Health Care Productivity

McKinsey Health Care Practice

McKinsey
Global
Institute

**with assistance from
our Advisory Committee
Ken Arrow
Martin N. Baily
Axel Börsch-Supan
Alan M. Garber**

Los Angeles

October 1996

Contents

Chapter 1: Context, objectives, and approach

In this chapter, we summarize the background and context for this project, its specific objectives, and the project approach and methodology.

BACKGROUND AND CONTEXT

Every major industrialized country treats health care as a unique and important sector, one whose considerable economic significance is heightened by its impact on social welfare. A high fraction of the gross domestic product (GDP), ranging from approximately 7 percent in the U.K. to 14 percent in the U.S., is devoted to health care; and health care expenditures can indirectly influence a country's level of GDP by reducing lost labor productivity from injury and illness. It is highly valued by society at large as well as by individual consumers, arguably above almost all other goods and services in most developed countries. Many citizens and policymakers believe that access to quality health care is a right.

Despite its privileged status as an essential social good, health care production imposes trade-offs like every other good and service; its production uses resources that are scarce relative to competing consumer wants. Every country must therefore have mechanisms – explicit or implicit, market-based or regulatory – for determining how much health care to produce, how to produce it, and how to distribute it across the population.

In recent years, the health care systems of almost all major industrialized countries have come under significant pressure to improve performance, particularly to better manage cost growth. Health care costs and the fraction of GDP they account for are rising rapidly, forcing many countries to begin trimming health care benefits or other social services. With growing numbers of elderly men and women – a demographic shift that increases heavily the proportion of the population using medical care as it decreases the proportion of net taxpayers – pressures on expenditures will continue to build. Continued medical innovation, ranging from incremental improvements in existing imaging technologies to the products of dramatic advances in molecular biology, are likely to increase the scope of effective but costly medical care. Few countries will be able to maintain their economic vitality while supporting even their current level of health care benefits unless they improve the economic performance of their health care systems.

Productivity is a critical determinant of health care system performance and is relevant to every nation. Productivity in health care 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. That is, 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 some of the spending pressure on their systems or may avoid making difficult *allocative* choices to reduce or redistribute benefits. Thus, interest in the level and the causes of productivity in health care systems is growing on the part of policymakers and other health care system stakeholders as per capita expenditure levels are rising.

International comparisons of health care system performance have usually focused on aggregate (or macro) analysis of health care expenditures and access, and have not disaggregated performance into productivity and other key performance drivers. Although such comparisons often include health outcomes, these data at the international level are usually crude and limited to measures like mortality rates and life expectancy. The findings of this body of research therefore pose many new questions: *What are the sources of the differences in spending among countries? Why do those differences appear to be unrelated to differences in overall life expectancy?*

For example, **Exhibits 1 through 4** show spending levels for the U.S., the U.K., and Germany as well as trends in health care expenditures as a percent of GDP, in dollars per capita, and in local currency real per capita spending; all demonstrate significant and growing differences in overall spending levels. **Exhibits 5 and 6** show comparative life expectancy and mortality data for selected conditions for these three countries. These data suggest that life expectancy is similar across the countries and mortality rates exhibit unclear and confusing patterns across selected diseases, but health expenditures vary widely.[1]

International comparisons of health care system performance conducted at the aggregate level usually cannot address the questions posed above because of two major limitations:

¶ **Spending levels.** Most aggregate analyses do not isolate differences in care input levels from relative input price levels, nor do they distinguish the direct medical inputs (such as physician time, hospital beds, and pharmaceuticals) used in disease treatment from

---

[1] Life expectancy in the U.S. is similar to that of Germany and the U.K. when the effects of infant mortality are removed. This adjusted measure of life expectancy is the most appropriate relative outcome measure because the inclusion of infant mortality effects may bias the results. Definitions of health status at birth vary widely between the U.S., Germany, and the U.K. For example, infants that may be considered stillborn (and thus do not contribute to infant mortality) in Germany and the U.K. may be included in the U.S. statistics, resulting in higher infant mortality and lower life expectancy at birth in the U.S. Thus, to produce comparable results for life expectancy, the effects of infant mortality are removed.

the administrative inputs consumed in managing and regulating the health care system. In many industrialized nations, health care is provided by the government sector or the prices of health care products and inputs are either administered or regulated. Under these circumstances, price data do not have the usual interpretation of opportunity costs. The combination of nonmarket pricing and failure to disaggregate renders macro results difficult to interpret.

¶ **Health outcomes.** Most aggregate-level analyses either do not assess outcomes at all or express outcomes in terms of units of service (e.g., physician visits, hospital-days). Implicit in this approach is an assumption that either the units of service are the products that are valued, or that units of service bear a direct and clear relationship to the health outcomes that patients and society value. Tracing the links from health services to health outcomes is difficult under any circumstances, since it is difficult to separate health effects of treatment from the influence of lifestyle, socioeconomic, or environmental factors. In addition, outcome measurement problems reflect the heterogeneity of people –whose risk factors and severity of disease may vary across nations – and the heterogeneity of health care itself. Furthermore, health care encompasses the prevention, diagnosis, and treatment of many diseases and medical conditions and is delivered in numerous care settings by varied providers who adopt different treatment approaches.

Given these limitations, prior aggregate-level research has revealed an apparent paradox: variation in health outcomes does not correspond to variation in per capita health expenditures. This paradox suggests that there is substantial variation in the productivity of health care. But aggregate comparisons are subject to the limitations of the data they use, and have neither convincingly demonstrated that the apparent productivity differences are real, nor pointed to strategies that policymakers and health care organizations around the globe can adopt to improve the economic performance of their health care systems.

Nevertheless, health care policymakers and other stakeholders are aggressively pursuing initiatives to reform their country's health care systems, given intensifying performance pressure. Efforts are numerous and varied, ranging from central planning and direct regulation of supply to the infusion of more market-based approaches. While many recognize the need for some mixture of regulation and market mechanisms, countries and systems have arrived at very different blends. And no one system is recognized as having the most productive system or as having achieved the right blend. Nevertheless, health care policymakers and organizations around the globe are now asking similar questions, including:

¶ Can market-based approaches work well in health care, given the high propensity for market failure in certain areas (in contrast to the

view that market failures in health care necessitate extensive regulation or even nationalization)?

¶ What specific market structures work best in health care coverage and care provision? Specifically, do more competition and greater "integration" in care provision or payment improve productivity?

¶ Do economic incentives influence the behaviors of physicians, hospitals, and other providers and, if so, which incentives work best?

¶ Are limits on aggregate spending and/or hospital and physician supply needed to avoid overconsumption of resources? If so, where and how should these be established and enforced?

PROJECT OBJECTIVES

In order to better understand differences in health care system performance and address the questions posed above, this joint project between McKinsey's Global Institute and McKinsey's Health Care Practice had three major objectives:

1. Assess differences in relative productivity at the *disease level* among the health care systems of three major industrialized countries – the U.S., Germany, and the U.K.

2. Examine the major sources and drivers of these differences in terms of variations in health care treatment approaches and in the underlying provider incentives and supply constraints that arise from the structure and regulations of each country's health care system.

3. Based on this examination, as well as on available aggregate-level analyses, identify implications for policymakers and health care organizations around the globe in their ongoing search for performance improvement.

Our focus was on productivity, not on the overall performance of the health care system. We did not seek to assess the allocative efficiency of each country's health care delivery, which would require difficult value judgments about such issues as the optimal level and distribution of health care spending. 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. Although health policy decisions cannot be made on the basis of productivity alone, a system must be productive to be economically efficient; failure to achieve productive efficiency means that there are ways to produce more health from the same amount of resources.

This project represents an attempt to move beyond prior international health care comparisons by combining aggregate-level analyses with disease-level productivity analyses. It was motivated by the belief that disease-level productivity analyses could provide useful, novel insights into the causes of variations in both health expenditures and outcomes at the aggregate level, and that it would reveal potential strategies that policymakers, providers, payors, and other interested parties in each country could adopt to improve health care productivity and overall performance.

PROJECT APPROACH AND METHODOLOGY

We sought to achieve the project objectives in four ways, by:

¶   Assessing the relative productivity of the three health care systems by comparing the health benefits achieved relative to the physical inputs used in disease treatment. We consider disease treatment to be analogous to a production process in which the output is a set of specific health outcomes, such as lower mortality.

¶   Examining productivity differences in the treatment of four specific diseases – breast cancer, lung cancer, diabetes, and cholelithiasis (gallstones). Each disease is common, costly, and causes substantial mortality and/or morbidity. In addition, there are several approaches to treating each disease, leading to international variation in treatment patterns.

¶   Defining the product of each disease treatment process as the health status (outcome) achieved in the patient population, selecting the most appropriate available measures and timeframes for each disease (e.g., 5-year survival rates for lung cancer).

¶   Defining and aggregating the care inputs used in the production process for each disease in terms of the physical "activity-based" units of labor, capital, and supplies (e.g., number of physician and nursing hours, doses of pharmaceuticals) rather than as monetary expenditure levels.

For each disease and each country, we measured resources consumed and health outcomes achieved to assess the relative productivity of the three health care systems. We then analyzed the underlying causes of these differences by characterizing treatment patterns and provider behaviors in each country; we linked these different treatment patterns to the incentives and constraints acting on providers as well as to the structural characteristics of each health care system. We also assessed the impact of regulation, which shapes system structure as well as supply constraints. (See **Exhibit 7** for a summary of our causality framework.) Finally, our data came from the late 1980s; health care

markets and delivery systems in each country have changed since then. Our discussion, therefore, also addresses these changes and the implications of the analysis for present-day health care systems.

Because we focused on productivity, this study does not address the impact of differences in access to health care or in other socioeconomic factors across the three countries. Similarly, it does not address directly the drivers of administrative costs, a potentially important cause of international variation in overall health expenditures.

Below we describe in more detail the approaches used to assess relative productivity and to examine the major sources and drivers of observed differences.

Assessing relative productivity

The lack of an output measure that is both meaningful and easily quantified makes it more difficult to assess relative productivity in health care than in other industries. The desired product of health care is improved "health" rather than units of service. The treatment process itself is complex, and health outcomes are strongly influenced by patient characteristics; it is therefore difficult to isolate the contribution of health care to health outcomes. We can directly measure levels of inputs used in each country, along with disease outcomes, but without further assumptions these numbers are insufficient to calculate relative productivity or to draw conclusions about the contribution of health care to health outcomes. Thus, we cannot measure relative productivity directly, but can only estimate whether one country is more productive relative to another: this is the economic concept of "productive efficiency." This term is therefore used throughout the remainder of this report.

Our methodology for estimating relative productive efficiency involved three major steps: 1) estimating per-case inputs used in each country; 2) estimating per-case outcomes in each country; and 3) comparing differences in input and outcome levels to assess relative productive efficiency.

1. Estimating inputs used. To estimate the inputs used, we 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, 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 possible). 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 (LOS) for treating this disease (see Appendix 1B). Because the units of measurement for each input vary, we standardized inputs 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. Appendices 1A through 1C provide more detail on our input methodology.

2. Estimating outcomes. We applied outcome measures pertinent to each disease and adjusted for differences in disease incidence across countries. Like the input measures, outcome measures were derived from literature reviews, database analyses, and clinical expert interviews.

An ideal comparative health outcome measure would assess the difference between health outcomes of otherwise identical individuals treated in different countries. That is, such a measure would not be confounded by differences in the severity or incidence of disease in the two countries and would only reflect differences in the effectiveness of treatment. One way to derive such a measure is to compare the expected 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 **Exhibit 8**. Since the outcome represents a change in health status, it is necessary to quantify health status expected for each disease as well as to determine the improvement in health that results from the disease treatment process.

¶ **Quantifying health.** Outcomes for each disease can be quantified using either survival rates or calculations modeling the quality of life. Survival rates, which are easily assessed, are appropriate measures for lung cancer and breast cancer, in which the primary goal of treatment is to reduce mortality. Outcomes for the cancers can thus be measured as years of life expectancy or life years (LYs). For diabetes and cholelithiasis, 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 quantified quality of life outcomes with the widely used Kaplan-Bush Index of Well-Being and applied it to calculate outcomes in quality adjusted life years (QALYs). While quality of life is also relevant in the cancers, it is quite difficult to measure with available data and is less relevant than in diabetes and cholelithiasis. Details on our outcome methodology

are described in Appendix 1D and in the individual disease chapters and their associated appendices (Chapters 3 through 6).

¶ **Measuring improvement in health from treatment.** Quantifying the effects of treatment on health outcomes is inherently difficult. Outcomes without treatment are usually unknown and can be influenced by the patient's baseline health status, which 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. Available data support this assumption.

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

3. Determining levels of productive efficiency (**Exhibit 9**). If we knew all of the input and outcome combinations of each country's treatment process – in other words, the country's entire production function as illustrated in **Exhibit 9** – productive efficiency could simply be assessed by observing the position of the production function: the higher the function, the more productive. However, the data available to us gave essentially one point – not an entire production function – for each country in the treatment of each disease, consisting of the average input level and average health outcome. Thus, we can only use the positions of two points to infer whether two countries were on the same production function in the treatment of a specific disease, or whether one country's treatment process was more productive.

The simplest case is illustrated by a comparison of Countries A and B in **Exhibit 9**: Country A achieves better outcomes while using fewer inputs, so Country A must be more productive. Countries A and C in **Exhibit 9** depict the more common situation, in which one country uses more resources and has better outcomes than another. In this case, knowledge about the disease treatment process itself is required: Country A is more productive than Country C if *the production function does not exhibit increasing returns* and Country A has greater *average productivity* than Country C.

A production process does not exhibit increasing, but rather diminishing (marginal) returns when the production function is shaped like either of the curves displayed. The key property is that when more inputs are used, the output of health increases, but the incremental increase in health from each incremental increase in resources diminishes with the level of resources (i.e., the slope of the curve diminishes as the level of resources increases). Because

the production process is defined at the per-case level, it is likely that diminishing returns characterize the treatment of each of the diseases studied in the three developed countries examined for two reasons:

¶ **Patients who are most likely to benefit are the first to be treated.** A rational allocation of resources first assigns treatment to those patients who are most likely to be cured or otherwise benefit from intervention. As treatment extends to patients with lower chances of cure or lesser benefit, inputs increase proportionately, but successive gains in outcomes decline. For example, screening programs to achieve early detection will have the greatest returns when applied to patients with the greatest *a priori* risks of having the disease. As the target population for screening is expanded to include patients with lower risks, inputs increase proportionately, but the additional benefits from early detection diminish.

¶ **The most cost-effective technologies are the first to be used.** For some diseases, a range of therapeutic or diagnostic technologies may be available. Extending treatment may require the addition of less effective or more expensive technologies, leading to diminishing returns.

Note that Countries A and C can only lie on the same production function if production exhibits increasing marginal returns. The dashed lines drawn from the origin to the points for each country give us the additional information needed to conclude that Country A is more productive than Country C. The slope of these dashed lines, or the simple ratio of health output to level of resources, represents the average productivity of the country in the treatment of the disease. When production does not exhibit increasing returns and the average productivity of the country that uses more inputs (Country A) exceeds the average productivity of the country that uses less inputs (Country C), productivity in the former country must exceed that of the latter.[2]

However, if the country with higher inputs and outcomes has lower average productivity for a disease (as shown in the comparison of Country C versus Country D), then its productive efficiency relative to the country with lower inputs and outcomes is indeterminate without detailed knowledge of the production function. Its lower average productivity may reflect either lower overall productive efficiency or a rational choice to operate at a portion of the disease treatment (production) function with small marginal returns to

---

[2] Any curve that satisfies the properties of a production function that does not exhibit increasing returns and that passes through the origin – and that passes through the point for Country C – must lie below a curve with the same properties that passes through the point for Country A. Note that the origin need not be limited to a point with zero input and zero outcome; this property will hold for any input and output combination bounded between zero and the input/output combination of the country with the lowest input level.

additional inputs in order to achieve better outcomes. For this comparison, only more detailed knowledge of the production function (i.e., some sense of the shape of the curve) can allow us to determine which country is more productive.

For those case comparisons that do not meet any of the productive efficiency tests described above (as shown in the comparison of Country B versus Country C), we cannot determine which country is more productive. Two countries could be producing health care using the same production function, but the one that spends more will have lower average productivity simply because it is operating at an area of rapidly diminishing marginal productivity. Country B uses more inputs than Country C because it allocates more funds to the treatment of the disease, not because it is less productive. Even if the two countries are equally productive in the sense that they use the same production function, however, it may be possible to infer that the country that spends more is unlikely to be obtaining reasonable value for its health expenditures. Cost-effectiveness analysis offers perhaps the most widely used method for determining whether the benefits of a health intervention justify its costs. Below we summarize this assessment methodology.

Assessing cost-effectiveness

As shown in **Exhibit 10**, if the additional outcome is worth more than the additional inputs needed to achieve it (i.e., marginal benefits exceed marginal costs), then the combination of better outcomes and greater inputs is preferred. The converse is also true; if the additional inputs are worth more than the improvement in health outcomes, the combination of lower outcomes and lower inputs is preferable.

Cost-effectiveness analysis does not itself determine what an outcome – an additional LY or QALY – is worth. However, the cost-effectiveness ratios (defined as a ratio in which the numerator is the difference in costs between the intervention under study and another treatment, while the denominator is the difference in outcomes for the intervention and the alternative) of commonly accepted interventions can be used as benchmarks against which the cost-effectiveness ratio of an intervention under study can be compared.

In our analysis, we used the U.S., the U.K., or German prices to value the input units as appropriate for the pairwise country comparison (i.e., both the U.S. and the U.K. prices were used to check the ratio when comparing the U.S. and the U.K. results). Since the U.S. prices were higher than the U.K. and German prices, ratios using the U.S. prices will always yield a cost per QALY estimate higher than those using the U.K. or German prices; these calculations thus provide a more stringent test of additional U.S. spending. If the ratio of additional inputs to additional outcomes is very low, then the country with higher outcomes and higher inputs likely has the preferred outcome/input combination in its treatment approach. Conversely, if the cost per QALY ratio is very high, then the

process with better outcomes likely represents a bad bargain, and the lower cost, lower outcome treatment approach is likely preferred.

Prior studies have produced some rough benchmarks for what is a "very high" or "very low" cost-effectiveness ratio in cost per QALY. Generally speaking, these studies have claimed that health care interventions in the U.S. that cost less than about $30,000 (in 1990 U.S. dollars) per QALY can be considered reasonably cost-effective, while those that cost more than about $100,000 per QALY are questionable. Between $30,000 per QALY and $100,000 per QALY, opinions can vary widely on whether an intervention is cost-effective. (More detail on cost-effectiveness benchmarks is provided in Appendix 1E.)

Examining major sources and drivers of observed productive efficiency differences

After we assessed productive efficiency (and cost-effectiveness where appropriate) of the three countries, we sought to understand why productive efficiency differed.

First, we identified and quantified the most significant variations in health care treatment approaches (i.e., provider care choices and behavior). While we recognized that significant variations are possible within each country, we focused on across-country variation, attempting to measure the "average" care delivery approach for each disease in each country.

Then, we determined the most salient differences in provider incentives and constraints that appeared to drive these treatment variations, and identified the specific structural characteristics of each system that appeared to shape these incentives and constraints (see again **Exhibit 7**). Of critical importance was understanding the specific nature of the major economic interactions among the health care system participants – particularly in the health coverage and care provision markets. In each of these markets, we examined the nature of the products and services that were being exchanged, focusing on the level of integration in the health care products bought and sold and the degree of competitive intensity associated with this exchange. We paid particular attention to the impact of regulation on these market structures.

Recognizing that health care system structure and the resulting incentives and constraints for providers vary within each country, we attempted to summarize the average existing in each country at the time of our assessment and to draw conclusions from this summary.

* * *

The next chapter (Chapter 2) provides more detail on our framework for health care system structure and uses this framework to describe the three predominant systems included in this study (the U.S., the U.K., and Germany) at the time of our assessment (the mid to late 1980s). The impact of system structure and different provider incentives and constraints on observed productive efficiency differences is discussed in each of the disease case chapters (Chapters 3 through 6) as well as in the cross-disease synthesis (Chapter 7).

Appendix 1A: Input measurement methodology

Countries use many types of inputs in the production processes for disease treatment. For example, treatment of a single disease may require labor from physicians, nurses, and technicians; capital inputs in the form of facilities, diagnostic equipment, and therapeutic equipment; and supplies such as pharmaceuticals and disposables. Different countries use these various inputs in different quantities and in different proportions, depending on the specific production process each employs to treat a disease.

Our comparison of the inputs used in a disease treatment process across countries uses a cost-function approach that measures or estimates the *individual physical inputs* used per case in each country to treat the disease, then multiplies the inputs by price weights to derive the total input measure. These inputs comprise physician-hours, nurse-hours, and other labor (e.g., we determined how many physician-hours were used to treat an average lung cancer case in the U.K.); capital inputs; and supplies. Data availability sometimes limited our ability to measure certain inputs directly, in which case we employed various estimation techniques. Some of these estimates were disease-specific; these are described in the chapters documenting the individual disease analyses. One general issue we faced in all diseases was the need to estimate labor inputs for hospital stays, and capital and supplies inputs from indirect information on the disease treatment process. Our approach to these estimates is described in Appendix 1B and Appendix 1C, respectively.

We then combined these different individual physical units into a *single* measure of resource usage in order to compare overall productive efficiency across countries. To do so, we developed an *aggregate measure of total resources used* per case by each country. This aggregate measure is described below, followed by a discussion of the estimation of standard input prices.

AGGREGATE MEASURE OF TOTAL RESOURCE USAGE

The necessity of developing an aggregate measure of total resource usage as well as the problems in developing such a measure are best shown using an example. If the U.K. uses 2 physician-hours and

4 nurse-hours to treat a particular disease, and Germany uses
1 physician-hour and 8 nurse-hours, which country uses more labor
resources? In developing a suitable aggregate measure of inputs,
we had three major objectives:

1. *Economic validity.* The measure should be consistent with basic
   economic principles. For example, it should increase if any single
   input increases; and it should be zero if, and only if, an *essential*
   input is zero.

2. *Sensitivity to prices.* Although it should measure real usage of
   physical inputs rather than monetary values, the measure should
   take into account differences in factor prices within a country; for
   instance, if a nurse costs less than a physician, inputs of physician
   time should be counted more heavily than inputs of nurse time. At
   the same time, the measure should permit its computation under any
   given set of prices; for instance, it should permit the "correction for"
   intercountry price differences.

3. *Ability to analyze sources of difference.* The measure should allow us to
   explain how differences in each individual input contribute to
   differences in aggregate resource usage across countries.

There is no unique aggregate measure of resource input that will fit all
possible situations. Unless the process of combining the inputs to the health
outcome is exactly known, all aggregate measures of resource input are,
therefore, approximations. In our approach, we used a weighted sum of the
physical input quantities, where the weights reflect average relative factor
prices in the countries studied. This approach meets the objectives outlined
above, as it: 1) is economically sound because it is based on costs, and the
prices used reflect real relative input prices; 2) takes local factor prices into
account by weighting inputs on the basis of relative factor prices and corrects
for intercountry price differences by applying a standard set of prices to inputs
in each country; and 3) simplifies analysis of sources of difference in the input
measure across countries since differences in total cost depend linearly on
differences in physical inputs.

ESTIMATING STANDARD PRICES OF INPUTS

To estimate a price-weighted index of inputs to health care in our approach, we
needed to determine the standard prices of inputs. We describe here our general
approach to modeling inputs using labor inputs as examples, and describe in
Appendix 1C how the approach was modified for capital and supplies.

A natural starting point is the actual prices in each country. Suppose that input
usages and country factor prices in the U.K., Germany, and the U.S. given in

Exhibit A-1 are used. Although it is straightforward to determine total costs of treatment in each country and the relative price ratio between physician-hours and nurse-hours, it is not clear which country uses "more resources." First, total costs in different countries are expressed in local currency units (£, DM, and $) and are, therefore, not directly comparable. Second, the relative price ratios between two inputs may vary across countries.

We explored four approaches:

¶   Conversion by the exchange rate

¶   Conversion by a countrywide purchasing power parity (PPP) ratio

¶   Conversion by a health care sector-specific PPP ratio

¶   Conversion by a standardized set of relative factor prices.

The first three approaches have the advantage of expressing all costs in the currency of a specific base country, but also have a significant disadvantage: commonly used countrywide currency conversion factors such as the exchange rate or PPP ratios do not adequately reflect intercountry differences in health care input factor prices. Reliable health care sector-specific PPP ratios – although theoretically superior as a basis for conversion – are not available.

Difficulty with exchange rate

As a basis for comparing the resources used to produce health care or any other consumption or investment goods, exchange rates are subject to short-run distortions arising from a variety of financial signals. These financial signals reflect fluctuations in expectations about employment levels, interest rates, the conditions of financial markets, and numerous other macroeconomic conditions that may influence future exchange rates, but may have little relevance to relative prices in the health care sector.

Difficulty with countrywide PPP ratios

Although the PPP ratio (e.g., the GDP PPP, or the household expenditure PPP, both published by the OECD) is less likely to be influenced by short-term fluctuations in macroeconomic conditions, it is subject to distortion by price regulation. Because the prices of many health care inputs are determined by government regulation rather than by market forces, the prices of health care inputs relative to other goods in an economy may vary significantly across countries, and this variation may reflect differences in regulated prices rather than in the resources actually used. A PPP-based currency conversion may therefore distort our measurement of the real resources used in health care production processes.

An example will illustrate this point. Suppose, for simplicity, the following differences between the U.S. and the U.K. economies:

¶   The U.S. prices (in dollars) of all goods are twice the U.K. prices (in pounds) of the same goods, except for health care inputs.

¶   Through government regulation, the U.K. sets prices (in pounds) for all health care inputs at only one-fourth of the U.S. prices (in dollars) for the same inputs.

¶   True resource usage for any disease treatment process is identical between the two countries. Thus, for example, if the U.S. uses 1 physician-hour, 3 nurse-hours, and 100 units of radiation to treat a particular disease, the U.K. uses the identical levels of inputs.

Under this scenario, a PPP-based comparison of health care costs would result in the following:

¶   The PPP currency conversion ratio would be $2 per £1. (This assumes that health care costs were not included in the PPP calculation. If they were included, the PPP ratio would be slightly higher, with the exact difference depending on how heavily health care costs were weighted in the PPP calculation.)

¶   Local currency costs for each disease treatment process would be four times higher in the U.S. than in the U.K.

¶   After conversion at PPP, adjusted disease treatment costs in the U.S. would be double the costs of the U.K.

This PPP-based comparison, therefore, reflects the lower relative input price levels of the U.K., rather than the true pattern of physical input usage. In terms of "physical productive efficiency," in our scenario the U.S. and the U.K. should be considered equivalent. Because our objective is to measure and compare this physical productive efficiency of disease treatment processes, this approach is misleading.

Limitations of health care sector-specific PPP ratio

In theory, use of a health care sector-specific PPP ratio would allow a more meaningful conversion of health care input costs to a common currency. Such a PPP ratio would compare the cost of a standardized "basket" of health care products and services across countries. For instance, in the above example a health care sector-specific PPP ratio between the U.S. and the U.K. would be $4 per £1; use of this PPP ratio would, therefore, correctly show that resource usages for disease treatment are equal in the two countries.

Unfortunately, although some health care sector-specific PPP ratios exist, reliable health care sector-specific PPP data are not available. In addition, even a health care sector-specific PPP ratio could generate misleading results if relative prices *within* the health care sector vary widely between countries. To see why, consider the following (hypothetical) conditions in the U.S. and the U.K.:

¶ A physician-hour in the U.S. costs $125, and a nurse-hour $25. In the U.K., a physician-hour is £50 and a nurse-hour £25. (Thus, physicians are five times as expensive as nurses in the U.S., but only twice as expensive in the U.K.)

¶ A health care sector-specific PPP ratio is developed using a basket of 1 physician-hour and 1 nurse-hour. This yields a PPP ratio of $150/£75 = 2 $/£.

¶ A particular disease treatment process uses 1 nurse-hour (and no physician time) in both the U.S. and the U.K., at local costs of $25 and £25, respectively.

¶ Using the health care sector-specific PPP ratio, the U.S. dollar cost of the U.K. process is £25 X 2 $/£ = $50; thus the U.K. appears to use twice the resources of the U.S., even though physical inputs for the two countries are identical.

Use of relative factor prices

To avoid the problems that arise from application of currency conversions, we adopted a fundamentally different approach. We compared what total costs *would be* in each country if the factor price ratios for inputs were identical across countries, e.g., if a nurse-hour in the U.K. costs the same fraction of a physician-hour as it does in Germany. These assumed, common *relative factor prices* thus become a set of common weights for the inputs, and the weighted sum of the inputs becomes our measure of aggregate resource usage. By using the same set of relative prices, this measure reflects only differences in input quantities and is not sensitive to currency conversion factors.

While any arbitrary set of relative factor prices could be used to develop an aggregate cost measure, in practice we should use factor prices that reflect the real relative prices observed in the three countries studied. Continuing the example from **Exhibit A-1**, using the relative factor prices given in **Exhibit A-2,** we observe the next problem: no set of common factor prices is technically "correct," since relative prices do differ across the countries. Although it seems reasonable to use a price of nurse-hours (in terms of physician-hours) between 0.36 and 0.50 in this case, there is no "natural" choice of a specific value in this range.

The most intuitive, although arbitrary, approach is to calculate total costs per country using *average factor prices* across the three countries. These average factor prices are calculated by choosing one factor as the unit of measurement ("numeraire"), in our case a physician-hour, and then relating the price of all other factors (e.g., nurse-hours) to the price of a physician-hour. These normalized prices are then averaged across the three countries. In our example, we obtain an average relative price of nurse-hours, which is 0.42 physician-hours. We treat other labor inputs similarly.

These average factor prices have no rigorous economic interpretation. They simply represent a pragmatic set of weights for combining physical inputs into a single overall measure of resource usage.

Continuing our example, we arrive at the total resource costs, measured in terms of physician-hours (**Exhibit A-3**). According to this aggregate measure of total resource usage, Germany spends the most resources (4.36 physician-hours) and the U.K. the least resources (3.68 physician-hours).

Not only may the set of price weights be arbitrary, but the choice of price weights may affect the rankings of the countries. That means that under its own price weights, one country may appear to use fewer resources than another country, while the ranking may reverse with the other country's price weights. This phenomenon can be appreciated by changing relative prices in **Exhibit A-2**. Suppose that nursing hours were significantly less costly. This would reduce Germany's total resource costs more than the other countries' resource costs because Germany uses nurse-hours relatively more than the other two countries. For instance, if the nurse-hour price is changed from 0.5 to 0.2 physician-hours, Germany appears to use fewer resources than the U.S., and even fewer resources than the U.K. (**Exhibit A-4**).

If switching the factor prices from the U.K. prices to the German prices also caused the ranking of the U.K. and German total costs to switch, we would not be able to conclude from this data whether either system was more "efficient" in its use of inputs.

In order to address this sensitivity to the choice of a set of relative factor prices, we compute our aggregate measure of resource usage not only using the average factor price but also using all three sets of relative factor prices actually observed in the three countries. Only if the ranking is the same in all four comparisons can we conclude without ambiguity that one country uses fewer resources than another one.

Using each country's actual factor prices has an additional advantage because the results have a clear and natural interpretation: they reflect what that country could achieve in total costs if it used the input quantities from other countries for its disease treatment processes. These factor prices are also "realistic" since they represent the real prices in a functioning economy.

In our example, we arrive at the following four comparisons given in **Exhibit A-5**. The comparison of total costs using the U.K. factor prices shows Germany to be 17 percent (= 107.7/92.3) more expensive than the U.K. This implies that if the U.K. adopted German levels of inputs for this disease treatment process, the U.K. costs would increase by about 17 percent. Using the German factor prices instead, Germany still appears more expensive, but now by 25 percent (= 111.1/88.9). This implies that Germany's costs are 25-percent higher than they would be if Germany adopted the U.K. levels of inputs for this disease treatment process. Applying the U.S. prices, Germany uses 13-percent (= 106.0/94.0) more resources than the U.K., while under average prices, Germany resource usage is 18-percent (= 108.4/91.5) higher than in the U.K.

Unlike the extreme example in **Exhibit A-4**, all four price systems in **Exhibit A-5** generate the same ranking of the three countries with respect to total resource usage. Hence, we can conclude definitively that the U.K. disease treatment process is the least, and the German treatment the most expensive. However, the relative magnitude of the difference in costs among the countries depends on which factor prices we use for the comparison. For this example, the differences are close – between 13 percent and 25 percent. *This will be true in general if the relative prices of different inputs are similar across countries and not as extreme as in* **Exhibit A-4**. Note that if relative prices for inputs were the same in all countries, use of any country's prices for the common factor prices would give identical results for relative total costs.

In turn, *if relative factor prices are as different as in the example of* **Exhibit A-4**, *we would expect to see a rank reversal.* If each country is operating its disease treatment processes at minimum cost (for a given level of output), it will tend to use more of the inputs with low relative prices, and fewer of the inputs with high relative prices, assuming that there is some degree of substitutability among inputs. Each country should therefore have an input mix that is tailored to its own relative prices for lowest cost production. In the extreme example above, Germany uses a cost-effective mix of more nurse-hours and fewer physician-hours than the U.K., because Germany's nurse price is only one-fifth its physician price (compared to two-fifths for the U.K.).

If a country does not have the lowest cost using its own relative prices, then it has a (theoretical) opportunity to lower its costs using an input mix from another country. Our disease case studies demonstrate that countries have not always taken advantage of these opportunities, since some countries clearly have highest cost input mixes even under their own price system. Several factors could explain why a country maintains a higher cost input mix, for instance:

> ¶ The higher cost position is associated with superior outcomes. The country that spends more gets more. As our case studies will show, higher costs are not always associated with better outcomes; this rationale is, therefore, not always valid.

- ¶ Political pressures or constraints preclude adoption of a lower cost production process.

- ¶ The health care production process is organized to achieve social goals – such as income redistribution – rather than to maximize the achievement of health outcomes from a given resource allocation.

Summary of cost comparisons

To summarize, the possibility that differences in factor prices across countries can affect the relative ranking of total costs implies that we need to compare costs under each country's factor prices in addition to the average factor prices.

If comparisons under all three price systems show a consistent rank ordering of countries' total costs, we can conclude definitively that there are observable differences in resource usage. If there are reversals in countries' rankings, then we have to view their resource usage as indistinguishable given our methodology.

We report cost comparisons under *average factor prices* as our point estimate of overall resource usage. Since the comparisons under individual country factor prices are used to check consistency of results, these comparisons are only reported if they indicate reversals that prevent us from making definitive conclusions on the rank ordering of resource usage across countries (**Exhibit A-6**).

A consistency check was conducted for each of the four diseases. The results indicated that the rank ordering of resource usage across the countries was the same for each country's set of factor prices.

Appendix 1B: Estimating labor inputs for hospital stays

Although we modeled many different events in the calculation of total physical inputs, one of the common components across the four diseases was the hospital stay associated with an inpatient procedure. In addition to the specific inputs for the procedure itself (such as surgeon time for a surgery), a hospital stay consumes many resources while a patient is being prepared for, or is recovering from, the procedure. In all diseases, the hospital stay was a major component of total input use, so it was important to follow a consistent approach to its measurement within the confines of data limitations.

LABOR PER HOSPITAL STAY

To estimate the labor inputs associated with a hospital stay, we used a simplified model based on hospital staffing ratios in each country. (Appendix 1C details the methodology for estimating capital and supplies inputs.) The staffing levels used in the input calculations represent the average hospital staffing levels across all diseases and during the entire LOS. As illustrated in **Exhibit B-1**, staffing levels are expected to be greater for certain diseases and during the initial days of a hospital stay, when more intensive care is delivered. We assumed that the average staffing level per bed-day was the same for all diseases and all days of an inpatient stay. Although this assumption was necessary because more detailed staffing information (by disease and day of hospitalization) is not available at the national level, it is likely to be approximately correct, particularly since an average level is used for each of the countries.

Therefore, our estimate of the total labor input for a hospital stay was the average staffing level per bed-day for the country multiplied by the LOS for the disease in that country. The LOS estimates were disease-specific, but the staffing ratios were common across diseases.

STAFFING LEVELS

Per bed-day

We determined the average staffing levels per bed-day for four personnel categories (physicians, qualified nurses, nonqualified nurses, and medical technicians) from aggregate data on labor levels and bed-day utilization in each country. As illustrated in **Exhibit B-2,** the U.S. exhibited the highest staffing intensity for each personnel category. Germany exhibited the lowest staffing intensity per bed-day except in the case of physicians, where the number of physicians per bed-day was slightly greater than that of the U.K.

These labor inputs were weighted by their relative salaries to calculate the total labor input per bed-day (as described in Appendix 1A). With these weights, the U.S. had the highest total staffing level intensity with 0.76 standardized input units, followed by the U.K. with 0.55 standardized input units, and Germany with 0.47 standardized input units (**Exhibit B-3**).

Relationship between staffing levels and LOS

Although staffing levels per bed-day were multiplied by LOS to determine labor inputs, this simple approach obscures the fact that these two factors are not independent. Staffing levels may well rise when policies are implemented to reduce LOS. Compressing hospital care into fewer days may require that the intensity of care be higher during those days; thus, LOS and staffing levels per bed-day may be inversely related to some extent. For example, in the U.S. shorter LOS may have required higher staffing levels per bed-day in order to facilitate patient throughput. On the other hand, longer LOS in Germany may have allowed lower staffing intensity per bed-day to provide the same service level to patients.

The relevant measure of relative labor usage among countries is the total labor used for an inpatient stay – i.e., the product of LOS and staffing levels per bed-day. It is this product that we calculate and compare in our disease case models to assess relative labor inputs for inpatient stays.

Although we discuss these input differences in detail in the disease case chapters, we can begin to develop some insight into the relationship between LOS and labor usage here. In order to disaggregate the effects of staffing levels per bed-day and LOS, we require a different way of understanding staffing levels. One such approach is to compare average staffing per inpatient stay in each country. These labor levels *per admission* are simply the product of average staffing levels per bed-day and average LOS across all diseases for each country.

As illustrated in **Exhibit B**-**4**, labor levels per average hospital admission in Germany (5.7 standardized input units) were slightly greater than those in the U.S. (5.5 standardized input units), while staffing levels in the U.K. were significantly lower (3.5 standardized input units). Germany and the U.S. had similar staffing levels per hospital admission, but Germany had higher physician staffing levels and the U.S. had higher staffing levels for all other hospital personnel.

This implies that although LOS and staffing per bed-day are indeed inversely related, there are other differences among countries driving differences in hospital staffing levels (because the products of LOS and staffing ratios are not equal). We discuss potential drivers for these staffing differences in the disease case chapters.

Appendix 1C: Estimating capital and supplies inputs

Our overall approach for modeling inputs is described in Appendix 1A. In this appendix, we describe the modifications to our approach in modeling capital and supplies.

The large number and variety of capital inputs and supplies made it infeasible to measure the specific quantities of all such items at a detailed level. Hospital care, for example, can include use of a wide variety of supplies (such as drugs, blood, plasma, various solutions, bandages, gauze, sutures, surgical instruments, and X-ray film) and general supplies (such as sheets, food, and office items). The types of capital used during hospitalization include diagnostic equipment, laboratory equipment, patient monitors, specialized surgical tools, computers, and the facility itself.

USE OF PROXY VARIABLES

Rather than attempting to measure each of these specific inputs, we approximated the capital and supplies inputs used by modeling capital/supply consumption on a per-service event basis. The number of these events in treating a disease in each country then served as proxy variables for capital and supply inputs. For example, each occurrence of an X ray drives usage of radiographic film, so the number of X rays taken was one of our proxies for supply usage. Each day of stay in a hospital also drives consumption of a number of supplies (such as IV fluids, bandages, and food), so the LOS was also one of our proxies for supply usage.

The complete list of proxies used for capital and supplies by disease case is given in **Exhibit C-1**.[3] By using these proxy variables, we implicitly assumed that the amount of capital and supplies consumed per event (e.g., per computerized tomography [CT] scan) was the same across all three countries, and that the variance in capital and supplies usage was due to variance in the number of these resource-consuming events. Our methodology, therefore, may not capture some of the differences among countries in the intensity of capital and supplies usage. For example, if the U.S. used more sophisticated CT scanners than the U.K., then the amount of capital consumed by each scan in the U.S. would likely be greater

---

[3]   We did not estimate capital or supplies inputs for diabetes due to data limitations.

than in the U.K. Our methodology does not capture this difference – it only reflects differences between the two countries in the number of scans performed.

RELATIVE PRICES FOR PROXY VARIABLES

In order to include the capital and supplies proxy variables in our overall input measure, we had to assess their relative prices. We adopted a simplified approach to estimating relative price, involving three steps: 1) estimating the percentage of total health care spending attributable to each variable, based on aggregate data where available; 2) deriving relative price weights for each disease that would generate the same proportions of spending in each category as these overall percentages; and 3) testing the sensitivity of our results to the specific price weights.

Step 1: estimating the percentage of total
health care spending attributable to each variable

We used aggregate data on health care spending in various categories to assign a "percentage of total spending" to each proxy variable. We made this assignment based on aggregate data on health care spending in various categories. First we allocated 20 percent of total spending to supply variables as a whole, and 10 percent to capital variables as a whole. This allocation was based on information on the total inpatient spending on supplies and capital in each country's health care system overall, which showed that in each country about 70 percent of costs were attributable to labor, 20 percent to supplies, and 10 percent to capital (**Exhibit C-2**).[4]

Within capital, we allocated the 10 percent to two groups of variables – *facilities* costs and *equipment* costs. (For example, increased LOS led to increased facilities costs, while a CT scan would appear as an equipment cost.) We used a breakdown of capital spending into facilities and equipment of 61.3 percent for facilities and 38.7 percent for equipment, which was based on the percentages of capital spending (interest and depreciation) allocated to these categories in U.S. hospitals in 1990.[5]

Within supplies, we allocated half of the 20 percent to pharmaceuticals and half to other supplies (such as laundry) used throughout hospitalization. Our results were insensitive to this division of the supply spending.

---

[4] We also verified that our indexed inputs were not significantly affected by this division, provided that labor accounted for at least 50 percent of the total inputs. This requirement is easily satisfied.

[5] Source: Health Care Financing Administration (HCFA). Facilities costs include permanent buildings and fixtures, while equipment costs include major movable capital equipment.

These allocations created four groups of proxy variables with the following cost percentages:

- ¶ Capital – facilities: 6.1 percent
- ¶ Capital – equipment: 3.9 percent
- ¶ Supplies – pharmaceuticals: 10 percent
- ¶ Supplies – other: 10 percent.

Finally, within each of these categories, we allocated the cost percentages *equally* among the variables. For example, if we had three subcategories of capital-facilities costs, each subcategory would be allocated 2 percent of total costs. Although the assumption of equal contributions to total costs within each category may be oversimplified, this approach captures the impact of each proxy variable while scaling the total contribution of the category to the appropriate percentage of total costs.

Step 2: deriving relative price weights
from percentage of spending allocations

Given the percentage of spending allocations to each proxy variable from Step 1 above, we calculated relative price weights for each variable so that the spending breakdown for each disease (averaged across countries) would match these percentage allocations. This methodology used the total input measure associated with *labor* inputs to establish a baseline from which capital and supplies price weights could be derived.

To illustrate this derivation, suppose that we had assigned 5 percent of total spending to the X-ray proxy variable (which counted the number of X rays). We needed to determine a price weight for X ray that resulted in:

(Price weight) X (X ray) = 5% X (total input cost).

when averaged across all three countries. Given that we had already calculated total labor inputs (as described earlier), our allocation of 10 percent of costs to capital and 20 percent to supplies implies that:

(Labor inputs) = 70% X (total input cost).

This allows us to express the X-ray cost in terms of labor inputs, as follows:

(Price weight) X (X ray) = (5%/70%) X (labor inputs).

We can now solve for the X-ray price weight:

Price weight = (5/70) X (labor inputs)/(X ray).

This derivation uses the *average* of labor inputs and X rays across the three countries for the specific disease being measured.

To illustrate further with a numerical example, suppose that our labor input measurement and X-ray proxy measurement yielded the values given in **Exhibit C-3**. We would then set the price weight of the X-ray variable to be:

X-ray price weight = (5/70) X (7/2) = 0.25 "physician-hour equivalents" per X ray.

This price weight forces the average cost of X rays in the overall input measure to be:

0.25 X 2 = 0.50, which is in the proper ratio of 5:70 relative to the average labor cost of 7.

Because this methodology forces the three-country average total cost of labor, capital, and supplies to be in a 70:10:20 ratio in each disease, it does not reflect possible differences in the relative usage of labor, capital, and supplies *across diseases*. For example, it is possible that treatment of breast cancer uses a higher proportion of supply inputs than treatment of cholelithiasis, possibly because of chemotherapy treatment in breast cancer. Our methodology does not reflect this possible difference. However, our methodology does reflect differences in capital and supplies usage *across countries* within a particular disease, since the proxy variables are estimated for each country's treatment process.

Step 3: testing sensitivity to price weights

Because we made a number of assumptions in deriving relative price weights for the proxy variables, we tested the robustness of our overall input measure to changes in these weights. In all cases, the input measure was not sensitive to even moderate changes in these price weights. This occurred for two reasons: 1) with only 30 percent of total costs allocated to capital and supplies, labor is the dominant driver of input usage; and 2) most of the capital and supplies proxy variables were positively correlated with labor inputs. (For example, the country that used the most labor also consumed the most hospital-days.) The scaling of capital and supplies relative to labor therefore did not greatly affect the relative input comparisons across countries.

ISSUES IN MEASURING CAPITAL USAGE

Measuring the capital resources used in disease treatment presents a number of conceptual difficulties, for example:

¶ Capital equipment and facilities are generally used to treat many diseases, so their costs cannot be uniquely assigned to a specific disease.

¶ The marginal cost of using capital equipment or facilities depends on the current utilization of their capacity. If capital is underutilized, the marginal cost of additional use may be zero; if capital is fully utilized, the marginal cost of additional use may be very high, since it may require capacity expansion.

¶ The economic depreciation associated with additional use depends on physical patterns of wear, which may have little relationship to the accounting measures of depreciation that are recorded as capital costs. (As described below, we used accounting data in our estimate of total capital costs.)

¶ The cost of funds invested in equipment and facilities depends on the capital structure of the firms and agencies making the investments. Health care investments in particular are rife with explicit and implicit subsidies (such as tax exemption for nonprofit hospitals in the U.S.).

Addressing these issues in an economically rigorous way was beyond the scope of our project and the limitations of our data. Our methodology used a simplified approach that estimated the *average accounting cost* associated with capital usage:

¶ Our aggregate measure of capital (the 10 percent used above) was based on the total interest and depreciation charges for hospitals in each country. Although these accounting measures are based on historical costs and do not correct for cost-of-capital subsidies, they are the most comprehensive data available on capital charges.

¶ Our methodology for deriving relative price weights for the capital proxy variables (described above) essentially amounted to a two-step allocation process for these total accounting charges:

1. Total (accounting-based) capital costs were allocated across diseases based on the average labor input (across countries) used in treating each disease.

2. These per-disease capital costs were allocated across the proxy variables for capital usage events using an "average cost per event."

While this methodology did not determine the true economic cost-of-capital usage, it is comparable to typical accounting-based capital measures and is likely to approximate actual capital costs well. Moreover, since capital accounts for

only about 10 percent of the total cost of health care, our simplified estimates of capital cost are unlikely to introduce significant distortions in our overall input measures.

Appendix 1D: Outcome measurement methodology

Health status reflects a number of different factors – including rates of survival and complications as well as levels of patient pain, symptom frequency, and functionality – influencing both the length and quality of life. While quantitative factors such as survival and complication rates may be obtained, qualitative information on factors such as patient pain and symptom frequency are not readily available. Estimates may be made, however, to obtain a measure of "quality of life."

We selected two different measures to estimate health outcomes: LYs for lung cancer and breast cancer, and QALYs for cholelithiasis and diabetes.

LIFE YEARS

Although it does not capture all aspects of health, survival – usually expressed in years of life expectancy – is widely accepted as a fundamental measure of health status. A survival curve, where the percentage of surviving patients are recorded over time following the initial diagnosis, is a popular method for depicting survival (**Exhibit D-1**). The total life expectancy from diagnosis, measured in units of LYs, is equal to the area under the survival curve.

Such survival curves must often be based in part upon extrapolations, because clinical studies track survival for limited time periods. For the disease cases studied here, survival data are generally available for all three countries only for a 5-year period. The expected LYs over the 5-year period, however, serves as only a partial measure of health status, since survival beyond the 5-year period is not captured. Although this truncated view is not a complete measure of overall health status, it provides a relative measure of mortality. The LY estimate obtained using 5-year survival curves was used as our basic outcome for breast cancer and lung cancer, where mortality is the key issue.

Since the 5-year outcomes do not account for health status beyond the 5-year period, we may not capture longer term differences in outcomes. This limitation may be important for breast cancer, in which approximately 60 percent of the patients survive beyond 5 years. It is not as important for lung cancer, since only 10 percent of patients with this condition live for 5 years after diagnosis.

To test the sensitivity of this 5-year truncation of the breast cancer outcomes, a second estimate was obtained using projected 10-year survival curves. The 10-year projections were generated through extrapolation of the 5-year survival curves, since actual survival statistics were unavailable. A number of extrapolations were tested,[6] each yielding similar results. These 10-year approximations were not reported as the main outcomes, but were employed in the cost per LY calculations (described in Chapter 1) to provide a conservative upper bound to the U.S. versus the U.K. outcomes. For these calculations, we used a flat line extension from the
5-year survival rate to approximate 10-year survival. This method underestimates the U.S. advantage over the U.K. in both breast cancer and lung cancer outcomes. Thus, our cost per LY results for 10-year outcomes underestimate the amount by which the U.S. outcomes exceed those of the U.K. in breast cancer, providing a stringent test of productive efficiency of breast cancer management in the U.S. Although it is possible that survival trends reverse after 5 years, available data suggest that this scenario is unlikely.

QUALITY ADJUSTED LIFE YEARS

Although survival is a critical dimension of health status, improvement in quality of life is also an important benefit of health care. Quality of life was the critical issue in our analysis of cholelithiasis and diabetes – cholelithiasis, because gallstones generally cause pain rather than death, and diabetes, because the complications we studied had major effects on quality of life.

Measuring quality of life raises a number of theoretical and practical questions. Individuals would be expected to have widely varying attitudes about the value of life and risks to health, and about the suffering that would result from experiencing pain, disability, or another form of morbidity. However, studies have shown that attitudes toward quality of life and toward conditions that detract from it are more generalizable than might have been expected. Several approaches to measuring the value of life have been employed. These approaches usually survey how a large number of subjects value the ability to perform various daily tasks and how much different types of restrictions (i.e., pain or the inability to walk) detract from life's value. Although these surveys are not exact measurements, they do generate consistent results across a wide range of physical conditions and geographic locations. These results then allow quantitative scales of value per health state to be built.

We used a widely recognized scale to measure quality of life, the Kaplan-Bush Index of Well-Being, also known as the Quality of Well-Being (QWB) scale. This

---

[6] These extrapolations included extension of the 5-year curve using a flat line, a linear fit, and a modified exponential decay.

scale ranges from 0 = death to 1 = full health, with all other health states lying in between. This scale allows us to express the health benefits of treatment in terms of QALYs, where a QALY is defined as 1 year in perfect health. In this sense, a QALY-based view of health benefits is a natural generalization of our simple survival-based measure of LYs.

As with the cancers, the timeframe for QALYs also affects the absolute value of health benefits. We used an average life expectancy (from disease onset) for both diabetes and cholelithiasis (see also Appendices 3A and 4A).

Appendix 1E: Benchmarks
for cost-effectiveness analysis

In Chapter 1, we described how we compare productive efficiency among countries. In some cases we cannot determine which country is more productive; we can, however, perform a cost-effectiveness assessment to help determine which country had the preferred input/outcome combination and may therefore be "better off" from an overall economic efficiency perspective. To do this, we compare the value of the additional inputs and additional outcomes, creating a ratio of the cost per LY or cost per QALY.

Chapter 1 lists benchmarks that have been used in various publications. These benchmarks were derived from cost-effectiveness studies of common practices and from observations of what current medical spending was generally considered acceptable. Although a truly standardized scale is unavailable, we describe in this appendix the origin of some of these benchmarks, as well as other potential benchmarks.

KAPLAN-BUSH BENCHMARKS

Kaplan and Bush, who developed a QWB scale to measure QALYs, also gave approximate rules for interpreting the QALY results.[7] (The Kaplan-Bush QWB scale and the QALY measure are discussed in Appendices 3A and 4A.) Their guidelines for the cost per QALY analyses relied on comparisons with other disease treatments and how the cost-effectiveness of these treatments was generally perceived. Based partly on the Kaplan-Bush figures, we identified three categories for health care spending options. Costs are expressed here in 1990 U.S. dollars. Below $30,000 per QALY, the treatment was considered "cost-effective by current standards." Between $30,000 and $100,000, the treatment was "possibly controversial, but justifiable by many current examples." And treatments costing more than $100,000 per QALY were "questionable in comparison with other health care expenditures."[7]

---

[7] Source: Kaplan RM, Bush JW. Health Psychology 1982; 1:61-80.

PRIOR STUDIES

In **Exhibit E-1** we show results from previous studies of other diseases.[6, 8] The cost-effective category included coronary artery bypass surgery for the left main coronary at $4,922 per QALY and treatment of mild hypertension in males age 40 at $10,896 per QALY. In-center hemodialysis ($43,952 per QALY) and coronary artery bypass graft for two-vessel disease ($39,770 per QALY) were both at the low end of the "possibly controversial, but justifiable" category. But cholestipol treatment for high cholesterol neared the "questionable" category. Total hip replacement was well into this upper category, with a ratio of $293,029 per QALY. These examples were taken from a variety of studies that employed different methods to measure QALY outcomes. Thus, comparisons considered a wide spread in the exact cost per QALY ratio. These examples and the Kaplan-Bush benchmarks therefore provided some signposts for interpreting our U.S. *versus* U.K. results. Note that the cost-effectiveness ratios for a procedure or other form of care depend greatly on the characteristics of the patient or population in which it is applied; a treatment can be highly cost-effective for one person and a very poor value for another.

AVERAGE WAGE AS ALTERNATIVE BENCHMARK

Another simple, first-order approach for interpreting costs per QALY is to set the minimum acceptable price per QALY at the country's average annual wage. This approximation is crude since equating wage levels to the worth of a life could raise a number of social and philosophical questions. Nonetheless, wage levels can serve as proxies or lower bounds for cost-effectiveness ratios. The minimum acceptable cost per QALY could differ by country, in accordance with the country's wealth. For example, the average wage was lower in the U.K. than in the U.S., suggesting that the U.K. system was willing to spend less per QALY than the U.S. system. In 1990, the average wage levels were $31,572 in the U.S. and $22,375 in the U.K. (1990 U.S. dollars, converted with GDP PPP). This estimate for the upper boundary of clearly cost-effective treatment in the U.S. ($31,572) was quite similar to the Kaplan-Bush estimate of $30,000, but the U.K. threshold is lower, as expected, at $22,000.

---

[8] Source: Torrance GW, J Health Economics 1986; 5:1-30.

LIMITATIONS OF BENCHMARKS

While benchmarks provide a pragmatic standard for comparing relative cost-effectiveness and preferred input/outcome combinations, several caveats must be kept in mind:

¶ They are not absolute boundaries, but rather very rough guides based on comparing the cost-effectiveness of several treatments in common medical practice.

¶ The preferred input/outcome combination strictly depends on societal preferences, which are likely to vary among countries (and even within a country). In particular, it might be completely rational for a poorer country to prefer somewhat worse outcomes with lower inputs, while a richer country might prefer the opposite (**Exhibit E-2**).

¶ Prices for health care inputs are considerably higher in the U.S. than in the U.K. or Germany (see Chapter 8). In local currencies, the cost of additional inputs to achieve an additional LY might, therefore, be lower using prices from the U.K. and Germany rather than from the U.S., potentially offsetting differences in preferences described above. Therefore, we must value the inputs in *each* of the three country's prices in order to comment on preferred input/outcome combinations.

Exhibit 1

HEALTH CARE EXPENDITURES
1990

Expenditures as a percent of GDP

Expenditures per capita
U.S. dollars, purchasing power parity (PPP)*

	Percent of GDP
U.S.	12.7
Germany	8.3
U.K.	6.0

	Per capita
	2,439
	1,473
	1,113
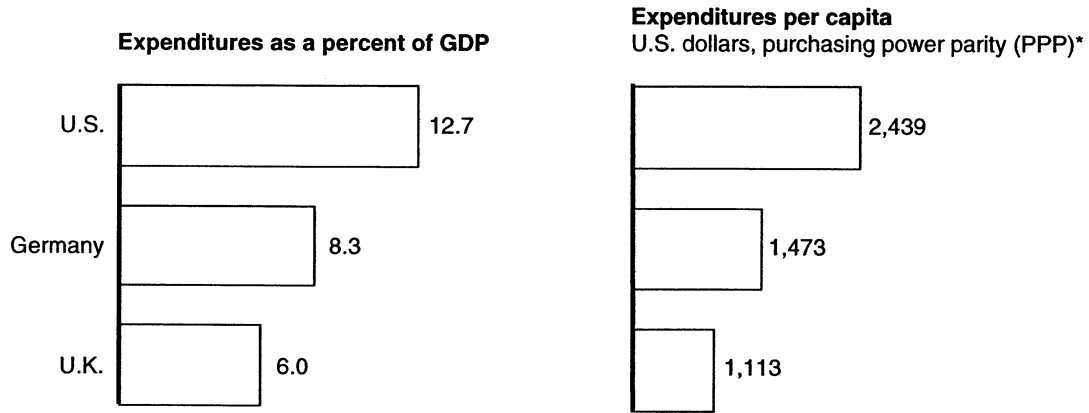
* Local currencies converted to U.S. dollars using GDP PPP for 1990 from OECD
Source: OECD; BASYS

Exhibit 2

GROWTH IN HEALTH CARE SPENDING
IN THE U.S., U.K., AND GERMANY
Spending as a percent of GDP
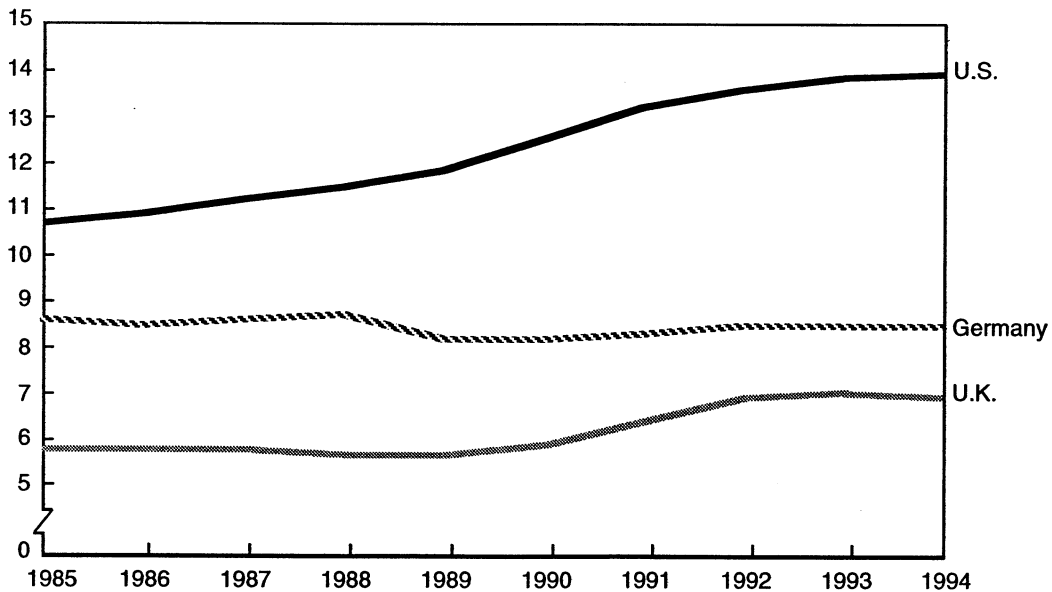
Exhibit 3

GROWTH IN HEALTH CARE SPENDING
IN THE U.S., U.K., AND GERMANY

1990 dollars per capita, PPP
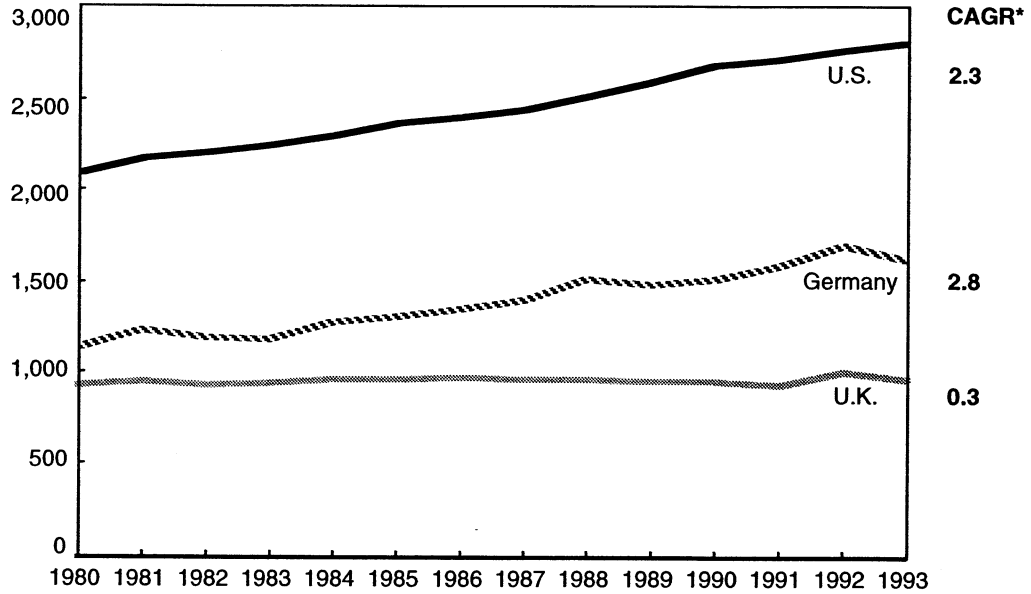
Exhibit 4

GROWTH IN HEALTH CARE SPENDING
IN THE U.S., U.K., AND GERMANY

1980 index = 100, local currency, deflated
to real values in 1990 local currency units
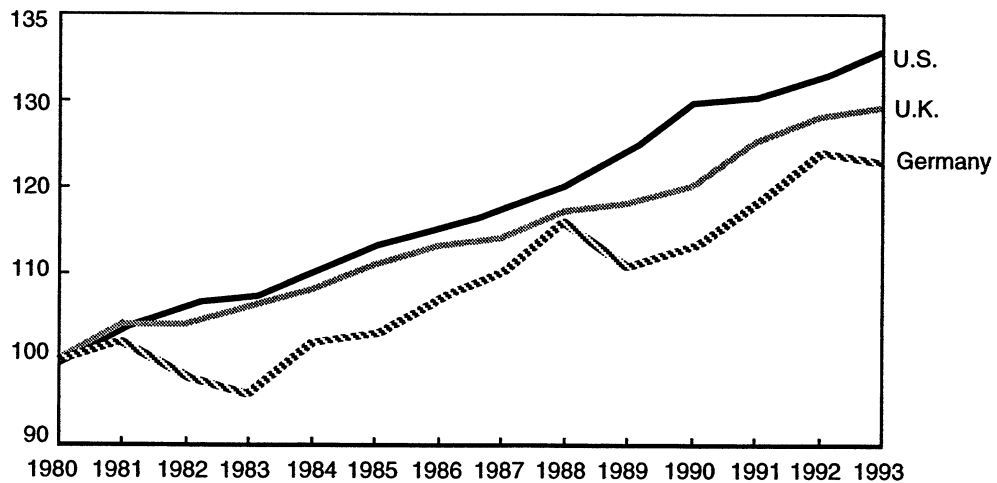
Exhibit 5

LIFE EXPECTANCY AND INFANT MORTALITY
1990

	Life expectancy at birth Years	**Infant mortality per 100 live births** Deaths per 100	**Life expectancy at age 40** Years
U.S.	75.4	0.92	77.6
U.K.	75.8	0.79	77.5
Germany	75.8	0.71	77.9
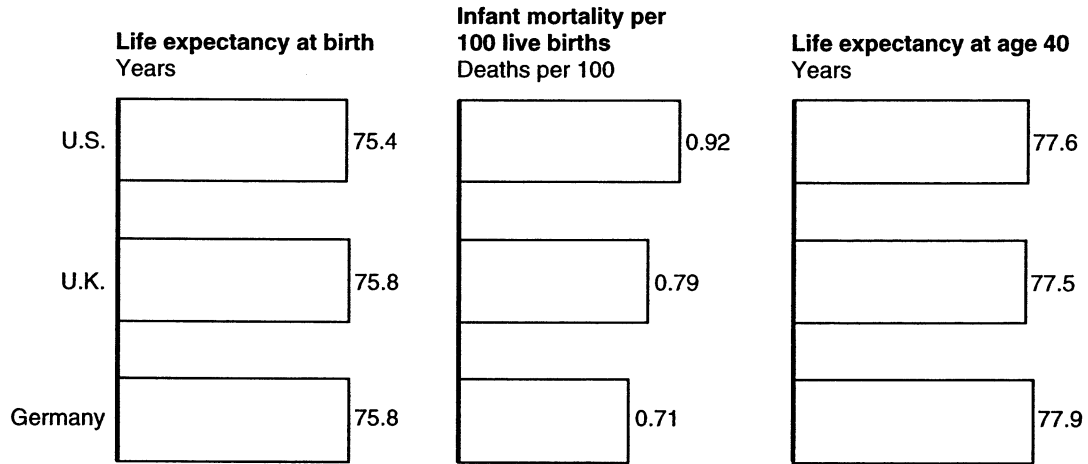
Note: Life expectancy here is an average of male and female levels

Source: OECD; Statistisches Bundesamt; Government Actuaries Department, U.K.; National Center for Health Statistics

---

Exhibit 6

MORTALITY RATE FOR SELECTED CONDITIONS – 1990-92
Average for age groups 35-44, 45-54, 55-64
Deaths per 100,000 people per year

Country	All cancers	Breast cancer	Lung cancer	Myocardial infarction
U.S.	306.5	58.3	72.6	68.5
U.K.	214.2	56.9	50.9	89.1
Germany	214.4	47.1	48.3	70.3
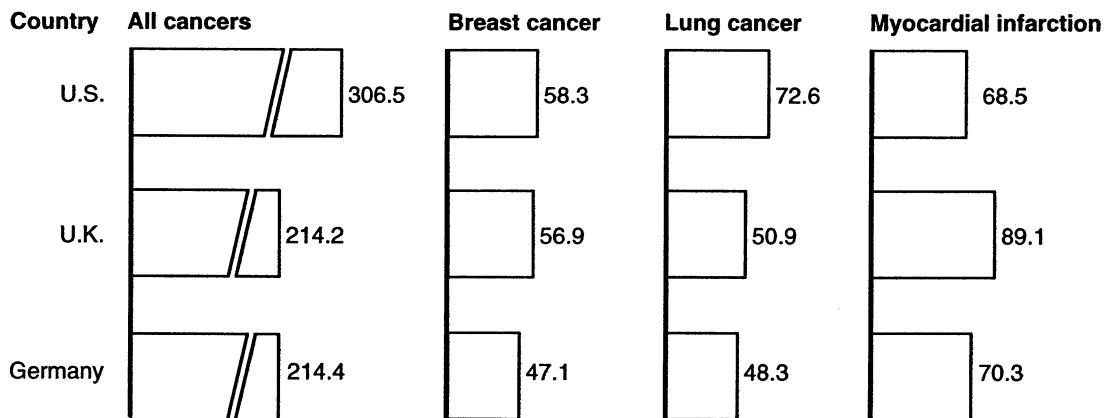
Note: Average mortality for ages 35-64 assuming population is 50% male, except for breast cancer; figures are the average of the death rate per 100,000 for each age group

Source: United Nations (UN), World Health Organization (WHO) 1993

022 P 134534/1

Exhibit 7

FRAMEWORK FOR CAUSAL ANALYSIS

Productive efficiency • Inputs • Outcomes		

▲

| Provider behavior
• Care triaging
• Treatment duration
• Staffing levels
• Setting choice
• Team-based approach
• Technology adoption | Direct driver of productive efficiency | } Vary by disease |

▲

| Provider incentives and constraints
• Physician incentives
• Hospital incentives
• Physician supply
• Hospital supply
• Capital constraints
• Substitution constraints | Direct drivers of provider behavior | |

▲

| Health care system structure
• Product integration and pricing
• Competitive intensity | Economic interactions among participants in health care system that drive incentives and constraints | } Generally common across diseases |

▲

| Regulation | The "rules and regulations" that shape interactions among participants | |

022 ST 140103/1

Exhibit 8

SURVIVAL CURVES WITH AND WITHOUT TREATMENT

Per case

**Percentage
of survivors**

100%

Improvement in
total life expectancy

With treatment

Without treatment

Time after diagnosis
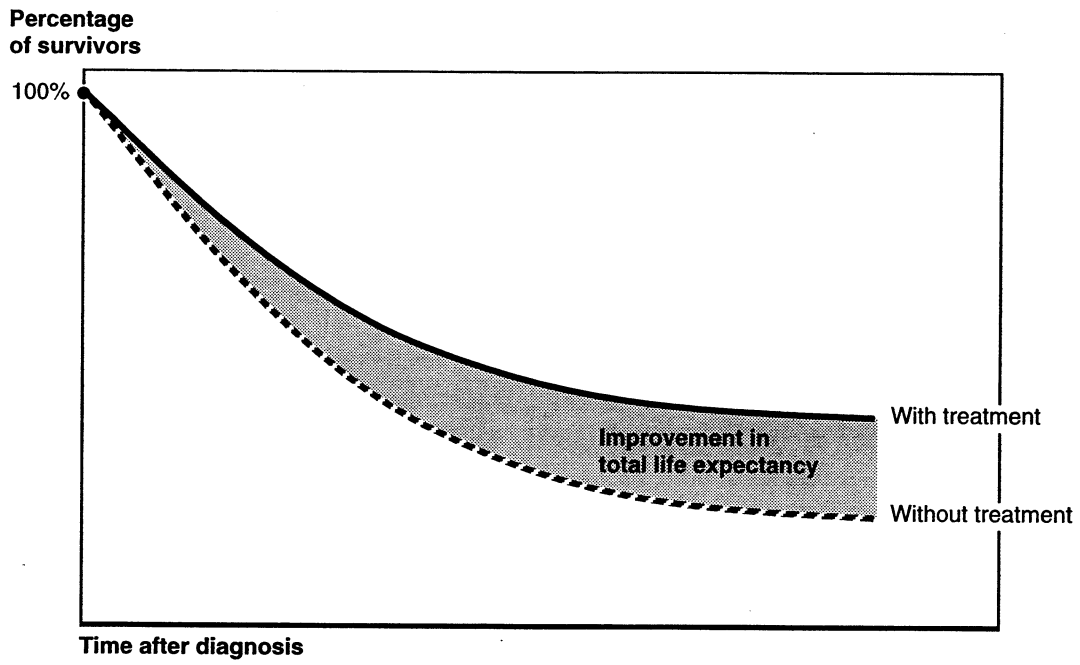
022 EF 136256/1

Exhibit 9

ASSESSING PRODUCTIVE EFFICIENCY

Per case

Outcomes*

Inputs

Comparison 1: A vs. B
A is more productive as it achieves better or equal outcomes with less inputs

