease and infect others, which lowers the welfare of everybody else. The situation describes the classic case of positive externalities in consumption, where the marginal private benefits (MPB) fall short of the marginal social benefits (MSB).
Traditional Sources of Market Failure

Even though absolute market power in medical markets may be hard to find, lack of competition can still be a significant problem. Most metropolitan areas are served by more than one hospital due to the simple fact that economies of scale in the hospital industry are exhausted at relatively low levels of capacity. Even in communities as small as 180,000 people, two or three hospitals providing most general services could coexist. In smaller communities, the lack of competition presents a greater challenge for market proponents. In these small markets, some inpatient services must be shared to avoid substantial inefficiencies (Kronick et al., 1993).

Even in larger communities with multiple facilities, some providers may have a degree of market power. There are some services and procedures that exhibit significant economies of scale, such as organ transplantation and various imaging technologies that include CT scans (computerized tomography) and MRIs (magnetic resonance imaging).

Although a pure monopoly may be difficult to find, firms often engage in collusive behavior to avoid competition. Recognizing that it is in their collective interest not to engage in price competition, providers differentiate their products to make direct price comparisons difficult. There is competition along the lines of quality and the number of services offered, but not price. Differentiation is often accomplished when providers agree to specialize, for example, with one hospital offering cardiac care and another obstetric care. This type of market segmentation is relatively easy, because most medical care is provided locally.

Externalities arise in medical care in a number of circumstances. The most obvious type of externality is associated with public health programs. Modern society can be a breeding ground for all sorts of communicable diseases. The ability of the Public Health Service to enforce health regulations and monitor contagious diseases serves to improve public health. Related activities include the provision

of clean water, clean air, and adequate sewage disposal, which greatly reduce the incidence of diseases such as cholera and dysentery. In addition, immunization against mumps, measles, small pox, polio, and whooping cough offers protection for more than one individual. The benefits extend to the entire population by eliminating potential carriers of the diseases. In other words, the incremental value to society is greater than the value to the individual alone. In a private market, fewer vaccinations would occur than is socially optimal and may call for collective action in the form of mandates or subsidies or both.

The Lessons from SARS

For those traveling to Canada, Europe, and Asia during the spring of 2003, SARS became a household word. Severe acute respiratory syndrome, or SARS for short, leaped onto the front pages of newspapers from Toronto to Singapore to Tokyo. Reminiscent of the 1995 movie Outbreak, in which a lethal virus spreads from an African monkey, SARS challenged the ability of the public health community to react to the real-life outbreak of a deadly disease.

How easy is it to control a new infectious disease? That depends on how it is transmitted, how hard it is to catch, whether apparently healthy individuals can spread the disease, and whether the organism can find an appropriate host in a nonhuman species. The SARS challenge was complicated by the fact that the disease originated in China, and Chinese authorities failed to report the existence of SARS for months, and then tried to hide the extent of the spread of the disease.

With no treatment yet available, efforts to control the disease have been very crude: identifying everyone infected, tracing everyone they have come into contact with, and isolating them all. Tracing everyone may be impossible, so the only option may be mass quarantines, school closures, and cancelled vacations and holidays.

All things considered, SARS was relatively mild as far as epidemics go. As of June 11, 2003, the World Health Organization (WHO) had received reports of 8,435 probable cases from 29 countries, including 70 from the United States. There were 789 deaths, translating into a mortality ratio of 9.4 percent. The world community was lucky this time. We may not be so lucky when the next deadly bug comes along. Designer facemasks may not be enough to protect us from a bug that is more contagious than SARS. In the meantime, the public health community must come up with better systems and procedures to effectively enforce large-scale quarantines to give the medical research community time to study the infectious disease agents and come up with the appropriate medical response.

The next time you hear reports of some strange illness in a remote corner of the globe, don’t think it will have no impact on your life. Let us hope that the health authorities have learned valuable lessons from SARS.

foundations and medical organizations. The annual Jerry Lewis telethon provides individuals with the opportunity to unite in the fight against muscular dystrophy. Personal contributions to the United Way, the Ronald McDonald House, the Children's Miracle Network, the American Cancer Society, and numerous other national and local organizations advance the fight against certain diseases and provide access to medical treatments that might otherwise be prohibitively expensive.

Given the nature of the externality, even those who refuse to contribute enjoy the benefits of knowing that medical research is finding cures for certain diseases and that certain medical services are available for those who cannot afford to pay for them. If this consumption externality exists and is significant, then collective action through government can be used to provide medical care to that segment of the population that cannot afford to buy it privately. Those who would not contribute privately now share the responsibility through mandatory taxation. Collective action determines the nature of the subsidy, the level of taxation, and the method of distribution.

The medical subsidy is almost always an in-kind transfer rather than a cash payment. Beneficiaries prefer cash rather than services. They almost always find themselves better off with the cash. Donors generally prefer in-kind benefits because of the lack of guarantees that cash would be used for medical care. In fact, Waldo and colleagues (1989) indicate that a cash transfer to the elderly equivalent to their per capita share of Medicare would do more to improve their welfare than the current subsidy for medical services. It seems that donors—in this case, taxpayers—care about health differently than other aspects of the recipient's well-being, which includes whether the food they eat is healthy or whether the house they live in is adequately heated and cooled.

Externalities may also be associated with exceptionally large medical expenditures. Frequently, those with incomplete or no health insurance coverage have medical bills that exceed their ability to pay. Faced with this event, they default on their obligation, and the community must pick up the tab. In other words, providers are forced to write off the expenses as bad debts and shift the costs of care onto privately insured patients. The fact that we are unable, or at least unwilling, to exclude anyone from access to medical care for financial reasons gives rise to the problem of the free rider. For this reason, many advocate mandatory health insurance covering catastrophic (high-cost) episodes of illness. In this way, everyone would be forced to participate in the cost of providing medical care, and the free-rider problem would be moderated.

Pure medical research that has no easily captured commercial value fits the definition of a public good. This is the type of medical research that is packaged and published primarily in medical journals. Much of the information that is shared in this manner shows other medical practitioners ways of combining activities and procedures into a particular mode of treatment. Unless patentable medical devices are included in the procedures, it is difficult for those responsible for the discovery to capture the benefits of their research. Good examples include radial keratotomy and the use of lasers in ophthalmological surgery.

Many will argue that medical research should be treated as a public good and financed collectively through government. In this way, basic advances financed

---

5 Medical care providers usually report the delinquent debtor to the appropriate credit bureau. This has become so common that many lenders, such as commercial banks and consumer credit companies, regularly ignore a default on a would-be borrower's credit history if the debt was associated with medical care (private conversation with Bart Cooper, GMAC).
by the taxpayer would belong in the public domain, freely available to potential users. The other side of the argument recognizes that academicians conduct much of our medical research. Working within the university and medical school setting, they are able to capture the benefits of their discoveries through the rules of promotion and tenure, so at least a portion of the benefits is translated into career enhancement opportunities and personal prestige. Some may choose to keep their findings out of the public domain in order to earn royalties or other payments.

To the extent that medical care has characteristics associated with market power among providers, externalities in production and consumption, and public goods, the level of services provided will fall short of the optimal level as defined by competitive markets.

**POLICY ISSUE**

Is medical research a public good, thus strengthening the argument in favor of government financing of basic medical research?

**ISSUES IN MEDICAL CARE DELIVERY**

**Advertising Professional Services: The Case of Optometry**

In most private markets, consumers gain access to important information through advertising. The argument against advertising professional services is based on the belief that advertising may mislead consumers, undermine quality, and ultimately raise prices. Professional associations representing medical practitioners have led the battle defending the long-standing restrictions on price advertising in medical markets.

Economic theory argues in favor of advertising in markets characterized by asymmetric information between buyers and sellers, where sellers have all the information and buyers have none. For example, advertising provides consumers with information on alternative sources of supply. This results in lower prices, because consumer demand for individual providers becomes more elastic. In addition, one of the goals of advertising is to increase consumer demand. To the extent that advertisers realize this benefit, they can take advantage of economies of scale in production and actually lower prices to consumers.

Although advertising may result in lower prices, its effect on product quality is less certain. In theory, high-quality providers have more to gain by advertising through repeat purchasers. In practice, however, low-quality providers advertise more. Ultimately, the effect of advertising on quality will be determined by consumer demand for quality and provider determination to produce quality, with the latter governed in large part by ethical standards established by the specific profession.

