

MARTIN NEIL BAILY

McKinsey & Company, Inc.

ALAN M. GARBER

*Department of Veterans Affairs and
Stanford University*

Health Care Productivity

IN ALL OF THE INDUSTRIAL COUNTRIES, a high fraction of gross domestic product (GDP), ranging from approximately 7 percent in the United Kingdom to 14 percent in the United States, is devoted to health care. In recent years policymakers have been forced to try to trim health care benefits or other social services, and the health care systems of almost all the industrial countries have come under significant pressure to control expenditures and improve performance.

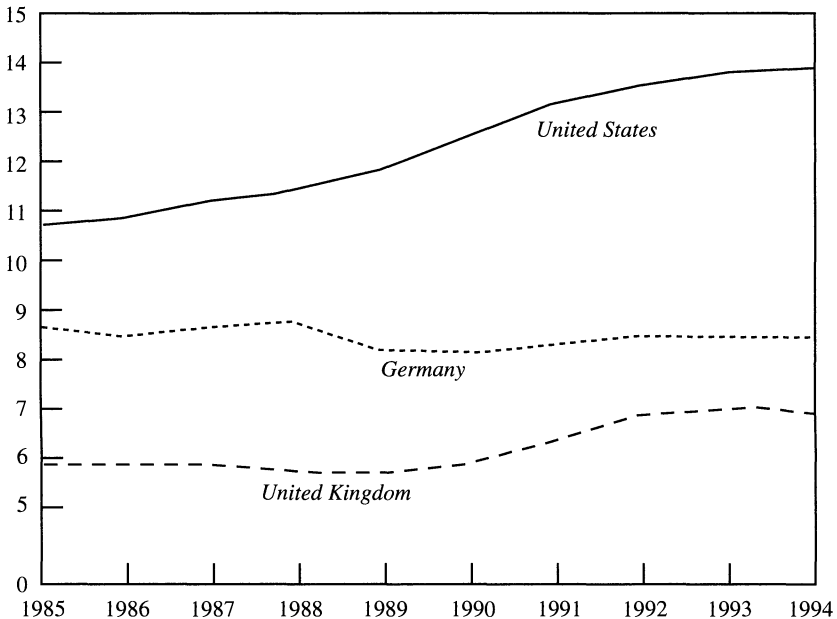
Although each nation's health care system operates with a mixture of regulation and market mechanisms, there are great differences among them. And these differences suggest that policymakers could learn important lessons by comparing performances across countries. Thus far, however, no single system is recognized as being the most productive or as having achieved the right blend of competition and regulation. No system provides the paradigm for others.

Some observers of the U.S. health care system argue that aggregate data indicate poor performance. Figure 1 shows aggregate spending data for Germany, the United Kingdom, and the United States; with the highest level of GDP and the largest fraction of spending, the United States spends much more per capita on health than the other two countries. Figure 2 provides a simple aggregate performance measure. Average life expectancy at birth in the United States is lower than in Germany and the United Kingdom, although, as the figure also shows,

The authors would like to thank the participants in the Brookings meeting for their comments. Debbie Kelsey and Gretchen Bossert provided substantial assistance in preparing this paper.

Figure 1. Growth in Health Care Spending in the United States, United Kingdom, and Germany

Percentage of GDP

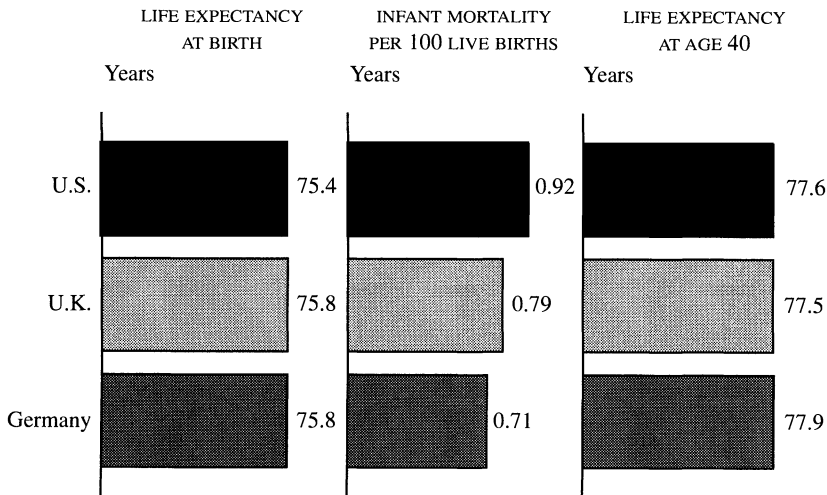


Source: McKinsey Global Institute and the McKinsey Health Care Practice (1996).

Note: The purchasing power parity (PPP) exchange rates used in this figure are for all of gross domestic product (GDP), not for health care. The difficulties that confront international price comparisons for health services led us to conclude that the overall GDP PPP is a better measure of the real opportunity cost of health care spending in the three countries.

higher infant mortality is the principal cause of the lower life expectancy in the United States. Life expectancy for adults is more nearly equal. Even so, with higher spending and outcomes that are apparently worse or no better, it would be easy to conclude that the health care delivery system in the United States is unproductive relative to other countries.

In fact the available aggregate evidence does not establish the validity of this conclusion. Other factors, such as differences in lifestyles, diet, and health practices, differ among countries and can have major impacts on mortality. Many forms of disease treatment improve the quality of life even if they do not extend it. Moreover, observed differences in mortality may be associated with differences in access to care, rather than with differences in the productivities of the health care

Figure 2. Life Expectancy and Infant Mortality, 1990

Source: McKinsey Global Institute and the McKinsey Health Care Practice (1996).

delivery systems themselves. Thus, aggregate-level evidence does not reveal whether the U.S. health care system is more costly than other systems because Americans pay health care providers more, because productivity is different, or because Americans simply demand more treatment.

To gain a better understanding of this issue, McKinsey and Company launched a project that had two major objectives:

- To assess differences in relative productivity at the *disease level* among the health care systems of three industrialized nations—the United States, Germany, and the United Kingdom.
- To examine the major causes of these differences by focusing on variations in diagnostic and treatment approaches and relating such variations to provider incentives and supply constraints that arise from the structure and regulations of each country's health care system.

The project focused on productivity, not on the overall performance of the health care system.¹ Productivity, a critical determinant of health

1. The working team consisted of Lynn Dorsey, Cuong Do, Andrew Gengos, Elise

care system performance, can be defined as the physical inputs used (labor, capital, and supplies) to achieve a given level of health outcomes in treating a specific disease. In other words, the concept of productivity can be applied to health care by viewing the management or treatment of a disease as the fundamental “production process” in health care. By improving productivity, countries can alleviate spending pressure on their systems while maintaining the level of outcomes, or they may be able to improve outcomes without increasing spending.

Four important qualifications to the scope of the project must be kept in mind. First, productive efficiency does not always imply allocative efficiency; one country can produce a great deal of health from limited resources, demonstrating high productivity, yet provide too little health care for its population overall. Conversely, it might achieve little additional outcome with the use of greater resources, doing so with productive efficiency (that is, no alternative production process would result in the same level of health with less resource use), but producing more than the socially optimal quantity. Although the project did not study this important element of the policy choice, understanding the impact of economic incentives on health care productivity can help policymakers understand any trade-off between allocative and productive efficiency that may exist.

Second, the project focused on only four specific diseases—diabetes, cholelithiasis (gallstones), breast cancer, and lung cancer—so the observations do not paint a complete picture of the health care systems in the three countries. Nevertheless, these four conditions are common, important, and representative. Thus we believe many of the lessons that emerge from the cases can be generalized.

Third, the figures presented for inputs and outputs are national averages, which obscure the large variations that exist within countries. In the United States, for example, health care can be reimbursed on an indemnity (fee-for-service) or on a capitated basis (fixed annual payment to the provider organization for all covered services for each

Russi, John Goree, Frank Basile, Paul Brody, David Crawley, Alexis Dormandy, Thomas Gertsner, Nicolaus Henke, Dolores Heras, Michele Holcomb, Debbie Kelsey, Keiko Kin, Karl Krista, Joan Mehn, Uma Muthu, Vikram Narasimhan, Sheryl Sandberg, Mary Ann Aitken, Julie Eskay, Diane Gutheil, Donna Gregory, Ruby Kapadia. Kathy Knauss, and Doreen Welborn. The Outside Advisory Committee included Kenneth Arrow, Axel Börsch Supan, and Ted Hall as well as the two authors of this article. Bernard Ferrari, Bill Lewis, and Charles Schetter oversaw the completion of this project.

enrollee), or some blend of the two. The mix of capitation (or managed care) and traditional fee-for-service care varies greatly around the country. Just as the financing mix varies, so do clinical practices. Treatments vary by city and region.² The project team was well aware of these differences but lacked the data necessary to carry out a more disaggregated formal analysis. Based on interviews with health care practitioners in the three countries, we think that the country differences that emerge in the cases are large and real and are representative of the three countries.

Fourth, the case study data were drawn from the mid- to late 1980s. Since then the health care systems in all three countries have been changing; structural and regulatory reforms have occurred in the United Kingdom and Germany, and managed care has grown rapidly in the United States. Some of the patterns of care that we observed may thus no longer be the norms in their respective countries. Conclusions about the impact of economic incentives on provider behavior are likely to remain valid, however, and many of the institutions and regulations that influenced patterns of care then remain in place today.

Measurement and Assessment Issues

Levels of inputs used in each country, along with disease outcomes, can be measured directly, but these are not in themselves sufficient to calculate relative productivity. That is because the treatment of a given disease is fundamentally a diminishing returns or, to state it more weakly, a nonincreasing returns activity. For the disease cases, production can be considered on a per-case or a per-patient-treated basis. When the quantity of health services is increased, it can be on one of two margins: greater resource intensity in the treatment of each patient, or an increase in the number of patients treated. Along both of these margins, diminishing returns are likely.³

Along the treatment-intensity margin, at least in the relevant range

2. The U.K. data are for southern England, not the entire country.

3. Ernst Berndt argued that we are really talking about returns to scale, not diminishing returns to a single factor. We think, however, that applying incremental doses of capital, labor, and supplies to a given disease population is more intuitively thought of as a diminishing returns activity. The idea seems clear irrespective of terminology.

of output, diminishing returns are likely to be the rule—most conditions will respond to additional units of treatment or resources devoted to diagnostic and other management with successively smaller units of health output. If patients who derive the greatest benefit from care are the first to receive it, diminishing returns are also likely to characterize an expansion in the number of patients treated. In a system that effectively “triages” candidates for any specific treatment, the patients most likely to benefit from that treatment will receive care first, the next most likely patients second, and so on. In the United States, this form of triage is modified by the market, where those with comprehensive insurance or the greatest willingness to pay may be more likely to receive, say, gallbladder surgery than those patients without explicit coverage. But because most of the uninsured have some access to treatment in the United States, overall there will be diminishing returns in the production of health care in all three countries.

Diminishing returns in medical treatments carry several implications. The country that devotes more resources to a disease will have lower average productivity than the other country.⁴ This is not a reflection of inefficiency, but a condition inherent in diminishing returns production. To evaluate productive efficiency, one must consider not only how to measure inputs and outcomes, but also how to assess the results. A specific definition of productive efficiency is thus needed in order to rank outcomes.

Estimating Inputs Used

To estimate the inputs used, the McKinsey team developed a detailed model of each disease treatment process. The model incorporated the important steps in the process, the key choices and decisions that providers face at each step, and the resulting resource implications. The sources of data used to explain the steps of the treatment process and associated inputs included published descriptions in the medical literature, analyses of national databases (such as hospital discharge information), and interviews with practitioners and administrators in each country.

Physical inputs included labor (from physicians, nurses, technicians,

4. This is the case for two countries operating on the same production function. See the discussion of figure 4.

and other health care providers), supplies (such as medications, surgical instruments, and X-ray film), and capital (such as diagnostic equipment and hospital facilities, where data were available). For the labor inputs associated with an inpatient stay, we used a simplified model that multiplied each country's average staffing level per day of hospital stay by the average length of stay for treating the specific disease. Because the units of measurement for each input vary, inputs were standardized using a base unit cost, which was an hour of a surgeon's time. (Note that the choice of the base unit is arbitrary and has no effect on the results.) We then calculated the weighted sum of the labor, supplies, and capital used to obtain an aggregate measure of physical inputs for each disease treatment process in each country. More detail on our input methodology is given in the final report of the project.⁵

Administrative Costs

Omitted from the case analysis is any estimate of administrative costs. We are focusing on the inputs used to treat the diseases, not on the health care systems in total. This issue is addressed explicitly later in the paper when the case results are related to the aggregate data. Administrative costs are estimated to be about 24 percent of total spending in the United States, 13 percent in Germany, and 16 percent in the United Kingdom.

Estimating Outcomes

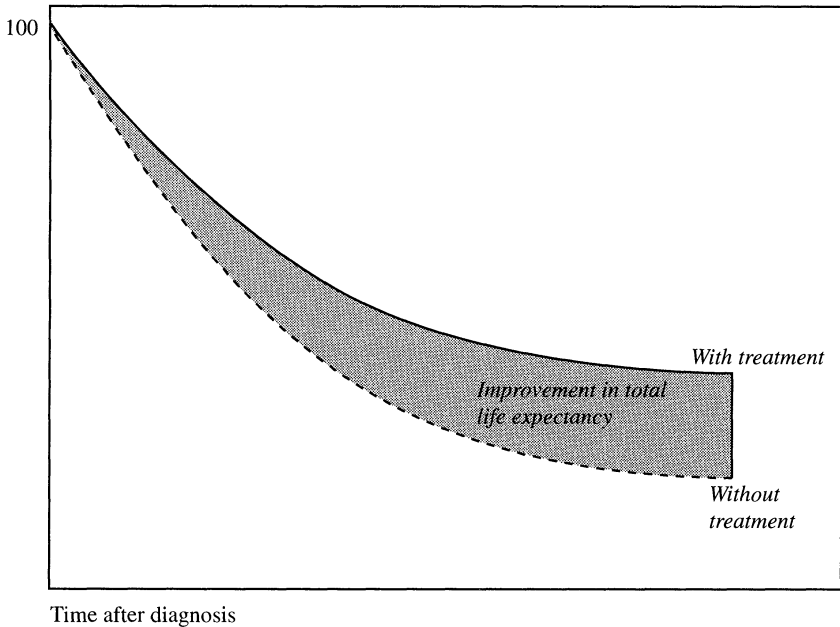
Outcome measures pertinent to each disease were adjusted for differences in disease incidence across countries. Like the input measures, outcome measures were derived from literature reviews, database analyses, and interviews with clinical experts. We derived our outcome measures by comparing the expected health outcomes with treatment in each country to the outcomes without treatment, which are presumably similar in each country. An example using mortality as the outcome measure is shown in figure 3. Because the outcome represents a change in health status, it is necessary to quantify health status expected for each disease and to estimate the improvement in health that results from the disease treatment process.

5. McKinsey Global Institute and the McKinsey Health Care Practice (1996).

Figure 3. Survival Curves with and without Treatment

Per case

Percentage of survivors



Source: Authors' conceptualization.

QUANTIFYING HEALTH. Outcomes for each disease can be quantified using either survival rates or calculations modeling the quality of life. Survival rates are relatively easy to assess and are appropriate measures for lung and breast cancers, in which the primary goal of treatment is to reduce the high level of mortality. Outcomes for the cancers can thus be measured as years of life expectancy or life years (LYs). For diabetes and cholelithiasis, whose mortality rates are much lower, the primary treatment goal is to reduce the incidence and severity of disabling or painful but nonfatal complications of the disease. Because treatment is intended to improve the quality of life—not only its duration—survival is an inadequate measure of health outcomes for these diseases. For these diseases, we used the Kaplan-Bush Index of Well-Being to calculate outcomes in quality-adjusted life years (QALYs). We believe that our major findings would have been similar if we had applied

another method for quantifying quality-of-life effects. Although quality of life is also relevant in the cancers, it is quite difficult to measure with available data and accounts for less of the intended benefits of treatment.

MEASURING IMPROVEMENT IN HEALTH FROM TREATMENT. Outcomes without treatment are usually unknown and can be influenced by the patient's baseline health status, which in turn reflects lifestyle, cultural factors, genetics, and so on. For some of the disease cases, we assumed that the baseline or untreated health outcome would be the same in each country, so that the absolute levels of health in treated patients would be a valid basis for comparing the outcomes of treatment in each country. Although available data are not conclusive, they are consistent with this assumption. Our studies were modified for situations in which this assumption might not have been valid; for example, in the diabetes case, we compared the outcome for the United Kingdom to that for U.S. whites because nonwhites are known to have higher rates of diabetes and of diabetic complications than whites.

Baseline health status was estimated for some diseases in order to calculate the change in outcomes with treatment. As mentioned earlier and described in greater detail later, we used this approach to assess relative productive efficiency in those cases in which one country achieved better outcomes using more inputs.