Comparison 2: A vs. C
A is more productive as 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

* Relative to baseline outcome with no treatment
Note: Solid lines represent treatment processes, or "production functions," while slope of each dotted line represents average productivity
022 CE 117817/1

Exhibit 10

DETERMINING PREFERRED INPUTS/OUTCOMES WHEN
BOTH ARE HIGHER IN ONE COUNTRY VS. ANOTHER

Value to consumers/ policymakers	Preferred input/outcome combination
Value of difference in outcome > value of difference in inputs	Ⓑ
Value of difference in outcome < value of difference in inputs	Ⓐ

Ⓑ
▲ Difference
 in outcomes
 per case

Ⓐ
Difference
in inputs
per case

022 ST 140155/1
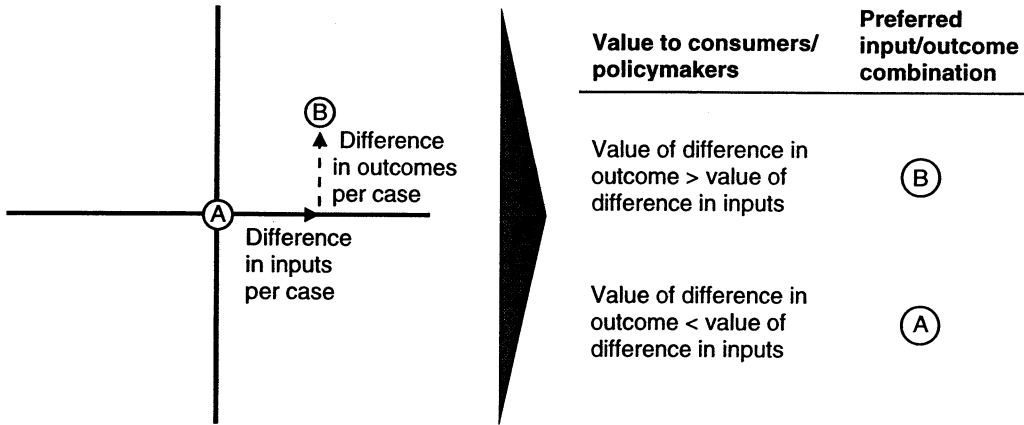
Exhibit A-1

TOTAL COSTS IN LOCAL CURRENCIES

Hypothetical values

Country	Physician-hours used	Nurse-hours used	Price of physician-hour	Price of nurse-hour	Total cost
U.K.	2	4	£50	£20	£180
Germany	1	8	80 DM	40 DM	400 DM
U.S.	1.5	6	$100	$36	$366

022 EF 136249/1

Exhibit A-2

RELATIVE FACTOR PRICES

Hypothetical values

Country	Price of physician-hour	Price of nurse-hour	Relative price of nurse-hours in terms of physician-hours
U.K.	£50	£20	0.40
Germany	80 DM	40 DM	0.50
U.S.	$100	$36	0.36

Exhibit A-3

TOTAL COSTS IN TERMS OF PHYSICIAN-HOURS

Hypothetical values

Country	Physician-hours used	Nurse-hours used	Price of physician-hour	Price of nurse-hour	Total cost	Cost index
U.K.	2.0	4.0	1.0	0.42	3.68	91.5%
Germany	1.0	8.0	1.0	0.42	4.36	108.4%
U.S.	1.5	6.0	1.0	0.42	4.02	100.0%

Exhibit A-4

RANK REVERSAL AT EXTREME PRICES

Hypothetical values

Country	Physician-hours used	Nurse-hours used	Price of physician-hour	Price of nurse-hour	Total cost	Cost index
U.K.	2.0	4.0	1.0	0.2	2.8	109%
Germany	1.0	8.0	1.0	0.2	2.6	96%
U.S.	1.5	6.0	1.0	0.2	2.7	100%

Exhibit A-5

TOTAL COSTS IN AVERAGE AND ACTUAL RELATIVE PRICES

Hypothetical values

Country	Physician-hours used	Nurse-hours used	Price of physician-hour	Price of nurse-hour	Total cost	Cost index
Average relative prices						
U.K.	2.0	4.0	1.0	0.42	3.68	91.5%
Germany	1.0	8.0	1.0	0.42	4.36	108.4%
U.S.	1.5	6.0	1.0	0.42	4.02	100.0%
U.K. relative prices						
U.K.	2.0	4.0	1.0	0.40	4.6	92.3%
Germany	1.0	8.0	1.0	0.40	4.2	107.7%
U.S.	1.5	6.0	1.0	0.40	3.9	100.0%
German relative prices						
U.K.	2.0	4.0	1.0	0.50	4.0	88.9%
Germany	1.0	8.0	1.0	0.50	5.0	111.1%
U.S.	1.5	6.0	1.0	0.50	4.5	100.0%
U.S. relative prices						
U.K.	2.0	4.0	1.0	0.36	3.44	94.0%
Germany	1.0	8.0	1.0	0.36	3.88	106.0%
U.S.	1.5	6.0	1.0	0.36	3.66	100.0%

Exhibit A-6

CONDUCTING CONSISTENCY CHECKS

	Under U.K. input prices	Under German input prices	Under U.S. input prices	Under average input prices
Cost of U.K. inputs				
Cost of German inputs	↕	↕	↕	↕
Cost of U.S. inputs				
	Compare	*Compare*	*Compare*	*Compare and report*

◄————*Use for consistency check*————►

◄————*Report if different result*————►
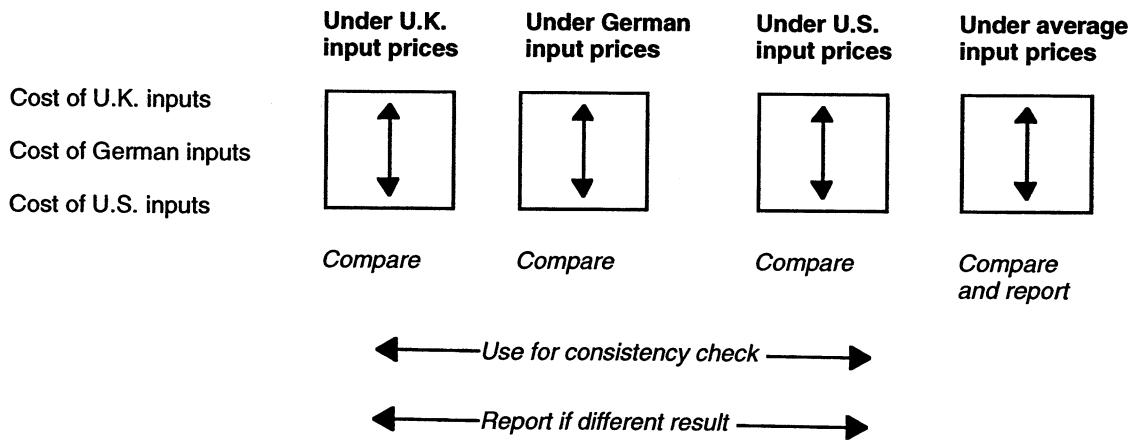
Exhibit B-1

STAFFING LEVEL DURING AN INPATIENT STAY

More care-intensive diseases

Staffing level

Model

Actual

Days of hospital stay
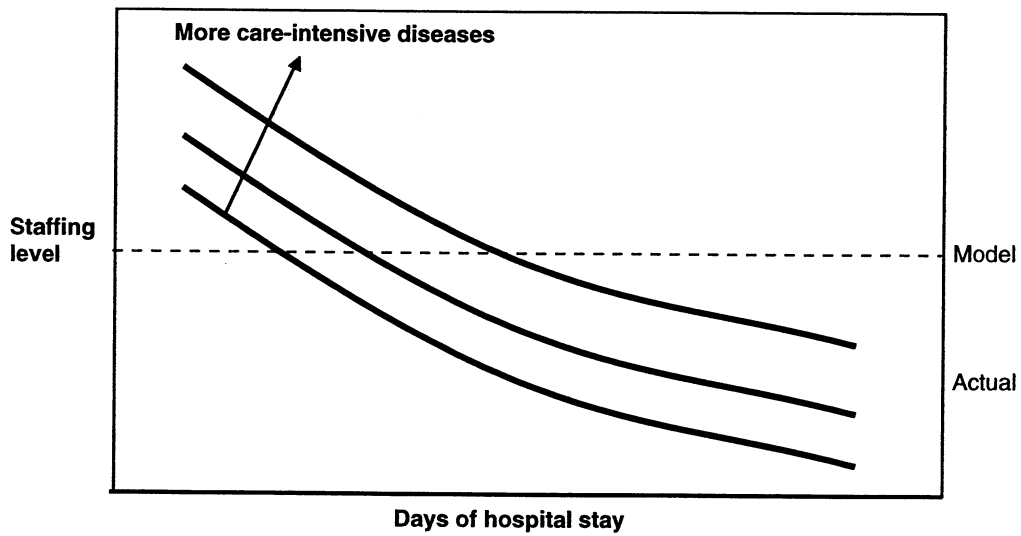
Exhibit B-2

STAFFING LEVELS PER BED-DAY
Standardized input units per bed-day

	Physicians	Qualified nurses	Nonqualified nurses	Medical technicians
U.S.	0.28	1.60	0.56	0.34
U.K.	0.22	1.08	0.34	0.27
Germany	0.24	0.81	0.22	0.13
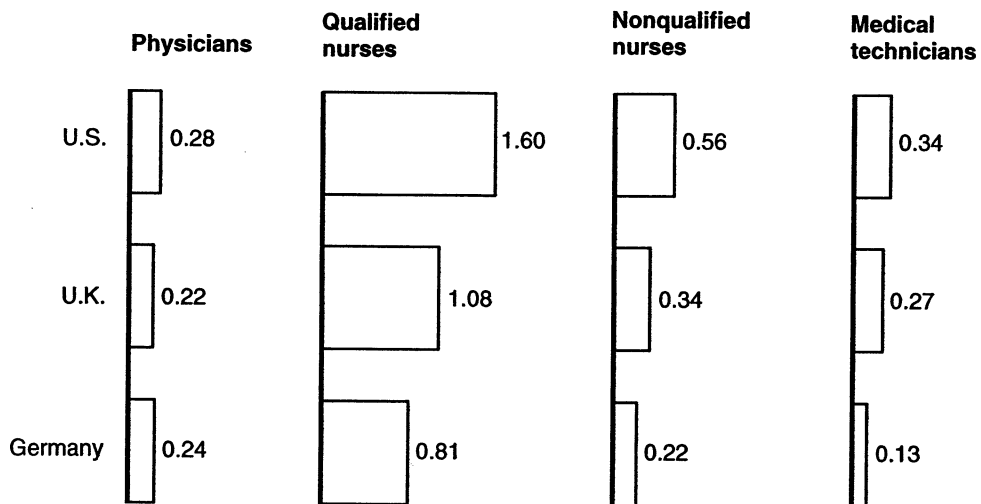
Source: McKinsey analysis
022 P 134544/1

Exhibit B-3

WEIGHTED STAFFING LEVELS PER BED-DAY

Standardized input units per bed-day

Country	Value
U.S.	0.76
U.K.	0.55
Germany	0.47
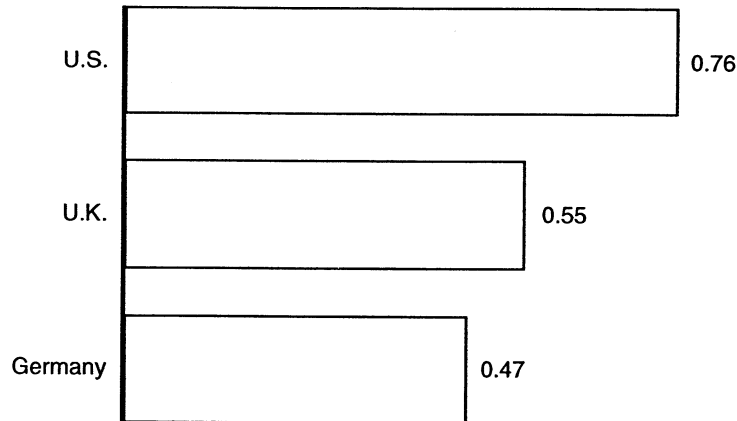
Note: The above estimates are the weighted sums of staff per bed-day of physicians, qualified nurses, nonqualified nurses, and
medical technicians where the weights represent the cost of each personnel category relative to that of a surgeon
Source: McKinsey analysis

Exhibit B-4

WEIGHTED STAFFING LEVELS PER ADMISSION

Standardized input units per admission

■ Medical technicians
▨ Nonqualified nurses
▒ Qualified nurses
□ Physicians

Country	Medical technicians	Nonqualified nurses	Qualified nurses	Physicians	Total
Germany	0.4	0.3	2.1	3.0	5.7
U.S.	0.6	0.4	2.4	2.0	5.5
U.K.	0.4	0.2	1.4	1.4	3.5
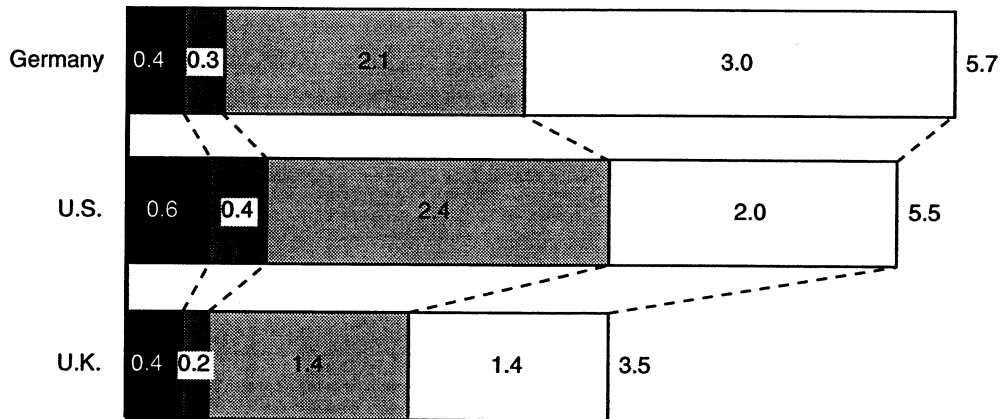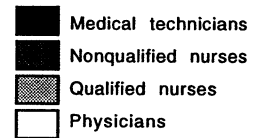
Note: Each category is weighted by the cost of that category's labor relative to that of a surgeon
Source: McKinsey analysis

Exhibit C-1

PROXY VARIABLES FOR CAPITAL AND SUPPLIES INPUT

Proxy	Driver of	Breast cancer	Lung cancer	Chole-lithiasis
For capital				
LOS	General facility and equipment	X	X	X
Operating room (OR) hours	OR equipment	X	X	X
X rays	X-ray equipment	X	X	X
CT scans	CT equipment	X	X	
Radiotherapy fractions	Radiotherapy equipment	X	X	
Mammograms	Mammographic equipment	X		
Ultrasounds	Ultrasound equipment			X
Outpatient visits	Clinic facility and equipment			X
Endoscopic procedures	Endoscopic equipment			X
EKGs	Electrocardiogram equipment			X
For supplies				
LOS	General inpatient supplies	X	X	X
Outpatient visits	General outpatient supplies	X	X	
OR hours	OR supplies	X	X	
Diagnostic tests	Diagnostic supplies (lab, film, etc.)	X	X	
Units of chemotherapeutics	Pharmaceuticals	X	X	

022 EF 136254/1

Exhibit C-2

COST STRUCTURE OF INPATIENT CARE

1989

Percent

	100.0		100.0		100.0
Capital	7.7		7.3		8.1
Supplies	28.4		24.4		22.1
Labor	63.9		68.3		69.8
	U.S.*		U.K.		Germany

Exhibit C-3
EXAMPLE VALUES FOR LABOR AND X RAY PROXY

Country	Labor index	X rays
U.S.	8	1
U.K.	4	3
Germany	9	2
Average	*7*	*2*

022 EF 136282/1

SURVIVAL CURVE

**Percentage of
survivors**

100%

Total life
expectancy

Time after diagnosis
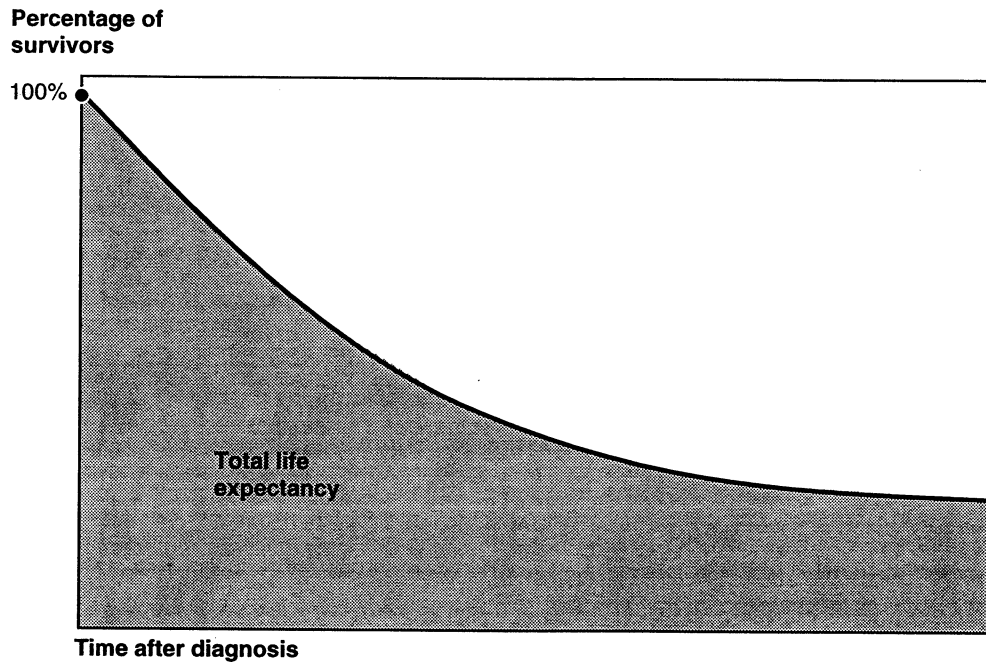
022 P 134550/1

Exhibit E-1

EXAMPLES OF COST-EFFECTIVENESS BENCHMARKS

$ Thousands/QALYs (1990 U.S. dollars)

Treatment of mild
hypertension in
males age 40

In-center
hemodialysis

Total hip
replacement
($293,029)

Coronary artery
bypass surgery for
left main coronary

Coronary artery
bypass graft (surgery)
2-vessel disease

Cholestipol for
high cholesterol

0   10   20   30   40   50   60   70   80   90   100

**Kaplan-Bush
benchmarks**

"Cost-effective by
current standards"

"Possibly controversial, but justifiable
by many current examples"

"Questionable
in comparison
with other
health care
expenditure"

Source: R.M. Kaplan and J.W. Bush, 1982; G.W. Torrance, 1986
022 P 134551/1

Exhibit E-2

DIFFERENCES IN PREFERENCES BETWEEN COUNTRIES

Poorer country preferences

Richer country preferences
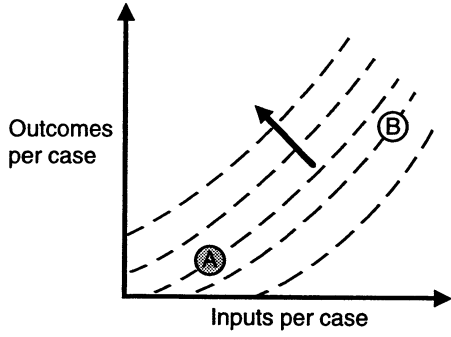
**Poorer country prefers lower
input/lower outcome combination**
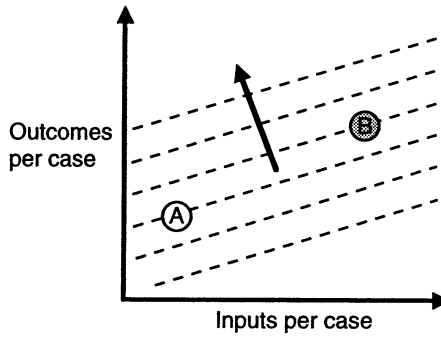
**Richer country prefers higher
input/higher outcome combination**

022 P 134548/1

Executive summary

Health care accounts for a large and rapidly growing portion of the gross domestic product (GDP) in the developed world. As a result, governments and health care organizations are increasingly interested in ways to rethink and reform their health care systems (**Exhibit 1**). Yet we are early in the process and many fundamental questions remain: *What are the sources of the differences in spending among countries? Why do those differences appear to be unrelated to differences in overall life expectancy?*

To help provide a foundation for future reform, we examined and compared the health care systems in the U.S., Germany, and the U.K. We did so by assessing productivity in the treatment of four diseases during the late 1980s: diabetes, cholelithiasis (gallstones), breast cancer, and lung cancer. We looked at the different day-to-day actions of doctors and hospitals and tried to connect these actions to differences in longevity and the quality of life. Surprisingly, different actions were mainly due to differences in how doctors and hospitals were paid and constraints they faced in providing treatment.

Each country had a different system structure, levels of spending, and levels of health care productivity. These differences stemmed from different kinds of regulation. In particular, we found competitive intensity and care integration to be very important in explaining productivity. Recent changes in the U.K. and the U.S. systems – which increase competition and integration – are likely to help more than those in Germany.

Our principal findings are (**Exhibit 2**):

¶ The U.S. spends the most (per capita) on health care followed by Germany and then the U.K. Higher spending in the U.S. was largely due to higher compensation for doctors and other personnel and higher administrative costs (**Exhibit 3**).

¶ The U.S.'s higher spending was ***not*** due to low productivity; in fact, it led Germany in all cases and led the U.K. in lung cancer and gallstones. It trailed the U.K. only in diabetes (**Exhibit 4).**

- The U.S. led in lung cancer and gallstones because it adopted productive technologies more quickly and broadly and had shorter hospital stays.

- Germany was the least productive because it used less outpatient care and kept patients in the hospital longer.

1

- The U.K. led the U.S. in diabetes because it focused treatment on patients who could benefit most and integrated the care of multiple specialists better over the course of a patient's lifetime.

WIDE VARIATIONS IN TREATMENT AND PRODUCTIVITY

Assessing relative productivity in health care is more difficult than in other industries because treating patients is a far more complex process than that typically found in the manufacturing or service sectors.[1] It involves numerous inputs, including the patient's own behavior. In addition, the output of the treatment process – improved health – is very hard to quantify.

It is, however, possible to cut through these complexities and reach a number of conclusions. For example, the conventional wisdom that the U.S. wastes resources in health care is challenged by the finding of its higher productivity relative to Germany and its mixed productivity relative to the U.K. While the U.S. did have productivity improvement potential in at least one disease (diabetes), the causes of the U.S.'s higher aggregate spending were its high compensation for doctors and other personnel and high administrative costs (**Exhibit 5**). Although this is a significant result, assessing the underlying causes was outside the scope of this research effort.

Again contrary to common wisdom, the large cross-country differences in productivity for each disease stemmed from dramatic variations in how doctors and hospitals treated patients. Despite similar clinical training and access to similar medical expertise and technology, there were surprisingly large differences in selection of patients for treatment, how long it took to treat a disease, when and how broadly technologies were adopted, and where treatment was given (**Exhibit 6 – top two sections**).

For example, more selective delivery of care and slower technology adoption in the U.K. led to 23-percent fewer resources used, but mixed productivity relative to the U.S. The U.K.'s lower productivity in gallstones resulted from later adoption of highly productive laparoscopic surgery (video-guided, small incision). Its lower productivity in lung cancer resulted from a more restricted patient selection process and, in particular, less use of computerized tomography (CT) scans in diagnosis and staging of cancer progression. In diabetes, however, the U.K.'s superior care integration led to lower complication rates and greater productivity, resulting in less resource use through aggressive management and team-based care in specialized clinics. Finally, in breast cancer, the U.K.'s lack of the broad-based mammographic screening program used in the U.S. appears to have increased productivity in some aspects of treatment.

---

[1] *Service Sector Productivity*, McKinsey Global Institute, Washington, D.C., October 1992; *Manufacturing Productivity*, McKinsey Global Institute, Washington, D.C., October 1993.

Germany's greater use of inpatient care led to 39-percent higher resource use on average and lower productivity relative to the U.S. In all three disease comparisons, Germany favored inpatient treatment over the less resource-intensive outpatient treatment and had significantly longer hospital stays. In gallstones, Germany also had much longer patient recovery periods, even with similar adoption rates of laparoscopic technology. The relatively consistent pattern suggests that underlying health care system characteristics strongly influence provider behavior and treatment approaches.

In all countries studied, doctors and hospitals responded predictably and consistently to their economic incentives and constraints within the boundaries of acceptable medical practice. Incentives and constraints were, in turn, determined by the structure of the health care system and by the ways in which the most important markets were regulated, particularly those for health insurance coverage and for hospital and physician services. The three countries in our assessment arrived at very different structures by the late 1980s, particularly in terms of the degree of care integration and competitive intensity. Because of the incentives and constraints they created, differences along these two dimensions led to varying productivity by disease. No country was most productive across all diseases.

COUNTRY-TO-COUNTRY COMPARISONS

U.K.-U.S differences. The U.K.'s more selective delivery of care and slower technology adoption primarily resulted from its economic incentives for doctors and its tight constraints on the supply of physicians, hospitals, and capital. These differences, in turn, were a product of the U.K. health care system's fixed physician salaries which contrast sharply with fee-for-service (FFS) payments in the U.S., its lower physician competition for patients and payor contracts, and stronger regulation (**Exhibit 6**).

The U.K. trailed in treating gallstones, because the U.S.'s higher competitive intensity and FFS physician incentives led to faster and broader adoption of the laparoscopic technology; U.S. physicians had economic reasons to be more responsive to consumer demand and insurance companies readily accepted the more cost-effective surgical substitute. In breast cancer, this same combination led to lower productivity in the detection of the disease because it encouraged a broad-based screening program in the U.S. In lung cancer, the U.K.'s restrictions on physician referral processes and its cap on capital investments led to a more intense triaging process with far less use of CT scans for diagnosis and staging. This resulted in a less than optimal group of patients selected for surgery and, when coupled with its longer hospital stays after surgery and more hospital stays with chemotherapy, led to lower productivity in the U.K.

3

In diabetes, however, the U.K.'s more integrated care approach and lower competitive intensity led to higher productivity. The National Health Service (NHS), which provided lifetime health coverage for the entire population, identified diabetes as a priority, provided dedicated funding, and actively encouraged providers to organize in specialized clinics with aggressive preventive care and disease management. Also, the limited supply of general practitioners (GPs) and demands on their time forced them to be better at selecting diabetics for treatment, referring the most severe to the diabetic clinics, and encouraging the less severe to follow self-treatment protocols. This led to much lower complication rates overall, less resource use, and better outcomes.

In contrast, the U.S. system provided disincentives for pursuing such an integrated approach. The high member turnover for U.S. insurers – up to 40 percent annually – and their fear of attracting too many diabetics if they offered integrated treatment, coupled with FFS physician incentives, led to highly fragmented care approaches and to less willingness to invest in specialized or preventive care.

Germany-U.S. differences. Germany's greater use of inpatient services and longer length of treatment can be directly linked to three factors: strong incentives for German hospitals and for some specialist physicians to fill hospital beds, regulations on hospital supply that actually led to surplus capacity, and regulation that discouraged substitution of outpatient care for inpatient care (**Exhibit 6**).

Specifically, German physicians and hospital services were, by law, compensated by their sickness funds, Germany's equivalent of U.S. insurers, on a per-day basis. In contrast, U.S. hospital services were compensated based on case rates (a set payment for the entire hospital episode) through Medicare and through a mixture of approaches from private insurers – including FFS, per day, and case rates. And while both German and U.S. hospitals competed aggressively for patients, only the U.S. hospitals faced any competition in their negotiations with insurers; by law, German hospitals negotiated with all sickness funds as a block for annual per-day rates.

In addition, U.S. private insurers faced price-based competition for members and, therefore, had some incentive to manage hospital costs and lengths of stay (LOS), whereas German sickness funds were essentially precluded from competing on price and from bundling hospital care in different ways. German hospitals also faced the threat of regulatory review and potential capacity cuts if their occupancy fell below 85 percent. Furthermore, German hospital department chiefs had incentives to increase the workload of their hospitals because they could earn substantial FFS income from private patients to supplement their hospital salaries, and because their department was allowed bed capacity for private patients in a relatively fixed ratio to its utilized public beds.

Ironically, Germany's attempts to regulate hospital capacity actually helped create surplus capacity. State governments had incentives conflicting with productivity, since they could create jobs and receive federal transfers from federal payor funds into their state economies by increasing local hospital bed capacity. In addition, Germany's regulatory barrier between inpatient and outpatient care – with separate providers and specified services, payment, governance, and oversight – precluded shifting care to more cost-effective outpatient settings as well as coordinating care across these care settings.

Overall, these regulatory constraints, coupled with the regulated per-day hospital price and lower competitive intensity, led to Germany's much higher resource use and lower productivity relative to the U.S.

RECENT CHANGES AND IMPLICATIONS

Combining the productivity findings of the individual disease cases with aggregate analyses suggests several implications for policymakers and health care organizations interested in reform. The critical *first step* is to clearly identify the problem or opportunity in precise terms: Is productivity low? Are compensation, other medical prices, or administrative costs too high?

Since the late 1980s, the time period covered by our cases, the health care systems in each country have changed significantly toward higher competitive intensity and greater integration of care (**Exhibit 7**). While the impact of these changes on productivity, input prices, and administrative costs has yet to be determined, our study findings allow us to assess the extent to which the changes will more than likely improve productivity.

In the U.S., there is evidence that the largely market-based system is leading to greater competitive intensity and higher care integration in at least some diseases without any significant regulatory changes. More integrated managed care products, such as health maintenance organizations (HMOs) and preferred provider organizations (PPOs), have emerged as a result of intensifying employer demands and increasingly competitive health coverage and care provision markets. In addition, both insurers and health care providers have created disease "carve-out" products that integrate care more effectively in such areas as cancer and diabetes. Not surprisingly, these developments have also led to a decline in specialist physician compensation and to actual price reductions for health coverage in some markets. The effects on administrative costs are unclear.

In the U.K., the 1991 reforms introduced some competition at the local level between payors and providers and fostered somewhat more integrated care, but left the integrated lifetime coverage and monopoly power of the NHS intact. Many NHS-owned hospitals were also privatized as self-governing

trusts with greater control over their capital purchases. However, the overall budget constraint remained.

While the system changes have actually increased administrative costs, their productivity impact is still unclear. According to some estimates, 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. In addition, there is some evidence that technology adoption rates have quickened. For example, a targeted breast cancer screening program based on mammography was established and adoption of laparoscopic technology has approached U.S. levels.

While some supply and capital constraints remain for hospitals and their associated specialists and while system competition has been limited to date, we would expect some improvement in the U.K. system productivity over time, at least in the diseases studied.

In Germany, major reforms have been made in the health coverage and, to a lesser extent, in the care provision markets. As of 1996, the 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 individual providers or to bundle care in different ways (e.g., by disease or case) have been left intact. While regulated case rate payments for hospitals have been introduced to substitute for per-day payments, 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.

It is unlikely that recent changes in the German system will do much to improve productivity, unless they somehow lead to removal of the regulatory barrier between inpatient and outpatient substitution, greater flexibility in sickness funds' negotiations with providers, or the adoption of case rate hospital payments across the board.

Thus, the U.S. and the U.K. appear to be moving in the direction of productive change in their health care systems, with each adopting some productive characteristics of the other. Given the questionable productivity impact of the German reforms, it is likely that Germany's productivity gap with the U.S. and possibly the U.K. is widening.

Exhibit 1

GROWTH IN HEALTH CARE SPENDING
IN THE U.S., THE U.K., AND GERMANY

Spending as a percent of GDP
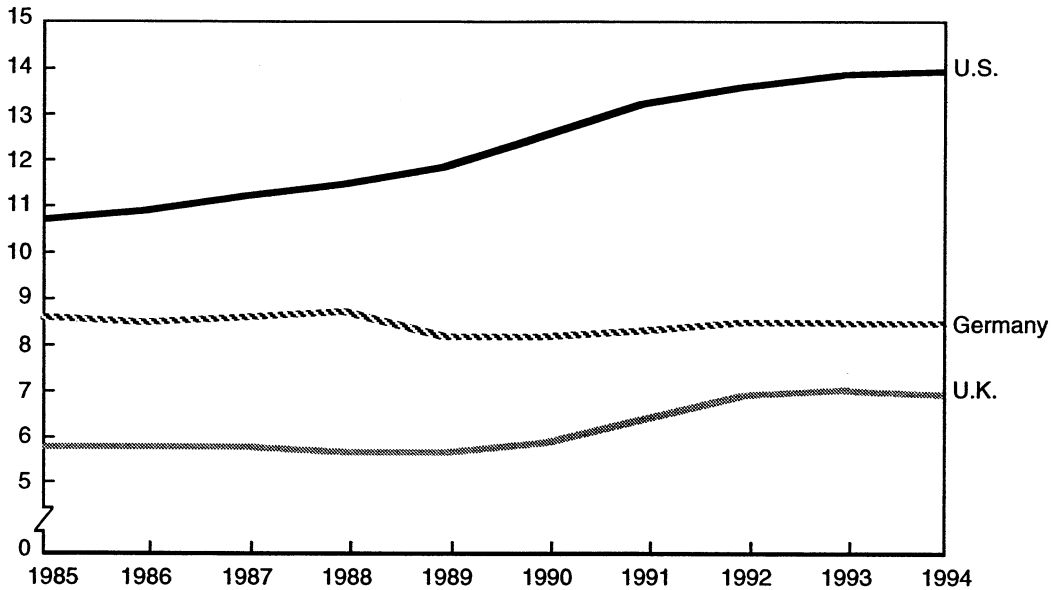
Source: OECD; Health Care Financing Administration; Office of National Health Statistics

---

Exhibit 2

RELATIONSHIP BETWEEN HEALTH CARE SYSTEM
STRUCTURE AND OVERALL PERFORMANCE

1990

High

**Competitive
intensity**

**U.S. –
Highest aggregate spending**
• More productive than Germany, mixed vs. the U.K.
• Higher inputs than the U.K., but lower than Germany
• Highest administrative costs
• Highest input prices

**Germany –
Modest aggregate spending**
• Mixed productivity
• Highest inputs
• Lowest administrative costs
• Lower input prices

**U.K. –
Lowest aggregate spending**
• Mixed productivity
• Lowest inputs
• Lower administrative costs
• Lowest input prices

Low

Low   High

Health care product integration

Exhibit 3

HEALTH CARE SPENDING DISAGGREGATION
1990

Relative productivity			Price of medical inputs*	Administrative costs Percent of total health care costs	Total health care spending per capita Dollars, PPP
Medical input level*	Disease treatment outcomes	Estimated productivity			
U.S. 100	Generally higher	Higher than Germany; mixed vs. the U.K.	100	24	2,439
U.K. 75	Mixed	Mixed	56	16	1,113
Germany 116	Mixed	Lower than U.S.; mixed vs. the U.K.	70	13	1,473

X   +

* Indexed values with U.S. = 100; medical input levels determined as a weighted sum of labor (physicians, nurses, and medical technicians), supplies (pharmaceuticals), and capital; price of medical inputs determined using the OECD price index

Source: McKinsey analysis

Exhibit 4

RELATIVE PRODUCTIVITY BY DISEASE CASE
1990

Disease	U.S. vs. U.K.			U.S. vs. Germany			Germany vs. U.K.		
	Inputs	Out-comes	More productive country	Inputs	Out-comes	More productive country	Inputs	Out-comes	More productive country
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		

Exhibit 5

SOURCES OF DIFFERENCE IN HEALTH CARE SPENDING

U.S. dollars per capita, 1990, PPP

U.S. vs. Germany

| | | | 259 | 2,439 |

1,473   −390   737   360   Other   2,439

Inputs used (U.S. ~15% lower)

Prices of inputs (U.S. ~40% higher)*

Administrative overhead (estimate U.S. 24%; Germany 13%)

Germany   U.S.

U.S. vs. U.K.

437   −185   2,439

686   Other

388   Administrative overhead (estimate U.S. 24%; U.K. 16%)

1,113   Prices of inputs (U.S. ~75% higher)*

Inputs used (U.S. ~30% higher)

U.K.   U.S.

Germany vs. U.K.

226   65   −324

393   Prices of inputs (Germany ~25% higher)*   Administrative overhead (estimate Germany 13%; U.K. 16%)   Other   1,473

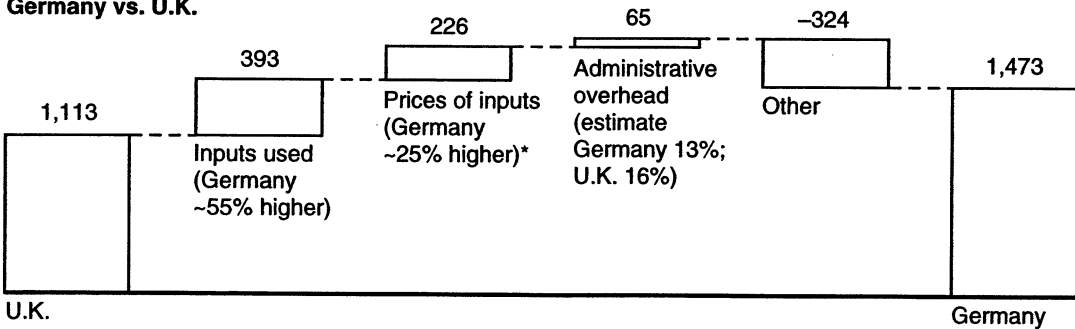1,113   Inputs used (Germany ~55% higher)

U.K.   Germany

* OECD price index used, which was directionally consistent with relative prices obtained using physician salaries and prices of pharmaceuticals and inpatient hospital stays

Note: Input levels, price levels, and administrative costs of the country with the higher spending level were used as a base for determining the expected spending of the second country and the corresponding spending gap in the above calculation. Relative input levels were determined using information on the amount of labor (physicians, nurses, and medical technicians), supplies (pharmaceuticals), and capital consumed in each country, assuming a breakdown of 70% labor, 20% supplies, and 10% capital; relative prices were determined using a price reported by the OECD; administrative costs were estimated as 24% in the U.S., 13% in Germany, and 16% in the U.K. based on available literature; education costs are not included

Source: BASYS; McKinsey analysis

022 ST 140106/ES

Exhibit 6

CAUSAL ANALYSIS OF PRODUCTIVITY DIFFERENCES

● Major differentiating factor
◐ Secondary differentiating factor
○ Undifferentiating

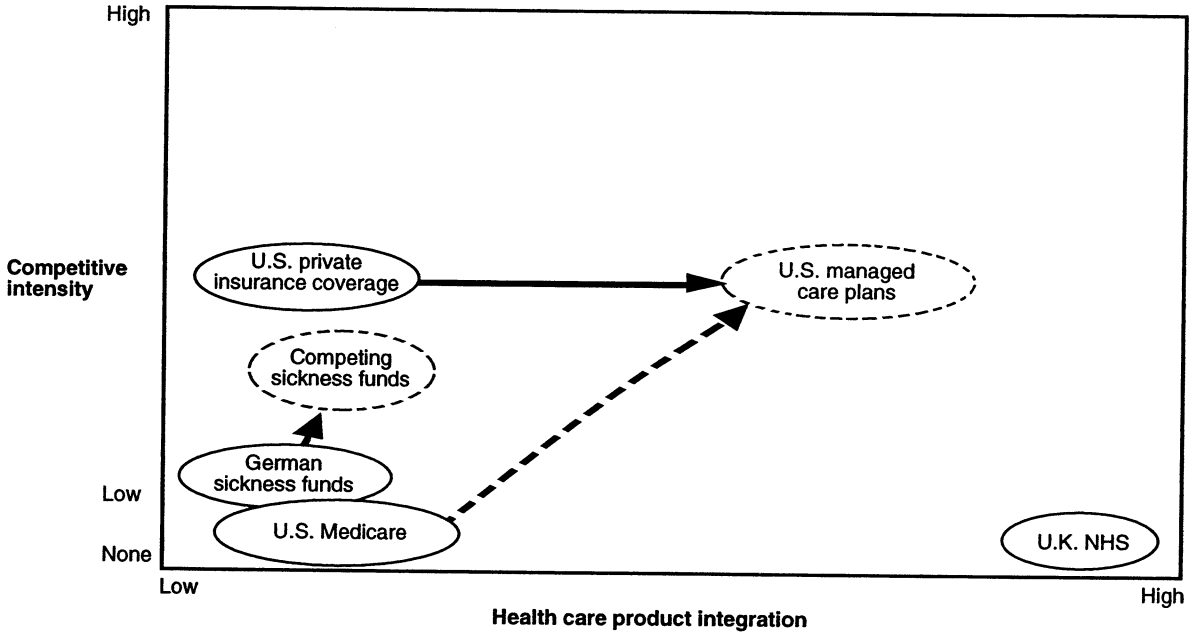|	U.S. vs. U.K.					U.S. vs. Germany			
	Diabetes	Choleli-thiasis	Breast cancer	Lung cancer	Overall	Choleli-thiasis	Breast cancer	Lung cancer	Overall
Benchmark	U.K.	U.S.	Indeter-minate	U.S.		U.S.	U.S.	U.S.	
Productivity									
• Inputs	●	●	●	●	●	●	●	●	●
• Outcomes	●	●	◐	●	●	◐	●	◐	◐
Provider behavior									
• Care triaging	●	○	●	◐	◐	○	○	◐	○
• Treatment duration	○	◐	◐	◐	◐	●	●	●	●
• Staffing levels	○	◐	◐	◐	◐	◐	◐	◐	◐
• Setting choice	○	○	◐	◐	◐	○	●	●	◐
• Team-based approach	●	○	○	○	○	○	○	○	○
• Technology adoption	○	●	●	◐	◐	◐	○	○	○
Provider incentives and constraints									
• Physician incentives	●	●	●	○	●	◐	◐	◐	◐
• Hospital incentives	○	◐	◐	◐	◐	●	●	●	●
• Physician supply	●	○	◐	◐	◐	○	○	○	○
• Hospital supply	○	◐	◐	○	◐	●	●	●	●
• Capital constraints	○	◐	●	◐	◐	○	○	○	○
• Substitution constraints	○	○	○	◐	○	○	●	●	◐
Health care system structure									
• Product integration and pricing	●	◐*	◐	○	◐	●	●	●	●
• Competitive intensity	●*	●	●*	◐	●	◐	◐	◐	◐
Regulation	●	●	●	●	●	●	●	●	●

* In these cases, a higher degree of the respective causal factor led to lower productivity

Exhibit 7

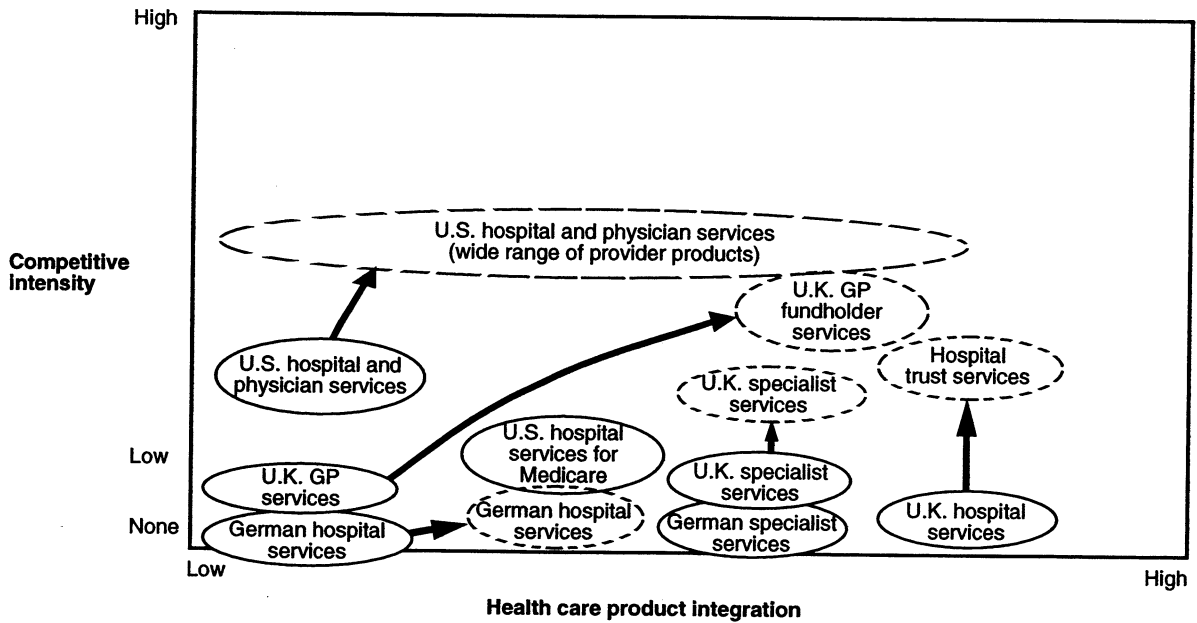## RECENT CHANGES TO HEALTH CARE MARKET STRUCTURES

◯ Mid to late 1980s

⬭ Current

Health coverage market

High

Competitive intensity

U.S. private insurance coverage → U.S. managed care plans

Competing sickness funds

German sickness funds

Low

U.S. Medicare

None

U.K. NHS

Low — High

Health care product integration

Care provision markets: between insurers and providers

High

Competitive intensity

U.S. hospital and physician services (wide range of provider products)

U.S. hospital and physician services

U.K. GP fundholder services

Hospital trust services

U.K. specialist services

Low

U.S. hospital services for Medicare

U.K. GP services

U.K. specialist services

German hospital services

German specialist services

U.K. hospital services

None

German hospital services

Low — High

Health care product integration

Chapter 2: Summary of health care system structure and regulation

This chapter describes the structure of the health care systems of Germany, the U.K., and the U.S. in order to provide a frame of reference for our examination of the underlying drivers of the productive efficiency differences observed in the disease cases (Chapters 3 through 6) and the cross-disease synthesis (Chapter 7).

We begin by outlining a framework for describing the structure and dynamics of any health care system in terms of the economic interactions that occur among system participants and how these interactions are shaped by regulation. We then use this framework to describe the predominant health care systems existing in Germany, the U.K., and the U.S. at the time of our assessment (approximately 1985 to 1991) and conclude by summarizing the key differences among these three systems. Throughout this chapter, we also comment on major changes in each of these systems since 1991.

FRAMEWORK FOR DESCRIBING HEALTH CARE SYSTEM STRUCTURE AND REGULATION

Health care is similar to other goods and services in that it is produced with resources that are scarce relative to competing human wants. Every country must have mechanisms, explicit or implicit, for determining how much health care to produce, how to produce it, and how to distribute it across the population. In this way, "markets" do exist for health care services in one form or another in all health care systems. In structuring and regulating their health care markets, the central challenge for policymakers is achieving economic efficiency as well as an acceptable level of social welfare.

It is important to recognize, however, that the provision and payment for health care services in any country or system is not the result of a single, simple market transaction between buyers and sellers. Rather, it is characterized by multiple exchanges of health care products and services between many different system participants. The presence of (and felt need for) health insurance coverage in most health care systems – given both uncertainty in the incidence of illness and social welfare considerations – creates the need for a risk-bearing intermediary (or payor) between the buyers (i.e., consumers or employers) and the sellers (i.e., providers) of medical services. This, in turn, creates interim products and services within the health care system. Additionally, the complexity and highly

emotional content of many medical decisions, constantly evolving medical technology, and uncertainties about the efficacy of various treatments often put the physician in the position of the prime decision maker (or "agent") for the patient (consumer) in the consumption of individual medical services.

The provision and payment for health care services can therefore best be described as an interdependent set of economic interactions, explicit or implicit, for different health care-related products and services (including health insurance coverage and care provision services) that occur among various health care system participants. These participants include consumers and/or employers, payors or other intermediaries, hospitals and other institutions, and physicians. In many systems, the government or other central authority plays an active role either as one or more of these participants, or by regulating one or more of these interactions. Furthermore, government can directly regulate supply.

The specific interactions that exist in any health care system and the product or service that is exchanged include (see **Exhibit 1** for a graphic illustration):

1. Interactions between consumers (or employers) and payors for *health care coverage*.

2. Interactions between payors and providers (including hospitals or other institutions, physicians, and other providers) for *care provision services*, including the guarantee of payment for care provision and possibly the guarantee for actual care provision services.

3. Interactions between consumers (or patients) and providers for *care provision services*, including hospital and physician services.

Each of these interactions can be highly regulated, market-based, or a blend – depending on how policymakers choose to structure them to achieve their economic efficiency and social welfare goals. And although these markets are highly interdependent, each can be structured and regulated very differently in a health care system. Many health care policymakers try to combine the strengths of market-based approaches in increasing economic efficiency with the strengths of regulation in correcting for market failure and ensuring social welfare. Different countries have arrived at very different blends and many are continuing to experiment.

The nature of these markets and their associated economic interactions *collectively* create specific incentives and constraints for providers, which in turn drive different care treatment approaches and result in different levels of productive efficiency across systems. Therefore, a critical step in examining the differences in productive efficiency of health care systems across countries is to understand how each interaction is structured and the specific incentives and constraints that result.

To do this, we use two major descriptive factors, both of which can be influenced through regulation: the level of competitive intensity in the market, and the degree of integration and pricing mechanisms in the health care products.

Competitive intensity

The level of competitive intensity in the market will depend on the relative concentration of "buyers" and "sellers," on ownership structure (e.g., government-owned or private, for-profit or not-for-profit), on the degree of flexibility allowed in the interaction, and on information transparency and symmetry. While competitive intensity can create incentives for system participants to improve performance, health care markets are often characterized by lower levels of competitive intensity than markets for other goods and services because of the high propensity for market failure. Specifically, asymmetrical information exists between system participants, since providers often have more information about treatment than consumers and payors; in addition, consumers are often insured against some or most of the costs of care. Furthermore, social welfare considerations often motivate governments to intervene in health care markets in ways that limit competitive intensity.

Degree of integration and pricing mechanisms

The second factor is the degree of integration and pricing mechanisms in the health care products exchanged in the market, with specific reference to the nature and type of "bundling" of individual health care procedures or services into larger units or packages. This factor is important in health care given the complexity of medical care and disease treatment processes and the fact that consumers rarely have sufficient knowledge of medical conditions and treatment options to act as their own care integrators. Providers and/or payors, therefore, play a significant role in assembling and packaging health care services, and the nature and extent of this role, as well as the product forms resulting, vary considerably (**Exhibit 2**). In addition to taking on different forms, health care products can be priced in very different ways by payors, providers, or government regulators.

The degree of product integration can be used to describe products in both the health coverage and care provision markets. In the care provision market, any package of services that providers offer as a single unit to payors for a specified price represents a more integrated product. In the health coverage market, the payor can actively coordinate the services of specific providers within a number of care settings and thus achieve a degree of integration. Importantly, product integration does not imply a single, vertically integrated payor/provider entity such as the U.K. National Health Service (NHS). Rather, health care product

integration refers to the integration of specific care elements and can therefore be achieved through the actions of independent payors, providers, or both.

When classifying degrees of integration in the health coverage market, it is important to distinguish between the degree of *product* or *care integration* offered by a payor and the extent of the *insurance coverage* offered by that payor. Health insurance pools the risks for multiple diseases across multiple consumers, while product integration refers to the assembly of health care services for a single individual. For example, while traditional indemnity insurance provides coverage for a wide range of diseases over a fixed time period (typically 1 year), it provides little or no health care product integration.

Health care product integration can be achieved along two dimensions: the breadth of the disease or diseases covered in the package over time, and the scope of the care components included, such as physician services, hospital services, and pharmaceuticals. Using these two dimensions, the degree of health care product integration can be described along a spectrum, ranging from very low to quite high (**Exhibit 3**).

¶ For example, the lower left corner represents *fee-for-service (FFS)* care products and the least integrated level of care, since these products involve providers delivering specific medical services to a patient in specific encounters. This would include FFS products offered by hospitals or physicians, and traditional indemnity-style health coverage offered by payors to consumers or employers, in which consumers select their own providers and are relatively free to determine their own treatment paths.

¶ Moving further out on the integration spectrum would be *hospital case rates*, in which hospitals bundle all of the hospital services required to treat the acute case or the acute phase of disease. With this product, hospitals receive a single payment for the case, regardless of the intensity of services provided or the number of days the patient stays in the hospital. Such products can also include physician as well as hospital services, representing a higher degree of integration.

¶ *Disease carve-out* products offered by providers to payors or by payors to consumers represent an even higher level of integration. In this product, the provider or payor bundles a range of care services for all occurrences of a disease over time (1 year or multi-year), and prices it as an integrated package. Such products can only be constructed for those diseases that are relatively well-defined and distinct from other diseases, such as some cancer care.

¶ Even further out on the integration spectrum are *bundled care services for multiple diseases over time* (1 year or multi-year), which address the

overall health of the patient over time and, therefore, include diagnostic and preventive care by a range of providers in addition to disease treatment by specialists and hospitals. Such products generally take the form of a prepaid, "capitated" payment per patient from payors to provider organizations, for which the providers agree to provide directly or to coordinate the full range of care to the defined patient population. This would also correspond to managed care health coverage products such as health maintenance organizations (HMOs) in which payors accept a prepaid capitated amount per member from employers or consumers, and coordinate and manage care for these members. Such products can also be constructed to cover the lifetime of the population rather than a single or multi-year time period, as in single payor systems such as NHS coverage in the U.K.

In addition to the examples described above, care products can take on various alternative forms and pricing mechanisms along the two dimensions of the breadth of diseases covered and the scope of care components included in the package (see again **Exhibit 3**).

* * *

The *level of competitive intensity* and the *degree of product integration* are important factors in understanding the structure of a health care system and the incentives and constraints created for its participants, particularly providers. In the remainder of this chapter, we use this framework to describe the structure of the three health care systems included in this study – Germany, the U.K., and the U.S.

For each country, we discuss the degree of product integration and level of competitive intensity of the three major health care markets: 1) interactions between consumers and/or employers and payors in the health coverage market; 2) interactions between payors and providers for care provision services; and 3) interactions between consumers and providers for hospital and physician care provision services. We also comment on the overall regulatory environment and its impact on the health care system structure in each country. While we focus on system structure at the time of our assessment (approximately 1985 to 1991), we also comment on major changes that have occurred since that time.

Overall, there are substantial differences in the structure of the three health care systems. The U.S. was the most market-based system (except for the government-controlled Medicare and Medicaid programs), with substantial degrees of freedom for payors and providers; the U.S., therefore, had relatively moderate to high levels of competitive intensity, but generally low product integration. The U.K. system was the most centrally controlled of the three and had low competitive intensity but high product integration. And while the German system was highly regulated, it had relatively low levels of

both competitive intensity and product integration (**Exhibit 4**). We now turn to a detailed description of each of these health care systems and the specific markets within them.

THE GERMAN SYSTEM[1]

The health care system in Germany, until recently, was strongly regulated. Ninety percent of consumers were insured through public, nonprofit payors known as "sickness funds," with the rest covered by private insurance. The public system was with few exceptions jointly funded (50/50) by consumers and employers and ensured coverage for everyone. The sickness funds were required by law to offer a precisely defined, comprehensive package of goods and services to all members. Most of the sickness funds faced little competition given that they were segmented by geography or consumers' profession and had to contract with all hospitals and physicians within a defined region. And while most sickness funds covered patients for extended periods of time, neither payors nor providers played an active role in integrating care for patients.

Using the framework, the markets in the German health care system and regulatory environment in the mid to late 1980s can be summarized as follows:

1. *Interactions between consumers and/or employers and payors for health care coverage:* low degree of product integration, given that payors paid medical expenses without assembling care; low competitive intensity, with payors required to offer a government-defined benefit package to all members and membership restricted by geography or profession, with limited or no member choice.

2. *Interactions between payors and providers for care provision services:*

   - *For inpatient care provision services:* low degree of product integration, with providers receiving per diem payments from payors; low competitive intensity, given that all payors had to contract with each hospital on virtually the same terms.

   - *For outpatient care provision services:* low degree of product integration, with ambulatory physicians essentially receiving FFS reimbursement; low competitive intensity as payors had to contract collectively with all physicians in a region on similar terms.

3. *Interactions between consumers and providers for care provision services:* low degree of product integration, as providers offered specific hospital or physician services to patients (coordinated only through the referral system); moderate competitive intensity, as consumers had some choice of physician and hospital.

---

[1] The basic sources used for this section (except where specifically noted) include the OECD, *Statistisches Bundesamt*, Wissenschaftliches Institut der Ortskrankenkassen (WIdO), National Economic Research Associates (NERA), and interviews.

These interactions were strongly influenced by *regulation* in Germany, which dictated the segregation of hospital-based and outpatient care, as well as defined the political processes through which both hospital and physician supply were controlled.

In 1996, several changes were introduced to increase the competitive intensity of the system at the payor level and to increase integration of hospital care to some extent. All consumers can now choose among any public sickness fund and are able to switch annually, forcing payors to compete for members. And hospitals are no longer paid strictly a per diem rate, as case rates have been introduced for about 15 to 20 percent of cases.

In the remainder of this section, we provide more detail on each of the three interactions listed above, as well as on the regulations shaping these interactions.

1. Germany: interactions between consumers or employers and payors for health care coverage

As stated, 90 percent of consumers were insured through public, nonprofit payors (*Körperschaften Öffentlichen Rechts*), known as "sickness funds" (or *Krankenkassen*), while the rest were covered through private insurance. About 7 percent of sickness fund members also bought supplementary private health insurance (e.g., single room hospital coverage). Most consumers had to choose from a limited number of sickness funds offering a standard product, as these funds for the most part operated in relatively small geographic areas or were restricted to certain professions. Most members stayed with their sickness fund for their lifetime, and coverage included all dependent family members until they were eligible to become members in their own right. Consumers and employers made equal contributions to the funds.

There were three main groups of sickness funds: the general regional sickness funds (AOK), collectively covering 30 million consumers; the company sickness funds (BKK) set up by individual companies (such as the Siemens BKK); and some large federal sickness funds (such as the *Barmer Ersatzkasse* [BEK], *Deutsche Angestellten Kasse* [DAK], and *Techniker Krankenkasse* [TKK]). There were approximately 60 private health insurers, whose share was more fragmented.

Payments to the sickness funds, paid half by the consumers and half by the employers, comprised both an "insurance premium" and income redistribution. This contribution rate, calculated as a percentage of the consumer's gross income, was determined based on the expected medical and administrative costs of the sickness fund in that year. It generally varied between 11 and 14 percent.[2] As no

---

[2] This percent contribution rate was applied to the gross income of each employed family member, up to a maximum of 72,000 deutsche marks [DMs]; income above this amount was not taxed; 50 percent of this contribution was paid by the employee and 50 percent by the employer.

significant reserves were built, the contribution rates were adjusted to account for decreasing or increasing costs. Thus, while payors were essentially precluded from directly sharing financial risk with consumers or employers, they could easily pass on higher unit costs to employers and consumers when calculating their next year's contribution rates.

Degree of product integration: low. The sickness funds, by law, had to offer an indemnity-like health coverage product in which the funds were essentially bearing the financial risk for the incidence of illness and for the cost of treatment for their defined population. All sickness funds had to offer essentially the same comprehensive product of reimbursable goods and services; this product was precisely defined by law and included hospital, physician, and dental services, pharmaceuticals, medical supplies, medical devices, vision care and eye glasses, and even wage substitution for sick leaves longer than 6 weeks. There were virtually no co-payments or deductibles for all major services.

Private payors had to offer a medical package basically identical to that of public payors. However, they had some flexibility to cover additional services (e.g., one-bed room at the hospital, chief physician's consultation at the hospital, dental fillings of gold/ceramic). Private payors could also shift more risk to consumers through co-payments and deductibles.

While sickness funds kept members for extended periods of time, they did not play an active product integration role. Payors did not intervene in care delivery, but rather paid all medical expenses incurred by their members according to the defined government benefit package and co-payment levels. The primary payor role in interactions with consumers and employers was, therefore, to process and pay claims. Given that German payors faced little annual turnover in their membership, however, they had some incentive to invest in and fund medical care with longer-term (vs. near-term) benefits.

Level of competitive intensity: low. Prior to the recent reforms, there was very little competition in the interactions between consumers or employers and payors. Most payors were (and still are) public, and the level of flexibility in product design and pricing was extremely low. The law defined who must insure, who could select which payor, the benefits payors had to offer, and the pricing mechanisms that could be employed.

Everyone living in Germany had to be insured at the start of employment or higher education. The payor had to accept every applicant and to offer the government-defined benefit package to the member for his or her lifetime. The payors from which a consumer could choose depended on his or her employment status. High-earning, white-collar employees could opt out of the public system and insure with a private payor. The remaining white-collar employees could choose among all the public sickness funds with few restrictions (e.g., TKK was only for members of technical professions). Traditionally, blue-collar employees could choose from their employer's BKK (if one existed) or their regional AOK.

While all payors were required to offer the same benefits package, they were able to compete to some extent through risk selection (e.g., convincing the lower risk, high-earning, white-collar workers to join). To reduce the incentive for sickness funds to seek good risks while maintaining some incentive to manage medical cost, a risk equalization scheme was established in 1994 based on age, sex, income, and number of family members. This scheme, however, has not been completely effective due to the difficulties of predicting and adjusting for differential risk. In contrast to public sickness funds, private payors charged a risk premium per insured person, which was differentiated by age of entry and sex of the insured.

As of 1996, significant changes were introduced in this interaction. All blue- and white-collar employees can now choose among any public sickness fund and switch annually, increasing competitive intensity by forcing funds to compete for members on price and other factors for the first time as well as to compete more aggressively on risk selection.

2. Germany: interactions between payors and providers for care provision services

As defined by law, payors purchased inpatient hospital and physician services through negotiations with hospitals, with payments for inpatient physician services included in the per diem hospital payments. Outpatient services were purchased through negotiations with regional associations of ambulatory physicians. Both of these interactions were strongly regulated.

For inpatient services, payors were required to negotiate as a *collective group* vis-à-vis each regional group of the 2,000 hospitals, and payors had very limited information and intervention rights.[3] Of Germany's 2,000 hospitals, 45 percent were government owned, 16 percent were privately owned, and 39 percent were owned by nonprofit institutions such as the Red Cross. About 92,500 hospital physicians were employed by these hospitals on a salaried basis, with a few exceptions for attending ambulatory physicians in hospitals, particularly in rural areas or smaller hospitals.

For outpatient services, payors negotiated with West Germany's 74,000 ambulatory physicians in private practice, largely through 19 regional public associations of ambulatory physicians – the *Kassenärztliche Vereinigungen* (KVs).

We discuss inpatient and outpatient care provision services in turn.

Inpatient care provision services

---

[3]   Information reflects West Germany in 1990.

Degree of product integration: low. Per diem payments were made from payors to providers for a hospital-day to cover both hospital and hospital-based physician services. Until 1996, the contract between all payors and a hospital essentially stipulated this per diem payment and a budgeted utilization (i.e., implicitly, a DM global hospital budget). However, in case of a loss (or profit) by a hospital, the per diem of the current year was adjusted to compensate for some of the loss (or profit) of the previous year, when it was determined that utilization was the cause. Thus, hospital services were managed primarily at the per-day level, with little incentive to integrate care or manage costs across the entire hospital episode (or case) or across multiple cases.

The disincentives to integrate care and manage costs that resulted from per diem hospital payments were magnified by hospital-based chief physician incentives to increase activity levels. While most hospital physicians were salaried, a department's chief physician obtained significant additional income from private consultations in the hospital. By law, chief physicians were exclusively allowed to negotiate a contract with a hospital that allowed them to offer services to privately insured patients in the hospital. Income derived from these additional consultations sometimes amounted to multiples of the physician's public salary. Because the number of private beds in a department was often related to the number of public beds, which in turn depended on the budgeted utilization rate (see above), there was a strong incentive for the chiefs, as key influencers of discharge decisions, to maintain high activity levels.[4]

In 1996, however, a change was introduced to increase the level of inpatient product integration and encourage hospitals to behave more economically. The uniform per diem rate per hospital was replaced by a prospective payment (for approximately 15 to 20 percent of the total, mostly surgical, procedures performed), department-specific per diems (for approximately 80 to 90 percent of the total procedures performed), plus a hospital-specific room rate. Gradually, more procedures are expected to be covered by such prospective payments. In addition to these reforms, further changes designed to better integrate inpatient care and manage hospital spending are under discussion; the most far-reaching proposal is to budget total hospital expenses using the previous year's budget with an adjustment for inflation.

Level of competitive intensity: low/none. There was (and still is) essentially no competition between hospitals and payors in their negotiations for care provision services. There was no flexibility allowed in individual agreements between payors and a hospital on payment terms, information rights, or ways of intervening in the care process. All payors had to accept claims from every hospital and had to negotiate as a collective group for hospital payment terms;

---

[4] While there was no official policy regulating the number of private beds relative to the number of public beds, university hospitals usually allocated one for every five and community hospitals one for every seven.

although the rate could and did vary by individual hospital, all payors negotiated the same rate with a given hospital. Even large payors (the share of the largest payor with a hospital was frequently as high as 30 to 50 percent) had to operate under the joint contract. Thus, there was no opportunity for selective contracting with hospitals on the part of payors, and therefore no opportunity nor incentive for the payors to compete with one another.

In addition, there was a significant lack of information as well as asymmetry in this interaction. The data that hospitals had to disclose to payors were very limited and payors' intervention rights were highly limited. Until 1996, the data disclosed were primarily cost and rough utilization information; payors therefore had little leverage with hospitals in per diem and budget negotiations, despite acting as a joint buying group. While payors had the right to request a hospital productivity review by an external evaluator or consultant, the result of these reviews generally had little or no impact.

As well, medical utilization review was a shared service of all sickness funds and was limited in scope and effectiveness; neither payors nor the utilization review service had access to medical records. Utilization review was generally performed via questionnaires sent to hospital physicians. Furthermore, sickness funds had little incentive to decrease hospital activity levels, given that only 25 percent of the hospital reimbursement depended on actual utilization (so called "flexible budgeting"). Length of stay (LOS)-reducing efforts would simply result in a hospital generating less income, which could require the payor to increase the per diem payment for the next year to account for some of the budget shortfall from the prior year.

In 1996, two changes were introduced that may increase to some extent the level of competitive intensity between payors and providers: 1) profits and losses are no longer fully compensated through automatic per diem adjustments; and 2) payors are able to obtain more data on patients, although less data on cost. The impact of these changes is difficult to predict.

Outpatient care provision services

Degree of product integration: low. At the highest level, the outpatient care provision product was an aggregate annual budget for ambulatory physician services across all diseases negotiated annually between a KV and a sickness fund. Physicians in a KV were thereby effectively capitated as a group for all ambulatory physician costs, and thus bore some risk for the cost of ambulatory physician services if activity levels exceeded budgeted amounts. However, the annual contract negotiated with the payors also implicitly determined a price per activity and an expected activity level. All reimbursable physicians' services were listed in a catalogue, which assigned a value to each service according to the "unified value scheme" *(Einheitlicher Bewertungsmaßstab*, or EBM). The fees in the EBM catalogue were set at the federal level in negotiations between the association of all sickness funds and the association of all KVs. If the activity

level was greater than budgeted, the unit value for a service listed in the EBM catalogue fell in direct proportion to the overutilization (e.g., if the activities delivered were double the contracted number, the price per service unit fell by 50 percent).

Therefore, while ambulatory physicians were collectively capitated through the global budget of the KV, each physician had incentive to increase his or her own activity in order to gain the largest possible share of the KV budget. Although in many KVs the regional KV budget was broken down by specialty, this step still did not create meaningful incentives for individual physicians to manage activity levels or to coordinate care with other physicians. As payors had virtually no opportunity to assemble care beyond providing individual services, the degree of product integration was low.

For the private insurers, members' claims for ambulatory physician services were reimbursed as long as they fell within a specified range above the amount that public sickness funds paid for the same services. Similar to the sickness funds' payment schemes, these FFS payments did not provide physicians with an incentive to assemble care, but rather, to retain the private patient.

Level of competitive intensity: low/none. There was no meaningful competition in payor-physician interactions for outpatient services. Virtually all ambulatory physicians had to become members of KVs. Each KV covered a particular region, and an annual budget per payor for all physicians in that region was fixed and distributed through the KV to physicians according to the nationally regulated, activity-based scheme described above. The KVs contracted on behalf of their member physicians with each individual payor. Payors could not select physicians; once a member of a KV, the physician was automatically accredited by all sickness funds that contracted with that KV. In addition, the structure of the contract between a KV and a sickness fund was defined by law, as described above. Furthermore, the regional KV also regulated ambulatory physician capacity, within ceilings set by the state government on the number of new physicians that could be accredited.

Finally, neither payors nor consumers systematically accessed claims data. The KV collected activity data coded by physicians, evaluated the payment depending on the budget and the activity level, compensated the physicians, and received compensation from the sickness funds. From time to time, the sickness funds conducted audits at the KVs; these audits, for example, helped enforce the use of generic versus branded prescriptions. However, the KVs were responsible for all coding and utilization control; this significantly limited the ability of the sickness funds to gain access to, analyze, and distribute to its members information on physician quality and cost.

3. Germany: interactions between consumers and providers for hospital and physician care provision services

Degree of product integration: low. The care provision products exchanged in the consumer-provider interaction were the specific hospital-based or ambulatory physician services provided to the patient. While ambulatory physicians cared for patients over time and over diseases, their integration role was limited to referring patients to other physicians for specialized care, primarily through the exchange of letters. In particular, there was a weak "gatekeeper" function for administrative reasons: a sickness fund issued *one* certificate of coverage for a member, which the consumer gave to a self-selected physician; if the consumer wanted to see another ambulatory physician or specialist, he had to obtain a referral from the holder of his certificate. Recently, the certificate was replaced by a smart card, simplifying "patient tourism."