Considering all the evidence, a federal appeals court ruled in 1980 that physicians and other medical professionals could advertise prices and services. Kwoka (1984), studying the market of optometric services, was one of the first to examine the impact of advertising on prices and quality. Results indicated that advertisers’ prices and quality were lower and that nonadvertisers’ prices also fell. However, the quality of the services offered by nonadvertisers actually increased. Given a sufficiently large number of nonadvertisers, overall quality in the market increased. Kwoka estimated that quality-adjusted prices for optometric services fell by 20 percent as a result of advertising, so loosening restrictions on advertising in optometry actually improved economic welfare.


**asymmetric information**

A situation in which information is unequally distributed between the individuals in a transaction. The person with more information will have an unfair advantage in determining the terms of any agreement.

**POLICY ISSUE**

Should physicians and other health care providers be allowed to advertise?
Imperfections in Medical Markets

Other imperfections contribute to the failure of medical markets to provide the socially optimal level of service (see Pauly, 1988). These imperfections include imperfect information, barriers to entry, and the prevalence of third-party payers.

Imperfect Information

Lack of information presents serious problems in a market economy. In medical markets, the problems that arise may be even more serious. Most patients are poorly informed about virtually every aspect of the medical transaction. They are usually aware of their symptoms and syndromes, but seldom do they understand the underlying causes of their medical conditions. They have scarcely an opportunity to form a learned opinion about the physician’s diagnosis or the prescribed treatment. In most cases, anything other than a complete recovery is not part of the expected outcome.

The overall lack of information available to patients is compounded by the difficulty in securing the information, measured in terms of time and expense. As a result, most patients rely almost exclusively on their provider to keep them informed on matters dealing with their medical condition, its diagnosis, and treatment alternatives. Patients also have little knowledge about price and quality differences among alternative providers. This imbalance of information between patient and provider, referred to as asymmetric information, has led to two important market defects.

First, patients are not able to judge price and quality differences among providers. As a result, providers can charge prices that are higher than the prevailing prices in the market for a given level of quality, or they may choose to offer a lower level of quality for a given price. The impact of this phenomenon can be seen in the variation in prices paid and the quantities of medical care provided to similar groups of patients. Evidence for these variations has been compiled by examining, for example, surgery rates for common procedures. In cases where alternative intervention strategies are not available—such as appendectomy, hernia repair, and hysterectomy—the variation in surgery rates is relatively low. But in cases where alternative treatments are available—such as tonsillectomy, disc surgery, and coronary artery bypass grafts—variation is high—up to four times the rate of the low-variance surgeries (Phelps, 1992).

The second problem may be described as an agency problem. The physician serves as the agent of the patient, and the patient delegates most of the decision-making authority to the physician. The expectation, in turn, is that the patient’s best interests will be the top priority. The dual role of provider of services on the one hand and the agent in charge of information on the other creates a dilemma: The physician is in a position to induce the patient to purchase more medical care than is actually needed. Physicians can recommend not only medical care with little marginal value, medical care on the flat of the curve, but also medical care that may actually harm the patient. At the other extreme, enrollees in managed care organizations may find themselves denied care that offers positive net benefits, because it is not in the financial interest of the provider to offer that care.

This information problem does not mean that medical markets are hopelessly noncompetitive. Market mechanisms have arisen to minimize the impact of these information differences. The medical community has created licensing, certification, and accreditation requirements for physicians, specialists, hospitals, and medical schools to assure minimum quality standards. Professional organizations establish ethical standards. And if this is not enough, the threat of a malpractice lawsuit is always a reminder of the importance of promoting the best interests of the patient.
Keep in mind that other markets also exhibit this information problem and are relatively competitive. The market for personal computers is a good example. Except for a small segment of the market, the general public is woefully ignorant of the differences between RAM and ROM, the number of meg in a gig, and the merits of Pentium and Celeron processors. Are there good reasons to buy a Mac instead of a PC? Do I want a zip drive or a DVD? Do I need an internal fax modem? Even with all this consumer ignorance, the market for personal computers is extremely competitive. Why? Because an informed minority provided the initial market discipline. They wrote the newsletters, contributed to the magazines, and spent endless hours on the Internet participating in forums and posting on bulletin boards. The demand for information fostered by this group created awareness among all consumers.

When consumers perceive that acquiring and using information best serves their own interests, there will be a demand for information. Consumers in medical markets do not perceive that their interests are served by spending time and money to acquire information. The third-party payer—the insurance company or the government—expropriates any savings from the search. Change that aspect of the medical marketplace, and consumers will have an incentive to become informed. Virtually all types of medical care, except emergency care, would be purchased in markets with enough informed consumers to ensure economic discipline. The demand for information is evident in the managed care marketplace, where many organizations and networks are reporting to their constituencies on how well they perform in certain critical areas, including primary and preventive care, surgical outcomes, and cost (Kenkel, 1994).

**Barriers to Entry** An important characteristic found in competitive markets is easy entry and easy exit of suppliers. Profits serve as a signal to prospective providers. If profits are greater than normally expected for a given level of risk, firms will enter the market and drive down prices, and profits will adjust to normal levels. Lower-than-normal profits will result in the opposite response, with marginally profitable firms leaving the market and driving up prices and profits for those who remain.

Entry barriers restrict resource movements and result in imperfect competition. Examples of barriers in medical markets are found in numerous restrictions on tasks performed and investments made. The licensing and certification of practitioners are two of the most common ways to restrict entry into the medical profession. The stated purpose of this policy is consumer protection, and its aim is to keep uninformed patients from seeking services from incompetent providers. **Certificate-of-need (CON)** laws require hospitals to secure approval from government planning agencies before adding new capacity or investing in expensive equipment. CON legislation seeks to eliminate the duplication of costly programs within a service area. Restrictions may sound good in theory, but one of the unintended consequences of any limits placed on a market is the elimination of competition. Reduced competition leads to market power, and market power leads to market failure.

**Third-Party Payers** In traditional markets, individuals spending their own money provide the discipline that culminates in the efficient provision of goods and services. One of the main reasons medical markets are not efficient is that consumers do not spend their own money. Only about 3 cents of every dollar spent on hospital services, and 20 cents out of every dollar spent on physicians’ services, comes directly from patients’ out-of-pocket spending. The rest is paid by third parties, primarily health insurance companies and the government. Therein lies the major problem in medical markets. Typically, pricing reflects the interaction of...
POLICY ISSUE

Conventional health insurance virtually eliminates any cost-conscious behavior on the part of the parties involved in the medical care transaction.

fee-for-service
The traditional payment method for medical care in which a provider bills for each episode of care.

cost-plus pricing
A pricing scheme in which a percentage profit is added to average cost.

retrospective payment
Payment determined after delivery of the good or service. Traditional fee-for-service medicine determines payment retrospectively.

Part 1 >>> The Relevance of Economics in Health and Medical Care

consumers’ willingness to pay for goods and services and their ability to buy them. Medical markets regularly ignore the desires of those without insurance and those without the ability to pay for care out-of-pocket. The desires of those who have insurance are distorted by the subsidy provided by their insurance.

A system financed primarily through retrospective fee-for-service insurance reimbursement is open-ended. Providers are able to pass through all their costs, no matter how inefficient the production of services. The system can be described as a cost-plus pricing system (Goodman and Musgrave, 1992). In a cost-plus environment, there is no incentive for providers to search for more efficient methods of production, and patients have no incentive to search for providers who offer lower prices. In competitive markets, providers are rewarded for offering quality products at the lowest price. In cost-plus markets, providers are rewarded by offering more services at higher prices, passing on the additional costs to the third-party payers.

Several factors led to the growth and expansion of the cost-plus system from the end of the Second World War through the 1980s. The American Medical Association (AMA) controlled medical licensing. This not-for-profit institution effectively limited competition in the medical profession by requiring that anyone wishing to practice medicine must graduate from an AMA-approved medical school. Not-for-profit and government-run institutions dominated the hospital sector. Without the economic discipline provided by the profit motive, hospitals competed for physicians. Operating surpluses were directed toward investment in new services and expensive equipment by physician-dominated boards. As a result, excess capacity in beds, nursing staffs, and allied personnel were used to maximize the ability of physicians to generate income for themselves. Finally, Blue Cross and Blue Shield dominated the health insurance industry, and the addition of Medicare and Medicaid in the 1960s meant that not-for-profit payers were financing over half of all medical care provided. This dominance created an atmosphere in which cost was a secondary consideration. Without a cost constraint, the only thing that mattered was the patient’s health. Whether the procedure provided a net benefit was not an issue.