Determining Levels of Productive Efficiency

We defined productive efficiency by the relative positions of the health production functions for each country. Because we observed only one point for each country for each case, we could not trace out the shapes of the production functions. Yet the assumption of diminishing returns and the observations of inputs and outcomes in each country enabled us to draw inferences about relative productive efficiency in most cases.⁶

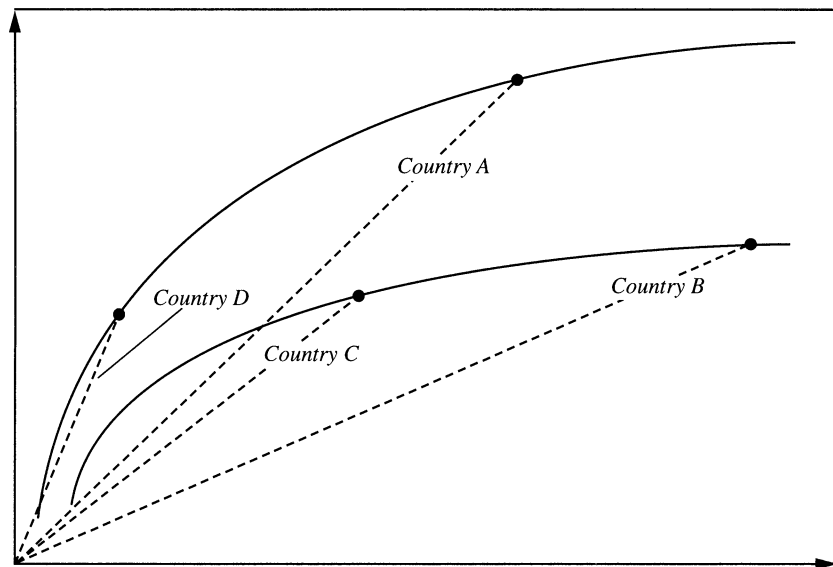
The simplest case for our comparisons is illustrated by Countries A and B in figure 4: Country A achieves better outcomes while using fewer inputs, so Country A must be more productive. Countries A and C depict the more common situation, in which one country uses more

6. Ernst Berndt points out that our assumption is stronger than we need. As long as there are not increasing returns to scale, our cross-country conclusions go through.

Figure 4. Assessing Productive Efficiency

Per case

Outcomes



Inputs

Comparison 1: A vs. B

- A is more productive because it achieves better or equal outcomes with less inputs

Comparison 2: A vs. C

- A is more productive because it has higher average productivity (ratio of outcomes to inputs) and treatment process does not show increasing returns with additional care inputs

Comparison 3: C vs. D

- C has higher inputs and outcomes but lower average productivity; productive efficiency can only be determined based on detailed knowledge of treatment process

Comparison 4: B vs. C

- No apparent difference in relative productive efficiency; one country may have preferred input/outcome combination based on cost-effectiveness analysis

Source: Authors' conceptualization.

resources and has better outcomes than another. In the situation shown, Country A has better outcomes than Country C and greater average productivity. This means that, with diminishing returns everywhere, Country A must be on a higher production function, and so it has greater productive efficiency.

Note that this is a pure productive efficiency comparison and does not represent a judgment about allocative efficiency. Without putting a value on the outcomes, we cannot state that Country C would choose the outcome-input combination in Country A rather than the combination it currently has. But because the production function of Country A lies above the production function for Country C, we can argue that Country C has the potential to improve its productivity. By operating on the same production function as Country A, Country C could have had better outcomes with no more inputs. If there were increasing returns, points A and C could be on the same production function. Without further information, we could not infer that Country C would have potential for improvement unless it was willing to raise its input level.

The third type of comparison is illustrated by points C and D, where Country D appears to be on a higher production function. Without more information, however, that conclusion cannot be drawn with certainty. The lower average productivity in Country C may reflect either lower overall productive efficiency or the result of market demand that caused production to operate at a portion of the production function with small marginal returns to additional inputs, in order to achieve better outcomes. The fourth type of comparison is between Country B and Country C. As shown in figure 4, these are on the same production function and hence have the same productive efficiency. But again, because we do not know the shape of the production function, in practice we cannot tell this case from the third type.

Cost-Effectiveness Analysis

The inability to rank the third and fourth types of comparisons does not preclude drawing conclusions about their relative desirability. The literature contains rules of thumb for “reasonable” cost-effectiveness ratios, although the cutoffs are inherently arbitrary. Typical practice is to consider interventions that cost less than \$30,000 (1990 dollars) per

QALY to be cost effective, to consider those that cost more than \$100,000 per QALY as cost ineffective, and to treat intermediate costs per QALY as a “gray area.”⁷ Comparisons based on cost-effectiveness analysis are not as clear-cut for the productivity analysis. Fortunately it turns out that in all but one case, we can compare productivity without needing to resort to cost-effectiveness comparisons. But we do report cost effectiveness for the one case.

We turn now to a brief discussion of each of the four diseases, describing how they are treated and what we found in terms of inputs, outcomes, and productive efficiency. In these summaries we try to explain the productivity differences in terms of provider behavior. What are the doctors and hospitals doing that is different in the three countries? Later in the paper, we consider how the economic incentives and system constraints give rise to these behavioral differences.

Case Study Findings: Diabetes

Diabetes mellitus is a chronic condition that impairs or destroys the body’s ability to regulate glucose levels. It affects a significant fraction of the population—about 2 to 3 percent—in the United States and the United Kingdom and accounts for at least 4 to 6 percent of total health care costs in both countries.⁸ (Because information was not available on treatment in Germany, we excluded it from this disease comparison.) Diabetes mellitus is really two different conditions.⁹ Type I, “juvenile onset” diabetes, occurs early in life and destroys the body’s ability to produce insulin and therefore regulate glucose. Type II, “adult onset” diabetes, develops later in life and results in insulin secretion that is insufficient for the body’s needs, in part because sensitivity to insulin is diminished. Type II is the more common of the two, accounting for approximately 90 percent of diabetics in the two countries. Although they are different diseases and can be treated differently, many aspects of their treatment are similar and use the same providers.

7. Attempts to estimate “optimal” cost-effectiveness ratios based on a specific family of utility functions suggest that double annual income is a reasonable cutoff under plausible circumstances. See Garber and Phelps (1997).

8. British Diabetic Association (1996); National Diabetes Data Group (1995).

9. Our study excluded gestational diabetes, which is diabetes with onset (or first recognition) during pregnancy.

There is no cure for diabetes. Instead, early, basic treatment of diabetes is directed toward maintaining blood glucose levels in a near-normal range. For all Type I and many Type II diabetics, regular insulin injections are required. For some Type II diabetics, management consists primarily of controlling the patient's diet and exercise habits and, often, taking oral medications that control blood glucose levels. Once a patient is diagnosed with the disease, some form of diabetes management is needed for the rest of the patient's life.

Complications of diabetes are frequent, can significantly diminish quality of life, and can even be life threatening or fatal. Common complications include heart and kidney disease; visual impairment, which may lead to blindness; and foot ulceration, which may require amputation. Effective management of the diabetic's condition can significantly delay or prevent some of these complications.

Inputs

Patients themselves provide the most important labor input into the treatment of diabetes in the form of self-care. Self-care includes insulin injections, self-testing of blood and urine, and diet and exercise control. Although the patient's labor in performing these functions is an input into the production process, we did not attempt to measure it for this analysis.

As our omission of Germany from this category demonstrates, there is a paucity of accurate data on capital and supply usage in diabetes treatment. Almost all treatment directed toward continuing management of the disease is delivered in the outpatient setting, where data are typically unavailable. Because labor represents roughly 70 percent of the total cost of health care in both the United States and the United Kingdom, we believe that it is an acceptable simplification to restrict our analysis only to the labor inputs required in treatment. Although the cost of supplies for self-care, particularly insulin, can be significant, the largest cost component of diabetes is the inpatient care associated with treating complications. Labor is clearly the major input for inpatient care.

The diabetes treatment steps requiring the most provider labor are the routine visits to manage the disease and inpatient treatment of complications. Our analysis estimated the labor inputs into both of these

treatment steps. We did not include costs of initial diagnosis in our measurements because these diagnostic tests are usually inexpensive and do not appear to vary significantly between the two countries. Neither country had a formal screening program for diabetes. Nor did we include outpatient visits to specialists and tests beyond those handled in routine clinic visits (although such referrals are routinely performed to check for complications). We focused only on the inpatient treatment generated by those referrals. Follow-up visits to specialists after inpatient treatment were also excluded from the measurement.

Measurement of Outcomes

Although diabetes cannot be cured, treatment can prolong life and improve its quality. Because complications are chiefly responsible for both the morbidity and mortality of the disease, we focused our analysis of outcomes on the relative rates of developing selected complications in the two countries. All other factors being equal, a health care system delivers better outcomes in diabetes by preventing and successfully managing diabetic complications.

We estimated complication rates by using national databases, surveys, and the available medical literature.¹⁰ Specifically, we evaluated complication rates for diabetic ketoacidosis and hyperosmolar coma (a pair of similar complications that often occur in association with an acute illness such as kidney or lung infection), retinopathy (abnormalities of the retina that can lead to progressive visual loss), blindness, and lower extremity amputation. For each of these complications, we were able to obtain comparable estimates of the incidence rate in both countries.¹¹

To develop an overall measure of outcomes for diabetes treatment, we estimated the impact of each complication on a diabetic's quality of life. When quality of life is incorporated into cost-effectiveness analyses, a weighting or utility is assigned to each state, so that a value of one is equivalent to best imaginable health, and a value of zero is assigned to the worst imaginable state, which is usually assumed to be

10. McKinsey Global Institute and the McKinsey Health Care Practice (1996, appendix).

11. The two rates may not have been precisely comparable, because the definitions of complications may not have been identical. Furthermore, the average duration of diabetes may not have been the same.

equivalent to death. Kaplan and Bush devised a method to assign utility ratings to a large number of health states, based on interviews and surveys where population samples expressed their relative preferences for these health states. Their health state ratings were comprehensive enough to allow us to assign utility scores to each state of diabetic complications. Using these scores and the incidence rates for complications, we developed an “expected quality-of-life score” for an average diabetic in each country. (In essence, this expected value weights a complication’s effect on the quality of life by the probability of developing the complication.) This expected value, which we used as our basic outcome measure for diabetes, is expressed in QALYs.

To derive an expected QALY score for each country, we made several assumptions about a diabetic’s potential health states, the quality scores of these states, and the probabilities of being in these states over time. Although some of these specific assumptions could be challenged, and other models of expected QALY could be developed, the final result of our outcome comparison is unlikely to be sensitive to the particular utility assessment method used. Essentially, any reasonable set of assumptions and methodology yields an outcome measure that shows the United Kingdom having superior outcomes for diabetes treatment, because diabetics there are less likely than diabetics in the United States to develop each of the complications.

PRODUCTIVE EFFICIENCY DIFFERENCES. On a weighted average basis, the United Kingdom used 34 percent fewer inputs than the United States, and U.K. diabetics had 1.35 more QALYS.¹² Compared with the baseline case of no treatment, U.K. diabetics achieved 27 percent greater improvement in outcomes due to treatment than U.S. diabetics did.¹³

With better outcomes and fewer inputs, the United Kingdom was clearly more productive than the United States in diabetes treatment. Its productive efficiency advantage stemmed from its consistently lower complication rates. Although these rates were relatively low in both

12. Input usage was 40 percent less for Type I diabetics and 32 percent less for Type II diabetics. Type I diabetics in the United Kingdom had 2.5 more QALYs than diabetics in the United States; Type II had 1.2 more.

13. Baseline outcome with no treatment was conservatively assumed to be death within one year for Type I diabetics; Type II diabetics were assumed to have the same QALYs as the lowest outcomes with treatment (U.S.).

countries (roughly 1 to 3 percent for most complications), the total impact of these annual rates during a diabetic's lifetime created a significant difference in overall outcomes. The United Kingdom's advantage for Type I diabetes outcomes was greater, primarily because that type occurs at a younger age, so the United Kingdom's advantage in complication rates compounded over a larger number of years.

REASONS FOR THE DIFFERENCES. The lower rate of complications appeared to derive from two aspects of provider behavior in the United Kingdom: more intense treatment for those who could benefit most (stricter triaging), and a team-based approach. Although the effect each of these factors had on complication rate differences between the two countries cannot be quantified, according to interviews with clinicians in the two countries, both seemed to be important.

In contrast to the more uniform approach to the treatment of diabetes in the United States, the United Kingdom was highly selective in assigning aggressive treatments. For some diabetics, generally those considered to have the least severe conditions, the United Kingdom provided less treatment than the United States; more than 40 percent of noninsulin-using Type II diabetics in the United Kingdom received only home care, whereas 93 percent of these diabetics in the United States were treated by a physician. For the two-thirds of diabetics in the United Kingdom who received some form of physician-guided care, routine visits with providers occurred about five times a year, compared with an average of 3.5 visits a year in the United States. For the one-third of U.K. diabetics seen in a diabetic clinic, visits were also more comprehensive than comparable visits in the United States. The United Kingdom thus provided more intensive treatment to the diabetics with the most severe conditions. The United Kingdom's diabetic clinics not only offered more provider attention to certain diabetics, but they also offered care from many different types of providers in a multidisciplinary team that might have included a diabetologist, an ophthalmologist, a chiropodist, a dietician, and a nurse specialized in diabetes. This team likely was more effective than a single physician in assessing the diabetic's condition, developing a self-care program, and educating and counseling the diabetic.¹⁴

14. The United Kingdom's better outcomes for diabetes could also have been partially caused by behavioral differences between U.K. and U.S. patients; if the U.K.

Cholelithiasis (Gallstones)

Cholelithiasis, or the presence of stones in the gallbladder, is very common in Western nations. Approximately 11 percent of the population of the United States, the United Kingdom, and Germany, totaling more than 42 million people, have cholelithiasis. Nearly 2 million new cases are diagnosed in these countries each year.¹⁵ Although gallstones can cause abdominal pain and other symptoms, most of them are asymptomatic. Only 1 to 4 percent of patients with gallstones develop symptoms or complications each year; 10 percent of all patients with cholelithiasis develop symptoms five years after diagnosis, and 20 percent develop symptoms after twenty years. Even though serious complications of cholelithiasis are infrequent, a great deal of effort and resources are spent in treating this condition, amounting to about \$7 billion in 1992 in the three countries. Consequently, cholelithiasis is one of the costliest, as well as most common, digestive diseases.

Although gallstones can lead to life-threatening conditions such as acute cholecystitis,¹⁶ the most common symptom, abdominal pain, is usually mild, is often transient, and is not unique to cholelithiasis. The most common method for removing gallstones is cholecystectomy, or surgical removal of the gallbladder with its contents. Two approaches to cholecystectomy are now common: traditional, or open, cholecystectomy; and laparoscopic cholecystectomy. Surgical removal is usu-

diabetics were "better" patients (patients who take better care of their conditions), that could have led to the lower complication rates observed. Access issues could have contributed to worse overall outcomes in the United States; if a group of U.S. diabetics did not have access to care and, therefore, had poor outcomes, the population-based complication rates could have been driven up significantly. Because no national data are available to compare treatment compliance in the two populations or to evaluate the impact of uneven access to care in the United States, we were unable to determine the role these factors might have played in the relative complication rates. Additionally, higher levels of obesity in the United States could partially explain higher complication rates there for Type II diabetics; Type I diabetics, who are generally younger and not as subject to obesity, would be unaffected by this difference.

15. Graves (1995); National Health Service (1995a, b); Krämling and others (1993).

16. Additional potentially life-threatening conditions include empyema of the gallbladder, common bile duct stones with or without cholangitis or pancreatitis, gallstone ileus, or, rarely, gallbladder cancer. Life-threatening gallstone complications almost always merit acute care, but these are uncommon. In addition, the risk of gallbladder cancer in patients with gallstones is very low (currently estimated at 1 of 1,000 patients a year). This cancer risk, therefore, does not ordinarily justify prophylactic treatment.

ally recommended on the basis of the severity and frequency of symptoms, the presence of coexisting diseases, and the risk that the patient will suffer complications from the procedure. Cholecystectomies are relatively free of complications, however, and nonsurgical alternatives are less effective at preventing recurrence of symptoms. Thus, the safety and efficacy of surgery have made it the treatment of choice for symptomatic cholelithiasis.

Management and Treatment

We evaluated relative productive efficiency using two measures: outcomes per unit of inputs on a per-operation basis (that is, productivity when a cholecystectomy was performed) and on a per-case (per patient with cholelithiasis) basis. The per-operation results highlighted the differences in resource allocation per operation; the per-case results measured the overall input usage when treating the disease in each country. A country that was not particularly efficient in the performance of surgery could have high productive efficiency on a per-case basis by assigning patients to surgery in a highly selective manner.