Level of competitive intensity: moderate. Ambulatory physicians and hospitals were – from the perspective of the consumer – highly fragmented, and consumers had significant freedom in choosing among them with virtually no cost impact. Although this interaction was somewhat regulated, regulations were not enforced and considerable flexibility existed. For example, while consumers were supposed to choose one of the two hospitals nearest to their homes, hospitals admitted any patient and the cost was covered by payors.[5] For consumers, however, information transparency was rather poor regarding physician and hospital quality. To select a hospital for an elective visit, consumers relied mostly on the counsel of their ambulatory physician, who referred to his or her own "network."

Providers had strong incentives to increase activity levels and to attract more patients; they aggressively built networks of referring physicians and hospitals and attempted to differentiate themselves on service, quality, perception, and other dimensions.

Overall, the level of competitive intensity between providers and consumers was, therefore, moderate.

Germany: regulation

In addition to the regulatory influences on the various markets described thus far, the German government also imposed other influential regulations on the health care system. The state regulated hospital capacity levels and, as described above, the KV regulated ambulatory physician capacity. In addition, Germany's physicians were strictly segregated by law into two separate sectors –

---

[5]   In 1994, however, a small patient co-payment (10 DM per hospital-day) was implemented.

ambulatory and hospital-based physicians – and were not encouraged to coordinate their activities across inpatient and ambulatory care settings.

Capacity regulation. Hospital capacity was regulated by the regional government and influenced by payors; if a hospital's utilization fell below 85 percent, capacity was reviewed and possibly adjusted. Thus, each hospital department had the incentive to keep utilization at 85 percent or higher to avoid capacity cuts. Hence, the hospital capacity utilization across German states was 85 percent, irrespective of significant variations in the number of beds per capita.

Despite their role as regulators of hospital capacity, the 11 state governments did not bear any risk for hospital costs. Rather, the higher costs that could stem from maintaining higher capacity were paid by employers and employees of other states through the sickness funds. Thus, a significant reduction in hospital capacity would simply result in local job losses and, therefore, eventually the need for local social security payments. Given this, German state governments had incentives to increase, rather than decrease, hospital capacity levels in their region.

Segregation between ambulatory and hospital-based physicians. As a result of the strict segregation between ambulatory physicians and hospital-based physicians, there was virtually no systematic exchange among providers of aggregated information on cost or other performance dimensions. In addition, crucial information on patient records flowed slowly between hospitals and ambulatory physicians. When ambulatory physicians referred a patient to the hospital, they simply mailed a letter stating the diagnosis to the hospital. Conversations between hospital and ambulatory physicians were usually rare and brief. As a result, hospitals often repeated diagnostics already performed by the ambulatory physician. And while hospitals were supposed to inform the ambulatory physician about the patient's condition upon discharge, this was usually done by letter, often weeks after discharge. This practice created the opportunity for missed follow-up and may have resulted in unnecessary readmissions.

There were also some perverse incentives created by this ambulatory and inpatient segregation. In particular, a general practitioner (GP) that referred a patient to an ambulatory specialist frequently lost the patient; therefore, he or she had an incentive to refer a patient to a hospital rather than to an ambulatory specialist who could perform the procedure in an outpatient setting at a lower cost. For example, a hospital usually referred diabetics back to their GP after discharge; a diabetologist might not have done so. Very recently, some payor/KV pilots were launched to give physicians greater incentives to manage patients across the inpatient and ambulatory care settings and to share some hospitalization risk; the impact is yet to be determined.

* * *

In summary, the German health system was characterized by low levels of both competitive intensity and product integration in the mid to late 1980s. This system structure – which was strongly influenced by regulation – created strong activity-increasing incentives, especially for hospitals, and imposed significant supply constraints on capacity and substitution of alternative care settings (**Exhibit 5**).

THE U.K. SYSTEM (PRE-NHS REFORMS)[6]

The U.K. has had a centrally planned NHS since 1948, providing lifetime coverage for the entire U.K. population. Prior to the 1991 reforms, providers were generally tightly integrated with the district payor entities. There was a fixed budget, supply was carefully controlled, and there was little, if any, competition among providers for patients or for payor contracts. Given that the NHS was a single payor and specialist physicians were employed by the hospitals, there was a relatively high degree of health care product integration throughout most of the system.

The markets in the U.K. health care system and the regulatory environment during the mid to late 1980s can be summarized as follows:

1. *Interactions between consumers and/or employers and payors for health care coverage*: high degree of product integration and no competitive intensity, as the NHS provided comprehensive health care coverage to the entire U.K. population.

2. *Interactions between payors and providers for care provision services:*

    - *For hospital and specialist care provision services:* moderate/high degree of product integration and low competitive intensity, given that the entire U.K. hospital system was funded and managed by the government and that specialist physicians were required to practice in designated hospitals.

    - *For primary care provision services:* low degree of product integration, with GPs working in small, independent practices; low competitive intensity given that GPs were generally assigned a patient population to serve through the NHS.

3. *Interactions between consumers and providers for care provision services:* moderate degree of product integration given the strong gatekeeper role of the GPs; low competitive intensity, with consumers having little choice of providers, given supply constraints.

These interactions were strongly influenced by *regulation* in the U.K., which imposed an overall budget constraint, centralized control of physician, hospital, and capital supply, and established rigid referral/gatekeeper and capital allocation processes.

In more recent years, the NHS has been supplemented by a small, but rapidly growing, private insurance sector, which in 1989 covered about 11 percent of the

---

[6] The basic sources used for this section (except where specifically noted) include the OECD, the NHS, NERA, an article in *The Milbank Quarterly* (Ham C. Population-centered and patient-focused purchasing: The U.K. experience. The Milbank Quarterly 1996; 74: 191-214), and interviews.

population and accounted for about 4 percent of health expenditures. In addition, reforms since 1991 have introduced significant market mechanisms and competition at the local payor-provider level, along with opportunities for greater physician and hospital autonomy. While the description that follows focuses on the pre-reform period, it highlights major changes that have occurred since 1991.[7]

In the remainder of this section, we provide more detail on each of the three interactions listed above, as well as on the regulations shaping these interactions.

1. U.K.: interactions between consumers or employers and payors for health care coverage

Degree of product integration: high. The NHS was (and is) a highly centralized, near-monopoly payor that provided comprehensive health care coverage to the U.K. population. All citizens of the U.K. received the same package of lifetime care from the NHS, which included coverage for a broad range of GP and hospital services. Even citizens with private insurance coverage regularly used the NHS for at least some services, such as GP services and complex hospital care. In addition, the NHS exercised significant influence over care provision, directly providing many services (see payor-provider interaction description). The NHS thus served as the prime care integrator for the majority of the U.K. population.

Level of competitive intensity: none. The interaction between the NHS and consumers was highly regulated, with no market influence. The NHS was publicly funded through general taxation (79 percent of NHS expenditures), national health insurance contributions for all employed persons based on income (16 percent), and charges to consumers (5 percent) through small co-payments on prescription drugs, dental services, and vision services. Consumers had little choice but to fund the NHS through the general taxation fund and national health insurance contributions.[8] While employers acted as "tax collectors" for these contributions, they did not play an active role in determining the level or nature of NHS coverage.

Since the NHS offered a standard set of health care services to all consumers, consumers had very little choice in their health care coverage. Consumers' only

---

[7] Separate settlements were agreed for England, Wales, Scotland and Northern Ireland: the description that follows applies predominantly to England where 83 percent of the U.K. population reside, consuming 81 percent of NHS expenditure (revenue and capital).

[8] In 1988 and 1989, general taxation funds received from the Treasury accounted for 79 percent of NHS expenditures while mandatory national health insurance contributions accounted for 16 percent. Employed persons were required to make mandatory health insurance contributions of 0.95 percent of earnings while employers contributed an additional 0.80 percent. Self-employed individuals were required to contribute 1.75 percent of earnings.

choices within the NHS involved making additional payments for a few select amenity services such as private rooms. And, given that changes to the NHS were generally brought about only through political channels, it was difficult for consumers to directly influence NHS coverage.

Some consumers supplemented their NHS care with private health insurance, which provides, for example, shorter wait times for elective surgeries.[9] Both employers and individuals were free to purchase private insurance, which operated in a relatively competitive market. Even in the private insurance market, however, there was a relatively high degree of concentration, with the British United Provident Association (BUPA) commanding the largest share. As private market competition has increased, consumers have had broader product and pricing choices.

2. U.K.: interactions between payors and providers for care provision services

Before the 1991 reforms, the interaction between the payor function of the NHS and NHS providers was internalized, with payment and provision essentially integrated. Public health expenditure budgets were set annually for the U.K. by the Cabinet (national government). These funds were then allocated to the Department of Health separately for hospital and community health services and for family health services (GPs, outpatient prescription drugs). The Department of Health, in turn, funded 14 regional health authorities (RHAs) for hospital, specialist, and community services, and 90 Family Practitioner Committees to contract with independent GPs. We discuss hospital and specialist care provision services and primary care physician services in turn.

Hospital and specialist care provision services

Degree of product integration: moderate/high. The entire U.K. hospital system was nationalized and regionalized in 1948 with the creation of the NHS. Prior to the 1991 reforms, the RHAs funded approximately 190 district health authorities (DHAs), each of which was responsible for hospital and community health services for between 100,000 and 800,000 people. The DHAs funded and managed the approximately 1,720 public British hospitals, including over 250,000 beds (5.2 beds per 1,000 population). Independent from the NHS were over 200 private hospitals with almost 11,000 beds. The DHAs allocated funds to hospitals by setting an annual budget for each hospital in their district. Contracts between the DHA and out-of-area hospitals were negotiated only if required services could not be provided locally. In addition, specialist physicians were salaried and employed by the RHAs and required to practice in their designated hospital. Some specialists supplemented their income by also practicing outside

---

[9]   Premiums for private insurance were age-adjusted and for group health products, experienced-rated.

of the NHS hospital, in the FFS private market. Furthermore, RHAs received and funded all capital projects. With hospitals needing to manage against an annual budget and with specialist physicians on salary and closely linked to hospitals, inpatient care in the U.K. was relatively integrated. Furthermore, some diseases were singled out by the NHS for more integrated care approaches in specialized clinics, such as diabetes (see Chapter 3).

Level of competitive intensity: low. Neither the NHS hospitals nor specialist physicians employed by the NHS competed for NHS payor contracts. Under the jurisdiction of the DHAs, hospital budgets were set, in part, based on historical usage patterns and may have reflected long-standing preferential relationships. Budgets typically provided for expected increases in pay and supply prices.

Hospitals were not rewarded for improved efficiency nor penalized for below-average performance, in part because no performance data were available to evaluate the providers. Once allocated to a provider, capital was "free" and no return was expected. Within the private market, however, both hospitals and specialists did compete for private insurance contracts.

Since 1991, however, significant changes have been introduced in the care provision market between payors and providers to create a more competitive and efficient internal market. Hospitals were given financially independent status and were required to finance operations through contracts with purchasers. At the same time, the role of the DHA was essentially changed from administrator/manager to local purchaser, with funds now allocated based on population size and need rather than on the level of care provided. Thus, a system was created with the aim of balancing local supply and demand within the constraints of a global budget system.

<u>Primary physician care provision services</u>

Degree of product integration: low. The NHS had approximately 31,500 GPs in 1990. GPs were highly fragmented in independent practices, usually partnerships of 2 or more; approximately one-third worked in practices of 5 or more partners.

The care provision product provided by GPs was an annual per capita set of services for a fee for each patient on their list or "panel" (adjusted for age and sex and prior experience). Funds allocated to GPs were divided into three categories: capitated primary care payments for each person registered with the GP, payments to cover a proportion of the GP's practice expenses, and payments for carrying out certain tasks or achieving predetermined targets. Funds were allocated using forecasts for the number of patients, the expected practice expenses and the demand for additional physician services such as night visits. As no maximum limit was imposed on these expenditures, supplementary allocations were required to compensate for unexpected increases in demand or prices. Additional income (sometimes more than 60 percent of their annual income) was obtained by GPs who performed minor

surgical procedures in their offices. Even with some capitated component in the payments, GPs were essentially reimbursed on an activity or service basis.

Actual fee and allowance levels for GPs were set by the government, following the advice of an independent Pay Review Body. This body made its recommendations after hearing representatives from the relevant professional associations. These recommendations were not binding on the government, and in a number of recent years, recommendations have either been delayed or not implemented in full. Family Practitioner Committees had to pay each GP in their district (local area) according to the nationally determined formulas.

GPs were also required to submit basic information to the Family Practitioner Committees, including the age and sex of patients under their care, number of treatments, and number of referrals (other data elements were added post-reform, including the number of PAP smears and immunizations). In this way, the NHS could monitor how well each GP performed and how well each GP implemented the gatekeeping role by limiting referrals.

Since 1991, the reforms have introduced the potential for greater product integration at the GP level with the creation of risk-assuming "GP fundholders." These groups of GPs receive and manage more of an integrated budget for broader care services, including some nonemergency hospital services, across a range of diseases.

Level of competitive intensity: low. Prior to the reforms, U.K. GPs faced little or no competition for NHS or other payor contracts. GPs were assigned a patient population through the NHS, which was roughly evenly distributed among area GPs. In this way, each GP had explicit responsibility to care for a given set of patients.

Since 1991, competition has increased with the introduction of GP fundholders described above, allowing primary care physicians to compete alone and in group practices for a broader range of care provision services. These GP fundholders can also preferentially contract with hospitals outside their district. GP fundholders in the U.K. now represent approximately 35 percent of all GPs and cover approximately 40 percent of the population. In addition, publicly funded local purchasers were established to contract with competing providers for services, including GPs, with the introduction of the internal market described earlier.

3. U.K.: interactions between consumers and providers for hospital and physician care provision services

Since the aggregate and local supply of hospitals and physicians was controlled centrally and was generally tight, consumers' ability to choose providers was rather limited (see section below on regulation of supply). In addition,

consumers generally followed physician recommendations regarding treatment, as they shared little risk for the cost of treatment.

Degree of product integration: moderate. From the perspective of the consumer, care provision services were moderately integrated, given the strong gatekeeper role of the GP in managing an individual patient's total care and given the fact that most consumers retained the same GP for a long period of time. This gatekeeping was likely necessary given the central controls on physician and hospital supply (see section below on regulation). In addition, hospital services appeared to the patient to be fairly integrated. Furthermore, some diseases such as diabetes were targeted for more integrated care approaches in specialized care settings in clinics, as discussed earlier.

Level of competitive intensity: low. Consumers had little meaningful choice of provider, given supply constraints and other regulations. While in theory patients were free to choose from among GPs in their district, there was very little incentive to do so. The transfer procedure was long and complex, many GPs had long patient lists, and the quality and variety of services offered by GPs did not vary significantly. As a result, once patients chose a GP they rarely, if ever, switched.

In addition, consumers had little choice of hospitals or specialists since they were required to be admitted to the hospital and treated by the specialist referred by their GP. Thus, providers – including GPs, specialists, and hospitals – faced little competition for patients. Since the 1991 reforms, however, competitive intensity among GPs for patients has increased, as GPs are now allowed to offer a greater range of services to patients. To date, however, there is no evidence that patients are switching GPs on this basis.

U.K.: regulation

Since the NHS is a nationalized system of health care, government regulation plays an important role in all of the health care system interactions, as noted in the prior descriptions. In this section we highlight and provide more detail on those regulations that most directly shaped the U.K. system structure.

Physician supply. The central government directly regulated physician supply. Numbers of students entering medical school were controlled by the central government on the advice of the Medical Manpower Advisory Committee, which was convened on an ad hoc basis prior to the 1991 reforms. This committee was made a permanent body in 1991 to continuously monitor the market for physician services and to make recommendations to the Secretary of State for Health.

Entry to general practice was tightly regulated, both in terms of total supply and geographical distribution, by the Medical Practices Committee. This committee

was established in 1948 to ensure that all citizens had access to GP services. It centrally approved (or refused) all applications to enter general practice, and rationed GP services through the use of a strict target of patients per physician for each locality (ranging in the U.K. from 1 GP:1,730 in 1986 to 1 GP:1,688 in 1990). The government also regulated where hospitals and their accompanying specialist physicians were located, again, to ensure that all citizens had access to these specialized services.

Global budget and capital restrictions. As described earlier, the NHS global budget was determined in the annual national budget negotiations, and then passed down to RHAs and DHAs. The RHAs allocated capital funds for large construction projects and distributed the remaining funds to DHAs which, in turn, allocated funds for equipment acquisition, smaller construction projects, and operating budgets for hospitals. Individual hospitals, therefore, did not have the ability to make capital investments on their own. With this approach, capital spending in the U.K. health care system was strongly influenced by regulatory and political forces. With the 1991 reforms, however, hospital trusts, and to some extent GP fundholders, have been given greater control over their capital purchases, with funds essentially loaned to them from the government with interest, much like a commercial transaction.

Gatekeeper/referral restrictions. Through the NHS, the government also imposed restrictions on patient referral processes. Specifically, all patients were required to first go through their GP before receiving any further, higher-level care. Furthermore, referral sequences among GPs, local specialists, and regional specialists were also regulated.

* * *

Overall, the highly centralized U.K. health system prior to the 1991 reforms was characterized by a moderate to high degree of product integration and very low competitive intensity. This system structure – strongly influenced by regulation – created activity-neutral and sometimes input-reducing incentives for payors and providers, and created significant aggregate supply constraints, particularly regarding capital and capacity (**Exhibit 6**).

THE U.S. SYSTEM[10]

The health care system in the U.S. was relatively competitive with few regulations. Distinct systems coexisted: a private health care system and government-sponsored systems, including both Medicare and Medicaid.

The U.S. private system (which covered about 73 percent of the population) could be characterized as very complex, highly flexible, and moderately competitive at all market levels, but with relatively low levels of product integration. The markets for health coverage and care provision services were largely defined at the local level.

In contrast, the government-sponsored Medicare program for the aged and disabled (which accounted for about 34 percent of total costs and covered about 3 percent of the population in 1990, including all those over 65 years old and certain persons with disabilities or kidney failure) was not very competitive and was highly regulated. Similarly, the Medicaid program for the poor (which accounted for about 8 percent of total costs and covered about 10 percent of the population, and was jointly financed by federal and state governments) was also very competitive and highly regulated. Interactions between consumers and providers in the Medicare and Medicaid systems were similar to those in the private system, with physicians competing for patients and typically providing low levels of product integration. About 14 percent of the population (of which approximately 75 percent were employees or their dependents) had no health coverage other than indigent care.

The markets in the U.S. health care system and regulatory environment in the mid to late 1980s can be summarized as follows:

1. *Interactions between consumers and/or employers and payors for health care coverage:* low degree of product integration for both systems, given the dominance of indemnity insurance; high competitive intensity for the private system, with multiple payors offering a wide range of products; low competitive intensity for the Medicare and Medicaid systems.

2. *Interactions between payors and providers for care provision services:*

   - *For hospital care provision services:* low degree of product integration for the private system, with payors typically contracting with hospitals on an FFS or per diem basis; moderate degree of product integration for the Medicare system, given case rate reimbursement; low to moderate competitive intensity for all

---

[10] The basic sources used for this section (except where specifically noted) include the OECD, NERA, and interviews.

systems, as payors contracted with most hospitals at essentially the same rates within a region.

- *For specialist and primary care physician care provision services:* low degree of product integration for all systems, given the predominance of FFS reimbursement; low competitive intensity for all systems, as payors in the private system generally reimbursed all accredited providers and physicians could freely decide whether or not to contract with the government for Medicare and Medicaid patients.

3. *Interactions between consumers and providers for care provision services:* low degree of product integration for all systems, with most physicians in solo or small group practice providing services on an FFS basis; high competitive intensity in all systems, given that consumers were free to choose physicians and hospitals.

These market interactions were relatively uninfluenced by *regulation* for the private system, but somewhat influenced by regulation in the government-sponsored systems. Overall, there was no meaningful regulation of physician or hospital capacity, or of capital investments or allocation.

Since the late 1980s, both the private and government-sponsored systems have undergone significant change, driven primarily by market forces. In all systems, there has been significant growth in HMOs and other products in which payors take a more active role in coordinating and managing care. Providers are also increasingly establishing and practicing in integrated systems and physician group practices.

In the remainder of this section, we provide more detail on each of the three interactions listed above, as well as on the regulations shaping these interactions.

1. U.S.: interactions between consumers or employers and payors for health care coverage

In the U.S. private market, health care coverage was voluntary for consumers, obtained either through their employers (about 60 percent of the population) or through direct purchase (about 13 percent of the population). There were more than 1,000 private nonprofit and for-profit payors offering health care coverage. Primary payors included commercial insurers (for-profit firms or mutual companies owned by policyholders), Blue Cross and Blue Shield plans (nonprofit insurers composing approximately 30 percent of the market with about 70 plans nationwide), and various for-profit and nonprofit managed care organizations, including HMOs and preferred provider organizations (PPOs). In most

geographic markets, there was high fragmentation among payors, with only a few having greater than a 10-percent local share.

Employers could choose whether or not, and at what level, to provide health care coverage, as well as how much to charge their employees and their dependents for coverage. Most employers viewed health care coverage as a benefit they had to provide to be competitive in attracting workers, and as basically a substitute for cash wages (in 1989, health benefits were about 6 percent of total compensation and 36 percent of total benefits). In addition, employees were able to avoid income taxes as well as social security taxes when they received wages in the form of health benefits. Self-employed or unemployed individuals, as well as those who did not receive coverage from their employers, could freely decide whether or not to purchase coverage and what type to purchase. Most people who had insurance were covered for inpatient hospital services and physician services. Industries with strong unions (e.g., steel, automobiles) tended to have the broadest benefit coverage, while service industries (e.g., restaurants) provided little or none.

In contrast to the private insurance market, Medicare coverage was highly regulated and administered by the federal government and financed by a combination of payroll taxes, general federal revenues, and premiums paid by beneficiaries. Medicare coverage comprised two parts: Part A for inpatient care services and Part B for physician and other ambulatory services. Part A was earned through payment of a payroll tax during one's working years, while Part B was funded through voluntary payment of a premium, once eligible for Medicare. The payroll tax for Part A, paid by virtually all employed individuals, was 1.45 percent of payroll for both the employer and employee. The premiums from beneficiaries were about $30 per month, which collectively covered about 25 percent of the total Part B program cost. Medicare beneficiaries had to also pay co-payments and deductibles (accounting for about 17 percent of services covered and consuming about 6 percent of patients' per capita income).

Because of these high out-of-pocket costs as well as some gaps in Medicare coverage (e.g., outpatient prescription drugs and long-term nursing home care), approximately 70 percent of Medicare beneficiaries purchased private supplemental coverage (purchased either individually or by their current/prior employer). This supplemental market behaved similarly to the private market described above. Since 1982, the federal government has allowed private managed care organizations to market to Medicare beneficiaries within specified guidelines.

Similar to the Medicare system, the Medicaid program was highly regulated. It was jointly funded by the federal and state governments and administered by the states under broad federal guidelines governing the scope of services, the level of payments to providers, and the population groups eligible for coverage. To be eligible for Medicaid, a person had to be poor and/or aged, blind, disabled, pregnant, or the parent of a dependent child. States further defined

eligibility levels (e.g., maximum income and asset levels) within certain broad parameters. About 60 percent of those below the federal poverty level were excluded from Medicaid, given the strict eligibility criteria, with significant state-by-state variation. Medicaid was the only public program that covered long-term nursing home care, with about 20 percent of Medicaid expenditures spent on nursing home care in 1990.

Degree of product integration: low for all systems. In the private system, payors had significant flexibility in designing, pricing, and marketing their products, with only moderate regulation by state insurance commissioners. Products covered a wide spectrum, including FFS coverage in which consumers could freely choose their providers; PPOs in which consumers received higher coverage levels when going to providers in a defined network; and HMOs in which consumers received generally more comprehensive benefits such as preventive services, but were generally restricted to receiving care from a more narrow provider network. Commercial insurers and the Blue Cross and Blue Shield plans tended to offer a range of these products. In 1989, FFS coverage accounted for about 73 percent of the privately insured market, PPO plans for 10 percent, and HMO for 17 percent; there were significant variations in these relative percentages across local markets.[11]

In general, payors offered employers and consumers fairly little product integration in their products, at least for the indemnity and PPO products that dominated at the time. Although the insurance coverage may have covered hospital and physician services for all diseases in a given time period (e.g., 1 year), the health coverage product usually contracted and reimbursed for each medical event separately, and therefore did not represent an integrated approach. This was also true for the Medicare and Medicaid systems.

In the last 5 years, however, there has been significant growth in HMOs and other managed care options in the private as well as government-sponsored markets; these products provide higher degrees of product integration, with payors taking a more active role in coordinating and managing care with providers.

Level of competitive intensity: high for private system, low for Medicare and Medicaid systems. As discussed, the U.S. private health care system comprised more than 1,000 private nonprofit and for-profit payors offering health care coverage, with a wide range of benefit structures, premiums, and approaches for paying the insured and the providers. This highly competitive system allowed for greater range of consumer choice as well as product and pricing flexibility.

In pricing their products, payors could freely underwrite risks; employ various rating methodologies; refuse to offer coverage to certain groups, segments, or

---

[11] Source: Employee Benefit Research Institute (EBRI). In 1991, FFS coverage had dropped to 50 percent of the privately insured market, while PPO plans had risen to 26 percent and HMO plans to 23 percent.

individuals; exclude coverage for preexisting conditions; and engage in other marketing or pricing practices to attract better risks to their plans and thereby offer more competitive rates. Payors could also freely adjust their benefit packages, including deductibles and co-payments as well as services covered.

Given that most commercial coverage was obtained through one's employer, competing payors generally pursued a two-stage, group marketing and sales process – first, to the employer, with the aim of being included in the choice of plans given to employees; and second, to the employees, to directly encourage them to choose their product because employers often offer a choice. While payors determined the price charged to the employer, the employer often influenced the price of different coverage options to the employees, depending on their overall health coverage strategy.

Unlike the competitive private market, there was little competition historically in the government-sponsored Medicare and Medicaid programs. Under the jurisdiction of federal and state governments, these programs functioned essentially as "single payor systems" for their covered populations.

However, some competition has been introduced in the Medicare system since 1982 when HMOs were allowed to compete for Medicare consumers in some states. While these organizations had to provide specified coverage and operate within certain federal guidelines, they were able to compete for members aggressively on the basis of additional coverage (e.g., prescription drugs) as well as on price (i.e., premiums paid by the beneficiary). By 1989, penetration of Medicare HMOs had only reached 3 percent of the Medicare population; by 1996, penetration increased to approximately 10 percent, ranging from 0 percent to almost 50 percent across different local markets. Competition has also been increasing in the Medicaid system, with "bidding out" to private managed care organizations on a pilot basis in some states.

2. U.S.: interactions between payors and providers for care provision services

For inpatient services, there were about 6,700 hospitals in the U.S. in the late 1980s, including approximately 5,500 community acute-care hospitals, 900 specialty hospitals (e.g., psychiatric, rehabilitation, long-term care), and 240 federal hospitals open only to military personnel, veterans, or Native Americans. Of the 5,480 community hospitals, about 60 percent were nonprofit, 26 percent were local government hospitals, and 14 percent were for-profit. There were approximately 3.9 community hospital beds per 1,000 residents (with an average occupancy rate of approximately 65 percent) with significant variations in supply across local markets.

For physician services, there were approximately 600,000 physicians in the U.S. in practice, or 2.3 per 1,000 population, with significant variations in supply

across local markets (e.g., 0.9 in rural populations). Of physicians in active practice, about one-third were in primary care and the remainder were specialists. During the late 1980s, most physicians were in solo versus group practice. In recent years, local and for-profit regional and national physician groups and integrated physician-hospital systems have emerged and grown rapidly. In addition, physician ownership (usually through joint ventures) of medical labs, diagnostic centers, and outpatient centers has increased.

This section focuses on the care provision market during the late 1980s and discusses hospital care provision services and physician care provision services in turn.

Hospital care provision services

Degree of product integration: low for private system, moderate for Medicare and Medicaid systems. In the private system, payors contracted with hospitals in a variety of ways, including FFS payments, per diem rates, case rates, and some capitation-based payments.[12] At the time of our assessment, the predominant form was FFS and per diem in the private system.[13] For PPO and HMO products, there was more case rate and some capitation-based payment, and hospitals could decide with which payors and on what terms they wanted to contract their services. As managed care penetration has increased, so has the extent of case rate and capitation-based payment for hospitals.

The Medicare system had an overall higher degree of product integration in the hospital product than the diverse private system. Since 1983, hospital services have been bundled and reimbursed on the basis of diagnosis-related groups (DRGs), which treat medical care provision on a more integrated disease level relative to FFS and per diem payments. Although introduced by Medicare, these case rates spread to the private market, as described above.

The Medicaid system, in contrast, utilized primarily per diem payments to hospitals. Since 1990, however, most states have adopted payment schemes similar to the Medicare system, resulting in an increasing degree of product integration.[14]

Level of competitive intensity: low for all systems. While there was some competition between payors and hospitals in the private system, it was rather

---

[12] In 1989, payments for inpatient costs were comprised of FFS (45 to 50 percent), per diem (15 percent), and per case (35 to 40 percent) payments, based on estimates. Medicare accounted for the majority of per case payments.

[13] In 1989, payments by private insurers for inpatient costs were comprised of FFS (87 percent), per diem (11 percent), and per case (2 percent) payments, based on estimates.

[14] In 1991, 45 states used payment systems based on a specified rate for each service rendered. Of these states, 20 had adopted the DRG-based system while 14 adopted a system based on a specified rate per case. Source: Medicaid Source Book: Background Data and Analysis (A 1993 Update). Congressional Research Service. CRS Report for Congress No. 61-899. Washington, 1993.

limited as most payors contracted with, and paid for services delivered in, most hospitals.

With the growth of managed care in more recent years, payors have been more selective in hospital contracting, focusing on achieving hospital cost discounts in exchange for channeling patient inflow. And as the level of excess hospital capacity has increased and the for-profit hospital chains have grown in recent years, the level of competitive intensity has increased dramatically in many U.S. metropolitan markets.

In the government-sponsored Medicare and Medicaid programs, competitive intensity was low since the federal or state government contracted with most accredited hospitals based on clear eligibility criteria. Similar to the private market, however, the level of competitive intensity has increased with the growth of managed care.

Physician care provision services (specialist and primary care)

Degree of product integration: low for all systems. Private payors contracted with physicians in a variety of ways, including FFS, capitation, or salary-based payment. In the late 1980s, FFS was the dominant reimbursement mechanism. Under FFS, payors paid physicians a specified level of coverage, generally 80 percent of "usual and customary fees." Because this method paid physicians "piecemeal" for each service rendered, the level of product integration in this product was low. The FFS-based physician product also dominated in the Medicare and Medicaid systems; payments to physicians for specific services varied based on geographic differences only.

In both the private and government-sponsored systems, the degree of product integration in physician products/services has increased with the rapid growth of managed care.

Level of competitive intensity: low for all systems. In payor-physician interactions, there was a relatively low level of competitive intensity in the private and government-sponsored systems. While payors in the private system had almost complete freedom in deciding with whom to contract, for which services, and on what terms, the de facto contract in the predominant FFS physician product meant that the payor would reimburse every accredited physician at a specified level of coverage. In addition, payors had very little information – such as physician cost, service, or clinical outcome performance – on which to base contracting decisions. Furthermore, U.S. physicians in some specialties and markets were able to exert considerable power and influence through formal and informal associations, such as the American Medical Association (AMA), despite being in solo practices.

There is some evidence, however, that competitive intensity in this interaction in the private market has increased and the balance of power has shifted between

physicians and payors with the growth of managed care. For example, physician salaries, particularly for specialists, have recently declined in real terms.

In the Medicare and Medicaid systems, competitive intensity was also low since physicians could freely decide whether or not to contract with the government. If they decided to contract, they had to meet defined criteria and agree to accept payment according to the set fee schedule. As in the private system, the level of competitive intensity in the government-sponsored systems has increased somewhat with the growth of managed care in recent years.

3. U.S.: interactions between consumers and providers for hospital and physician care provision services

Degree of product integration: low for all systems. As described above, most physicians in the late 1980s were in solo or small group practice, with the predominance of an FFS reimbursement. In this environment, providers generally did not offer integrated care services to consumers; patients were treated on an as-needed basis and often by a number of separate providers. With the recent increase in managed care and the associated development of risk-assuming physician group practices and integrated care systems, however, the degree of product integration has increased significantly.

Level of competitive intensity: high for all systems. Competitive intensity was high between consumers and providers, as consumers were relatively free to choose their physicians and hospitals in all systems. Since consumers generally had minimal co-payments and deductibles, providers tended to compete more on the basis of services, quality perception, and other factors. Within HMOs and other managed care offerings, however, consumers may have more limited choice of provider.

U.S.: regulation

As described above, interactions between payors, providers, and consumers were relatively unregulated for the private system. In contrast, government-sponsored programs including Medicare and Medicaid were highly regulated by both the federal and state governments. While some regulation of hospital supply occurred at the state level through "certificate of need" programs and some regulation of physician supply occurred through the medical school admission process, there were no meaningful federal supply-level controls.

* * *

Overall, the U.S. health care system of the mid to late 1980s was characterized by a moderate level of competitive intensity and a relatively low degree of

product integration. This system, which was relatively unregulated, created predominantly activity-increasing incentives for providers and created no meaningful supply constraints (**Exhibit 7**).

SUMMARY OF MAJOR DIFFERENCES
AMONG THE THREE SYSTEMS

The predominant health care systems of the three countries included in this assessment differed along many dimensions, most notably in the level of competitive intensity and product integration in the care provision and health coverage markets, as well as in the nature and extent of regulation. The U.S. was the most competitive, market-based system (except for the government-sponsored Medicare and Medicaid programs), with substantial degrees of freedom for payors and providers. The U.K. system was the most centrally controlled of the three, leading to high product integration and low competitive intensity. And while the German system was relatively regulated in terms of payors and their negotiations with providers, it had low levels of both competitive intensity and product integration.

To facilitate reading and interpretation of the four disease case studies (Chapters 3 through 6) and the cross-disease synthesis (Chapter 7), we have developed two summary exhibits highlighting and contrasting the characteristics of the three systems at the time of our assessment (**Exhibit 8**) as well as with more recent changes (**Exhibit 9**).

At the time of our assessment

¶ In the *health coverage market*, there were major differences in the level of competitive intensity and product integration among the three systems. The U.K., through the NHS, had a highly integrated health coverage product, whereas the products of the U.S. and German systems were relatively unintegrated. Payors in the U.K. had virtually no competition, some German payors competed to some extent for members, and U.S. payors competed quite aggressively for members.

¶ For the *interaction between payors and providers in the care provision market*, competitive intensity was relatively low or nonexistent across all three countries, but the U.K. had more integrated care products relative to Germany and the U.S., particularly for hospital and specialist physician services. The U.S. also had more integrated hospital products for Medicare relative to Germany.

¶ For the *interaction between consumers and providers in the care provision market*, competitive intensity was moderate in the U.S., slightly lower in Germany, and very low in the U.K. (due primarily to supply/capacity constraints). U.K. provision products, however, were more integrated than those of Germany and the U.S., particularly for hospital and GP services.

Summarizing these levels of competitive intensity and product integration across the care provision and health coverage markets yields a highly simplified, two-dimensional characterization of each system (refer back to **Exhibit 4**). This characterization shows that the U.S. had a moderate level of competitive intensity, but relatively low product integration; the U.K. was very low in competitive intensity, but high in product integration; and Germany was relatively low on both dimensions.

With more recent changes

As discussed throughout this chapter, each of the three health care systems is undergoing significant change in the health coverage market, care provision markets, or both. In general, the three countries are increasing the level of competitive intensity and product integration in these markets (**Exhibit 9**):

¶ *In the health coverage market*, Germany has significantly increased the level of competitive intensity among sickness funds, while competition in the U.S. Medicare and Medicaid systems has increased with the opening of these markets to private managed care organizations. In addition, the growth of managed care in the U.S. has led to the development of more integrated health coverage products.

¶ *In the care provision market* (specifically the interaction between payors and providers), the U.K. has seen a significant increase in competitive intensity since the 1991 reforms for both hospitals (with the introduction of private trusts) and GPs (with the introduction of risk-bearing GP fundholders). In the U.S., the growth of managed care and associated development of physician group practices and risk-bearing provider systems has led to an increase in the degree of product integration in provider services. In Germany, the introduction of case rates for some hospital services (covering approximately 15 to 20 percent of revenues or services) represents some increase in the degree of hospital product integration.

Although our productive efficiency assessment focuses on system structure in the three countries during the mid to late 1980s, more recent changes are important to understand in determining potential implications for policymakers and health care organizations today.

* * *

As can be seen in the next four disease-specific cases (Chapters 3 through 6) and the subsequent cross-disease synthesis (Chapter 7), differences in the regulation of the health care system and the resulting structure of the health coverage and care provision markets within the U.S., the U.K., and Germany created very

different incentives and constraints for providers. These differences, in turn, led to different treatment approaches and resulting productive efficiency differences.

Exhibit 1

HEALTH CARE SYSTEM
MARKET INTERACTIONS

⟷ Interaction

◀- -▶ Interaction that can be internalized

▭ Product/service that is exchanged

Government regulation

1. Health care coverage

2. Care provision services

| Consumers/employers | Payors, e.g.
 • Single/government-run
 • Multiple, quasi-public
 • Multiple, private | Hospitals and other institutions |
| | | Physicians and other care providers |

3. Care provision services
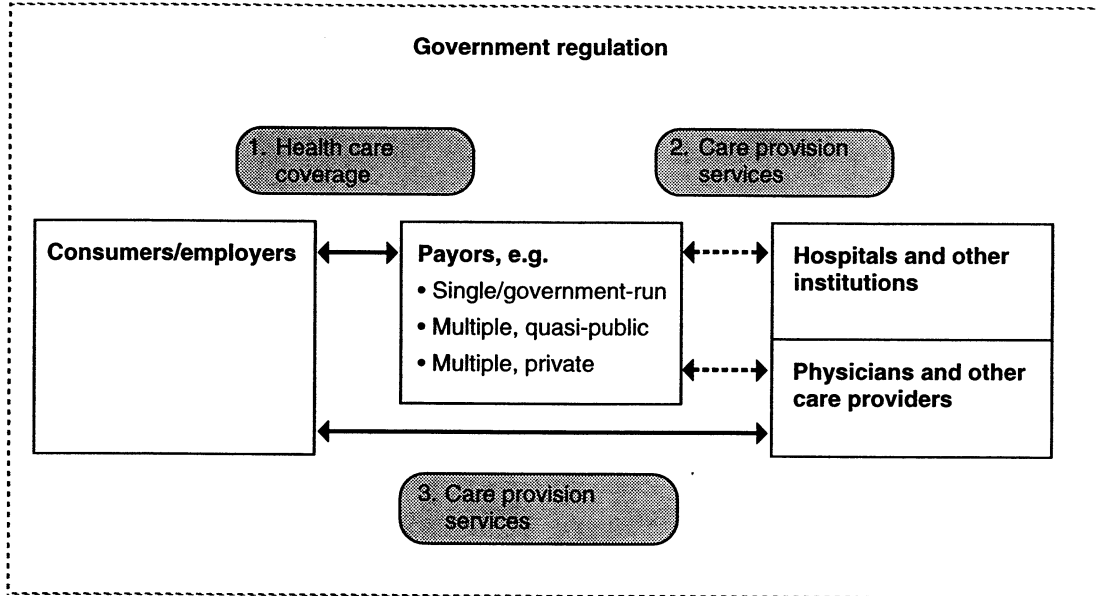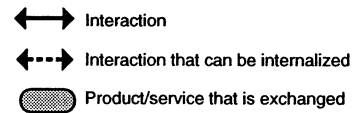
- Interactions can be highly regulated, market-based, or a blend, depending on how they are structured and regulated

- While interactions are interdependent, each can be structured differently

- Collectively, different market structures create different **incentives and constraints** for providers, which in turn drive different treatment patterns and productive efficiency levels

022 MI 95131/2

Exhibit 2

HEALTH CARE PRODUCT INTEGRATION
BY PROVIDERS OR PAYORS

Individual services	Provider integration	Payor integration	Consumer

Office visit No. 1 ⟶

X ray ⟶ → Diagnosis "package"

Lab test ⟶

Office visit No. 2 ⟶ → Treatment package

Surgical procedure ⟶

Follow-up visit ⟶ → Package to address complete condition or phase of disease →

Prescription ⟶

Inpatient stay – first day ⟶

Lab test in hospital ⟶ → Hospital care package →

X ray in hospital ⟶

Second day in hospital ⟶

Follow-up visit ⟶

Lab test ⟶

022 EF 136233/2

Exhibit 3

DEGREE OF HEALTH CARE PRODUCT INTEGRATION

Breadth of diseases covered over time

022 ST 140109/2

Exhibit 4

OVERALL MARKET STRUCTURE OF EACH SYSTEM
Mid to late 1980s

* "Average" over specific care provision and health coverage markets in each country
022 ST 125810/2

Exhibit 5

SUMMARY OF GERMAN HEALTH CARE SYSTEM CHARACTERISTICS

Mid to late 1980s
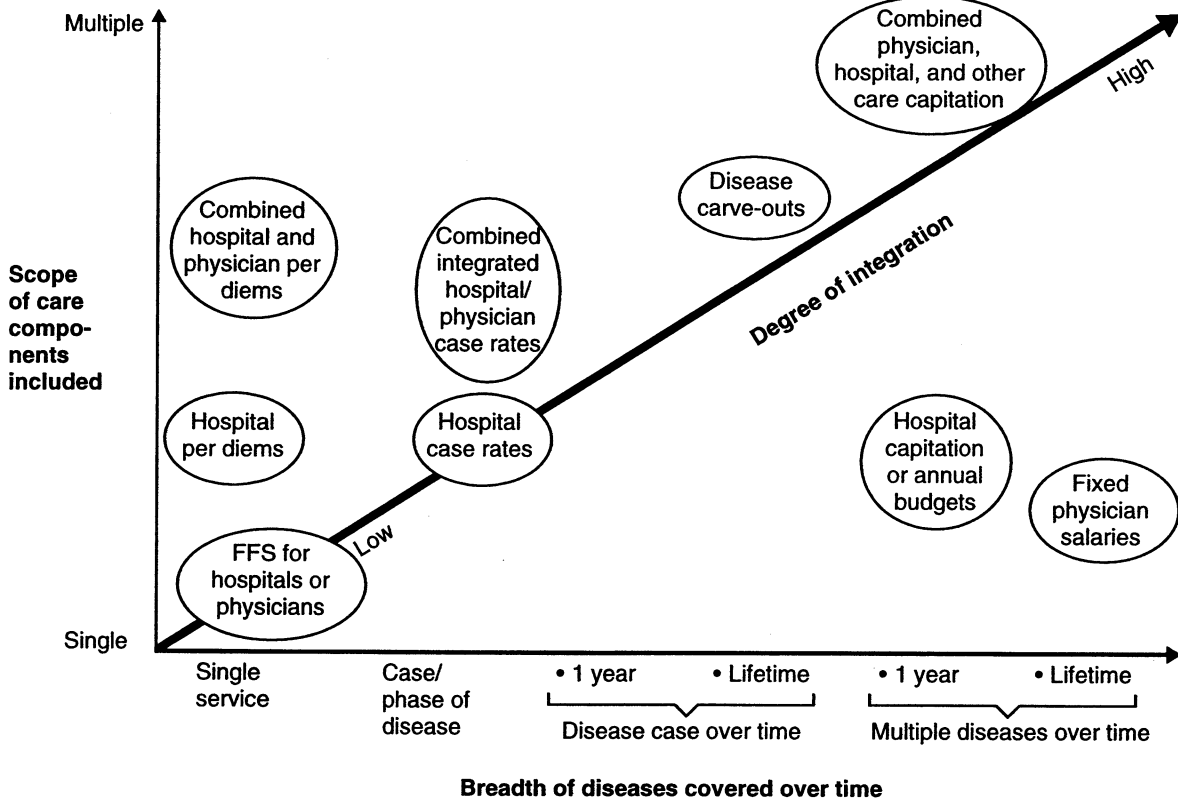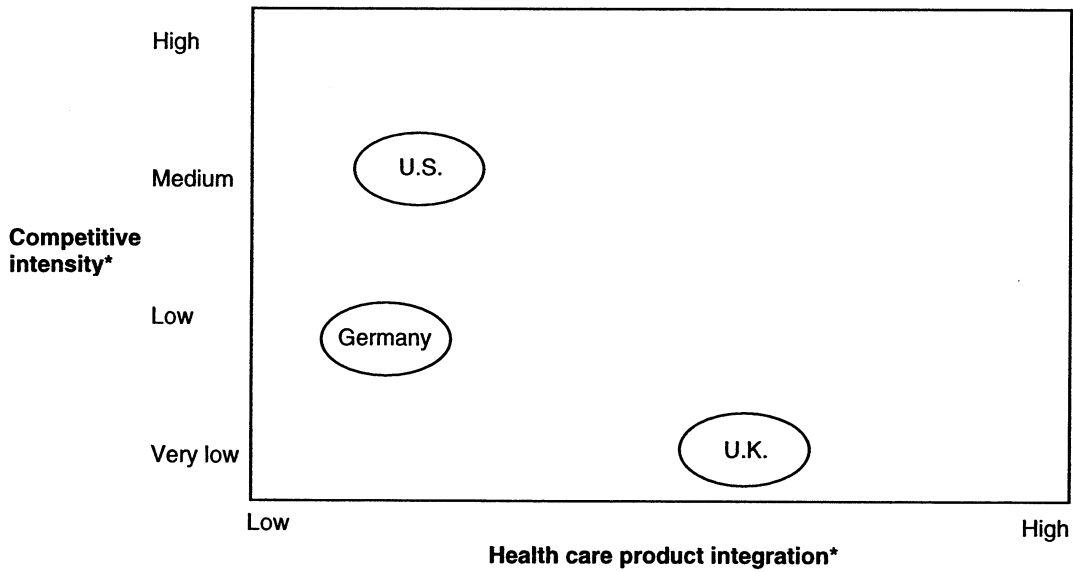
Market/interaction	Degree of product integration	Level of competitive intensity	Highlights	Implications for provider incentives and constraints
1. Health coverage: consumer/ employer-payor interaction	Low	Historically low (but moving toward moderate/ high)	• Mandated coverage; historically very limited consumer choice and product/pricing flexibility (but changing) • Relatively concentrated, nonprofit sickness funds accounted for ~90% of population • Payors and employers bore most risk; employers and consumers split cost of coverage 50:50	
2. Care provision: payor-provider interaction				
• **Payor-hospital (including hospital-based specialist physicians)**	Low	Low/none	• Per diem payments to hospitals (case rate payments emerged for selected diagnoses as of 1996) • Many hospitals (~2,000), but organized at regional levels for "collective" payor negotiations • No flexibility for individual negotiations or intervention; all payors negotiated as a group with each hospital • Limited and regulated information available to payors	• **Strong activity-increasing incentives, especially for hospitals**
• **Payor-physician (ambulatory)**	Low	Low/none	• Approximately 74,000 ambulatory private physicians organized into 19 regional public associations (KVs) for collective payor negotiations • While KVs bore collective risk through total budget payment, individual physician incentives existed to increase activity through FFS reimbursement for services • While annual rate/budget negotiated varied by payor, each payor had to negotiate with regional KV and could not select physicians • Limited information available to payors	• **Significant supply constraints on capacity and substitution between alternative care settings**
3. Care provision: consumer-provider interaction	Low	Moderate	• Relatively high flexibility in choice of hospital and physician for most services • Relatively poor information transparency for consumers • Virtually no cost/risk shared by consumers, given low co-payments and deductibles	
Regulation			• Physician capacity controlled by KVs; hospital capacity controlled by the state with payor influence • Strict segregation of hospital-based physicians and ambulatory physicians	

Exhibit 6

SUMMARY OF U.K. HEALTH CARE
SYSTEM CHARACTERISTICS

Mid to late 1980s

Market/interaction	Degree of product integration	Level of competitive intensity	Highlights	Implications for provider incentives and constraints
1. Health coverage: consumer/ employer-payor interaction	High	None for public coverage; high for private coverage	• Highly concentrated: NHS provided all coverage to ~90% of population through 14 RHAs and ~200 local DHAs (for hospital services) and 90 Family Practitioner Committees (for GP services) • Mandated coverage; no choice and/or product/pricing flexibility for NHS coverage • NHS bore all risk; consumers bore cost through general taxation, mandatory employee contributions, and some direct charges	
2. Care provision: payor-provider interaction				
• Payor-hospital (including hospital-based specialist physicians)	Moderate/ high	Historically low (but moving toward moderate with 1991 reforms)	• Annual budgets for hospitals; hospital-based physicians paid on salaried basis, but with significant opportunities to make additional income from private FFS practice • Largely "internalized" (before 1991) at local level; DHAs and RHAs directly provided or contracted for care subject to budget	• Activity-neutral and sometimes input-reducing incentives
• Payor-physician (GPs)	Low	Historically low (but moving toward moderate with 1991 reforms)	• Blend of FFS and some capitation-based reimbursement methodology for GPs • Largely internalized (before 1991) at local level; Family Practitioner Committees contracted with local GPs; limited flexibility, given government-set fee and allowance levels • Highly fragmented physicians, with ~1/3 in larger group practices	• Significant supply constraints, particularly for capital and hospital and physician capacity
3. Care provision: consumer-provider interaction	Moderate	Low	• Although high flexibility (theoretically) in choice of hospital and physician services, effectively low given supply constraints • GPs referred patients to consultants/ hospitals; no direct access • Relatively poor information transparency for consumers • Virtually no cost/risk shared by consumers, given low co-payments and deductibles	

| Regulation | | | • Centralized control of global budget, aggregate supply levels, capital expenditures, and referral/gatekeeper processes
• Some "centralized" coordination or active promotion in a few target diseases (e.g., diabetes) | |

022 MI 95133/2

Exhibit 7

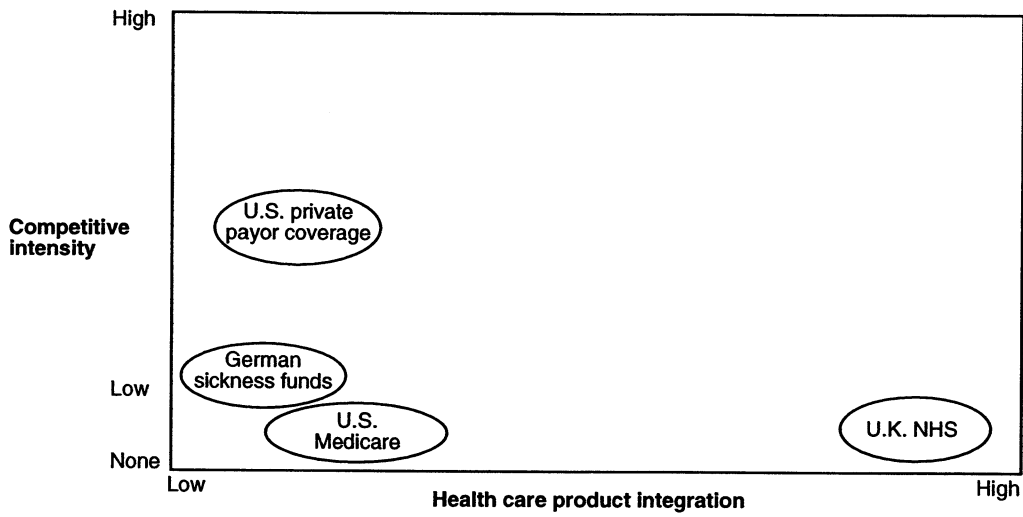# SUMMARY OF U.S. HEALTH CARE SYSTEM CHARACTERISTICS

Mid to late 1980s

Market/interaction	Degree of product integration	Level of competitive intensity	Highlights	Implications for provider incentives and constraints
1. Health coverage: consumer/ employer- payor interaction	Low (shifting to moderate)	High (private) Low (Medicare/ Medicaid)	• Highly fragmented payors and employers/ consumers with mix of for-profit and not-for-profit payors (except for Medicare) • High consumer choice and product/pricing flexibility • Payors and employers bore most risk; employers and consumers share in cost of coverage (government/taxes funded Medicare) • Low information transparency, given complexity	
2. Care provision: payor-provider interaction				
• Payor- hospital	Low (shifting to moderate)	Low/ moderate (shifting to moderate/ high)	• Mix of per diem and case rate payments (all case rate for Medicare); some capitation emerged in 1990s • Many fragmented hospitals negotiated with payors predominately on individual basis (except for Medicare) • High flexibility in rate/term negotiations and degree of intervention by payors	• **Predominantly activity- increasing incentives** • **No major supply constraints**
• Payor- physician	Low (shifting to moderate)	Low (shifting to moderate/ high)	• Predominance of FFS payments; growing use of capitation and other risk-based payments in 1990s • Over 600,000 private physicians; highly fragmented with most in solo or small group practice, but groups emerged in 1990s • High flexibility in rate/term negotiations and degree of intervention	
3. Care provision: consumer- provider interaction	Low	High	• Relatively high flexibility in choice of hospital and physician for most services (but decreasing with growth of managed care) • Relatively poor information transparency for consumers • Virtually no cost/risk shared by consumers, given low co-payments and deductibles	

| **Regulation** | | | • Some state-level regulation on capacity, but no meaningful federal-level controls | |

022 MI 95134/2

Exhibit 8

HEALTH CARE MARKET STRUCTURES: MID TO LATE 1980s

Health coverage market

High

Competitive
intensity

U.S. private
payor coverage

German
sickness funds

Low

U.S.
Medicare

U.K. NHS

None

Low

High

Health care product integration

Care provision market: between payors and providers

High

Competitive
intensity

U.S. hospital and
physician services

Low

U.K. GP
services

U.S. hospital
services for
for Medicare

U.K., German
specialist
services

U.K. hospital
services

German ambulatory
physician and hospital
services

None

Low

High

Health care product integration
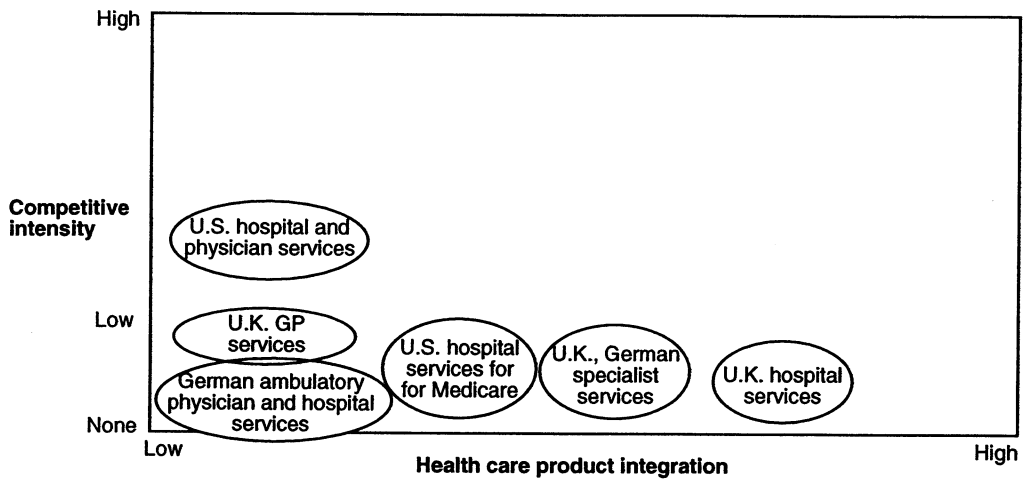
Care provision market: between providers and consumers

High

Competitive
intensity

U.S. hospital and
physician services

U.S. hospital
services for
for Medicare

German ambulatory
physician and hospital
services

German
specialist
services

U.K. specialist
services

U.K. GP
services

U.K. hospital
services

Low

None

Low

High

Health care product integration

022 ST 140146/2
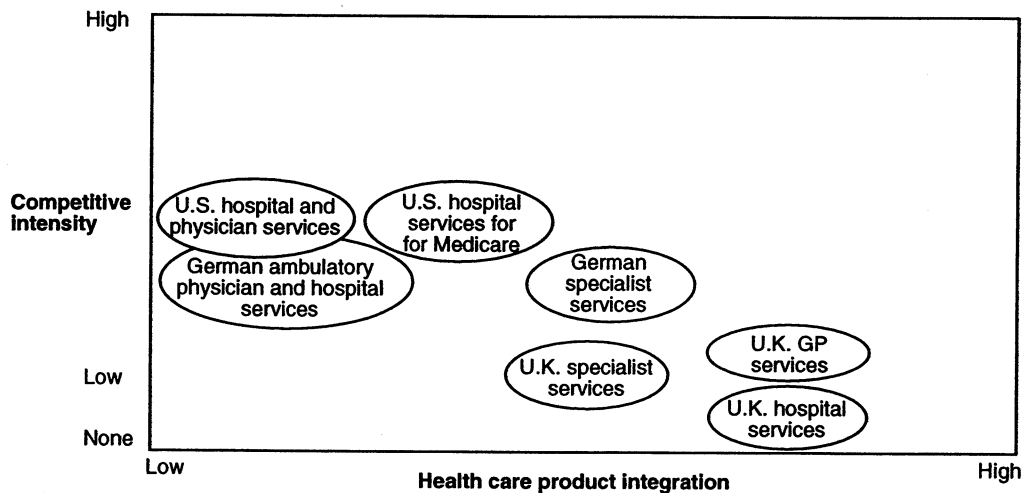
Exhibit 9

RECENT CHANGES TO HEALTH
CARE MARKET STRUCTURES

◯ Mid to late 1980s

⬭ Current

Health coverage market

High

Competitive
intensity

Low

None

```
                              U.S. private                    U.S. managed
                              payor coverage  ──────────────▶  care plans

                                    Competing
                                 ┆ sickness funds ┆

                              German
                           sickness funds
                                      ▲
                                       ＼
                                        ＼
                              U.S. Medicare                              U.K. NHS
```

Low   High

Health care product integration

Care provision market: between payors and providers

High

Competitive
intensity

Low

None

```
              U.S. hospital and physician services
              (wide range of provider products)

                                              U.K. GP
                                           fundholder
                                            services
                                                            Hospital
         U.S. hospital and                                trust services
         physician services           U.K. specialist
                                          services

                        U.S. hospital
                        services for
         U.K. GP        Medicare        U.K. specialist
         services                         services        U.K. hospital
         German hospital  German hospital  German specialist  services
         services          services          services
```

Low   High

Health care product integration

022 ST 140147/2

Chapter 3: Diabetes case

This chapter discusses the relative productive efficiency of the U.K. and the U.S. in the treatment of diabetes.

We begin with an overview of the disease, a discussion of the productive efficiency measure used, and a description of the treatment process. After assessing the relative productive efficiency of these countries in the treatment of diabetes, we analyze the provider behaviors driving these productive efficiency differences. Finally, we discuss how different health care system structures and regulatory environments affected provider incentives and constraints and, therefore, productive efficiency.

BRIEF DISEASE OVERVIEW

Diabetes is a chronic condition that impairs or destroys the body's ability to regulate the level of glucose in the body. It affects a significant fraction of the population – about 2 to 3 percent – in the U.S. and the U.K.[1] (Because information was not available on treatment in Germany, we excluded it from this comparison). Diabetes accounts for at least 4 to 6 percent of total health care costs in both countries; this cost burden is expected to grow as the prevalence of diabetes increases with aging populations.

Diabetes is really two different conditions.[2] Type I diabetes (or "juvenile onset") occurs early in life and results in the destruction of the body's ability to produce insulin and, therefore, regulate glucose. Type II diabetes (or "adult onset") develops later in life and results in decreased insulin secretion and decreased sensitivity to insulin. Type II diabetes is far more common than Type I; approximately 90 percent of diabetics in the U.S. and U.K. are Type II. Type II diabetes develops gradually, while Type I diabetes develops abruptly. Although Type I and Type II are different diseases and can be treated differently, their treatment processes involve many of the same providers and clinical protocols.

There is no cure for diabetes. Diabetes is treated by managing blood glucose levels. For Type I and some Type II diabetics, management involves regular

---

[1] Sources: Office of Population Censuses and Surveys, 1993; British Diabetic Association, 1996; NIH Publication No. 95-1468, 1995.

[2] Our study excluded *gestational* diabetes, which is diabetes with onset (or first recognition) during pregnancy.

insulin injections. For many Type II diabetics, management consists primarily of controlling the patient's diet and exercise habits, as well as the use of oral agents. For all diabetics, ongoing management is required throughout the patient's life.

Diabetics frequently develop complications from the disease. Some of these complications are life-threatening, while others significantly diminish quality of life. Common complications include heart and kidney disease; visual impairment, which may lead to blindness; and foot ulceration, resulting from poor circulation and nerve damage which, if severe enough, may require amputation. With effective management of the diabetic's condition, some of these complications can be prevented or significantly delayed.

DEFINITION OF PRODUCTIVE EFFICIENCY MEASURE

Because the two conditions are clinically distinct, we measured the productive efficiency of Type I and Type II treatment separately. For each type of diabetes, we measured both the inputs into the disease treatment process and the outcomes from this process.

Timeframe of analysis

Our analysis of diabetes focused on treatment practices between 1985 and 1990. Because no comprehensive studies provide a snapshot of diabetic care in a specific year, we combined data from several different years and time periods to build an aggregate picture of diabetes treatment. Most of our sources focus on the late 1980s, although some – particularly studies of and data on outcomes – are from earlier or slightly later time periods. Because diabetic care practices have evolved relatively slowly over the past 10 to 20 years, reflecting the accumulation of gradual improvements in care rather than dramatic breakthroughs in treatment, our use of data from multiple years should not introduce significant biases or inconsistencies. (See Appendix 3D for a description of major sources used.)

Summary of disease management and treatment phases

The treatment of diabetes, whether Type I or II, can be divided into two phases (**Exhibit 1**): 1) management and 2) complications. In the management phase, patients and providers seek to maintain near-normal blood glucose levels while screening for early signs of complications. In the complications phase, complications of diabetes are treated. After being treated for a complication, a patient will return to ongoing management and may require further treatment of any subsequent complications.

Our productive efficiency analysis focused on specific aspects of the diabetes disease treatment process that are both measurable and likely to reflect

differences between the U.S. and the U.K. This analysis evaluated the labor productive efficiency of the ongoing, physician-guided care programs in the management phase, and the inpatient treatments for selected complications (**Exhibit 2**). The following sections explain our measurement of inputs and outcomes in detail.

Measurement of inputs

Input measurement focuses on provider labor inputs. The patients themselves provide the most important labor input into the treatment of diabetes through self-care in the management phase of the disease. Self-care includes insulin injections when needed, self-testing of blood and urine, and diet and exercise control. Economically, the patient's labor in performing these functions is an input into the production process. However, we did not try to estimate the patient labor input for two pragmatic reasons. First, data on actual patient labor are nonexistent. Second, it would be difficult to assign an opportunity cost to this patient labor, since time spent in self-care probably does not reduce the patient's working hours. (Instead it probably reduces the patient's leisure time, which has a utility impact for the patient but is very difficult to quantify.)

Capital and supplies inputs not evaluated. Accurate data on capital and supply usage in diabetes treatment are unavailable. Almost all treatment in the management phase is delivered in the outpatient setting, where data are typically harder to collect and less widely available. In fact, the relative scarcity of outpatient data in Germany led us to exclude it from the comparison entirely.

Because labor represents roughly 70 percent of the total cost of health care in both the U.S. and the U.K., we believe this restriction to labor inputs represents an acceptable simplification. Diabetes does have one unique characteristic in that the cost of supplies for self-care, particularly insulin, can be significant over a patient's lifetime. However, as we will discuss below, the largest cost component of diabetes is the inpatient care associated with complications treatment; for this inpatient care, labor is clearly the major input.

Evaluation includes the major treatment steps, but excludes minor inputs. The diabetes treatment steps requiring the majority of provider labor are the ongoing, routine visits during the management phase and the inpatient treatment of complications. Our analysis estimated the labor inputs into both of these treatment steps. We did not analyze the following relatively minor inputs, both because they accounted for little cost (especially relative to the costs attributed to labor during the management phase) and because we did not suspect that there were major differences between the two countries.

> ¶ **Diagnosis of diabetes.** Initial diagnosis of diabetes may require several physician visits and lab tests. Diagnosis can sometimes occur during routine medical exams or during treatment of other conditions.

In other cases, patients present with symptoms indicative of diabetes and tests are required to confirm the diagnosis. Because the patterns of diagnosis are so variable, we did not include diagnosis in our measurement. We have no reason to believe that diagnosis protocols (e.g., which tests to perform) vary significantly between the U.S. and the U.K.; in particular, neither country has a formal screening program for diabetes.

¶ **Diagnosis and treatment planning for complications beyond routine visits in the management phase.** One function of routine physician visits in the management phase is to check for complications. Although primary care physicians ordinarily have responsibility for such monitoring, diabetics may be referred for additional tests or consultations if certain complications are discovered or suspected. (For example, a referral may be made to an ophthalmologist if a diabetic shows indications of retinopathy.) We focused only on the inpatient treatment generated by these referrals; outpatient physician visits and tests beyond those handled in routine clinic visits were excluded.

¶ **Follow-up visits after treatment for complications.** Similarly, after inpatient treatment, a diabetic may make follow-up visits to a specialist to monitor his condition beyond the normal routine care in the management phase. Such visits were also excluded from our 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.

Evaluation addresses several important complications. Ideally, diabetic complication rates between the U.S. and the U.K. would be compared by conducting simultaneous population-based studies on diabetics in the two nations that would permit adjustment for factors such as age, sex, race, and duration and type of diabetes. As a study such as this has never been done, we estimated complication rates by using national databases, surveys, and the available medical literature. Specifically, we evaluated complication rates for DKA/hyperosmolar coma, retinopathy, blindness, and lower extremity amputation. For each of these complications, we were able to obtain a comparable estimate of the incidence rate for the complication in the U.S. and the U.K. However, the comparability of the complication rates between the two nations may be subject to some margin of error

as a result of differences in the definitions of complications used or the age and duration of diabetes in the populations studied. (Specific data sources, estimates, and the comparability of each complication rate are described in Appendix 3C.)

Because of data limitations, we were not able to address other complications, such as end-stage renal disease, ischemic heart disease, and stroke. For these complications, data relating to their incidence and treatment do not adequately distinguish diabetes from other causes. It was, therefore, impossible to detect differences between the U.S. and the U.K. for these complications. Nevertheless, we believe that the complications we measured provide a good indicator of the relative inputs and outcomes for diabetes treatment between the two countries.

Outcomes measurement derived from complication rates. To develop an overall measure of outcomes for diabetes treatment, we estimated the impact of each complication on a diabetic's "quality of life." This estimate was based on the "Kaplan-Bush Index," a widely used scale that defines a range of possible health states and assigns a numerical score to each state. The health state score measures the relative quality of life for that state. The Kaplan-Bush scores were derived from interviews and surveys where population samples expressed their relative preferences for these health states.

With these quality of life impact scores and the incidence rates for complications, we developed a measure of the "expected quality of life score" for an average diabetic in each country. (In essence, this expected value weights the quality of life impact of a complication by the probability of developing the complication.) We used this expected value as our basic outcome measure for diabetes. It is expressed in quality adjusted life years (QALYs), which are normalized so that a complications-free year of life is worth 1.0 QALYs. (Details of our methodology for defining and calculating this expected QALY measure are provided in Appendix 3A.)

Results not sensitive to specific assumptions and methodology. To derive an expected QALY score for each country, we made a number of assumptions about a diabetic's potential health states, the quality scores of these states, and the probabilities of being in these states over time. While some of these specific assumptions could be challenged, and other models of expected QALY could be developed, the final result of our outcome comparison between the two countries is quite robust. Essentially, any reasonable set of assumptions and methodology yields an outcome measure that shows the U.K. having superior outcomes for diabetes treatment. The reason for this robustness is that diabetics in the U.K. are less likely to develop each of the complications (**Exhibit 3**); any process for combining these complication rates into an overall outcomes measure will, therefore, show that diabetics in the U.K. had better outcomes.

Although the ranking in outcomes between the countries is not sensitive to specific assumptions, the absolute magnitude of the expected QALY measure for each country is affected by some of these assumptions.

Therefore, our comparison of outcomes should be interpreted primarily as a qualitative ranking and only secondarily as a quantitative measure.

Outcomes measurement excludes some aspects of diabetes treatment. Our expected QALY measure reflects the effectiveness of each health care system in preventing complications, which is a major objective of diabetes treatment. However, this measure does not capture some other potentially important aspects of outcomes from diabetes treatment:

¶ **Impact of the burden of self-care on the diabetic's quality of life.** The health system can influence the extent of this burden through education, design of self-care protocols, availability of supplies, and potentially even through technology (e.g., use of insulin pumps). While we did not incorporate these factors into our measure, we have no reason to believe that either country differs significantly from the other in any of these areas.

¶ **Risks of hypoglycemia.** Insulin-using diabetics who tightly control their blood glucose levels may succeed at preventing complications, but have a greater likelihood of experiencing episodes of hypoglycemia. (Hypoglycemia occurs when the diabetic has a blood glucose level that is too low; it can result in a range of acute symptoms and, possibly, in long-term neurological impairment.) These incidents can have a significant effect on quality of life, and potentially even on mortality through an increased chance of accidents. We have not incorporated hypoglycemia into our measure because data on its incidence are not available. Again, though, we have no reason to believe that either country performs better than the other in preventing hypoglycemic attacks. Furthermore, recent evidence from the Diabetes Control and Complications Trial (DCCT) suggests that the higher rate of hypoglycemic attacks that occur with tighter glucose control have minimal or no effects on quality of life.[3]

¶ **Overall mortality risk for diabetics.** Comprehensive mortality curves for diabetes are not available; therefore, we were not able to measure the overall mortality risk for diabetics in the two countries. However, we were able to compare some indicators of mortality, such as death rates per capita from diabetes and death rates from selected diabetic cohort studies between the two countries. These indicators suggest that the U.K. has lower mortality from diabetes, although the results are not conclusive. (Details of these mortality indicators are presented in Appendix 3A.) The qualitative result from our complications-based measure – that outcomes in the U.K.

---

[3]   Source: *Diabetes Care*, 1996.

are superior – would, therefore, likely remain true if mortality data were included in the measure.