Restraint was not present on the demand side either, because insurance was paying the bills. Conventional health insurance distorts the decision-making process by making it appear that medical care is cheap at the point of purchase. Medical care, of course, is not cheap. But cost-plus reimbursement by third-party payers provides an incentive for people to demand interventions that provide little benefit.

The cost-plus system began to run into problems during the 1980s. No matter how prosperous a nation is, there is a limit to how much its people are willing to spend on any single item. As health care spending approached and exceeded 10 percent of gross domestic product, showing no signs of slowing down, policy makers and planners began to address concerns about the “health care crisis.” Thus began the bureaucratic struggle to slow the growth in health care spending.

In its early stages, this struggle focused on reimbursement strategies and restrictions on access to services. Medicare and Medicaid placed restrictions on providers by creating fee schedules and changing the method of reimbursement from retrospective payment to prospective payment. Private payers did the same, using the strategy of managed care. In both cases, the focus was not on changing buyer behavior but on limiting unnecessary procedures and services.

The move to prospective payment creates incentives on the supply side to limit care. The desires of patients becomes a secondary consideration, subordinated to the desire to control costs. The stage is set for the next phase of the cost-plus
cycle. Either the system will evolve into one in which individuals are motivated by the economic discipline of the market or into one dominated by the bureaucratic discipline of the government.

**Government Intervention in Medical Markets**

Government involvement in the medical marketplace is extensive. This involvement includes financing, direct provision, regulation, and subsidization. More than 45 percent of all health care spending comes directly from government sources, including Medicare, Medicaid, and the various health plans covering government employees and their dependents, both civilian and military. Government regulators are responsible for licensing, occupational health and safety, the administration of food and drugs, environmental protection, public health, and other oversight functions. Finally, the government uses features of the tax code to subsidize and encourage the provision of group insurance in employer-sponsored plans.

**Regulation**

The health care industry is one of the most heavily regulated industries in the United States economy. Price controls, entry restrictions covering both providers and hospitals, and regulations on the development and introduction of new drugs and medical devices are the major areas of regulatory control affecting the health care economy.

**Price Controls** The United States has a long history of placing restrictions on markets in the form of wage and price controls. World War II, the Korean War, and the wage-price freeze that was part of the stabilization program enacted during the Nixon Administration are a few of the instances in which government has attempted to fight inflation by freezing prices. Since the inception of Medicare and Medicaid, medical markets have been subject to price controls of one variety or another. In the beginning, physicians’ fees were limited to usual, customary, and reasonable (UCR) charges. Under UCR, physicians could charge the minimum of the doctor’s usual fee, defined by the median fee during the past year, and the customary fee, defined by the fees charged by other doctors in the area. The use of UCR resulted in a steady escalation of physicians’ fees. The formula left no reason for a physician’s usual fee to be lower than the customary fee charged in the area. If the usual fee was the minimum in the formula, Medicare paid the usual fee. As individual fees escalated, area fees escalated. The underlying incentive was always to make sure that your usual fee was not the minimum.

Medical prices continued to rise faster than the rate of overall inflation. As prices increased, spending increased. Efforts to limit spending growth shifted to the hospital sector in the early 1980s with the introduction of prospective payment. This new approach paid hospitals for an episode of treatment instead of using the usual cost-plus method. Under prospective payment, hospitals were paid according to the expected cost of treating a particular patient based on the principal diagnosis. If the actual cost of treatment was less than the payment, the hospital kept the surplus. If actual costs were greater, the hospital absorbed the loss or shifted the costs to other patients. Prospective payment changed the incentive structure completely. Hospitals were no longer rewarded for providing more services at a higher cost, and it was actually in their best interest to limit the amount and quality of services.

---

Observe other factors included in the reimbursement formula are the percentage of free care provided to indigents, whether the institution is a teaching hospital, and whether it is located in an urban area.
of services offered and to limit admissions and discharge patients as quickly as possible. Although hospital admissions moderated, and average length of stay fell dramatically, the use of outpatient services increased dramatically, leading some to question whether the potential for savings has been exhausted (Schwartz, 1987).

Attributing the spending restraint to the method of paying hospitals, the focus shifted back to physicians’ fees. The 1990s saw the advent of the relative-value scale for determining allowable physician fees. Basing fees on resource use, the relative-value scale is an attempt by bureaucrats to mimic markets. If the value scale is set correctly, prices will be set at levels that would exist in a competitive market. The relative-value scale has redefined the payment structure, treating evaluation and patient management services to higher relative fees while lowering relative fees paid for invasive procedures.

**Entry Restrictions** The government has a long history of licensing, certifying, and accrediting medical care providers. Although the stated purpose of these restrictions is consumer protection, some evidence exists that the self-interest of the providers may be the driving force behind the practice (Kessel, 1958; Moore, 1961). Licensing attempts to limit the likelihood that incompetent providers will treat uninformed patients. Originally, licensing merely placed restrictions on who was allowed to open a medical practice. As time passed, restrictions were expanded to cover a wide range of activities deemed unethical by practicing physicians. These activities included advertising, price cutting, and other conduct considered unprofessional. Clearly, licensing laws serve not only to protect patients but also to limit the number of practitioners, thus protecting physicians from would-be competitors.

**Limits on New Product Development** Congress established the Food and Drug Administration (FDA) in 1938 to oversee the entry of new drugs and medical devices into the medical market. The FDA does not allow new drugs on the market until they have been thoroughly tested and ultimately proved safe and effective. Even though the FDA has had several major successes in the past (the most notable was keeping the tranquilizer, thalidomide, off the U.S. market), the FDA approval process is the reason the time from the discovery of a promising chemical compound to drug approval averages 12 years.

The welfare effects of overly restrictive policies regarding new drug introduction are not always clear. Eliminating all risk is impractical, because using and consuming any drug carries with it some level of risk. The optimal level of risk is not zero, but the potential costs and benefits of introducing a new drug must be weighed. Regulators must consider the two types of statistical errors, referred to as Type I and Type II errors, when evaluating the safety and efficacy of a new drug. For simplicity, assume that a drug is either safe or unsafe and that the FDA either approves the drug for use or rejects it.

Type I error occurs when a safe drug is rejected; in other words, the review process results in a false negative. Type II error occurs when an unsafe drug is approved, a false positive. Regulators are much more concerned about avoiding Type II errors, approving drugs that harm patients. The consequences of approving an unsafe drug are obvious; patients suffer complications, get sicker, and die. The consequences of rejecting a safe drug are hidden; patients do not have access to a drug that might improve their health. Critics argue that the bias inherent in the regulatory process is harmful to the most vulnerable patients, those who are

---

7 In 1971 proof of efficacy was added as a requirement for new drug approval. In other words, the drug not only had to be safe, it had to work as claimed.
critically ill and have few alternative treatments available to them. Clearly, the FDA serves an essential function in the new-drug approval process. Allowing the market to be the sole determinant in drug availability would result in market failure by subjecting poorly informed patients to undue levels of risk.

**ISSUES IN MEDICAL CARE DELIVERY**

**FDA Regulation: The Case of the Cardiopump**

How can a patient who has no pulse give informed consent? Developers of the cardiopump, a cardiopulmonary resuscitation device for heart-attack victims, must find a satisfactory answer to this question before the FDA will allow further testing. Manual CPR exerts downward pressure on the chest and must rely on the chest to reexpand naturally. The cardiopump, which looks like a modified toilet plunger, exerts pressure in both directions, pulling blood back into the heart and oxygen back into the lungs.

The product is available elsewhere around the world, including England, Germany, Sweden, Canada, Australia, Japan, and Chile. In fact, it is a standard device in ambulances in Austria and France. But the FDA considers it a “significant risk device” that requires informed consent before it can be used on anyone in a medical trial. For the developers of the device, this designation represents a catch-22. Before the device can be used in a trial, the patient must give informed consent. But how can a patient with no pulse give informed consent?

The FDA is literally protecting patients to death. Approximately one million Americans have heart attacks every year. Of the 700,000 who are given CPR, only 20,000 survive to leave the hospital. Based on a limited sample in St. Paul, Minnesota, survival rates could increase by as much as 35 percent with the use of the cardiopump. That estimate fits comfortably within the range of a 10 to 50 percent improvement in expected survival rates. Extrapolating that number nationally implies that the device could save 7,000 lives annually.