We divided the management of cholelithiasis into three phases: diagnosis, treatment, and recovery. In the diagnosis phase, patients and physicians decide whether and how to treat. If surgery is selected, the patient receives pre- and postoperative tests, the operation itself, and any additional procedures required to treat complications. Finally, each patient enters a period of convalescence, primarily at home, before resuming work and other usual activities.

Inputs

We accounted for the actual use of labor, supplies, and capital in the treatment. Because recovery time is a significant component of the cost of treating cholelithiasis, we also included the opportunity cost of patient time, measured by weighting the number of work hours the patient spent in the hospital and during recovery by the average hourly wage in the country. We summed the per-operation use of labor, supplies, and capital separately for the open and laparoscopic operations in each country and used the relative number of each to obtain weighted inputs. Adding together these weighted inputs, we obtained the total input usage per operation in each of the three countries.

The input total per case is simply the input total per operation multiplied by the surgical frequency per case. Surgical frequency is the percentage of cholecystectomies per capita divided by the prevalence of cholelithiasis in the same country. Unless specified otherwise, we discuss results on a per-case basis.

To the extent possible, the analysis incorporates the inputs used in each step of surgical treatment, including the treatment of complications, common bile duct exploration, and stone removal. The analysis did not include nonsurgical treatments, which were rarely used, nor did it incorporate diagnostic tests and analgesia for patients who did not receive further treatment. These costs are likely to be very low, and accurate estimates are unavailable.

Outcomes

These outcomes reflect both the long-term effectiveness of the operation and operative morbidity and mortality. Significant complications from these operations were infrequent and similar in all three countries: 3.0 to 5.0 percent for open cholecystectomy; and 3.5 to 4.4 percent for the laparoscopic operation.¹⁷ Because both surgical options had similarly high success rates, it is reasonable to presume that each country produced similar outcomes per operation.

Outcomes *per case* were somewhat more complex. The relative success of cholelithiasis treatment depended crucially on the decision to proceed with surgery. Both the potential benefit from surgery and the success of each individual operation affected the per-case outcome. Ideally, both of these factors would be incorporated into the outcome measure to assess the overall quality of treatment. Although the degree of surgical success was approximately equal in the three countries, the potential benefit to the patient who underwent surgery depended on the severity of symptoms.

We incorporated symptom relief into our model of health outcomes by estimating the effects of surgery on each patient's QALYs. Pain was the major symptom, and each pain episode reduced the patient's quality of life. Thus, before and during surgery, a patient's quality index was less than 1, and after surgery, the patient was restored to a quality index

17. MacIntyre and Wilson (1993); Roslyn and others (1993); Dunn and others (1994); National Center for Health Statistics (1995).

of unity. Outcomes differed depending on the extent to which surgery alleviated severe symptomatic disease or treated disease that only marginally detracted from quality of life. Because the frequency of symptomatic pain episodes varies from patient to patient, we calculated outcomes using a range of frequency of symptoms, from recurrence every fourteen days to recurrence every sixty days; when choosing a single point estimate, we assumed that symptoms occur every thirty days.¹⁸

Productive Efficiency: U.K.-U.S. Comparison

On a *per-operation basis*, the United States and the United Kingdom had similar outcomes, but the United States used 71 percent fewer inputs. It used fewer resources per operation than the United Kingdom for both open and laparoscopic operations, and a higher proportion of the U.S. operations were laparoscopic.¹⁹

Although it used fewer inputs per operation, the United States performed more operations. Thus, on a *per-case basis*, it used 56 percent more inputs than the United Kingdom. This higher rate of surgery yielded outcomes that were 76 percent better than those in the United Kingdom on a per-case basis. Because its improvement in outcomes was greater than its increase in inputs, the United States had higher average productivity than the United Kingdom, an advantage that did not vary with the frequency of symptoms. For example, at fourteen days between symptoms, the United States was 72 percent higher than the United Kingdom; at sixty days, it was 76 percent higher.

Productive Efficiency: German-U.S. Comparison

On a *per-operation basis*, the United States used 52 percent fewer resources than Germany while obtaining similar outcomes. Even though Germany used 8 percent fewer inputs per open cholecystectomy, the resource use per operation was lower in the United States, both because a greater percentage of procedures were laparoscopic and because the United States used about 40 percent fewer resources per laparoscopic

18. Clinician interviews.

19. Inputs per laparoscopic cholecystectomy are indexed to the U.S. open surgery inputs to highlight the lower input usage of laparoscopic surgery.

cholecystectomy. Because outcomes per operation were equal, the United States was more productive in the laparoscopic technique and on an overall per-operation basis.

Because Americans with cholelithiasis were less likely to receive an operation and because the United States used fewer inputs per operation, the United States consumed lower inputs per case and had lower outcomes on a per-case basis relative to Germany. The United States had 52 percent higher average productivity than Germany over the entire range of symptoms occurring between fourteen and sixty days.

In this situation, in which the country with the lower input-lower outcome combination (that is, the United States) was the country with higher average productivity, we need detailed knowledge of the treatment process to determine which country was more productive. Shorter hospital stays, shorter recovery periods, and broader adoption of laparoscopy enabled the United States to use 72 percent fewer inputs per operation, with identical surgical outcomes. These advantages of the U.S. treatment process, coupled with the fact that higher German outcomes per case resulted solely from a higher surgical frequency, led us to conclude that the United States was more productive.

Reasons for the Differences

Three differences in provider behavior led to differences in productive efficiency between the United States and the United Kingdom: technology adoption, treatment duration, and staffing levels. The primary cause of the higher U.S. productive efficiency was faster adoption of the laparoscopic approach. Shorter hospital stays and postdischarge recuperation, which were at least partially related to the adoption of the laparoscopic operation, also increased U.S. productive efficiency relative to the United Kingdom and were only partially offset by higher levels of hospital staffing.

Differences in staffing levels and technology adoption also affected productive efficiency differences between the United States and Germany, but differences in treatment duration were especially striking—German patients experienced much longer hospitalizations and convalescence times after discharge. The slightly later adoption of laparoscopy further diminished productive efficiency in Germany. Despite the

added costs of higher levels of staffing in the United States, the net effect of these other differences led to lower productive efficiency in Germany.

Breast Cancer

Breast cancer is a leading cause of cancer mortality in all three countries, where between fifty-five and ninety cases per 100,000 women are diagnosed annually.²⁰ This incidence translates to a lifetime risk of disease on the order of 10 percent. Female breast cancer rarely occurs before the age of thirty and is most often diagnosed at fifty years of age and older. It is often fatal because it has a tendency to spread from the breast to distant tissues if left untreated. Currently, the only reliable cure for the disease is to remove it while it is still localized to the breast, an option that frequently is not possible at the time of diagnosis. There are no simple preventive steps that dramatically reduce individual risk.

Management and Treatment

We divided the management and treatment of breast cancer into four phases: screening; assessment; therapeutic; and follow-up. In the screening phase, patients with no prior indication of problems are examined for the presence of an abnormal tissue mass. The resources consumed by screening and diagnostic services were substantial in relation to those consumed by treatment alone. Furthermore, the patterns of screening and diagnosis vary among the countries studied. If screening indicates that disease may be present, a woman enters the assessment phase, where diagnostic testing and biopsies are performed to confirm or reject a malignant diagnosis. The therapeutic phase, in which patients are treated for the cancer, can include interventions designed to remove the primary tumor and to prevent or halt its spread. The follow-up phase includes all diagnostic testing to monitor the patient's progress after treatment, as well as therapeutic treatment upon any relapse.

20. SEER database.

Inputs

The input measure for breast cancer included all labor, capital, and supplies associated with the procedures performed in the four phases. We did not include elective reconstruction of the breast after a mastectomy. A preliminary analysis revealed that in the time period of our study, few women underwent breast reconstruction in any of the countries; the resources consumed by reconstruction were, therefore, likely to be small compared with the total cost of cancer care. In addition, the availability of reconstruction likely had little *differential* effect on the treatment approaches for cancer care in each country.

Outcomes

Our outcome measure is based on the percentage of women diagnosed with breast cancer who survive for five years following diagnosis. This measure is calculated from survival statistics for relatively large populations of breast cancer patients in each of the three countries during roughly the same time period. From these statistics, we constructed age-adjusted, five-year survival curves and compared the survival “profiles” of each country. Before any adjustments, these profiles show the highest survival rates in the United States, followed by the United Kingdom and then Germany.

Ideally, our outcome measure would reflect the increment to life years generated by breast cancer treatment in each of the countries, relative to the life years that would have occurred in the absence of treatment. This could not be done systematically, because we lacked information on the likely five-year survival rates of untreated individuals. Instead, we simply assumed that, left untreated, all patients would die right away. This procedure is very conservative in that it reduces the percentage output advantage of the best outcome country, namely, the United States. Effectively it means that the outcome measures are not much different in the three countries.

A problem with using five-year survival rates as the basis of the outcome measure is that screening introduces a bias. Suppose two women contract breast cancer in the same year, say 1980. The woman in Country A is diagnosed in 1980 by mammographic screening; the woman in Country B is diagnosed only in 1982, after she finds a lump. Both women then die six years after the onset of the disease, in 1986.

The woman in Country A who was mammographically screened is shown as having survived five years after diagnosis, but the woman in Country B who was diagnosed only with a lump is not. In fact the treatment did not affect the outcome. A bias is also introduced by women with slow-growing tumors who would have lived out their normal life spans even in the absence of treatment. Mammographic screening discovers and treats these cases and may count them as five-year successes of treatment when in fact the treatment did not affect their health outcome.

Because the United States had mammographic screening in the 1980s (the time frame of our study), while the United Kingdom had none and Germany had very little, we adjusted the observed five-year survival rates in the United States. Once again, our procedure was conservative, because it sharply reduced the substantial unadjusted survival advantage of the United States. It was assumed that one-third of all U.S. cases were detected by screening and that this introduced a lead-time bias of three years.²¹

Productive Efficiency Differences

Both the United States and the United Kingdom were clearly more productive than Germany in treating breast cancer. The United States used 38 percent fewer inputs and achieved 9 percent better outcomes than Germany, whereas the United Kingdom used 53 percent fewer inputs and achieved 6 percent better outcomes. The United States used 15 percent more inputs and achieved 3 percent better outcomes than the United Kingdom, which made it impossible for us to determine which nation had higher productive efficiency. This comparison is considered inconclusive in terms of productive efficiency.

In U.S. prices, the United States spent \$32,000 more per life year than the United Kingdom for treating breast cancer; that amount is generally considered cost effective. In U.K. prices, the United States spent only an additional \$13,000 per life year, which suggests that it

21. By using five-year survival as the outcome measure, we do not capture differences in the quality of life; data limitations prevented us from doing so. In recent clinical trials, researchers have been using disease-free survival rates, acknowledging that survival without the recurrence of cancer is potentially more useful as an outcome measure than raw survival. Unfortunately, disease-free survival rates were not widely recorded during the time period of our study.

would clearly make economic sense for the United Kingdom to increase the resources used to treat this disease. Keep in mind also that the procedures for outcome measurement tended to reduce the measured U.S. outcome advantage.

Reasons for the Differences

Once the adjustments were made to the measure, we found relatively small differences in outcomes among the three countries. Consequently in our analysis of the reasons for the productivity differences, we concentrate on explaining the differences in inputs.

Screening practices had a significant effect on differences in overall input consumption and productive efficiency. At the time of our analysis, the United Kingdom had no formal screening program, and therefore no resources were considered to be consumed in this phase.²² In comparison, the widespread adoption of screening in the United States came at a high cost. Screening through mammography and physical exam accounted for about 15 percent of the total resources consumed in breast cancer care, with mammography accounting for most of these resources. Much of this activity focused on premenopausal women who, in the absence of risk factors, were less likely to benefit from it than postmenopausal women. Physical breast exams were part of a typical gynecological exam, which means that women as young as eighteen years underwent this type of screening. Like the United States, Germany employed both mammographic and physical exam screening. Overall, Germany consumed slightly more resources than the United States on screening but, on balance, consumed more on physical exams than on mammography.

The broader the screened population (that is, the younger the age at which screening began), the more frequently screening resulted in false positive cases, leading to large additional “downstream” costs in the assessment phase. That is because younger women are much more likely than postmenopausal women to have noncancerous abnormalities that are then detected and assessed. This downstream cost was greatest

22. Forrest and others (1986); clinician interviews. A small number of women are likely to have received physical exam screens in the United Kingdom, but there is no good estimate for this level of care. Therefore, we did not include any screening in our analysis of the total resource consumption for breast cancer care in the United Kingdom.

in the United States because of its wide use of mammography on younger women, which when compared with the mostly physical exam screening in Germany, identified more nonpalpable masses, most of which were benign. By increasing costs without producing substantial benefit, this broad-based mammographic screening lowered the United States's productive efficiency in breast cancer treatment.²³

The protocols in the *assessment* phase also differed in the three countries. There can be either one-step or two-step procedures. In the one-step procedure the patient undergoes a surgical biopsy in a hospital while under general anesthetic, and if the mass is cancerous, it is removed together with a mastectomy or other treatment. In a two-step procedure, a biopsy is performed under local anesthetic, followed by surgery at a later time if there is a finding of cancer. A biopsy can be performed using either a surgical biopsy or a fine needle aspiration (FNA), and it can be performed either on an inpatient or an outpatient basis.

Over time, there has been a shift away from one-step procedures. During the time frame of the study, the United States had virtually completed the shift to the two-step procedure, but in Germany and the United Kingdom 80 percent of the procedures were still one step, and these were all carried out in hospitals. In the United States the first step was generally a surgical biopsy carried out in the doctor's office. The two-step procedures used in the United Kingdom involved an FNA in the first step, on an outpatient basis. Biopsies in Germany were carried out in the hospital even in the two-step cases.

Because outpatient biopsies consume fewer resources and because most patients who go on to the assessment phase turn out not to have malignancies, the United States was able to save substantial resources by performing all of the biopsies on an outpatient basis and using a two-step procedure. This advantage more than offset the fact that the United States performed more biopsies in total, mostly because of its screening program. Overall, the United States used 3 percent fewer

23. In 1987 the United Kingdom instituted a nationwide breast cancer screening program that did not become fully functional until 1991. Using mammography, the program is restricted to women over the age of fifty and currently calls for screening every three years.

resources than the United Kingdom and 20 percent fewer resources than Germany on assessment.²⁴

In the *therapeutic* phase, surgery is the most important treatment for breast cancer. We observed differences in the frequency of surgery, as well as in the mix of the two major types of surgeries performed. Overall, the frequency of surgery for the primary breast tumor was 91 percent, 75 percent, and 97 percent in the United States, United Kingdom, and Germany, respectively. Of those cases treated surgically, the frequency of breast-conserving surgery was 29 percent, 44 percent, and 39 percent for the United States, United Kingdom, and Germany, respectively.²⁵

Despite the differences in their frequency of surgery, the United States and United Kingdom consumed about the same level of resources for surgery and subsequent hospitalization; Germany consumed about 50 percent more. This is because total resource consumption depends not only on the frequency of surgery, but also on the lengths of hospital stay, and these were shortest in the United States.

Radiotherapy and chemotherapy were often part of breast cancer therapy. Chemotherapy was not a major part of total resource use in any of the countries (3 to 4 percent). Germany used 25 percent more resources than the United States in chemotherapy because of greater use of inpatient treatment. Radiotherapy was a somewhat more important component of total resource use (6 percent in the United States, 12 percent in the United Kingdom, and 5 percent in Germany). And both the United Kingdom and Germany used greater total resources than the United States (60 percent and 10 percent more, respectively). There were a variety of offsetting reasons for these findings.²⁶

- The United Kingdom gave radiation in fewer but larger doses

24. Because few data were available on assessment protocols, the analysis here is derived through interviews with clinicians in each of the three countries.

25. SEER database; Thames Cancer Registry database; Krebsregister–Saarland database; Arbeitsgruppe zur Koordination Klinischer Krebsregister database; Grosshadern dataset; Foreman and Rider (1995); clinician interviews. The frequency rate does not include surgeries performed upon recurrence of cancer.

26. Thames Cancer Registry database; Arbeitsgruppe zur Koordination Klinischer Krebsregister database; clinician interviews.

(saving resources), but used older equipment (requiring extra labor).

- The United States had higher staffing levels in hospitals, which raised resource use for inpatient radiotherapy. The United States performed less radiotherapy, largely because of its less frequent use of breast-conserving surgery.
- The United Kingdom did more radiotherapy, mostly because it did less surgery than either the United States or Germany.
- The United States and Germany each consumed about 30 percent of total resources in the therapeutic phase; the United Kingdom about 39 percent. The United Kingdom and Germany consumed 3 percent and 11 percent more resources, respectively, than the United States.²⁷

Lung Cancer

Lung cancer is the leading cause of cancer death in all three countries.²⁸ In 1995 lung cancer caused about 160,000 deaths in the United States alone.²⁹ The disease is associated with cigarette smoking and develops most often in scarred or chronically diseased lungs. Its poor prognosis reflects its advanced state at the time it is usually detected. Symptoms of lung cancer include persistent cough, breathing difficulty, abnormal sputum, chest pain, and repeated attacks of bronchitis or pneumonia. Lung cancers spread widely to other organs; the extent of spread is a critical element in determining overall prognosis and type of treatment offered.