Outcomes measurement compares U.K. population with U.S. white population.
Blacks have a significantly higher incidence of both diabetes and its associated complications than whites.[4] Because blacks comprise 12 percent of the U.S. population but only 1 percent of the U.K. population, comparing complication rates of the total U.S. population with the total U.K. population potentially biases the results; higher complication rates in the U.S. relative to the U.K. might be the result of the racial composition of the population rather than different treatment processes for diabetes. To eliminate this potential bias, we compared complication rates of the U.K. population to complication rates of the U.S. white population. (See Appendix 3C for details on the populations studied for each complication rate.)

DESCRIPTION OF THE MANAGEMENT AND TREATMENT PROCESS

The following sections describe each of the two phases of the diabetes management and treatment process in some detail and highlight the clinical and economic trade-offs implicit in the range of treatment options available in each phase. Later in this chapter, we discuss the specific practice patterns observed in each country and their implications for productive efficiency.

Overview

Management of diabetes after diagnosis consists of ongoing self-care by the diabetic and periodic interactions with providers to monitor the diabetic's condition. Self-care is a critical component of the management process. For Type I and some Type II diabetics, this involves daily administration of insulin. The diabetic must inject proper amounts and types of insulin at the right times to maintain near-normal blood glucose levels. Because glucose levels can also be influenced by diet and exercise, adherence to dietary restrictions and exercise regimes is also an important part of effective self-care. (For some Type II diabetics, diet and exercise are used exclusively to control their condition.) In addition, some Type II diabetics use other pharmaceuticals to assist in glucose control. Finally, diabetic self-care also includes periodic monitoring of the diabetic's condition with various home tests to assess blood and urine glucose levels.

---

[4]   Source: Carter et al., 1996.

The providers' role in the management phase is primarily to help and encourage the patient to conduct effective self-care. A provider will determine an initial self-care protocol for the diabetic (specifying, for example, frequency of insulin injections and blood glucose tests) and will educate the diabetic on how and why to carry out this care program. For some diabetics, particularly young Type I diabetics, education of the family is also critical. After an initial period of developing the care protocol and educating the patient, the provider will recommend a program of ongoing interactions with providers. These ongoing interactions will be used to monitor the diabetic's condition, adjust the care protocol as needed, continue to reinforce patient education, and check for signs of complications.

Some complications, such as retinopathy, may be discovered during routine visits as part of normal disease management. Other complications may occur abruptly, even in an emergency situation, such as DKA. Many of these conditions require inpatient treatment. Although the original provider(s) who care for the diabetic during the management phase may remain involved with the case once an inpatient complication occurs, often other specialists will treat the specific complications.

Treatment of complications accounts for roughly two-thirds of the total inputs for diabetes treatment (**Exhibit 4**). (Recall that this analysis only addresses nonpatient labor inputs.) Complications are infrequent (only about 1 to 3 percent of diabetics develop a particular complication each year), but they are expensive because they usually require inpatient treatment. In contrast, the care provided in the management phase is far more frequent – typically several visits per year for each diabetic – but is inexpensive because it is routine outpatient treatment.

Management phase

Management phase decisions. Four types of decisions are made in the management phase of diabetes treatment (**Exhibit 5**).

> ¶ **Self-care protocol for the diabetic.** This protocol encompasses a number of specific decisions on types and frequency of self-care therapies and tests. For example, for insulin-dependent diabetics, the protocol for insulin administration is a critical part of self-care; this protocol indicates how often insulin should be injected, as well as the type(s) and amounts of insulin to be used.

> ¶ **Clinical setting (if any) for ongoing care of the diabetic**. Most diabetics follow some type of physician-guided ongoing care program throughout their lives. This program consists of a regular series of encounters with physicians and other caregivers to monitor the diabetic's condition and adjust the self-care protocol as needed. This ongoing care can be performed in a number of settings, where

the setting determines not only the physical location, but more importantly, the type and specialization of providers involved in the diabetic's care. While a wide range of settings has been used for diabetic treatment, we have considered the following categories in our analysis of setting decisions in the two countries:

- *Home care only.* The diabetic does not have any regular program of interactions with providers to monitor his condition.

- *Physician's office.* The diabetic receives ongoing care from a single physician (with some nursing assistance). The physician is usually a general practitioner (GP), although some specialists (typically endocrinologists) provide this routine care in the U.S.

- *Hospital outpatient clinic.* A single specialist provides ongoing care for the diabetic (again, with some nursing assistance) in a general hospital outpatient setting.

- *Dedicated diabetic clinic.* A team of professionals dedicated to diabetes provides ongoing care for the diabetic. The physical "clinic" is generally part of a hospital's outpatient facilities. The diabetes team varies widely in composition, but it usually includes a physician with a special interest or skill in diabetes treatment. Other professionals on the team may include a specialized diabetes nurse, a dietitian, a chiropodist, and an ophthalmologist. (See **Exhibit 6** for staffing of U.K. diabetic clinics.)

Dedicated diabetic clinics are used extensively in the U.K., where they are common in community hospitals, and increasingly as an adjunct to services offered by GP groups. In the U.S., these clinics are generally found only in some academic medical centers or specialized children's hospitals. Some ongoing care programs involve several of these settings; for example, a diabetic might use a GP, as well as a diabetic clinic.

¶ **Frequency of visits for ongoing care.** Depending on the severity of their condition, some diabetics may need to visit providers in these settings many times per year. Average visit frequency typically ranges from about two or three to about six visits per year.

¶ **Specific interactions in ongoing care visits.** Whenever a diabetic interacts with a provider, several decisions are made as to how to monitor the diabetic's condition and how to influence the self-care program. Many different tests can be performed, including general physical examinations, blood glucose and urine tests, a test for long-term control of blood glucose using "glycosylated hemoglobin" (HbA1c), and checks for symptoms of complications such as eye and foot exams. These encounters can also be used to address the patient's questions and concerns, provide counseling and education on all aspects of diabetes management, and adjust the self-care protocol if necessary.

To a large extent the actions pursued during a patient visit are influenced by the clinical setting for ongoing care. In general, more tests and more extensive patient interactions occur in dedicated diabetic clinics than in other settings because these clinics have more resources and more specialized personnel. In particular, during the timeframe of our analysis, on a per-visit basis, the U.K.'s diabetic clinics were more likely than U.S. providers to screen diabetics for glycosylated hemoglobin, urine protein and glucose, and visual acuity (**Exhibit 7**). Because there were more visits per year in the U.K. than in the U.S., the disparity in tests per year was even greater.

Simplified model of management phase decisions. To measure and explain productive efficiency differences between the U.S. and the U.K. in the late 1980s, we used a simplified model of these decisions in the management phase (**Exhibit 8**). This model incorporates the following simplifications:

¶ It does not include the decisions on the diabetic's self-care protocol. These decisions primarily affect patient labor, which we excluded from our input measurement.

¶ It represents only the aggregate effect on staffing resources of specific care decisions during patient visits, rather than identifying these individual decisions. For example, our simplified model does not indicate whether diabetics receive an HbA1c test in their routine visits, although this is clearly a clinical decision that must be made by providers at the time of a visit. Instead, the model shows the staffing resources (provider-hours) applied at each visit. Because more testing and counseling require more provider time, this staffing variable is a reasonable aggregate measure of the intensity of care provided at each visit.

¶ For simplicity of description, we classified the possible settings for routine care into two generic categories – clinic care and office care. The clinic category represents hospital outpatient clinics in the U.S. [5] and specialized diabetic clinics in the U.K. Obviously, a clinic in the U.S. has very different characteristics from a clinic in the U.K. In our decision tree for the management phase, this difference in clinic characteristics is reflected in the staffing levels each country employs in the clinic setting – the U.K. clinics have multidisciplinary teams of providers, while the U.S. clinics have only a single physician and a nurse.

Relative costs of different options. Cost of treatment in the management phase is a function of the setting chosen, the staffing levels in that setting, and the frequency of patient visits. In both the U.S. and the U.K., clinic care is more expensive than office care on a per-visit basis (**Exhibit 9**). The percentage differential is greater in the U.K. (where clinic care is more than twice as expensive per visit as office care) for two reasons: 1) U.K. clinics have a more complete set of providers; and 2) U.K. office visits are almost all with GPs, whereas some U.S. office visits are with specialists, who are more expensive.

The aggregate cost of the management phase across an entire health care system depends, to a large extent, on the relative frequency of using these two types of

---

[5] Because dedicated diabetic clinics in the U.S. are not common in community hospitals or physician practices, we have assumed that the large majority of U.S. diabetics do not receive care in this setting. Our model, therefore, does not include the care provided to U.S. diabetics in those specialized medical centers that do offer dedicated diabetic clinics.

settings (or of using neither, as in a home care-only care program). As we will discuss later, there are significant differences in these relative frequencies between the two countries.

Therapeutic benefit of different options. More effective treatment in the management phase clearly plays a role in delaying or preventing complications. However, the cause and effect linkages are complex and not well-understood. In theory, more effective management can reduce or delay complications in two ways: 1) by improving the metabolic control that diabetics achieve through self-care; and 2) through earlier detection and treatment of complications.

¶ **Better metabolic control**. Several studies have suggested that better metabolic control (maintaining blood glucose closer to normal levels) can contribute to lower incidence of some complications. One comprehensive study of Type I diabetics, the DCCT, found a clear and substantial decrease in the incidence of several complications (retinopathy, nephropathy, and neuropathy) with better glucose control. Although the study did not explicitly measure incidence of blindness, foot ulceration, and amputation, it is reasonable to expect that improved control would also affect these complications, since it clearly affects some of their contributing causes (retinopathy for blindness, and neuropathy for ulceration and amputation).

The DCCT represented an extreme case of control in that diabetics were given intensive treatment and monitoring at a level far beyond what would be obtained in typical diabetic care. Nevertheless, the DCCT suggests that there may be a general relationship between degree of glucose control and some complication rates, at least for Type I diabetics. The relationship for Type II diabetics is less clear, particularly since the method of glucose control for most Type II diabetics is diet and exercise, rather than insulin therapy. However, many experts believe that better glucose control in Type II diabetics (via better compliance with diet, exercise, and pharmaceutical programs) would also lead to lower complication rates since the long-term effect of hyperglycemia (excess glucose in the bloodstream) may be similar in both types of diabetics.

¶ **Earlier detection and treatment**. Some complications can be prevented if symptoms are discovered early and intervention is timely. In particular, both blindness and lower extremity amputation are to some degree preventable if their early forms – retinopathy and ulceration – are detected and treated. Blindness can be prevented or delayed for some diabetics by treating retinopathy with laser therapy. Amputation can be prevented or delayed by treating foot ulcers promptly and by encouraging diabetics to examine their feet regularly for injuries and take preventive measures, such as using proper footwear.

Advantages of dedicated diabetic clinics. The U.K.'s dedicated diabetic clinics offered potential advantages for diabetics through both better metabolic control and earlier detection and treatment of complications.

¶ **Impact of dedicated diabetic clinics on metabolic control.** Because the patient's self-care drives metabolic control (i.e., maintenance of near-normal glucose levels), it is difficult to identify exactly what it is about a care program that causes the diabetic to behave desirably. However, the resources available in the U.K.'s diabetic clinics for education and counseling likely contributed to better patient understanding and compliance. One indirect indicator of the level of attention to metabolic control in each country is the frequency of insulin injections for Type I diabetics. In the U.K., about 90 percent of the Type I diabetics used multiple insulin injections per day, compared to about 60 percent in the U.S. (**Exhibit 10**). Since at least two insulin injections per day are usually needed to achieve tight metabolic control in diabetics, the frequent use of single, daily injections in the U.S. suggests that tight metabolic control could not have been achieved as frequently in the U.S. as in the U.K.[6] While this is only one indicator of metabolic control, it suggests that, for Type I diabetics at least, providers in the U.K.'s diabetic clinics may have encouraged more aggressive treatment of diabetes or obtained better compliance to insulin regimes during the late 1980s. (Some clinicians believe that in the past 5 years, diabetic treatment in the U.S. has begun to place more emphasis on glucose control than it did during the timeframe of our analysis.)

¶ **Impact of dedicated diabetic clinics on detection and treatment of complications.** It is clear that more extensive testing and intervention during routine visits can result in earlier detection and treatment of complications. The more extensive testing performed at dedicated diabetic clinics in the U.K. is, therefore, likely to have resulted in superior outcomes (see again **Exhibit 7**).

These generalizations must be tempered by a recognition that more intensive management of diabetes may be unnecessary for many diabetics. While some diabetics, such as those with acute conditions or severe behavioral issues, may benefit from intensive management, others may be very well served with relatively little provider interaction. Achieving the best balance between inputs and outcomes may, therefore, depend largely on matching diabetics to the right management programs.

---

[6] We are unable to determine whether the higher frequency of insulin injections in the U.K. was caused by greater provider effort or different preferences on the part of patients; the result in either case, however, is likely to be tighter metabolic control in the U.K.

As we discuss below in the section on provider behavior differences, our analysis provides some insight into the degree of matching achieved by each country since we were able to measure the management phase care provided to Type I and Type II diabetics separately. We would expect that Type I diabetics typically require more intensive management phase care because of the severity of their condition (although Type II diabetics might require more care for associated conditions). Within Type II diabetics, we were able to measure the management phase care for insulin-using Type II diabetics and noninsulin-using Type II diabetics separately. Most Type II diabetics on insulin would be expected to have more severe diabetes than noninsulin users; differences in the care provided to these two subgroups, therefore, give us some indication of the way providers and administrators are allocating management phase care across the diabetic population.

Complications phase

Once a diabetic develops a complication, medical treatment is generally required to correct the condition (if possible) or to prevent further deterioration in health status. We have addressed treatment of three complications: DKA/hyperosmolar coma, retinopathy, and lower extremity amputation. (We have also measured the incidence of blindness in diabetics; but because this condition is typically permanent and incurable, it does not require treatment.) As described earlier, our analysis of the complications phase only considers the inpatient treatment of complications; DKA/hyperosmolar coma and lower extremity amputation are treated almost exclusively in the inpatient setting; and although outpatient treatment is now available for retinopathy, it was not widely used during the time of our study. Our model does not include additional outpatient care or consultations before or after inpatient treatment.

Complications phase decisions. Treatment of each type of complication requires a number of clinical judgments about the diabetic's condition and the best program of intervention. For example, when a lower extremity amputation is being considered, physicians must judge whether amputation is absolutely necessary, and if so, how much of the patient's leg to remove; judgments must be made in retinopathy about whether laser therapy is indicated, and if so, how to perform it. With the exception of our simplifying assumption that 60 percent of patients with retinopathy received laser treatment in both countries, we did *not* analyze these specific clinical decisions for each complication. [7] We assumed that

---

[7]   As there are no data on what percentage of patients receive laser treatment in the year of incidence and there is no evidence that the two nations differ systematically in their approach to the treatment of retinopathy, we made a simplifying assumption that 60 percent of patients with retinopathy received laser treatment in the year of incidence. It is possible that patients with retinopathy in the U.K. received laser treatment at a lower rate than those in the U.S. If this were true, our assumption that the same percentage of patients receive treatment in both cases would cause us to overestimate input usage in the U.K. relative to the U.S.; our estimate of lower relative inputs in the U.K. is, therefore, conservative.

standards of care, skills, and availability of equipment were roughly similar between the two countries, so that, on average, providers made similar clinical judgments for diabetics with equivalent conditions.

Instead of analyzing specific clinical decisions, we focused on the economic consequences of the organization of inpatient care for these conditions. There are two major drivers of input usage in treating complications: hospital length of stay (LOS) and hospital staffing levels. Our analysis used the simplifying assumption that the labor provided per day to treat each diabetic complication is the same as the overall average for all inpatient care. We did not explicitly measure the physician or hospital inputs associated with specific procedures, such as surgeries, since, in general, these inputs are relatively few compared to the inputs associated with the overall hospital stay.

This simple economic model of the decisions in the complications phase is shown in **Exhibit 11**. Obviously these aggregate variables are not truly decisions in the sense of specific clinical judgments. A provider does not decide to have an LOS for DKA of 6.4 days; instead, the LOS reflects a number of decisions across a large number of patients, such as timing of tests and therapies, administrative policies for patient flow, and criteria for discharge.

Relative costs of different options. In our simple economic model, the inputs used for inpatient treatment of a complication are simply the product of LOS and average hospital staffing per occupied bed. (Recall that we are measuring only labor inputs.) Obviously, longer LOS or higher levels of staffing per bed increase the inputs for treating complications.

Because our objective was to assess the overall productive efficiency of treating diabetes, we measured all inputs on a per-diabetic basis. The total input usage in the complications phase, therefore, also depended on the fraction of diabetics who incurred complications each year:

Annual input usage in complications phase per diabetic =

Annual complication rate X LOS X hospital staffing

Because of this complication rate effect, the effectiveness of the management phase in preventing complications had a direct impact on the inputs used in the complications phase. If a health system performed well in the management phase, it would reap the economic benefit of reducing costs in the complications phase.

Therapeutic benefit of different options. We assumed that providers in the U.S. and the U.K. made similar clinical decisions for equivalent complications and had access to the same knowledge, skills, and equipment. We, therefore, expected similar clinical results for treatment of complications between the two countries, in spite of differences in LOS and hospital staffing levels. These

differences in the organization of inpatient care had clear *economic* consequences, but we have no reason to believe they affected clinical results.

ASSESSMENT OF RELATIVE PRODUCTIVE EFFICIENCY

The U.K. used significantly fewer inputs than the U.S. – 40-percent less for Type I diabetics and 32-percent less for Type II diabetics (**Exhibit 12**). On a weighted average basis (combining Type I and Type II diabetics), the U.K. used 34-percent fewer inputs than the U.S. for diabetics overall (**Exhibit 13**).

The U.K. also achieved better outcomes in diabetes than the U.S. Type I diabetics in the U.K. had 2.5 more QALYs than diabetics in the U.S. In Type II diabetes, the U.K. achieved 1.2 more QALYs than the U.S. (**Exhibit 12**). On a weighted average basis, U.K. diabetics had 1.35 more QALYs than U.S. diabetics. Looking at the improvement in outcomes over the baseline case of no treatment,[8] U.K. diabetics achieved 27-percent greater improvement in outcomes due to treatment than U.S. diabetics did (**Exhibit 13**).

With better outcomes and fewer inputs, the U.K. was clearly more productive than the U.S. in diabetes treatment. The U.K.'s productive efficiency advantage stemmed from its consistently lower complication rates. Although complication rates were relatively low in both countries (roughly 1 to 3 percent for most complications), the "compounding" of these annual rates over a diabetic's lifetime created a significant difference in overall outcomes. The U.K.'s outcome advantage for Type I diabetes was greater primarily because Type I diabetes occurs at a younger age, so the U.K.'s advantage in complication rates compounded over a larger number of years.

The input differences between the U.S. and the U.K. were also substantial. As discussed below, these input differences were also driven primarily by the U.K.'s lower complication rates, which led to lower consumption of resources in the complications phase.

MAJOR DRIVERS OF PRODUCTIVE
EFFICIENCY DIFFERENCES IN TERMS
OF PROVIDER BEHAVIOR DIFFERENCES

The productive efficiency differences observed were caused most directly by differences in provider behavior. In this section, we discuss the provider

---

[8] Baseline outcome with no treatment conservatively assumed to be that Type I diabetics die within 1 year and Type I diabetics have the same QALYs as the lowest outcomes with treatment (U.S.).

behaviors that were the major drivers of productive efficiency differences between the nations, focusing first on drivers of input differences.

Drivers of input differences

Almost all of the U.K.'s advantage in input usage was due to significantly lower resource consumption in the complications phase; the U.S. and U.K. used similar inputs for the management phase (**Exhibit 14**). Because the complications phase consumed far more resources than the management phase, this savings in the complications phase resulted in a large savings in total inputs.

While the differences in management phase inputs were relatively small, they suggest that the U.K. took a different approach to the management phase than the U.S. We will describe specific differences in approach with Type I, Type II insulin users, and Type II noninsulin users in the next section. In aggregate, our analysis suggests that the U.K. triaged care among diabetics, allocating more inputs to more severe cases, while in the U.S. the levels of care diabetics received appeared not to vary with their condition.

1. Management phase

We discuss the decisions made in management phase care in the U.S. and U.K. using the simplified decision model described earlier (see again **Exhibit 8**). Decisions for Type I and Type II diabetes are described separately; in addition, we distinguish between Type II insulin users and Type II noninsulin users. (Use of insulin in Type II diabetics can be viewed as a crude marker of more severe diabetes.) For each class of diabetic, we show each country's frequency of using each care setting, the average number of patient visits per year in each setting, and the staffing levels used per visit in each setting. We also show the impact of each of these decisions on the total inputs used (management plus complications) in treating this class of diabetic. (This total input impact is shown as a percentage increase or decrease for the U.K. relative to the U.S. total inputs.) The total input impact shown does *not* include the ultimate effect of management phase decisions on complication rates and, hence, on inputs in the complications phase. It only measures the direct effect of the decisions on management phase inputs.

Type I diabetes (Exhibit 15). Both the U.S. and the U.K. provided management phase care to the majority of their Type I diabetics in a clinic setting, with the U.K. treating 82 percent of Type I diabetics in diabetic clinics, and the U.S. treating 75 percent of Type I diabetics in hospital outpatient clinics. In addition, the U.S. treated many Type I diabetics in general office settings (50 percent), while the U.K. treated very few in this setting (13 percent). (Note that the U.S. percentages for clinic and office care add to more than 100 percent because some diabetics received care in both settings.) It is possible that the U.K. provided no ongoing physician-guided management care to a small minority of Type I

diabetics (10 percent), while this was not done in the U.S.[9] The net effect of these differences in setting choice was a decrease of about 4 percent of total inputs for the U.K. relative to the U.S., due primarily to significantly lower use of the physician office setting.

For patients seen by providers in a clinic setting, the frequency of visits per diabetic per year was higher in the U.K. (5.1) than in the U.S. (3.6). Similarly, visit frequency for office visits per year was higher in the U.K. (5.2) than the U.S. (3.4). This increased total inputs per diabetic in the U.K. by 13 percent relative to the U.S.

As we described earlier, staffing resources applied per visit were slightly higher in the U.S. general hospital outpatient clinic setting than in the U.K. diabetic clinic setting. Likewise, resources used in the U.S. office setting were higher than in the U.K. (see again **Exhibit 9**). These differences had a combined effect of decreasing inputs per diabetic in the U.K. by 8 percent relative to the U.S.

Overall, the U.K. used slightly fewer inputs (1 percent) in the management of Type I diabetics. While the U.K. had greater clinic use and greater visit frequency than the U.S., these factors were offset by higher staffing in U.S. outpatient clinics and higher use of office-based care in the U.S.

Type II diabetes (Exhibit 16). *Insulin-using Type II diabetics* – the U.K. treated insulin-using Type II diabetics very much like it treated Type I diabetics; most (82 percent) of these diabetics were managed in diabetic clinics, with about five visits per year per diabetic. In the U.S., however, office-based care was the dominant setting for the management of Type II insulin users, with fewer than half of these diabetics attending a clinic. As with Type I diabetics, visit frequency in the U.K. was consistently higher than in the U.S. for both office and clinic settings.

The net effect of these decisions was to increase total inputs in the U.K. by about 9 percent over U.S. inputs. Most of this increase (19 percent) was due to higher visit frequency in the U.K. From an input perspective, the greater use of clinic care for insulin-using Type I diabetics in the U.K. was offset by the lower use of office care.

Noninsulin-using Type II diabetics (**Exhibit 17**) – about 70 percent of all diabetics are noninsulin-using Type II diabetics. Decisions on how to treat this group of diabetics, therefore, had the largest impact on overall system input usage. The U.K.'s pattern for management phase care for noninsulin-using Type II diabetics was very different from its pattern for Type I diabetics and insulin-using Type II

---

[9] The figures indicate that some Type I diabetics did not receive a physician-guided care through the National Health Service (NHS). We were unable to determine whether this lack of care was a statistical artifact, patients received care outside the NHS (i.e., from private consultants) or from nonphysician providers, or patients truly did not receive care.

diabetics. Few of these diabetics (only 16 percent) were treated in a clinic setting. About half were seen in an office setting, but more than 40 percent received no physician-guided care at all. Visit frequency for those seen in clinics was also lower than it was for insulin-using diabetics, at about four visits per year versus about five per year for insulin users. In the U.S., almost all of the noninsulin-using Type II diabetics (93 percent) were treated in an office setting, with a small number receiving only home care.

Both countries provided less intensive service to this group of diabetics than to either Type I diabetics or insulin-using Type II diabetics. The differences were more dramatic in the U.K., however, where most insulin-using diabetics received clinic care, but many noninsulin-using diabetics received only home care. This use of home care instead of office care for many noninsulin-using Type II diabetics reduced total inputs in the U.K. by about 16 percent relative to the U.S. While this savings was offset by other factors (primarily greater visit frequency in the U.K. for those who do receive ongoing care), the net effect was that management phase inputs in the U.K. for Type II noninsulin users were 7-percent lower than those in the U.S. In contrast, management phase inputs for Type II insulin users were higher in the U.K.

Summary of management phase input distribution. Combining the input impact of management phase decisions across the different types of diabetics shows a clear triaging effect in the U.K. **Exhibit 18** shows the distribution of input usage per diabetic in each country. In the U.K., about one-third of the diabetics (mostly insulin users) received intensive care in a clinic setting; a third received a moderate level of care in an office setting; and a third (mostly noninsulin-using Type II diabetics) received no physician care at all. This pattern of triaging differs significantly from the pattern in the U.S., where most diabetics received a moderate level of care (primarily in an office setting). In addition, while the inputs used for "moderate" care were roughly similar between the two countries, the U.K. spent more for intensively managed diabetics because of its dedicated diabetic clinics and more frequent visits.

These different patterns of care allocation suggest a greater willingness or ability on the part of providers and administrators in the U.K. to discriminate among diabetics for the type of management phase care they received. Because most of those receiving intensive care in the U.K. were insulin users and most of those receiving only home care were noninsulin-using Type II diabetics, we can hypothesize that this discrimination was based on a perceived need for services due to the severity of the diabetic's condition. [10] In contrast, the U.S. discriminated relatively little among diabetics – there was both less investment in

---

[10] Note, however, that there may be other factors at work as well. For example, even some Type I diabetics in the U.K. receive only home care. While this may represent a sound clinical judgment that some Type I diabetics do not need a physician-guided care program, it is also possible that services may be constrained or unavailable for some segments of the population.

intensive management and less willingness or ability to forego physician-guided management care completely for some diabetics. Given the large differences in severity of diabetes across the diabetic population, we would expect that the spending pattern in the U.K. had the potential to be more cost-effective.

2. Complications phase

As with the management phase, we will discuss the decisions made in the complications phase and the impact of these decisions on inputs using the simplified model described earlier (see again **Exhibit 11**). Decisions for Type I and Type II diabetes will be described separately. However, our data sources for hospital LOS and average hospital staffing levels did not distinguish between Type I and Type II diabetics. We assume that, after a diabetic gets a complication, the inputs used in treating the complication did not depend on whether the diabetic was Type I or Type II. (This assumption may not be strictly correct from a clinical standpoint, since differences in type may affect co-morbidities, specific treatment protocols, and recovery rates. However, it seems to be a reasonable approximation.) The only difference in our complications phase model between Type I and Type II diabetics is, therefore, the *complication rates* incurred by these classes of diabetics. Because we did not have complication rate data for insulin-using and noninsulin-using Type II diabetics separately, we were not able to measure complications phase inputs for these subtypes separately, as we did for the management phase.

Type I diabetes (Exhibit 19 and Exhibit 3). Complication rates for Type I diabetics in the U.K. were about half the rates in the U.S. (with an average of 1.4 percent across the three diseases studied for the U.K. versus 3.0 percent for the U.S.). These lower complication rates reduced total inputs in the U.K. by 38 percent relative to the U.S. since there were fewer complications requiring inpatient treatment.

On an inputs-per-complication basis, the U.K. used slightly fewer (1 percent) resources than the U.S. This difference was the net effect of two offsetting factors. Hospital LOS for the three complications studied were longer in the U.K. than in the U.S. by about 32 percent (with an average LOS of 11.2 days in the U.K. versus 8.5 days in the U.S.). However, hospital staffing levels (per inpatient-day) were substantially lower in the U.K. than in the U.S. The U.K. averaged 4.1 specialist-hour equivalents per inpatient-day, as compared to 5.7 in the U.S.

The combination of lower complication rates and slightly lower input usage per complication resulted in a net savings of 40 percent for the U.K. relative to the U.S.

Type II diabetes (Exhibit 20). The complications phase analysis for Type II diabetes was very similar to Type I. As with Type I diabetics, complication rates for U.S. Type II diabetics were significantly higher (57 percent) than those in the U.K.; the U.K. had an average complication rate of 0.9 percent across the four

complications studied versus 1.6 percent for the U.S. These lower complication rates had a dramatic effect on total inputs; the U.K. achieved a 35-percent savings relative to the U.S. by treating fewer complications.

Since we used common data for Type I and Type II LOS and hospital staffing levels, we see the same effect in inputs per complication treated: lower staffing levels in the U.K. offset longer LOS for a slight net decrease in total inputs in the U.K. relative to the U.S.

For Type II diabetes, the net effect of lower complication rates and lower input usage per complication was a 29-percent reduction in total inputs for the U.K. relative to the U.S.

Summary of drivers of input differences

The major driver of input differences for both Type I (**Exhibit 21**) and Type II (**Exhibit 22)** diabetes was differences in complication rates. Lower complication rates reduced inputs in the U.K. relative to the U.S. by 38 percent for Type I diabetics and by 26 percent for Type II diabetics. Secondary factors included lower hospital staffing and longer LOS in the U.K.; these factors worked in opposite directions and, thus, had an insignificant net effect on total inputs.

Decisions in the management phase had a relatively insignificant *direct* effect on total inputs, since the management phase represents only a small portion of total diabetes treatment inputs. However, to the extent that management phase decisions affected complication rates, they were critical, indirect drivers of total input differences.

Summary of drivers of productive efficiency differences

Similar to input differences between the U.K. and the U.S., productive efficiency differences were driven by the U.K.'s lower complication rates. Our challenge is, therefore, to identify the aspects of the U.K.'s diabetes treatment process that led to these lower complication rates. [11]

Although it is difficult to demonstrate a direct cause-and-effect relationship, the organization of care in the management phase can clearly influence providers and patients to adopt more effective behaviors that reduce complication rates. From our analysis, we can identify two potential advantageous provider behaviors in the U.K.: more intense care triaging, and the use of a team-based approach (**Exhibit 23**). While it is not possible to determine quantitatively how

---

[11] As with all of the diseases, we are making an implicit assumption that the populations of the two countries are sufficiently similar that outcome differences reflect differences in treatment rather than population characteristics. As described earlier, we compared the U.K. population to U.S. whites to remove potential racial biases in our results.

much impact each of these factors had on complication rate differences between the two countries, both were potentially important.

Care triaging. In contrast with the U.S.'s more uniform approach to the treatment of diabetes, the U.K. differentiated among diabetics for the type and amount of care provided. This approach may have enabled the U.K. to achieve better outcomes by more effectively matching care programs to the diabetics' needs.

> ¶ **Less treatment for some diabetics in the U.K.** For some diabetics, generally those considered to have the least severe conditions, the U.K. provided less treatment than the U.S.; more than 40 percent of noninsulin-using Type II diabetics in the U.K. received only home care, while 93 percent of these diabetics in the U.S. were treated by a physician.

> ¶ **More intense treatment for some diabetics in the U.K.** For the two-thirds of diabetics in the U.K. who received some form of physician-guided care, routine visits with providers occurred about 5 times per year, as opposed to an average of 3.5 visits per year in the U.S. For the one-third of U.K. diabetics seen in a diabetic clinic, visits were also more comprehensive than comparable visits in the U.S. The U.K. thus provided more intense treatment to diabetics generally considered to have the most severe conditions.

Use of a team-based approach. The U.K.'s diabetic clinics not only offered more provider attention to certain diabetics but, importantly, they offered care from many different types of providers in a multidisciplinary team. The provider team in a U.K. diabetic clinic might have included a diabetologist, an ophthalmologist, a chiropodist, a dietitian, and a nurse specialized in diabetes. This team might have been more effective than a single physician in assessing the diabetic's condition, developing a self-care program, and educating and counseling the diabetic.

Other potential drivers of complication rate differences

We have attempted to explain the better outcomes achieved by the U.K. in diabetes by analyzing differences in provider behavior; our analysis points to the U.K.'s triaging of patients and use of multidisciplinary teams as the cause of the U.K.'s lower complication rates. Clinicians in both nations believe that significant gains can be made in diabetes treatment from providing intensive and multidisciplinary care to diabetics with severe conditions. Clinical trials, such as the DCCT, have also suggested that tighter management of certain diabetics can prevent or delay complications from occurring. Thus, allocation of intensive care to a population likely to benefit may explain the improved outcomes.

It is possible, however, that other factors, which we were unable to measure or control, are at least partially responsible for the U.K.'s lower complication rates from diabetes. If the U.K. had higher mortality rates from diabetes than the U.S., this could have caused the U.K. to have lower morbidity rates as their sickest patients would have died rather than presenting with complications. However, as discussed in Appendix 3B, while there is no definitive study that permits an unambiguous comparison of mortality rates between the two nations, evidence suggests that the U.K. had lower mortality rates than the U.S. for diabetes; we can, therefore, rule out the possibility that the U.K.'s lower complication rates were caused by a significantly higher mortality rate for diabetics in the U.K. relative to the U.S.

The U.K.'s better outcomes for diabetes could also have been partially caused by behavioral differences in the U.K. and U.S. population as patients; if the U.K. diabetics were "better" patients (patients who take better care of their conditions), this could have led to the lower complication rates observed. It might also be possible that access issues contributed to worse overall outcomes in the U.S.; if there was a group of diabetics in the U.S. who did not have access to care and, therefore, had very poor outcomes, the U.S.'s 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 U.S., we were unable to determine the role these factors might have played in the relative complication rates.

Additionally, if the U.S. diabetic population had higher levels of obesity than the U.K. diabetic population, which may be likely due to the generally held belief that the U.S. population has higher levels of obesity than most other nations, this could at least partially explain the higher U.S. complication rates for Type II diabetics; Type I diabetics, however, who are generally younger and not as subject to obesity, would be largely unaffected by this difference.

**CAUSAL ANALYSIS OF PROVIDER BEHAVIOR
DIFFERENCES: INCENTIVES AND CONSTRAINTS,
SYSTEM STRUCTURE, AND REGULATION**

The differences in provider behavior in the U.S. and U.K. can be explained by the incentives and constraints providers faced in each country and by underlying differences in the health care system structure and regulation, which are described in detail in Chapter 2. Below, we explain how these differences led to the provider behavior differences we observed in the treatment of diabetes. We focus on the two differences in management phase provider behavior – care triaging and the use of a team-based approach – that contributed to lower complication rates in the U.K. Although the LOS was longer in the U.K. and

hospital staffing was higher in the U.S., these two effects in combination had an insignificant impact on overall input and productive efficiency differences.

Incentives for U.K. physicians to triage care and use a team-based approach; incentives for U.S. physicians to treat all diabetics themselves

U.K. physicians had strong incentives to triage care – to provide intensive treatment to some diabetics while providing minimal treatment to others – and to use a team-based approach. U.S. physicians, on the other hand, had incentives to treat all diabetics and to provide treatment themselves (**Exhibit 24**).

The sources of these different physician incentives were major differences in health care system structure, particularly in the degree of competitive intensity and integration and pricing of care provision products negotiated between payors and physicians (**Exhibit 25**). Most physician services in the U.S., including both specialist and primary care, were negotiated and compensated on a fee-for-service (FFS) basis by payors. U.S. physicians also faced the threat of malpractice suits. U.S. physicians – both GPs and specialists, both of whom saw diabetic patients – therefore had incentives to treat all diabetics rather than recommending that the least severe cases provide treatment for themselves at home. In the U.K., on the other hand, specialists were paid flat salaries negotiated on an annual basis; this method of payment gave them little incentive to treat patients who were able to receive home care only.

Not only did physician incentives lead U.K. physicians to provide less care for some diabetics, but physician incentives also led U.K. physicians to provide more intensive and team-based care for other diabetics. The question is why the NHS chose to make this "investment" for the severe diabetics, while the U.S. system, as a whole, did not. One possible explanation is that the payback for an investment in management phase care for diabetes is relatively long term, since the return occurs many years later in reduced complication rates. As the NHS covers all health care expenses for the entire lifetime of the U.K. population, the NHS had the incentive to make these investments. And since the NHS integrated the payor and provider side of health care within one organization, it served as the infrastructure under which multidisciplinary teams could be formed. In 1954, the NHS issued a recommendation that one diabetic clinic be established per region; since then, the clinics have continued to develop with the NHS's funding and support.

In the U.S., however, the FFS product that providers offered patients and payors was not integrated in terms of care coordination or over time. In addition, the U.S.'s decentralized multiple payors did not generally coordinate the activities of providers, making the grouping of providers into teams far more difficult. As providers typically were only reimbursed

for medical care they directly provided to patients themselves, and not for patients' education or other services, they had no incentive to provide additional services to diabetics or to coordinate with other providers to provide multidisciplinary care. U.S. payors also had little incentive to provide intensive or multidisciplinary care to diabetics; since payors faced high churn in members (e.g., up to 40 percent annually), they likely would not benefit from making up-front investments to prevent complications that typically occur years later.

Differences in physician incentives in the U.S. and the U.K. were also driven by differences in the competitive intensity of the nations' health care system structure. As U.S. providers competed for patients, they had less incentive to coordinate care with other providers, but rather, had incentives to treat all patients themselves. In the U.K., however, neither specialists nor GPs competed in any meaningful way for patients and, therefore, were not wary of recommending home care only for patients or of referring patients to other providers or to diabetic clinics.

Differences in competition among payors in the two nations were also very important in driving physician incentives. The U.K.'s provision of less care to some diabetics was sustainable because the single-payor system of the NHS provided no alternative (except private insurance) for diabetics who were seeking more treatment. In the U.S., competition among payors for members made it more difficult for payors and, therefore, providers to triage, since consumers (or their employers) who received a smaller allocation of care resources could seek alternative health coverage.

Interestingly, while payor competition for members in the U.S. gave payors and providers incentive to treat all diabetics, it likely simultaneously led to a disincentive to provide very intensive or multidisciplinary care for diabetics. Given inadequate insurance market pricing for the risk of diabetics, U.S. payors who encouraged the use of specialty clinics for diabetes would have faced an adverse selection problem as diabetics – who would have had more expensive claims – moved to this payor for better care. The U.K., however, faced no such issue since the NHS was and is a single payor for the entire population.

Tight controls on physician supply in the U.K. through regulation

The U.K. exercised strict controls over the number of physicians through the NHS budgeting process and regulation. This resulted in a relatively limited supply of both GPs and specialists in the U.K. (see again **Exhibit 25**). This limited supply of GPs created pressure for GPs to recommend limited care (e.g., home care only) for cases without significant needs, even though they were compensated on an FFS basis (see again **Exhibit 24**). The constrained supply of specialist physicians in the U.K. limited the capacity of dedicated

diabetic clinics, creating pressure for GPs and specialists to limit treatment in diabetic clinics to those diabetics who most needed the clinic services. The constrained supply of specialists also made treatment in a centralized clinic more attractive as it enabled specialists to cover diabetics in a broad area more easily. Regulation in the U.K. of physician supply thus created incentives for both care triaging and the use of multidisciplinary teams in centralized diabetic clinics.

SUMMARY OF DIABETES CASE RESULTS

The U.K. was more productive than the U.S. in the treatment of diabetes. The source of this difference was differences in provider behavior between the two countries – namely more intense care triaging and use of a team-based approach in the U.K. – which led to lower complication rates for both Type I and Type II diabetics. These differences in provider behavior were caused by tight controls on physician supply in the U.K. and incentives for U.K. physicians to triage care and use a team-based approach, as opposed to incentives for U.S. physicians to treat all diabetics themselves.

These incentives and constraints were caused by major differences in the health care system structures and regulatory environments of the two countries. As the NHS covered all patients for life and integrated care across providers, the U.K. had the incentive and ability to establish diabetic clinics and provide intensive care to diabetics most in need of treatment. In the U.S., however, payors experienced high churn in members and, therefore, did not have an incentive to make long-term investments in diabetic care; in contrast, fear of adverse selection in the competitive U.S. payor market may have given U.S. payors incentive to avoid providing specialized care to diabetics and, thereby, attracting more of them. In addition, provider competition for patients in the U.S. created incentives for providers to treat patients themselves rather than coordinate with other providers.

Appendix 3A: Definition of outcome measure for diabetes

The diabetes care process impacts diabetics in two ways. First, it increases their life spans by preventing or delaying death from diabetes-induced conditions. Second, it improves the quality of their lives by preventing or delaying several types of early and late stage complications, such as foot ulceration or amputations. To define a single outcome measure for the diabetes treatment process, we combined these different aspects of the benefit of treatment into a single measure.

Before describing our methodology for this outcome measure, we address two preliminary issues below: first, we review the source data we used to measure individual aspects of the outcome of diabetes treatment. These data constrained the type of overall outcome measures we can realistically compute. Second, we review the relative ranking of the U.S. versus the U.K. on these individual aspects of treatment outcome. This comparison demonstrates that in essentially every aspect of treatment, the U.K. achieved superior outcomes. As a consequence, any aggregate measure of outcomes would show the U.K. to be superior. This is an important fact to consider when reviewing the assumptions behind our proposed aggregate outcome measure, since it demonstrates that although some of these assumptions could be challenged, the qualitative result (U.K. has better outcomes) is not sensitive to these assumptions.

After addressing these preliminary issues, we define an overall numerical outcome measure for the diabetes process and explain how it is calculated.

DATA AVAILABLE ON DIABETES OUTCOMES

We measured two aspects of diabetes treatment: 1) the incidence of certain *complications* in the diabetic population; and 2) indicators of the *mortality* rates associated with diabetes.

1. Complication incidence data

For some complications, data were available on the percentage of diabetics who incur each complication annually in both the U.S. and the U.K. As noted previously, we compared complication rates for the U.K. population with U.S. whites to eliminate potential racial biases in the outcomes. We used *annual*

incidence rates of complications as our basic measurement of the success of the diabetes treatment process. (Sources for complication rate data are described in Appendix 3C). While complication incidence rates are good indicators of diabetics' quality of life, it is important to recognize several limitations of these measures:

¶ They do not cover all complications of diabetes. In particular, we excluded occurrence of cardiac and renal disease from the complication measures. While these are important complications, they are difficult to measure; and many factors other than diabetes treatment can affect the incidence and outcomes of cardiac and renal conditions.

¶ Available data on complication rates are often aggregate measures across the entire diabetic population. We were usually not able to separate these incidence rates by age or by duration of diabetes since many studies of diabetic complications do not provide this level of detail. [12] Age- or duration-specific incidence rates would allow a more precise model of the impact of complications on a diabetic's health across his entire lifetime. We made a simplifying assumption of constant incidence rates over a diabetic's lifetime.

¶ We also did not have information on the joint occurrence of multiple complications (for example, the impact of having retinopathy on the likelihood of developing lower extremity ulceration). As we discuss below, we, therefore, assumed *independence* of different complications to simplify the modeling of outcomes.

2. Mortality indicators

Unlike lung cancer and breast cancer, there are no comprehensive registries of diabetics that would have allowed us to track mortality over time. We were, therefore, unable to develop mortality curves for diabetics or to calculate simple aggregate measures of mortality such as life years (LYs)saved over a 5-year period. However, several indirect (and imperfect) indicators of mortality can serve as the basis for comparisons between the U.S. and the U.K. They include recorded death rates from diabetes, death rates adjusted for diabetes prevalence and multiple causes of death, and mortality rates from diabetic cohort studies in the U.S. and the U.K. These mortality indicators are discussed in detail in Appendix 3B; in the following section, we present overall results from comparisons of these indicators.

---

[12] Type I and Type II incidence rates are often separately available; in other cases, we have estimated type-specific incidence. See Appendix 3C for details.

**COMPARISON OF U.S. TO U.K. ON
INDIVIDUAL ASPECTS OF OUTCOMES**

On every indicator of diabetes morbidity and mortality, the U.K. achieved better outcomes than the U.S. Complication incidence rates in the U.K. were consistently about half those of the U.S. (**Exhibit A-1**).

As discussed above, comprehensive mortality data for diabetics are not available. However, several indirect indicators of mortality also demonstrated better performance in the U.K. (**Exhibit A-2**). (These mortality comparisons, and the difficulties in drawing definitive conclusions from them, are described in detail in Appendix 3B.)

Because the U.K. had lower complication rates and apparently lower mortality, essentially any aggregate outcome measure that combines these complication and mortality measures would show the U.K. as having superior outcomes.

OUTCOME MEASURE DEFINITION

QALY model

Our outcome measure for the diabetes treatment process modeled *the expected QALYs* for an average diabetic after the onset of diabetes. This measure summed, over a diabetic's lifetime, the expected "utility" of his health state at any point in time, where utilities are normalized so that a completely "well" year has utility of 1 and death has a utility of 0. The specific health states used in our QALY model and the utilities assigned to each state are described below. The *expected* QALY for an average diabetic depends on the probabilities of being in each health state at each point in time; these probabilities depend on the *complication rates* for the specific country and type of diabetes being modeled. As discussed above, complication rates in the U.K. were consistently lower; in our QALY model, this translated into lower probabilities that diabetics in the U.K. were in "poor" health states with low utilities over time. Expected QALY for U.K. diabetics was, therefore, greater than expected QALY for U.S. diabetics.

Absolute QALY versus improvement in QALY

In theory, our outcome measure should identify the *improvement* in QALYs resulting from the diabetes treatment process. Untreated diabetics would not all die immediately (although Type I diabetics die quickly without insulin), and they would not all get complications, so it is not strictly correct to treat all of the QALYs observed for diabetics as an outcome of the diabetes treatment process. Data on the complication rates or mortality for diabetics without any treatment are not available, so we were unable to measure the impact of treatment. As we have no reason to believe that untreated diabetics in the U.S. would have had

different outcomes than untreated diabetics in the U.K., comparing the absolute QALYs achieved by diabetics in each country yields a valid measurement of the difference in the outcomes of treatment between the two countries.[13]

Health states modeled

Each of the four complications we studied – DKA/hyperosmolar coma, lower-extremity amputation, sight-threatening retinopathy, and blindness – affects a diabetic's quality of life. However, the first complication – DKA/hyperosmolar coma – has only a *temporary* effect on quality of life since it is a curable condition. The other three complications – retinopathy, amputation, and blindness – have a permanent effect on quality of life. [14] The impact of a "temporary" complication, such as DKA, on a diabetic's QALY over his entire lifetime is very small compared to the impact of a "permanent" complication, such as blindness; we, therefore, focused on the permanent complications in modeling diabetes outcomes. (We also tested more complex models that take the temporary complications into account. Because temporary complications have a relatively negligible impact on QALY, these more complex models provided results very similar to the simplified model presented below.)

Diabetics can experience multiple complications simultaneously. For example, a diabetic may be blind and also have an amputation. Our model of diabetes outcomes, therefore, considers the following seven health states:

1. Healthy (no complications)

2. Retinopathy only

3. Blindness only

4. Amputation only[15]

5. Retinopathy and amputation

6. Blindness and amputation

7. Death.

---

[13] Our assumption here is that the U.S. and U.K. diabetic populations do not differ in some fundamental ways, such as genetic characteristics or lifestyle, that would affect complication rates.

[14] Retinopathy can be treated to prevent it from advancing to blindness; however, we assume here that a diabetic with retinopathy experiences a permanent degradation in vision even if full blindness is prevented.

[15] Although amputation is permanent, a diabetic can incur multiple amputation events, such as amputation of a leg following amputation of a foot, or amputation of the other leg following amputation of the first leg. We have used a simplified health states model that considers only the first amputation and treats subsequent amputations as not changing the health state. We have also implicitly treated the amputation complication rate as representing the first amputation; again, this is a simplification since subsequent amputations would also be reflected in amputation rate data.

(Note that the combination of retinopathy and blindness is equivalent to blindness alone, since vision status cannot be worse than blindness.)

Assignment of QALY scores to health states

A diabetic who experiences complications necessarily has a lower quality of life than one who is complications-free. The *extent* of the reduction in quality of life is fundamentally a subjective issue; different individuals may assign very different utilities to the various health states that may result from diabetes. However, to make measurement of diabetes outcomes tractable, we needed to create a concrete model of how individuals' utilities were affected by complications.

Specific QALY scores or "disutility weights" for each health state were developed using the Kaplan-Bush Index of Well-Being scale.[16] This scale was developed from a survey of several hundred people in the San Diego area that measured relative preferences for various health states. It assigns a well-being score based on an individual's condition on several dimensions, such as mobility, ability to perform tasks, and presence of symptoms causing pain or impaired function. For each complication combination, we estimated what the Kaplan-Bush Index would be for a diabetic who experienced those complications and had no other health problems. (An actual population of diabetics would have a range of other conditions that would also influence their well-being indices.)

This analysis yielded the following QALY scores (disutility weights) for each health state. (Note that these weights are *annualized* levels.)

Health state	Assumptions	QALY Weight
Healthy (no complications)	No other health problems	1.0000 0.0000
Sight-threatening retinopathy only	Vision impaired; otherwise normal function	0.7623 0.2377
Lower extremity amputation only	Cannot drive; walk with limitations; some activities limited	0.4772 0.5228
Blindness only	Cannot drive; walk with limitations; some activities limited	0.5592 0.4408
Retinopathy and amputation	Cannot drive; walk with limitations; some activities limited	0.4772 0.5228
Blindness and amputation	Cannot drive; walk with limitations; limited in amount or kind of work, school, or housework	0.4620
Death		0.0000

---

[16] Source: Kaplan et al., 1976.

One possible concern with these scores is that they may overstate the impact of complications, particularly retinopathy, on diabetics' utility. For example, it seems implausible that a diabetic would be indifferent between living a full year with retinopathy and living about three-quarters of a year without retinopathy. In this sense, the Kaplan-Bush scale appears to be biased towards low utility scores, particularly for relatively minor health problems. Nevertheless, it provides a pragmatic and consistent basis for calculating expected QALY with a range of possible health states.

Calculating expected QALYs for 1 year

Our outcome measure for diabetes is the expected number of QALYs over a diabetic's lifetime from onset of disease. This expected value depends on the probabilities that the diabetic will be in each possible health state (as defined above) at each point during his lifetime. To calculate these probabilities, we modeled the diabetic's health status as a discrete-state, discrete-transition Markov process, with transitions among the seven health states at yearly intervals. The "healthy" state is the initial state for the process, and the "death" state is an absorbing state.

The transition probabilities between states were calculated as follows:

¶ For transitions to all states except the Death state, we calculated transition probabilities assuming *independence* of complications. For example, we assumed that the probability of transitioning from the "retinopathy only" state to "retinopathy and amputation" is simply the annual complication rate for amputation; in other words, we assumed that having retinopathy does not increase the conditional probability of getting an amputation. From a clinical perspective, this assumption is dubious; since many complications reflect an overall decline in the diabetic's condition, having one complication likely increases the probability of having another. However, our independence assumption is likely "conservative" in that it reduces the difference in expected QALY between the U.S. and the U.K. A more realistic model with positive correlation among complications would magnify the effect of the U.S.'s higher complication rates.

¶ For transitions to the Death state, we calculated transition probabilities by taking an assumed baseline mortality rate for diabetics (which is age-dependent) and adding an adjustment for the specific complications present. Baseline mortality rates for diabetics were estimated using average mortality rates by age in the U.S.,

adjusted upwards by 0.15 percent per year for Type I diabetics and by 2.5 percent per year for Type II diabetics.[17] We assumed that these baseline diabetic mortality rates were the same for the U.S. and the U.K. since we did not have diabetic-specific mortality data for the two countries. (Refer to Appendix 3B for details on mortality in the two countries.) Mortality adjustments for the presence of complications were as follows:

- If retinopathy or blindness was present (without amputation), we added 2.8 percent to the baseline mortality rate for Type I diabetics and 6.8 percent for Type II diabetics.[18]

- If amputation was present, we added 14.3 percent to the baseline mortality rate (regardless of the presence or absence of retinopathy or blindness).[19]

Given the state probabilities for each year of the diabetic's lifetime, the expected QALY measure was calculated by weighting each state's utility by the probability of being in that state, and summing over time:

$$E(QALY) = \sum_{n=Onset}^{Max} \sum_{i \in States} Prob \text{ (in State i in Year n)} * \text{Kaplan-Bush Utility of State i}$$

where:

> *States* is the set of seven health states described above,
> *Onset* is the age at which an average diabetic gets diabetes (we have used 15 for Type I diabetics, and 55 for Type II diabetics), [20]
> *Max* is the maximum life of a diabetic, which we have modeled as 100 years (although the number surviving past age 80 is negligible.)

(Note that we are making an assumption that each complication occurs, at most, once per year. While it is theoretically possible to incur some complications several times in a single year – for instance, multiple occurrences of DKA – the low annual probabilities of these complications imply that multiple occurrences are very rare; thus, our simplification is probably reasonable.)

---

[17] Source: Javitt and Aiello, 1996.

[18] Source: Javitt and Aiello, 1996. We have estimated that the impact of blindness on mortality is the same as the impact of proliferative retinopathy, since blindness is generally caused by severe retinopathy. The mortality model in Javitt uses both a multiplicative factor on age-adjusted mortality and an additive factor; we have simplified this model to use only an additive factor, normalizing so the net impact on mortality rate at an "average" age is the same as in Javitt's model.

[19] Source: Esterostom, 1989. Quoted average of three 5-year mortality studies of diabetics with amputation.

[20] Source: Klein et al., 1989.

Since we are assuming for simplicity that each complication occurs in the middle of a year, we have $t^P_C = 1/2$ for each complication. The above formula simplifies to:

$$E(QALY) = 1 - \sum_{complications} p_C(D^T_C t^T_C + D^P_C/2)$$

One important practical advantage of the additive disutility model is that this result holds regardless of how complications are interrelated (i.e., it does not matter whether different complications are independent events, mutually exclusive events, etc.). Our data sources only provide the individual complication probabilities; we do not know the joint probabilities of multiple complications, or even whether different complications are independent events. (On clinical grounds, we would doubt that different complications are, in fact, independent, since all complications reflect an underlying problem in metabolic control.) It is, therefore, necessary to use an expected utility model that relies on individual complication probabilities.

In contrast, suppose we used a simpler utility model that reflected only the absence of complications; i.e., it assigned:

    Utility with no complications = 1.0
    Utility with any complication = 0.0

The expected utility using this model would equal the probability of having *no* complications. While this is conceptually straightforward, it is not possible to calculate this probability without additional information about the joint probabilities of having multiple complications – information we do not have.

CALCULATING EXPECTED QALYs OVER LIFETIME

As noted before, we assumed that complication probabilities are the same in each year of a diabetic's life. These yearly complications will, therefore, equal the annual incidence rate for the complications across the entire diabetic population. This assumption is probably not realistic; as diabetics age, the chance of complications probably increases. However, since we have no information on the time profile of complications, we are unable to justify any specific models with greater complexity. Moreover, any models with different complication probabilities over time would have to be consistent with an age distribution for the diabetic population, since the observed annual incidence rates across the population reflect the weighted average of complication probabilities across age groups. Such a model would be further complicated by improvements in treatment over time: as diabetes care has improved, the expected future complication rates for younger diabetics may have declined relative to the actual complication rates experienced by older diabetics. Because of all of these issues, we opted for the assumption of equal complication probabilities over time.

(Again, we stress that because the U.K. has uniformly lower complication rates than the U.S., this simplifying assumption is unlikely to affect the relative outcome rankings.)

The impact of previous years' complications on the utility for a given year depends on the duration of its effect. Some complications, such as blindness, amputation, and retinopathy, permanently impair health status. Others, such as foot ulceration and DKA, can be treated and "cured," with no necessary long-term effect. Our model recognizes these differences by using "cumulative" probabilities to weight the post-treatment disutilities of "permanent" complications. For a permanent complication, the impact on a diabetic's QALYs in a specific year can, therefore, be calculated as follows:

Impact of permanent complication C on expected QALYs in year N=

$$p_C(D^T_C t^T_C + D^P_C/2) + \text{Prob (had complication in Years 1 through N-1)}*D^P_C$$

The first term reflects the QALY impact of getting the complication during the Nth year. The second term reflects the cumulative probability that the diabetic already has the complication going into the start of the Nth year. Note that the time weighting in this case is the full year, since the diabetic has the complication throughout the Nth year.

Most permanent complications, like blindness, only occur once. We, therefore, assume mutual exclusivity for getting the same complication in different years. (Note that this is not true for amputation, since as a portion of a toe or foot can be amputated, it is possible to experience this complication more than once. Although we are ignoring this possibility in our model, the single amputation event we are including represents the average severity of amputation across the population.) As a result of this assumption of mutual exclusivity, the cumulative probability of having incurred the complication in previous years is simply the sum of the annual probabilities over these years:

Prob (got complication C in Years 1 through N-1)

$$= \text{Prob (got complication C in Year 1)} +$$
Prob (got complication C in Year 2) +
Prob (got complication C in Year 3) + ...
... Prob (got complication C in Year N-1) +

$$= (N-1)p_C \text{ (where } p_C \text{ is the annual incidence rate for C)}$$

Therefore, the expected QALYs for the Nth year of a diabetic's life can be expressed as follows:

E(QALYs in Year N) =

$$= 1 - \sum_{\substack{all \\ complications}} p_C(D^T_C t^T_C + D^P_C/2) - \sum_{\substack{permanent \\ complications}} (N-1)p_C D^P_C$$

The expected QALYs over a diabetic's lifetime is, therefore:

$$E(QALYs) = \sum_{N=1}^{L} E(QALYs \text{ in Year N}), \text{ where L is the life expectancy}$$

$$= L \left(1 - \sum_{\substack{all \\ complications}} p_C(D^T_C t^T_C + D^P_C/2) - (L-1)/2* \sum_{\substack{permanent \\ complications}} p_C D^P_C\right)$$

Since the post-treatment disutility for a nonpermanent complication is zero, this formula simplifies as follows:

$$E(QALYs) = L \left[1 - \sum_{\substack{all \\ complications}} p_C(D^T_C t^T_C + D^P_C/2) - (L-1)/2* \sum_{\substack{permanent \\ complications}} p_C D^P_C \right]$$

(Recall that we are assuming equal life expectancies for diabetics between the U.S. and the U.K. Thus, the parameter L above is the same for both countries.)

This form permits a reasonably simple interpretation of the effect of permanent complications on QALYs: a permanent complication C reduces expected QALYs by $(Lp_C)D^P_C(L/2)$. Lp_C is the lifetime probability of incurring the complication. D^P_C is the disutility that results from having the complication. (L/2) can be interpreted as the average amount of time a diabetic who gets a complication will have it; this model implies that, under our assumptions, this time is half the diabetic's life expectancy.

We should also note that, under this model, the impact of permanent complications on QALYs far exceeds the impact of nonpermanent complications because of the (L/2) multiplier in the above formula. This result makes intuitive sense, because permanent complications affect the diabetic for a far longer period of time.

RESULTS OF QALY MODEL

The following table shows the calculations for expected QALYs for Type I and Type II diabetics in the U.S. and the U.K.

Type I		Type II	
U.S.	U.K.	U.S.	U.K.

E (QALY)	51.5	54.0	13.7	14.9
Difference from U.S.		2.5		1.2

Because of its consistently lower complication rates, the U.K. had a higher expected QALY for both Type I and Type II diabetics. The difference between the two countries was larger for Type I diabetics primarily because of their longer life expectancy (after onset of diabetes); the effect of higher complication rates in the U.S. had more time to "compound" in Type I diabetics.

Appendix 3B: Comparison of diabetes mortality between U.S. and U.K.

There is no definitive study or information source that permits a direct comparison of mortality rates between the U.K. and the U.S. However, currently available evidence suggests that the U.K. had lower mortality rates from diabetes during the timeframe of our analysis. This conclusion is consistent with the U.K.'s superior performance in preventing complications and indicates that the U.K.'s diabetic treatment process achieved better overall outcomes regardless of how the various measures of process output (early stage complications, late stage complications, and mortality) were weighted.

Our tentative conclusion of lower mortality in the U.K. is based on several indirect indicators:

Comparison	Type	Results	Caveats
Raw death rates per capita from diabetes	All	U.K. 30-70% of U.S. in most age/sex groups	1. Death certificate data unreliable 2. Does not adjust for prevalence of disease
Raw death rates from diabetes per diabetic	Type I	U.K. 47-81% of U.S. in age groups 15-34	1. Death certificate data unreliable
Death rates per capita with diabetes as cause or contributing factor	All	U.K. 76% of U.S.	1. Death certificate data better with multiple causes, but still not definitive 2. Does not adjust for prevalence of disease
Cohort studies on death rates of diabetics over defined study period	Type I	U.K. 28% of U.S.	1. Cohorts from different studies, not directly comparable

The following sections describe these specific comparisons and discuss the issues with each.

"RAW" DEATH RATES FROM DIABETES

The World Health Organization (WHO) reports rates of death from diabetes for entire populations by sex and age category. These data are based on analyses of death certificates to determine the reported cause of death. These data show that the U.K. had consistently lower death rates from diabetes in all age/sex categories (except for an insignificant anomaly in males 75+).

DEATH RATES FROM DIABETES, 1990

Rate per 100,000 population, by age and sex

Country	Sex	15-24	25-34	35-44	45-54	55-64	65-74	75+
U.S.	M	0.4	1.8	4.8	11.5	28.6	61.8	146.4
	F	0.3	1.2	3.1	9.2	27.5	62.1	148.9
U.K.	M	0.2	0.7	1.4	4.7	16.7	43.4	147.1
	F	0.2	0.5	1.0	3.2	11.8	38.0	124.5
Ratio	M	50%	39%	29%	41%	58%	70%	100%
	F	67%	42%	32%	35%	43%	61%	84%

(The World Health Organization (WHO) reported per capita death rates for the age group below 15 is 0.1 per 100,000 population in both nations. We have not included this age group because the death rates reported for this group are too low to allow us to reach any conclusions.)

These data must be interpreted very cautiously. Researchers have consistently noted that death certificates are unreliable indicators of causes of death, particularly for diabetes. Diabetics frequently die of complications, such as cardiac failure or renal failure, and these complications may be listed as the cause of death rather than diabetes. These death rates may, therefore, understate the true mortality of diabetics. Physicians' practices in recording cause of death also vary widely among countries and even within countries; the apparent difference between the U.K. and the U.S. could, therefore, be due to differences in reporting methods.

RAW DEATH RATES FROM DIABETES PER DIABETIC

Differences in death rates from diabetes could also be due to difference in the prevalence of the disease between the countries rather than differences in treatment effectiveness. Our analysis indicates that diagnosed diabetes is slightly more prevalent in the U.S. than in the U.K.[21]

PREVALENCE OF DIAGNOSED DIABETES, 1992

Percent of population

---

[21] Sources: Office of Population Censuses and Surveys, 1993; NIH Publication No. 95-1468, 1995.

Country	Type I	Type II
U.S.	0.29%	2.64%
U.K.	0.24%	2.20%

Given these prevalence rates of diabetes, there is not a significant difference between the two nations in mortality rates from diabetes per diabetic.

DEATH RATES FROM DIABETES PER DIABETIC

Percent

Country	Mortality rate
U.S.	0.56%
U.K.	0.60%

We can draw a somewhat more precise conclusion by focusing only on the Type I diabetics. In contrast to Type II diabetes, Type I diabetes is usually an unambiguous diagnosis. In addition, since untreated Type I diabetes causes severe symptoms, we can be certain that there are few undiagnosed cases. Because Type I diabetics represent the majority of the diabetes cases in the under-35 age group, we can make the simplifying assumption that the reported deaths from diabetes under age 35 are from Type I diabetes only. Using this assumption and the reported prevalence for Type I diabetes, the U.K. clearly has lower mortality rates for Type I diabetics under the age of 35.

MORTALITY RATES FROM TYPE I DIABETES, 1990

Rate per 100,000 diagnosed Type I diabetics

Country	Sex	15-24	25-34
U.S.	M	138%	621%
	F	103%	414%
U.K.	M	83%	292%
	F	83%	208%
U.K./U.S.	M	60%	47%
	F	81%	50%

DEATH RATES MENTIONING DIABETES
AS A CONTRIBUTING FACTOR

The problem of unreliable or inconsistent death certificate reporting can be partly corrected if death certificates mention both "cause of death" and other "contributing factors." Some studies have attempted to capture this information from death

certificates and analyze it to determine the frequency of diabetes appearing as either a cause or a contributing factor of death. In particular, one U.K. study from 1985 to 1986 found approximately 47 deaths per 100,000 population with diabetes as a primary or contributing factor. A U.S. study from the same time period (1986) found 62 deaths per 100,000 with diabetes as a primary or contributing factor. These studies support our conclusion of lower mortality rates in the U.K., with the U.K.'s rate at 76 percent of the U.S. rate.

Again, however, these results are not definitive. The calculated death rates were based on standard age distributions for the entire countries; while similar, these are not identical and could affect the results. As with the raw death rates (diabetes as primary cause of death), these rates could also be affected by the higher prevalence of diabetes in the U.S.

COHORT STUDIES

The potential errors mentioned above can be avoided by comparing the actual mortality experience of cohorts of diabetics in the two countries, controlling for factors such as age, sex, duration of disease, and type of diabetes. We have identified two separate cohort studies from 1985 (one in the U.S. and one in the U.K.) that can be compared to some extent to observe differences in mortality rates. Because these were separate studies, the cohorts are not directly comparable; however, they are reasonably similar.

Country	Region	Age range	Type of diabetes	Time cohort followed	Death rate per 100,000 LYs
U.S.	Allegheny Co., PA	25-37	Type I only	5 years	812
U.K.	Aberdeen	15-44	Mostly Type I (90%)	5 years	225

The U.K. cohort had a mortality rate of only 28 percent of the U.S. cohort's rate, supporting our conclusion of superior outcomes in the U.K. However, it is possible that the U.K. mortality rates resulted from the wider age range of the U.K. cohort. The small number of Type II diabetics in the U.K. cohort could not account for much of this difference in mortality rate – even if the Type II diabetics had a death rate of 0, the death rate for the Type I diabetics in the U.K. cohort would have been 250 per 100,000 LYs (225/90 percent).

These data must also be approached skeptically. These cohorts are unlikely to be truly representative of the entire U.S. and U.K. populations, since no diabetes treatment is known to have such a large effect on mortality.

Appendix 3C: Sources for diabetes complication rates

This appendix documents the sources used in the diabetes case for complication rates in the U.S. and the U.K. Complication rate estimates were developed for four complications: DKA/hyperosmolar coma, retinopathy, blindness, and lower extremity amputation. We report each complication rate as an annual incidence per Type I and Type II diabetic.

The following sections provide the sources and values for diabetes prevalence and each complication rate estimate, and discuss the issues with each.

DIABETIC PREVALENCE

Number of diagnosed diabetics, 1992

Age	U.K.		U.S. whites	
	Diabetics	Prevalence	Diabetics	Prevalence
Under 15	11,793	0.1%	56,398	0.1%
15-44	331,514	1.3%	976,602	1.0%
45-64	524,686	4.1%	2,238,000	5.4%
65+	549,000	6.0%	2,816,000	9.7%
Total	1,416,993	2.4%	6,087,000	2.9%
Type I	141,910	0.2%	636,849	0.3%
Type II	1,275,083	2.2%	5,450,151	2.6%

Discussion of sources

1. **U.K. data**. Number of diagnosed diabetics is from the *1993 Health Survey for England* carried out by the Office of Population Censuses and Surveys, reported in *Diabetes in the United Kingdom – 1996*, published by the British Diabetic Association, pages 3 through 5. Number of Type I and Type II diabetics calculated assuming

10.01 percent of U.K. diabetics are Type I, as found in 1983 Poole study. U.K. population figures from *Annual Abstract of Statistics*.

2. **U.S. data.** Number of white diagnosed diabetics in 1992 from *Diabetes in America, 2nd edition*, published by the National Institutes of Health, page 63. Number of Type I and Type II white diabetics calculated assuming 10.50 percent of white diabetics are Type I. U.S. population figures from U.S. Bureau of the Census.

DKA/HYPEROSMOLAR COMA

Annual incidence rate per 100,000 diabetics

Ref.	Country	Source	Data	Type I	Type II
1	U.S. (whites only)	National Hospital Discharge Survey, 1992.	Hospital admissions for DKA/hyperosmolar coma per diagnosed diabetic	8,550	90
2	U.K.	Hospital Episode Statistics, 1992	Hospital episodes for DKA/hyperosmolar coma per diagnosed diabetic	3,930	70

DKA and hyperosmolar coma are technically two different disease processes. However, they are closely related; both are characterized by complex metabolic derangements with a substantially elevated glucose level. From a treatment standpoint, the two processes are similar. DKA occurs primarily in Type I diabetics, and hyperosmolar coma occurs primarily in Type II diabetics.

Since diabetes and its associated complications occur more frequently with age, differences in complication rates could reflect differences in the age distribution of the populations examined. To test this possibility, we compared DKA and hyperosmolar incidence rates in each country within age groups.

DKA
Annual incidence rate per 100,000 Type I diabetics

Country	0-14	15-44	45-64	65+
U.S. (whites only)	10,390	7,540	8,770	17,220
U.K.	8,270	3,110	3,620	5,850

HYPEROSMOLAR COMA
Annual incidence rate per 100,000 Type II diabetics

Country	0-14	15-44	45-64	65+
U.S. (whites only)	n/a	60	60	110
U.K.	n/a	60	40	90

As the U.K. has lower or equivalent DKA and hyperosmolar coma incidence rates for each age group, we can conclude that lower incidence rates for the population overall are not reflective of age differences.

Discussion of sources

1. U.S. data are from the *National Hospital Discharge Survey (NHDS)*, January to December 1992. NHDS collects data on hospital discharges from a sample of short-stay, nonfederal hospitals in the U.S. The complication rate is the number of admissions for DKA (ICD-9 CM code 250.1) and hyperosmolar coma (ICD-9 CM code 250.2) as the primary diagnosis for whites. (The rationale for comparing the U.K. population to U.S. whites is explained in Chapter 3.) Complication rates per diabetic were calculated using the number of diagnosed white diabetics in the U.S. by age group.