The caution of the FDA is understandable. Regulators are sensitive to the criticisms that resound in the halls of Congress when a drug or medical device harms a single person during its testing. The agency’s success in keeping the tranquilizer thalidomide off the market in the 1960s is an excellent case in point. In contrast, the 7,000 people whose lives could be saved every year with the approval of the cardiopump are silent in their protest. When we are talking about life-or-death situations, would it not be wise to reconsider the requirement for informed consent?


**Tax Policy**

Policy makers and planners often use tax subsidies to encourage certain types of behavior. (Those who do not qualify for them call these subsidies “loopholes.”) Federal and state income tax provisions subsidize the purchase of health insurance. A key ruling by the Tax Court after the Second World War exempted certain nonwage benefits from being included in an employee’s taxable income. It was during this period of wage and price controls that government policy makers chose to use the power to tax—or in this case, the power not to tax—to encourage
employers to offer group health insurance to their workers. Since that time, group health insurance has been a nontaxable benefit for employees and, at the same time, a tax-deductible expense for employers.

Sheils and Hogan (1999) estimate that the subsidy in terms of forgone tax revenues exceeded $100 billion in 1998. The value of the subsidy to the individual is equal to the annual insurance premium paid by the employer multiplied by the individual’s marginal tax bracket. The benefits of the tax subsidy increase as a person’s income increases. If the annual premium paid by the employer is $4,000, a person in the 15 percent marginal tax bracket saves $600 a year in taxes by receiving the benefit instead of the income. In contrast, a person in the 42 percent tax bracket saves $1,680 on the same policy.\(^8\)

One of the major consequences of this tax subsidy is that individuals demand more health insurance when it is purchased by their employers than if they had received the income and bought it themselves. Most economists will agree that paying insurance premiums with before-tax dollars leads to overconsumption of medical care. Paying for expensive insurance with before-tax dollars makes more sense than paying for expensive medical care with after-tax dollars. As a result, indemnity insurance policies traditionally have had low deductible and copayment requirements.

---

\(^8\) The self-employed did not always enjoy the same tax preference. The Tax Reform Act of 1996 allowed the self-employed to deduct only 25 percent of the cost of personal health insurance (up to a maximum of total self-employment income). The percentage increased over time and reached 100 percent in 2003.
a dramatic reduction in the rate of increase in pharmaceutical prices in the past several years. Price increases during the 1980s regularly topped two to three times the rate of inflation in general. During the 1990s, drug inflation has moderated to as low as 5.7 percent in 1992. With the annual rate of inflation stabilizing at just over 3 percent, pharmaceutical companies still have a way to go before their goal of stability in real prices is within reach.

Responsibility for this price slowdown may be attributed to the buying power of the big institutional purchasers, such as Kaiser Permanente, a large West coast health maintenance organization (HMO). Using their monopsony power (see Back of the Envelope earlier in this chapter), the HMOs have been able to negotiate deep price discounts of as much as 40 to 60 percent below list prices and rebates of up to 75 percent of average wholesale prices. In return, the look-alikes are accepted on the HMO formulary, the list of drugs covered by the HMO.

Market forces will continue to exert downward pressure on drug prices as long as buyers have reliable alternatives to the established drugs. We may even expect that makers of the established drugs may soon begin discounting their products to discourage competition.


Government Failure

Even markets that work perfectly offer no guarantee that the efficient allocation of resources will satisfy the public’s desires for equity in the distribution of goods and services. On the other hand, no credible evidence supports government remedies as the answer for the perceived inequities either. It is debatable whether government solutions will always improve welfare. Markets may fail, but governments may be just as prone to failure. And correcting government failure is inherently more difficult than correcting market failure.

Few will question the intentions of government involvement in medical care. Everyone is in favor of improved access and lower costs. But careful consideration of the unintended consequences of government intervention is equally important. Choosing a health care strategy for yourself and your family is a difficult task. Choosing some other agent to make that decision for you is not only difficult, it can be dangerous. Transferring decision making from the private sector to the public sector substitutes bureaucratic discipline for economic discipline.

The notion of perfect competition in markets is just as rare as the notion of perfect democracy in political science (Becker, 1958). Criticism directed at market failure—without at least admitting the possibility of government failure—is dishonest, or at minimum naive. Voters face considerable obstacles in getting their collective voices heard. The interval between elections is long: two to six years. The viable choices are limited, usually to the two major-party candidates, and agreement with every aspect of a candidate’s platform is highly unlikely. Special interest groups, through subsidized lobbying efforts, have disproportionate influence on the decision-making process. And at the same time, protecting minority desires when government is by majority rule poses a problem.

These cautions should not discourage us from using government intervention as a strategy to ensure efficient market performance and equitable outcomes. But they should stand as a warning against relying too heavily on government to solve all our problems. Frequently, solutions proposed by well-meaning government

Policy Issue

Does imperfect government address the issues of equity and efficiency in health care delivery better than imperfect markets?
policy makers ignore the realities of the real world. We may not be able to create heaven on earth, but we may be able to improve the circumstances of millions of Americans with the right mix of market discipline and bureaucratic oversight.

The appropriate perspective in this debate is not whether the proposed system is efficient or fair (Pauly, 1997). No matter which alternative approach is chosen, it will be imperfect in its implementation. The appropriate perspective is whether efficiency and fairness are best addressed by imperfect government or imperfect markets.

Traditional microeconomics views the price mechanism as the invisible hand that leads to economic welfare maximization in a perfectly competitive market. In this chapter, we have examined the requirements necessary for competitive markets to result in equitable and efficient outcomes. Sources of market failure—including market power, externalities, and public goods—were described and discussed. Other sources of failure were applied to medical markets, including information problems, barriers to entry, and third-party payers.

The invisible hand is not able to perform its usual function in a system dominated by government decision makers. When government oversees production and consumption, it is the visible, tangible hand, or its equivalent, that determines prices. With complete knowledge of consumer preferences and producer capabilities, the efficiency problems could be solved. Following the reasoning of Lerner (1944), the planning agency must obtain the prices of all inputs and outputs, publish and distribute a list containing this information, and instruct all decision makers to act as if they were maximizers in a perfectly competitive market. In other words, substitute the superior wisdom of the planners for the collective wisdom of the masses.

Markets sometimes fail to produce the optimal level of output. The challenge facing policy makers is to intervene, not when they see market failure, but whenever government actions will actually take us closer to the social optimum.

If medical markets are to work, that is, if they are to produce acceptable levels of efficiency and equity, the following conditions must be present (Enthoven, 1988):

- Decisions must be made by well-informed, cost-conscious consumers. Motivated by self-interest, and adequately informed about treatment alternatives, cost-conscious consumers will economize because they will personally benefit from such behavior. The patient/buyer must be an active participant in the decision-making process if cost containment is to be achieved.
- Competition among providers is essential. Competition guards against undue concentration, because substitutes are readily available. Coupled with the first condition, consumer demand is sensitive to price changes.
- Cost-conscious decisions are possible only if consumers who desire to enter the market have money to spend. Often phrased in terms of equity, the real issue is economic self-sufficiency. As such, medical care markets require either universal insurance coverage or universal access to insurance. The choice depends on whether the majority of the populace is concerned with equal outcomes or equal opportunities. Satisfying this condition ensures that the system is morally acceptable to a majority of the people.
If one journal article can launch a career, Mark Pauly has shown us how it can be done. His 1968 article in the *American Economic Review*, entitled “The Economics of Moral Hazard,” has become essential reading for anyone desiring to understand the effects of health insurance on health care utilization and cost. After receiving his Ph.D. in 1967, Pauly catapulted himself into the epicenter of health economics with his classic treatise.

After brief academic appointments at Northwestern University and his alma mater, the University of Virginia, Pauly moved to the University of Pennsylvania’s Wharton School, where he became the Executive Director of the Leonard Davis Institute of Health Economics. Founded in 1967, the Leonard Davis Institute (LDI) has maintained a commitment to health services research and education in an interdisciplinary setting. Pauly was named Bendheim Professor in 1990 and is currently chairperson of the Health Care Systems Department.