Lung cancers are typically grouped into two categories according to cell type. Small cell lung cancer accounts for 20 to 25 percent of the

27. Although there are many options relating to the procedures available for monitoring patients for relapse and for treating upon relapse, the follow-up phase itself does not consume many input resources relative to the other phases. Because the overall cost is small, any practice differences among the three health care systems resulted in relatively insignificant resource consumption and productive efficiency differences. Thus, the treatment differences observed were less important in explaining input and productive efficiency variations than differences in the other phases.

28. World Health Organization (annual).

29. Wingo and others (1995).

cases and has a particularly poor prognosis.³⁰ Non-small cell lung cancer accounts for the balance of the cases and can be cured if detected early. Although the approaches to treatment vary between the two groupings of cancers, in general, both are managed through one or more interventions—surgery, radiotherapy, chemotherapy, and supportive care.

Because lung cancer is often incurable, therapy often is directed toward more limited goals than curing the disease. Therapy can be divided into three classes: curative; palliative (amelioration of symptoms only); or supportive (maintaining patient comfort without active therapy).

The intent of treatment and specific treatment options are decided after discussion between physician and patient. The extent of the cancer, its cell type, and the patient's physical and emotional condition determine which treatment is appropriate.

Management and Treatment

We divided the management and treatment of lung cancer into three distinct phases: diagnosis and staging; curative care; and palliative care. The purpose of the diagnosis and staging phase is to identify the condition as lung cancer, assess the cell type of the disease, and determine the size of the primary tumor and the extent of spread to distant parts of the body. The information gained in this phase is used to assess the appropriate course of treatment—whether curative care or palliative care. These two treatment options represent the second and third phases in the management of lung cancer. Curative care, warranted in only a minority of cases, is aggressive and attempts to eradicate the cancer and return the patient to full health. Palliative care offers an alternative when a patient has little chance of cure or when curative care has failed to eradicate the disease. Palliative care takes two different forms: anticancer palliative care (which includes any noncurative intent surgery, chemotherapy, or radiotherapy directed at a tumor site) and supportive care (which includes any other palliative care).

30. Metastases are tumors that form in parts of the body remote from the primary tumor and are the product of cancer spreading from the primary site.

Inputs

The input measure covers all the labor, capital, and supplies associated with the procedures performed in the three phases of management. We excluded “best supportive care” in the palliative care phase because no reliable data could be found to cover it. We believe that the resource consumption involved was small and that differences among countries were likely insignificant.

Outcomes

The median survival of lung cancer patients is about a year, and only about 10 percent of cases survive five years after diagnosis. A five-year survivor has a high likelihood of being cured of the disease, and so, for the basis of our comparison, we chose an outcome measure of life years saved based on the cumulative five-year survival curve.

Most outcome measures for lung cancer, like those for breast cancer, are problematic. Analysis based solely on survival duration does not adequately take into account the quality of life. Undoubtedly, an outcome measure adjusted for quality of life would handle this potential problem, but we were unable to use such a metric because the required data were unavailable. We believe, however, it is reasonable to assume that no significant differences in treatment preferences existed among countries and that therefore our use of five-year survival provides a reasonable basis for outcome comparison.³¹

Productive Efficiency Differences

Germany used 21 percent more inputs and achieved 12 percent worse outcomes than the United States in the treatment of lung cancer. With better outcomes and fewer inputs, the United States was clearly more productive than Germany in lung cancer treatment.

31. The availability of treatment options for terminal patients may affect the shape of the five-year survival curve but should not affect the percentage of cases that actually survive. This difference in curve shape occurs because in a resource-constrained system, terminal patients are less likely to gain access to treatments such as chemotherapy and thus may die sooner, which changes the shape of the survival curve. These conditions may have been present in the United Kingdom, so a small portion of the outcome difference between the United Kingdom and the other two countries may be due to the availability of such life-extending, but not life-saving treatments.

The United Kingdom used 24 percent fewer inputs and achieved 58 percent lower outcomes than the United States in the treatment of lung cancer. In this case, measuring average productivity requires comparing each nation's outcomes *with* treatment to outcomes *without* treatment. Average productivity was thus calculated using five-year survival curves to determine each nation's outcome with treatment and a baseline estimate of 3.8 months survival without treatment. That measure showed that average productivity was 82 percent higher in the United States than in the United Kingdom.³² Because the disease treatment process did not appear to exhibit increasing returns at the positions of the United Kingdom and the United States, we conclude that the United States was more productive than the United Kingdom in lung cancer treatment.

The United Kingdom used 37 percent fewer inputs and achieved 52 percent lower outcomes than Germany.³³ Based on five-year survival, Germany had 33 percent higher average productivity than the United Kingdom, and so we concluded that Germany was more productive than the United Kingdom in lung cancer treatment.

Reasons for the Differences

The United States demonstrated greater productive efficiency than the United Kingdom for two main reasons: shorter hospital stays for surgery; and substitution of outpatient for inpatient chemotherapy. The United States also used CT (computerized tomography) scans more frequently in diagnosis and staging than the United Kingdom did; these scans made it possible to target treatment toward the patients who would benefit most and ultimately improved outcomes. Although higher staffing levels diminished the United States's productive efficiency relative to that of the United Kingdom, the net result of differences in treatment was higher productive efficiency in the United States.

32. SEER database; Joslin and Rider (1993). In our data search, we found no examples of a clinical trial that compared outcomes for treated versus untreated cases. We did, however, find survival curves for untreated cases (that is, patients who received only basic support care; these results are the basis for our estimate of 3.8 months. The cases that underlie these untreated curves obviously do not reflect an adequate cross section of all lung cancer cases. Thus, survival curves and our estimate likely understate the true average survival for untreated cases. We believe that this understatement is small and contributes insignificantly to our outcome calculation.

33. Calculation based on ratio of German results to those in the United Kingdom.

Germany's productive efficiency relative to the United States was lowered by its longer hospital stays and its greater use of the inpatient setting for chemotherapy. Although its lower staffing levels raised Germany's productive efficiency relative to the United States, the net effect of provider treatment differences led to higher productive efficiency in the United States.

Differences in the frequency and type of diagnostic testing had a significant effect on differences in overall input consumption and productive efficiency.³⁴ In general the United Kingdom performed fewer diagnostic tests per lung cancer patient than did the United States or Germany. The most important differences in behavior were in the areas of CT scans, endoscopic exams, and biopsy, where the United Kingdom appears to underinvest relative to the United States and Germany. Only about 20 percent of cases in the United Kingdom were assessed with a CT scan, compared with 80 percent in the United States and close to 100 percent in Germany.

In the United States the diagnosis and staging phase accounted for about 21 percent of all resources devoted to lung cancer, compared with about 18 percent in the other two countries. The United Kingdom consumed 8 percent fewer resources than the United States; Germany consumed 1 percent more. The resources consumed during the curative care management phase, on average, accounted for about 40 percent of total resources devoted to lung cancer care. Surgery was responsible for more than half of these resources. Radiotherapy played a lesser though important role, accounting for about 20 percent of these resources. Chemotherapy, which was used infrequently, rounded out the care, consuming about 10 percent of the resources devoted to this phase.

The total resources committed to surgery differed significantly across the three countries. Resource consumption was driven by the frequency of surgery, the length of hospital stay during recovery, and the level of hospital staffing. The surgical frequency was highest in Germany, with about 30 percent of all lung cancer patients receiving surgical treatment. The United States and the United Kingdom followed with 22 percent and 13 percent, respectively. Shorter lengths of stay in the United States were partially offset by higher hospital staffing levels. Accounting both

34. Edinburgh Lung Cancer Group (1987); Humphrey and others (1990); Scotland data, unpublished; clinician interviews. The frequencies of CT scanning, bronchoscopy, and mediastinoscopy are reported in these sources.

for differences in frequency of surgery and in resources per operation, we concluded that the United Kingdom consumed 25 percent fewer resources than the United States, and Germany 60 percent more.

The drivers of differences in resource consumption associated with radiotherapy were similar to those driving differences in surgery. Most radiotherapy was performed in the outpatient setting, however, which meant that the length of hospital stay and staffing factors were of less importance than for surgery.

Chemotherapy was a relatively minor part of the curative care management phase in that its frequency was quite low in all three countries. The setting of chemotherapy care differed considerably, however, with the United Kingdom and Germany using an inpatient setting far more than did the United States. As a result, overall resource consumption for the United Kingdom and Germany was about 120 percent and 90 percent greater, respectively, than that in the United States. The palliative care management phase, on average, accounted for about 40 percent of total resources devoted to lung cancer care. In general, patterns of palliative care paralleled patterns of curative intent care in each country.

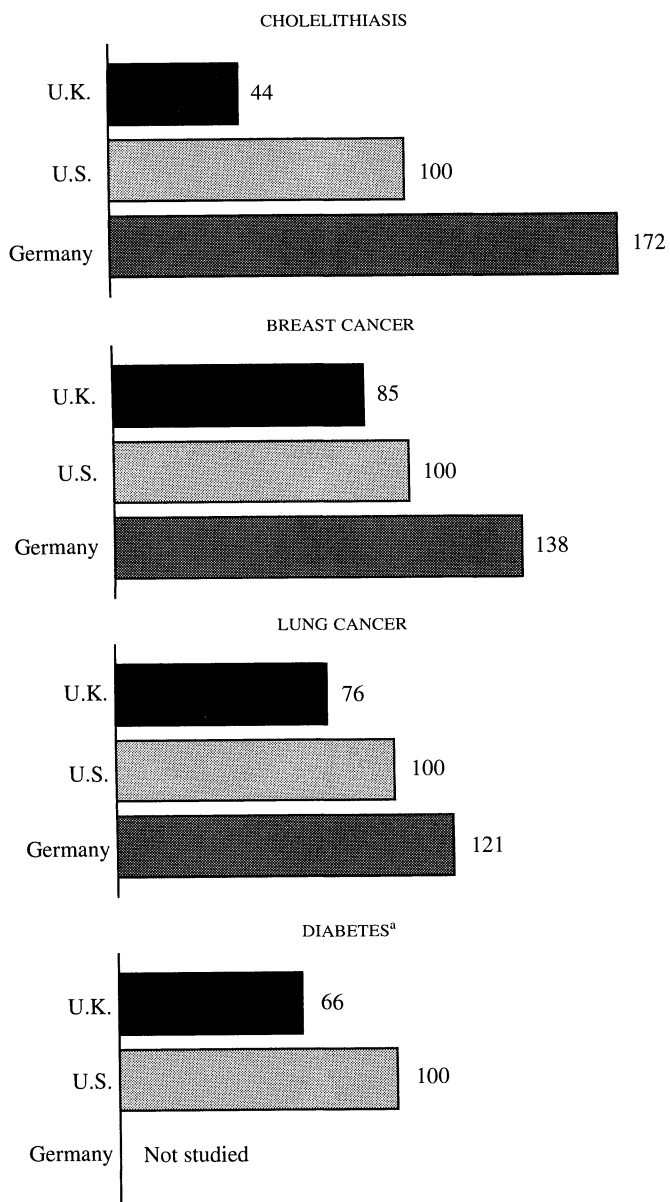
Summary of the Case Findings

Figures 5 and 6 summarize the outcomes and inputs for the four diseases, which are indexed so that the U.S. outcomes and inputs equal 100. In all of the diseases, the United Kingdom used the smallest amount of inputs, the United States used more, and Germany used the most. The United States had the most favorable outcomes for breast and lung cancer; the United Kingdom for diabetes, and Germany for cholelithiasis. Table 1 summarizes the resulting productive efficiency findings. The United States appears to have the highest productivity for lung cancer and cholelithiasis; the United Kingdom has the highest productivity for diabetes. For breast cancer the outcome is indeterminate between the United States and the United Kingdom, which has lower outcomes but also lower inputs. It does appear, however, that the United Kingdom devoted too few resources to this disease.

Table 2 summarizes the differences in provider behavior that account for the input and outcomes differences. We have divided these behav-

Figure 5. Inputs by Disease

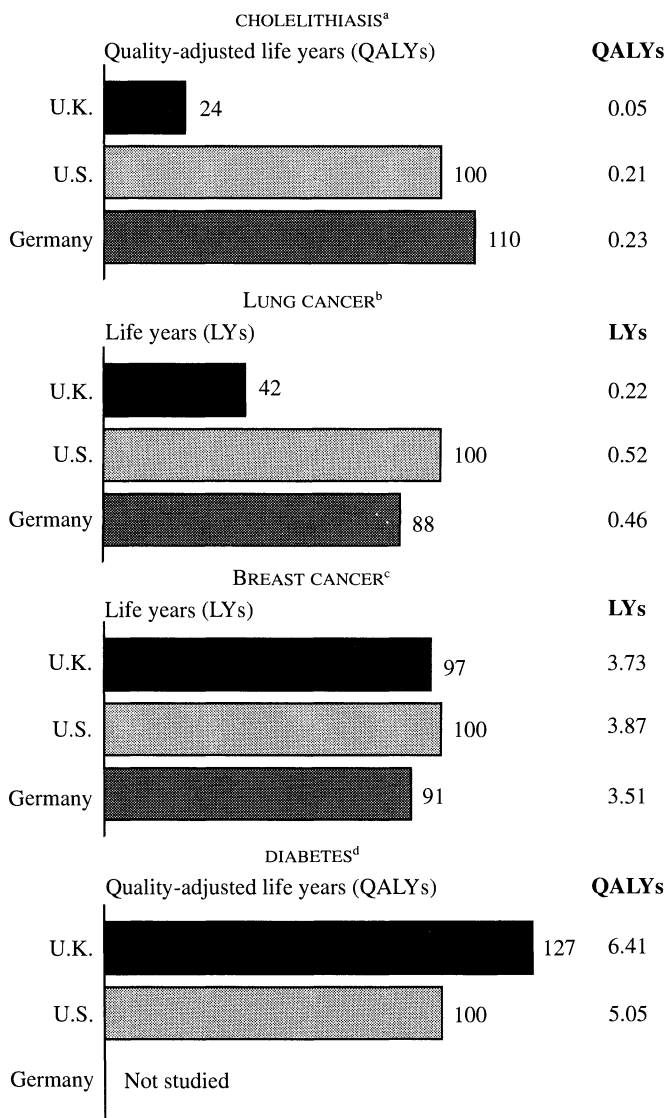
U.S. = 100



Source: McKinsey Global Institute and the McKinsey Health Care Practice (1996).

a. Weighted average of Type I and Type II diabetes.

Figure 6. Outcomes by Disease: Improvement in Outcomes Because of Treatment
U.S. = 100



Source: McKinsey Global Institute and the McKinsey Health Care Practice (1996).

Note: Improvement is relative to baseline outcome with no treatment.

a. Outcomes based on thirty-day frequency of symptoms.

b. Outcomes based on five-year survival.

c. Outcomes based on five-year survival; baseline outcome with no treatment assumed to be immediate death.

d. Weighted average of Type I and Type II; baseline outcome with no treatment assumed Type I diabetics die within a year and Type II diabetics have same QALYs as lowest outcomes with treatment (U.S.).

Table 1. Relative Productive Efficiency by Disease Case

<i>Disease</i>	<i>U.S. vs. U.K.</i>			<i>U.S. vs. Germany</i>			<i>Germany vs. U.K.</i>		
	<i>Inputs</i>	<i>Outcomes</i>	<i>More productive country</i>	<i>Inputs</i>	<i>Outcomes</i>	<i>More productive country</i>	<i>Inputs</i>	<i>Outcomes</i>	<i>More productive country</i>
Cholelithiasis	↗	↗	U.S.	↘	↘	U.S.	↗	↗	Germany
Lung cancer	↗	↗	U.S.	↘	↗	U.S.	↗	↗	Germany
Breast cancer	↗	↗	Indeterminate	↘	↗	U.S.	↗	↘	U.K.
Diabetes	↗	↘	U.K.	Not studied			Not studied		

Source: McKinsey Global Institute and the McKinsey Health Care Practice (1996).