2. U.K. data are from Brown 1996 (unpublished). Data from Brown are based on the *Hospital Episode Statistics (HES) database*, April 1, 1992 to March 31, 1993. HES is a U.K. reporting system that collects information on all inpatient and outpatient activity in England. This information is aggregated at the national level by the Ministry of Health and at the local level by district health authorities (DHAs) or regional health authorities (RHAs). The complication rate is the number of episodes for DKA (ICD-9 code 250.1) and hyperosmolar coma (ICD-9 code 250.2) as the primary diagnosis. Complication rates per diabetic were calculated using the number of diagnosed diabetics by age group. These data only reflect hospital episodes in England; in using these data to compare the U.S. to the U.K., we assumed that there are no significant differences in complication rates between England and the rest of the U.K. England contains over 82 percent of the population of the U.K.; even if complication rates were significantly different in other parts of the U.K., our conclusion of lower complication rates in the U.K. would remain unchanged.

   Rather than counting hospital admissions as in the U.S., HES counts Finished Consultant Episodes. These episodes are the duration of care for a patient under a particular consultant (senior hospital-based physician) and, therefore, might lead to overcounting compared to U.S. admissions data. For example, if a patient with a heart attack is admitted one night by a gastroenterologist on duty and then transferred to the care of a cardiologist the next day, this would count as two episodes in the U.K. but one admission in the U.S. This means that the U.K. data may overestimate the complication rates in the U.K. compared to the U.S., making our estimate of higher complication rates in the U.S. conservative.

RETINOPATHY

Annual incidence rate per 100,000 diabetics

Ref.	Country	Source	Data	Type I	Type II
1	U.S. (98% whites)	Klein et al., "The Wisconsin Epidemiological Study of Diabetic Retinopathy; X. Four-Year Incidence and Progression of Diabetic Retinopathy When Age at Diagnosis Is 30 Years or More," *Arch Ophthalmology*, Vol. 107 (Feb. 1989).	Annual incidence derived from prevalence of retinopathy in baseline examination of study population	n/a	4,910
2	U.K.	Neil et al., "Diabetes in the Elderly: The Oxford Community Diabetes Study"	Annual incidence derived from prevalence of retinopathy in 60+ population	n/a	2,780

Note that we only have comparable data for Type II diabetics.

Discussion of sources

1. Data are from the "Wisconsin Epidemiologic Study of Diabetic Retinopathy."[22] Retinopathy was studied in a population-based sample of 1,370 people in Southern Wisconsin, with diabetes diagnosed at 30 years of age or older (assumed to be Type II diabetes). The study population was 98-percent white.

   Annual incidence rates were calculated from the prevalence of "any retinopathy" found in the 1980 to 1982 baseline examination of the entire study population. The mean duration of diabetes for this population was 11.87 years. (The annual incidence rate found in a 4-year follow-up examination of 50 percent of this population was 7,618 per 100,000. Instead of using this number, we derived the annual incidence rate from the prevalence found at the baseline examination because it included the entire study population.)

2. Data are from the "Oxford Community Diabetes Study."[23] Retinopathy was studied in a population-based sample of

---

[22] Source: Klein et al., 1989.

[23] Source: Neil et al., 1989.

431 diabetic patients in Oxford in approximately 1986. Of these, 193 patients were 60 years of age or older, with 97 percent of this group having Type II diabetes.

Annual incidence rates were calculated from the prevalence of "retinopathy of any grade" for the population 60 years of age or older. The mean duration of diabetes for this population was 9 years.

Comparability issues

In using these two studies to compare retinopathy in the U.S. and the U.K., we are comparing a segment of the U.S. population with a segment of the population 60 years of age and over in the U.K. The U.K. population sample is thus substantially older than the U.S. population studied, as shown below.

Percentage of population studied by age group

Age	U.K.	U.S.
20-59	0%	27.4%
60-69	35.2%	33.1%
70-79	50.3%	27.8%
80+	14.5%	11.6%
Mean Age	*72.21*	*62.61*

Because the prevalence of retinopathy increases with age, we would expect an older population to have a higher prevalence of retinopathy; the use of an older population in the U.K. relative to the U.S. overestimates the prevalence of retinopathy in the U.K. relative to the U.S. Our use of these studies to compare retinopathy between the two nations, therefore, gives us a very conservative estimate of better performance in the U.K.

Incidence rates were calculated for both studies based on the prevalence of "any" retinopathy. Differences in what is clinically considered to be retinopathy may exist between nations or between individual practitioners; by using any retinopathy rather than a definition of a specific type of retinopathy, such as "proliferative" retinopathy (which would necessitate a more precise definition and thus perhaps give rise to even more comparability difficulty), we attempted to minimize the definitional differences in our complication rate comparison.

BLINDNESS

Annual incidence rate per 100,000 diabetics

Ref.	Country	Source	Data	Type I	Type II
1	U.S. (98% white)	Klein et al., "The Wisconsin Epidemiological Study of Diabetic Retinopathy," Articles IX and X. *Arch Ophthalmology*, Vol. 107 (Feb. 1989).	Annual incidence derived from prevalence of blindness in baseline examination of study population	240	330
2	U.K.	A. Houston, "Retinopathy in the Poole Area: An Epidemiological Inquiry," *Advances in Diabetes Epidemiology*, INSERM Symposium No. 22 (1982).	Annual incidence derived from prevalence of blindness in study population	120	160

Discussion of sources

1. Data are from the "Wisconsin Epidemiologic Study of Diabetic Retinopathy." Eyesight was studied in a population-based sample of diabetics in Southern Wisconsin, 996 Type I and 1,370 Type II. (Type I diabetics were assumed to be those first diagnosed before the age of 30; Type II diabetics were assumed to be those first diagnosed over the age of 30.) The study population was 98-percent white.

   Blindness was defined as visual acuity of 20/200 or less in the better eye. The annual incidence rate was calculated from the prevalence of blindness found in a 1980 to 1982 baseline examination of the entire study population. The mean duration of diabetes for the Type I population was 14.67 years; the mean duration of diabetes for the Type II population was 11.87 years.

2. Data are from "Retinopathy in the Poole Area: An Epidemiological Inquiry."[24] Eyesight was studied in 714 known diabetics in the Poole area. Blindness was defined as visual acuity of 20/200 or less in the better eye. The annual incidence rate was calculated from the prevalence of blindness found in 1979 to 1980 eyesight examinations of the study population.

---

[24] Source: Houston, 1982.

The U.K. population was not identified as Type I or Type II. We assumed that the ratio of blindness of Type I and Type II diabetics was the same in the U.K. as in the U.S. As the incidence of blindness in the total U.K. population was higher than the incidence of both the Type I and Type II populations studied in the U.S., our finding of better outcomes in the U.K. would remain accurate regardless of the methodology used to break the U.K. population into Type I and Type II subgroups.

Comparability issues

The mean duration of diabetes was not given for the population studied. In order to calculate the annual incidence of blindness, we assumed that the mean duration of diabetes was the same in the U.K. population studied as in the U.S. population in the Wisconsin study. Because U.K. population was older than the U.S. population (see below), this assumption caused us to overestimate the incidence of blindness in the U.K. Our comparison of the relative incidence of blindness between the two nations, therefore, underestimates the degree of better outcomes in the U.K.

Percentage of population studied by age group

Age	U.K.	U.S.
0-19	5.5%	11.5%
20-59	35.1%	45.2%
60-69	27.1%	20.3%
70-79	27.3%	16.2%
80+	5.0%	6.7%
Mean Age	*58.85*	*50.84*

These data measure the incidence of blindness in the diabetic populations of the U.S. and the U.K., but do not specify the reason for blindness. If one nation had a greater rate of other causes of blindness (e.g., cataract, mascular degeneration, glaucoma) than the other, this could bias the results; however, we have no reason to believe that this would be true.

LOWER EXTREMITY AMPUTATION (LEA)

Annual incidence rate per 100,000 diabetics

Ref.	Country	Source	Data	Type I	Type II
1	U.S. (whites only)	*National Hospital Discharge Survey*, 1992	Hospital admissions for amputation per diagnosed diabetic	220	210
2	U.K.	*Hospital Episode Statistics*, 1992	Hospital episodes for amputation per diagnosed diabetic	140	130

As diabetes and its associated complications occur more frequently with rising age, differences in complication rates could reflect differences in the age of the populations compared. To test this possibility, we compared LEA incidence rates in each country within age groups.

LEA

Annual incidence rate per 100,000 diabetics

Country	0-14	15-44	45-64	65+
U.S. (whites only)	0	110	310	420
U.K.	0	20	130	270

As the U.K. has lower LEA incidence rates for each age group, we can conclude that lower incidence rates for the population overall are not reflective of age differences.

The numbers presented above measure the number of amputations per diabetic in the population, not the number of diabetics who receive amputations. It is conceivable, however, that multiple amputations are performed on the same patient (in an attempt to remove only as much of the toe or foot as is absolutely necessary) to a greater extent in the U.S. than in the U.K.; this would be true if the U.S. had a greater emphasis on limb-sparing than the U.K. If this were the case, this presents one reason why the higher rate of amputation in the U.S. relative to the U.K. might be exaggerated.

Discussion of sources

1. Data are from the *National Hospital Discharge Survey (NHDS),* 1992. (See description in the DKA section of this appendix.) The complication rate is the number of admissions for lower extremity amputation, with diabetes as the primary diagnosis for whites. (The rationale for comparing the U.K. population to U.S. whites is explained in Chapter 3.) Complication rates per diabetic were calculated using the number of diagnosed white diabetics in the U.S. by age group.

2. Data are from Brown, 1996 (unpublished). Data from Brown are based on the *Hospital Episode Statistics (HES) database*, April 1, 1992 to March 31, 1993. (See description in the DKA section of this appendix.) The complication rate is the number of admissions for lower extremity amputation with diabetes listed as the primary diagnosis. Complication rates per diabetic were calculated using the number of diagnosed diabetics in the U.K. by age group.

As with the DKA complication rates, the HES' practice of counting hospital episodes rather than hospital admissions makes our estimate of higher complication rates in the U.S. conservative. (See explanation in the DKA section of this appendix.)

Both sources: neither the NHDS nor the HES identifies diabetics as Type I or Type II. In order to estimate incidence rates for Type I and Type II diabetics separately, we, therefore, had to turn to another data source. A 1985 study of 2,023 diabetics in Rochester, New York, reported incidence rates of lower extremity amputation for Type I and Type II diabetics.[25] The ratio of LEA for Type I diabetics to LEA for Type II diabetics found in this study was 0.729. We used this ratio to assess incidence of LEA in Type I and Type II diabetics in both countries.

As expected, the effect of the post-treatment phase from permanent complications dominates the results. Because all complication probabilities are relatively small, the difference between the U.S. and the U.K. in expected QALYs is also relatively small. This difference is greater for Type I diabetics because of their longer life expectancies – this increases the effect of the permanent complications since Type I diabetics experience the effects of these complications for more years.

[25] Source: Humphrey et al., 1994.

Appendix 3D: Sources

This list details data sources used in the diabetes case study. We cover most of the main topics here, but this list is not exhaustive of all of the articles and government statistics that were employed throughout our work. In addition, we performed interviews with clinical and health care experts at a number of points during our study. Through these interviews, we collected qualitative and quantitative data on treatment patterns and checked our key assumptions and conclusions.

Below, we give the main sources used by topic.

KAPLAN-BUSH QUALITY OF WELL-BEING SCALE

Kaplan RM, Bush JW, Berry CW. Health status: Types of validity and the Index of Well-Being. Health Services Research 1976; Winter: 478-507.

DIABETIC CLINICS IN THE U.K.

Spathis GS. Facilities in Diabetic Clinics in the United Kingdom: A Report to the Medical Advisory Committee of the British Diabetic Association, October 1984.

Spathis GS. Facilities in diabetic clinics in the United Kingdom: Shortcomings and recommendations. Diabetic Medicine 1986.

Williams DRR, Spathis GS. Facilities in diabetic clinics in the U.K.: How much have they changed? Diabetic Medicine 1992.

GENERAL INFORMATION

Carter et al. Noninsulin-dependent diabetes mellitus in minorities in the United States. Annals of Internal Medicine 1996; 125.

Diabetes in America, National Institutes of Health, NIH Publication No. 95-1468, 1995.

Diabetes in the United Kingdom – 1996, British Diabetic Association, 1995.

Diabetes in the United States: A Strategy for Prevention, Centers for Disease Control, U.S. Department of Health and Human Services, 1994.

Diabetes Surveillance, Centers for Disease Control , U.S. Department of Health and Human Services, 1991.

Diabetes Surveillance, Centers for Disease Control , U.S. Department of Health and Human Services, 1993.

Diabetes Vital Statistics, American Diabetes Association, 1991.

Esterostom J. Amputation in the diabetic population: Incidence, causes, cost, treatment, and prevention. Ther 1989; 16: 247-55.

Geffner D. Diabetes care in health maintenance organizations. Diabetes Care 1992; 15: Supplement 1.

Gerard K, Donaldson C, Maynard AK. The cost of diabetes. Diabetic Medicine 1989; 6: 164-170.

Health Survey for England, Social Survey Division, Office of Population Censuses and Surveys, 1993.

Houston A. Retinopathy in the Poole area: An epidemiological inquiry. Advances in Diabetes Epidemiology, Elsevier Biomedical Press B.V., 1982.

Humphrey et al. The contribution of noninsulin-dependent diabetes to lower-extremity amputation in the community. Archives Internal Medicine 1994; 154: 885-892.

Influence of intensive diabetes treatment on quality of life outcomes in the diabetes control and complications trial. Diabetes Care 1996; 19.

Javitt, Aiello. Cost-effectiveness of detecting and treating diabetic retinopathy. Annals of Internal Medicine 1996; 124 .

Klein et al. Wisconsin epidemiologic study of diabetic retinopathy: Four-year incidence and progression of diabetic retinopathy when age at diagnosis is less than 30 years. Archives Ophthalmology 1989; 107: 608-613.

Klein et al. Wisconsin epidemiologic study of diabetic retinopathy: Four-year incidence and progression of diabetic retinopathy when age at diagnosis is 30 years or more. Archives Ophthalmology 1989; 107: 608-613.

Korn AM. Case management and quality of care for diabetic patients. Diabetic Care 1992; 15: Supplement 1.

Neil et al. Diabetes in the elderly: The Oxford community diabetes study. Diabetic Medicine 1989; 6: 608-613.

Siebert C, Lipsett L, Greenblatt J, Silverman R. Survey of physician practice behaviors related to diabetes mellitus in the U.S. Diabetic Care 1993; 42.

Somers AR. How to control diabetes and its costs. Business and Health 1990.

Standards of Clinical Care for People with Diabetes, Clinical Standards Advisory Group, London, July 1994.

Exhibit 1

PHASES OF DIABETES
MANAGEMENT AND TREATMENT PROCESS

	Management	**Complications**
Purpose of phase	Control patient's blood glucose level through insulin, diet, exercise, and monitoring Screen for onset of complications	Treat complications to restore health (if possible) or prevent further decline/death
Comments	Typically a combination of regular physician or clinic visits and extensive self-care Ongoing process throughout diabetic's life	Usually requires inpatient treatment Some complications can be successfully treated (e.g., diabetic ketoacidosis); others are permanent (e.g., blindness)
Average percent of inputs consumed in each phase*	**Type I** 37 **Type II** 63	**Type I** 41 **Type II** 59

* Average inputs consumed per phase in the U.S. and U.K.

Exhibit 2

OVERVIEW OF DIABETES
MANAGEMENT AND TREATMENT PROCESS

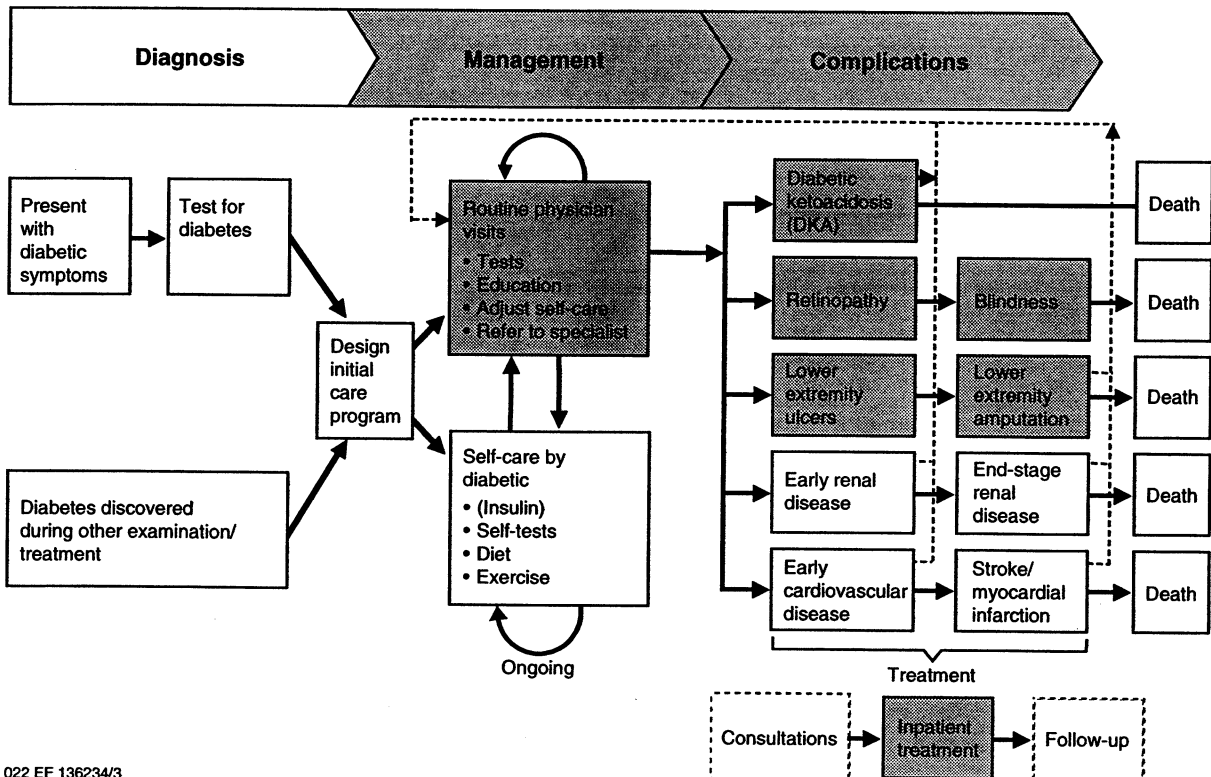▓▓ Included in analysis

022 EF 136234/3

Exhibit 3

INCIDENCE OF DIABETIC COMPLICATIONS

Annual incidence rate per 100,000 diabetics

Complication	Type I diabetics			Type II diabetics		
	U.S.	U.K.	Ratio U.K./ U.S. Percent	U.S.	U.K.	Ratio U.K./U.S. Percent
DKA/hyperosmolar coma	8,550	3,930	46	90	70	78
Retinopathy*	n/a	n/a		4,910	2,780	57
Lower extremity amputation	220	140	64	210	130	62
Blindness	240	120	50	330	160	48
Average	*3,003*	*1,397*	*47*	*1,385*	*785*	*57*

* Comparable data were not available on retinopathy for Type I diabetics in the U.K. and the U.S., so we were unable to determine these incidence rates
022 EF 136269/3

Exhibit 4
INPUTS BY PHASE*
Percent

Type I diabetes
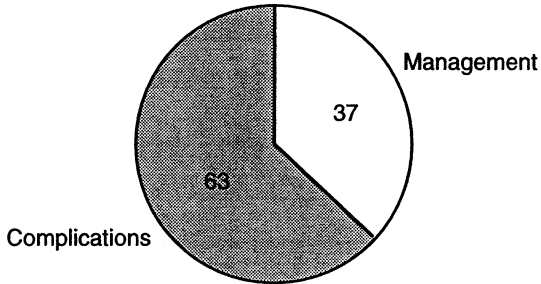
Management
37

63

Complications

Type II diabetes

Management
41

59

Complications

Exhibit 5
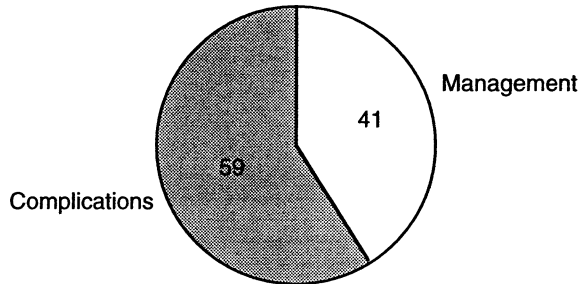**OVERVIEW OF DIABETES
MANAGEMENT AND TREATMENT PROCESS**
Management phase

① What **self-care protocol** should be followed?

② What **clinical setting** should be used for ongoing care?

③ What **frequency** of visits should be used?

④ What **specific interactions** should occur during patient visits?

Insulin therapy?
┌No
└Yes ┬ Type(s) of insulin?
     ├ Amount?
     ├ Frequency?
     └ Method?
Other pharmaceuticals?
Dietary restrictions?
Exercise program?
Other behavioral changes (e.g., smoking cessation)?
Home tests?
  ├ Blood glucose?
  ├ Urine?
  └ Frequency?

Options
• Home care only

• Physician's office

• Hospital outpatient clinic

• Dedicated diabetic clinic

• Combinations of above

From annual only to
6+ visits per year

What tests to perform?
  ├ General physical (weight, blood pressure)?
  ├ Blood test?
  ├ Urine test?
  ├ HbA1c?
  ├ Eye exam?
  └ Foot exam?
Review of self-care protocol?
Counseling/education?
  ├ Dietary advice?
  ├ Suggestions on self-care?
  └ Psychological counseling?

Exhibit 6
STAFFING IN U.K. DIABETIC CLINICS

Type of provider	Clinics with this resource Percent
Consultant (specialist)	85
Nurse	100
Dietitian	97
Chiropodist (podiatrist)	86
Ophthalmologist	18
Registrar (hospital physician)*	55

* A registrar in the U.K. is a hospital physician who is senior to a houseman (a newly qualified physician), but junior to a consultant (a senior physician)
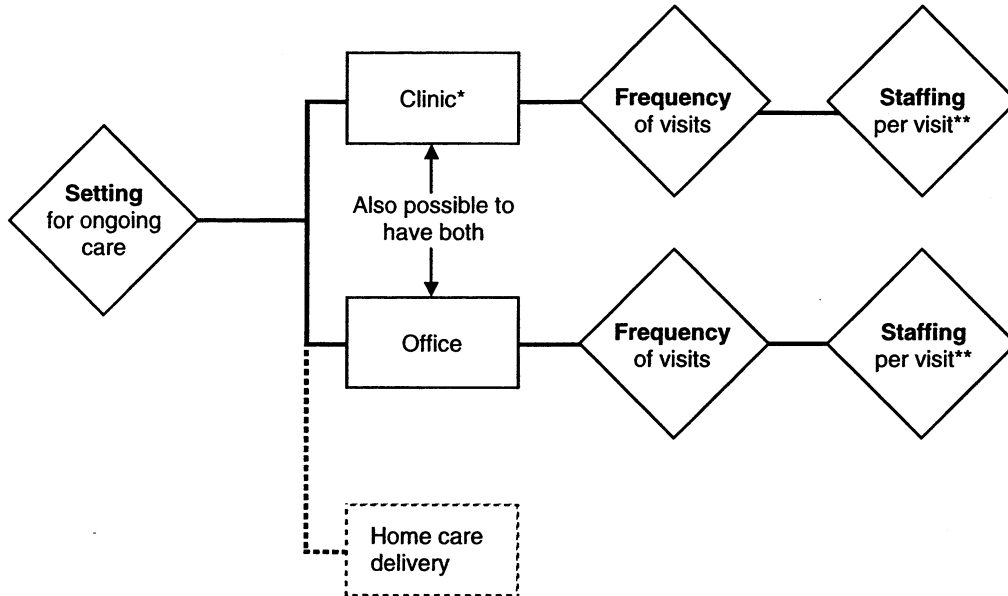
Exhibit 7
TESTS PERFORMED DURING PHYSICIAN VISITS
Percent of diabetics receiving test; normalized for type of diabetes*

U.S. (office setting)
Percent

U.K. (diabetic clinic)
Percent

Test	U.S.	U.K.
Urine glucose	65	74
Blood glucose	77	74
HbA1c	9	39
Blood pressure	80	23
Visual acuity	40	54
Fundoscopy	42	37
Foot exam	55	57
Urine protein	30	74
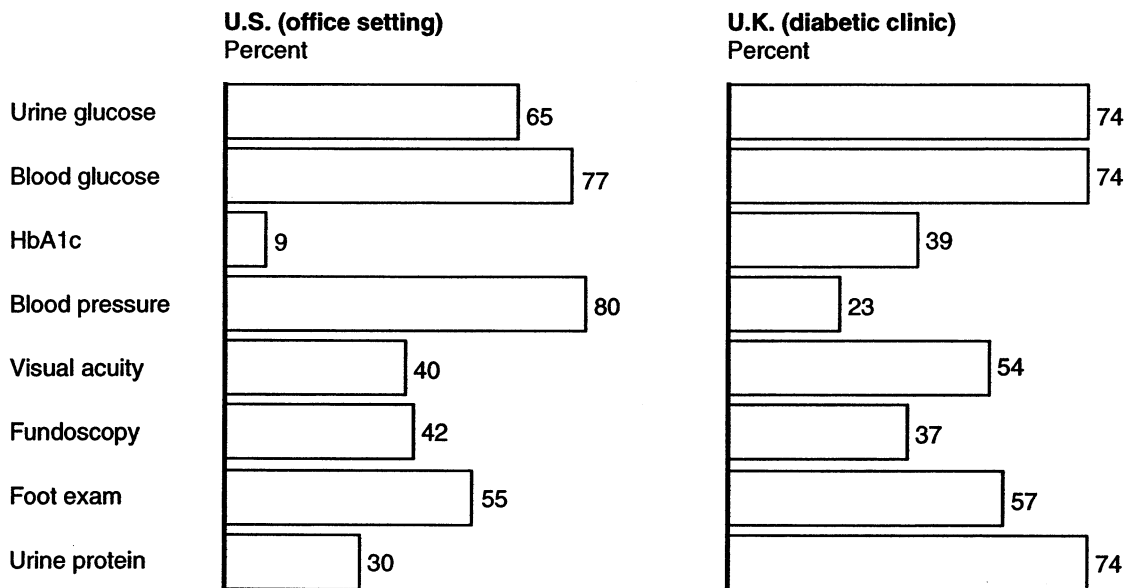
* U.K. normalized using U.S. frequencies for Type I, Type II insulin users and Type II noninsulin users
Source: Literature search; McKinsey analysis
022 RE 122742/3

Exhibit 8

SIMPLIFIED MODEL OF
MANAGEMENT PHASE DECISIONS
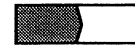
* Represents a hospital outpatient clinic in the U.S. and a dedicated diabetic clinic in the U.K.
** Proxy for specific tests and patient interactions during a visit

---

Exhibit 9

INPUTS BY SETTING

Standardized input units per visit

U.S.

Clinic	0.31
Office	0.17

U.K.

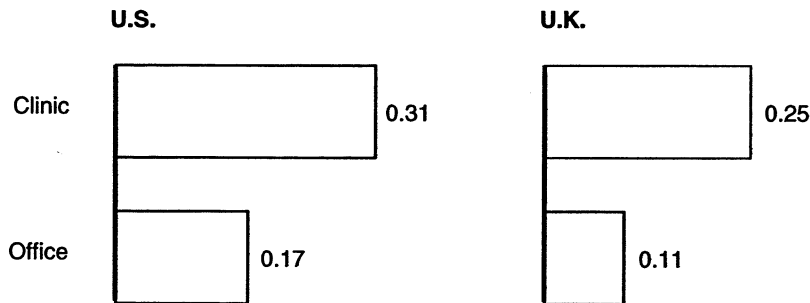|	
Clinic	0.25
Office	0.11

Exhibit 10

INSULIN INJECTION FREQUENCY – TYPE I DIABETICS

Percent of diabetics

U.S.

1 injection per day	38
2 injections per day	48
3 or more injections per day	13
Pump	1

U.K.

1 injection per day	10
2 injections per day	75
3 or more injections per day	14
Pump	1

022 CL 111722/3

Exhibit 11
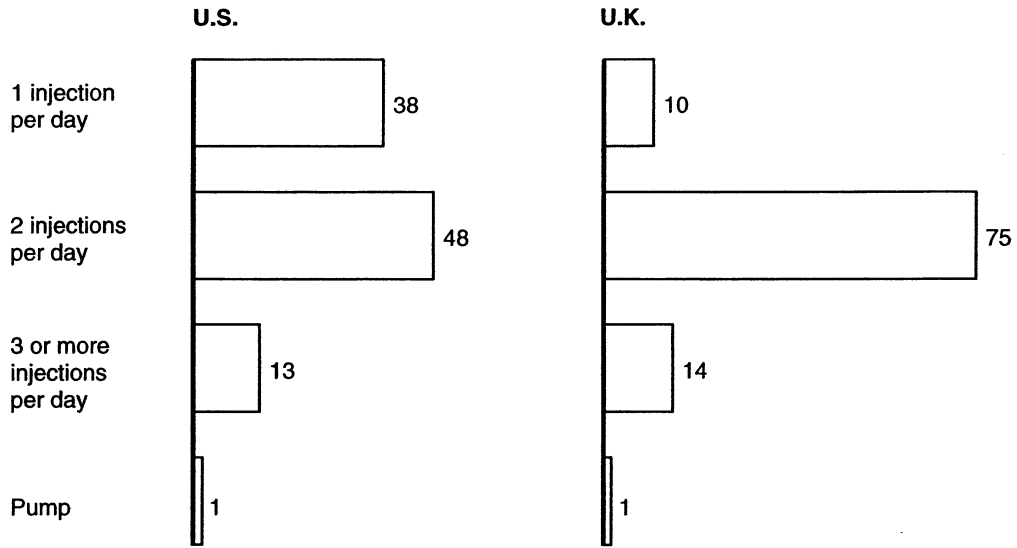
SIMPLIFIED MODEL OF COMPLICATION PHASE DECISIONS

Management phase

Complication occurs requiring inpatient treatment
• DKA
• Sight-threatening retinopathy
• Foot amputation

What LOS is required for inpatient treatment?

What staffing levels are used at the hospital?

022 ST 125211/3
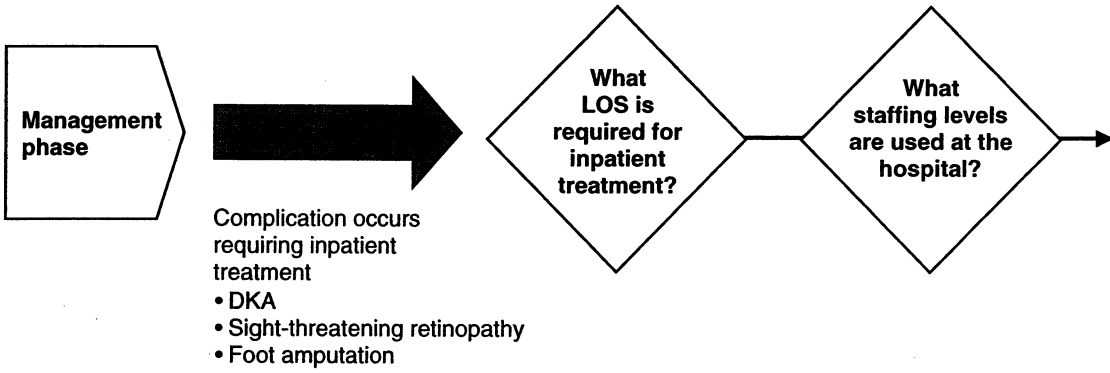
Exhibit 12

INPUTS AND OUTCOMES PER CASE OF DIABETES

Type I diabetes

Outcomes per case
QALYs
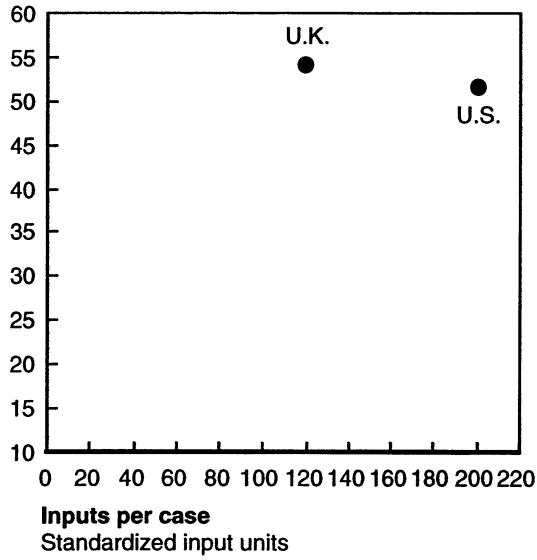
Inputs per case
Standardized input units

Type II diabetes

Outcomes per case
QALYs*

Inputs per case
Standardized input units

Weighted average of Type I and Type II diabetes
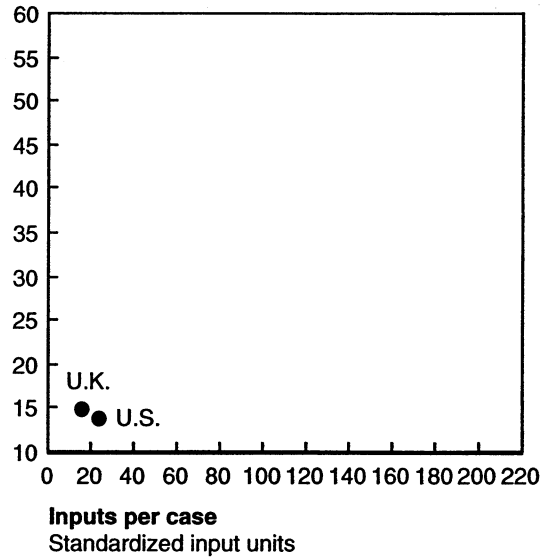
Outcomes per case
QALYs*

Inputs per case
Standardized input units
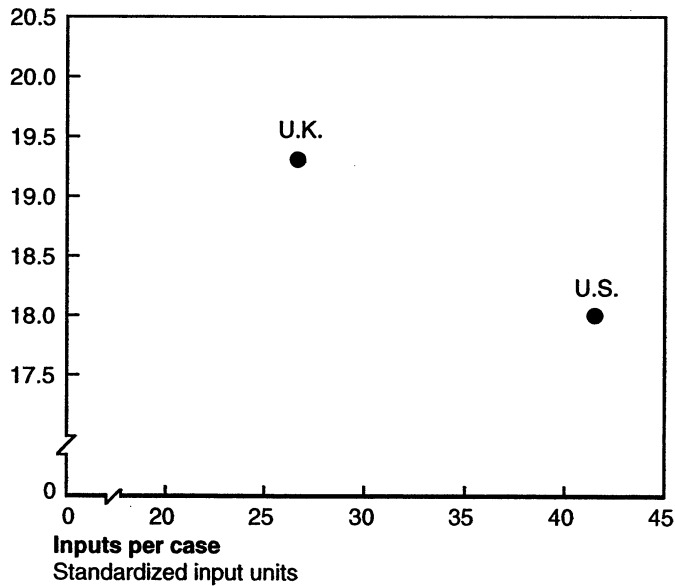
Exhibit 13

INPUTS AND OUTCOMES PER CASE OF DIABETES

Weighted average of Type I and Type II diabetes

Inputs
Index, U.S. = 100%

Outcomes
Improvement in outcomes due to treatment*
Index, U.S. = 100%

QALYs

	Inputs	Outcomes	QALYs
U.K.	66	127	6.41
U.S.	100	100	5.05
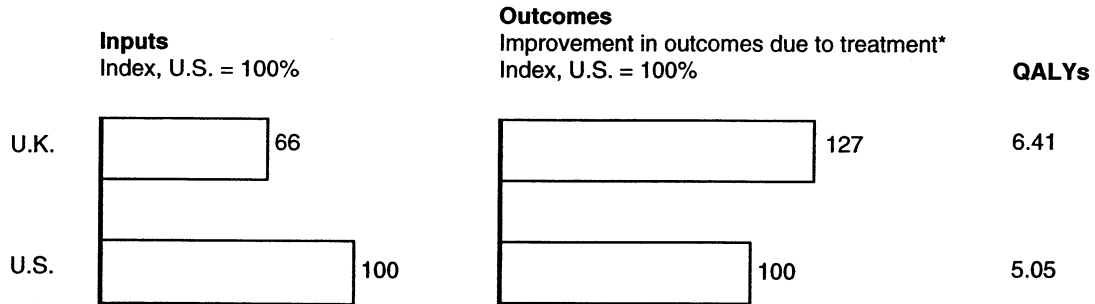
* Baseline outcome with no treatment assumed to be Type I diabetics die within 1 year and Type II diabetics have same QALYs as lowest
  outcomes with treatment (U.S.)

Exhibit 14

INPUT DIFFERENCES BY PHASE
Standardized input units

U.S.
U.K.

Management phase

	U.S.	U.K.
Type I	1.1	1.1
Type II insulin users	0.9	1.1
Type II noninsulin users	0.6	0.5

Complication phase

	U.S.	U.K.
Type I	2.7	1.2
Type II average	3.7	1.3

(Complication data not available for subtypes of Type II)

022 ST 140141/3

Exhibit 15

MANAGEMENT PHASE DECISIONS
AND IMPACT ON INPUTS

Type I diabetes

(+1%)

Setting for ongoing care

Clinic

U.S. 75%
U.K. 82%

Percentage
using setting

(-5%)

Office

U.S. 50%
U.K. 13%

Frequency of
visits at clinic

U.S. 3.6 (+9%)
U.K. 5.1

Visits per
diabetic per year

Staffing per
visit at clinic

U.S. 0.31 (-5%)
U.K. 0.25

Standardized
input units per visit

Frequency of
visits at office

U.S. 3.4 (+4%)
U.K. 5.2

Staffing per
visit at office

U.S. 0.17 (-3%)
U.K. 0.11

**Home
care* only**

U.S. 0%
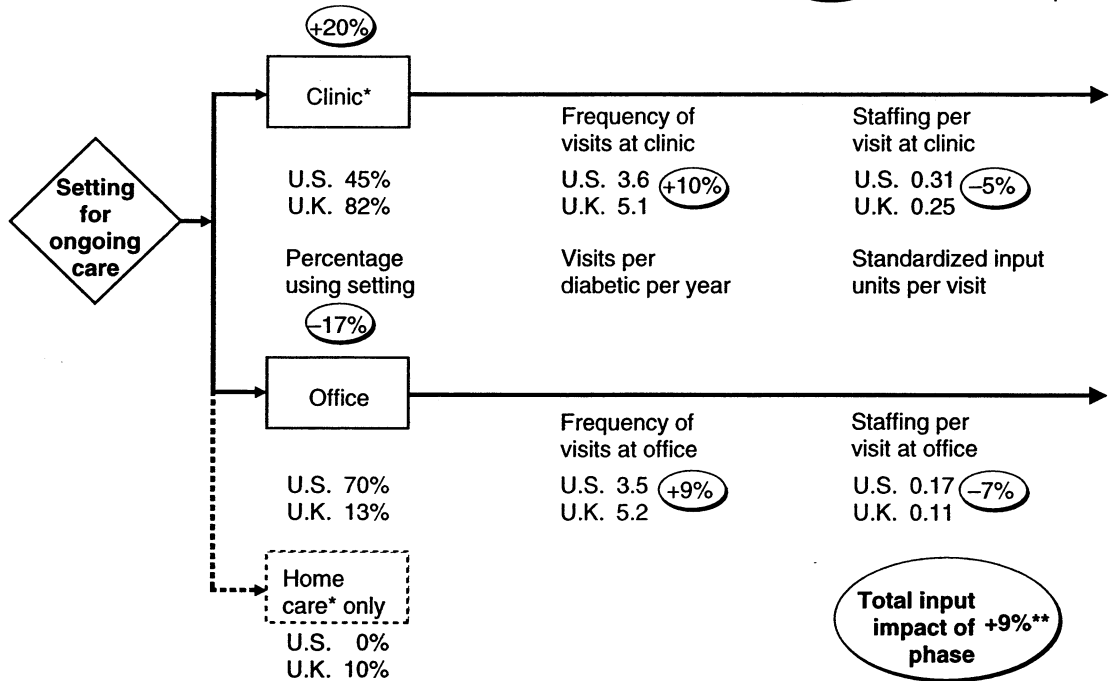U.K. 10%

(Total input
impact of -1%
phase)

* Input impact of home care frequency reflected in inputs of using clinic and office settings
022 ST 125212/3

Exhibit 16

MANAGEMENT PHASE DECISIONS AND IMPACT ON INPUTS

Type II insulin users

Percent impact on U.K. inputs relative to U.S. total inputs

(+20%)

Clinic*

Setting for ongoing care

Frequency of visits at clinic

Staffing per visit at clinic

U.S. 45%
U.K. 82%

U.S. 3.6 (+10%)
U.K. 5.1

U.S. 0.31 (-5%)
U.K. 0.25

Percentage using setting

(-17%)

Visits per diabetic per year

Standardized input units per visit

Office

Frequency of visits at office

Staffing per visit at office

U.S. 70%
U.K. 13%

U.S. 3.5 (+9%)
U.K. 5.2

U.S. 0.17 (-7%)
U.K. 0.11

Home care* only
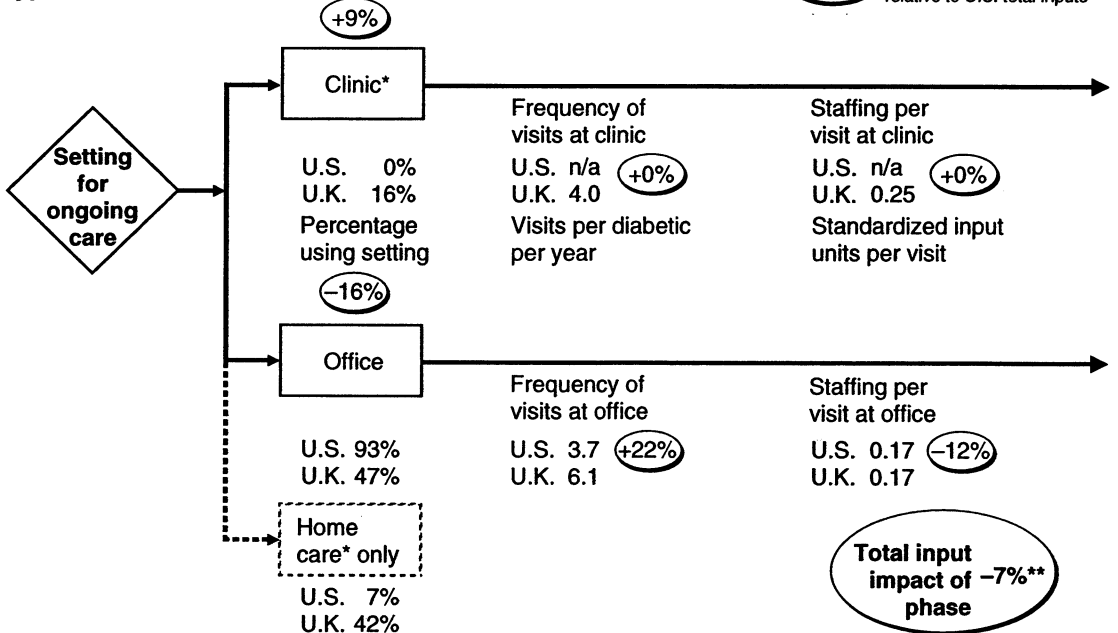
U.S. 0%
U.K. 10%

Total input impact of +9%** phase

* Input impact of home care frequency reflected in inputs of using clinic and office settings
** Includes joint effects of multiple factors of –2%

Exhibit 17

MANAGEMENT PHASE DECISIONS AND IMPACT ON INPUTS

Type II noninsulin users

Percent impact on U.K. inputs relative to U.S. total inputs

(+9%)

Clinic*

Setting for ongoing care

Frequency of visits at clinic

Staffing per visit at clinic

U.S. 0%
U.K. 16%

U.S. n/a (+0%)
U.K. 4.0

U.S. n/a (+0%)
U.K. 0.25

Percentage using setting

(-16%)

Visits per diabetic per year

Standardized input units per visit

Office

Frequency of visits at office

Staffing per visit at office

U.S. 93%
U.K. 47%

U.S. 3.7 (+22%)
U.K. 6.1

U.S. 0.17 (-12%)
U.K. 0.17

Home care* only

U.S. 7%
U.K. 42%

Total input impact of –7%** phase

* Input impact of home care frequency reflected in inputs of using clinic and office settings
** Includes joint effects of multiple factors of –14%

022 EF 136237/3

Exhibit 18

DISTRIBUTION OF MANAGEMENT PHASE
RESOURCES ACROSS DIABETIC POPULATION

——— U.K.
- - - - - U.S.

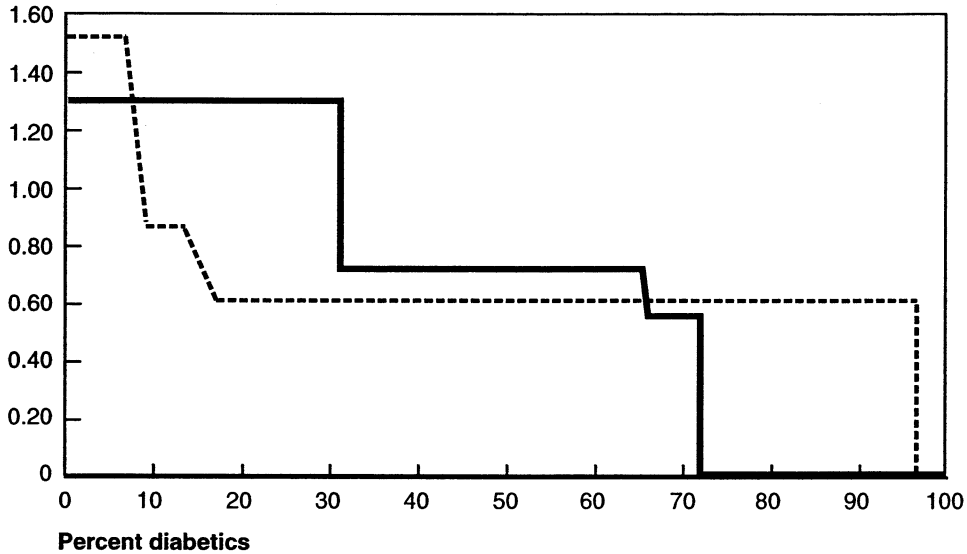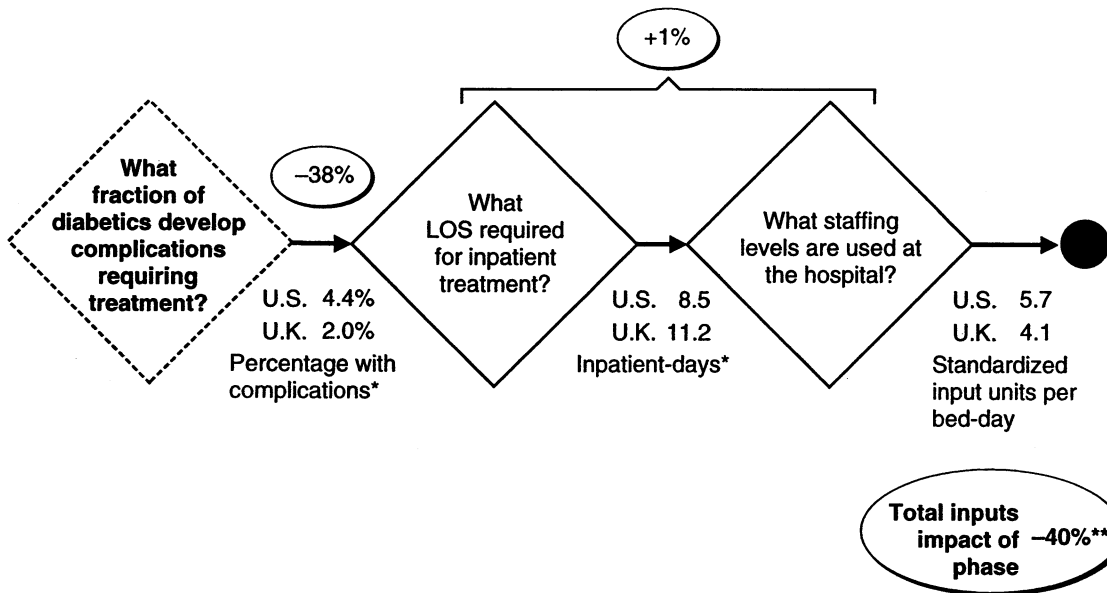**Standardized input units per diabetic**

Percent diabetics

Exhibit 19

COMPLICATIONS PHASE DECISIONS
AND IMPACT ON INPUTS

Type I diabetes

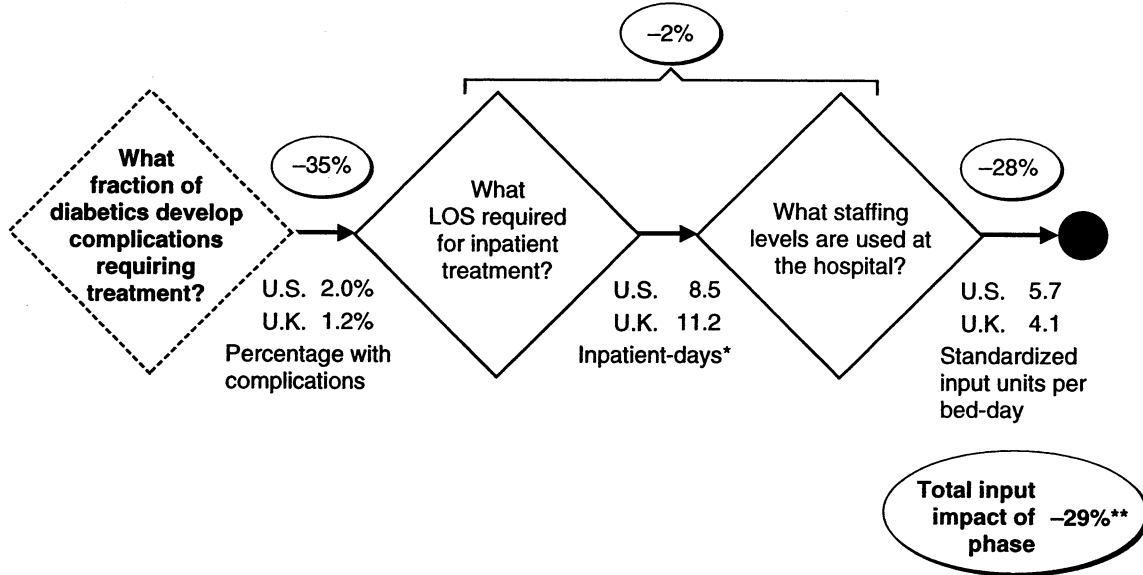% Percent impact on U.K. inputs relative to U.S. total inputs

+1%

What fraction of diabetics develop complications requiring treatment?

−38%

U.S. 4.4%
U.K. 2.0%
Percentage with complications*

What LOS required for inpatient treatment?

U.S. 8.5
U.K. 11.2
Inpatient-days*

What staffing levels are used at the hospital?

U.S. 5.7
U.K. 4.1
Standardized input units per bed-day

Total inputs impact of −40%** phase

* Averaged across 3 complications
** Includes joint effect from multiple factors of –3%
022 HC 133161/3

Exhibit 20

COMPLICATIONS PHASE
DECISIONS AND IMPACT ON INPUTS
Type II diabetes – insulin and noninsulin users

% Percent impact on U.K. inputs
relative to U.S. total inputs

−2%

What fraction of diabetics develop complications requiring treatment?

−35%

What LOS required for inpatient treatment?

−28%

What staffing levels are used at the hospital?

U.S. 2.0%
U.K. 1.2%
Percentage with complications

U.S.   8.5
U.K.   11.2
Inpatient-days*

U.S.   5.7
U.K.   4.1
Standardized input units per bed-day

Total input impact of −29% phase**

* Averaged across 3 complications
** Includes joint effect from multiple factors of −2%

Exhibit 21

SOURCES OF DIFFERENCE IN U.S. AND U.K. INPUTS PER CASE –
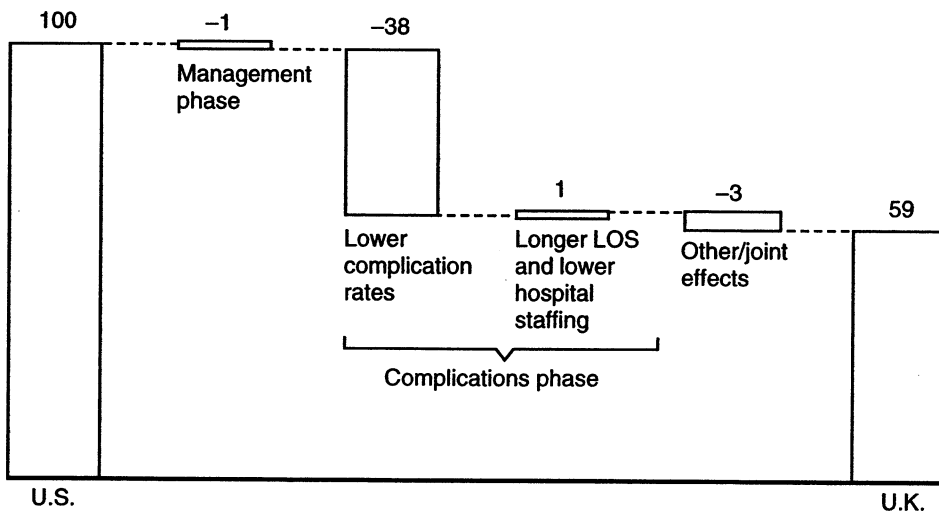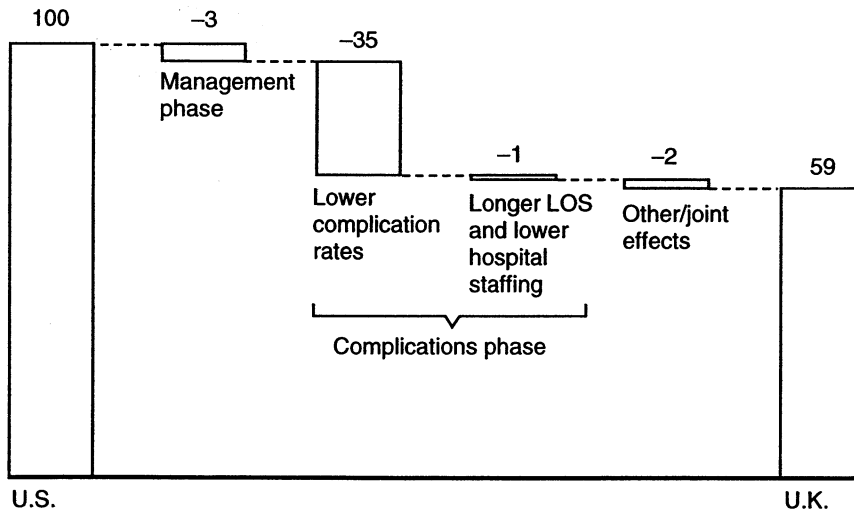## TYPE I DIABETES
Percent of total difference

100   −1   −38

Management phase

Lower complication rates

1

Longer LOS and lower hospital staffing

−3

Other/joint effects

59

Complications phase

U.S.   U.K.

Exhibit 22

SOURCES OF DIFFERENCE IN U.S. AND U.K. INPUTS PER CASE – TYPE II DIABETES

Percent of difference

100	−3	−35	−1	−2	59

Management phase

Lower complication rates

Longer LOS and lower hospital staffing

Other/joint effects

Complications phase

U.S.   U.K.

* Weighted average of Type II insulin users and Type II noninsulin users; weights reflect U.S. relative frequencies of these subtypes
022 ST 125839/3

Exhibit 23

IMPACT OF PROVIDER BEHAVIOR DIFFERENCES ON PRODUCTIVE EFFICIENCY

● Major differentiating factor
◐ Secondary differentiating factor
○ Undifferentiating

Provider behaviors	U.S. vs. U.K.	
Care triaging	●	U.K. – care matched to patients' needs U.S. – "one-size-fits-all" approach • In some cases, U.K. provided less care than U.S. • In some cases, U.K. provided more intense care than U.S.
Treatment duration	○	
Staffing levels	○	
Setting choice	○	
Team-based approach	●	U.K. – diabetic clinics provided care from multidisciplinary providers U.S. – no team approach
Technology adoption	○	

Exhibit 24

CAUSAL ANALYSIS: INCENTIVES AND CONSTRAINTS
U.S. vs. U.K.

● Major differentiating factor
◐ Secondary differentiating factor
○ Undifferentiating

	Extent of causal impact	Provider behaviors		
		Care triaging		Team-based approach
		Limits to care in the U.K.	Intense care in the U.K.	
Incentives				
Physician	●	U.K. physicians had less incentive to treat all patients than U.S. physicians	U.S. physicians had incentive to perform treatment and procedures, not education and other services	U.S. physicians had little incentive to coordinate with other providers
Hospital	○			
Constraints				
Physician supply	●	U.K. limited supply of GPs and specialists; so did not treat everyone	U.K. GPs busy, so referred tough cases to clinics	U.K. specialists busy with regional area to cover, easier to treat diabetics in specialized clinic
Hospital supply	○			
Capital	○			
Substitution	○			

022 ST 125841/3

Exhibit 25

CAUSAL ANALYSIS: SYSTEM STRUCTURE AND REGULATION
U.S. vs. U.K.

● Major differentiating factor
◑ Secondary differentiating factor
○ Undifferentiating

	Extent of causal impact	Provider incentives and constraints	
		Physician incentives	Physician constraints
Health care system structure			
Product integration and pricing	●	NHS coverage for total population for life created incentive to invest in intensive treatment for some diabetics; U.S. payors' high churn provided limited incentive to invest NHS funded diabetic care and encouraged diabetic clinic formation; U.S. nonintegrated approach did not encourage multidisciplinary care U.S. FFS vs. U.K. salary for specialist physicians encouraged treatment of all diabetics in fragmented vs. team-based approach	
Competitive intensity	●	U.S. provider competition for patients encouraged treatment of all diabetics and more fragmented vs. team-based care U.S. payor competition caused fear of adverse selection and high churn, precluding intensive diabetic treatment	
Regulation			U.K. physician supply strictly controlled

022 ST 125842/3

Exhibit A-1

INCIDENCE RATES FOR DIABETIC COMPLICATIONS

Annual incidence rate per 100,000 diabetics

Complication	Type I Diabetics			Type II Diabetics		
	U.S.	U.K.	Ratio U.K./U.S. Percent	U.S.	U.K.	Ratio U.K./U.S. Percent
DKA/hyperosmolar coma	8,550	3,930	46	90	70	78
Retinopathy	n/a	n/a		4,910	2,780	57
Lower extremity amputation	220	140	64	210	130	62
Blindness	240	120	50	330	160	48
Average	*3,003*	*1,397*	***47***	*1,385*	*785*	***57***

Exhibit A-2

MORTALITY INDICATOR COMPARISONS

Comparison	Type	Results
Raw death rates per capita from diabetes	All	U.K. 30-70% of U.S. in most age/sex groups
Raw death rates from diabetes per diabetic	Type I	U.K. 47-81% of U.S. in age groups 15-34
Death rates per capita with diabetes as cause or contributing factor	All	U.K. 76% of U.S.
Cohort studies on death rates of diabetics over defined study period	Type I	U.K. 28% of U.S.

022 EF 136271/3

Chapter 4: Cholelithiasis (gallstones) case

This chapter discusses the relative productive efficiency of Germany, the U.K., and the U.S. in the treatment of cholelithiasis.

We begin with an overview of the disease, a discussion of the productive efficiency measure used, and a description of the treatment process. After assessing the relative productive efficiency of these countries in the treatment of cholelithiasis, we analyze the provider behaviors driving these productive efficiency differences. Finally, we discuss how different health care system structures and regulatory environments affected provider incentives and constraints and, therefore, productive efficiency.

BRIEF DISEASE OVERVIEW

Approximately 11 percent of the population of the U.S., the U.K., and Germany, totaling more than 42 million people, have cholelithiasis (gallstones). Nearly 2 million new cases are diagnosed in these countries each year.[1] Most gallstones, however, never cause symptoms. Only 1 to 4 percent of patients with gallstones develop symptoms or complications each year; 10 percent of all patients with cholelithiasis develop symptoms 5 years after diagnosis, and 20 percent develop symptoms after 20 years. In 1992, an estimated total of $7 billion was spent on treatment of gallstones in the U.S., the U.K., and Germany, making this one of the costliest, as well as most common, digestive diseases.

The primary function of the gallbladder is to store bile, a substance that aids in the digestion of fats. Because the liver provides a nearly continuous supply of bile, the gallbladder is not essential to digestive function. The complications that arise from stones in the gallbladder range from abdominal pain to life-threatening conditions such as acute cholecystitis.[2] The most common symptoms caused by gallstones are nonspecific forms of abdominal pain that may result from a number of other conditions, most of them self-limited. When gallstones are discovered in the evaluation of such symptoms, or are known to be present from previous

---

[1] Sources: National Hospital Discharge Survey; Hospital Episode Statistics; Krämling et al., 1993.

[2] Additional potentially life-threatening conditions include empyema of the gallbladder, common bile duct (CBD) 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 rather uncommon. In addition, the risk of gallbladder cancer in patients with gallstones is very low (currently estimated at 1 of 1,000 patients per year). This cancer risk, therefore, does not normally justify prophylactic treatment.

evaluations, determining whether or not the gallstones are the source of the symptoms is a challenging task for physicians.

Symptomatic gallstones are usually treated by cholecystectomy, or surgical removal of the gallbladder. One of two surgical procedures is commonly used, laparoscopic or open cholecystectomy. The decision to elect surgery is made after taking into account the severity of symptoms, the risk of complications, and any co-morbidities. Cholecystectomies are highly successful and exhibit very low complication rates, in contrast to less successful nonsurgical alternatives. Thus, the efficacy of surgery has made it the treatment of choice for symptomatic cholelithiasis.

DEFINITION OF PRODUCTIVE EFFICIENCY MEASURE

We evaluated relative productive efficiency in each country from both a per-operation (averaging among all patients that underwent a cholecystectomy) and a per-case (averaging among all patients with cholelithiasis) perspective. The per-operation results highlighted the differences in resource allocation per-operation, while the per-case results most comprehensively measured the overall input usage when treating cholelithiasis in each country. By exploring both results, we were able to obtain important insights into each health care system's relative efficiency.

Timeframe of analysis

Our study of cholelithiasis and its treatment addressed conditions in the U.S., the U.K., and Germany during the years 1991 to 1993. When necessary, we drew from studies outside this timeframe.[3] The percentage of operations performed laparoscopically (rather than by the traditional open approach) varied greatly during these years, however, so that the results depended heavily on which year was chosen. All three countries adopted the laparoscopic technique rapidly between the years of 1988 to 1994. We studied the effects of the rapid dissemination of the new technique, ensuring that data from the same years were used for intercountry comparisons. (See Appendix 4D for a summary of major sources used.)

Summary of disease management and treatment phases

We divided the management of cholelithiasis into three phases (**Exhibit 1**):
1) diagnosis; 2) treatment; and 3) recovery. Although a large number of clinical

---

[3] For example, prevalence studies from the late 1980s were used. The prevalence data, however, were insensitive to the small differences in time period (less than 5 years).

decisions were made throughout the process (**Exhibit 2**), we highlighted the key decisions in a simplified model of the management and treatment process shown in **Exhibit 3**. In the diagnosis phase, patients and physicians decide whether or not to proceed with surgery, undergo nonsurgical therapy, or have no treatment. If surgery is selected, the patient receives preoperative 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.

Measurement of inputs

We assessed inputs on the basis of four units of measurement: per case of cholelithiasis, per operation (weighted total of open and laparoscopic), per open operation, and per laparoscopic operation. On average, of the total input usage of cholelithiasis treatment, 74 percent was consumed for the treatment phase and 26 percent for the recovery phase (**Exhibit 4**). This 3:1 ratio varied somewhat among countries, with the U.S. spending a slightly smaller percentage on the recovery phase.

Our input methodology is described in Appendix 1A. In brief, we accounted for the actual use of labor, supplies, and capital in the treatment of cholelithiasis. Because recovery time is a significant component of the costs of treating cholelithiasis, we also included the opportunity cost of patient time. We measured this opportunity cost by weighting the number of work hours spent by the patient in the hospital and during recovery by the average hourly wage in the country.

In **Exhibit 5** we describe the model employed to understand the inputs used in treating a patient with gallstones at various levels of aggregation. First, we summed the per-operation use of labor, supplies, and capital separately for the open and laparoscopic operations in each country. Next, we used the relative number of open and laparoscopic surgeries 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 of cholelithiasis is simply the input total per operation multiplied by the surgical frequency per case of cholelithiasis. We estimated this figure by dividing the percentage of cholecystectomies per capita by the percentage of the population having cholelithiasis (calculated separately for each country). Unless specified otherwise, we discuss results on a per-case basis.

To the extent possible, our analysis incorporates measures of inputs used in each step of surgical treatment, including the treatment of complications, CBD exploration, and stone removal. We excluded the following two areas of treatment:

¶ **Nonsurgical treatments.** Nonsurgical treatments are excluded because they were rarely used. Among the excluded treatments are ESWL (extracorporeal shockwave lithotripsy), oral and contact dissolution therapies, and mechanical extraction of the stones through a catheter placed in the gallbladder (either percutaneously or endoscopically). Although these nonsurgical methods held some promise for patients with particular profiles, those treated without surgery frequently developed recurrent gallstones. Consequently, one-third to one-half of all patients who initially underwent noninvasive treatment underwent a cholecystectomy within 5 to 10 years.[4] Nonsurgical approaches, therefore, were not commonly used except under special patient conditions.

¶ **Diagnostic tests and analgesia for symptomatic patients who received no further treatment.** For patients who underwent cholecystectomy, any of several diagnostic tests might have been performed, including ultrasound, X ray, and blood chemistry tests. These tests, as well as prescribed pain medications, were included in the analysis, but they were excluded for symptomatic patients who did not receive further treatment. The input usage from these factors among untreated patients was relatively low and accurate estimates difficult to obtain.

Measurement of outcomes

Outcomes, like inputs, are calculated both on a per-case and a per-operation basis. To determine the outcome *per operation*, we investigated the morbidity and mortality rates for a cholecystectomy. We found that the incidence of complications from these common operations was quite low and was similar in the U.S., the U.K., and Germany: 3.0 to 5.0 percent for open cholecystectomy and 3.5 to 4.4 percent for the laparoscopic operation.[5] Thus, the open and the laparoscopic surgical options had similarly high success rates. The expectation of a successful surgery was both high and equivalent across the countries, leading to equal estimated outcomes per operation (i.e., the relative per-operation outcome = 1 for all three countries).

Outcomes *per case* were somewhat more complex. The relative success of cholelithiasis treatment depended crucially on the decision to proceed with surgery. Both the patient's potential benefit from surgery and the degree of success per operation affected the per-case outcome. Ideally, both of these factors should be incorporated into the outcome measure to assess the overall quality of cholelithiasis treatment. While the degree of surgical success was

---

[4]   Sources: National Hospital Discharge Survey; Hospital Episode Statistics; Krämling et al., 1993.

[5]   Sources: McIntyre and Wilson, 1993; Roslyn, 1993; *Ann R Coll Engl*, 1994; U.S. Hospital Survey.

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 quality adjusted life years (QALYs). QALYs were calculated using a tool for measuring a patient's health – the Kaplan-Bush Index of Well-Being, also known as the Quality of Well-Being scale (QWB).[6] This index ranges from 0 = death to 1 = full health, with all other health states lying in between. For a cholelithiasis patient, 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 the totally healthy state of 1. (Details of the conceptual basis for measuring QALYs and our implementation of this measure are described in Appendix 4A. Results from our calculations are discussed later in the text.)

Outcomes differed depending on the extent to which surgery alleviated severe symptomatic disease or treated disease that only marginally detracted from quality of life. For example, if pain episodes were frequent, then performing surgery to remove the gallbladder significantly improved outcomes. But if pain episodes were infrequent, then outcomes were diminished little if fewer operations were performed, as in the U.K. Because the frequency of symptomatic pain episodes varies from patient to patient, we calculated outcomes using a reasonable frequency range of symptoms every 14 days to symptoms every 60 days; when choosing a single point estimate, we assumed that symptoms occur every 30 days.[7]

DESCRIPTION OF THE MANAGEMENT AND TREATMENT PROCESS

The following sections describe each of the three phases in the cholelithiasis management and treatment process in some detail and highlight the clinical and economic trade-offs implicit in the range of treatment options available in each phase. Later in this chapter, we discuss the specific practice patterns observed in each country and their implications for productive efficiency.

Diagnosis phase

The detection of gallstones was usually prompted by the new onset of episodic abdominal pain. The "classic" symptom of gallstones, biliary colic, consists of

---

[6] Source: Kaplan et al., 1976.

[7] Source: Clinician interviews.

severe epigastric pain (located in the upper middle abdomen) or pain located in the right upper quadrant of the abdomen, lasting 1 to 5 hours and often (but not always) waking the patient at night. Many patients may have experienced abdominal pain whose characteristics were vague and less clearly related to gallstones. Diagnostic investigation of suspected gallstone disease nearly always included an upper abdominal ultrasound, although other tests, such as liver function tests (inexpensive, widely available blood assays) and imaging procedures (e.g., HIDA scan, a radionuclide technique for imaging the gallbladder) were also used. Ultrasound was most common since it is noninvasive, generally accessible, and accurate in identifying the presence of stones. These tests were usually ordered by the primary care physician or general practitioner (GP), who usually made the surgical referral. Identification of stones in the gallbladder and/or CBD confirmed the diagnosis.

Most symptomatic gallstone patients were treated, as were nearly all patients who exhibited complications. Two clinical features determined whether surgical treatment was indicated (**Exhibit 6**): the strength of the relationship between the patient's symptoms and the presence of gallstones, and the magnitude of the morbidity they caused. Key features of morbidity were the severity and frequency of symptoms, the risk of complications, and the patient's other co-morbid conditions. In some cases, the diagnosis of symptomatic gallstones was not clear cut. With as much as 11 percent of the population having gallstones and with only 20 percent of these developing symptoms or complications over a 20-year period, the incidence of false positives could have been quite high.[8] In these cases, a number of alternative diagnoses were considered and investigated, leading to further evaluation.

Physicians determined the probability that symptoms were due to gallstones based on the symptom characteristics and other clinical features and laboratory tests. Based on this evidence, each physician then decided whether or not to recommend surgical treatment. Presumably, physicians in the three countries applied similar reasoning and used similar indications for surgery (e.g., removing gallstones that were clearly symptomatic). However, the rates of surgical treatment differed substantially by country (**Exhibit 7**); in the U.K. about 0.48 percent of cholelithiasis cases were treated surgically, while nearly four times as many cases were treated in the U.S. (1.87 percent), and even more in Germany (2.12 percent).[9]

Surgical rates might vary simply because the prevalence of gallstone disease varied by country. Arguably, such variation in prevalence could arise from variation in risk factors for gallbladder disease, such as advanced age, female gender, and obesity. However, as described in Appendix 4B, the prevalence of gallstones was approximately equal in the three countries (11 percent of the

---

[8]   Sources: National Hospital Discharge Survey; Hospital Episode Statistics; Krämling et al., 1993.