One article can launch a career, but the reputation of a scholar is based on continuous research output. Continuous may not be the appropriate term to describe Pauly’s contribution to the health economics literature—unbelievable is probably better. Along with numerous books, articles, and monographs, his research interests encompass medical economics and the role of markets in medical care, national health care policy, and health insurance. In addition, he is a member of the editorial boards of the *Journal of Health Economics*, the *Public Finance Quarterly*, and the *Journal of Risk and Uncertainty*. He is also an elected member of the Institutes of Medicine of the National Academy of Science.

Pauly is one of a handful of health economists worldwide who argue that competition, when appropriately defined and understood, can work effectively in medical markets. Contrast this belief with the mainstream thought that gives little consideration to market solutions for the problems of medical care delivery and finance, and you begin to understand why many of his colleagues consider him an anomaly within the profession.

His belief that the incentive structure can shape both the behavior of patients and providers has resulted in his teaming with John C. Goodman, director of the National Center for Policy Analysis, in publishing the article “Tax Credits for Insurance and Medical Savings Accounts” in the Spring 1995 issue of *Health Affairs*. This innovative approach to health care reform recommends the use of tax credits, *medical savings accounts*, and high-deductible health insurance to improve both efficiency and equity in the health care sector. A colleague who does not share Pauly’s faith in market solutions referred to his belief in markets as a “disease.” If Pauly’s insistence on a place for markets in health care delivery and finance is a disease, he is not likely to accept the cure without a struggle, especially when the proposed cure is a government-run system.

On more than one occasion, after a previous speaker had stirred the audience into a feeding frenzy on the various evils of the U.S. medical care delivery system, Pauly has stepped to the podium only to quiet the crowd with his clear analytical approach and keen insight into the underlying issues, providing balance to a discussion in which balance is often lacking.

If the essential ingredients for making enlightened choices are knowledge and academic inquiry, Mark Pauly has advanced our ability to make enlightened choices through his outstanding contribution to the field of health economics and the economics of insurance.

Source: Mark V. Pauly, *curriculum vitae* and personal communication.
1. What is market failure? What are the major reasons that a free, unregulated market in medical care might not be optimal?

2. Proponents of a government-run health care system argue that the market does not work well in the medical care industry. What evidence do they use to support this claim?

3. Explain how market failure can be used to justify government intervention in medical care markets.

4. How do price controls affect the workings of a perfectly competitive market? Use a supply-demand diagram as part of your answer.

5. What assumptions of the perfectly competitive marketplace are violated in medical markets? How does each affect equilibrium price and quantity?

**REFERENCES**


The Economics of Consumer Choice

To explain consumer behavior economists use a simple model based on the concept of utility. The theory posits that individuals derive satisfaction, or utility, from consuming goods and services. The more goods and services consumed the higher the level of satisfaction achieved. A consumer's ability to satisfy his or her desire for goods is limited by the amount of money income to spend and the prices of the goods available for purchase. The three prerequisites for the development of a theory of consumer choice are: (1) there must be goods to buy, (2) consumers must have money to spend, and (3) they must be able to rank their preferences.9

As in all neoclassical economics, consumers are assumed to be maximizers. In the case where there are two goods available for consumption, consumers are interested in maximizing utility subject to a budget constraint, or

Maximize \( U = U(X, Y) \)

subject to \( M = P_X X + P_Y Y \)

where \( U \) is the level of utility, \( X \) and \( Y \) are the two goods in question, \( M \) is the money income available for spending on the two goods, and \( P_X \) and \( P_Y \) are their respective prices.

| CONSUMER PREFERENCES: INDIFFERENCE CURVES |

Economists depict consumer preferences graphically with indifference curves. An indifference curve illustrates the various combinations of goods that are equally satisfying to the consumer. In Figure 3A.1, having \( X_0 \) of good \( X \) and \( Y_0 \) of good \( Y \) places the consumer at point \( R \) on the indifference curve labeled \( U_0 \). Points \( S(X_1 \) and \( Y_1) \) and \( T(X_2 \) and \( Y_2) \) are likewise on \( U_0 \), indicating that these three combinations of \( X \) and \( Y \) provide the same level of satisfaction. The consumer is said to be indifferent as far as these three alternatives are concerned.

Higher levels of satisfaction are depicted by higher indifference curves. A combination of goods on indifference curve \( U_1 \), such as \( V \) is preferred to \( R \), \( S \), and \( T \). Similarly, \( W \) on indifference curve \( U_2 \) is preferred to \( V \). Because \( W \) is preferred to \( V \) and \( V \) is preferred to \( R \), \( S \), and \( T \), the transitive nature of preferences implies that \( W \) is also preferred to \( R \), \( S \), and \( T \).

When the consumer is able to rank all available alternatives, the set of indifference curves represents a preference map. Indifference curves serve the same purpose on this preference map that contour lines

9 The model does not require that consumers have the ability to attach numerical values to the utility levels. The requirement is that they be able to rank their preferences in an ordinal sense; e.g., most preferred to least preferred.
serve on a topographical map. As you move along an indifference curve, the level of utility stays the same. As you move along a contour line, the elevation stays the same. Move from one indifference curve to another and the level of utility changes. Move from one contour line to another and you move to a different elevation.

Indifference curves have certain properties that are important in the development of the theory of consumer choice. They are all negatively sloped, indicating that combinations of goods that have more of one good and the same or more of the other good are preferred. This property indicates that the goods in question are desirable. The consumer prefers more to less.

Indifference curves are typically drawn convex to the origin (they bow in, as shown in Figure 3A.1). Convexity implies that consumers are more willing to give up good Y for some amount of X when Y is plentiful. If the consumer has only a small amount of Y, it will take more X in the exchange to keep the consumer at the same level of satisfaction. The marginal rate of substitution (MRS) is defined as the amount of Y that the consumer would be willing to give up for a small increase in X and maintain the same level of utility. In other words, MRS is the importance attached to an additional unit of good X in terms of the amount of Y given up.

Movement from R to S on indifference curve U₀, results in a different combination of X and Y; point S has more X, but less Y than point R. The slope of U₀, defined as the change in the amount of Y relative to the change in the amount of X, is also the marginal rate of substitution. The movement from R to S may be broken down into two distinct moves. A move from R to A lowers the level of utility by reducing the amount of good Y. For small movements along U₀, this change in utility is equal to the marginal utility of Y (the change in utility resulting from a unit change in Y) multiplied by the total change in Y, or \( MU_Y \times (\Delta Y) \). Similarly, a move from A to S restores utility to its previous level due to the increase in the amount of good X. Using the same logic, that change is equal to \( MU_X \times (\Delta X) \). These two changes offset each other and are thus equal in magnitude, so \( \Delta Y/\Delta X = MU_X / MU_Y \). In other words, the slope of the indifference curve \( (\Delta Y/\Delta X) \), the MRS good X for good Y, equals the ratio of the marginal utilities of the two goods \( (MU_X / MU_Y) \).¹⁰

Indifference curves do not intersect one another. Intersecting curves would present a logical inconsistency. Points on any one indifference curve provide the consumer with the same level of utility. Points on a separate indifference curve are equally satisfying to the consumer but at a different level of utility. If two indifference curves intersect, the point of intersection would be on both curves simultaneously. The implication is that points on the two indifference curves represent the same and different levels of utility simultaneously.

### Consumer Constraints: The Budget Line

Consumers have a limited capacity to satisfy their preferences. Because of limited money income and positive prices for the goods and services, the ability to achieve the desired level of consumption is constrained. The consumer’s money income constraint may be written \( M = P_X X + P_Y Y \). By rearranging terms, the constraint may be written in the form of an equation, or budget line, as follows

\[ Y = (M/P_Y) - (P_X/P_Y)X \]

\( M/P_Y \) is the value of Y when \( X = 0 \) and is equal to the Y intercept. The corresponding X intercept, \( M/P_X \), is the value of X when \( Y = 0 \). The slope of the budget line, \( P_X/P_Y \), is the relative prices of the two goods. The budget line represents all combinations of goods X and Y the consumer is able to buy. Any combination of X and Y that is on or below the budget line is attainable. Given the prices of the two goods, the consumer does...