Table 2. Specific Differences in Provider Behavior Identified in Case Studies

<i>Provider behavior difference</i>	<i>U.S. vs. U.K.</i>			
	<i>Diabetes</i>	<i>Cholelithiasis</i>	<i>Breast cancer</i>	<i>Lung cancer</i>
Care triaging	Clear triaging of diabetics for care management resources in U.K.	Much less frequent surgery in U.K.	No screening in U.K.; more lumpectomy and radiotherapy in U.K.; more chemotherapy in U.S.	More use of "supportive care only" in U.K.; less radiotherapy in U.K.; more diagnostic testing in U.S. and less surgery in U.K. given complex and lengthy referral process
Length of hospital stay	Longer stay for complications treatment in U.K.	Longer stay, longer patient recovery time in U.K.	Longer surgical stay in U.K.	Longer surgical stay in U.K.; shorter radiotherapy stay in U.K.
Staffing levels	Lower hospital staffing in U.K.	Lower hospital staffing in U.K.	Lower hospital staffing in U.K.	Lower hospital staffing in U.K.
Setting choice	More inpatient biopsy in U.K.; maintenance of one-step biopsy-surgical treatment protocol	More inpatient chemotherapy in U.K.; more outpatient chemotherapy in U.S.
Team-based approach	Multispecialty diabetic clinics in U.K.
Technology adoption	. . .	Later laparoscopic adoption in U.K.	Use of fine needle aspiration in U.K. vs. surgical biopsy; no mammographic screening in U.K.	Less use of computerized tomography scan for staging in U.K.

(continued)

Table 2. (continued)

<i>Provider behavior difference</i>	<i>U.S. vs. Germany</i>		
	<i>Cholelithiasis</i>	<i>Breast cancer</i>	<i>Lung cancer</i>
Care triaging	Slightly more frequent surgery in Germany	Broader screening and more lumpectomy in Germany	More frequent surgery in Germany
Length of hospital stay	Longer stay, longer patient recovery time in Germany	Longer surgical stay in Germany	Longer surgical stay in Germany
Staffing levels	Lower hospital staffing in Germany	Lower hospital staffing in Germany	Lower hospital staffing in Germany
Setting choice	. . .	Inpatient biopsy in Germany vs. outpatient in U.S.; more inpatient chemotherapy in Germany	More inpatient chemotherapy in Germany
Team-based approach
Technology adoption	Slightly later laparoscopic adoption in Germany	More mammographic screening in U.S.	. . .

ioral differences into six categories. *Care triaging* reflects the approaches to screening and to the allocation of resources once diagnosis has been made. The United Kingdom is notable for doing less screening and less surgery for the cancers and less surgery for cholelithiasis. Patients with cholelithiasis and lung cancer in the United Kingdom are more likely to be assigned to palliative care than in the United States. The situation is different for diabetes, where triaging in the United Kingdom contributed to both lower inputs and higher outcomes. The United Kingdom and particularly Germany had *longer hospital stays* than did the United States. The United States had higher *staffing levels* per hospital day than the other two countries; these staffing levels, however, only partially offset the input advantage to the United States of the short lengths of stay.

Choice of setting was also a factor, with treatment taking place outside hospitals in the United States more often than in the United Kingdom and much more often than in Germany. In our cases, the *team approach* was used only by the United Kingdom and only for diabetes. Nonetheless, we judge this approach to be an important example of treating a chronic disease, using a specialized clinic and fewer resources to achieve better outcomes. Finally, *technology adoption* was important. For lung cancer (CT scans) and for cholelithiasis (laparoscopic surgery), the more rapid adoption of technology improved U.S. performance. Technology adoption was notably slower in the United Kingdom and somewhat slower in Germany. Conversely, the rapid spread of mammographic screening technology, applied to patients where the effect on outcomes is minimal, hurt U.S. performance.

Differences in Provider Incentives, Constraints, and Regulations

These differences in treatment protocols gave rise to different productive efficiencies. But what causes differences in the approaches to treatment? Differences in medical knowledge are unlikely to be responsible. The medical literature is available to all, and methods of training doctors are fairly similar. The differences in the incentives that providers faced in the three countries were striking, however, and those differences could have led to very different rates of hospital utilization

and differences in other aspects of medical practice. We concluded that the variations in incentives—as well as differences in regulatory and other constraints—explained much of the difference in productive efficiency that we observed.

Major Differences between the United States and the United Kingdom

Most physician services in the United States, including both specialist and primary care, were negotiated and compensated on a fee-for-service basis by payors. Physicians aggressively competed for patients, and to a lesser extent, for payor contracts. The U.S. physicians also faced the threat of malpractice suits.

Although U.S. payors faced some price-based competition and could have bundled and negotiated services in a variety of ways, they were not an effective force to counterbalance incentives to increase physician activity and technology during the time period we were studying. With indemnity coverage, locally prevailing practices determined physician payments. The U.S. physicians as a group were able to use local medical associations and specialty societies to promote changes in these practices, leading to increases in standards of care and thus to health insurance coverage for higher activity levels or new technology adoption. U.S. payors were often forced to adopt such coverage in their health insurance products to meet employer and consumer demands for new treatment approaches and thereby to compete effectively; the tax-preferred treatment of health insurance premiums magnified this pressure.

In contrast, *specialist* physician services in the United Kingdom were negotiated in the form of an annual salary for a range of services performed by the National Health Service (NHS) through its regional health authorities; however, specialists could also earn additional income by treating private patients. Thus the U.K. specialists had no economic incentives to increase the amount of care provided to NHS patients. Indeed, to the extent that they had alternative income sources (from private practice, for example), they may even have had an incentive to limit the time devoted to NHS patients. In addition, the U.K. specialists had few incentives to adopt new technologies, unless the technologies freed up constrained care resources.

General practitioner services in the United Kingdom took the form primarily of fee-for-service contracts, with rates negotiated on the basis of a complex formula through the NHS. In principle, therefore, the general practitioners had an incentive to increase treatment intensity. In practice, however, this was not the case because the NHS controlled both the supply of doctors and the fee payments. Given the tight physician supply and the structure of the NHS contracts, neither the UK specialists nor general practitioners competed in any meaningful way for NHS patients.

As the organizing force for health care in the country and as the employer of many of the physicians, the NHS was able to influence physicians through training, dissemination of information and guidelines, and, if necessary, through direct authority. The NHS, therefore, contributed to physicians' greater concern for cost effectiveness in the United Kingdom and thereby to their greater willingness to adopt technology more slowly and selectively. By internalizing the interaction between payors and physicians, the United Kingdom may have been better able to apply these controls than the U.S. payors were able to do through arm's-length, competitive interactions with physicians. The U.S. payors lacked market power relative to physicians, primarily because the payors' customers—employers—did not aggressively resist cost increases until the early 1990s.

A consistent finding of international productivity comparisons is that competition promotes high productivity. Competitive intensity in both care provision and health coverage was much greater in the United States, which helps explain its relatively high productivity for lung and breast cancer and cholelithiasis compared with that in the United Kingdom. The health care market is subject to several types of market failure, however, that can distort the effect of competition. Third party payment encourages excessive treatment (the moral hazard problem), a problem that can be exacerbated by the tax treatment of health insurance premiums. We did find that the United States provided more treatment than the United Kingdom, consistent with this activity-increasing incentive.

The methodology used in these international comparisons is not designed to determine the optimal level of expenditures. We restrict our analysis to the empirical comparisons of resources used and the associated outcomes. On this basis, however, we conclude that the higher

levels of treatment in the United States generally led to better outcomes relative to the United Kingdom. In particular, the improvement in outcomes for lung cancer, breast cancer treatment (as opposed to screening), and cholelithiasis in the United States were large when compared with the increments in inputs. Moreover, one alleged symptom of excessive treatment is the use of new technology that is not cost effective. We found, however, that in the case of cholecystectomy, the rapid adoption of new technology improved productivity relative to the United Kingdom.

An important qualification to this conclusion occurred when third-party payment was combined with incomplete information. For example, providers offered mammographic screening to premenopausal women because patients demanded it and insurers paid. We found little or no improvement in outcomes in the United States associated with such screening despite the substantial costs involved.³⁵ Many of the costs resulted from false positives from the screening.

Adverse selection is a form of market failure that could not occur in a universal health care system but that has major consequences in the United States. U.S. payors have an incentive to avoid patients with chronic or expensive diseases, whose expected health care costs are higher than their premiums. Specifically, payors have an incentive to avoid diabetic patients because they generate above-average claims. Adverse selection has thus made it less attractive for U.S. providers to establish diabetes clinics of the type developed in the United Kingdom. Diabetes clinics have been established in the United States, but they have had trouble obtaining patients with insurance coverage. The repayment schedules established by private payors for doctor visits for diabetic patients exacerbated the difficulties of treating this disease in the United States. And the schedules for the private sector were reinforced by similar schedules for Medicare and Medicaid, which did not cover the cost of a diabetologist. Historically, many diabetic patients in the United States have been treated by general practitioners who did not carry out routine foot exams—an essential step in avoiding foot ulcers and possible amputation.

The success of the United Kingdom's approach to diabetes did not

35. Screening of postmenopausal women is likely to be more cost effective. The United Kingdom has now instituted such a mammographic screening program.

come from offering more treatment to all, but in large part from training patients in methods of self-care. Such an approach made sense for the NHS, which had lifetime responsibility for patients. The U.S. payors generally provided health coverage for one-year terms and faced relatively high annual turnover in their member populations (20 to 40 percent). This reduced their incentive to make investments in preventive or education-oriented care that had a longer-term payback.

Differences in physician and hospital supply were also important. The United Kingdom exercised strict controls over the number of physicians and the number and capacity of hospitals through the NHS budgeting process and regulations. In the United States the supply of physicians and hospitals was relatively unconstrained, although licensing requirements serve as an entry barrier. Supply constraints contributed to the differences in the amount and intensity of care provided in the two countries. Physician and hospital capacity constraints in the United Kingdom forced providers to be more selective in choosing patients to treat and to substitute procedures that conserved scarce resources. Thus, for example, they adopted fine needle aspiration for breast cancer biopsy, which requires neither a hospital admission nor the services of a surgeon. The selection process could either be explicit (for example, through providers' decisions to limit care or resources) or implicit (through patient queuing, for example).

The NHS budgets also explicitly limited funding for capital investments. Most funds were controlled at the regional or district level rather than incorporated into local hospital annual budgets. In these allocation decisions, the NHS considered the cost-effectiveness of a new technology in treating a specific disease, as well as the effect of a given technology on the overall system; for example, the regional and district health authorities could consider the effects of increased availability of CT scans for lung cancer diagnosis and staging on systemwide usage and costs. These funding limits and allocation processes contributed to the slower adoption and narrower use of capital-intensive technology, such as mammographic equipment, CT scans, and laparoscopic equipment, in the United Kingdom relative to the United States. In addition, funding limits may have precluded substitution of more capital-intensive resources, such as CT scans, for other care resources.

In the United States individual hospitals and physicians decided how much to invest in new capital; they could thus respond to—or drive—

demand for new technology on the part of both patients and payors, with reasonable confidence that payors would reimburse patients treated with these technologies.

Major Differences between the United States and Germany

German hospitals had strong incentives to increase their lengths of stay; the U.S. hospitals had incentives to reduce them. These incentive differences led to significantly lower productive efficiency in Germany, observed in all three case study comparisons.

German hospital services, including physician services, were, by law, negotiated and compensated on a per diem basis with the payors. In contrast, the U.S. hospital services were negotiated and compensated on a case rate basis from Medicare (through the diagnosis-related group, or DRG, system) and through a mixture of approaches from private insurers that included fee-for-service, per diem charges, and case rates. Case rates accounted for 35 to 40 percent of an average U.S. hospital's total revenues.

The incentives U.S. hospitals faced depended on whether they were being reimbursed on a case-rate or per diem basis; they could, in principle, have used different lengths of stay for the two classes of patient. In practice it is difficult for doctors to apply different protocols to patients in adjacent hospital beds. Thus the incentives created by case-rate reimbursement influenced the treatment protocol for everyone.

The use of per diem rates gave German hospitals an immediate incentive to extend lengths of stay. This incentive was reinforced by the fact that German hospitals faced the threat of regulatory review and capacity cuts if their utilization fell below 85 percent. By maintaining high occupancy, hospitals avoided this threat.

Specialist physicians in Germany were employed by their hospitals and paid a flat salary; thus these physicians appeared not to have a direct economic incentive to increase the amount of care provided. They had clear "noneconomic" incentives to further the interests of their employers, the hospitals, however, and therefore had a relatively strong incentive to increase the amount of inpatient care they provided. Besides incentives to maintain high enrollment, there was an incentive to select inpatient rather than outpatient care and to prolong hospital stays. The chief physicians of German hospital departments were rewarded

for increasing the workload of their hospitals; each department was allowed bed capacity for private patients in a relatively fixed ratio to its utilized public beds, so that the workload of the hospital from publicly funded patients had an indirect but significant effect on the chief's private income.

CONSTRAINTS ON HOSPITAL CAPACITY IN GERMANY. Hospital capacity in Germany was seemingly constrained, whereas the U.S. capacity was relatively unconstrained; yet Germany had more hospital beds per capita than the United States. The German constraint, therefore, had the perverse effect of increasing supply and (in combination with the above incentives) encouraging longer and more frequent use of inpatient treatments.

Capacity was regulated by state governments, which had an incentive to maintain or increase the number of hospital beds because they created jobs and resulted in transfers from federal payor funds into state economies. In addition, regulations required that payors partially fund losses at hospitals; thus, unlike hospitals in the United States, hospitals in Germany did not face the threat of closure if they were not covering costs. Furthermore, the regulations and system structure that increased hospital capacity in Germany also increased the number of hospital-based physicians.

REGULATION OF INPATIENT AND OUTPATIENT SEGMENTS. These segments of care in Germany were strictly separated, governed by different organizations and regulatory authorities, and the type of care that each could provide was specified by law. This constraint created a barrier to substitution and coordination between the two sides and specified many services to be performed in the inpatient setting, leading to greater use of inpatient services. In particular, because of regulation, substitution of less resource-intensive outpatient procedures for inpatient procedures did not occur in Germany to the extent it did in the United States, where providers were relatively free to use whatever care settings they chose. For example, the U.S. providers typically used an outpatient surgical biopsy for breast cancer assessment, whereas German providers used an inpatient surgical biopsy; similarly, the United States replaced inpatient chemotherapy with outpatient chemotherapy more quickly than Germany did.

Overall, the constraints on hospital supply and substitution in Germany, resulting from its system structure and strong regulation, led to

its greater use of inpatient services as well as longer treatment lengths, lowering its productive efficiency relative to the United States.

Reconciling Aggregate and Disease Case Level Performance

We noted at the beginning of this paper that the performance of each country's health care system has been assessed by comparing life expectancy (as the measure of outcomes) to the level of health care spending (the measure of inputs). Our disease-level analysis has assessed relative productive efficiency in terms of (quality-adjusted) life years per quantity of input usage at the disease case level. And the results from the case studies seem quite different from the aggregate picture. The diseases we studied are common, important, and, we believe, representative, so the discrepancies in findings are surprising. Several factors may explain the discrepancy with aggregate data.

On the outcomes side, the disease-level analysis generally concluded that U.S. outcomes compared favorably with those in the other countries, while the aggregate data on life expectancy slightly favored Germany and the United Kingdom. On the input side, aggregate spending per capita is much lower in Germany than in the United States, even though the United States used fewer resources per case. The discrepancy between the aggregate and disease-level results is not as wide for the United Kingdom, given that both our results and the aggregate data show low input levels.

There are four main explanations for the differences in results between the aggregate and the case study analysis. First, the incidence of diabetes, breast cancer, and lung cancer might be higher in the United States than in the other countries. Second, the factors of production, notably doctors' salaries, are priced much higher in the United States. Third, the United States carries a substantial administrative burden relative to the other countries. Administrative costs were not included in our disease analyses. Fourth, life expectancy is heavily influenced by neonatal mortality, which is higher in the United States than in the other two countries. Although impaired access to health services and a lack of productivity in this medical activity could contribute to the less favorable birth outcomes in the United States, neonatal mortality is heavily influenced by social and economic factors, along with individ-

ual health behaviors, that are not strongly related to health care delivery. Overall life expectancy at birth, then, may be an unsuitable measure of health outcomes for the purpose of measuring productivity of health services.

Inputs at the Aggregate and Disease Case Levels

Although data limitations precluded direct study of input usage on a national level, proxies for the most important components exist.³⁶ Comparison of various medical inputs used per capita, including physicians, hospital medical personnel, hospital bed-days, and drug prescriptions, showed a pattern across the three countries similar to our findings at the disease case level (figure 7).³⁷ Germany used more of each of these inputs per capita than the United States, which in turn used more than the United Kingdom (except for prescription drugs). Thus the aggregate data are directionally consistent with the disease-level findings. The relative magnitude of the input differences at the aggregate and disease case levels is also very similar in the U.S.-U.K. comparison (figure 8).

Our case results, however, show considerably higher input use in Germany than the level suggested by aggregate data. Among the possible explanations for this discrepancy are higher disease incidence in the United States, the inpatient focus of the sample of diseases studied, and data limitations.