[9]   The sources for the number of cholecystectomies in each country are described in Appendix 4C.

population). Such evidence suggests that the difference in surgical rates resulted from other factors.

The major factor driving surgical frequency was whether the physician considered the patient's gallstones to be symptomatic (and, hence, required surgical removal). Some physicians recommended surgery after only one symptomatic episode, whether mild or severe. A more conservative approach was to wait for surgery until after the patient experienced several mild episodes or more than one severe bout of biliary colic. Physicians may vary in assessing the severity of symptoms or the likelihood that symptoms are due to gallstones before recommending or performing surgery. How physicians interpreted patient pain varied among physicians within a country and, in aggregate, varied among physicians of different countries; and the health care system may have influenced their approach to patient management when the benefits of treatment were uncertain. We discuss the effect of both medical and economic factors in the following sections.

Treatment phase

As noted before, the poor efficacy of the nonsurgical treatments (e.g., ESWL, dissolution therapies) rendered them a rarely used treatment option. We therefore focused our treatment discussion on surgical options only (see again **Exhibit 3**). For the past 100 years, the prevailing surgical treatment of symptomatic gallstones had been an open cholecystectomy through an abdominal incision to remove the gallbladder. If necessary, the surgeon also explored the CBD for stones and removed them as necessary. Traditional cholecystectomy resulted in a 6- to 11-day hospital stay and a 3- to 6-week period of convalescence. [10] Both the mortality (less than 0.05 percent) and morbidity (less than 5 percent) of this well-established operation were quite low.[11]

Late in the 1980s, laparoscopic cholecystectomy became available as a surgical alternative to the open operation.[12] This operation was first performed successfully in France in 1987 and in the U.S. in 1988. (The original laparoscopic cholecystectomy was performed in Germany several years earlier, but this operation was not entirely successful and, therefore, received only muted publicity.) Visualization of the gallbladder and surrounding vital structures

[10] Sources: CPHA; HCIA; Royal College of Surgeons of England Confidential Audit; Federal Office of Statistics; clinician interview.

[11] Sources: Scott; McIntyre and Wilson, 1993; Federal Office of Statistics; clinician interviews.

[12] Other surgical treatments were not widely used. Minilaparotomy cholecystectomy is a modified open operation that removes the gallbladder through a much smaller incision. Published data were limited to less than 200 cases, such that this small number precluded meaningful discussion of the technique. Cholecystectomy, gallbladder drainage, and stone removal could also be achieved either percutaneously or under local anesthesia. However, these methods were employed rarely and only for high-risk or debilitated patients.

using a laparoscope allowed surgeons to perform this operation without the large incision necessary for open surgery. After filling the abdominal cavity with carbon dioxide gas, laparoscopic imaging and surgical instruments were introduced through multiple, small (about half-inch) incisions for visualization, manipulation, and dissection. The operation was viewed on a videoscreen with magnification. Once dissected, the gallbladder was removed via one of the small incisions as were the laparoscope and other instruments.

On average, the laparoscopic operation initially required slightly longer time in the operating room (approximately 80 to 90 minutes versus approximately 60 to 80 minutes for open), although operative time decreased as surgeons gained experience.[13] Both the open and the laparoscopic operations required general anesthesia and were subject to approximately the same risks and complications. However, the rate of CBD injury was higher for laparoscopy than for the open approach (0.7 to 1.0 percent for laparoscopic compared to 0.3 percent for open), while laparoscopic infection rates were lower (1.0 to 1.4 percent versus 2.0 to 2.8 percent).[14] Patients had less pain following the laparoscopic operation. Thus, the hospital stay (1 to 3 days) and convalescence (1 to 2 weeks) were usually shorter than after open cholecystectomy.[15] If CBD stones were suspected, a preoperative ERCP with sphincterotomy may have been performed to remove them. Similarly, a postoperative ERCP and sphincterotomy may have been performed if CBD stones were discovered during the laparoscopic cholecystectomy operation.

In both the open and laparoscopic operations, an intraoperative cholangiogram could be used to better ascertain whether stones were present in the biliary ductal system. This procedure was essentially an X ray of the bile ducts, which have been injected with a special dye. There was some debate in the medical community as to whether a cholangiogram should be routine and whether its use helped prevent complications. For the time period covered by this study, cholangiograms were more prevalent in open than in laparoscopic surgery. We included their use in our model of the production process.

A small percentage (less than 5 percent) of the laparoscopic operations were converted to open cholecystectomies, intraoperatively. This conversion may have occurred for several reasons, including difficulty in discerning a patient's anatomy. Medically, a conversion was not considered a complication, but rather a medical decision based on new evidence of the patient's condition at the time of surgery. We included these conversions, counting them as open surgeries, but costing them appropriately as a combination of laparoscopic and open inputs

---

[13] Sources: Herbet, 1993; clinician interviews.

[14] Sources: McIntyre and Wilson, 1993; Roslyn, 1993; *Ann. R. Coll. Engl.*, 1994; U.S. Hospital Survey.

[15] Sources: CPHA; HCIA; Royal College of Surgeons of England Confidential Audit; Federal Office of Statistics; Scott; McIntyre and Wilson, 1993; clinician interviews.

(e.g., higher capital inputs as for a laparoscopic surgery, but longer LOS as for open).

Because medical indications for laparoscopic cholecystectomy were similar to those for open cholecystectomy, most patients who could be treated surgically were candidates for the laparoscopic operation. The choice between open or laparoscopic surgery was principally driven by a patient's underlying co-morbid conditions and the surgeon's experience.[16]

National data show that the overall rate of cholecystectomy remained relatively constant after the introduction of laparoscopic cholecystectomy in the U.S., the U.K., and Germany, with some small increases observed for 1989 to 1991 (**Exhibit 8**). However, most experts in the field (in all three countries) believed that rates of surgery increased significantly since the adoption of laparoscopy. Several regional studies from the U.S. and Canada supported this assertion. In particular, statewide studies in Connecticut and Maryland, health maintenance organization (HMO)- and Medicare-based studies in Pennsylvania, Blue Cross/Blue Shield claims from the Washington, D.C. area, and a provincewide study from Ontario all showed increases in the rate of cholecystectomy.[17] These increases were as large as 20 to 30 percent in 1 year, with the largest jumps occurring in the years 1989 to 1991, the period immediately following the widespread adoption of laparoscopic cholecystectomy.

The apparent discrepancy between national and regional reports of the growth in cholecystectomy rate is only partially understood. One possible weakness of national hospital discharge data is that the data only included information from inpatient stays. Many laparoscopic operations were performed on an outpatient basis and, therefore, were not counted as hospital discharges. In addition, the advent of a new technique often causes some confusion in procedural coding. Thus, some laparoscopic cholecystectomies may have been miscoded during the initial years of the technique.

Many experts believed that the laparoscopic technique would, and did, increase the number of cholecystectomies because it offered a less invasive, less morbid alternative to the traditional open cholecystectomy. Less severe pain and shorter recovery time meant that the new surgical intervention would disrupt the patient's life less than the open operation. Because the physical and economic costs of laparoscopic surgery were lower, the pool of candidates expanded.

---

[16] Patients who were not candidates for laparoscopic cholecystectomy included those with generalized peritonitis, septic shock from cholangitis, severe acute pancreatitis, end-stage liver cirrhosis with portal hypertension, and patients with known gallbladder cancer. Patients with acute cholecystitis, quiescent acute gallstone pancreatitis, prior upper abdominal surgery, and symptomatic gallstones in the second trimester of pregnancy may have been candidates depending on the surgeon's experience with complex laparoscopic procedures.

[17] Sources: Orlando et al., 1993; Steiner et al., 1994; Legoretta et al., 1993; Escarce et al., 1995; Klar and Kongstvedt, 1994; Choen et al., 1996.

Patients who were frail, obese, or had other co-morbidities that precluded open surgery may have been able to undergo the laparoscopic treatment. In addition, patients who were in the "gray area" – whose symptoms were not severe enough to make open surgery a necessity – may have chosen (or been urged) to undergo laparoscopic cholecystectomy. In short, while the textbook indications for surgery may not have changed (symptomatic gallstones), the interpretation of what symptomatic meant may have become more aggressive in light of the availability of less invasive surgery. The surgical rate was expected to plateau at a new, higher rate rather than continuing its upward climb. Some evidence for this leveling off was evident in the Maryland study in which 1991 and 1992 rates were constant, following increases in 1989 and 1990.

After selection of the surgical technique, several other choices had to be made during the treatment phase. Medical decisions in the treatment phase included the set of procedures performed (e.g., X ray, ERCP) and the patient's LOS in the hospital. In accordance with the less disruptive nature of the operation, the LOS for a patient undergoing a laparoscopic cholecystectomy was considerably shorter (2 to 6 days) than for the open option (6 to 11 days).[18] Complications, when encountered, also generated medical decisions in the treatment phase. Rates of complications were low (less than 5 percent) for both open and laparoscopic operations, and the decision to treat was generally straightforward.

Recovery phase

Following the treatment phase, the patient returned home to recover. We specifically incorporated recovery time into the model for cholelithiasis (but not for the other diseases) to further highlight the differences between the open and laparoscopic surgical options. In this phase, the important decision was how long a patient recuperated before returning to work or normal activities. Generally, this time was determined through physician/patient consultation. For reasons similar to those motivating differences in LOS, recovery time following laparoscopic surgery was considerably shorter (1 to 2 weeks) than after an open operation (3 to 6 weeks).[19] The less invasive laparoscopic operation caused the patient less pain with faster resumption of normal mobility and diet. In our model, we did not specifically separate pharmaceuticals (e.g., pain medications) into those administered during the hospital stay and those used during at-home recovery. Since patients took the majority of medications while still in the hospital, all pharmaceuticals were counted as part of the treatment phase.

---

[18] Sources: CPHA; HCIA; Royal College of Surgeons of England Confidential Audit; Federal Office of Statistics; Scott; McIntyre and Wilson 1993; clinician interviews.

[19] Sources: CPHA; HCIA; Royal College of Surgeons of England Confidential Audit; Federal Office of Statistics; Scott; McIntyre and Wilson 1993; clinician interviews.

Decisions as drivers of input differences

We included all of the above decisions in our model of the management and treatment process.   For example, diagnostic tests; pre-, post-, and intraoperative operations; surgical inputs; pharmaceuticals; and recovery time were accounted for in the input measure for patients who underwent a cholecystectomy. Our analysis did not incorporate the resources used for the care of patients who were not treated surgically.

Among the many choices made by physicians during treatment, five decisions were especially likely to lead to differences in the inputs used for treatment in the U.S., the U.K., and Germany (**Exhibit 9**). Four of these decisions were made primarily through the physician-patient interaction: surgical frequency (how often a cholecystectomy was performed); surgical technology (whether open or laparoscopic techniques were used); LOS (how long a patient remained in the hospital); and recovery time (how long a patient recuperated at home). In addition to these decisions made by the physician and patient, a fifth choice also affected resources used – hospital staffing levels. As described in Chapter 1, staffing levels are closely related to LOS; short LOS, since it involves more patient turnover, typically requires higher staffing levels. Because decisions regarding complication management and which diagnostic operations to perform (e.g., X ray, ultrasound) did not differ significantly across the three countries, they also did not drive input differences.

Generally, clinical decision making – even when it appeared to be the joint decision of physician, patient, and hospital – was necessarily influenced by nonmedical (e.g., economic) factors. In subsequent sections, we explore these economic incentives and constraints and how they filtered through the structure of each country's health care system.

ASSESSMENT OF RELATIVE PRODUCTIVE EFFICIENCY

U.S. versus U.K.

Productive efficiency per operation. The U.S. was more productive than the U.K. on a per-operation, per-laparoscopic cholecystectomy, and per-open cholecystectomy basis because it used fewer inputs in each case while achieving the same outcomes (**Exhibits 10 and 11**). On a per-operation basis, the U.S. and the U.K. had similar outcomes, but the U.S. utilized 71-percent fewer inputs than

the U.K. This was because the U.S. used fewer resources per operation than the U.K. for both open and laparoscopic operations, but achieved similar outcomes.[20]

Productive efficiency per case. Although the U.S. used fewer inputs per operation than the U.K., the U.S. performed more operations. Thus, on a per-case basis, the U.S. used 56-percent more inputs than the U.K. (**Exhibit 10**). This higher rate of surgery in the U.S. yielded outcomes that were 76-percent better than those in the U.K. on a per-case basis (**Exhibit 11**).

As discussed in Chapter 1, when one country had both higher inputs and outcomes, we assess which nation was more productive by comparing average productivity. If the country with the higher inputs and better outcomes also had higher average productivity, and there is no reason to believe that the treatment process at the per-case level showed increasing returns to scale, the country with higher average productivity was more productive. For cholelithiasis, the U.S. had higher average productivity than the U.K., and this advantage did not vary with the frequency of symptoms. For example, at 14 days between symptoms, the U.S. was 72-percent higher than the U.K.; and at 60 days between symptoms, the U.S. was 76-percent higher (**Exhibits 12 and 13**). As the disease treatment process in cholelithiasis did not show increasing returns, we conclude that the U.S. was more productive than the U.K. in cholelithiasis treatment.

U.K. versus Germany

Productive efficiency per operation. On a per-operation basis, Germany was more productive than the U.K. because it used fewer inputs while achieving similar outcomes (**Exhibits 10 and 11**). For open cholecystectomy, the U.K. used 15-percent more inputs than Germany, while for laparoscopic cholecystectomy, the U.K. used 13-percent fewer inputs than Germany. Surgical outcomes were equal; thus, the U.K. was more productive in open operations, while Germany was more productive in laparoscopic operations. Germany was thus more productive per operation than the U.K. based largely on its higher use of laparoscopy.

Productive efficiency per case. Germany, which used more inputs to obtain better outcomes, had greater average productivity than the U.K. Assuming that the disease treatment process did not show increasing returns, this implies that Germany was more productive. While there were not large differences in inputs per operation between the two nations, Germany performed surgery more than four times as frequently, so that inputs per case of cholelithiasis were 128-percent higher in Germany (**Exhibit 10**). The greater frequency of surgery also implied, however, that Germany had higher outcomes per case than the U.K. (**Exhibit 11**).

---

[20] Inputs per laparoscopic cholecystectomy shown in Exhibit 10 are indexed to the U.S. open surgery inputs to highlight the lower input usage of laparoscopic surgery.

If symptoms occurred every 14 days on average, the average productivity in Germany was 13-percent higher than in the U.K. If symptoms only occurred once every 60 days, average productivity was 16-percent higher in Germany than in the U.K. (**Exhibits 12 and 13**). Because increasing returns to disease treatment are very unlikely at the per-case level, Germany was more productive than the U.K. in cholelithiasis treatment.

U.S. versus Germany

Productive efficiency per operation. On a per-operation basis, the U.S. employed 52-percent fewer resources than Germany while obtaining similar outcomes (**Exhibits 10 and 11**). Even though Germany used 8-percent fewer inputs per open cholecystectomy, the cost per operation was lower in the U.S., both because a greater percentage of procedures were laparoscopic and because the U.S. used about 40-percent fewer resources for laparoscopic cholecystectomy. Since outcomes per operation were equal, Germany was more productive than the U.S. in open surgery, while the U.S. was more productive in the laparoscopic technique and on an overall per-operation basis.

Productive efficiency per case. Since Americans with cholelithiasis were less likely to receive an operation and since the U.S. used fewer inputs per operation, the U.S. consumed lower inputs per case (72 percent as compared to Germany) and had lower outcomes on a per-case basis relative to Germany (**Exhibits 10 and 11**). The U.S. had 52-percent higher average productivity than Germany over the entire range of symptoms occurring between 14 and 60 days (**Exhibits 12 and 13**). In this case, we need detailed knowledge of the treatment process to determine which country was more productive because the country with the lower input/lower outcome combination (i.e., the U.S.) was the country with higher average productivity. Because of shorter hospital stays, shorter recovery periods, and broader adoption of laparoscopy, the U.S. used 72-percent fewer inputs per operation, with identical surgical outcomes. These advantages in 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 U.S. was more productive.

MAJOR DRIVERS OF PRODUCTIVE EFFICIENCY DIFFERENCES IN TERMS OF PROVIDER BEHAVIOR DIFFERENCES

The productive efficiency differences observed were caused most directly by differences in provider behavior. In this section, we discuss the provider behaviors that were the major drivers of productive efficiency differences between the nations, focusing first on drivers of input differences.

Summary of drivers of input differences

As discussed, the three systems exhibited considerable divergence in input usage. The following discussion describes the four main drivers of differences in input usage on a per-case and per-operation basis, including: 1) surgical frequency; 2) choice of surgical technology; 3) inpatient procedures; and 4) recovery time.

1. Surgical frequency. The U.K. achieved most of its reduction in inputs by operating on fewer patients than in the U.S. and Germany (**Exhibits 14 and 18**). The per capita rate of cholecystectomy was 0.21 percent in the U.S., 0.23 percent in Germany, and 0.05 percent in the U.K. (Appendix 4C). The prevalence of cholelithiasis is about 11 percent in all three countries (see the discussion in Appendix 4B). Dividing surgical incidence by gallstone prevalence yielded a surgical frequency (per case of cholelithiasis) of 1.87 percent in the U.S., 2.12 percent in Germany, and 0.48 percent in the U.K. (see again **Exhibit 7**).

The data indicated that a patient with cholelithiasis was about four to five times as likely to undergo surgery in the U.S. or Germany as in the U.K. This difference in care triaging reduced input usage on a per-case basis by about 74 percent in the U.K. relative to the U.S. In the absence of any medical or epidemiological evidence to the contrary, the lower U.K. surgery rate suggested that, in aggregate, U.K. physicians used more stringent criteria than American or German physicians for determining which gallstones were "symptomatic" (and, hence, required surgery). Clinician interviews indicated that while on the average U.S. and German physicians recommended surgery after one episode of biliary colic, U.K. physicians may have waited for a second episode. This anecdotal evidence supported the hypothesis that U.K. physicians generally recommended surgery less readily or less often than their American and German counterparts. We describe specific causal links to health care system structure in the following sections.

The cholecystectomy rate is likely to have influenced the outcomes of cholelithiasis by changing the prevalence and duration of pain. Recurrent bouts of painful stones may have afflicted many of the patients with cholelithiasis who did not have an operation. The frequency and severity of these episodes determined the magnitude of the potential decrease in outcomes. Outcomes are discussed further below.

2. Choice of surgical technology. During the time of this study, the U.S. and Germany each used the laparoscopic approach for about 80 to 90 percent of all cholecystectomies.[21] In contrast, the U.K. used the laparoscopic approach for only about 60 percent of all cholecystectomies.[22]

---

[21] Sources: National Institutes of Health, 1993; Orlando et al., 1993; Krämling et al., 1993; clinician interviews.

[22] Source: Russell, 1993.

The choice between open and laparoscopic surgery had a major impact on total inputs per case (**Exhibits 15 and 18**). This choice was also the major driver of input difference between the U.S. and U.K. on a per-operation basis (**Exhibits 15 and 20**). Laparoscopic cholecystectomy required fewer inputs than open surgery because both hospital stays and postdischarge recovery time were shorter. Thus, the greater use of laparoscopy in the U.S. lowered U.S. per-case inputs compared to the U.K. (14 percent) and Germany (10 percent). Some additional up-front capital costs were required for laparoscopy, but these did not outweigh the benefits of the operation. About $36,000 to $45,000 was necessary for the initial laparoscopic equipment setup (insufflation and video equipment).[23] And when a laser was used rather than electrocautery (which was already available in operating rooms), a further investment of $60,000 to $100,000 was needed.[24] Studies showed that operations performed by electrocautery or by laser met with equal success and, therefore, the more costly laser technology was not usually employed.

Despite the major differences in use of the laparoscopic technique at the time of the study, all three countries adopted the laparoscopic technique more rapidly than usual for new surgical technology. The rates of adoption varied, however. As shown in **Exhibit 17**, penetration of laparoscopic cholecystectomy in the U.S. rose most quickly, followed closely by Germany. The increase in laparoscopic operations in the U.K. occurred more slowly, but even in the U.K. approximately 85 percent of cholecystectomies were laparoscopic by 1996. This quick uptake was spurred, in part, by surgeons' desires to maintain state-of-the-art techniques. The rapid uptake caused fewer prospective clinical studies to be performed, which initially raised some questions about the efficacy of the laparoscopic technique. However, as the operation became more widely used, more data comparing laparoscopic and open complication rates became available. Today the safety and efficacy of a laparoscopic cholecystectomy is widely accepted as equivalent to the open method. The adoption of laparoscopy is discussed further below in the section on incentives and constraints.

3. Inpatient procedures: LOS and hospital staffing. As discussed in Chapter 1, the decisions about the duration of a hospital admission and the staffing required per day of the admission are intertwined. Because these factors are not independent, we analyzed their combined impact on input usage. Although the short LOS in the U.S. lowered input usage, it was associated with high staffing levels. How these factors jointly influenced overall input usage depended on the particular disease; but for cholelithiasis, the net effect was to raise input usage. Specifically, the joint impact of LOS and staffing caused the U.S. input usage to be 5-percent higher than the U.K. and 2-percent higher than Germany.

---

[23] Sources: McIntyre and Wilson, 1993; *Minimal Access Surgery: Implications for the NHS*, 1993; Wagner, 1990.

[24] Sources: Gadacz et al., 1990; Sadler et al., 1992; Talamini and Gadacz, 1992.

Exhibits 15 and 18 through 21 show these combined effects of LOS and staffing levels on inputs. Below we discuss each of these factors in turn.

¶ **LOS**. The average LOS for laparoscopic cholecystectomy was 2.97 days (U.S.), 3.80 days (U.K.), and 5.85 days (Germany); these numbers are significantly lower than for open cholecystectomy, which had an LOS of 6.77 days (U.S.), 10.92 days (U.K.), and 10.14 days (Germany).[25]

The U.S. had the lowest LOS for both laparoscopic and open surgery (2.97 and 6.77 days, respectively). A small portion of the LOS differences for laparoscopic surgery resulted from the differences in use of outpatient facilities. Outpatient cholecystectomy was rare in the U.K. and Germany, but as many as 20 percent of the U.S. laparoscopic cholecystectomies were performed as outpatient operations.[26] Note that the average LOS for inpatient laparoscopy in the U.S. (3.46 days) was still lower than the LOS in the U.K. (3.80 days) and in Germany (5.85 days). For open surgery, LOS in the U.K. and Germany were about 50-percent longer than the U.S., and are comparable at 10.92 days for Germany and 10.14 days for the U.K. For laparoscopic surgery, the U.K. LOS (3.80 days) was 25-percent longer than the U.S., while LOS in Germany (5.85 days) was almost 100-percent longer than in the U.S.

¶ **Staffing levels.** Staffing levels can be described in units of staff levels per bed-day or per admission. In Chapter 1 we discussed in detail these two ways of measuring inpatient staffing. That discussion highlighted two trends. Staffing per bed-day was highest in the U.S., followed by the U.K., then Germany. However, looking at staff per admission, Germany was the highest, with the U.S. roughly equal to German staffing, and the U.K. lowest by a considerable margin. In order to disaggregate the effect of LOS on staffing levels per bed-day, we discuss the causal factors behind the staff per admission trends in the following section.

4. Recovery time. Recovery time was a significant component of input usage in gallstone treatment, generating approximately 25 percent of total inputs. Inputs during recovery time were dominated by patient labor, which was measured as the work hours missed by a patient during recovery, priced at an average hourly wage. The average recovery time for laparoscopic cholecystectomy was 6 days

---

[25] Sources: CPHA; HCIA; Royal College of Surgeons of England Confidential Audit; Federal Office of Statistics; clinician interview.

[26] Sources: HCIA; hospital reimbursement information; clinician interview.

(U.S.), 10 days (U.K.), and 16 days (Germany), compared with open recovery times of 28 days (U.S.), 40 days (U.K.), and 32 days (Germany).[27]

The U.S. had the shortest recovery time for both open surgery and laparoscopic (28 and 6 days, respectively). For open surgery, recovery time in Germany (32 days) was slightly longer than the U.S., while the U.K. (40 days) was about 40-percent longer than the U.S. For laparoscopic surgery, the U.K. recovery time of 10 days was 60-percent longer than the U.S., while Germany, at 16 days, was considerably longer (over 150 percent) than the U.S. (**Exhibit 16**). The impact of these differences on input usage was significant, with German inputs 23-percent higher than the U.S. (**Exhibits 16 and 19**), while U.K. inputs were 5-percent higher (**Exhibits 16 and 18**) on a per-case basis.

Summary of drivers of productive efficiency differences

U.S. versus U.K. Three differences in provider behavior led to differences in productive efficiency between the U.S. and the U.K.: technology adoption, treatment duration, and staffing levels (**Exhibit 22**). The primary cause of the higher U.S. productive efficiency in the treatment of cholelithiasis was faster adoption of the laparoscopy procedure. Lower LOS and recovery time in the U.S., which were at least partially related to the adoption of the laparoscopic operation, also increased U.S. productive efficiency relative to the U.K. And while the relative productive efficiency of the U.S. was lowered by higher levels of staffing, the net effect of different U.S. provider behavior was higher productive efficiency than in the U.K. In the following section, we discuss the sources of each of these provider behavior differences. In addition, we discuss differences in the frequency of surgery, which did not necessarily affect relative productive efficiency, but had a significant impact on relative resource consumption and on overall outcomes.

U.S. versus Germany. Productive efficiency differences between the U.S. and Germany were caused by the same differences in provider behavior as in the U.S. versus U.K. comparison: treatment duration, staffing levels, and technology adoption (**Exhibit 22**). In this comparison, however, the primary cause of differences in productive efficiency was treatment duration, with the longer LOS and recovery time in Germany significantly decreasing relative productive efficiency. Productive efficiency in Germany was also lowered by slightly later adoption of laparoscopy. And while the U.S.'s productive efficiency was lowered by higher levels of staffing, the net effect of different U.S. provider behavior was higher productive efficiency than in Germany. In the following section, we discuss the sources of each of these provider

---

[27] Sources: Scott; McIntyre and Wilson, 1993; Federal Office of Statistics; clinician interviews. These figures represent the total number of recovery-days, but they were multiplied by 5/7 in the calculation of patient input time used, in order to reflect the average days of work missed.

behavior differences. In addition, we discuss differences in the frequency of surgery, which did not necessarily affect relative productive efficiency, but had a significant impact on relative resource consumption.

CAUSAL ANALYSIS OF PROVIDER BEHAVIOR DIFFERENCES: INCENTIVES AND CONSTRAINTS, SYSTEM STRUCTURE, AND REGULATION

The differences in provider behavior in the U.S., U.K., and Germany can be explained by the incentives and constraints providers faced in each country and by underlying differences in the health care system structure and regulation, which are described in detail in Chapter 2.   Below, we explain how these differences led to the provider behavior differences we observed in the management and treatment of cholelithiasis.

U.S. versus U.K.

As previously discussed, laparoscopic operations today account for roughly the same percentage of cholecystectomies in all three countries. At the time of our analysis, however, substantially fewer cholecystectomies in the U.K. were performed laparoscopically. Yet a highly price-sensitive health care system, like that of the U.K., would be expected to adopt laparoscopic cholecystectomy earlier, since it was less resource-intensive than the open option (because of shorter LOS and recovery time). Several factors may explain this apparent paradox: physician incentives, hospital incentives, and capital and hospital supply constraints. Below, we discuss these incentives and constraints and how they arose from the structure and regulation of the different health care systems. In addition, we discuss incentives for U.S. hospitals to manage LOS to a greater extent and have higher staffing than U.K. hospitals; regulations leading to longer recovery time in the U.K.; and incentives for U.S. physicians to perform more operations than U.K. physicians.

1. Incentives for U.S. physicians to adopt laparoscopy faster than U.K. physicians (Exhibits 23 and 24). U.S. physicians had incentives to adopt laparoscopy to meet patient demand. The less invasive nature of laparoscopy, along with the positive portrayal of this technique in the mass media, led to a rapid growth in patient demand for it in the U.S. As U.S. providers and payors competed for patients, those who did not offer or cover laparoscopic cholecystectomy could have lost patients to competitors. In addition, most physician services in the U.S. were negotiated and compensated on a fee-for-service (FFS) basis, and some physicians initially received slightly higher fees for performing laparoscopic surgery. The threat of malpractice suits may also have fueled physician incentives to perform the less invasive laparoscopic procedure. Finally, because the laparoscopic

technology was more cost-effective, U.S. payors that competed partly on price readily accepted the laparoscopic substitute for the open procedure. The U.S. health care system structure and regulatory environment thus gave physicians strong incentives to adopt the new technology.

U.K. specialists, on the other hand, did not compete for patients and were paid flat salaries regardless of the amount or type of procedures they performed. U.K. specialists, therefore, had little incentive to adopt the laparoscopic operation. In addition, the adoption of laparoscopy was hindered by capital and hospital supply constraints, which we discuss below.

2. Incentives for U.S. hospitals to adopt laparoscopy; constraints on U.K. capital and hospital supply that led to slower adoption of laparoscopy (**Exhibits 23 and 24**). Like U.S. physicians, U.S. hospitals had incentives to adopt laparoscopy rapidly due to the competitive intensity of the U.S. health care system. Hospitals responded to demands from both patients and physicians (led by surgeons) to invest in the necessary equipment. The cost of a laparoscopic setup ($36,000 to $45,000) was relatively modest, compared with other new technologies; and most American hospitals were not restricted by external budget controls. Thus, hospital investment in the equipment, fueled by competition for patients and physicians, was rapid. Because hospitals were originally reimbursed for laparoscopic cholecystectomy at the same rate as the open operation, hospitals stood to gain from the lower LOS requirement of a laparoscopic cholecystectomy, as discussed below.

In contrast to the incentives of U.S. hospitals to adopt laparoscopy, U.K. hospitals faced constraints that slowed its adoption. Limited U.K. hospital supply, which was tightly controlled through regulation, gave hospitals incentives to keep the traditional open operation, since it used less operating room time and resulted in lower rates of some complications (e.g., bile duct injury) during initial adoption of the technology. In addition, U.K. hospitals faced capital constraints that led to slower adoption of laparoscopy. As discussed in Chapter 2, U.K. hospitals applied through a regional allocation process under the National Health Service budgeting process in order to receive additional capital funding. In some cases, additional funds to purchase new, unbudgeted equipment were unavailable until the next budget was set.

Further stalling the capital investment was initial skepticism in the U.K. about the actual savings potential of the new technique and fears that laparoscopy would increase usage of already scarce hospital resources. While some studies estimated lower costs for the laparoscopic operation, others suggested that an open cholecystectomy was the more cost-effective option. As laparoscopy became more established and physician experience increased, the operating time for the operation decreased, reducing the overall inputs required for laparoscopy. Even when the reduced input usage of laparoscopic surgery was recognized, many in the U.K. also feared that the less invasive technique

would increase the number of total cholecystectomies performed and, hence, increase overall spending and use of hospital resources on cholelithiasis.

It is interesting to note, however, that laparoscopic surgery was adopted quickly in the small private health care sector of the U.K., which was similar in many respects to the U.S. health care system.

3. Incentives for U.S. hospitals to manage LOS to a greater extent than U.K. hospitals (**Exhibits 23 and 24**). The U.S. had shorter LOS for open operations (10.9 days in the U.K. versus 6.8 days in the U.S.) and for laparoscopic cholecystectomy (3.8 versus 3.0 days). This lower U.S. LOS was due both to the more rapid adoption of laparoscopy, from which patients recovered faster than the open operation, and to differing hospital incentives in the U.S. and U.K.

At the time of our study, most U.S. payors individually contracted with hospitals for services. Payments stipulated by these contracts were generally of three basic types: FFS (about 45 to 50 percent), per case (about 35 to 40 percent) or per diem (about 15 percent). FFS payments paid hospitals a fixed amount per operation or service. Per-case payments paid hospitals a fixed amount per admission based on the patient's diagnostic-related group (DRG).[28] Under this system, the reimbursement was defined solely by the operation or diagnosis, not the number of days the patient was in the hospital. While FFS reimbursement did not influence hospital decisions on LOS, DRG-based reimbursement gave hospitals a clear incentive to decrease costs by reducing LOS per patient or managing the intensity of services provided, since the hospital itself assumed the risk for the cost of the patient's stay. Hospitals thus had an incentive to lower LOS by streamlining admission and discharge processes and shortening pre- and postoperative time.

The per diem contract, in which reimbursement was tied to the hospital-day rather than the entire admission, was not as common as FFS or per-case payment. Payors following this arrangement coordinated physician care and managed patient throughput (e.g., through a utilization review process). Thus, under a per diem contract, the payor assumed the risk for hospital stay and there was no incentive to lower LOS.

Regardless of the type of contractual arrangement, the risk-bearing entity (payor or provider) had an incentive to manage hospital costs in the U.S. Thus, competition among U.S. payors in the health coverage market likely helped to accelerate the adoption of case rates beyond the Medicare market. Since private payors competed for customers in part on the basis of price, they saw the adoption of case rate pricing as a way to control increasing hospital costs.

---

[28] The Medicare program in the U.S. adopted case rate (DRG) based payment in 1983; many other payors subsequently adopted this payment system as well.

In contrast, the U.K. health care system structure and regulation created economic incentives for U.K. hospitals that differed from those in the U.S. and were not aligned with lowering LOS. In the U.K., fixed hospital budgets, coupled with limited competition among hospitals for patients or payor funds, created no incentives for U.K. hospitals to manage LOS. Lower LOS implied that patient population would turn over more rapidly (i.e., greater admission and discharge rates). Thus, caring for new patients and performing additional operations was more costly than maintaining patients in recovery at the end of a longer LOS. In other words, cycling through more patient cases in which treatment required higher input usage would have raised *total* hospital costs without generating a corresponding increase in revenue. In addition, if capacity reallocations were made, then efficient hospitals with high throughputs could have been viewed as overresourced. These efficient hospitals would have had high *total* input usage (as opposed to inputs per patient) and might have been penalized through the imposition of resource reductions.

4. Incentives for U.S. hospitals to have higher staffing than U.K. hospitals (Exhibits 23 and 24). The lower LOS in the U.S. likely led to higher U.S. staffing levels, since the intensity of care per average hospital-day increased with decreasing LOS. System structure and regulation also likely led to the different levels of staffing in the two countries. U.K. hospitals did not compete for patients as in the U.S. and, therefore, did not need to offer higher levels of service to attract patients. U.K. hospitals also had limited budgets with which to fund all their hospital needs, including staffing. In addition, U.K. staffing levels were also affected by the availability of physicians, which was regulated and limited by the NHS. Thus, U.K. hospitals were under more pressure to lower staffing levels than those in the U.S.

5. Regulations leading to longer recovery time in the U.K. (Exhibit 24). The average recovery time in the U.S. was shorter than in the U.K. for open (28 days versus 32 days) and laparoscopic (6 days versus 10 days) operations, leading to fewer days of work absence in each case. Since the medical procedure was the same in both countries, the range of recovery times was surprising and suggests that nonmedical factors were largely responsible for the variation in recovery times. The most important of these factors was the regulatory environment – how patients and physicians were affected by differences in disability policies.

In both countries, physicians made recommendations about how much time each patient needed to recuperate. In the U.K., the patient's GP oversaw recovery time, while in the U.S. the surgeon made the recommendation. Although surgeons may have systematically urged earlier return to full activity than GPs or other primary care physicians, such differences should have been small for a well-established procedure like cholecystectomy. Physicians did not estimate recovery time based solely on personal expertise and the characteristics of each patient; but instead physicians tended to adopt common practices, adhering

closely to a "normal" recovery time, which tended to follow government and employer disability policies.

In the U.S., the individual states determined disability payments. Employers often supplemented government disability policies by allowing additional disability-days. Disability payments in the U.S. were approximately 50 to 60 percent of the patient's wages, and the limit of duration was approximately 6 weeks. In general, whatever the physician recommended was allowed, within certain limits. For the U.K., disability payments were determined by individual companies and their employees, and they were often more generous than the usual U.S. government limits. In fact, most people in the U.K. initially received their full wages (for at least the first 30 days), with 6- to 100-percent payment continuing for up to 6 months, depending on the employer.

6. Incentives for U.S. physicians to perform more operations than U.K. physicians (Exhibits 23 and 24). While differences in the frequency of surgery did not necessarily impact relative productive efficiency, they were the most important driver of input differences between the two countries. Physicians in the U.S., paid on an FFS basis, as described above, had strong incentives to perform operations. Since surgery relieved symptoms, and was straightforward and relatively safe, it was easy to justify surgical treatment for cholelithiasis. Payor competition for members led payors to cover these operations, as those who did not would be unable to continue attracting new members.

In the U.K., however, surgeons were salaried, and hospital budgets were not linked to the number of procedures. Physicians and hospitals derived little financial gain from operating more frequently. In general, cholecystectomy was an elective operation that could be deferred to make room for the treatment of emergency conditions, which may have been necessary given the supply constraints in the U.K. health care system.

The more complex U.K. system of referrals, established to accommodate the limited supply of specialists, may also have limited surgical frequency. GPs, the primary point of patient contact, served as the decision making channel between the patient and the surgeon; patients could not receive an operation without seeing a surgeon, and they could not see a surgeon without a referral from a GP. Similarly, access to a radiologist for a diagnostic ultrasound also required a referral from a GP or surgeon. This arrangement, with its more stringent referral requirements (compared with the U.S.), may have inhibited the use of cholecystectomy by implicitly requiring a higher "burden of proof" before a patient was investigated and treated. As a result, patients often waited to see if their symptoms recurred and worsened, potentially leading to an emergency surgery. Or, if their symptoms did not recur, surgery was no longer considered.

It is interesting to note that some evidence suggests that group model HMOs in the U.S., which faced incentives and constraints similar to those in the U.K., performed cholecystectomy at rates closer to those in the U.K. (**Exhibit 25**). One such health plan, with 300,000 members, performed 238 cholecystectomies in 1994, translating to a surgical frequency per total population of 0.08 percent. Because the age and gender distribution of the HMO population was similar to the total U.S. distribution, we assumed that the prevalence of gallstones (11 percent) in the HMO was the same as for the U.S., generating a surgical frequency for cholelithiasis of 0.72 percent in this HMO. This rate was slightly greater than for the U.K. (0.05 percent per capita and 0.48 percent per case), but was approximately one-third of the rate seen in the U.S. overall (0.21 percent per capita and 1.87 percent per case). (See again **Exhibit 7.**)

U.S. versus Germany

As discussed, the primary causes of differences in productive efficiency between the U.S. and Germany were differences in treatment duration, staffing levels, and technology adoption. Below, we discuss the incentives and constraints that led to these differences in provider behavior and how they arose from the structure and regulation of the different health care systems. We focus on: incentives for German physicians and hospitals to raise LOS and incentives for U.S. hospitals to lower LOS; incentives for U.S. hospitals to have higher staffing than German hospitals; incentives for German hospitals to adopt laparoscopy slightly slower than the U.S; and regulations allowing longer recovery time in Germany. In addition, we discuss incentives for German physicians to perform more operations than U.S. physicians, which led to Germany's higher resource use, but not necessarily to its lower productive efficiency.

1. Incentives for U.S. hospitals to lower LOS; incentives for German physicians and hospitals to raise LOS (Exhibits 26 and 27). Germany had longer LOS for both open and laparoscopic operations (10.1 versus 6.8 days for open and 5.9 versus 3.0 days for laparoscopic). Above, we have discussed how competition in the U.S. payor and hospital markets led to incentives to manage LOS in the U.S. Therefore, in this section, we explore German incentives to increase hospital utilization.

The German health care system and regulation gave German hospitals strong incentives to increase LOS. German hospitals were compensated on a per diem basis by the sickness funds, whereby they received additional reimbursement for each day of a patient's stay. Annual hospital budgeting was based on this per diem reimbursement, with the per diem level of the previous year being increased or decreased as necessary to compensate for inflation and for some of the increase or decrease in hospital utilization. Each hospital negotiated as a block with all payors for these per diem rates.

State authorities, as regulators of hospital capacity, had little incentive to reduce the use of local health care resources, as discussed in Chapter 2; in fact, they 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. Consequently, regulations generally stipulated very high occupancy rates, usually about 85 percent. If this target was not met, hospitals were at risk for being reviewed and having their capacity cut. The combination of this regulatory threat and the large supply of hospital beds created a strong incentive for hospitals and, therefore, the physicians they employed, to keep LOS long in order to keep more beds occupied. Physicians had further incentive to maintain high utilization of public beds because hospitals typically specified that beds had to be used for public and private patients in a relatively set ratio. German department chiefs could, therefore, add private bed capacity and earn higher private patient fees by increasing the use of public beds.

Limited competition in the German payor market also contributed to long German LOS. German sickness funds (payors) had relatively restricted memberships along geographic or occupational lines, and retained their members for a relatively long period. Payors did not compete for members on price and could not negotiate differentially from other payors with each hospital; therefore, they had no real incentive or ability to pressure hospitals to lower costs or manage LOS.

German payors and regulators have become aware of the potential shortcomings associated with per diem reimbursement, and they are currently addressing them by changing contractual policies. In 1996, cholecystectomy (and some other procedures) began to be reimbursed on a case rate basis. This case rate, by pushing some of the risk onto hospitals, is expected to shorten the average LOS considerably, since hospitals now have some incentive to reduce hospital stays as in the U.S.

2. Incentives for U.S. hospitals to have higher staffing than German hospitals (Exhibits 26 and 27). Hospitals in both Germany and the U.S. competed for patients, giving hospitals incentive to provide quality service through high staffing levels. While both Germany and the U.S. had higher staffing than the U.K., the U.S. had the highest staffing level for an inpatient visit overall. This was due to the fact that the U.S.'s shorter LOS likely necessitated higher staffing, as previously discussed.

3. Incentives for German hospitals to adopt laparoscopy slightly slower than U.S. hospitals (Exhibits 26 and 27). German adoption of laparoscopic cholecystectomy closely followed the U.S. pattern, resulting in relatively quick adoption by both countries. The operation was first used in the U.S. about a year before it was used in Germany, and the initial spread of the technique occurred earlier in the U.S. Providers in both countries competed for patients and were thus encouraged to compete with each other and maintain state-of-the-art facilities and skills in order to attract patients. Hospitals in both countries were

relatively unconstrained in making funds available for purchasing laparoscopic equipment, in contrast to the capital constraints seen in the U.K. With the help of training centers established by product manufacturers, physicians received training in the new techniques. Provider incentives, therefore, allowed for a rapid uptake of the laparoscopic operation in both the U.S. and Germany.

While both countries adopted laparoscopy quickly, uptake in the U.S. was slightly faster. In part, this may have been a response to earlier marketing efforts by equipment suppliers. In addition, since German hospitals had incentives to maintain high occupancy rates, the shorter LOS resulting from laparoscopic treatment would have represented an important disadvantage of the technique. The system structure and regulatory reasons that led to long German LOS (described above) may have also served to slightly slow the adoption of laparoscopy in Germany relative to the U.S.

4. Regulations leading to longer recovery time in Germany. Like cholecystectomy patients in the U.K., patients in Germany received substantially more generous benefits when recuperating from surgery than in the U.S. In Germany, sickness fund reimbursement to employers was mandated by law; regulation stipulated that employees could receive 6 weeks of disability time, with 100-percent wage replacement. These benefits were far more generous than those mandated in the U.S. As described earlier, these disability policies influenced provider-recommended recovery times; recovery times in Germany were, therefore, much longer than those in the U.S.

5. Incentives for German physicians to perform more operations than U.S. physicians. The cholecystectomy rate in both Germany and the U.S. was high, with the rate in Germany being about 10- to 15-percent greater than in the U.S. While the frequency of operations did not necessarily influence relative productive efficiency, it did affect relative input usage.

U.S. physician FFS and hospital reimbursement, as well as per diem reimbursement in Germany, created incentives for providers to perform the operation. In combination with these hospital and physician incentives, the abundant supply of physicians and hospital beds in both countries made it possible to perform a large number of cholecystectomies. As a low-risk procedure with substantial discretion in its application, along with a large pool of potential patients, cholecystectomy was a procedure that could be performed widely. While both nations widely used cholecystectomy to treat symptomatic cholelithiasis, Germany performed operations at a higher rate. [29] This was likely due to the enhanced incentives to perform operations in Germany due to German hospital incentives to maintain high capacity, as discussed earlier.

---

[29] In addition, some German physicians may also have removed asymptomatic gallbladders as a prophylactic measure (about 5 percent of all German cholecystectomies).

SUMMARY OF CHOLELITHIASIS CASE RESULTS

In the treatment of cholelithiasis, the U.S. was the most productive, Germany was the next most productive, and the U.K. was the least productive. The primary source of these differences in productive efficiency was the rate at which the nations adopted laparoscopy: the U.S. most quickly adopted laparoscopy and was the most productive, whereas the U.K.'s slow adoption caused it to be the least productive. Treatment duration also influenced productive efficiency, especially in the case of Germany, which had the longest LOS and recovery time from surgery. While U.S. productive efficiency was slightly reduced relative to both other nations by its higher staffing levels, the impact of this difference was minimal compared to the impact of its quick technology adoption and shorter treatment duration and recovery.

The slower adoption of laparoscopy in the U.K. relative to the other nations was caused by differences in the health care system structure and regulation. While physicians and hospitals in both the U.S. and Germany competed for patients, and thus had incentive to respond to patient demand for laparoscopy, U.K. providers did not face this pressure, given the structure and function of the NHS. In contrast, tight controls on capital and hospital supply impeded the adoption of the new technology in the U.K.

While Germany adopted laparoscopy almost as quickly as the U.S., its productive efficiency was decreased by its extensive treatment duration, both in terms of LOS and recovery time. German hospitals and physicians had strong incentives to keep hospital utilization high, given per diem reimbursement, the threat of capacity cuts if utilization fell below regulated percentages, and the fact that department chiefs could increase private bed capacity if more public beds were filled. While U.K. hospitals did not have these direct incentives to increase LOS and had a more limited supply of hospital beds than the other nations, the U.K. annual hospital budget created incentives to increase LOS as a method of preserving resources through "bed-blocking." Only in the U.S., where payors faced competition for members and some case rate payment existed for hospitals, were there direct incentives to lower LOS. Recovery time was also lowest in the U.S., as the U.S. regulatory environment provided somewhat less generous disability coverage than in the other two nations.

Appendix 4A: Outcome measure and consequences for relative performance

As described in the body of the chapter, outcomes for the U.S., the U.K., and Germany were estimated to be equal on a per-operation basis. This was a reasonable assumption because the operation was highly successful at relieving gallstone symptoms; furthermore, complication rates were low and essentially equal across the three countries. However, a more detailed outcome measure was necessary to compare outcomes on a per-case basis. For per-case outcomes, it was necessary to measure the reduction of pain achieved by a cholecystectomy. For example, if a patient's pain was severe before the surgery, a successful operation would improve the patient's quality of life substantially. We measured this effect in units of QALYs, as in the diabetes case. Although we were unable to obtain direct data on pain levels before and after surgery, we approximated these levels by estimating the frequency of pain episodes and using the Kaplan-Bush QWB scale to quantify the "disutility" of pain. Before showing the results, we first describe how QALYs would ideally be measured in cholelithiasis, if detailed pain data were available. We then discuss qualitative – but theoretical – examples of possible relative outcomes for the U.S., the U.K., and Germany. Finally, we present our quantitative model and results.

Ideal outcome definition

The *ideal* outcome measure for cholelithiasis is illustrated in **Exhibit A-1**. The assumptions were as follows:

1. The expected QALY for a patient with no pain was high and constant for all countries.

2. Having surgery slightly reduced the quality of life. This negative impact on expected QALYs was constant for all countries.

3. It was possible to distinguish the severity of pain that different patients experienced and to rank them by the severity of pain. Here severity means both the level and frequency of pain and/or complications caused by the patient's gallstones.

4. Patients received surgery in descending order of the severity of their symptoms, i.e., more severe patients underwent surgery before less severe ones.

5. More severe patients had lower expected QALYs than patients with less severe cholelithiasis. The expected QALY for each severity level allowed a "severity curve" to be constructed.

6. The severity curve was the same for all three countries. Caveat: the curves shown and discussed here are illustrative examples, not definitive results.

The ideal outcomes result was the improvement in expected QALYs (comparing before and after surgery), summed over all cases. Those cases not undergoing surgery contributed zero to this sum. If too many cases received surgery, then the contribution to the outcomes could be negative. This negative contribution occurred if the number of surgeries performed exceeded the point where the severity curve crossed the "after surgery" curve (**Exhibit A-1**). For each country, the line denoting the number of cases receiving surgery was marked on the graph. The outcome for the country was the area bordered by the y-axis, the after surgery line, the country's number of surgeries line, and the severity curve. As an illustrative example, the outcome for the U.S. is shown by the shaded area in **Exhibit A-1.**

Shape of severity curve

The ideal outcome measure was limited by lack of data. In order to obtain these outcomes, the severity curve was essential. Constructing this curve, however, required knowing the severity of every case of cholelithiasis in the country. Severity of cases could be described by the level of pain in each patient. But pain levels were difficult to measure since descriptions and threshold levels varied from patient to patient. Furthermore, even if these pain levels were measurable, this information was not available on any large-scale basis.

We did not know the shape of the severity curve. Instead, we described some examples and what they would have indicated about relative outcomes. Three possible severity curves and the resulting performance (input/outcome) charts are shown in **Exhibit A-2**. Here we used the appropriate surgical frequencies to order the countries, but all other values were illustrative. Our analysis showed that, depending on the shape of the curve, any of the three countries could have had the highest outcome. If only a few cases were severe (**Exhibit A-2A**), then the U.K. physicians made the appropriate choice of surgical frequency, while the Americans and Germans overproduced. If many but not all cases could benefit from surgery, as in **Exhibit A-2B**, then the U.S. maximized the outcome and Germany performed the operations too frequently. In this case, the U.K. did not achieve the full benefits possible, but used only a limited amount of resources to treat cholelithiasis. For a flat severity curve (**Exhibit A-2C**), every surgery contributed equally to the outcome. This flat curve suggested that as many surgeries as possible should be performed (provided that the first surgery is cost-effective). Germany achieved the best outcome, while the reduced inputs

in the U.S. and U.K. led to fewer QALYs. Thus, the shape of the severity curve was critical in determining how outcomes would be affected by the different surgical rates. More surgery was not always better, but it improved outcomes when it was directed toward the most severe cases.

Although our data did not permit us to derive an ideal rate of surgery for treating cholelithiasis, we were able to make some general observations about how this number could affect overall performance and productive efficiency. The quality of the performance depended on two factors: the frequency of surgery and the use of inputs per operation. As demonstrated in the cases for the U.K. and Germany, the inputs per operation (particularly laparoscopic surgery) were clearly higher than in the U.S. Thus, regardless of how many surgeries were performed in these two countries, the U.K. and German systems could not reach their maximum productive efficiency potential until they each reduced their per-operation inputs. For the U.S., the input usage per operation may also have been higher than necessary, but how far these inputs could be reduced was not clear from this simple evaluation. The important message, then, was that both the surgical frequency *and* the inputs per operation should be optimized. In addition, the optimal number of surgeries to perform was unknown without knowledge of the severity curve.

Quantitative outcome model

Although we did not know the exact severity curve for cholelithiasis patients, we developed a model of severity based on the probability of experiencing pain and other symptoms, the frequency of symptomatic episodes, and the length of time during which a patient may have experienced pain (e.g., while waiting to undergo surgery). The resulting estimates of symptom status were then combined with ratings for health states from the Kaplan-Bush QWB scale to convert the health states into QALYs. Other preference-assessment techniques for rating health states would not give identical results, but our general conclusions should not be affected by which index was used.

As shown in **Exhibit A-3**, the cholelithiasis outcome model divided patients into four categories: 1) "emergency," those undergoing emergency surgery; 2) "elective," those undergoing elective surgery; 3) "borderline," patients who underwent surgery in Germany but not in another country; and 4) "no surgery," patients who did not undergo surgery in any country and could be considered asymptomatic. Germany was used as the defining country for borderline patients because its high rate of surgery was interpreted to mean that it had the highest percentage of "discretionary" indications for cholecystectomy. "No surgery" patients were not included in the model since they would have contributed equally in all countries. In all cases, patients were assumed to have

no additional debilitating conditions beyond cholelithiasis.[30] This assumption allowed us to obtain outcomes that reflected differences only in cholelithiasis treatment.

From the first three categories, patients were further divided into symptom groups of differing severity. The symptom groups corresponded to health states described by Kaplan-Bush symptom states, to facilitate calculations. The functional and symptom states, with their corresponding Kaplan-Bush indices, are given in **Exhibit A-4**. Emergency patients may have experienced fever, vomiting, and pain stemming from inflammation or other complications. (Here we used "emergency" to describe truly complicated cases.) Elective patients may have had vomiting and pain, pain only, or pain with medication and/or diet modification. Borderline patients may have had pain or pain with medication and/or diet modification. Two additional health states were also possible. During surgery, and for a brief period following surgery (1 day for laparoscopic and 3 days for open), the patient was confined to a hospital bed, with only a limited ability to walk and carry out daily tasks. This lowered functional state was accounted for in the model according to the Kaplan-Bush index. In addition, because of the erratic nature of gallstone symptoms, all presurgery patients may have had some days when they experienced no symptoms. These pain-free days were also included, each contributing a healthy-day score of 1 to the total outcome.

With the patients divided into different symptom groups, and each symptom category assigned a Kaplan-Bush QWB score, the next step was to sum up the different contributions to the QALY score. For each health state, the outcome was the product of the Kaplan-Bush index and the number of days in the state. The total outcome was the sum of the health state contributions, each weighted by the probability of being in that health state (see **Exhibit A-5**). The number of days in each health state and the probability of being in each state are discussed below. The resulting sum of daily outcomes (quality adjusted life *days*) was indexed by the total probability and converted to QALYs (quality adjusted life *years*). Differences in outcomes were also calculated, relative to the U.S.

The total timeframe modeled in our calculation was 27 years, the difference between the average age at surgery (49 years) and average life expectancy of the population (76 years). For borderline patients not undergoing surgery, the days in these years were either pain-free or contained a pain episode. The exact number of days in each condition was determined by the frequency of pain attacks. For patients undergoing surgery, a small part of the 27 years was spent in the presurgery phase, as determined by the wait-time between diagnosis and surgery. Wait-times for emergency surgery were 1 day for the U.S. and Germany and 3 days for the U.K. For elective patients, average wait-times were 7 days in

---

[30] This assumption was carried out by using, for all patients, the same functional state (from the Kaplan-Bush QWB scale).

the U.S. and Germany and 53 days in the U.K., as shown in **Exhibit A-6**. Of these wait-time days, some included a symptomatic episode and some were pain-free. As for the borderline patients, the exact number of pain and pain-free days was determined by the frequency of pain episodes. Following surgery (and immediate recovery) the emergency and elective patients experienced pain-free, totally healthy-days for the remainder of the 27 years.

The frequency of pain episodes was difficult to measure directly. In fact, if this frequency was known, then we could have constructed a crude approximation to the severity curve, described above in the ideal outcome measure. However, estimates of symptom frequency were made, based on clinician experience. In emergency cases, pain episodes occurred daily. For patients undergoing elective surgery, pain episodes likely occurred every 2 weeks. And for borderline patients, the frequency of symptoms ranged from once every 2 weeks to once every 3 months. We calculated the outcome for the high- and low-frequency ranges using the same frequency for each country. As symptom frequency for borderline patients decreased, the U.K. outcome decreased. This was expected since, as symptom frequency decreased, the need for surgery in borderline patients also decreased. Thus, the more restricted U.K. use of surgery did not diminish outcomes. This scenario corresponded to the severity curve pictured in **Exhibit A-2A**, with the U.K. limited-surgery approach performing cholecystectomy only in patients who were highly symptomatic.

In reality, the frequency of patient symptoms will vary over the pool of surgery candidates. Since the three countries perform surgeries at different rates, the frequency of pain for patients near the surgery "cutoff" may differ by country. Unfortunately, data on frequency of pain episodes are unavailable. However, the directionality of our results would be unchanged by this additional subtlety.

The probability of being in one of the health states was approximated by the percent of cases who did or did not receive surgery, according to the surgical rates of each country. The division into symptomatic groups was based on data from published studies and from clinician interviews. Comparing data from different studies showed, as in **Exhibit A-6** for emergency cases, that the probability of being in one or another symptom group was roughly equal in each country, to within data availability. The exception was that in Germany, an estimated 5 percent of the surgeries were performed on asymptomatic patients, presumably as a prophylactic measure. The estimated percentages of patients in each symptom group is given in **Exhibit A-7**, along with the Kaplan-Bush index for each health state.

Appendix 4B: Sources for prevalence of cholelithiasis

This appendix describes the sources we employed to determine the prevalence of cholelithiasis in the U.S., the U.K., and Germany. These sources were divided into two types: those that gave an estimate for overall prevalence and those that reported detailed findings from ultrasound investigations on selected populations. In the latter studies, ultrasound was used to determine the presence of both symptomatic and asymptomatic gallstones. Results were generally reported by age and gender groups and often divided into patients with asymptomatic gallstones, with symptomatic gallstones, or patients having had a previous cholecystectomy.

We first list the sources that gave broad estimates for prevalence, as well as the estimate itself. Next we describe the ultrasound studies and their findings. Finally, we discuss the value we used for the prevalence of cholelithiasis in each of the three countries. Only those studies that were specific to the U.S., the U.K., or Germany were considered.

PREVALENCE ESTIMATES

The sources given in **Exhibit B-1** estimate the overall prevalence of cholelithiasis. Prevalence was not explicitly defined in these papers. In general, prevalence included all gallstones (symptomatic and asymptomatic), as well as previously performed cholecystectomies. However, the term *prevalence* was also used more loosely to denote only existing gallstones (symptomatic and asymptomatic). These estimates were reported in the articles without a full explanation of their derivation. More detailed, explicit studies of prevalence are described in the following section.

PREVALENCE STUDIES

The previous estimates (**Exhibit B-1**) suggest that the prevalence of cholelithiasis in the U.S., the U.K., and Germany is somewhere between 5 and 20 percent. However, more explicit studies of prevalence have also been conducted. These studies gave prevalence data, determined from ultrasound screening, for

population groups based on age and gender. We list these studies in **Exhibit B-2,** then describe their findings in more detail below.

Discussion of sources

(9) Oxford/Pixley study. This study combined results from a survey and from ultrasound testing to compare the prevalence of gallstones in two groups of women (vegetarians and nonvegetarians), aged 40 to 69. The study found positive correlation between the prevalence of gallstones and obesity, age, and a family history of gallstones. In addition, the vegetarians had a lower risk of developing gallstones (1.9:1, after controlling for age and other factors).

For our purposes, the nonvegetarian results were the most interesting, since they better represent the general population. These women were randomly selected from those registered at two Oxford general practices; 632 nonvegetarians participated. Of these women, 113 (17.8 percent) had gallstones and 43 (6.8 percent) had previously undergone a cholecystectomy, yielding an overall prevalence of 24.6 percent. Because the study included only those women between the ages of 40 and 69, the demographic group with the highest rate of gallstones, the high prevalence was considerably greater than the prevalence for the total population.

The results of the study for the nonvegetarian women are given in **Exhibits B-3 and B-4**.

(10) Bristol/Heaton study. The aim of this study was to expand the available prevalence data to include men and younger women from a British population sample. A random sample of men (40 to 69 years of age) and women (25 to 69 years of age) was drawn from the lists of 19 GPs in East Bristol. All subjects were white, and a total of 1,896 were included. The study consisted of an ultrasound test for the presence of gallstones, which helped to classify the participant in one of three categories: normal gallbladder, gallbladder absent and cholecystectomy scar present, or gallstones. (If the scan was technically unsatisfactory then the test was repeated until a final conclusion about the presence of stones was reached.) Prior to the ultrasound, the participants were questioned regarding any history of abdominal pain. The questions were designed to assess the correlation between pain and possibly symptomatic gallstones.

This study was unusual in that the cholelithiasis prevalence, i.e., gallstones or a previous cholecystectomy, was nearly equal for men (7 percent) and women (8 percent). (In most other studies, prevalence in women was two or three times greater than prevalence in men.) In addition, the overall prevalence (7.5 percent) was on the lower end of estimates for the population as a whole. This study included only a portion of ages in the general population. However, these ages were those with the highest risk of gallstones. A probable explanation for the somewhat low prevalence was a regional fluctuation, possibly accentuated by

the ethnic bias of the population sample (all participants were white). In fact, the authors caution, "The results should not be extrapolated to the whole of Britain because there may be regional variations." However, by combining these results with those of the Oxford/Pixley study, we formed a more complete picture of prevalence throughout the U.K.

The results of the study for men and women are given in **Exhibits B-4 and B-5**.

(11) Düsseldorf Peter study. Like the British studies, the goal of this study was to evaluate the prevalence of gallstones. Before undergoing an ultrasound examination, the participants filled out a form regarding their eating habits and any possible symptoms of gallstone disease. The ultrasound then confirmed or denied the existence of stones. Participants were classified as having symptomatic or asymptomatic stones, or a previous cholecystectomy.

This study included both women and men, but all of the participants were hospital patients (1,512 in total). The authors cautioned that this sampling of hospital patients may have yielded prevalence estimates that were higher than would be expected in the general population. The overall prevalence of gallstones (including any patients with a previous cholecystectomy) for this study was 26 percent (17 percent with stones and 9 percent with previous gallbladder removal).

The results of the study for men and women are given in **Exhibits B-4, B-6, and B-7**.

(12) Schwedt & Neuruppin/Nurnberg & Berndt studies. Although published in 1991, this study was initiated during 1986 through 1988. Participants were recruited from the East German cities of Schwedt and Neuruppin. Neuruppin was a rural area and Schwedt was experiencing growth from a relatively small city to a larger, more modern, industrialized city. The Schwedt participants were taken at random from roles of inhabitants, but the Neuruppin participants were drawn from specific citizen or employment groups. Thus, the Neuruppin study group was less representative of the overall Neuruppin population.

As in the studies above, these German investigations employed both an ultrasound screening for gallstones and a series of questions to identify risk factors such as age, weight/height, and family history. The data for the men were reanalyzed and published anew as a combination of the Schwedt and Neuruppin data in order to better understand the risk factors. From Schwedt, 700 men and 700 women were tested, yielding prevalence data of 14 percent for men, 23 percent for women, and 18 percent overall. From Neuruppin, 683 men and 1,143 women participated, with prevalence results of 12 percent for men, 26 percent for women, and 21 percent overall. The total combined prevalence from this study was 20 percent, where prevalence includes both patients with gallstones and any previous cholecystectomy patients.

The results of the studies for men and women are given in **Exhibits B-4, B-6, and B-8**.

COMPARISON OF AGE/GENDER
GROUP PREVALENCE DATA

In this section we compare the results from the prevalence studies to determine whether the age/gender group prevalence distribution was similar in the U.S., the U.K., and Germany.

Our review of the literature suggested that the precise prevalence was not well-known in any of the countries, but that it probably lay within the 5- to 20-percent range. The important question was two-fold: was the prevalence roughly equal among the three countries of interest, and what was an appropriate estimate to use for prevalence in our calculations?

To determine whether the overall prevalence was roughly equal among the countries, we separated the overall prevalence into two components. The first was the prevalence distribution according to age and gender, and the second was the population distribution by age and gender. In other words, the overall prevalence was determined by the prevalence per age/gender group and by the distribution of the country's population as a whole. The risk of gallstones increased with age and was far higher in women than in men. Thus, a population that was weighted toward older women would have a higher prevalence than a more evenly distributed population, even if dietary and other factors kept the per age/gender group prevalence the same.

In order to compare the different studies, we first determined the 95-percent confidence intervals around the prevalence results. In this way, we made a first approximation to the error imposed by the finite sample in each study. We employed the standard formula for the error: $s = sqrt[p(1-p)/n]$, which was then weighted by $z=2$ to give the 95-percent confidence interval. Here n was the number of participants in the age/gender group and p was the prevalence for that group. By adding and subtracting the error ($2s$) to the prevalence (p), we obtained high and low estimates for the prevalence in each age/gender group. The results of this analysis are shown in **Exhibits B-4 and B-6**.

From **Exhibits B-4 and B-6** we see that the two German studies gave prevalence data that overlapped for all age groups in both genders. Furthermore, the prevalence from the British studies overlapped the German results in both genders. The Bristol/Heaton results overlapped with those from Düsseldorf/Peter for the men (**Exhibit B-6**). And for the women, the Oxford/Pixley results overlapped with those from both of the German studies (**Exhibit B-4**). The Bristol/Heaton data were somewhat lower for the women, but still shared significant overlap with the Oxford/Pixley and German data.

Because of the substantial overlap at all age groups and for both genders, we concluded that the prevalence age/gender distribution was roughly equal in the U.K. and Germany.

We did not analyze prevalence data from the U.S. Although some studies had been conducted, these were usually for nonrepresentative target groups, such as the Pima tribe of Native Americans or Mexican-American populations. Both of these groups have significantly higher prevalence of gallstones than the aggregate U.S. population. Thus, we did not use the results from these studies as a basis for the total U.S. prevalence. However, the U.S. and U.K. are expected to have similar prevalence based on broad similarities in diet, lifestyle, and other factors.[31] Consequently, we assumed that the U.S. prevalence distribution by age and gender was the same as for the U.K. and Germany.

OVERALL PREVALENCE BASED ON POPULATION DISTRIBUTION

In this section we show that the overall prevalence in the U.S., the U.K., and Germany was approximately equal because the age/gender population distributions of the three countries were not sufficiently different to significantly change the overall prevalence.

As described in the previous section, the total prevalence of cholelithiasis for a country depends on the age/gender distribution of both the prevalence and the general population. We showed that the prevalence distribution across age/gender groups was equal in the three countries. Next we show that the different age/gender population distributions of the three countries imposed no significant differences in the aggregate prevalence.

We used population data separated into age and gender groups for each of the three countries. We then calculated the overall prevalence by weighting the age/gender prevalence data by the population distribution and summing over all population groups. In this way, we found that the three countries shared the same overall prevalence to within 1 percent. This was easily within the range of error of the prevalence data.

Thus, we concluded that the U.S., the U.K., and Germany had approximately equal prevalence of cholelithiasis, despite small differences in population distribution (e.g., Germany was more heavily weighted toward an older population than the U.S.).

---

[31] Source: Sadler et al., 1992.

FINAL COMMENTS

Sensitivity of results to prevalence

Although it is possible that the overall prevalence was not exactly the same in the U.S., the U.K., and Germany, any reasonable differences would have been too small to significantly affect the results of our input calculations.

The input calculations were driven primarily by the surgical frequency per case of cholelithiasis. Since we used the same prevalence for all three countries, the ratio among the countries of the surgical frequencies per case of cholelithiasis followed the same ratio as for the surgical frequency per population (see **Exhibit 6,** Chapter 4). The differences in the surgical frequencies per population were quite significant, with the U.K. performing approximately four times fewer surgeries than the U.S. or Germany. In order for the prevalence to make a significant effect on the surgical frequency per case of cholelithiasis, the prevalence, for example, in Germany would have to be three to four times that in the U.K. However, such a large difference in prevalence is extremely unlikely, based on the current literature.

Choice of overall prevalence

From the assembled sources, the prevalence estimates of existing gallstones in the total population of the three countries were in the range of 7 to 15 percent. Here we used the prevalence of *existing* gallstones only and did not include those patients who had already undergone a cholecystectomy. The reasoning was simple: those who have already had their gallbladders removed will not enter the production process of diagnosis, treatment, and recovery that we outlined for the treatment of cholelithiasis. The prevalence was approximately the same for all three countries. Thus, the exact number we used, within the proper range, was not crucial to our analysis, i.e., it would not affect the overall rankings of the countries. We employed the midpoint of the prevalence range, 11 percent, as the prevalence of (existing) gallstones for the U.S., the U.K., and Germany.

Appendix 4C: Sources for surgical frequency

In this appendix we briefly describe the sources used to determine the surgical frequency for cholecystectomy in the U.S., the U.K., and Germany. In order to derive the surgical frequency, we used the number of cholecystectomies performed in the country per year, divided by the total population of the country. The resulting surgical frequencies were: 0.21 percent for the U.S., 0.05 percent for the U.K., and 0.23 percent for Germany. Population figures for 1992 are given in **Exhibit C-1**.

Determining the number of cholecystectomies performed per year was relatively straightforward for the U.S. and the U.K., but this figure was not as well-known for Germany. We first state the values that we used, as well as the other German estimates, and then give a brief description of each of the German sources (**Exhibit C-2**).

The number of cholecystectomies performed each year in Germany is uncertain and subject to controversy. The most commonly quoted estimates were in the range of 60,000 to 100,000. The origin of this figure, although widely cited, was not known and was generally regarded as too low. The figure from the Statistisches Bundesamt was also likely to be too low. This number (111,510) represented the count from forms filled out by physicians regarding the procedures performed. This form was not tied to reimbursement. Thus, the accuracy of this count was questionable (as had been shown for similar forms in the U.S. that were not required for reimbursement).

The figure from the Krämling article (190,000) was based on a survey of general surgeons that probed the state of laparoscopic cholecystectomy in Germany. Approximately 20 percent of all German surgeons (853) were chosen at random.[32] From these, the survey was sent only to those types of surgeons with the greatest probability to have had cholecystectomy experience. Of these 449 surveys, 325 were returned. Removing those who had no experience with gallstone surgery left the remaining 204, upon whose answers the survey results were based. The number of surgeries performed by the respondents was 37,850 per year. These respondents came from a pool representing 20 percent of all German surgeons; thus, extrapolation to all surgeons was trivial (i.e., multiply by 5), yielding approximately 190,000 cholecystectomies per year. Credence for this estimate was enhanced by the fact that *Der Chirurg*, which published this study, is a highly regarded journal for German surgeons. This number was further substantiated by

---

[32] *Deutsche Chirurgie*, 1992.

industry experts, whose estimates of the number of German cholecystectomies per year were similar (170,000).

Appendix 4D: Sources

This list details data sources used in the cholelithiasis case study. We cover most of the main topics here, but this list is not exhaustive of all of the articles and government statistics that were employed throughout our work. In addition, we performed interviews with clinical and health care experts at a number of points during our study. Through these interviews, we collected qualitative and quantitative data on treatment patterns and checked our key assumptions and conclusions.