---

¹⁰ This derivation may be shown more formally using the Lagrangian multiplier method. The consumer’s effort to maximize utility \( U = U(X,Y) \) is constrained by limited money income, \( M = P_X X + P_Y Y \). The problem becomes one of maximizing \( L = U(X,Y) + \lambda (M - P_X X - P_Y Y) \). Setting the partial derivatives of \( L \) with respect to \( X, Y, \) and \( \lambda \) equal to zero gives

\[
\frac{\partial L}{\partial X} = aU/aX - \lambda P_X = 0 \\
\frac{\partial L}{\partial Y} = aU/aY - \lambda P_Y = 0 \\
\frac{\partial L}{\partial \lambda} = M - P_X X - P_Y Y = 0 
\]

Solving the first two equations for \( \lambda \) and setting them equal to each other yields

\[
\lambda = (aU/aX)/P_X = (aU/aY)/P_Y 
\]

In other words,

\[
\lambda = MU_X / P_X = MU_Y / P_Y 
\]
not have enough money to reach points above the budget line. In our model, we assume the consumer spends all budgeted money for the two goods, and thus ends up on the budget line, not below it.

Holding prices constant, changes in income will shift the budget line. Using Figure 3A.2, it can be seen that increases in income shift the curve to the right and decreases in income shift it to the left. Changes in relative prices will cause the curve to rotate. Holding \( P_Y \) constant, if \( P_X \) increases, the curve will rotate to the left. If \( P_X \) decreases, it will rotate to the right.

**CONSUMER CHOICE: THE CONCEPT OF EQUILIBRIUM**

Consumer preferences, graphically depicted by indifference curves, represent what the consumer is willing to buy. The money income constraint, depicted by the budget line, represents what the consumer is able to buy. Determining consumer choice is a matter of bringing together these two concepts—willingness to buy and ability to buy. The consumer’s decision on how to allocate scarce money income between the two goods is an attempt to match preferences with spending power—wants with affordability, willingness to buy with ability to buy—and in the process attain maximum satisfaction.

Individuals adjust their consumption behavior to the point where they cannot increase total utility without increasing their budget. Graphically, the choice may be shown as one of finding a point of tangency between the consumers' budget line and the highest attainable indifference curve. This point is identified by superimposing the preference map over the budget line and determining the unique point of tangency. This point of tangency represents an equilibrium because it is the only point where the slope of the indifference curve equals the slope of the budget line.

The consumer maximizes utility at point \( B \) in Figure 3A.3. Points like \( A \) do not represent equilibrium since the consumer can reach a higher level of utility simply by moving down the budget line toward point \( B \), spending the same amount of money, purchasing a different combination of \( X \) and \( Y \), and reaching a higher level of utility. Likewise, the consumer could move down indifference curve \( U_1 \), maintain a constant level of utility, and spend less money. At point \( B \), the slope of the indifference curve, \( MU_X/MU_Y \), is equal to the slope of the budget line, \( P_X/P_Y \). Thus, the equilibrium condition as already stated is satisfied. In equilibrium, \( MU_X/P_X = MU_Y/P_Y \). This condition may be rewritten \( MU_X/P_X = MU_Y/P_Y = \cdots = MU_n/P_n \).
It may be said the consumer maximizes utility when the last dollar spent on each good consumed provides the same increment to utility as the last dollar spent on every other good. This equilibrium condition provides one point on the individual’s demand curve for each good consumed, $X_0$ at price $P_X$. Changing the price of the good and finding the new level of consumption identifies additional points on the demand curve. Connecting all these price-quantity pairs in a separate graph traces out the actual demand curve.

**Implications of the Model**

The shapes of indifference curves depend on the consumer’s own assessment of the desirability of the available alternatives. Consumers with a strong preference for $X$ will have relatively steep indifference curves. Those with strong preferences for $Y$ will have indifference curves that are relatively flat. One possible extension of the model might be to examine the consequences of preference switching. The left-hand side of Figure 3A.4 shows the equilibrium between physicians’ office visits ($V$) and other uses of income ($I$). The healthy consumer will have a relatively flat preference map, indicating a strong desire to spend money on goods other than visits to the physician. With equilibrium at point $A$, this consumer will spend $Y_1$ income on all other goods and visit the physician $V_1$ times per year, resulting in a utility level of $U_1$.

The onset of an illness results in a preference switch, depicted by a steeper preference map on the right. The consumer now places more importance on visits to the physician relative to other spending. The result is a new equilibrium at point $B$, spending $Y_2$ on other goods, $V_2$ visits to the physician, and utility on indifference curve $U_2$. If the consumer cannot afford to reduce spending on other goods below $Y_1$, the preferred equilibrium cannot be attained. Instead the consumer will remain at point $A$, spending $Y_1$ on other goods, visiting the physician $V_1$ times, and attaining a lower level of utility, $U_0$.

**Conclusion**

The model of consumer choice discussed in this appendix is used to explain and predict consumer behavior. Even though consumers may not consciously apply this decision calculus in each and every situation, this does not mean that the model serves no useful purpose. Remember the model was developed to explain and predict. If it helps us accomplish these tasks, it serves us well.

---

11 The marginal utility of the last dollar spent on every good is equal to the $\lambda$ in the previous footnote.
In a world of competitive markets, firms that are successful in minimizing costs will earn a normal profit. Cost minimization is accomplished by the efficient use of resources. In this appendix, we will examine production and cost in a competitive market where firms attempt to maximize profits.

Production with Two Variable Inputs

Economists describe the production process as a functional relationship between inputs and outputs. The so-called production function shows the maximum output that can be produced from a given level of inputs using the available technology. Unlike utility, output is a measurable concept—bushels of grain, tons of steel, barrels of oil, or number of appendectomies performed. The inputs include land, natural resources, machinery, labor, and the entrepreneurial energies used to combine them and produce a product or service that people wish to buy. The production process with two variable inputs, labor \( L \) and capital \( K \), may be depicted in its generalized form

\[
Q = Q(L, K)
\]

where \( Q \) represents the amount of the good produced and \( Q(\ldots) \) the mathematical relationship describing the production process. Production functions are usually presented in one of three forms: a table, an equation, or a graph.

Figure 3B.1 summarizes the output levels that may be attained when labor and capital are combined according to the production function \( Q = 100 \, LK \). The amount of labor used in the production process is listed across the bottom of the table, and the amount of capital is listed along the left-hand side. Interpreting the data in the table is straightforward. For example, when five units of capital are combined with six workers, the firm is able to produce 548 units of output. Different combinations of labor and capital will result in different levels of output. As long as the inputs are used efficiently, the firm will produce exactly the level of output shown in the table.

Production Isoquants

It is possible to produce the same level of output using different combinations of the two inputs. For example, the firm may produce 316 units of output using ten units of capital and one unit of labor. The same level of output can be produced using five units of capital and two units of labor, two units of capital and five units of labor, or one unit of capital and ten units of labor. A similar observation may be made about 200 units of output, or 400 units, or any one of many different levels of output. The curves drawn in the body of the table represent the different combinations of \( L \) and \( K \) that produce the same level of output. These equal quantity curves are called isoquants, and serve the same purpose in production theory as indifference curves in consumer theory.

Plotting the isoquants in Figure 3B.2 provides a clear picture of the production levels that are attainable using the various combinations of labor and capital. The firm may use a number of different combinations of labor and capital to produce \( Q_1 \) units of output. Although only three are shown below, an infinite number of isoquants exist, one for every possible level of output. Because isoquants farther from the origin represent higher levels of output, \( Q_1 > Q_2 > Q_3 \).

Isoquants are usually drawn convex to the origin. The slope of the isoquant measures the ability to substitute one input for the other while maintaining the same level of output. As the firm adjusts its input mix, the ability to substitute, called the
marginal rate of technical substitution (MRTS), changes. When the production process uses a large amount of capital relative to labor, the marginal productivity of labor is high relative to that of capital. One additional worker can easily make up for the reduction of capital. Substitution of labor for capital is relatively easy and the marginal rate of technical substitution labor for capital (MRTS\_LK) is relatively high.

When the amount of capital employed is low relative to the number of workers, the marginal
productivity of labor is low relative to that of capital. It takes many more workers to make up for a reduction in capital. In other words, substitution of labor for capital is more difficult when capital is scarce relative to the number of workers competing for its use. Thus, as we move down an isoquant, using more labor and less capital, the $MRTS_{L,K}$ declines.

All along the isoquant, the marginal rate of technical substitution is the slope of the isoquant. It can be shown that $MRTS_{L,K}$ is the ratio of the marginal product of labor to the marginal product of capital ($\frac{MPL}{MPK}$).\(^{12}\) If labor and capital are perfect substitutes, $MRTS_{L,K}$ will be the same regardless of the amount of labor and capital used in the production process. In this case, the isoquant will be a downward-sloping straight line. If instead labor and capital are perfect complements, always used in fixed proportions, the isoquants are L-shaped.