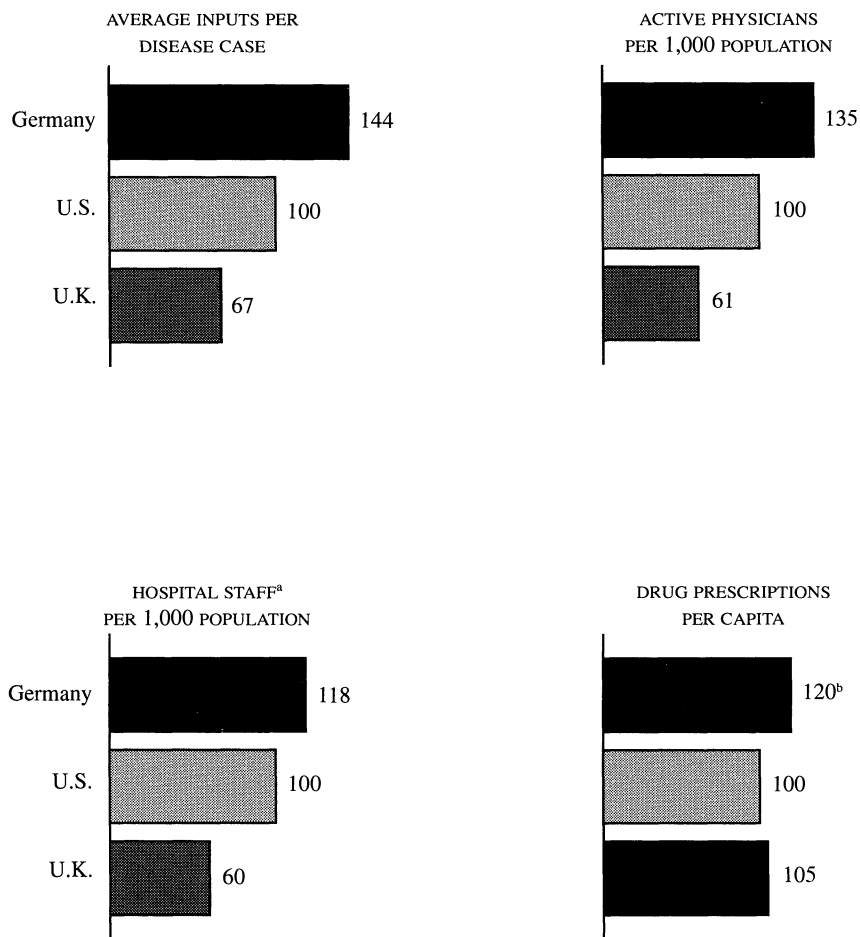
HIGHER DISEASE INCIDENCE IN THE UNITED STATES. Use of medical inputs per capita is driven by both disease-level productive efficiency (inputs per case) and the incidence and mix of diseases in each country (cases per capita). Incidence rates for breast and lung cancer are 27 to 36 percent lower in Germany than in the United States. Thus, the higher input usage per case in Germany is slightly offset by the greater number of cases of lung and breast cancer in the United States. The incidence of diabetes in the United Kingdom is less than half the U.S. rate. The two cancers have greater input usage per case than diabetes and cholelithiasis, so their weight in the total is magnified. Thus, different incidence rates can explain part of the inconsistency in magnitude between aggregate and disease-level input usage.

36. Data on hospital supplies and capital usage were not available on a national basis.

37. The category called hospital personnel includes medical technicians and nurses (qualified and nonqualified).

Figure 7. Medical Inputs by Country, 1990

U.S. = 100

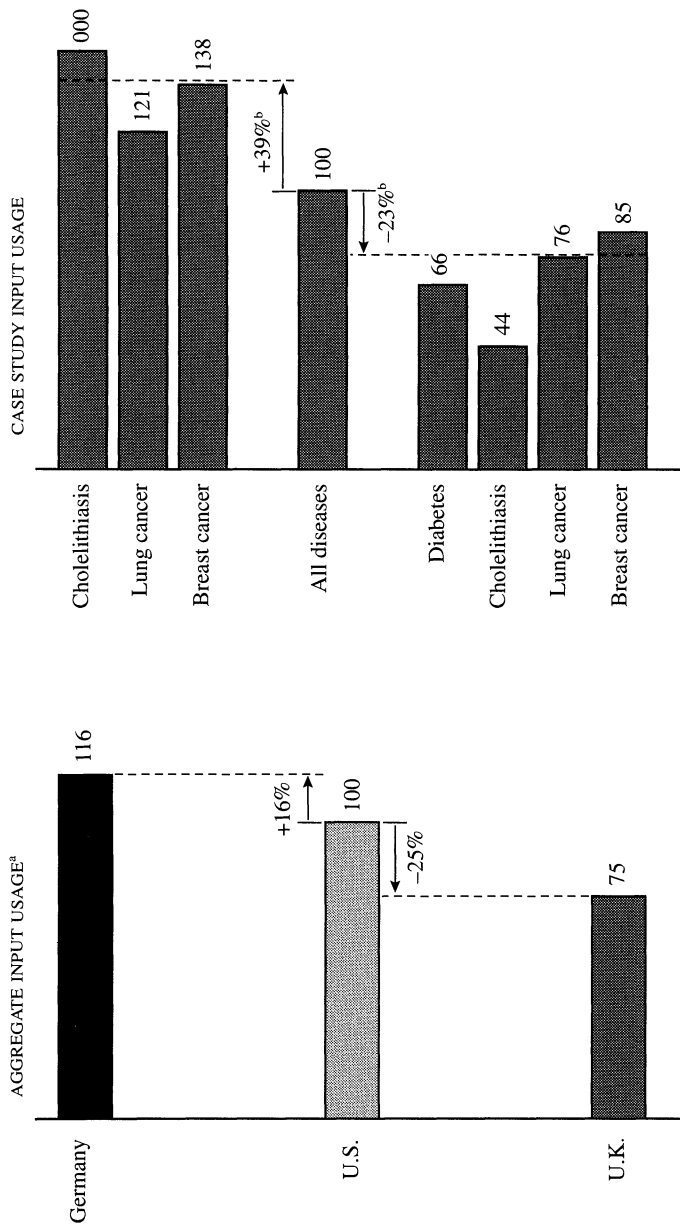


Source: McKinsey Global Institute and the McKinsey Health Care Practice (1996).

a. Hospital staff includes nurses (qualified and nonqualified) and medical technicians for 1989. U.K. data is for England only.

b. Exceeding over-the-counter drugs for which prescriptions are written.

Figure 8. Aggregate Input Differences Compared with Disease Case Input Differences
U.S. = 100



Source: McKinsey Global Institute and the McKinsey Health Care Practice (1996).

a. Weighted average of inputs for labor (70 percent), prescriptions (20 percent), and capital (10 percent) for each of three countries.

b. Weighted average of input usage across the diseases.

INPATIENT FOCUS OF THE DISEASES STUDIED. All three diseases studied in Germany were frequently treated with surgery, and all required significant inpatient stays. These differences in treatment patterns may have biased our results to the extent that Germany's greater use of inputs relative to the United States was concentrated in surgeons and hospital capacity. It is therefore possible that a comparison of treatment processes for outpatient procedures, or for nonsurgical care, would have found smaller differences in inputs between the two countries.

DATA LIMITATIONS. Although we do not think it is a major factor, our input data did not include detailed information on capital costs. Because they were less than 10 percent of total cost in all three countries, this omission should not be a large issue, but it could have led to some discrepancy in aggregate and disease-level comparisons. Germany's supply of hospital capacity per capita far exceeded the U.S. supply, despite higher occupancy levels in Germany. The United States used more of some expensive technologies, such as CT and MRI (magnetic resonance imaging) scanners.

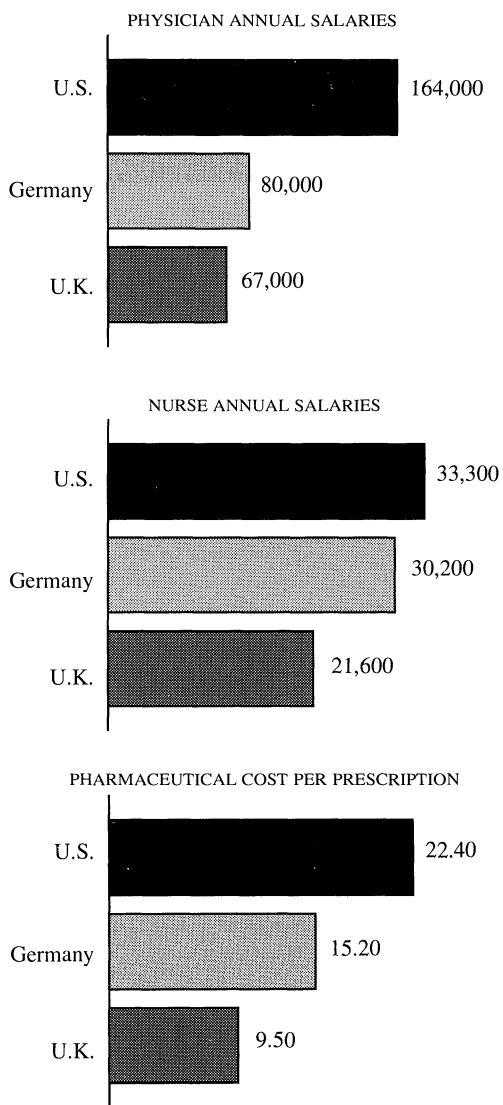
Relative Input Prices

The prices of many medical inputs were higher in the United States than in either Germany or the United Kingdom. Figure 9 shows average input prices in the three countries for physicians, nurses, and prescriptions.³⁸ The most striking differences are in physician incomes. The U.S. physicians earned on average about twice as much as physicians in Germany and about two-and-a-half times as much as physicians in the United Kingdom, reflecting both a higher wage premium for physicians in the United States relative to other professional workers and somewhat higher average wages in the United States.

The pattern of higher input prices in the United States corresponds to the structure of the three health care systems. Both Germany and the United Kingdom featured central administration of their health care financing systems. Their governments and agencies may therefore have acted like monopsony buyers of medical services and used their market power to drive down prices. Although the United States had some

38. These prices are converted to U.S. dollars at GDP PPP (purchasing power parity) ratios for comparability. This price comparison methodology is consistent with our comparison of per capita health care spending in U.S. dollars at GDP PPP.

Figure 9. Input Prices across the United States, Germany, and the United Kingdom
U.S. dollars^a



Source: McKinsey Global Institute and the McKinsey Health Care Practice (1996).

a. 1990 U.S. dollars, adjusted for purchasing power parity.

market power in purchasing (mostly through Medicare), many input prices were set in markets without dominant buyers but with strong sellers. Thus the relative concentration and market power of buyers and sellers of medical services in the three countries may have contributed to the observed differences in input prices. In addition, differences in relative provider skill or experience levels may have contributed to observed price differences, which in turn could have contributed to different productive efficiency levels. Furthermore, physician incomes in the United States reflect to some extent the significant education costs borne directly by the physician. A comprehensive analysis of pricing levels, their causes, and their potential effect on productive efficiency was outside the scope of our study.

Relative Administrative Spending

Administrative spending includes four distinct cost categories that are difficult to disaggregate: (1) payor, provider, and government agency costs for administering the insurance and provider reimbursement system; (2) provider costs associated with managing health care facilities and practices; (3) payor costs for selling and marketing health coverage products to purchasers and members; and (4) payor and provider costs for care management, including utilization review and quality assurance. We combined information on administrative costs incurred by payors and hospitals in the three countries, together with suggestive data from a U.S.-Canada comparison study to estimate total administrative spending at 24 percent of total costs in the United States, 13 percent in Germany, and 16 percent in the United Kingdom.³⁹

Several factors may have contributed to higher administrative costs in the United States. For example, the relative fragmentation of providers and payors and the resulting complexity of the insurance and reimbursement system may have played a major role; a single-payor system can simplify the providers' interface with the reimbursement system by eliminating much of the claims processing and can reduce or even eliminate marketing and sales expenses.

It is possible that administrative costs cannot be separated precisely

39. These figures are rough calculations and may be a slight overestimate because the hospital administrative cost percentage appears to be slightly greater than the percentage for all health care services.

from patient care. Higher administrative costs in the United States may have resulted from a more significant care management function on the part of payors and providers, which in turn could have contributed to the higher U.S. productive efficiency observed in the disease cases. There may be a trade-off between productive efficiency and the cost of running the system.

The reasons for studying the disease treatment on its own are that we did not have administrative inputs by disease, that we remain unsure of the validity of our administrative cost estimates even in the aggregate, and that we have little evidence about the size, or even existence, of a trade-off between administrative costs and treatment costs. As we discuss later, Germany and the United Kingdom could probably raise their treatment productivity without adding much to administrative costs; and the United States is already finding ways to cut such costs without changing the basic competitive structure of the system, notably through the proliferation of managed care.

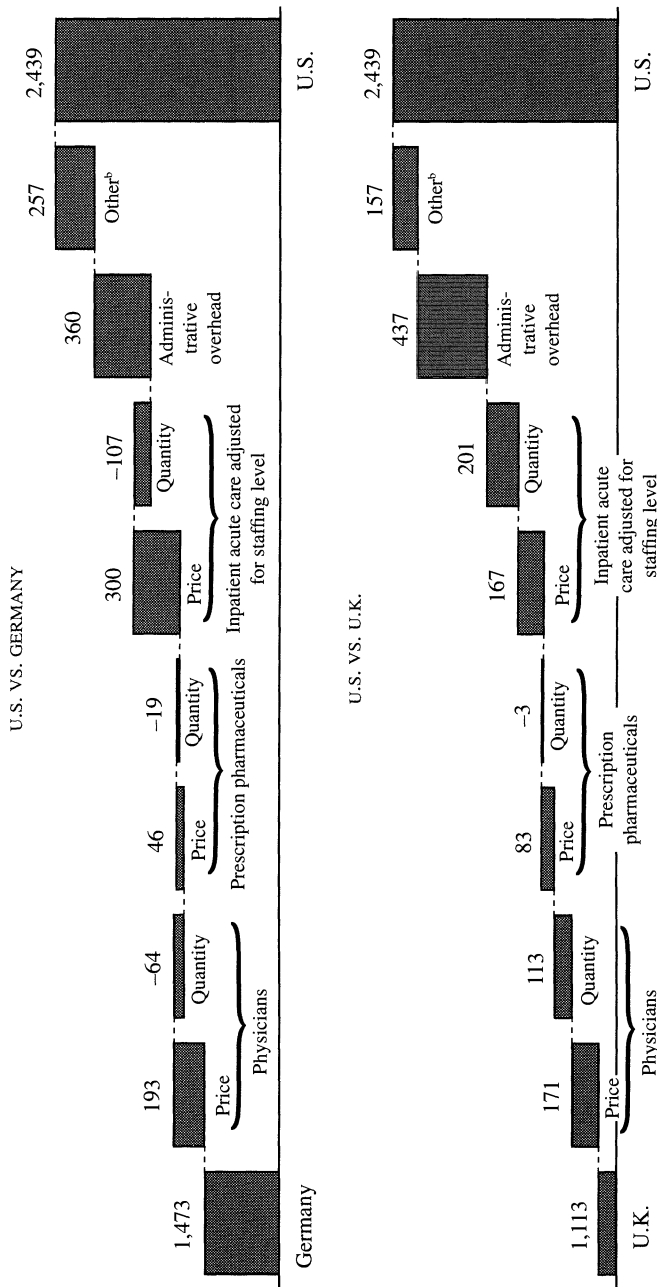
To see how sensitive our results are, however, one can make a crude adjustment for administrative costs by adding 24 percent, 13 percent and 16 percent, respectively, to the inputs in the three countries. This adjustment would leave the productive efficiency advantage with the United States for lung cancer and cholelithiasis, although it would close the gaps (raising relative productivity by about 10 percent for Germany and 7 percent for the United Kingdom). The breast cancer results would continue to be ambiguous, and the United Kingdom would look even stronger in diabetes.

Decomposing the Spending Differences

To help understand the aggregate spending numbers better and to judge the relative importance of the alternative reasons for the discrepancy between these numbers and the case studies, we decomposed the aggregate spending differences into price, quantity, administration, and residual factors. The results are shown in figure 10. We started with expenditures on physicians, drugs, and hospitals. We then broke these down into the number of physicians and their salaries, the quantity of prescription drugs and average prices for drugs, and the number of bed-days and the price per bed-day. The last category of hospital spending was adjusted for the higher staffing level in the United States. The

Figure 10. Spending Differences by Physicians, Prescriptions, and Hospitals

U.S. dollars per capita^a



Source: McKinsey Global Institute and the McKinsey Health Care Practice (1996).

a. 1990 U.S. dollars, adjusted for purchasing power parity.

b. Other includes overhead costs in physician practices, over-the-counter drugs, outpatient hospital care, nursing homes, dental care, and medical devices; in addition, it includes the price and input effects of categories other than physicians, prescription pharmaceuticals, and inpatient care.

component of per diem charges attributable to higher staffing in the United States was counted as a quantity, not a price difference. The administrative cost figures are based on the information discussed earlier.⁴⁰ Finally, there is a residual, or “other,” factor, reflecting the remaining differences in per capita spending that are not part of the three main categories. Examples include dental care and doctors’ assistants.