Below, we give the main sources used by topic.

PREVALENCE OF CHOLELITHIASIS

See Appendix 4B.

NUMBER OF CHOLECYSTECTOMIES

National Hospital Discharge Survey, U.S. Centers for Disease Control.

NHS Hospital Episode Statistics, NHS.

Krämling HJ, Lange V, Heberer G. Aktueller stand der gallensteinchirurgie in Deutschland. Der Chirurg 1993; 64: 295-302.

NONSURGICAL TREATMENT OPTIONS

Gallstones and Laparoscopic Cholecystectomy, Consensus Development Panel, National Institutes of Health, 1993. (as cited in JAMA 1993; 269: 1018.)

ADOPTION OF LAPAROSCOPIC TECHNOLOGY AND RELATED DISCUSSIONS

Bloor K, Maynard A. Health Science Journal 1994; Nov 17: 24-26.

Choen MM, Young W, Theriault ME, Hernandez R. Has laparoscopic cholecystectomy changed patterns of practice and patient outlook in Ontario? Can Med Assoc J 1996; 154: 491-500.

Escarce JJ, Chen W, Schwartz JS. Falling cholecystectomy threshold since the introduction of laparoscopic cholecystectomy. JAMA 1995; 273: 1581-5.

Fullarton GM, Darling K, Williams J, MacMillan R, Bell G. British Medical Journal 1994; 81: 124.

Klar RM, Kongstvedt PR. JAMA 1994; 271: 500-1.

Legoretta AP, Silber JH, Costantino GN, Kobylinski RW, Katz SL. Increased cholecystectomy rates after the introduction of laparoscopic cholecystectomy," JAMA 1993; 270: 1429-32.

Orlando R, Russell JC, Lynch J, Mattie A. Laparoscopic cholecystectomy: A statewide experience. Arch Surg 1993; 128: 494-8.

Steiner CA, Bass EB, Talamini MA, Pitt HA, Steinberg EP. Surgical rates and operative mortality for open and laparoscopic cholecystectomy in Maryland. NEJM 1994; 330: 403-8.

U.S.

Table 22, National Hospital Discharge Data, Centers for Disease Control.

Gallstones and Laparoscopic Cholecystectomy, Consensus Development Panel, National Institutes of Health, 1993. (as cited in JAMA 1993; 269: 1018.)

Germany

Krämling HJ, Lange V, Heberer G. Aktueller stand der gallensteinchirurgie in Deutschland. Der Chirurg 1993; 64: 295-302.

Clinician interviews.

U.K.

McCloy R. Through the keyhole. Health Science Journal 1992; Nov 12: 26-7.

Russell RCG. General surgery: Biliary surgery. BMJ 1993; 307: 1266-9.

Industry interviews.

COST OF LAPAROSCOPIC TECHNOLOGY

Gadacz TR, Talamini MA, Lillemoe KD, Yeo CJ. Surgical Clinics of North America 1990; 70: 1249.

McIntyre IMC, Wilson RG. Br J Surg 1993; 80: 552.

Sadler GP, Shandall A, Reed BI. Laparoscopic cholecystectomy (Review). Br J of Hospital Medicine 1992; 48: 462.

Talamini MA, Gadacz TR. Advances in Surgery 1992; 25: 1.

Wagner M. The many reasons why hospitals buy technology. Modern Health Care 1990; Aug. 6: 21.

Minimal Access Surgery: Implications for the NHS (London: HMSO Publications Centre, 1993).

ESTIMATE OF OUTPATIENT LAPAROSCOPIC SURGERIES

U.S.

Outpatient Utilization Profile, HCIA.

U.K.

Clinician interview.

Germany

Hospital reimbursement information.

OPERATIVE TIME

Herbet, Surg L&E, 1993.

Clinician interviews.

FREQUENCY OF DIAGNOSTIC PROCEDURES

Hodgson.

Kelly.

McIntyre IMC, Wilson RG. Br J Surg 1993; 80: 552.

Martin.

Phillips, Am Surg, 1993.

Scott, Ann R Coll Phy, 1992.

Stoker.

Clinician interviews.

POSTOPERATIVE HOSPITAL-DAYS (LOS)

U.S.

CPHA.

HCIA.

Clinician interviews.

U.K.

Royal College of Surgeons of England Confidential Audit.

Clinician interview.

Germany

Statistisches Bundesamt, Federal Office of Statistics.

Clinician interview.

COMPLICATIONS DATA

McIntyre IMC, Wilson RG. Br J Surg 1993; 80: 552.

Krämling HJ, Lange V, Heberer G. Aktueller stand der gallensteinchirurgie in Deutschland. Der Chirurg 1993; 64: 295-302.

Roslyn. Open cholecystectomy, a contemporary analysis of 42,474 patients. Ann Surg 1993; Aug.

Laparoscopic cholecystectomy in England and Wales: Results of an audit by the Royal College of Surgeons of England. Ann R Coll Engl, 1994.

U.S. Hospital Survey.

RECOVERY TIMES

U.S.

Clinician interview.

U.K.

Scott, Ann. R. Coll. Phys.

McIntyre IMC, Wilson RG. Br J Surg 1993; 80: 552.

Germany

Statistisches Bundesamt, Federal Office of Statistics.

Clinician interviews.

WAIT-TIMES FOR CHOLECYSTECTOMY

NHS Hospital Episode Statistics, NHS.

Clinician interviews.

FREQUENCY OF SYMPTOMS FOR PATIENTS WITH CHOLELITHIASIS

Clinician interviews.

KAPLAN-BUSH QUALITY OF WELL-BEING SCALE

Kaplan RM, Bush JW, Berry CC. Health status: Types of validity and the Index of Well-Being. Health Services Research 1976; Winter: 478-507.
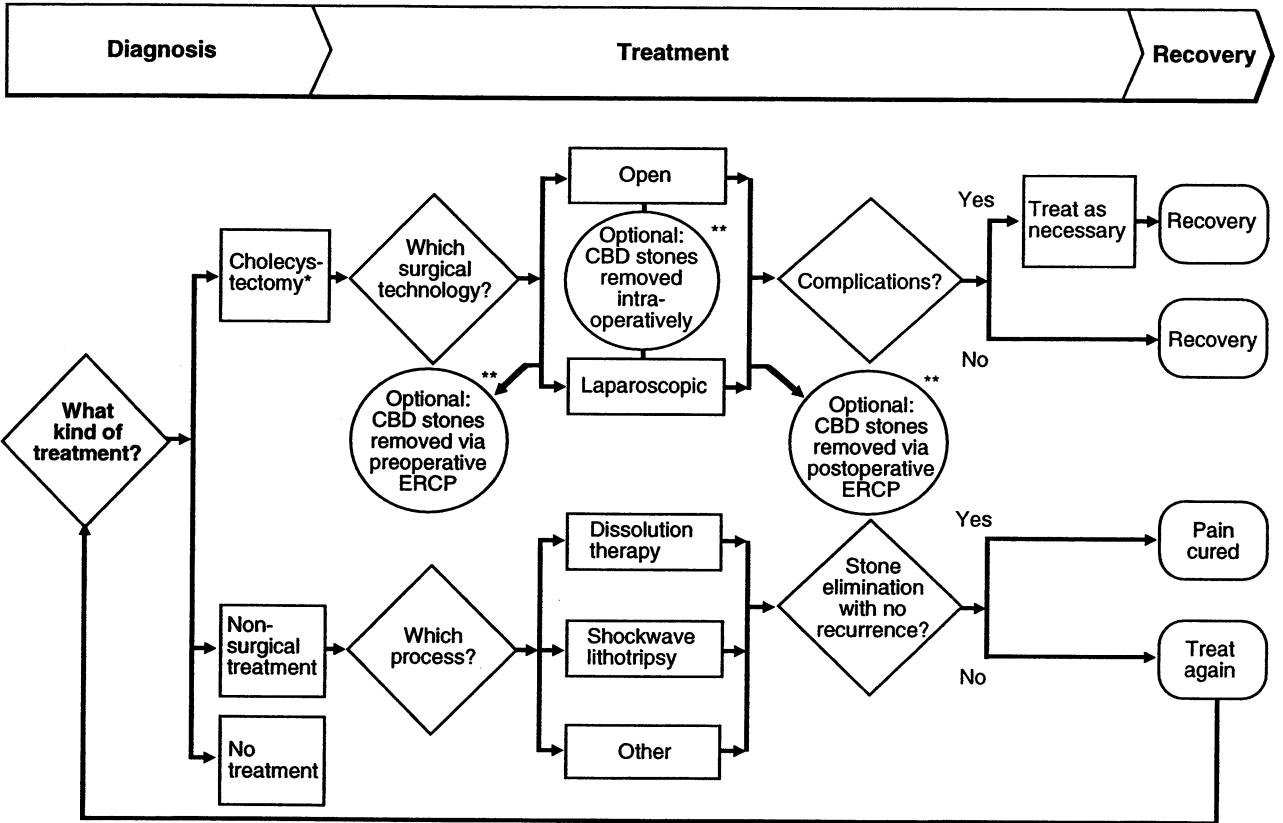
Exhibit 1

**PHASES OF CHOLELITHIASIS
MANAGEMENT AND TREATMENT PROCESS**

	Diagnosis	Treatment	Recovery
Purpose of phase	Determine whether treatment is necessary Determine treatment type (surgical vs. nonsurgical treatment)	Eliminate gallstones by surgical removal of the gallbladder (or by nonsurgical techniques) Treat and manage complications if they arise	Allow time to heal and recuperate before returning to work/activity
Comments	Decision made after considering frequency and severity of symptoms and patient co-morbidity	Cholecystectomy (open or laparoscopic) the norm, with nonsurgical techniques relatively uncommon Low complication rates (<5%) Possible CBD stone removal using preoperative, post-operative, or intraoperative endoscopic retrograde cholangiopancreatography (ERCP)	Laparoscopic cholecystectomy requires shorter recovery time and fewer postoperative drugs than open surgery
Average percent of inputs consumed in each phase*	–	74%	26%

* Average inputs consumed per phase in the U.S., U.K., and Germany; diagnosis stage not considered; treatment phase includes direct input usage of those cases receiving surgery

022 EF 136239/4

Exhibit 2

OVERVIEW OF CHOLELITHIASIS
MANAGEMENT AND TREATMENT PROCESS

| Diagnosis | Treatment | Recovery |

```
                                                    ┌──────────┐
                                                    │   Open   │
                                                    └──────────┘
                                                  ┌─────────────┐
                        ┌──────────┐   Optional:**│
         ┌──────────┐   │  Which   │  CBD stones  │         ┌──────────────┐  Yes  ┌──────────┐  ┌──────────┐
         │Cholecys- │   │ surgical │   removed    │         │Complications?│──────▶│Treat as  │─▶│ Recovery │
         │ tectomy* │   │technology?│  intra-      │         └──────────────┘       │necessary │  └──────────┘
         └──────────┘   └──────────┘  operatively │              │                  └──────────┘
                                                                 │ No
  ┌───────────┐          Optional:**                                           ┌──────────┐
  │   What    │         CBD stones    ┌──────────────┐          Optional:**    │ Recovery │
  │kind of    │          removed via  │ Laparoscopic │         CBD stones      └──────────┘
  │treatment? │          preoperative └──────────────┘          removed via
  └───────────┘             ERCP                                postoperative
                                                                   ERCP

                                      ┌──────────────┐
                                      │ Dissolution  │                              Yes  ┌──────────┐
         ┌───────────┐   ┌──────────┐ │   therapy    │         ┌──────────────┐        ──▶│   Pain   │
         │   Non-    │   │  Which   │ └──────────────┘         │    Stone     │           │  cured   │
         │ surgical  │   │ process? │ ┌──────────────┐         │ elimination  │           └──────────┘
         │ treatment │   └──────────┘ │  Shockwave   │         │   with no    │
         └───────────┘               │ lithotripsy  │         │ recurrence?  │      No   ┌──────────┐
         ┌───────────┐               └──────────────┘         └──────────────┘        ──▶│  Treat   │
         │    No     │               ┌──────────────┐                                     │  again   │
         │ treatment │               │    Other     │                                    └──────────┘
         └───────────┘               └──────────────┘
```

* Surgical removal of the gallbladder
** CBD stones are removed only if found
022 EF 137257/4

Exhibit 3

**SIMPLIFIED MODEL OF DIAGNOSIS,
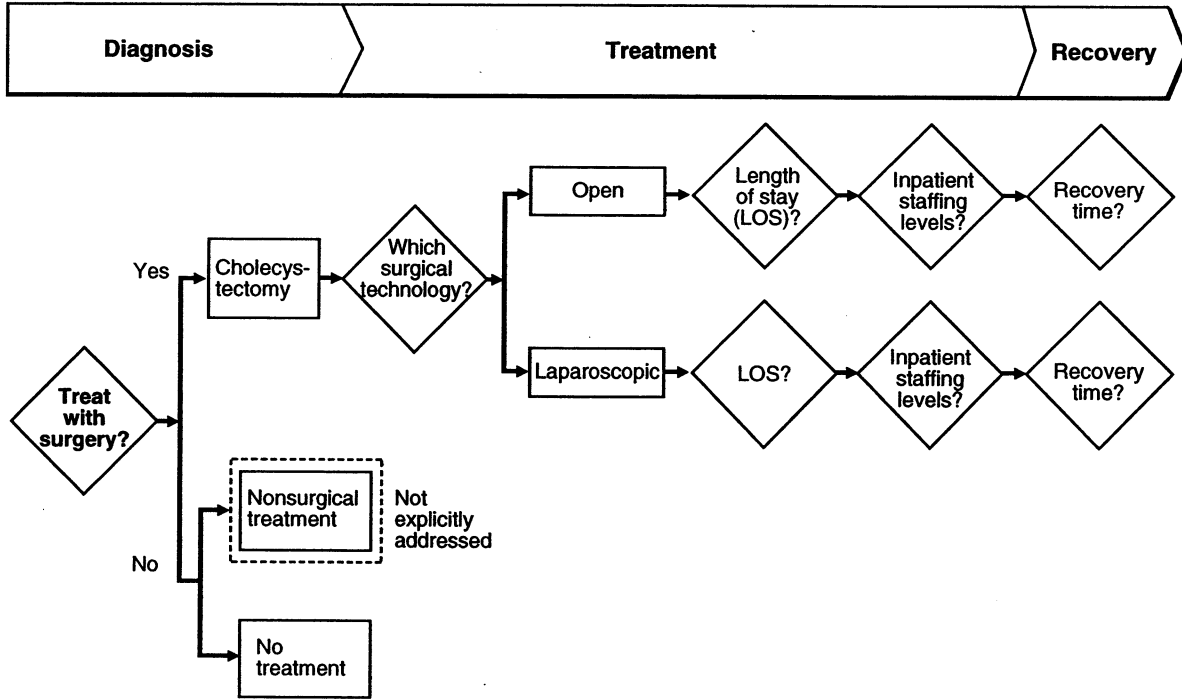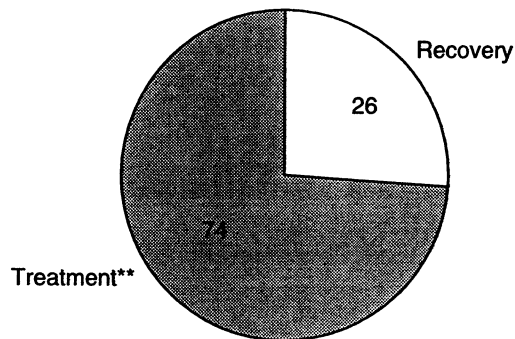TREATMENT, AND RECOVERY PHASE DECISIONS**

Exhibit 4

INPUTS BY PHASE*

Percent

* Average inputs consumed per phase in the U.S., U.K., and Germany
** Direct input usage of treatment planning for those cases receiving surgery are included in the treatment phase
Source: McKinsey analysis

022 ST 125964/4
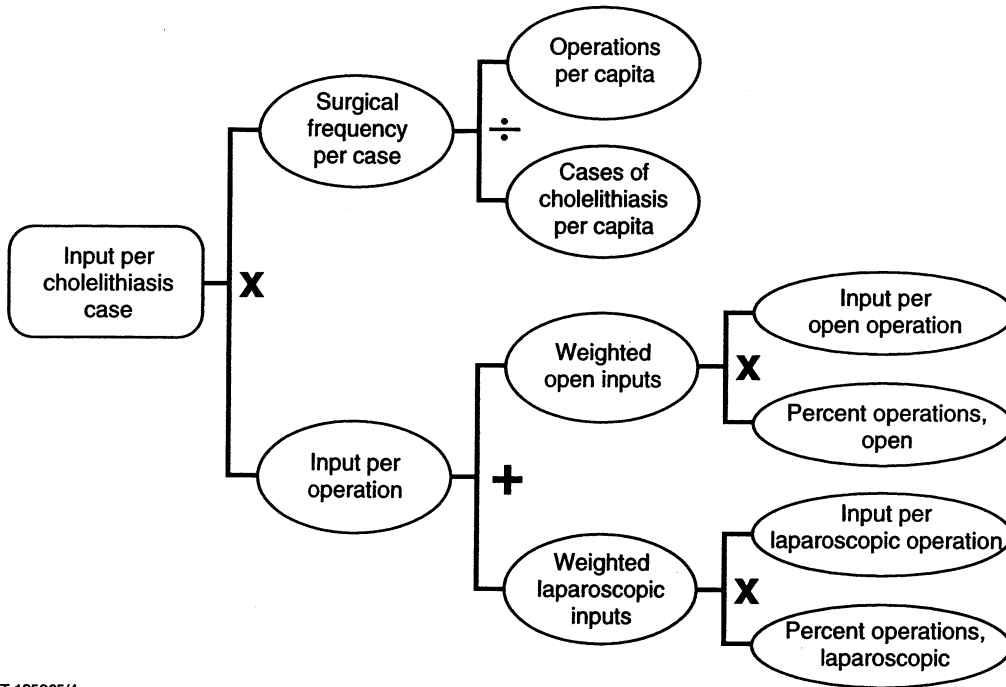
Exhibit 5

DERIVATION OF INPUT PER CASE OF CHOLELITHIASIS
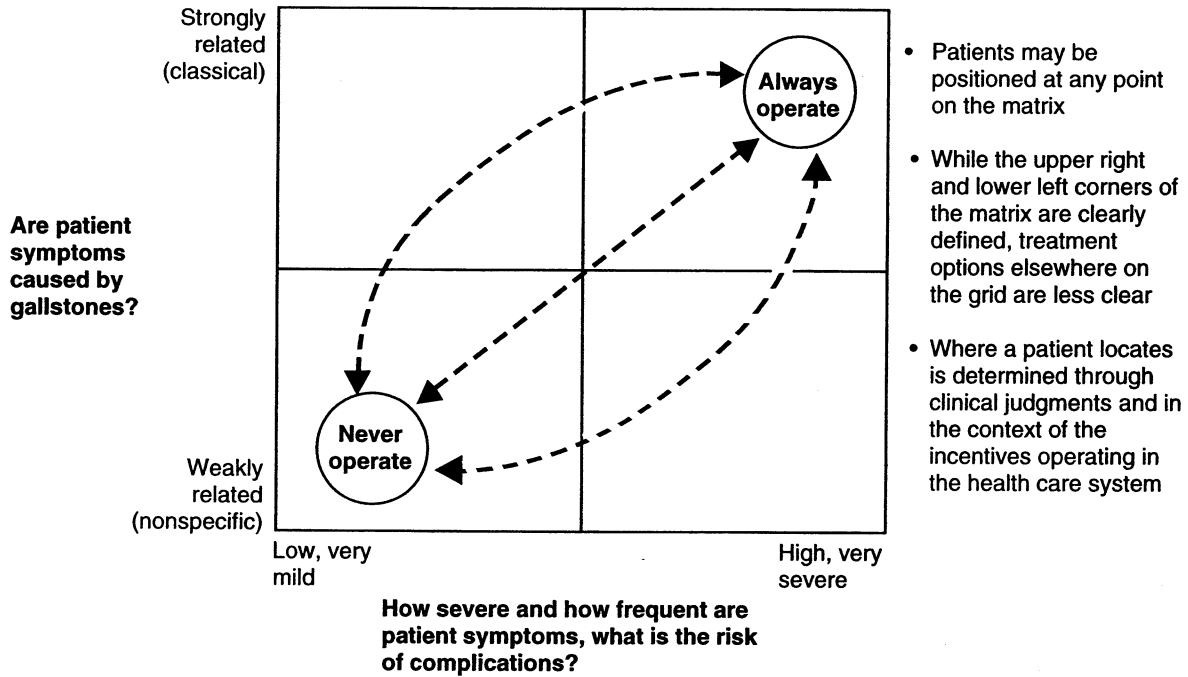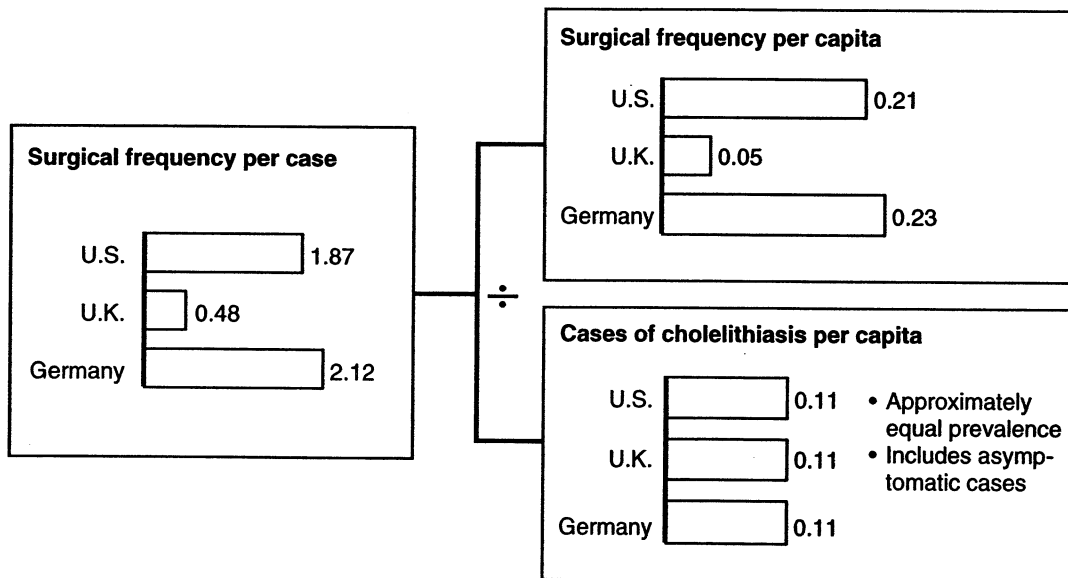
Exhibit 6

DIAGNOSIS AND TREATMENT DECISION MAKING PROCESS

Strongly related (classical)

Are patient symptoms caused by gallstones?

Weakly related (nonspecific)

Always operate

Never operate

Low, very mild

High, very severe

How severe and how frequent are patient symptoms, what is the risk of complications?

- Patients may be positioned at any point on the matrix

- While the upper right and lower left corners of the matrix are clearly defined, treatment options elsewhere on the grid are less clear

- Where a patient locates is determined through clinical judgments and in the context of the incentives operating in the health care system

Exhibit 7

SURGICAL FREQUENCY PER CASE OF CHOLELITHIASIS
Percent

Surgical frequency per case

U.S.	1.87
U.K.	0.48
Germany	2.12

÷

Surgical frequency per capita

U.S.	0.21
U.K.	0.05
Germany	0.23

Cases of cholelithiasis per capita

U.S.	0.11
U.K.	0.11
Germany	0.11

- Approximately equal prevalence
- Includes asymptomatic cases

Exhibit 8

CHOLECYSTECTOMY RATE AND IMPACT OF LAPAROSCOPY

Cholecystectomies per 1,000 population
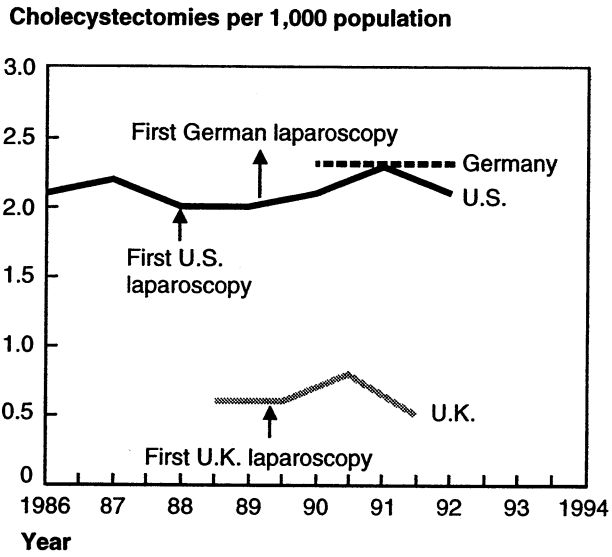
• Small increases (<10%) in the numbers of cholecystectomies are seen for the years following the adoption of laparoscopic techniques

• Increased national and regional surgical rates suggest that laparoscopy widened the indications for gallbladder removal

Exhibit 9

KEY DECISIONS DRIVING OBSERVED INPUT DIFFERENCES

Decision	Primary decision maker(s)	Relative impact on differences in input usage
Surgical frequency	Physician	High
Surgical technology	Physician	Medium
LOS	Physician/hospital	Low
Staffing levels	Hospital	Low
Recovery time	Physician/patient	Medium

022 EF 136240/4

Exhibit 10

INPUTS FOR TREATMENT OF CHOLELITHIASIS

Index, U.S. = 100%

Per case of cholelithiasis		Per cholecystectomy (operation)	
U.K.	44	U.K.	171
U.S.	100	U.S.	100
Germany	172	Germany	152

Per open cholecystectomy		Per laparoscopic cholecystectomy	
U.K.	107	U.K.	37
U.S.	100	U.S.	31
Germany	92	Germany	50
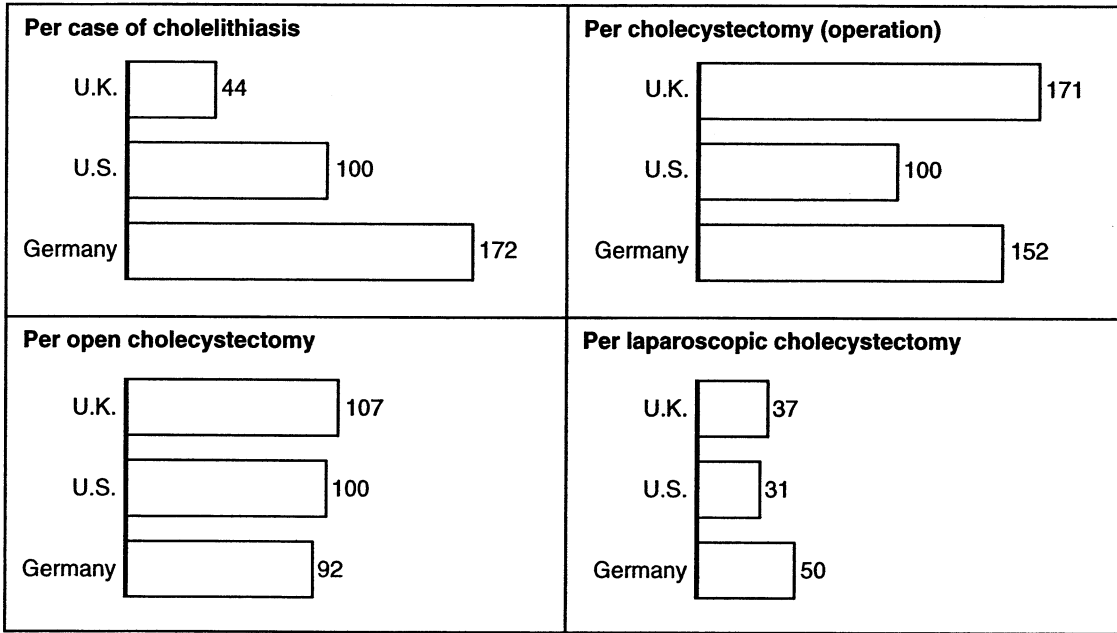
Exhibit 11

OUTPUTS FOR TREATMENT OF CHOLELITHIASIS

Improvement in outcomes due to treatment*

Index, U.S. = 100%

Per case of cholelithiasis		QALYs**	Per cholecystectomy (operation)	
U.K.	24	0.05	U.K.	100
U.S.	100	0.21	U.S.	100
Germany	110	0.23	Germany	100

Per open cholecystectomy		Per laparoscopic cholecystectomy	
U.K.	100	U.K.	100
U.S.	100	U.S.	100
Germany	100	Germany	100
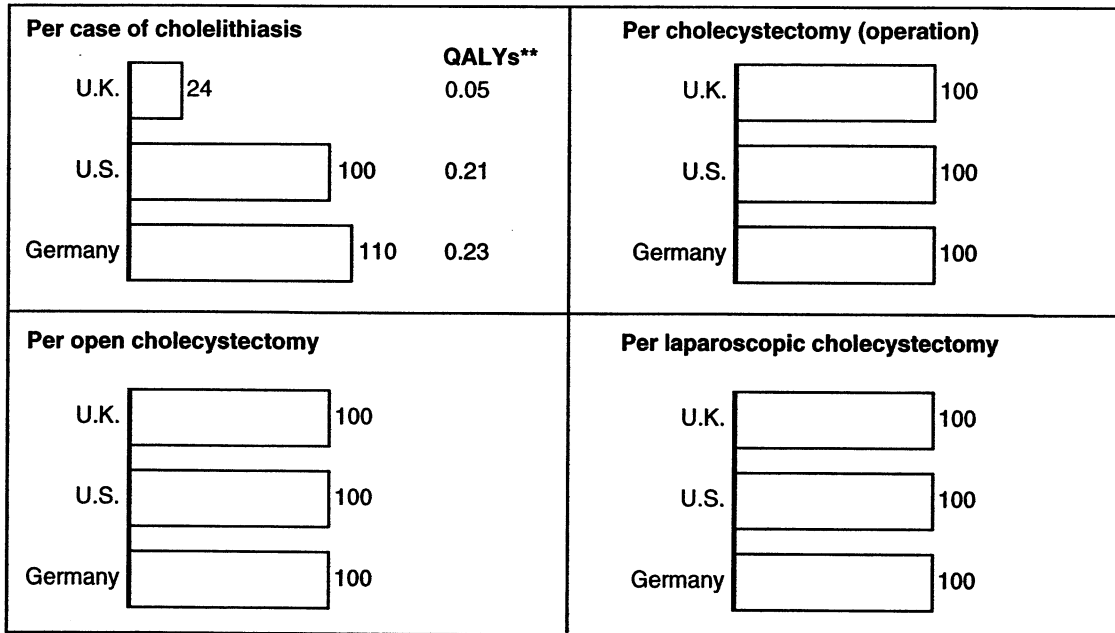
* Outcomes based on 30-day frequency of symptoms; relative to baseline outcome with no treatment
** QALYs calculated in analysis for outcomes per case of cholelithiasis; QALYs for other cases not required for analysis
Source: McKinsey analysis
022 ST 140005/4

Exhibit 12
INPUTS AND OUTCOMES PER CASE OF CHOLELITHIASIS

Outcomes per case
QALYs

Inputs per case
Standardized input units

	Inputs Index, U.S. = 100%	**Outcomes** Improvement in outcomes due to treatment Index, U.S. = 100%	**QALYs**
U.K.	44	24	0.05
U.S.	100	100	0.21
Germany	172	110	0.23
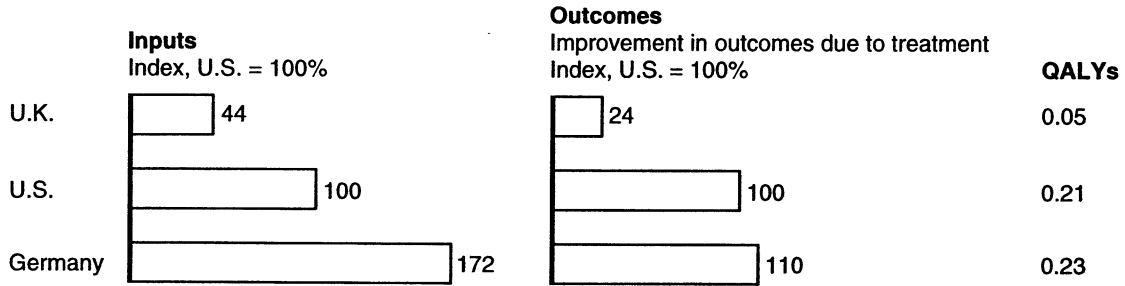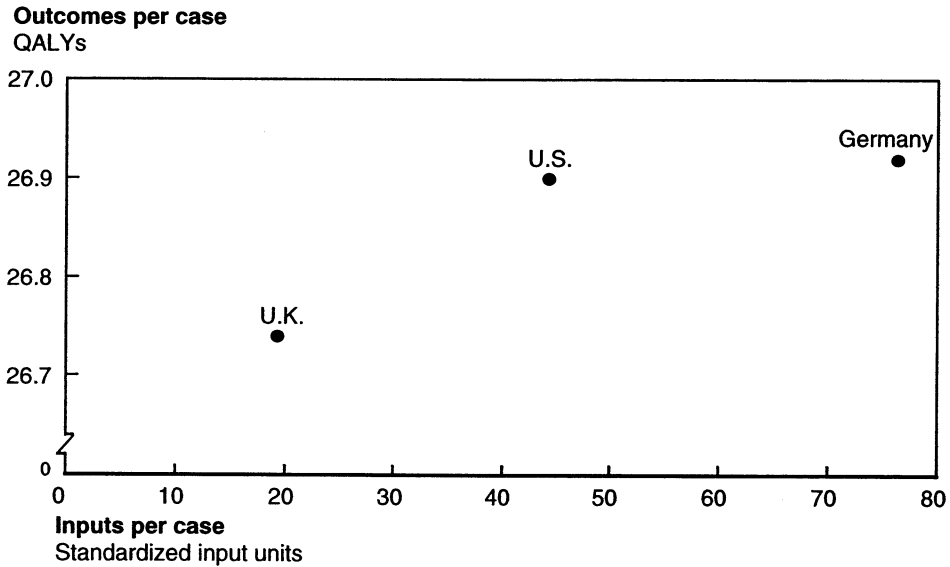
Note: Outcome measure assumes a frequency of symptoms of 30 days; baseline outcomes with no treatment estimated to be outcomes when no surgeries are performed

---

Exhibit 13
ASSESSMENT OF RELATIVE PRODUCTIVE EFFICIENCY

	Difference in inputs[1] Standardized input units	**Difference in outcomes**[1] QALYs[2]	**Average productivity advantage**[3] Percent	**Conclusion**
U.S. vs. U.K.	25	0.15	60 days / 14 days — 72 / 76	U.S. more productive
Germany vs. U.K.	56	0.18	14-60 days 52[4]	Germany more productive
U.S. vs. Germany	−32	−0.03	14 days / 60 days 13 16	U.S. more productive
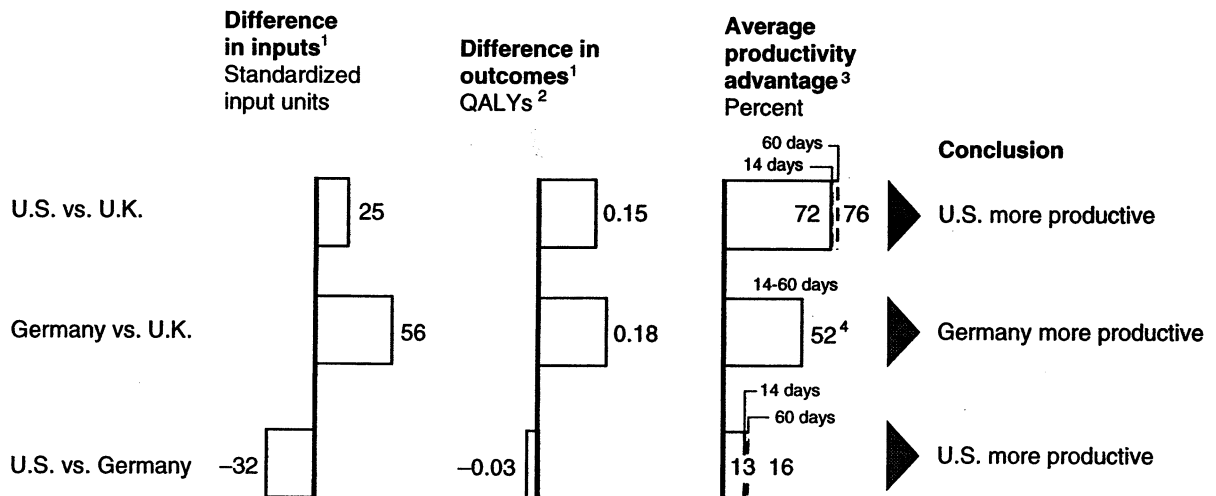
1 Difference between number of units in first country to those in second country
2 Calculated assuming 30-day frequency of symptoms
3 Average productivity advantage of first country over second
4 For the U.S. vs. Germany comparison, average productivity advantage is 52% for symptoms every 14 and 60 days

022 ST 140148/4

Exhibit 14

DIAGNOSIS PHASE DECISION AND IMPACT ON INPUTS

	Surgical frequency Percent	Impact on inputs per case (U.S. = 100%)
Cholecystectomy	(U.S.) 2.94	
	(U.K.) 0.75	(U.K.) –74
	(Ger) 3.33	(Ger) 2

Treat with surgery?
- Yes → Cholecystectomy
- No → Nonsurgical treatment
- → No treatment
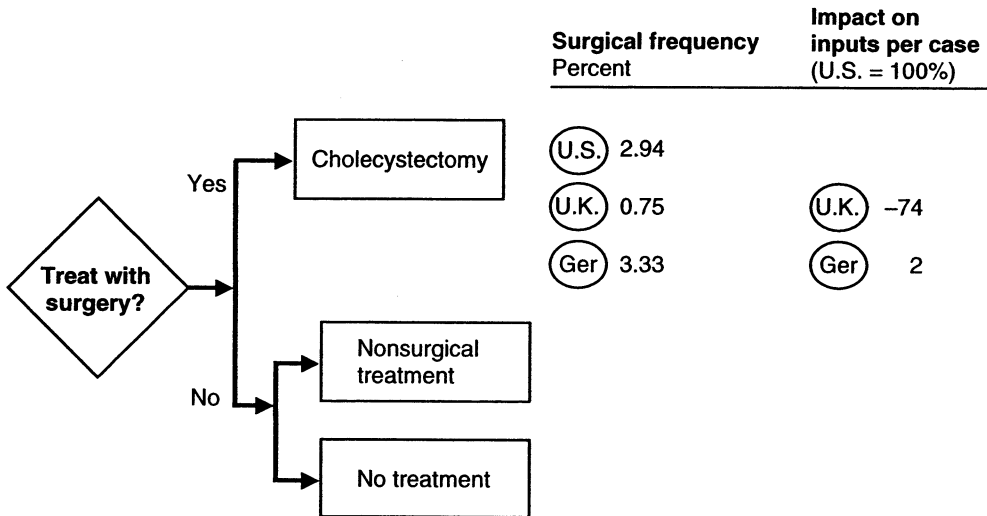
Source: National Hospital Discharge Survey (U.S.); Hospital Episode Statistics (U.K.); H.J. Krämling, *Der Chirurg*, 1993; articles; interviews
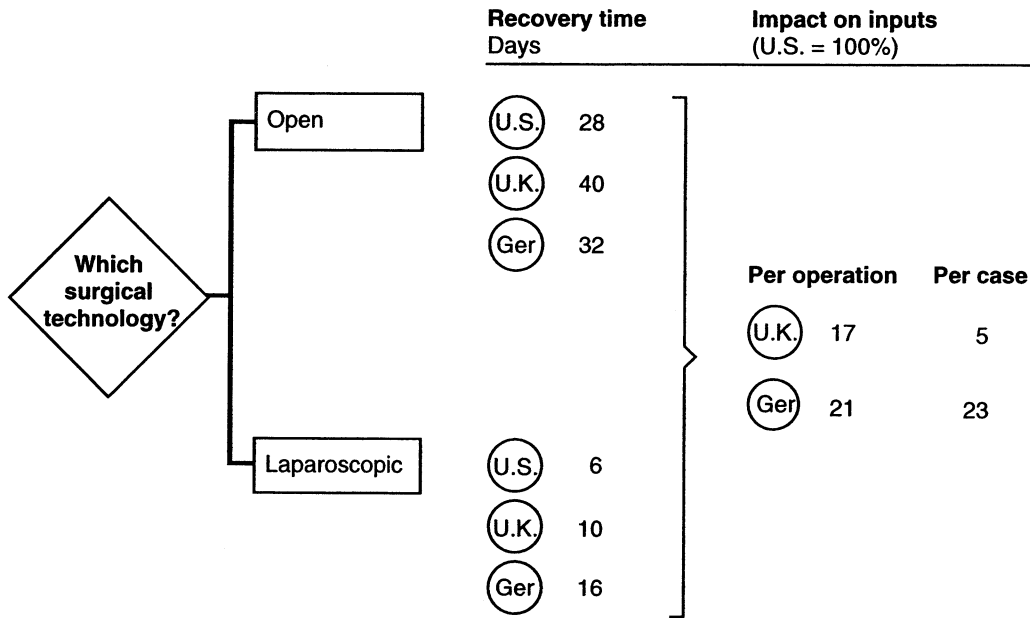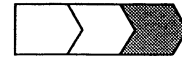

Exhibit 15

TREATMENT PHASE DECISION AND IMPACT ON INPUTS

	Percent of surgeries	LOS Days	Staffing levels* Standardized input units per bed-day
Open → LOS → Staffing levels?			
(U.S.)	10	6.77	0.76
(U.K.)	40	10.92	0.55
(Ger)	20	10.14	0.47
Laparoscopic → LOS → Staffing levels?			
(U.S.)	90	2.97	0.76
(U.K.)	60	3.80	0.55
(Ger)	80	5.85	0.47

Which surgical technology?

	Per operation	Per case	Per operation	Per case
Impact on inputs (U.S. = 100)	(U.K.) 53	14	(U.K.) –5	–5
	(Ger) 10	10	(Ger) –2	–2
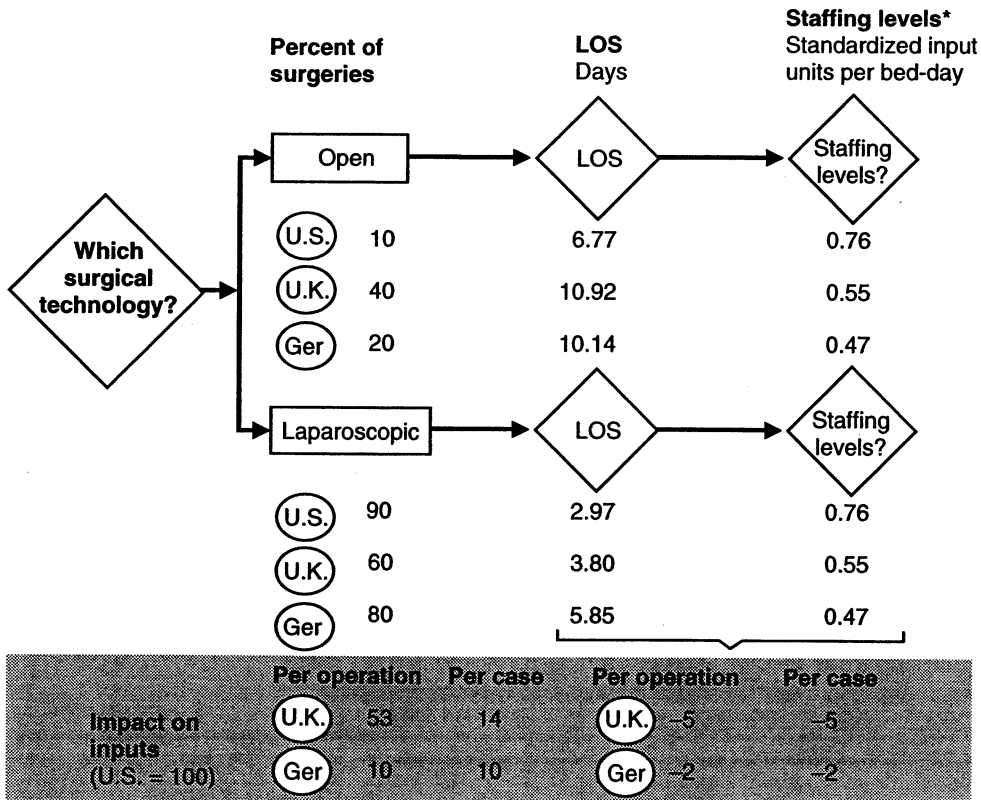
* Staffing levels given as a weighted sum of physician, nurse (qualified and nonqualified), and medical technician staffing levels

Source: CPHA; National Hospital Discharge Survey (U.S.); Royal College of Surgeons of England Confidential Audit; NIH Consensus Development Panel on Gallstones and Laparoscopic Cholecystectomy, 1993; articles; interviews; McKinsey analysis

022 CL 111921/4

Exhibit 16
RECOVERY TIME DECISION AND IMPACT ON INPUTS

	Recovery time Days	**Impact on inputs** (U.S. = 100%)

Open — (U.S.) 28, (U.K.) 40, (Ger) 32

Which surgical technology?

Laparoscopic — (U.S.) 6, (U.K.) 10, (Ger) 16

	Per operation	Per case
(U.K.)	17	5
(Ger)	21	23

Source: Annals of Royal College of Physicians; articles; interviews; McKinsey analysis

---

Exhibit 17

RATE OF ADOPTION OF LAPAROSCOPIC TECHNOLOGY
Percent

Percent of total cholecystectomy procedures that were laparoscopic

U.S.
Germany
U.K.
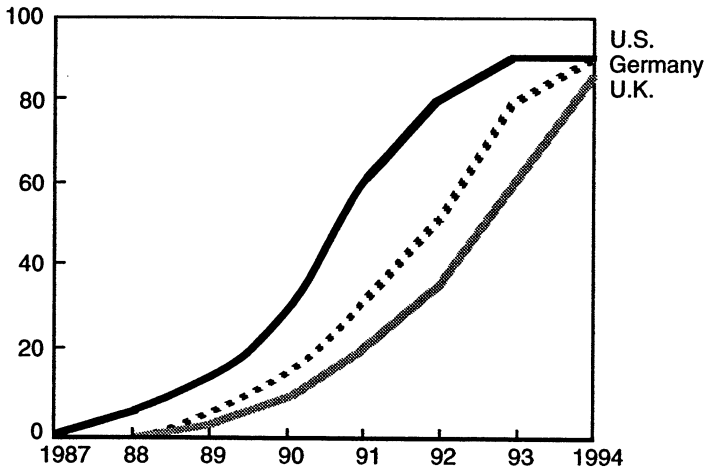
- Rapid uptake of new technology led by U.S.

- Penetration rate in U.K. and Germany was faster than is typical for new technology, in part because adoption has been spurred by endoscopic equipment manufacturers who operate training centers in the laparoscopic technique

- Penetration of laparoscopic cholecystectomy estimated for years 1988-91

Source: R. Orlando III et al., 1993 (U.S.); NIH Consensus Development Panel on Gallstones and Laparoscopic Cholecystectomy, 1993; R. McCloy, 1992 (U.K.); R.C.G. Russell, 1993 (U.K.); industry interviews (U.K.); H.J. Krämling et al., 1993 (Germany); clinician interviews (Germany)

Exhibit 18

SOURCES OF DIFFERENCE IN U.S. AND U.K.
INPUTS PER CASE – CHOLELITHIASIS

Percent of difference

```
100        -74*
                                                                    4          44
                        14         -5*        5
                                                                  Other

           Surgical     Open/      LOS/staffing  Recovery
           frequency    laparoscopic levels      time
                        mix
U.S.                                                                         U.K.
```
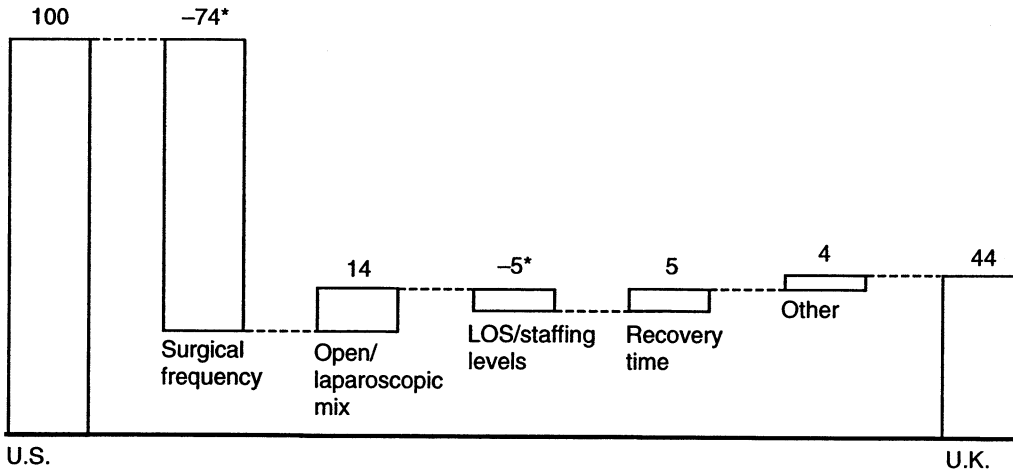
 * Impact on inputs equals difference in U.S. inputs if U.S. uses U.K. level for this variable. All others calculated as the difference
   in U.K. inputs if U.K. uses U.S. levels for the variables
Source: McKinsey analysis

Exhibit 19

SOURCES OF DIFFERENCE IN U.S. AND GERMANY
INPUTS PER CASE – CHOLELITHIASIS

Percent of difference

```
                                                                  21          172

                                                       23
                        10         -2*
                                                                Other
            20
                        Open/                  Recovery
                        laparoscopic LOS/staffing time
 100        Surgical    mix          levels
            frequency

U.S.                                                                      Germany
```
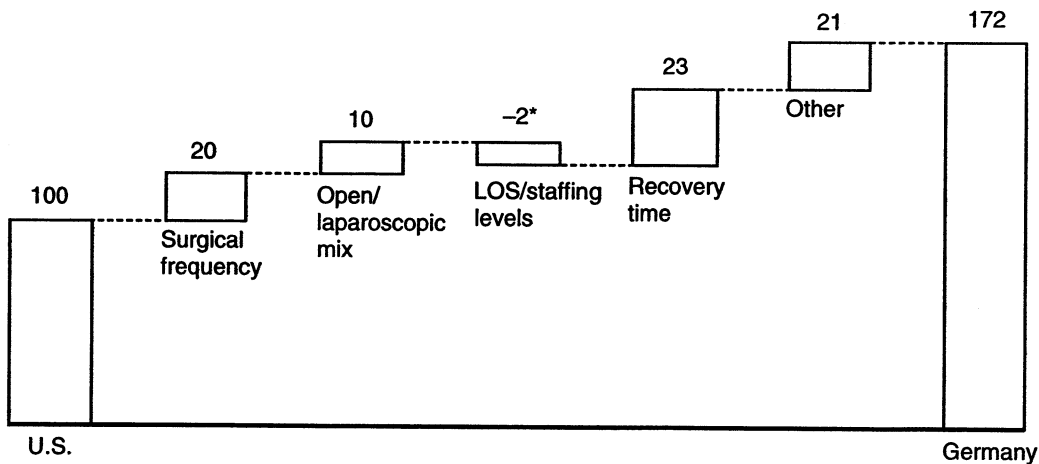
 * Impact on inputs equals difference in U.S. inputs if U.S. uses German staffing levels. All others calculated as the difference
   in German inputs if Germany uses U.S. levels for the variables
Source: McKinsey analysis
022 SL 128922/4

Exhibit 20

SOURCES OF DIFFERENCE IN U.S. AND U.K.
INPUTS PER OPERATION – CHOLELITHIASIS
Percent of difference

Exhibit 21

SOURCES OF DIFFERENCE IN U.S. AND GERMANY
INPUTS PER OPERATION – CHOLELITHIASIS
Percent of difference

Exhibit 22

IMPACT OF PROVIDER BEHAVIOR
DIFFERENCES ON PRODUCTIVE EFFICIENCY

● Major differentiating factor
◑ Secondary differentiating factor
○ Undifferentiating

Provider behaviors	U.S. vs. U.K.	U.S. vs. Germany
Care triaging	○	○
Treatment duration	◑ U.K. longer LOS and recovery time than U.S.	● Germany longer LOS and recovery time than U.S.
Staffing levels	◑ U.K. lower staffing than U.S.	◑ Germany lower staffing than U.S.
Setting choice	○	○
Team-based approach	○	○
Technology adoption	● U.K. adopted laparoscopy later than U.S.	◑ Germany adopted laparoscopy slightly later than U.S.

022 ST 125826/4

Exhibit 23

CAUSAL ANALYSIS: INCENTIVES AND CONSTRAINTS
U.S. vs. U.K.

● Major differentiating factor
◐ Secondary differentiating factor
○ Undifferentiating

	Extent of causal impact	Provider behaviors		
		Treatment duration	Staffing levels	Technology adoption
Incentives				
Physician	●			U.S. physicians had incentive to adopt laparoscopy
Hospital	◐	U.S. hospitals had incentive to manage LOS	U.S. hospitals had incentive to keep LOS low, requiring higher staffing	U.S. hospitals had incentive to adopt laparoscopy
Constraints				
Physician supply	○			
Hospital supply	◐			U.K. feared increased use of hospitals with laparoscopy
Capital	◐			U.K. capital constraints precluded broad adoption of laparoscopy technology
Substitution	○			

Exhibit 24

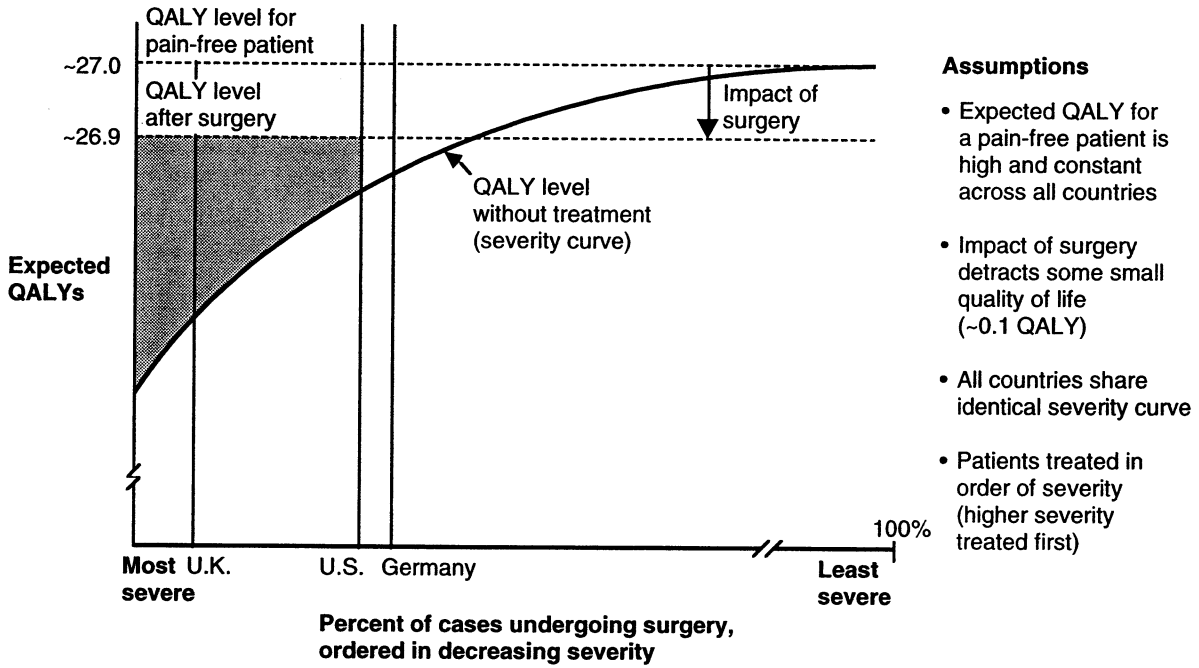## CAUSAL ANALYSIS: SYSTEM STRUCTURE AND REGULATION
U.S. vs. U.K.

● Major differentiating factor
◐ Secondary differentiating factor
○ Undifferentiating

	Extent of causal impact	Provider incentives and constraints			
		Physician incentives	Hospital incentives	Hospital supply	Capital constraints
Health care system structure					
Product integration and pricing mechanisms	◐	U.S. FFS vs. U.K. salary for specialist physicians encouraged fast adoption of laparoscopy and more surgeries	U.K. annual hospital budget encouraged "bed-blocking" and therefore longer LOS; hospitals feared laparoscopy would take resources from other diseases; U.S. hospital case rates for Medicare encouraged shorter LOS		
Competitive intensity	●	U.S. physician competition for patients and payor competition for members encouraged adoption and coverage for laparoscopy	U.S. hospital competition for patients and payor competition for members encouraged adoption and coverage for laparoscopy U.S. payor price-based competition for members put pressure on hospitals to manage LOS		
Regulation	●	Threat of malpractice suits for U.S. physicians encouraged more surgeries More generous disability policies in U.K. encouraged longer recovery times	U.K. hospital budgets controlled and set through NHS global budget	U.K. hospital supply strictly controlled	U.K. capital supply strictly controlled

Exhibit 25

SURGICAL FREQUENCY COMPARISON

Percent of cholelithiasis cases

	1.87	0.72	0.48
	U.S. average	U.S. group model HMO example	U.K.

Less selective practices ← → **More selective practices**

Exhibit 26

CAUSAL ANALYSIS: INCENTIVES AND CONSTRAINTS
U.S. vs. Germany

● Major differentiating factor
◐ Secondary differentiating factor
○ Undifferentiating

	Extent of causal impact	Provider behaviors		
		Treatment duration	Staffing levels	Technology adoption
Incentives				
Physician	◐	German physicians had incentive to keep LOS long to occupy hospital beds		
Hospital	●	U.S. hospitals had incentive to manage LOS; German hospitals had incentive to occupy hospital beds	U.S. hospitals had incentive to keep LOS low, requiring higher staffing	German hospitals had incentive to keep longer LOS, so did not adopt laparoscopy as quickly
Constraints				
Physician supply	○			
Hospital supply	●	High number of German hospital beds led to long LOS		
Capital	○			
Substitution	○			

Exhibit 27

CAUSAL ANALYSIS: SYSTEM STRUCTURE AND REGULATION
U.S. vs. Germany

● Major differentiating factor
◐ Secondary differentiating factor
○ Undifferentiating

	Extent of causal impact	Provider incentives and constraints		
		Physician incentives	Hospital incentives	Hospital supply
Health care system structure				
Product integration and pricing mechanisms	●		German per diem vs. U.S. case rate product encouraged longer LOS	
Competitive intensity	◐		U.S. payor price-based competition for members and ability to negotiate differentially with individual hospitals encouraged hospital cost control; German payors had relatively restricted memberships along geographic or occupational lines and could not negotiate differentially with hospitals for price or other contract terms	
Regulation	●	German department chiefs could add private bed capacity if kept public bed capacity high	German hospitals feared capacity cuts if utilization fell below 85%; long recovery time mandated by German law	German hospital supply regulated through state government with political motives

Exhibit A-1

## IDEAL OUTCOME MEASURE: CHANGE
IN QALYs PROVIDED BY TREATMENT

Assumptions

- Expected QALY for a pain-free patient is high and constant across all countries

- Impact of surgery detracts some small quality of life (~0.1 QALY)

- All countries share identical severity curve

- Patients treated in order of severity (higher severity treated first)

Percent of cases undergoing surgery, ordered in decreasing severity

Note: As an example, the U.S. change in expected QALYs is shown as the shaded area

---

Exhibit A-2

SHAPE OF SEVERITY CURVE AFFECTS OUTCOME ORDERING

QALYs
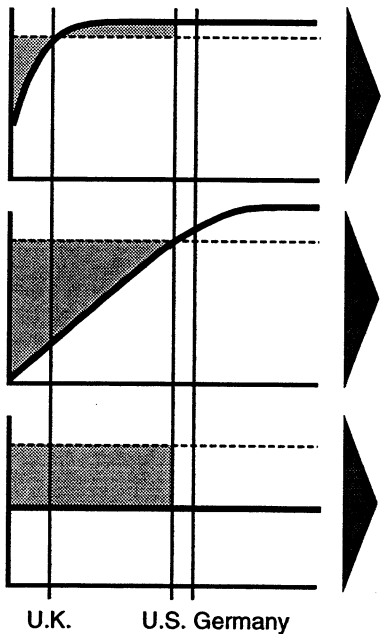
(A) Very few cases require surgery
  - The U.K. rations appropriately

(B) Many cases require surgery, but not all
  - The U.S. does just enough surgeries (Germany too many; the U.K. could do more)

(C) All cases receive the same benefit from surgery
  - Germany reaches the highest outcome

Outcomes per case

Inputs per case
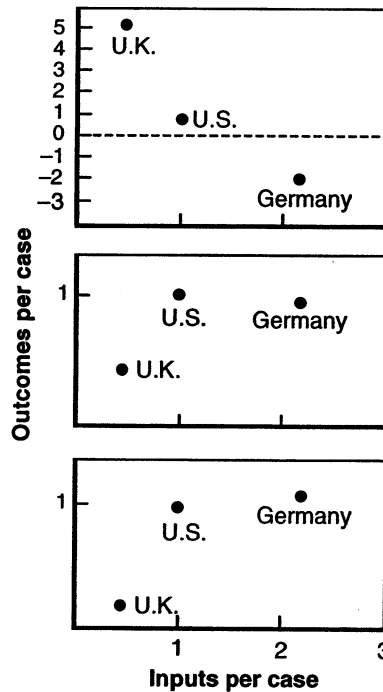
Exhibit A-3

MODEL OF QUALITY OF LIFE
OUTCOMES FOR CHOLELITHIASIS

Not included
in analysis

Treatment groups   **Health state**

Emergency
surgery

Elective
surgery

**Cholelithiasis
cases**

Borderline
patients

No surgery

Each health state is
represented by a Kaplan-
Bush Index of Well-Being

The Kaplan-Bush Index of
Well-Being is weighted by
• Number of patients
  in the health state
• Number of days in
  the health state

The weighted indices
are summed to yield
total outcome

022 SL 128924/4

Exhibit A-4
PATIENT SYMPTOM GROUPS

Functional states	Symptom states	Patient group, phase	Kaplan-Bush index	Number of patients*
No functional impairments	Pain, fever, vomiting	Emergency, presurgery	0.7433 - 0.1104 = 0.6329	10% of surgeries
No functional impairments	Pain and vomiting	Elective, presurgery	0.7433 - 0.0317 = 0.7116	20% of surgeries
No functional impairments	Pain only	Elective, presurgery	0.7433 - 0.0382 = 0.7051	50% of surgeries**
No functional impairments	Pain and diet or medication	Elective, presurgery	0.7433 +0.0742 = 0.8175	20% of surgeries
No functional impairments	No pain	All surgeries, presurgery	0.7433 + 0.2567 = 1.00	100% of surgeries
Work, mobility limited	Generally tired, weak	All surgeries, during surgery	0.5250 - 0.0027 = 0.5223	100% of surgeries
No functional impairments	No pain	All surgeries, postsurgery	0.7433 + 0.2567 = 1.00	100% of surgeries
No functional impairments	Pain only	Borderline patients	0.7433 - 0.0382 = 0.7051	70% of borderlines
No functional impairments	Pain and diet or medication	Borderline patients	0.7433 +0.0742 = 0.8175	30% of borderlines
No functional impairments	No pain	Borderline patients	0.7433 + 0.2567 = 1.00	100% of borderlines

* Actual percentages of patients in each symptom group were found by weighting the percentage in this column by the percent of surgery on borderline patients. With surgery + borderline population equal to 100%, the surgery percentages were: 88% for the U.S., 23% for the U.K., and 100% for Germany

** For Germany, the percent of surgery patients having pain-only was 44%, not 50%, with 5% of surgery patients having no pain at all times

---

Exhibit A-5
QUANTITATIVE OUTCOMES

022 AR 130396/4
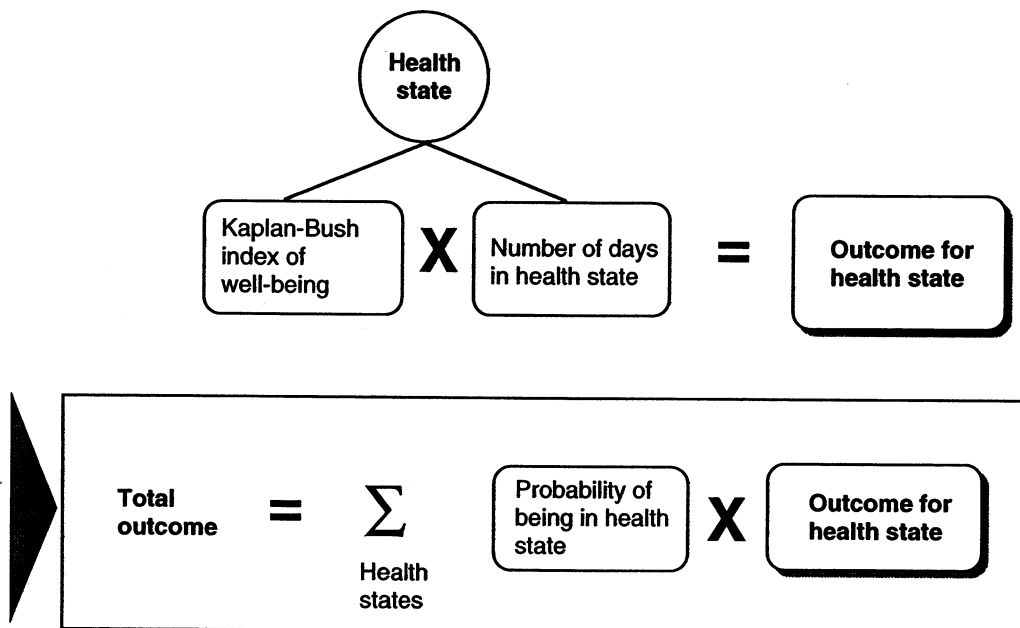
Exhibit A-6

WAIT TIMES FOR CHOLECYSTECTOMY
Days from decision-to-admit until admission

Emergency
Elective

U.S.
1
7

U.K.
3
53

Germany
1
7
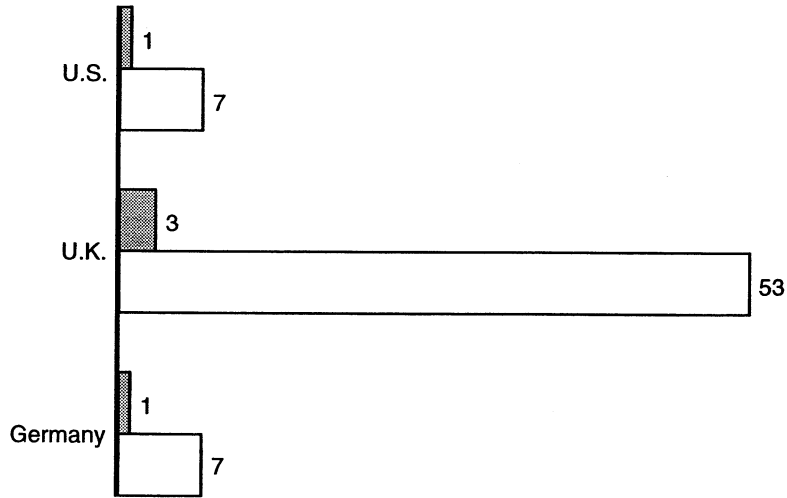
Source: Clinical reviews; Hospital Episode Statistics (U.K.)

Exhibit A-7

CASES WITH COMPLICATED INDICATIONS
Percent of total operations

Range of estimates from sources

Not included in average

Weighted average

U.S.
14%
≥ 65 years old study

U.K.
19%

Germany
15%

0   10   20   30   40   50
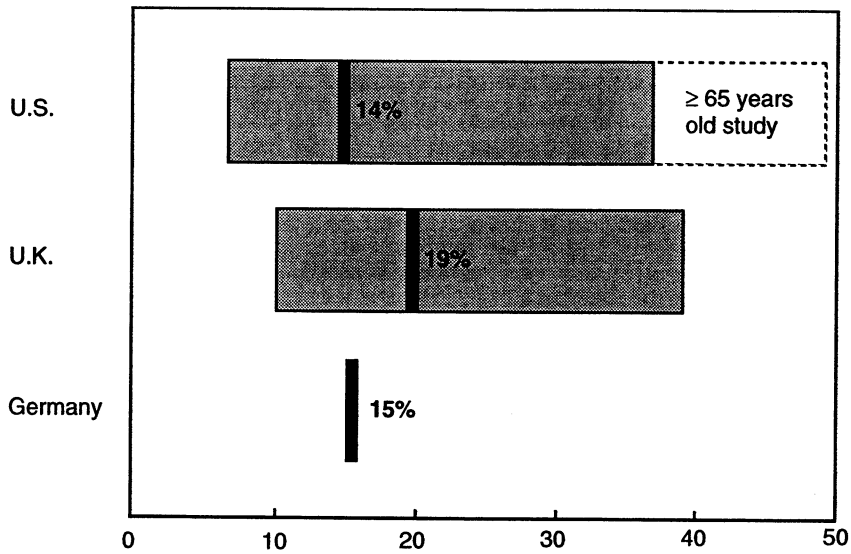
Source: Articles (U.S., U.K.); clinical reviews (U.S., U.K., Germany)
022 AR 130397/4

Exhibit B-1

PREVALENCE ESTIMATES

Reference	Estimate	Country	Article
1	10-15% of adults 8-12% of population	U.S.	*Gallstones and Laparoscopic Cholecystectomy*, Consensus Development Panel, National Institutes of Health (1993), as cited in *JAMA*, Vol. 269 (1993): 1018.
2	7-9%	U.S.	*Digestive Diseases Statistics*, NIH.
3	15-20% of adults 10% of population	U.S.	G.J. Farha and R.L. Beamer, "New Options for Treating Gallstone Disease of Adults," *American Family Physician*, Vol. 44 (1991): 1295.
4	11.6%	U.S.	M.M. Lieber, "The Incidence of Gallstones and Their Correlation with Other Diseases," *Ann. Surg.*,Vol. 135 (1952): 394.
5	5-7%	U.S. and U.K.	G.P. Sadler, A. Shandall, and B.I. Reed, "Laparoscopic Cholecystectomy (Review)," *Br. J. of Hospital Medicine*, Vol. 48 (1992): 462.
6	5-20%	W. Europe	V. Pearson, *Minimal Access Surgery - A Review*, Health Care Evaluation Unit, Univ. of Bristol Dept. Epidemiology & Public Health Medicine (1994).
7	7% men 12-14% women	Europe	V.A. Holstege and E.J. Kohlberger, "Epidemiologie und Pathogenese von Gallenstein, Kenntnisstand und therapeutische Konsequenzen," *Fortschr. Med.*, Vol. 107 (1989): 673.
8	12% (3:1, women:men)	Europe	Herold, 1990.