**Production in the Short Run**

When a firm uses its resources efficiently, the only way to increase output is to increase the amount of inputs used. In most cases, it is easier to increase the workforce than it is to add capital equipment. Inputs whose levels can be adjusted quickly, such as labor, are called variable inputs. Inputs that take more time to increase, such as machinery, are called fixed inputs. The time lags required for these adjustments further define the production process as either short run or long run. In the case of a two-input production function, the long run is defined as the time period where both inputs are variable. The short run is the time period where one of the inputs, usually capital, is fixed.

In the short run, the only way to change output is to change the amount of the variable input used. The amount of the fixed input cannot be changed. In other words, the size or scale of the operation is fixed in the short run. From Figure 3B.1, short-run production may be shown by fixing the capital input at, say, five units and varying the amount of labor used from one to ten units. Presented in tabular form that information is shown in Table 3B.1.

From the first two columns, production increases as the number of workers hired increases. The average product ($AP_L$) and the marginal product of labor ($MPL$) may also be derived from the data on the total product of labor ($TPL$). The average product, a measure of technical efficiency, is calculated by dividing the total product of labor by the number of workers, or $AP_L = \frac{TPL}{L}$. The marginal product is the change in total product when one additional worker is hired. It is calculated by dividing the change in the total product by the change in the

<table>
<thead>
<tr>
<th>Units of Labor</th>
<th>Total Product</th>
<th>Capital-Labor Ratio</th>
<th>Average Product</th>
<th>Marginal Product</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>∞</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>1</td>
<td>224</td>
<td>5.00</td>
<td>224</td>
<td>224</td>
</tr>
<tr>
<td>2</td>
<td>316</td>
<td>2.50</td>
<td>158</td>
<td>92</td>
</tr>
<tr>
<td>3</td>
<td>387</td>
<td>1.67</td>
<td>129</td>
<td>71</td>
</tr>
<tr>
<td>4</td>
<td>447</td>
<td>1.25</td>
<td>112</td>
<td>60</td>
</tr>
<tr>
<td>5</td>
<td>500</td>
<td>1.00</td>
<td>100</td>
<td>83</td>
</tr>
<tr>
<td>6</td>
<td>548</td>
<td>0.83</td>
<td>91</td>
<td>48</td>
</tr>
<tr>
<td>7</td>
<td>592</td>
<td>0.72</td>
<td>85</td>
<td>44</td>
</tr>
<tr>
<td>8</td>
<td>632</td>
<td>0.63</td>
<td>79</td>
<td>40</td>
</tr>
<tr>
<td>9</td>
<td>671</td>
<td>0.56</td>
<td>75</td>
<td>39</td>
</tr>
<tr>
<td>10</td>
<td>707</td>
<td>0.50</td>
<td>71</td>
<td>36</td>
</tr>
</tbody>
</table>

\(^{12}\) The MRTS at any point on an isoquant may be derived by taking the total differential of the production function $Q = Q(L, K)$, and setting it equal to zero.

$$dQ = (\frac{\partial Q}{\partial L})dL + (\frac{\partial Q}{\partial K})dK = 0$$

As the amount of $L$ and $K$ change along an isoquant, the level of output does not change, or $dQ = 0$. Solving this equality for the slope of the isoquant, $\frac{dK}{dL} = (\frac{\partial Q}{\partial L})/(\frac{\partial Q}{\partial K})$. Since $(\frac{\partial Q}{\partial L})$ equals $MP_L$ and $(\frac{\partial Q}{\partial K})$ equals $MP_K$.

$$\frac{dK}{dL} = \frac{MP_L}{MP_K}MRTS_{L,K}$$
number of workers used in the production process, or
\[ MP_L = \frac{\Delta TP_L}{\Delta L}. \]

The production function utilized in this discussion illustrates an important empirical observation in short-run production, the **law of diminishing returns**. Holding the amount of capital constant, each added worker has less capital on average to work with, as evidenced by a constantly declining capital-labor ratio \((K/L)\). So each additional worker contributes less to output than the previous worker. The law of diminishing returns is not based on an economic theory, it is physical law that holds true for production in general.

Although the law of diminishing returns characterizes every short-run production process, marginal product and average do not always decline from the outset. Some production processes display increasing marginal and average product initially due to the benefits derived from specialization and the division of labor. Figure 3B.3 presents a generalized short-run production function. As the number of workers increases, total product increases at an increasing rate up to point \(A\). Beyond point \(A\), production continues to increase as more workers are used, but at a decreasing rate. The rate of increase in output slows until a maximum output is reached at point \(B\). Beyond point \(B\), given the amount of capital available per worker, further increases in output are not possible. Adding workers actually decreases output.

Firms do not operate where the marginal product of an input is negative. Doing so would imply the firm could increase its output by decreasing the amount of the input used, increasing revenue and lowering cost. Thus, efficient production occurs when the marginal products of all inputs are positive.

**Figure 3B.3: Generalized Production in the Short Run**
**Optimal Input Use**

The profit-maximizing firm will attempt to maximize output from the resources committed to production. The firm faces a resource constraint determined by the cost of inputs and the amount of money it is willing to spend. When two inputs, labor \((L)\) and capital \((K)\), are used in production, the constraint may be written \(C = wL + rK\), where \(C\) is the total cost, \(w\) is the wage rate paid labor, and \(r\) is the unit cost of capital. This cost constraint may be rewritten as an isocost curve, or \(K = \left(\frac{C}{r}\right) - \left(\frac{w}{r}\right)L\). The isocost curve is shown in Figure 3B.4, and may be interpreted as all possible combinations of \(L\) and \(K\) that can be hired for a total cost equal to \(C\) when input prices equal \(w\) and \(r\). The more money the firm is willing to commit to production, the farther the isocost curve is from the origin and the greater the output that can be produced.

The slope of the isocost curve is the relative price of the inputs, or \(-\frac{w}{r}\). Combining the isocost map with the relevant isocost curve allows us to determine the combination of inputs the profit-maximizing firm will choose. Maximizing output at a given level of cost requires that the firm use the optimal or least-cost combination of the inputs. This is shown in Figure 3B.5 at point \(E\) where the isocost curve is just tangent to the isoquant \(Q_1\). At the point where the isoquant is tangent to the isocost curve, their slopes are equal. In other words, the slope of the isoquant, or \(\frac{MRT_{L,K}}{MPL/MPK} = \frac{w}{r}\), equals the slope of the isocost curve, or \(\frac{w}{r}\), when the firm is using the least-cost combination of inputs \(L\) and \(K\). Formally, this equilibrium condition may be written \(\frac{MRT_{L,K}}{MPL/MPK} = \frac{w}{r}\).

The equilibrium condition may also be written \(\frac{MPL}{w} = \frac{MPK}{r}\). In this form it is easily seen that firms adjust the amounts of labor and capital used until the marginal product from the last dollar spent on labor is equal to the marginal product from the last dollar spent on capital.

### Extensions of the Model

The optimal input mix for producing a given level of output will change as the relative prices of the inputs change. Figure 3B.6 illustrates the least-cost method.

---

13 The mathematical derivation of the equilibrium condition in production mirrors that of the equilibrium condition in consumer theory. Using the Lagrangian multiplier method, it can be shown that the firm’s effort to maximize output \(Q = Q(L, K)\) is limited by a total cost constraint, \(C = wL + rK\). The problem becomes one of maximizing \(L = Q(L, K) + \lambda(C - wL - rK)\). Setting the partial derivatives of \(L\) with respect to \(L, K,\) and \(\lambda\) equal to zero gives

\[
\frac{\partial L}{\partial L} = \frac{\partial U}{\partial L} - \lambda w = 0
\]

\[
\frac{\partial L}{\partial K} = \frac{\partial U}{\partial K} - \lambda r = 0
\]

\[
\frac{\partial L}{\partial \lambda} = C - wL - rK = 0
\]

Solving the first two equations for \(\lambda\) and setting them equal to each other yields

\[
\lambda = \left(\frac{\partial Q}{\partial L}\right)/w = \left(\frac{\partial Q}{\partial K}\right)/r
\]

In other words,

\[
\lambda = \frac{MP_L}{w} = \frac{MP_K}{r}.
\]
of producing \( Q^* \) medical care at two different prices for physicians’ services. When the price of physicians’ services is high \( (P_0) \), equilibrium will be at point \( H \), using \( S_H \). If physicians are paid less, holding the price of other medical inputs \( (P_o) \) constant, the same level of medical care will be provided using a different mix of physicians’ services and other medical inputs. At low physicians’ prices \( (P_1) \), equilibrium will be at point \( L \), using \( S_L \) physicians’ services.