This decomposition is necessarily rough, but it does suggest that price differences make up the largest element in explaining the higher expenditure in the United States. The gap in doctors’ salaries alone accounts for 20 percent of the spending gap between the United States and Germany and about 13 percent of the U.S.-U.K. gap.

On balance, therefore, we believe that the aggregate numbers are consistent with the results of the cases. The higher rate of spending in the United States is driven largely by higher prices and administrative costs. Higher disease incidence may also be a factor, although its importance in the aggregate was not something we could estimate. Overall, we judge that the high productivity of the U.S. health care delivery system works to offset these other reasons for high health care costs.

Recent and Future Changes in the Health Care Delivery Systems

The results of the case studies come from the 1980s; since then there have been important changes in all three countries. In the United States the largely market-based system is leading to greater competitive intensity and an increased ability to provide integrated care, even without significant regulatory changes. More integrated and managed care is being provided by HMOs and preferred provider organizations. These approaches have grown rapidly in importance as employers have demanded lower cost health care coverage for their employees.

40. Other adjustments were also made to the per diem hospital charges. Daily rates include payment for the administrative costs of the hospital, so we have tried to remove this part of the payment (so as not to double-count such costs). We have also tried to adjust for that part of the hospital costs paid for directly by governments.

Specialized clinics and more aggressive management for diabetic care have emerged, including an emphasis on self-care, as a result of actions on the part of integrated provider systems, managed care payors, and manufacturers of diabetic supplies. The benefits of such an integrated approach to care started to outweigh the potential adverse selection problem. Furthermore, “disease management” approaches to care as a way to manage costs and improve outcomes have grown in popularity among managed care organizations, integrated provider systems, and suppliers.

Not surprisingly, these developments have also led to a decline in compensation for specialist physicians and to actual price reductions for health coverage in some markets. The effects on administrative costs are unclear, however. Although recent consolidations among and between payors and providers have led to administrative cost decreases, there is some evidence that the share of administrative costs focused on care management (in the form of information systems, personnel, and so forth) have increased. These changes, however, may have improved productive efficiency.

In the United Kingdom, reforms passed in 1991 introduced some competition at the local level between the payor function and providers through the creation of an internal market, fostered somewhat more integrated care, but left the lifetime payor coverage and monopoly power of the NHS largely intact. More decentralized health authorities were given the responsibility of purchasing services from competing providers; general practitioners were allowed to become “fundholders” and thereby assume and manage the financial risk of a broader set of care provision services (such as drugs, outpatient care, diagnostic tests, and nonurgent surgical procedures); and many NHS-owned hospitals were effectively privatized into self-governing trusts. In addition, these hospital trusts were given greater control over their capital purchases, with funds loaned to them by the government with interest, much like a commercial transaction. The overall budget and many other supply constraints remain, however, and efforts to encourage the use of non-public financing sources have met with little success.

Although these system changes apparently have not increased administrative costs, their effects on productive efficiency are still unclear. As many as 50,000 nursing jobs and 60,000 hospital beds have been eliminated since 1990, but 20,000 more senior managers have

been added in the NHS, according to some estimates.⁴¹ And there is some evidence that adoption of technology has quickened (for example, a targeted breast cancer screening program based on mammography was established; adoption of laparoscopic technology for cholecystectomy has neared U.S. levels), resulting from better NHS evaluation and fiat as well as from increased provider responsiveness to demand. It is also possible that the general practitioner fundholders can now encourage and achieve more rapid incremental improvements in health care delivery by exerting more direct pressure on local specialists and hospitals. Although some supply and capital constraints remain for hospitals and their associated specialists, and competition has been limited outside the major metropolitan areas, we would expect some improvement in the system's productive efficiency over time, at least in the diseases studied.

In Germany major reforms have been made in health coverage, and to a lesser extent, the care provision markets. As of 1996 payors (sickness funds) are allowed to compete for members on the basis of price and other factors, but restrictions on their ability to negotiate price differentially with providers or to bundle care in different ways (by disease or case, for example) have been left intact. Regulated case-rate payments for hospitals have been introduced to substitute for per diem payments, but they cover only about 15 to 20 percent of cases. Regulatory barriers between inpatient and outpatient care remain, as do the regulatory processes for controlling hospital and physician supply. Payors are, not surprisingly, searching actively for and adopting the U.S. practices for managing care—such as hospital utilization management—but they face significant regulatory limitations in what they can implement. Additional reforms under discussion for 1997 are focused on managing hospital costs through, for example, the introduction of a regional- or state-level hospital budget.

Recent changes in the German system are unlikely to improve productive efficiency much, unless they eventually lead to removal of regulatory constraints on inpatient and outpatient substitution, greater flexibility in payors' negotiations with individual or groups of hospitals and physicians, or to the widespread adoption of case-rate hospital payments.

41. Whitney, Craig R. 1996. "Health Squeeze—A Special Report: Rising Health Costs Threaten Generous Benefits in Europe," *New York Times*, August 6, p. A-1.

Conclusions

The desire to limit government and private expenditures for health care while improving health outcomes makes health care productivity an important policy issue throughout the world. Although productivity is only one aspect of the performance of any health care system, improvements in productivity can make it easier to achieve other health system goals, such as greater access to care and protection from the financial losses resulting from ill health.

For three of the four cases examined here the productive efficiency of the treatment delivery part of the U.S. health care system compared favorably with that of Germany and the United Kingdom. The productive efficiency of the United States exceeded that of Germany and was never clearly inferior to that of the United Kingdom. Only in the management of diabetes, a chronic disease that can be treated better with the kind of integrated disease management implemented in the United Kingdom in the 1980s, did the United States fall behind.

Patterns of care were consistent with the incentives and constraints operating in each system. The United States had the most heterogeneous system, which during the late 1980s was characterized by fee-for-service reimbursement for health care, relatively low levels of integration of services, a high degree of competition among payors and providers, and relatively few regulatory constraints on the organization of services and the acquisition of new medical technologies.

The United Kingdom's governmental system of health care financing and delivery had a single payor and little or no direct competition among providers. The budgeting system used to reimburse providers led to constraints on resources, particularly for capital acquisition, with effective limitations on overall expenditures for health care. The system was relatively well positioned to implement integrated programs for managing chronic diseases like diabetes, yet underinvestment in both new and old technologies may have impaired productivity.

The German system was substantially more regulated than that in the United States, with little flexibility in the organization of services. Several features, such as rewards for longer hospital stays, served as disincentives to increase overall productivity.

Although the case studies did not demonstrate that any single form of organization of care was associated with uniformly greater produc-

tivity, we believe that they strongly suggest that flexibility in the organization of care, coupled with competition among providers and appropriate incentives, is most likely to promote productivity in the treatment process. The unanswered question is whether and to what extent higher prices and administrative costs are or are not results of this same flexible and productive system and hence offset much of the productivity advantage. It remains to be seen whether the advantages of the U.S. health care system can be obtained while holding down administrative costs and putting appropriate competitive pressure on prices. It is also unclear whether quality of care will deteriorate with declines in the prices of individual or bundled health services. Nevertheless, we believe that recent trends, including improved measurement of health outcomes, greater price sensitivity among purchasers, and various administrative efficiencies, will help the United States improve overall performance in health care in the coming years.

References

- Arbeitsgruppe zur Koordination Klinischer Krebsregister (Institute for Medical Information) database. Giessen, Germany.
- British Diabetic Association. 1996. *Diabetes in the United Kingdom-1996*. London.
- Dunn, David, and others. 1994. "Laparoscopic Cholecystectomy in England and Wales: Results of an Audit by the Royal College of Surgeons of England." *Annals of the Royal College of Surgeons* 76: 269-75.
- Edinburgh Lung Cancer Group. 1987. "Patients Presenting with Lung Cancer in South East Scotland." *Thorax* 42: 853-57.
- Foreman, David, and Lesley Rider. 1993. *Cancer in Yorkshire: Cancer Registry Special Report Series*, vol. 3, *Breast Cancer*. Leeds, Eng.: Leeds Graphic Press, Ltd.
- Forrest, Patrick, and others. 1986. "Breast Cancer Screening: Report to the Health Ministers of England, Wales, Scotland, and Northern Ireland." Department of Health and Social Security, London.
- Garber, Alan M., and Charles E. Phelps. 1997. "Economic Foundations of Cost-Effectiveness Analysis." *Journal of Health Economics* 16: 1-31.
- Graves, Edmund J. 1995. "National Hospital Discharge Survey: Annual Summary, 1993." *Vital Health Statistics*, series 13, no. 121. Centers for Disease Control and Prevention, U.S. Department of Health and Human Services, Hyattsville, Md.

- Großhadern Dataset. Tumorzentrum München, Ludwig Maximilian University and Technical University of Munich. Munich.
- Humphrey, Edward, and others. 1990. "National Survey of the Pattern of Care for Carcinoma of the Lung." *The Journal of Thoracic and Cardiovascular Surgery* 100: 837–43.
- Jensen, O. Møller, and others. 1990. "Cancer in the European Community and its Member States." *European Journal of Cancer* 26 (11/12): 1167–1256.
- Joslin, Charles, and Lesley Rider. 1995. *Cancer in Yorkshire: Cancer Registry Special Report Series*, vol. 1, *Lung Cancer*. Leeds, Eng.: Leeds Graphic Press, Ltd.
- Krämling, H.-J., and others. 1993. "Aktueller Stand der Gallensteinchirurgie in Deutschland." *Der Chirurg* 64: 295–302.
- Krebsregister Saarland database, Saarbrücken, Germany.
- MacIntyre, I. M. C., and R. G. Wilson. 1993. "Laparoscopic Cholecystectomy." *British Journal of Surgery* 80: 552–59.
- McKinsey Global Institute and the McKinsey Health Care Practice. 1996. *Health Care Productivity*. Los Angeles: McKinsey and Co., Inc.
- National Center for Health Statistics. 1995. "Detailed Diagnosis and Procedures, National Hospital Discharge Survey, 1993." *Vital Health Statistics*, series 13, no. 122. Centers for Disease Control and Prevention, U.S. Department of Health and Human Services, Hyattsville, Md.
- National Diabetes Data Group. 1995. *Diabetes in America*, 2d ed. NIH Publication 95-1468. Washington, D.C.: National Institutes of Health.
- National Health Service. 1995a. *Hospital Episode Statistics: Finished Consultant Episodes by Diagnosis, Operation and Specialty*, vol. 1. London: Department of Health.
- National Health Service. 1995b. *Hospital Episode Statistics: Finished Consultant Episodes: Waiting Time of Elective Admissions by Diagnosis and Operation; Injury/Poisoning by External Causes*, vol. 2. London: Department of Health.
- Roslyn, Joel, and others. 1993. "Open Cholecystectomy: A Contemporary Analysis of 42,474 Patients." *Annals of Surgery* 218 (2): 129–37.
- SEER (Surveillance Epidemiology and End Results), public use database. 1996. *Cancer Statistics Review 1973–93*. National Cancer Institute, National Institutes of Health, Washington, D.C.
- Thames Cancer Registry database, Sutton, England.
- Wingo, Phyllis A., and others. 1995. "Cancer Statistics, 1995." *CA—A Cancer Journal for Clinicians* 45: 8–30.
- World Health Organization. 1987–1992. *World Health Statistics Annual*. Geneva.

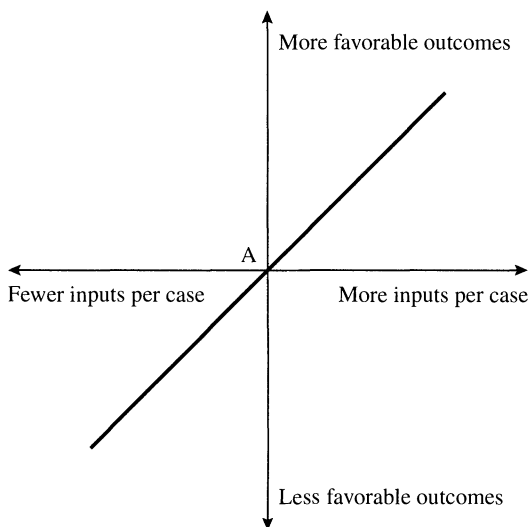
Comments

Comment by Ernst R. Berndt: Martin Baily and Alan Garber seek answers to two very important and timely questions: First, what are the sources of the apparently very large differences in health care spending among industrialized countries? And second, why do those spending differences appear to be unrelated to differences in overall life expectancy?

Rather than attempting to answer these two questions directly, Baily and Garber examine the treatment costs and protocols for four relatively common diseases or conditions in three countries. For each of the four, they quantify the total physical inputs used in treating a case of that disease, the medical and quality-of-life outcomes achieved by the health care sector in each country in treating that particular disease, and, thus, the productive efficiency or productivity levels realized by each of the three countries in treatments of these four diseases.

This approach to assessing the sources and consequences of variations in health care expenditures across countries is extremely useful, for it allows the investigators to combine aggregate-level analyses with disease-specific details and considerations. Moreover, such a focus on input, output, and outcomes measurement at the level of specific diseases builds on an emerging health economics literature.¹ The conceptual framework employed by the authors envisages disease treatment as a productive process involving inputs and outputs and one in which outputs can be positive or negative.

1. See, for example, Cutler and others (1996); Ellison and Hellerstein (1997); Frank, Berndt, and Busch (1997); and Triplett (1997).

Figure 1. Productivity Comparisons among Countries

Source: Author's calculations.

The basic idea is displayed in figure 1, where favorable outcomes are on the vertical axis, and an index of total input quantity per treated case is on the horizontal axis. To establish productivity rankings, Bailey and Garber undertake pairwise comparisons. Relative to point A in the middle of the figure, we see that in the northwest quadrant, greater favorable outcomes are associated with fewer total inputs per treated case than at point A; thus all points in this northwest quadrant unambiguously represent greater average productivity than at A. The southeast quadrant is symmetric, in that relative to point A, points here represent lower average productivity—favorable outcomes are fewer and total inputs per treated case are larger. Without further structural information on the shape of the production function, however, one cannot rank average productivity in the southwest or northeast quadrants relative to point A, for favorable outcomes are larger (smaller) and total inputs per treated case are larger (smaller) in the northeast (southwest) quadrant than at point A.

To resolve such potentially ambiguous cases, the authors assume that “the treatment of a given disease is fundamentally a diminishing returns

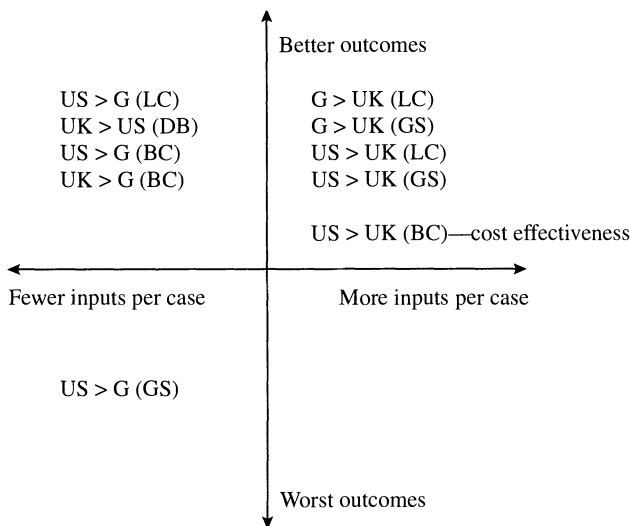
activity.” Under this assumption (and even under the weaker assumption of nonincreasing returns), as long as a country or disease lies on a point above (beneath) a 45-degree line emanating from A, that country has greater (lesser) productivity than A.

The plausibility of the assumption of global decreasing returns at the per-case level depends on the empirical validity of two assumptions: that patients who are most likely to benefit are the first to be treated; and that the most cost-effective treatments are the first to be used.

It is useful to consider cases where these assumptions might not be valid. If physicians do not choose which patients should be treated first (if triage is not operative), but instead typically treat those who present themselves with symptoms and illnesses with varying levels of severity on a first-come, first-served basis, then the first assumption would not necessarily hold. Moreover, for those medical illnesses and conditions that are substantially underdiagnosed, there is no a priori reason to expect a merit ordering of severity from those who seek treatment. Indeed, as the authors hint when discussing treatments for cancer patients, those beyond a reasonable chance of successful treatment could in many cases be the first to present themselves for diagnosis and treatment. With respect to the second assumption, given issues of moral hazard, third-party payment, and well-known geographical variations in medical practice within the United States, it is not at all clear that the most cost-effective treatments are, in practice, the first to be used.² In short, the decreasing returns assumption is one that can plausibly be called into question. Fortunately, in their empirical work, all that the authors need is the weaker assumption of nonincreasing returns to scale.

Although relatively straightforward conceptually, the actual conduct of this empirical research effort involved enormous resources; McKinsey apparently had as many as twenty people working on this project over a three-year period. This is a massive empirical study. Unfortunately, the sheer size and length of the study mean that when it is finally published, it runs the danger of being already somewhat out of date. The data used are drawn from the mid- and late 1980s and are therefore already a decade old, a rather long time in the rapidly changing health care marketplace. Indeed, even the borders of the country of Germany have changed since the data were collected.

2. See, for example, Wennberg (1984) and the studies summarized in Folland and Stano (1990).

Figure 2. Summary of Baily-Garber Conclusions

Source: Author's calculations.

Note: US is United States, UK is United Kingdom, and G is (West) Germany. LC refers to lung cancer, DB to diabetes, BC to breast cancer, and GS to gallstones.

The Baily-Garber conclusions are summarized in figure 2 here. In that figure, when country A has productivity superior to that of country B, the notation $A > B$ is employed. As seen in the northwest quadrant of figure 2, in four instances the pairwise country comparisons are unambiguous, because A has fewer inputs per case and better outcomes than does B. Assuming nonincreasing returns to scale, five more rankings can be obtained; these appear in the southwest and northeast quadrants. Only in one case, that involving a comparison of productivity between the United States and the United Kingdom in the treatment of breast cancer, is it impossible to reach a definitive conclusion without further information. Even here, however, considerations of cost effectiveness employed by the authors enable them to rank the U.S. productivity higher than that of the United Kingdom.

What are the sources of these differences? David Cutler elaborates more on these differences in his comments, but let me briefly state that relative to Germany, the United States tends to employ more outpatient and less inpatient days of services, and conditional on hospitalization,

average length of stay is shorter in the United States. Relative to the United Kingdom, the United States has been quicker to adopt such high-tech medical products and procedures as CT scanners and laparoscopic surgery, but its (until recently) more fragmented health care system has not taken as much advantage of team provider processes such as those involved in the United Kingdom's treatment of diabetes.

Baily-Garber use QALYs (quality-adjusted life years) in the gallstone and diabetes outcomes measurements, but not for breast and lung cancers. Survival rates for patients with lung cancer are very low, and neglect of QALYs is a defensible research strategy. For breast cancer, however, five-year survival rates in the United States are currently at least 50 percent and rising, and thus the *a priori* case for excluding QALYs is not very strong—and one I am sure that patient advocacy groups would vigorously challenge. Quality-of-life adjustments could be quite significant in the treatment of breast cancer. Substantial progress has been achieved, for example, in making chemotherapy treatments for breast cancer less burdensome for patients; recent noteworthy biotech and pharmaceutical product innovations reduce discomfort from nausea, leave patients less vulnerable to infections, and increase energy levels by reducing anemia. There may well be differences across countries in the use of such products.

Finally, the medical community has learned in the last few years that a very substantial proportion of women diagnosed with breast cancer experience depressive episodes during treatment, a condition that apparently has been considerably underdiagnosed. How treatment of breast cancer deals with associated comorbid conditions such as depression is a topic of considerable importance in interpreting international differences in treatment costs and quality-of-life-adjusted outcomes. Thus the *a priori* case for including QALYs when comparing outcomes of treatment for breast cancer across countries is a rather plausible one. As the authors note, quality-of-life data associated with the treatment of both breast and lung cancer apparently were not available when the data were collected.³

The Baily-Garber study is a massive one tackling a very important set of issues, but many issues remain to be explored. Two are particu-

3. There is, however, a substantial literature on the measurement of quality of life in cancer patients. See, for example, Barofsky (1996), Goodyear and Fraumeni (1996), and Padilla and others (1996).

larly interesting. First, it is well known that within the United States (as well as in other countries), there is a tremendous diversity and variation among hospitals and physicians in treatment protocols for particular diseases, illnesses, and conditions.⁴ Although the existence of the National Health Service in the United Kingdom might facilitate greater uniformity in treatments, and although Baily and Garber have undertaken a thorough analysis involving numerous interviews with clinicians and health care administrators, I would have liked to have seen at least a bit more discussion on literature dealing with the size of within-country variation relative to between-country variation.

Second, it is also well known in health economics that for many medical expenditure categories, the mean treatment expenditure is much larger than the median, because of the existence of a far right tail reflecting relatively small numbers of extremely high-cost or high expenditure cases. In this paper Baily and Garber examine means and compare means across countries. An alternative approach would involve focusing on the outliers—how do treatments of the very rare but costly cases differ across countries? Focusing on the extreme high-cost cases may have just as large an effect on explaining differences in total costs per case across countries as does a focus on the most representative or median treatments.

Baily and Garber conclude by providing four useful facts regarding the apparent “big picture” paradox of unproductive medical expenditures in the United States. First, for each of these four diseases, the United States has a higher incidence than either the United Kingdom or Germany. Why that is the case is not clear, nor is it clear whether this generalizes to other diseases or conditions. Are we Americans more disease prone? Or has our insurance system with all its principal-agent incentive distortions yielded greater amounts of diagnosed illnesses per capita?

Second, at least in the mid- and late 1980s, the factors of production in health care were all higher priced in the United States. Recent work suggests that at least initially the impact of managed care in the United States on treatments of certain conditions has involved greater reductions of prices than of treatment quantities.⁵ To the extent this is true,

4. See Wennberg (1984) and Folland and Stano (1990) for further discussion.

5. Cutler, McClellan, and Newhouse (1997); Berndt, Frank, and McGuire (1997).

the expenditure differences among the three countries in the period observed by the Baily-Garber study may well be decreasing as managed care diffuses more fully in the United States.

Third, the Baily-Garber study excludes administrative costs, which they estimate to be 24 percent of total costs in the United States, 16 percent in the United Kingdom, and only 13 percent in Germany. What the influence of managed care will be on administrative costs is not yet clear, particularly because cost-reducing developments in information technology would appear to give particularly large cost advantages to centrally administered single payor systems, such as that in the United Kingdom.

Fourth and finally, and now in the context of aggregate data rather than disease-specific cases, aggregate life expectancy appears to be affected quite critically by neonatal mortality, which is worse in the United States. Differences among countries in life expectancy conditional on surviving to a threshold year are considerably smaller. This leads Baily and Garber to put one red herring to rest for good: "Overall life expectancy at birth . . . may be an unsuitable measure of health outcomes for the purpose of measuring productivity of health services." That is a simple but important message well worth remembering.

Comment by David M. Cutler: This paper is an extremely nice look at a very difficult question. Measuring the productivity of the medical care system is hard enough; comparing productivity across countries is even more difficult. Yet that is the task of the paper. Some evidence on the sheer magnitude of the work involved comes from knowing that the longer study on which this summary is based is two inches thick, and the list of McKinsey personnel involved in the study is perhaps twenty lines of very dense text.

The methodology of the paper is to compare productivity for four conditions in the United States, the United Kingdom, and Germany and then to try to generalize from these cases. This seems exactly the right approach. Productivity comparisons of the medical system as a whole are just too difficult to be of much use in this market. To compare medical care productivity across countries, one must look at a more detailed level.

The paper concludes that with few exceptions, the United States is more productively efficient than either the United Kingdom or Ger-

many. Because the paper measures average productivity, not marginal productivity, this conclusion does not seem too surprising. But at a closer look, the results are quite surprising. Consider what is conventionally thought to be a description of the U.S. medical care system: If you are fortunate to be well insured, the medical care system in the United States is the best in the world. The care is lavish, the technology is sophisticated, and the doctors are the best. Outcomes are better in the United States than elsewhere. But you pay a lot for this. For most people, the additional care is worth it. Some people, however, get much more care than is appropriate. At the margin, therefore, there is a substantial waste of resources.

This is not what this paper finds. Although there are differences across diseases, consider the paper's analysis of productivity for cholelithiasis (gallstones). For patients who receive surgery, inputs are substantially less in the United States than in the other two countries. Some of this is due to technology; laparoscopic cholecystectomy (a less invasive form of surgery than the traditional open cholecystectomy) diffused much more rapidly in the United States than in either of the other two countries. But in some cases, the United States also uses fewer inputs for a given procedure, and hospital stays in the United States are far shorter than they are in the other two countries. On the outputs side, the rate of surgical complications is essentially the same across countries. The net effect is that the United States is more productive than either the United Kingdom or Germany, but the difference is entirely on the input side. Outputs per treatment are the same; the United States just provides those treatments with fewer resources.

Although there are certainly differences across diseases, the dominant conclusion of the Baily-Garber study is that most of the United States's higher productivity is because the medical care system uses fewer inputs, such as shorter hospital stays and more outpatient care. Outcome differences are much smaller across countries, and for some diseases, output conditional on treatment is the same across countries.

This finding is very different from the conventional wisdom. In my comments I want to bridge the gap between the findings in this paper and the conventional wisdom and then offer a few comments about the definition of productivity. I shall mix thoughts about the current paper with suggestions for future research.

Incorporating Morbidity

The most difficult issue in the paper is the measurement of outcomes. For breast cancer and lung cancer, the authors use survival to measure outcomes. This makes good sense; for a patient with cancer, survival is almost everything. For the nonfatal conditions (diabetes and cholelithiasis), however, the authors do not have particularly good measures of outcomes. For cholelithiasis, there is essentially no measure of morbidity (other than surgical complications, which are the same across countries). And for diabetes, the authors use data on some complications, but the most life-threatening complications (end stage renal disease, ischemic heart disease, and stroke) cannot be isolated well enough to be incorporated. In addition, there is no analysis of morbidity conditional on the level of complications.

The lack of good morbidity data for nonfatal illness is crucial. If one asks what a very intensive medical care system is likely to provide for a generally nonfatal illness, the answer is twofold: more sophisticated techniques to improve quality of life; and a greater chance of detecting very rare complications. The first of these is essentially ruled out because the authors cannot measure it. The second is likely to be too difficult to detect without a much larger sample of patients. Thus, it seems to me that the paper systematically undercounts morbidity benefits from very intensive medical practice. Because the United States has the most technology-intensive medical system, I suspect the paper systematically undercounts health outcomes in the United States.

Indeed, it is striking that for lung and breast cancer, outcomes conditional on treatment are generally better in the United States than they are in other countries, but that is not true for the nonfatal conditions. This may be because the appropriate morbidity data is not there to allow this comparison.

Raising the issue of better outcome measures is fine, but what should be done about it? It seems to me that a different sampling method is required from the one the authors use. The authors measure outcomes by using cross-section data where it is available. For example, complication rates for diabetes in the United States and the United Kingdom are based on the number of hospital admissions for particular complications divided by the number of people with diabetes. If the number

of hospital admissions cannot be measured accurately, there is no way to measure morbidity.

An alternative estimation strategy would be to sample patients with diabetes in different countries and follow these patients over time to determine whether they experience any adverse events. In some cases, the longitudinal samples have already been drawn (for example, the Framingham Heart Study on cardiovascular disease); indeed, this is the type of data the authors use for cancer. In other cases, one could get insurance records for people with a particular condition and monitor their disease progress over time.

Using a longitudinal sample from insurance records or other sources would have two additional advantages. First, it would allow the authors to adjust more accurately for the severity of disease across countries. For example, in claims records one can find out about the patient's entire medical record and thus construct an estimate of comorbid conditions. Baily and Garber generally wind up assuming that disease severity is the same across countries. Second, when patients have been ranked by severity, the authors could look at how patients with different levels of severity fare in different countries.

I suspect the lack of morbidity data is sufficiently large that it skews the outcome measures for the nonfatal diseases quite severely. My guess is that outcomes for the nonfatal conditions—and even for the potentially fatal ones—are better in the United States relative to the other countries than the paper suggests. But documenting this issue one way or the other will require a different approach to the problem than what the authors have chosen.

Input Differences across Countries

Perhaps the most consistent finding in the paper is that the United States uses substantially fewer inputs per medical treatment than do other countries. As noted earlier, most of the productivity differences across the three countries come from the level of inputs used. Some of these differences are clearly present and are important. The higher use of laparoscopic surgery in gallstone cases in the United States is an example.

But many more of the differences are in the length of hospital stay or in the use of inpatient versus outpatient care. That is particularly

true in the comparison of the United States and Germany (where length of hospital stays are much greater). I am less convinced of the degree of resource savings than are the authors. The important issue is how much hospital care is provided in the marginal day of care in Germany, and how many of these resources wind up being provided at home or in outpatient settings in the United States. For many surgical procedures, the last days of hospital care are relatively unintensive: monitoring the patient or controlling pain that could be done in nonhospital settings.

To be sure, some additional resources are used in the hospital compared with outpatient settings, but the resources are less than those used on an average day. Although the days in the hospital are adjusted for the average intensity of services provided, mismeasurement of average service use is a concern. As a result, shorter hospital stays might generate more (or less) apparent resource savings than are actually realized. Anecdotal evidence from managed care and hospital executives suggests that marginal services provided in the last few hospital days are not particularly large. Many managed care insurers have learned that the dollar savings they realized from reducing hospital stays were not particularly large. Because shorter lengths of hospital stays are virtually the entire reason why the United States is more productive than Germany, some caution about this conclusion is appropriate.

This factor and the previous one may offset each other, and the United States may well be more productive than the other countries. But I suspect the paper probably overstates the resource differences across countries and understates the outcome differences in a way that affects the conclusions about why productivity differs across countries.

Productivity Measurement

The final issue I want to address is the concept of productivity, particularly the authors' focus on *average* productivity across countries. International comparisons of medical systems raise two issues. First, is one country better on average than another? This is the question the paper answers. Second, are the resources being put into the medical system at the margin worth their cost? This answer may differ across countries; in the United States, my guess is the answer is "no." When insurers cut back on the care received, for example, a typical finding is

that spending can be cut by a significant amount without substantial adverse effects. This is consistent with a low marginal value to medical services.

This issue has important implications for the conclusions of the paper. Baily and Garber summarize their conclusions as “competition promotes high productivity.” I would revise the statement to “competition promotes high average productivity but low marginal productivity.” I suspect the ranking of countries is not nearly as obvious as the paper suggests.

The Baily-Garber paper, and the McKinsey study behind it, are a major advance in the understanding of the productivity of the medical system. They conclude that the United States is the most productive of the three countries they analyze. I suspect they are right. But I would guess that there is more to the conventional wisdom than the results indicate. On the basis of this paper, one would conclude that productivity is high in the United States because fewer inputs are used to produce the same output. My sense is that productivity is higher because the United States achieves greater outputs than the other two countries with the same or more inputs. To examine these questions, economists will need to make use of longitudinal data that the Baily-Garber paper does not use. Much work remains to be done.

Commentators' References

- Barofsky, Iva. 1996. “Cancer: Psychosocial Aspects.” In Spilker (1996).
- Berndt, Ernst R., Richard G. Frank, and Thomas G. McGuire. 1997. “Alternative Insurance Arrangements and the Treatment of Depression: What Are the Facts?” *American Journal of Managed Care* 3 (February): 135–43.
- Cutler, David M., Mark B. McClellan, and Joseph P. Newhouse. 1997. “Prices and Productivity in Managed Care Insurance.” NBER Working Paper. National Bureau of Economic Research, Cambridge. September.
- Cutler, David M., and others. 1996. *Are Medical Prices Declining?* NBER Working Paper 5750. Cambridge, Mass.: National Bureau of Economic Research.
- Ellison, Sara Fisher, and Judith Hellerstein. 1997. “The Economics of Antibiotics: An Exploratory Study.” NBER Working Paper. National Bureau of Economic Research, Cambridge. May.
- Frank, Richard G., Ernst R. Berndt, and Susan Busch. 1997. “Constructing

- Quality-Adjusted Producer and Consumer Price Indexes for the Treatment of Depression: An Exploratory Analysis.” National Bureau of Economic Research, Cambridge. May.
- Folland, Sherman, and Miron Stano. 1990. “Small Area Variations: A Critical Review of Propositions, Methods, and Evidence.” *Medical Care Review* 47 (Winter): 419–65.
- Goodyear, Michael D. E., and Michael A. Fraumeni. 1996. “Incorporating Quality of Life Assessment into Clinical Cancer Trials.” In Spilker (1996).
- Padilla, Geraldine V., and others. 1996. “Quality of Life: Cancer.” In Spilker (1996).
- Spilker, Bert, ed. 1996. *Quality of Life and Pharmacoeconomics in Clinical Trials*. 2d ed. Philadelphia: Lippincott-Raven.
- Triplett, Jack E. 1997. “Human Repair and Car Repair in the National Accounts.” National Bureau of Economic Research, Cambridge.
- Wennberg, John E. 1884. “Dealing with Medical Practice Variations: A Proposal for Action.” *Health Affairs* 3 (2): 6–32.