The model provides several interesting implications. When the fees paid physicians are relatively high, the physician-population ratio will be relatively low and patients will visit their doctors less often. Additionally, higher physicians’ prices encourage the use of other medical inputs. Thus, when physicians’ prices are higher, we expect medical care to be produced using more capital per patient.

**ESTIMATING PRODUCTION FUNCTIONS**

The simplest and most widely used production function in empirical work is the Cobb-Douglas variety. The Cobb-Douglas production function may be written as \( Q = AL^\alpha K^\beta \) where \( \alpha \) and \( \beta \) are positive parameters estimated from the empirical data. Using this functional form, the exponents represent output elasticities, or the percentage change in output for every 1 percent change in the quantity of the input used. In the case of the labor input, a 1 percent increase in \( L \) will result in an \( \alpha \) percent increase in \( Q \). Likewise for capital, a 1 percent increase in \( K \) will result in a \( \beta \) percent increase in \( Q \). If \( \alpha + \beta = 1 \), the production function exhibits constant returns to scale. In this case a 1 percent increase in the amount of both inputs used yields a 1 percent increase in output. If \( \alpha + \beta > 1 \), say 1.2, then a 1 percent increase in \( L \) and \( K \) results in a 1.2 percent increase in \( Q \) and the production function exhibits increasing returns to scale.

The Cobb-Douglas production function is estimated empirically by first taking the logarithm of both sides, resulting in

\[
\log Q = A + \alpha \log L + \beta \log K
\]

Regressing \( \log Q \) on \( \log L \) and \( \log K \) provides estimates of the output elasticities from the estimated coefficients (refer back to the statistical appendix to Chapter 2 for the discussion on regression analysis).

---

14 The marginal products of labor and capital for a Cobb-Douglas production function are determined as follows:

\[
\frac{MP_L}{Q} = \frac{\partial Q}{\partial L} = \alpha L^{\alpha-1}K^\beta = \alpha (Q/L)
\]

\[
\frac{MP_K}{Q} = \frac{\partial Q}{\partial K} = \beta L^\alpha K^{\beta-1} = \beta (Q/K)
\]

The output elasticities \( \varepsilon_l \) and \( \varepsilon_c \) are

\[
\varepsilon_l = \frac{(L/Q)/(Q/L)}{Q/L} = \frac{L/Q}{Q/L} = \alpha
\]

\[
\varepsilon_c = \frac{(K/Q)/(Q/K)}{Q/K} = \frac{K/Q}{Q/K} = \beta
\]
**Production to Cost**

Cost may be divided into two categories: fixed and variable. Costs associated with the fixed inputs, costs that do not change as the level of production changes, are **fixed costs**. Costs associated with the variable inputs, costs that change as the level of production changes, are **variable costs**. Using the two-input production function introduced above with capital representing the fixed input and labor the variable input, capital costs are fixed costs and labor costs are variable costs.

Total cost is the amount that must be spent on all inputs to produce a given level of output, including all applicable opportunity costs. Total cost is comprised of fixed costs and variable costs, all the costs associated with the capital inputs and all the costs associated with the variable inputs. Using the same notation developed earlier, the total cost function may be written

\[ C = rK + wL. \]

In other words, the production function and the prices of inputs determine the firm’s total cost function. The production function determines how much capital and labor are used in the production process, and the respective input prices determine the total amount spent on each input.

In practice, the short-run total cost curve may be derived from the short-run production function. With the amount of capital available fixed in the short run, \( rK \) is constant and represents fixed costs. In order to increase the level of output, the amount of labor used must increase. The production function determines the amount of labor needed to produce any given level of output. The short-run variable cost associated with each level of output \( (Q) \) is determined by the amount of labor required \( (L) \) multiplied by the cost of labor \( (w) \). Figure 3B.7 depicts the short-run total cost function associated with the production function shown in Figure 3B.3. Note the symmetry. In the range of output where production increases at an increasing rate (up to point \( A \) in Figure 3B.3), cost increases at a decreasing rate. When production increases at a decreasing rate, cost increases at an increasing rate.

This relationship is much clearer when viewed from the perspective of the short-run average and marginal cost curves. By definition, average variable cost \( (AVC) \) is the total variable cost \( (TVC) \) divided by the level of output produced \( (Q) \), or \( AVC = TVC/Q \).

Since \( TVC = wL, \) \( AVC = wL/Q \) or \( w(L/Q) \). Remembering that \( Q/L \) is the average product of labor \( (AP_L) \), we note \( AVC = w/AP_L \). As the average product of labor increases, average variable cost decreases. When \( AP_L \) reaches its maximum \( AVC \) reaches its minimum. As \( AP_L \) decreases, \( AVC \) increases.

Likewise, the relationship between marginal cost \( (MC) \) and the marginal product of labor \( (MP_L) \) can be determined: \( MC = \Delta TVC/\Delta Q \). Substituting \( wL \) for \( TVC \) yields \( MC = \Delta wL/DQ \). In competitive labor markets, the firm is a price taker, so the only way to change \( wL \) is to change \( L \), implying \( MC = \Delta L/\Delta Q \). Because \( DQ/\Delta L \) is the marginal product of labor, \( MC = w/MP_L \). As marginal product increases, marginal cost decreases. When \( MP_L \) reaches its maximum, \( MC \) reaches its minimum. As \( MP_L \) decreases, \( MC \) increases. Thus, we expect short-run average costs and short-run marginal costs to be U-shaped, initially decreasing, then reaching a minimum, and finally increasing.

The relationship between average costs and marginal costs is shown in Figure 3B.8. Average total cost is the sum of average fixed cost and average variable cost. As long as marginal cost is below average cost, notice that average cost decreases. When marginal cost rises above average cost, average cost

---

Oppportunity costs include both the explicit costs associated with actual payments to resources used in production and the implicit costs associated with the owners’ time and investment. Explicit costs are all those costs recorded by the firm for accounting purposes, including rent paid on buildings, salaries paid to workers, and interest paid on loans. Implicit costs are the opportunity costs of using resources owned by the firm, including forgone earnings on money invested in the business.
begins to increase. Thus, marginal cost intersects each average cost curve at its respective minimum.\textsuperscript{16}

**LONG-RUN COSTS**

Long-run costs are also U-shaped, but for different reasons. In the long run the firm has the option of increasing the size of its physical plant. Doing so often means the use of more efficient equipment, specialized labor, and lower average costs. The economic principle is called economies of scale. The long-run average cost curve may be thought of as an envelope curve, depicting the least-cost option for producing each level of output. Figure 3B.9 shows the long-run average costs associated with three different plant sizes: small ($AC_S$), medium ($AC_M$), and large ($AC_L$). The minimum cost of producing each level of output depends on the size of the physical plant. If the desired level of output is less than $Q_1$, the firm will minimize cost if it uses the small plant. For output levels between $Q_1$ and $Q_2$, costs are minimized using the medium-sized plant. For output levels greater than $Q_2$, the large plant minimizes costs.

The envelope curve in the diagram on the righthand side depicts all possible plant sizes. Competition will force the firm to use the plant whose costs are given by $AC_4$, the optimal plant. Firms that do not use this sized plant will find themselves with higher costs than their competitors, and they will lose money.

**CONCLUSION**

The theory discussed in this appendix provides a summary of the economic theory of the firm. The material is not intended to cover the full range of topics presented in a microeconomics course, but it should be sufficient to give the reader a broad overview of the standard neoclassical theory of the firm.

\textsuperscript{16} For those with a little knowledge of calculus, the intersection of average and marginal cost at minimum average cost may be shown by noting that the slope of the average cost curve is equal to zero at its minimum; that is, its first derivative is equal to zero at its minimum. For the average variable cost curve

\[
\frac{dAVC}{dQ} = \frac{d(TVC/Q)}{dQ} = 0
\]

Dividing both terms in the numerator and factoring out $1/Q$ results in

\[
\frac{1}{Q}(MC - AVC) = 0
\]

For the right side of the expression to equal zero, $MC - AVC$, or marginal cost equals average variable cost when the slope of average variable cost equals zero (when $AVC$ has reached its minimum).