Exhibit B-2

PREVALENCE STUDIES

Reference	Country/region	Article
9	U.K./Oxford	F. Pixley, D. Wilson, K. McPherson, and J. Mann, "Effect of Vegetarianism on Development of Gallstones in Women," *Br. Med. Journal*, Vol. 291 (1985): 11.
10	U.K./Bristol	K.W. Heaton, F.E.M. Braddon, R.A. Mountford, A.O. Hughes, and P.M. Emmett, "Symptomatic and Silent Gallstones in the Community," *Gut*, Vol. 32 (1992): 316.
11	Germany/Düsseldorf	P. Peter, A. Zeller, and A. Sonnenberg, "Haeufigkeit Asymptomatischer Gallenstiene bei Krankenhauspatienten," *Z. Gastroenterologie*, Vol. 23 (1985): 573.
12	Germany	N.D. Nurnberg, H. Pannwitz, and H. Berndt, "Neue Erkenntnisse zur Epidemiologie des Schwedt, Gallensteinleidens durch die Sonographie," *Z. Aerztl. Fortbild.*, Vol. 85 (1991): 747. H. Berndt, D. Nurnberg, and H. Pannwitz, "Epidemiologie der Gallenblasensteine bei Maennern," *Z. Aerztl. Fortbild.*, Vol. 85 (1991): 1201.

WOMEN (OXFORD/PIXLEY)

GS = Gallstones
CY = Previous cholecystectomy

Age Years	Number in group	Number with GS or CY (Percent)
40-44	134	21 (16)
45-49	92	21 (23)
50-54	115	30 (26)
55-59	110	22 (20)
60-64	101	32 (31)
65-69	80	30 (38)

Exhibit B-4

TOTAL PREVALENCE OF CHOLELITHIASIS BY AGE – WOMEN
Range of 95% confidence interval

High — Pixley — Low   High — Nurnberg — Low   High — Peter — Low   High — Heaton — Low
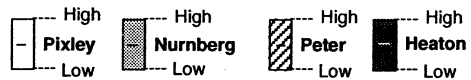
Percent gallstones + percent cholecystectomy
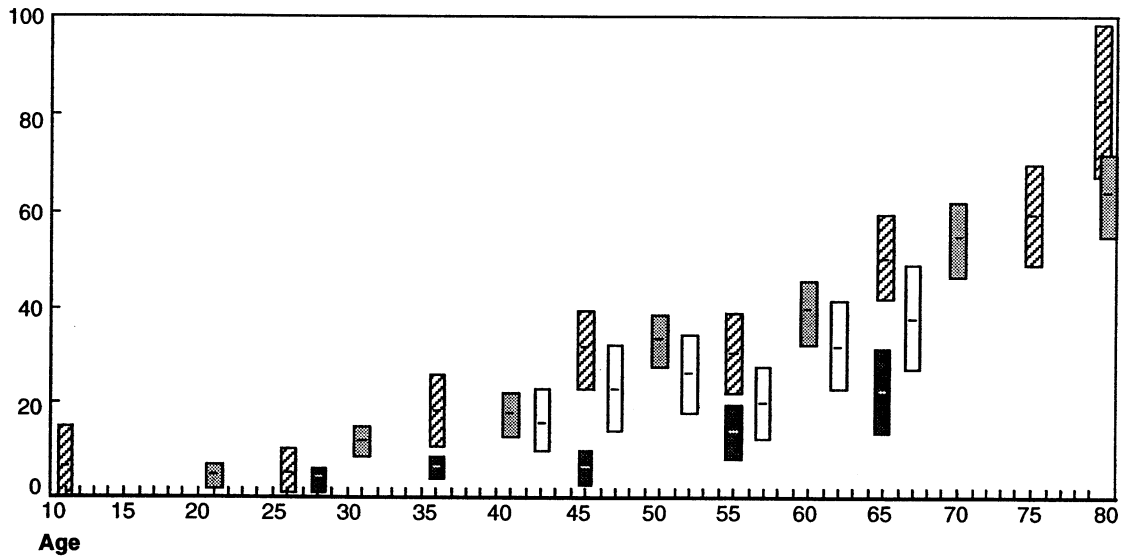
Source: F. Pixley et al., 1985 (U.K.); D. Nurnberg et al., 1991 (Germany); P. Peter et al., 1985 (Germany); K.W. Heaton et al., 1992 (U.K.)
022 RE 122889/4

Exhibit B-5

RESULTS OF BRISTOL/HEATON STUDY

GS = Gallstones
CY = Previous cholecystectomy

Age Years	Number in group	Percent of men with condition			Number in group	Percent of women with condition		
		GS	CY	GS or CY		GS	CY	GS or CY
25-29	n/a	n/a	n/a	n/a	305	3	1	4
30-39	n/a	n/a	n/a	n/a	328	3	3	6
40-49	430	4	1	5	199	4	3	7
50-59	226	5	2	7	141	8	6	14
60-69	182	9	3	12	85	12	11	22

Exhibit B-6

TOTAL PREVALENCE OF CHOLELITHIASIS BY AGE – MEN
Range of 95% confidence interval

High Berndt Low
High Peter Low
High Heaton Low

Percent gallstones + percent cholecystectomy
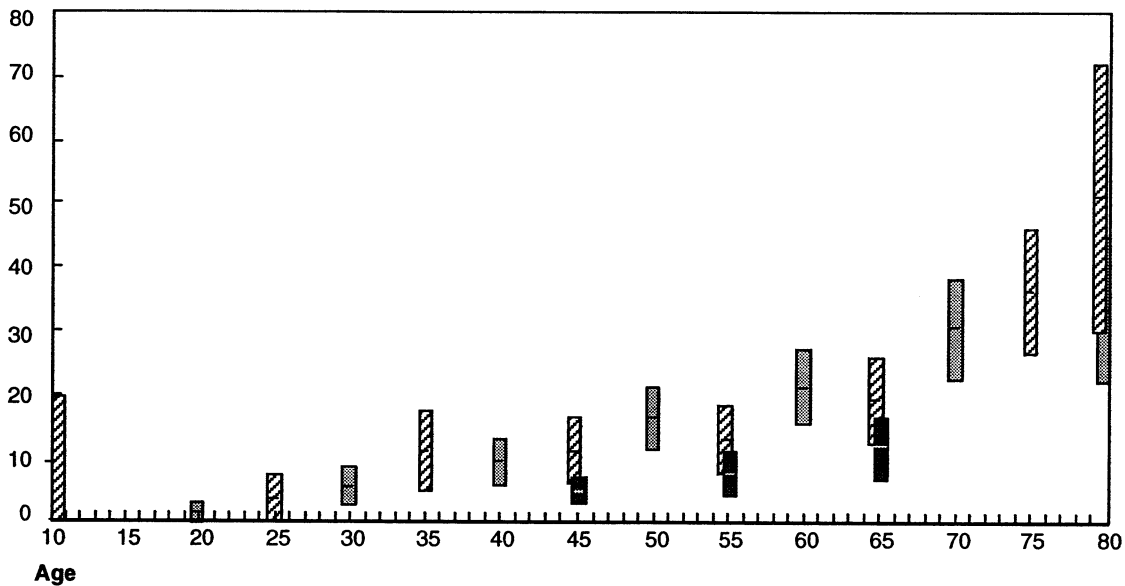
Source: H. Berndt et al., 1991 (Germany); P. Peter et al., 1985 (Germany); K.W. Heaton et al., 1992 (U.K.)
022 RE 122890/4

Exhibit B-7

RESULTS OF DÜSSELDORF/PETER STUDY

GS = Gallstones
CY = Previous cholecystectomy

Age Years	Percent of study participants	Percent of men with condition			Percent of women with condition		
		GS	CY	GS or CY	GS	CY	GS or CY
0-19	4	6	3	9	3	3	6
20-29	9	3	0	3	5	0	5
30-39	14	7	4	11	11	7	18
40-49	19	9	2	11	21	10	31
50-59	20	5	8	13	17	13	30
60-69	18	15	4	19	30	20	50
70-79	13	25	11	36	41	18	59
80+	3	51	0	51	65	18	83

022 EF 136266/4

Exhibit B-8

RESULTS OF SCHWEDT & NEURUPPIN/ NURNBERG & BERNDT STUDY

GS = Gallstones
CY = Previous cholecystectomy

Men

Age Years	Schwedt Number in group	Percent with GS or CY	Neuruppin Number in group	Percent with GS or CY
0-14	4	0	13	0
15-24	162	2	88	0
25-34	84	10	161	3
35-44	101	10	145	9
45-54	107	10	134	16
55-64	92	21	94	21
65-74	105	29	24	38
75+	45	24	24	58

Women

Age Years	Schwedt Number in group	Percent with GS or CY	Neuruppin Number in group	Percent with GS or CY
0-14	11	0	14	0
15-24	130	5	269	5
25-34	104	12	29	12
35-44	101	16	144	19
45-54	110	24	173	39
55-64	101	31	100	48
65-74	91	42	67	72
75+	52	54	77	70

022 EF 136267/4

Exhibit C-1

POPULATION IN THE U.S., U.K., AND GERMANY

Country	Population	Source
U.S.	255,462,000	Statistical Abstract of the United States, 1993.
U.K.	48,378,000	Office of Population Census and Surveys: England only.
Germany	81,492,000	European Marketing Data and Statistics, 1996

Exhibit C-2

NUMBER OF CHOLECYSTECTOMIES IN THE U.S., U.K., AND GERMANY

Country	Number of cholecystectomies	Source
U.S.	525,000	Table 22, National Hospital Discharge Data, Centers for Disease Control (CDC).
U.K.	25,463	Hospital Episode Statistics (HES) England only.
Germany	190,000	Krämling et al., 1993.

Other German estimates

170,000	Industry interviews.
111,510	Statistisches Bundesamt, Federal Office of Statistics.
60,000-100,000	Clinician interviews.

Chapter 5: Breast cancer case

This chapter discusses the relative productive efficiency of Germany, the U.K., and the U.S. in the treatment of breast cancer.

We begin with an overview of the disease, a discussion of the productive efficiency measure used, and a description of the treatment process. After assessing the relative productive efficiency of these countries in the treatment of breast cancer, we analyze the provider behaviors driving these productive efficiency differences. Finally, we discuss how different health care system structures and regulatory environments affected provider incentives and constraints and, therefore, productive efficiency.

BRIEF DISEASE OVERVIEW

Breast cancer is at or near the top cause of cancer-related deaths for women in the U.S., U.K., and Germany. It is estimated that in each of these three countries, between 55 and 90 cases per 100,000 women are diagnosed annually.[1] This incidence translates to a lifetime risk of disease on the order of 10 percent.

Female breast cancer is a disease that rarely occurs in women under the age of 30, with the majority of cases occurring in women 50 years or older. It is a life-threatening condition characterized by abnormal cell growth in the breast tissue which can spread to other distant tissue sites if left untreated. Currently, there is no definitive cure for the disease, and there are no simple preventive steps that dramatically reduce individual risk (e.g., such as smoking cessation for lung cancer). It is widely believed that the best response to the threat of the disease is its early detection and treatment. Once detected, breast cancer is treated and patients are monitored for recurrence and treated again if necessary.

Given the many options available in the management and treatment of breast cancer, there are significant differences in the delivery of care to breast cancer patients. In detection, there are choices with respect to whether to screen patients, the breadth of the population screened, the frequency of the screening, and the technology utilized. Once cases are identified through screening or the presence of symptoms, there are a number of procedures that can be used to

---

[1]   Sources: Surveillance Epidemiology and End Results (SEER); European Journal of Cancer, 1990.

assess suspect cases. Even after a definitive diagnosis, there are options as to what type of care is delivered and patient follow-up subsequent to treatment. The choices made by health care systems and providers in each of these areas have significant consequences for the outcomes achieved and resources used in breast cancer management and treatment.

DEFINITION OF PRODUCTIVE EFFICIENCY MEASURE

Timeframe of analysis

Our analysis of breast cancer focuses on treatment practices in roughly the mid to late 1980s timeframe. Because there are no national, population-based data on treatment and outcomes, we combined data from several different time periods and subpopulations to build an aggregate picture of breast cancer management and outcomes in each country. By piecing together information from many different sources, we gained a reasonable understanding of what constituted "typical" care for breast cancer patients in each country. (See Appendix 5B for a description of the major sources used.) As there were substantial variations in practice patterns within each country and the treatment of breast cancer changed in important ways from the early 1980s until the present, our results represent a comparison of "snapshots" in time of the average care practiced in each country, rather than a controlled comparison of static practice patterns.

Although the analysis presented here focuses on care in the 1980s, we also identify some changes in patterns of care that have occurred since that time. Where we identified such changes, we have discussed them and their implications for the causal analysis of the underlying drivers of provider behavior and resulting productive efficiency.

Summary of disease management and treatment phases

We divided the management and treatment of breast cancer into four phases (**Exhibit 1**): 1) screening; 2) assessment; 3) therapeutic; and 4) follow-up. In the screening phase, patients are tested for the presence of cancer. In order to be considered a screening examination, diagnostic testing in this phase of the management process includes only those tests for abnormalities of which the patient had no suspicion. If there is suspicion of disease, a woman enters the assessment phase, where diagnostic testing and biopsies are performed to confirm or reject a malignant diagnosis. In the therapeutic phase, patients are treated for the primary tumor and any conditions related to the spreading of the disease. Finally, the follow-up phase includes all diagnostic testing to monitor the patient's progress after treatment, as well as therapeutic treatment upon any relapse.

It is important to note that this broad definition of breast cancer management and treatment includes a high number of screening examinations of women who ultimately did not have cancer. We purposely included this group for two reasons. First, the amount of resources consumed by screening and diagnostic services was significant relative to those consumed by treatment alone. Second, we observed different practice patterns among countries regarding screening and diagnosis. Since these different approaches have large resource consumption implications and may actually affect outcomes through earlier detection of the cancer, we included the process of seeking and confirming a diagnosis of breast cancer in our analysis of breast cancer management and treatment.

Measurement of inputs

The input measure for breast cancer included all labor, capital, and supplies associated with the procedures performed in the four phases.[2] We did not include elective reconstruction of the breast after removal by a mastectomy. A preliminary analysis revealed that in the time period of our study, the frequency of breast reconstruction was quite low in each country; the resources consumed by reconstruction were, therefore, likely to be small in comparison to 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.

Measurement of outcomes

Our outcome measure is the 5-year survival of women diagnosed with breast cancer, with the 5-year period commencing at diagnosis. This measure is calculated from survival statistics for relatively large populations of breast cancer patients in each of the three countries over roughly the same time period. From these statistics, we constructed age-adjusted, 5-year survival curves and compared the survival "profiles" of each country. These comparisons highlight differences in survival times achieved by the countries, which presumably correlate strongly with differences in care that we identified in our analysis of treatment practices. (Details on the calculation of this outcome measurement and a discussion of its validity can be found in Appendix 5A.)

By using 5-year survival as our 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.

---

[2] Refer to the discussion of input methodology in Appendix 1A, "Input Methodology."

**DESCRIPTION OF THE MANAGEMENT
AND TREATMENT PROCESS**

The following sections describe each of the four phases of the breast cancer management and treatment process in some detail and highlight the clinical and economic trade-offs implicit in the range of options available in each phase. Later in this chapter, we discuss the specific practice patterns observed in each country and their implications for productive efficiency.

Screening phase

The fundamental premise for implementing a screening program is that, with appropriate care, the early detection of breast cancer will improve the prognosis. While there is controversy surrounding who benefits from screening, there is evidence to suggest that postmenopausal women (i.e., typically over 50 years old) can benefit in terms of increased survival. Premenopausal women may also benefit from early detection, but it is less clear that a screening program can meet this goal effectively.[3] This is due in part to the fact that early detection is more difficult in the premenopausal breast given the current state of technology.

Several screening methods and tests exist. The three most common are mammography, physical exam, and breast self-examination (BSE) (**Exhibit 2**).

¶ **Mammography** provides a radiographic image of the breast that highlights varying densities of breast tissue. In the postmenopausal breast, there is typically good contrast between the lower density breast tissue and the higher density suspect lesion. However, the premenopausal breast, in general, consists of higher density tissue (i.e., closer in density to that of a tumor) which can significantly reduce the image contrast and hinder detection.

¶ **Physical exam**, as typically performed by a physician, is a comprehensive examination that relies on the fact that many potential tumors are of a size and density that allow them to be palpably differentiated from the surrounding tissue. Like mammography, the effectiveness of physical exam is limited by the overall density of the breast tissue. In general, younger women have denser breast tissue, making the differentiation of the tumor from the healthy tissue difficult.

¶ **BSE** relies on the same principle as physical exam, except that the woman performs the examination on herself. The benefits of BSE have been debated. One argument is that a woman may lack the expertise needed to make this a reliable examination. Another

---

[3]   It is likely that premenopausal women with risk factors, such as family history, benefit from screening.

argument is that a woman can become very attuned to subtle changes after repeated examinations and, therefore, may be a very appropriate and accurate examiner. Since it is difficult to identify the degree to which BSE is performed and its effectiveness as a screening diagnostic is uncertain, we eliminated it from our assessment of breast cancer treatment.

Assuming that early detection is beneficial to the patient, wide-scale screening is useful if the diagnostic test adequately meets several criteria. We use the word adequately since these criteria involve trade-offs; therefore, the adoption of a screening program involves subjective consideration. First, the test needs to be sensitive enough to detect as many of the true cancers as possible. Second, the test should be able to detect true cancers without falsely identifying cancers. Third, the test should be cost-effective in the sense that it saves a sufficient level of costs and/or lives through early detection. That is, the test – including both the direct cost of the test itself and the cost of further procedures performed on a falsely identified group – cannot consume too many of the resources that could otherwise be used in some more productive medical or nonmedical endeavor. Finally, the test should not be highly invasive and/or painful as there is a trade-off between the benefit to the group correctly identified with breast cancer and the unnecessary morbidity or pain endured by those who do not have breast cancer. Since women voluntarily undergo screening, some individuals could choose not to participate because the small probability of cancer detection is offset by the large cost of morbidity or pain.

Neither mammography nor physical exam are perfectly sensitive or specific, even in combination. Under very controlled experimental situations, it is possible to compare the sensitivity and specificity of mammography and physical exam and to draw general conclusions. The research literature has a number of comparisons, many of which point to mammography as the "better" test. However, from a resource consumption point of view, it is likely that the actual amount of resources used in producing and interpreting a mammogram are greater than those required to produce a physical exam, thus complicating the trade-off decision.[4]

A complication of much greater concern is the fact that screening for breast cancer does not occur under ideally controlled conditions. In fact, the skill levels of the personnel involved in both types of tests can have a marked impact on the testing. Consider the example of a general practitioner (GP) annually performing a small number of very short-duration physical exams for mostly premenopausal women; the likelihood of this practitioner encountering a positive case, let alone

---

[4] In some cases, mammography is performed in conjunction with a physical exam and, therefore, the combination would undoubtedly consume more resources than physical exam alone. However, a significant amount of mammography is performed routinely without a physical exam unless the results of the mammogram are suspicious.

successfully identifying it based on a brief physical exam, is reduced. Likewise, a practitioner with little experience in interpreting mammographic films will likely miss some malignancies and falsely identify others, especially in harder-to-distinguish premenopausal cases.

Whether screening is performed, who it is performed on, what technology is utilized, and who administers and interprets the results of the diagnostics are issues that depend on how the health care system is structured and functions. For example, a centrally administered public system with no private market alternative may explicitly consider the trade-offs inherent to these issues in implementing a screening program. In contrast, a fully private market system may place the burden of sorting through these issues on the consumer. Whether the end result is the outcome of an explicit decision, a function of consumer preference, or subject to other considerations, it has considerable economic consequences for the resources consumed in breast cancer care.

Assessment phase

The assessment phase covers the actions performed to arrive at a definitive diagnosis for all women with a suspicious screening result or a symptomatic presentation of a potential breast abnormality. For those patients with an identified malignancy, the assessment phase also involves actions taken to determine the extent of disease and to plan for treatment (**Exhibit 3**). It is important to recognize that the majority of women entering this phase will emerge with a benign breast cancer assessment and, therefore, will require no further intervention except screening at appropriate intervals.

While biopsy is not the only procedure performed in the assessment phase, it may be the most important as it confirms the existence or absence of cancer. There are several biopsy techniques available that differ in terms of the technology employed and the setting of the procedure (i.e., the physical location of the procedure and its associated staffing and treatment duration requirements). During the time period of our study, we observed four different approaches to biopsy.

¶ **Inpatient surgical, one-step**. The first technique combines the assessment of the abnormal mass with the actual treatment of the breast cancer, should the mass be identified as malignant. It involves an inpatient surgical procedure where the abnormality is removed from the breast while the patient is under general anesthesia. Then the tissue is examined for cancer, and if found to be malignant, definitive surgery is performed to remove the cancer entirely and to assess or remove the lymph nodes in the axilla. This one-step surgical protocol requires that the surgeon attains contingent consent for definitive treatment from all patients should a malignancy be identified. In the malignant cases, the procedure obviously combines

aspects of both the assessment and therapeutic phases as we have defined them. It is important again to note that for the majority of patients undergoing this procedure, the tissue removed from the breast is benign and no further treatment is administered after the pathological confirmation is attained.

¶ **Inpatient surgical, two-step**. The second type of biopsy assessment is a slight variation of the first. The difference lies in the decoupling of the one-step surgical procedure into two distinct surgical events. The first event is an inpatient diagnostic biopsy and the second is the definitive surgical treatment should the biopsy be malignant. The major advantages of the two-step protocol relate to the decoupling of the assessment from the definitive treatment, and a more accurate assessment of the biopsy. By decoupling the assessment from the definitive treatment, the discussion of definitive cancer treatment can be deferred until after a diagnosis has been made. Thus, the majority of patients do not have to consider the implications of the various treatment options because they are diagnosed as benign. Only those with a malignancy need undergo the anxiety of selecting treatment options for the therapeutic procedures. In addition, the pathological assessment is more thorough in this two-step approach because the pathologist has the time to do more than the simple frozen section analysis performed in the one-step protocol described above.

From the resource consumption point of view, the one-step inpatient protocol consumes fewer resources than the two-step inpatient protocol because it combines the two procedures into one event for malignant cases.

¶ **Outpatient surgical, two-step**. The third type of biopsy is a variation on the two-step protocol described above. The initial biopsy is performed in an outpatient setting, which allows the patient to avoid confinement to a hospital. Like the prior two-step protocol, this protocol follows up malignant cases with an inpatient surgical procedure. Overall, this protocol consumes fewer resources than the inpatient two-step process because it avoids the resources associated with the inpatient stay required for the biopsy. It also requires fewer resources than the one-step protocol as it avoids an inpatient stay for all benign cases.

¶ **Fine needle aspiration (FNA), two-step**. The final biopsy protocol we observed was just beginning to emerge as an option during the time period studied and did not require surgery. FNA biopsy allows the abnormality to be pathologically assessed from a cell sample retracted using the bore of a needle. This procedure is relatively quick and can be performed in an outpatient setting. At the time, FNA could only be performed on abnormalities of the size and

density that could be felt, to assure that the needle in fact sampled the target tissue area. In addition, the pathological skill required to assess a cell sample is greater than that to assess a tissue sample retrieved through a surgical biopsy. Due to the nature of this procedure at the time of our study, a definitive diagnosis (i.e., either benign or malignant) was achievable in about 90 percent of the cases. The remaining 10 percent of the cases were inconclusive because the pathology was unclear.[5] It is important to mention that there is another needle-based technique that retrieves a small tissue sample by using a larger bore needle, known as core needle biopsy. However, this technique was not in use to any significant extent during the time of our study, and we, therefore, exclude it.[6]   Overall, the biopsy portion of the two-step FNA biopsy/definitive treatment protocol required the fewest total resources of all the biopsy techniques observed during the period of study.

Therapeutic phase

Once a case enters the therapeutic phase, there has been a definitive diagnosis of cancer. In each of the three countries, about 90 percent of patients had a tumor that was either confined to the breast or had spread to the nearby lymph nodes (referred to as "early stage") in the axilla (an area near the underarm).[7] The remaining patients had metastatic disease, involving the spread of cancer to remote areas of the body such as the brain, liver, lungs, or bone (referred to as "advanced stage"). While the management strategies varied by the extent of metastasis, they typically involved some combination of surgery, radiotherapy, chemotherapy, and hormonal therapy (**Exhibit 4**). Since the vast majority of patients did not have metastatic disease, we concentrated on potential differences in the treatment of this larger, early stage group.

Presuming strong enough health, most early stage patients were considered for surgical treatment. In the early 1980s, clinical research had begun to demonstrate that a new, less disfiguring, surgically based care option delivered essentially the same outcome for some early stage cases as the generally accepted, more radical surgical procedure.[8] We observed utilization of both options.

---

[5]   Sources: Clinician interview, NHS Breast Screening Programme.

[6]   Our analysis captures the amount of needle biopsy performed in total and assumes that most of this was performed via FNA. The resource requirement for core needle biopsy is similar to that for FNA, thus if the adoption of core needle was more significant than we estimated, this would not lead to much, if any, change in the total resource consumption we calculated.

[7]   Sources: SEER; Thames Cancer Registry; Krebsregister – Saarland.

[8]   Eligibility for this less disfiguring procedure was typically based on the size and location of the cancer in the breast. In general, cases of small and peripheral tumors were eligible.

The more radical treatment of the breast tumor through mastectomy required the complete removal of the breast and the lymph nodes in the axilla. The new research showed that for eligible cases, a breast-conserving procedure could be used that involved removal of the tumor and a small amount of healthy breast tissue surrounding it, followed by a sampling of the lymph nodes in the axilla. The protocol also called for irradiation of the breast and axilla to eradicate any small areas of cancer spread from the primary tumor site. This radiation therapy required 4 to 6 weeks of daily administrations of radiation, which made the breast-conserving procedure more resource intensive overall than the mastectomy alone.

While surgery was the typical and most frequent major intervention in the therapeutic phase, chemotherapy, radiotherapy, and hormone therapy played roles as well, especially in the treatment of more advanced stage disease. At the time period of our study, clinical trials had just begun showing that chemotherapy administered after surgery was beneficial, and we observed some amount of this adjuvant chemotherapy in practice. In addition, both chemotherapy and radiotherapy were frequently used separately or in combination for patients ineligible for surgery. Hormonal therapy is useful as a supplement to other therapies when the tumor is shown to have hormone receptors. The presence of receptors is typical for postmenopausal cases. Hormonal therapy consumes very few resources as the hormones are taken orally, are comparatively inexpensive, and have no major side effects that would be associated with a need for hospitalization. The administration of radiotherapy does consume resources due to the large number of visits required to deliver the total amount of radiation prescribed by the various protocols. In addition, a small number of cases must be treated in the hospital for a variety of reasons such as the frailty of the patient. For these inpatient cases, the resource requirement for therapy is large due to the number of days over which the radiation is delivered. Chemotherapy, too, is sometimes delivered in an inpatient setting because it sometimes weakens the body's ability to fight infection. There are obvious differences in resource requirements between an outpatient and inpatient setting.

Follow-up phase

Once a case has passed through the therapeutic phase, those patients in which the disease has been controlled successfully enter a monitoring period. Most recurrences occur within a 2-year period; thus, ongoing monitoring typically is more intense within this time period. However, lifetime follow-up, including examination of the contralateral breast is recommended by many. As in the screening phase, the follow-up phase is designed to detect the presence of cancer (**Exhibit 5**). Potential variation in the monitoring part of this phase arises from the range of available diagnostic options and how often they are administered. Should a recurrence occur, the patient receives additional treatment that we considered part of the follow-up phase. Relapse treatment varies and can

include different types of surgery, as well as radiotherapy and chemotherapy. In general, the follow-up phase consumed a relatively small portion of total resources devoted to breast cancer management and treatment.

* * *

Of all four phases, the assessment and therapeutic phases consumed the most resources, followed by the screening phase. The assessment phase confronted providers with several options for arriving at a definitive diagnosis, differing in terms of diagnostic technology employed and the procedure setting. Management in the therapeutic phase was reasonably similar in the sense that most women received a surgical treatment of some type; resource utilization differences stemmed from differences in the use of breast-conserving procedures and radical mastectomy, differences in in-hospital surgical recovery times, and differences in use and setting for other therapies such as chemotherapy. Finally, in the screening phase, the key considerations included whether or not screening was performed and, if performed, the breadth of the program and which technology was employed.

ASSESSMENT OF RELATIVE PRODUCTIVE EFFICIENCY

U.S. versus Germany

The U.S. used 38-percent fewer inputs and achieved 9-percent better outcomes than Germany in the treatment of breast cancer (**Exhibit 6**). With better outcomes and lower inputs, the U.S. was clearly more productive than Germany in breast cancer treatment (**Exhibit 7**).

U.K. versus Germany

The U.K. used 53-percent fewer inputs and achieved 6-percent better outcomes than Germany in the treatment of breast cancer (**Exhibit 6**). With better outcomes and lower inputs, the U.K. was clearly more productive than Germany in breast cancer treatment (**Exhibit 7**).

U.S. versus U.K.

The U.S. used 15-percent more inputs and achieved 3-percent better outcomes than the U.K. in the treatment of breast cancer (**Exhibit 6**). As discussed in Chapter 1, when one country had both higher inputs and better outcomes, we assessed which nation had the higher productive efficiency by calculating

average productivity and understanding the nature of the disease treatment process. Measuring average productivity requires comparison of a nation's outcomes *with* treatment to outcomes *without* treatment. As there were no data available on outcomes for breast cancer without treatment, we were unable to calculate a precise average productivity estimate for the U.S. and the U.K. in the treatment of breast cancer. Estimating average productivity over the entire range of possible outcomes without treatment, the average productivity advantage ranges from a 35-percent advantage for the U.S. to a 13-percent advantage for the U.K. (**Exhibit 7**).[9] We are, therefore, unable to determine which nation has higher productive efficiency in the treatment of breast cancer.

When we cannot determine which country is more productive, we perform a cost-effectiveness assessment to comment on which country has the preferred input/outcome combination (**Exhibit 8**), as described in Chapter 1. In U.S. prices, the U.S. spent an additional $32,000 per LY, which is below the $100,000 benchmark (discussed in Chapter 1) and can, therefore, be considered cost-effective. In U.K. prices, however, the U.S. spent only an additional $13,000 per LY; as health care expenditures under $30,000 are considered cost-effective using the benchmarks discussed in Chapter 1, the U.K. could likely have benefited from greater expenditures in the treatment of this disease.

MAJOR DRIVERS OF PRODUCTIVE EFFICIENCY DIFFERENCES IN TERMS OF PROVIDER BEHAVIOR DIFFERENCES

The productive efficiency differences observed were caused most directly by differences in provider behavior. In this section, we discuss the provider behaviors that were the major drivers of productive efficiency differences between the nations, focusing first on drivers of input differences.

Summary of drivers of input differences

U.S. versus U.K. Comparing the U.S. with the U.K. in breast cancer treatment, four practice pattern variations caused significant differences in input consumption, two that had a large influence and two that were less important (**Exhibit 9**). The most important factor was the less prevalent screening in the U.K. in contrast to the broad-based mammographic and physical exam screening in the U.S.; the other important factor was the different approach to biopsy between the two countries. Differences in the frequency of chemotherapy and

---

[9]   The entire range of possible outcomes without treatment is the range from immediate death to lowest outcomes with treatment (which correspond to the outcomes with treatment in Germany).

radiotherapy between the two countries were of lesser importance, but still had a significant impact.

U.S. versus Germany. Comparing the U.S. with Germany, four behavior differences caused most of the differences in input consumption (**Exhibit 10**). Three of these – the biopsy protocol, surgical frequency and length of stay (LOS), and chemotherapy setting – were all major contributors to the difference. Different approaches to screening were less important but still meaningful, if the resources consumed through increased biopsy due to the broader mammography screening in the U.S. were considered.

Summary of drivers of productive efficiency differences

U.S. versus U.K. Of the six provider behavior categories introduced in our casual framework in Chapter 1, five led to significant but offsetting differences in productive efficiency between the U.S. and the U.K.: care triaging, treatment duration, staffing levels, setting choice, and technology adoption (**Exhibit 11**). The U.S.'s broad-based mammographic screening program, use of surgical biopsy rather than FNA, and higher staffing levels lowered its productive efficiency relative to the U.K. On the other hand, the U.S.'s shorter LOS and use of outpatient biopsy versus inpatient biopsy increased its productive efficiency relative to the U.K. In combination, these offsetting behavioral differences led to indeterminate productive efficiency between the U.S. and U.K.

U.S. versus Germany. Again using our six categories, three major differences in provider behavior led to somewhat offsetting differences in productive efficiency between the U.S. and Germany: treatment duration, staffing levels, and setting choice (**Exhibit 11**). Germany's productive efficiency relative to the U.S. was lowered by its longer hospital LOS and its greater use of the inpatient setting for biopsy and chemotherapy. While Germany's productive efficiency relative to the U.S. was raised slightly by Germany's lower staffing levels, the net effect of the provider treatment differences led to higher productive efficiency in the U.S.

Below, we discuss in greater detail the major provider treatment variations that led to differences in inputs and productive efficiency by each phase of the breast cancer management and treatment process.

Screening phase

As described, differences in screening practices had a significant effect on differences in overall input consumption and productive efficiency. At the time of our analysis, the U.K. had no formal screening program, and therefore, no

resources were considered to be consumed in this phase.[10] In comparison, the widespread adoption of screening in the U.S. 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 in the U.S., with mammography accounting for the majority of these resources. It is interesting to note that 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. Breast physical exam in the U.S. was part of a typical gynecological exam, which means that women potentially as young as 18 years of age underwent this type of screening. Like the U.S., Germany employed both mammographic and physical exam screening. Overall, Germany consumed slightly more resources than the U.S. on screening but, on balance, consumed more on physical exam than on mammography.

As previously mentioned, the broader the screened population (i.e., how young an age group is screened), the more likely it was that the screening incurred large additional "downstream" costs in the assessment phase by raising the number of false positive cases requiring assessment. This 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 in the U.S. due to its wide use of mammography on younger women, which when compared with the mostly physical exam-based screening in Germany, identified more nonpalpable masses, most of which were benign. By increasing costs without producing substantial benefit, the U.S.'s broad-based mammographic screening lowered the U.S.'s productive efficiency in breast cancer treatment.

In 1987, the U.K. instituted a nationwide screening program that became fully functional by 1991.[11] Utilizing mammography, the program is restricted to women over the age of 50 and currently calls for screening every 3 years. This program differs from that of either Germany or the U.S. in that it is population based, administered centrally, and targeted at postmenopausal women. In addition, when screening produces a suspect case, the screening center closely coordinates with both the assessment and therapeutic functions as necessary. Undoubtedly, this program has raised the total consumption of resources for breast cancer care in the U.K. However, due to the targeted nature of the screening and the longer average time between screens, the U.K. likely consumes less on screening today under this program than either the U.S. or Germany. In terms of outcomes, it is too soon to tell if the program has had any effect on

---

[10] Sources: Forrest et al., 1986; clinician interviews. It is likely that a small number of women would have received physical exam screens in the U.K., but there is no good estimate for this level of care. Therefore, we included no resource consumption for screening in our analysis of the total resource consumption for breast cancer care in the U.K.

[11] Sources: NHS Breast Screening Programme, 1993; National Audit Office, February 1992.

mortality for breast cancer and, therefore, any effect on the U.K.'s productive efficiency in breast cancer treatment.

Assessment phase

In general, we observed a different combination of biopsy technology and procedure setting in each of the three countries. Historically, the combined one-step protocol of surgical biopsy and definitive surgical treatment contingent on malignancy was used in all three countries. However, since the early 1980s, there has been a trend toward decoupling the two procedures. Interestingly, each country has adopted a different strategy in decoupling.[12]

In the U.S., the move toward the two-step biopsy/definitive treatment protocol was completed by the time period of our analysis. Virtually all of the biopsies occurred as part of the two-step protocol, utilizing the outpatient setting for the surgical biopsy procedure. The U.K. had only begun to make the transition to a two-step protocol, based on FNA biopsy and inpatient surgical management of malignancies. Only about 20 percent of cases were following the two-step protocol, with the remaining 80 percent still being managed under the one-step, biopsy-contingent, definitive treatment protocol. Like the U.K., Germany had adopted a two-step protocol in only about 20 percent of all cases. However, cases following the two-step protocol had an inpatient surgical biopsy that differed from the cases in both the U.S. and U.K. The remaining 80 percent were managed through the one-step, biopsy-contingent, definitive treatment protocol.[13]

With these different approaches to biopsy, in conjunction with different frequencies of biopsy, we observed different levels of input usage and productive efficiency. Overall, the U.S. consumed the least on biopsy since it moved entirely away from the inpatient setting. Cases in the U.K. handled under the FNA biopsy two-step protocol were the least resource-consuming of all; however, as the majority of cases were handled under the more resource-consuming, one-step inpatient biopsy protocol, overall the U.K. consumed more resources than the U.S. in the assessment phase. Germany consumed the most resources, as both protocols in use were performed in the inpatient setting, and Germany was thus the least productive in the assessment phase.

While there were differences in the frequency of biopsy, the relative effects of these differences were not as great as differences in the type of biopsy protocol followed in each country. In fact, the U.S., which had the lowest resource

---

[12] As little research data was available on the use of different types of biopsies and settings for biopsy in each country, we have estimated these differences based on our interviews in each country with clinicians.

[13] As noted in the footnote above, these statistics relating to the percentage of the one- versus the two-step protocol are estimates derived through interviews with clinicians in each of the three countries.

consumption for biopsy, had the highest frequency. Biopsy frequency in Germany was slightly higher than in the U.K. due to the screening performed.

Of all the phases of breast cancer care, differences in provider behavior in the assessment phase introduced the largest amount of total variance to the differences observed in overall resource consumption among the three countries. The U.S. consumed 40 percent of total care resources on assessment compared with 50 percent in the U.K. and 43 percent in Germany. Relative to the absolute amount of resources consumed by the U.S. on assessment, the U.K. and Germany consumed 3-percent and 20-percent more, respectively.

Therapeutic phase

Surgery is the most important treatment for breast cancer. We observed differences in the frequency of surgery overall, as well as in the mix of the two major types of surgeries performed. Overall, the frequency of surgery for the primary tumor in the breast[14] in the U.S., U.K., and Germany was 91 percent, 75 percent, and 97 percent, respectively. Of those cases treated surgically, the frequency of breast-conserving surgery for the U.S., U.K., and Germany was 29 percent, 44 percent, and 39 percent, respectively.[15]

From the standpoint of resources consumed for the surgical procedure and the hospital recovery time, these frequencies resulted in the U.S. and U.K. consuming resources at about the same level, while Germany was about 50-percent higher. This is because total resource consumption comprises the above two frequencies, as well as the corresponding LOS for both the breast-conserving and mastectomy procedures; while the U.S. had the shortest LOS for both procedures, hospital stay for breast conservation in the U.K. was shorter than in Germany, and the U.K. and Germany had about the same stay for mastectomy. While the surgical differences between the nations affect input usage without impacting productive efficiency (i.e., performing more surgeries moves a nation further out on the same production function rather than to a different production function), the differences in LOS had a direct impact on each nation's productive efficiency (i.e., a shorter LOS with similar outcomes yields better outcomes for fewer inputs and is thus more productive).

Radiotherapy was a component of several different protocols for the management of breast cancer. Each of these protocols could have potentially differed in the number of delivery sessions prescribed (called "fractions") and the total amount of radiation to be delivered. In addition, the therapy was delivered, although infrequently, in an inpatient setting, which brought in the

---

[14] Does not include surgeries performed upon recurrence of cancer.

[15] Sources: *SEER*; Thames Cancer Registry; Krebsregister – Saarland; Arbeitsgruppe Zur Koordination Klinischer Krebsregister; Großhadern Dataset; Foreman and Rider, 1995; clinician interviews.

resource consumption effect of the LOS. Thus, total resource consumption related to radiotherapy was driven by many factors. Overall radiotherapy resource consumption in the U.S., U.K., and Germany was 6 percent, 12 percent, and 5 percent of all resources devoted to breast cancer care in each country, respectively. Compared with the total resources consumed by radiotherapy in the U.S., the U.K. and Germany consumed about 60 percent and 10 percent more, respectively.

There are several, sometimes offsetting, reasons for these overall differences in care. The U.K. tended to administer fewer fractions of greater total radiation than both the U.S. and Germany.[16] This tended to lower its relative resource consumption. However, according to clinicians, the delivery of a given fraction required more labor in the U.K. due to the age of radiotherapy equipment in use, which would tend to increase resource consumption. For inpatient radiotherapy, the U.S. consumed more resources than both the U.K. and Germany due to its higher level of hospital staffing. Germany, in turn, consumed more on inpatient radiotherapy than the U.K. due to the fact that the U.K. tended to use fewer fractions in its protocols.

The overall frequency of radiotherapy differed among the three countries and obviously had some effect on resource consumption. The U.S. was lowest, with about 55 percent of all cases receiving some type of radiotherapy.[17] The U.K. and Germany had frequencies of about 65 percent and 60 percent, respectively. The difference between the U.K. and the U.S. can be explained by the U.K.'s higher frequency of breast-conserving procedures, which are typically followed by radiotherapy. Adding to this difference, and also explaining the higher radiotherapy frequency of the U.K. compared with Germany, was the U.K.'s lower surgical frequency, which was offset partially by increasing radiotherapy frequency.

Chemotherapy, like radiotherapy, was a component of many of the protocols in use for the treatment of breast cancer. Most of the total amount of chemotherapy was administered in those cases that had a recurrence of cancer after primary treatment failed, and thus, in the follow-up phase. The most frequent use of chemotherapy in the therapeutic phase was adjuvant chemotherapy to support surgical care. Just prior to the time period of our analysis, the benefits of postoperative chemotherapy were becoming recognized. Thus, surgical care protocols in each of the countries were beginning to include chemotherapy.

Although the small differences in the frequency of adoption of adjuvant chemotherapy did not lead to significant differences in the total consumption of

---

[16] Sources: Thames Cancer Registry; Arbeitsgruppe Zur Koordination Klinischer Krebsregister; clinician interviews.

[17] Sources: SEER; Thames Cancer Registry; Arbeitsgruppe Zur Koordination Klinischer Krebsregister; clinician interviews.

resources or productive efficiency among countries, the setting of chemotherapy administration did. While both the U.S. and U.K. utilized the outpatient setting for all chemotherapy, Germany performed about one third of breast cancer chemotherapy in the hospital.[18] Inpatient administration made chemotherapy much more resource intensive, as the typical regimen included about six courses, each one over a 3- to 4-day period. Thus, patients undergoing this protocol would spend 18 to 24 days in the hospital over a 6-month period, lowering Germany's productive efficiency in the treatment of breast cancer compared to both the U.S. and the U.K. The U.S. and U.K. consumed only about 4 percent and 3 percent, respectively, of total resources on chemotherapy in the therapeutic phase, while the setting difference in Germany added about 25-percent more resources to the level consumed by the U.S.

As mentioned previously, hormone therapy consumed relatively few resources compared with other therapies. The slight differences in the frequency of its usage did not add substantially to the overall differences in resource consumption among the three countries.

Overall, the U.S. consumed about 30 percent of total resources in the therapeutic phase. The U.K. and Germany consumed about 39 percent and 30 percent, respectively. Relative to the U.S., the U.K. and Germany consumed 3-percent and 11-percent more resources, respectively.

Follow-up phase

While 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.

CAUSAL ANALYSIS OF PROVIDER BEHAVIOR DIFFERENCES: INCENTIVES AND CONSTRAINTS, SYSTEM STRUCTURE, AND REGULATION

The differences in provider behavior in the U.S., U.K., and Germany can be explained by the incentives and constraints providers faced in each country and by underlying differences in the health care system structure and regulation, which

---

[18] Source: Clinician interviews.

are described in detail in Chapter 2. Below, we explain how these differences led to the provider behavior differences we observed in the treatment of breast cancer.

U.S. versus U.K.

As discussed, productive efficiency differences between the U.S. and U.K. were caused by differences in three areas of breast cancer treatment: screening, biopsy, and inpatient procedures. Screening practices had a large impact on productive efficiency due to differences in care triaging (i.e., the extent of screening) and the technology used for screening. Biopsy practices had a large impact on productive efficiency due to differences in the choice of biopsy technology and care setting. Inpatient procedure practices influenced productive efficiency due to differences in LOS and staffing levels between the two nations. In this section, we discuss each of these in turn. We also discuss differences in radiotherapy and chemotherapy practices, which did not appear to influence relative productive efficiency, but led to differences in resource consumption.

Screening: extent of screening and use of technology

Provider incentives and constraints (Exhibit 12). The presence of widespread mammographic and physical exam screening performed in the U.S. was in stark contrast to the near absence of screening in the U.K.[19] In the U.S. during the mid 1970s, breast cancer awareness increased greatly, in part in response to the cases of high-profile public figures. As a consequence of this and an earlier clinical trial that suggested an outcome benefit to early detection and treatment, women became interested in breast cancer screening and providers began recommending it. At that time, mammography was being used frequently for diagnostic purposes, but had yet to be used routinely for screening.

Physical exam screening in the U.S. likely became more widespread due to physicians beginning to promote an annual gynecological visit. At the center of this visit was the cervical cancer screen (i.e., Pap smear), but it also included a brief physical exam, part of which was devoted to a breast exam.

While at one time mammographic screening in the U.S. was paid for directly by consumers, it has became covered by insurance over time, creating a strong incentive for physicians and hospitals (or other facilities) to increase the frequency of its use. Since there were essentially no major regulatory barriers to purchasing capital equipment, a large increase in the installed base of mammographic equipment occurred; and women underwent mammographic

---

[19] Beyond the occasional physical exam screening that a patient in the U.K. would receive, there were mammography trials underway in the U.K. at the time of our analysis. Since both activities wee relatively limited on a population wide basis, we did not consider them in our analysis. In any case, the higher resource intensity approach in the U.S. far overwhelmed the limited screening being performed in the U.K.

screening in increasing numbers. No capital equipment had to be purchased for physical exams, which also increased in frequency. As discussed earlier, this widespread screening, especially among premenopausal women, likely lowered the U.S.'s relative productive efficiency in breast cancer treatment due to the large resources it consumed both in screening and further downstream in the treatment process.

Adoption of mammography technology for screening for most women in the U.K. was subject to the approval of the NHS, which did not occur until after the time period of our study. Like the U.S., the U.K. had mammography equipment in place to perform diagnostic assessments to aid diagnosis once a suspicious symptom had been recognized. However, the equipment was housed entirely in the hospitals and, therefore, patients required a GP referral to the hospital in order to gain access. Since hospitals would not accept self-referrals for screening, very few, if any, mammographic screens were performed and no equipment was added strictly for screening purposes. Furthermore, since capital was controlled centrally by the NHS and there was no mammography capacity for widespread screening, providers and hospitals had no incentive to institute screening at the expense of taking capacity away from breast cancer diagnostic needs.

No economic incentive existed for GPs in the U.K. to perform physical exam-based screening. The NHS did not recognize the procedure in its GP reimbursement, and GPs were sufficiently busy that they had no incentive to perform additional work without compensating revenue. U.K. physicians thus had little incentive to institute broad-based screening and U.K. capital constraints limited the adoption of mammography.

System structure and regulation (Exhibit 13). The sources of these differences in physician and hospital incentives and in physician supply and capital constraints were major differences in health care system structure and regulation. Most physician services in the U.S., including both specialists and primary care, were negotiated and compensated on a fee-for-service (FFS) basis by payors. U.S. physicians thus had incentive to provide mammographic and physical exam screening to all of their patients, as they were compensated for these activities. In contrast, specialists in the U.K. were paid flat salaries negotiated on an annual basis, providing no incentive to perform screening on a broad section of the population.

Differences in physician and hospital incentives were also driven by differences in the competitive intensity of the nations' health care systems. U.S. physicians' and hospitals' competition for patients encouraged them to provide screening in response to patient demand; in addition, the threat of malpractice could have further encouraged U.S. physicians to be responsive to patient demand. Likewise, U.S. payors' competition for members led them to respond to consumer demand by reimbursing mammographic screening, as well as an annual gynecological exam. In contrast, the majority of the U.K. population had to rely on the NHS for all their health care needs, creating no competitive payor pressure (although there

might have been political pressure) in the U.K. system to respond to consumer demands.

Regulation also contributed to the more limited access that breast cancer patients had to mammographic screening in the U.K. The supply of physicians in the U.K. was controlled by regulation and, as it was more limited than the supply of physicians in the U.S., could not as easily accommodate an increased level of services. The NHS also tightly controlled hospital budgets and reimbursed hospitals in a lump sum for all operating costs. This fixed reimbursement placed the burden of resource allocation on the hospitals and physicians; these two entities had to decide how much would be spent on what disease and, within a given disease, who was treated at what level. This higher level of product integration, combined with a limited fixed budget, encouraged medical rationing according to need and resulted in capital being less readily available for mammographic equipment in the U.K. than it was in the U.S. The monopolistic position of the NHS allowed it to enforce this strict resource supply control, as well as mandates for system usage, like the rules for referral that also contributed to the limited use of mammography for screening.

Biopsy: choice of biopsy technology and setting

Provider incentives and constraints (**Exhibit 12**). The U.S. performed more biopsies per diagnosed patient than the U.K. Almost all biopsies in the U.S. were performed using the two-step outpatient biopsy as opposed to the mix of one-step inpatient protocol (80 percent of cases) that combined biopsy and definitive surgery and the two-step FNA outpatient protocol (20 percent of cases) that decoupled biopsy from definitive treatment. As mentioned earlier, these differences had meaningful, but offsetting, influences on the relative productive efficiency of the U.K. and U.S.

The higher frequency of surgical biopsy in the U.S. compared with the U.K. was primarily due to the widespread screening program in the U.S., which, as mentioned before, identified many suspicious, though ultimately benign, cases that were followed up with biopsy. As the incentives and constraints that led to the higher frequency of biopsy are the same as those that led to the higher frequency of screening, we will not discuss the higher frequency of biopsy in the U.S., but instead refer the reader to the prior analysis of screening differences between the two countries.

By the time of our analysis, the U.S. had completely departed from the one-step protocol that combined biopsy and surgery and had adopted the two-step outpatient biopsy. Surgeons in the U.S. had a clear economic incentive to decouple the one-step surgical protocol into the two-step surgical protocol; by moving to the two-step protocol, surgeons would receive more revenue under

FFS reimbursement.[20] Furthermore, no supply constraints on surgeons, operating room facilities, or hospital beds existed in the U.S., allowing the additional activities to be absorbed easily by the health care system. Arguably, the surgeon could justify the introduction of the additional surgical step as beneficial for the patient since it avoided the need for contingent consent for definitive surgery and offered the patient the maximum assurance of a definitive and accurate assessment of the abnormality. The position of the hospital was less clear. Although the move from the one-step procedure to the two-step procedure would have been favorable from a revenue standpoint if the two-step procedure was performed in the hospital, the U.S. moved to an outpatient biopsy, reducing the revenue of the hospital by eliminating inpatient care (in some hospitals, this was partially offset by having outpatient facilities as well). This hospital disadvantage, however, was probably overwhelmed by the surgeon incentive and the degree to which hospitals owned outpatient facilities.

The existence of this surgeon incentive could also have delayed the substitution of the lesser priced FNA technology for the two-step outpatient biopsy in the U.S. Since the surgeon typically did not perform the FNA biopsy, there was a disincentive for the surgeon to encourage its adoption due to the potential revenue loss. This disincentive, in combination with a belief that the surgical biopsy was a more definitive procedure, may have led surgeons to advocate surgical biopsy over FNA, thus lowering the U.S.'s productive efficiency in breast cancer treatment.

In the U.K., use of the two-step outpatient surgical protocol was likely inhibited by the supply constraints on surgeons, hospital beds, and outpatient surgical facilities. Moving to the two-step surgical protocol would increase the total number of surgeries by the number of malignancies found. Since the definitive surgical treatment procedure that was bundled with the surgical biopsy under the one-step protocol would have to be performed as a stand-alone procedure under the two-step protocol, the resource requirements in terms of specialists' time and hospital operating room time would increase. With few, if any, outpatient surgical facilities and limited surgeon capacity, the increased surgery would further tax the already constrained inpatient system. Thus, given the supply constraints, it made sense for the U.K. system to maintain the one-step protocol until an outpatient substitute could be adopted.

When the FNA technology was identified, it allowed the decoupling of the one-step protocol because it was an outpatient procedure that did not require surgeon time or surgical facilities. With FNA technology, U.K. physicians reduced the surgical load down to only those patients found to have malignancies. This actually freed up both surgeon and hospital resources, thus creating an incentive for the overburdened U.K. hospitals to adopt the technology.

---

[20] Sources: Health Care Financing Administration (HCFA); Medicare fee schedule; clinician interviews.

System structure and regulation (**Exhibit 13**). The sources of these differences in physician incentives and physician and hospital supply constraints were differences in health care system structure and regulation. Differences in product integration and pricing of physician services, in particular, played a significant role in driving different biopsy protocols. In the U.S., the biopsy was separable from the definitive treatment with respect to reimbursement, and the surgeon could increase revenues under FFS reimbursement by performing the additional surgical step. This naturally helped create the physician incentive to move toward the two-step surgical biopsy protocol. In the U.K., however, both the fixed annual payment system for hospitals and the fact that surgeons were paid on salary encouraged a higher degree of care integration because they forced providers to consider the treatment and costs for breast cancer in relation to the treatments and costs for all other conditions. This helps to explain the reluctance to leave the one-step protocol if the alternative was the higher hospital resource-consuming, two-step inpatient surgical protocol. It also explains the adoption of the FNA protocol when it became available since that protocol created capacity in constrained surgeon and hospital bed resources.

In addition, competition for patients in the U.S. encouraged surgeons to offer the two-step surgical protocol as a way to attract patients. Since the two-step protocol offered more certainty and better information with which to decide on definitive treatment of malignancy, patients may have preferred the protocol. Interestingly, practice in the U.S. today has moved partially away (about 50 percent) from the surgical protocol to the FNA (or core needle) protocol. This could be due, in part, to the increased pressure payors have put on physicians to lower costs, which potentially overwhelmed the surgeons' preference for the more resource-consuming surgical biopsy. This could also be due to patients seeking less invasive procedures and physicians striving to meet this demand. The practice in the U.K., as would have been expected, has moved almost entirely to the needle-based protocols.

Furthermore, competition for customers in the U.S. health care coverage market may have resulted in payors competing on the basis of, among other things, price. Thus, payors would be interested in encouraging surgeons to move the biopsy under the two-step protocol to the outpatient setting to capture the cost savings associated with the avoided hospital stay. The surgeon would likely be indifferent to this, as the surgical reimbursement would likely have been the same regardless of the setting.

Regulation also contributed to the differences between the nations in biopsy protocol and setting, as the regulatory-imposed controls on the supply of physicians and hospitals in the U.K. influenced provider incentives and constraints, as described in the previous section.

Inpatient procedures: LOS and hospital staffing

Provider incentives and constraints (**Exhibit 12**). As discussed earlier, differences in inpatient LOS and staffing levels between the U.S. and the U.K. led to offsetting differences in relative productive efficiency. Patients undergoing inpatient procedures in the U.S. had a shorter average LOS than those in the U.K. due to differing hospital incentives. The shorter LOS in the U.S., in turn, caused U.S. hospitals to keep staffing high, as a shorter average LOS means patients were discharged earlier in the U.S. than in the U.K.; as U.S. hospitals thus had a higher proportion of their patients in acute conditions, they had to keep staffing higher to meet the needs of those patients.

System structure and regulation (**Exhibit 13**). The sources of these differences in hospital incentives were differences in health care system structure and regulation. At the time of our study, most U.S. payors individually contracted with hospitals for services. Payments stipulated by these contracts were generally of three basic types: FFS (about 45 to 50 percent), per case (about 35 to 40 percent), or per diem (about 15 percent). FFS payments gave hospitals a fixed amount per procedure or service. Per-case payments gave hospitals a fixed amount per admission, based on the patient's diagnostic-related group (DRG).[21] Under this system, the reimbursement was defined solely by the procedure or diagnosis, not the number of days the patient was in the hospital. While FFS reimbursement did not influence hospital decisions on LOS or staffing, DRG-based reimbursement gave hospitals a clear incentive to decrease costs by either reducing LOS per patient or managing the intensity of services provided, since the hospital assumed the risk for the cost of the patient's stay. Hospitals thus had an incentive to reduce LOS by streamlining admission and discharge processes and shortening preoperative and postoperative time. As discussed above, this shorter LOS likely led to higher U.S. staffing levels, as the intensity of care per average hospital-day increased with decreasing LOS. The net effect of shorter LOS and higher staffing was lower input usage (and, thus, higher profit margins on a per-case basis) for the hospital.

The per diem contract, in which reimbursement was tied to the hospital-day rather than to the entire admission, was not as common as FFS or per-case payment. Payors following this arrangement coordinated physician care and managed patient care throughput (e.g., through a utilization review process). Thus, under a per diem contract (as in FFS contracts), the payor primarily assumed the risk for hospital stay (which could also be transferred to physicians via capitated contracts).

Regardless of the type of contractual arrangement, the risk-bearing entity (payor or provider) had an incentive to manage hospital costs in the U.S. Thus, competition among U.S. payors in the health coverage market likely helped to accelerate the adoption of case rates beyond the Medicare market; since private

---

[21] The Medicare program in the U.S. adopted case rate (DRG) based payment in 1983; many other payors subsequently adopted a derivative of this payment system as well.

payors competed for customers on the basis of price, among other things, they viewed the adoption of case rate pricing as a way to control increasing hospital costs.

In contrast, the U.K. health care system structure and regulation created incentives for U.K. hospitals that were entirely different from those in the U.S. and were not aligned with decreasing LOS. In the U.K., fixed hospital budgets, coupled with limited competition between hospitals for patients or payor funds, created no incentives for U.K. hospitals to manage LOS. As shorter LOS implied that patients would be turned over more rapidly (i.e., greater admission and discharge rates), caring for new patients and performing additional procedures was more costly than maintaining patients in recovery at the end of a longer LOS. In other words, cycling through more patient cases whose treatment required higher input usage would have raised *total* hospital costs without generating a corresponding increase in revenue. More efficient hospitals that utilized or exceeded their entire operating budgets could not easily be differentiated from inefficient hospitals that consumed their entire budgets. Thus, if the budget allocation process was not sensitive enough to detect superior performance, it could deprive operating funds from the more efficient hospitals and, in the worst case, reduce their operating budgets to curtail overspending.

This practice also contributed to the U.K.'s lower hospital staffing relative to the U.S. Since average LOS was longer, and patients arguably required less intense care per day on average, it was reasonable for the U.K. to have lower staffing levels. U.K. hospitals also did not compete for patients as in the U.S. and, therefore, did not need to offer higher levels of service to attract patients. U.K. hospitals also had limited budgets with which to fund all their hospital needs, including staffing. In addition, U.K. staffing levels were also affected by the availability of physicians, which was regulated and limited by the NHS. There was thus much more pressure to maintain lower staffing levels in the U.K. than the U.S.

Radiotherapy: extent of radiotherapy

Provider incentives and constraints. The higher frequency of radiotherapy in the U.K. was mostly a consequence of the broader adoption of breast-conserving surgical treatment in the U.K. compared with the U.S. This is due to the fact that radiotherapy is a necessary complement to the breast-conserving protocol in order to achieve outcomes equivalent to the mastectomy alternative. While the U.K.'s higher frequency increased the U.K.'s input use, it was not a key driver of relative productive efficiency differences.

There were several offsetting behaviors relating to radiotherapy that are worth explaining. As discussed, the average number of fractions delivered in U.K. protocols was lower than that of U.S. protocols; this tended to raise the resource consumption of the U.S. in relation to the U.K. However, through clinician interviews we learned that both the staffing levels and treatment duration were

higher in the U.K. than in the U.S. and Germany. Clinicians attributed both these differences to the much older equipment in use in the U.K. Presumably, the older equipment required more labor to operate and more setup time for each patient. This tended to raise the U.K. resource consumption compared with the U.S. While these differences are interesting, they are offsetting; overall differences in resource consumption introduced by differences in radiotherapy practice were small. Thus, we limit the discussion below to the surgical choice that caused the above described frequency difference.

Our search for explanations for this difference in behavior toward the two surgical protocols covered such things as differences in patient preferences, levels of surgical training, malpractice threat, distance to the nearest radiotherapy center, and availability of reconstructive surgery after mastectomy. Although each of these differences could have contributed to the overall difference in practice, none of them stood out as a particularly compelling explanation for the relatively large difference we observed. However, differences in the incentives and constraints faced by providers in each country could explain why the U.S. appeared to lag the U.K. in adopting the breast-conserving procedure. From the surgeon's point of view in the U.S., the reimbursement for mastectomy was greater than that for breast-conserving surgery.[22] In the U.K., where the hospitals were constrained for beds, there was an obvious incentive to adopt the breast conserving procedure since the associated LOS was shorter than that of mastectomy. The increased radiotherapy load in the U.K. did not require a large capital investment in new equipment, but merely used existing equipment more hours during each day.

There is a distinction in our analysis that should be made to reconcile this point with the previous point that U.K. hospitals had no incentive to manage LOS. First, we must acknowledge that U.K. hospitals had waiting lists. Thus, to the extent that the hospitals could become more efficient by reducing the LOS for a procedure, for instance, the more progress they could make against the waiting list. However, if this efficiency cost additional money, as may have been the case in adding staff to reduce LOS, the hospital likely would not implement it. This is because the implementation would cost more money in two ways (new staffing and additional expense for incremental patients brought in from the wait list), yet under a fixed global budget, no additional revenue would be generated.

However, if an innovation occurred that allowed the substitution of a procedure requiring less inpatient resources than the one it replaced, then the procedure could be adopted and the cost that was avoided (i.e., the capacity created) could be used to serve other patients on the wait list. In this way, no net expense was incurred, yet the patient throughput was increased. From a financial standpoint, the hospital would be indifferent toward adopting the innovation. But from a moral obligation or social welfare point of view, the hospital would obviously

---

[22] Sources: HCFA; Medicare fee schedule; clinician interviews.

pursue adoption. This was the case with the adoption of breast conserving surgery and, as discussed previously, the adoption of outpatient FNA biopsy.

It is interesting to note that since the time period of our study, the proportion of breast-conserving procedures in the U.S. has increased to be near equal with that of the U.K., likely due to increased patient demand as awareness of the breast-conserving procedure has grown.

System structure and regulation. As previously discussed, the FFS physician reimbursement scheme in the U.S. created incentives for the surgeon to choose the revenue-maximizing mastectomy procedure. In contrast, the U.K. specialists' flat salary did not provide this incentive. As also previously discussed, the NHS' practice of reimbursing hospitals in a lump sum for all operating costs placed the burden of resource allocation on the hospitals and physicians; adopting the breast-conserving procedure, which consumed lower inpatient resources, was thus beneficial.

Chemotherapy: extent of chemotherapy

Provider incentives and constraints. As discussed, the adoption of adjuvant (i.e., postoperative) chemotherapy was broader in the U.S. than it was in the U.K. This difference had a small but significant difference in input differences between the two nations, but was not a key driver of relative productive efficiency differences. This behavioral difference can best be explained by the presence of activity-increasing physician incentives in the U.S. In addition, the supply of physicians trained in the area of chemotherapy was likely higher in the U.S. than in the U.K.

System structure and regulation. As previously discussed, the U.S.'s FFS physician reimbursement policies created activity-increasing incentives for U.S. physicians, while the flat salaries for specialists in the U.K. did not. Likewise, the previously discussed lack of competition in the U.K. health care markets enabled the NHS to limit the supply of physicians who were trained to administer chemotherapy and to resist patient demands for more chemotherapy treatment. In contrast, competition in the U.S. led U.S. payors and providers to allow broader access to chemotherapy treatment.

U.S. versus Germany

As discussed, differences in productive efficiency between the U.S. and Germany were caused by provider behavior differences in three areas of breast cancer treatment: biopsy, chemotherapy, and inpatient procedures. Different biopsy and chemotherapy choices, particularly concerning care setting, had a large impact on productive efficiency. Differences in hospital LOS and staffing levels also caused significant productive efficiency differences. In this section, we describe each of these in turn. We also discuss differences in screening practices,

which did not significantly influence productive efficiency, but were important in influencing different decisions and procedures performed downstream in the treatment process. Since the incentives and constraints, system structure, and regulations that led to U.S. provider behavior were discussed in the preceding U.S. versus U.K. discussion, our focus in this section will be on Germany.

Biopsy: choice of setting

Provider incentives and constraints (**Exhibit 14**). While there was a small difference in the frequency of biopsies between the U.S. and Germany due to the higher U.S. use of mammographic screening, the behavior that influenced productive efficiency was the setting choice for biopsy.

As previously discussed, both the U.S. and Germany departed from practicing the one-step bundled biopsy and definitive surgery protocol in favor of the two-step protocol that involved a surgical biopsy. However, in the U.S. the transition to the two-step protocol was complete by the time period of our analysis and the biopsy was performed in the outpatient setting. In contrast, the move to the two-step protocol was still underway in Germany and, for those managed under the two-step protocol, the biopsy was still performed in the inpatient setting.

The incentives for U.S. surgeons to move to the two-step surgical protocol for biopsy were discussed in the U.S. versus U.K. comparison; essentially, the move was in response to an activity-increasing incentive that provided additional revenue to the surgeon. Along with this move, biopsy was performed in the outpatient setting in the U.S., explained in the system structure description below. German hospitals and physicians, on the other hand, had strong incentives to keep procedures in the inpatient setting to occupy hospital beds, which was made possible by the abundant supply of hospital beds in Germany. On a per capita basis, Germany has more hospital beds and hospital based physicians than the U.S.[23] Furthermore, since per diem reimbursement was based on meeting a target hospital occupancy that was quite high, the prosperity of all hospitals depended on increasing to reach, or maintaining, a high level of patient volume. These hospital and physician incentives help explain Germany's choice of inpatient setting for biopsy, as well as its move from the one-step to the two-step protocol, which consumed more total hospital resources. In addition, German regulation, discussed below, established substitution constraints that gave German hospital-based physicians strong incentives to keep procedures in the inpatient setting.

System structure and regulation (**Exhibit 15**). The sources of these differences in physician and hospital incentives, hospital supply, and substitution constraints were differences in health care system structure and regulation. In

---

[23] Note that while German hospitals had lower levels of staffing per bed than the U.S., they had more total beds and hospital physicians per capita than the U.S.

the U.S., some price-based payor competition for members encouraged payors to be cost-conscious; payors may have, therefore, encouraged surgeons to move the biopsy to the outpatient setting to avoid the cost of a hospital stay.

In Germany, however, regulation created strong incentives for physicians and hospitals to keep the biopsy procedure in an inpatient setting. German hospitals were compensated on a per diem basis by the sickness funds, whereby they received additional reimbursement for each day of a patient's stay. Annual hospital budgeting was based on this per diem reimbursement, with the per diem level of the previous year being increased or decreased as necessary to compensate for inflation and for some of the increase or decrease in hospital utilization. Each hospital negotiated as a block with all payors for these per diem rates.

State authorities, as regulators of hospital capacity, had little incentive to reduce the use of local health care resources, as discussed in Chapter 2; in fact, they 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. Consequently, regulations generally stipulated very high occupancy rates, usually about 85 percent. If this target was not met, hospitals were at risk for being reviewed and having their capacity cut. The combination of this regulatory threat and the large supply of hospital beds created a strong incentive for hospitals and, therefore, the physicians they employed, to keep biopsy in the inpatient setting in order to keep more beds occupied. Physicians had further incentive to keep the utilization of public beds high because hospitals typically specified that beds had to be used for public and private patients in a relatively set ratio; German department chiefs could, therefore, add private bed capacity, and thus earn higher private patient fees, by increasing the use of public beds.

Limited competition in the German payor market also contributed to the continued use of inpatient biopsy in Germany. German sickness funds (payors) had relatively restricted memberships along geographic or occupational lines, and retained their members for a relatively long period. Payors did not compete for members on price and could not negotiate differentially from each other with each hospital; therefore, they had no real incentive or ability to pressure hospitals to manage costs or to move biopsy to an outpatient setting.

Regulation in Germany also created incentives for biopsy to remain inpatient by establishing substitution constraints that limited the practice of physicians to either the hospital or office setting and specified which services were to be performed in different settings; most hospital-based physicians were not authorized to perform procedures in the outpatient setting.[24] If a hospital-based

---

[24] Physicians in the German public system were licensed to practice in either the hospital or office-based setting. An exception allowed academic department heads to practice in both settings.

physician converted inpatient biopsy to an outpatient event, the physician and the hospital would lose the revenue associated with the procedure.

Chemotherapy: choice of setting (Exhibits 14 and 15)

The major difference in chemotherapy practices between the U.S. and Germany that influenced relative productive efficiency was that Germany utilized the inpatient setting to a much greater degree for chemotherapy than did the U.S. The reason for this practice difference is the same as that provided for the use of inpatient versus outpatient biopsy in Germany in the previous section.

Inpatient procedures: LOS and hospital staffing

Provider incentives and constraints (**Exhibit 14**). As discussed above, the longer LOS for inpatient procedures in Germany, even offset by higher staffing levels in the U.S., led to lower productive efficiency for Germany relative to the U.S. in breast cancer treatment. As we have already explained, the incentives and constraints in the U.S. system that led to shorter LOS and higher staffing, we will focus on Germany in this section.

The longer stays in Germany resulted from the same incentives that caused biopsy to be performed in the inpatient setting. As previously explained, German hospitals had strong incentives to keep hospital beds occupied and, therefore, to raise or maintain the relatively long hospital LOS. This longer LOS in Germany was made possible by the high number of German hospital beds. The longer stays in German hospitals were only partially offset by the lower staffing levels for both nurses and physicians relative to the U.S. While hospitals in both countries had incentives to attract patients by maintaining high staffing levels, the significantly shorter patient stays in U.S. hospitals likely led to the need for somewhat higher levels of staffing, as previously discussed. German physicians, particularly department chiefs, also had an incentive to fill public hospital beds, as discussed earlier.

System structure and regulation (**Exhibit 14**). Differences in hospital and physician incentives and hospital supply resulted from differences in the health care system structure and regulation. In addition to the system structure and regulation differences discussed in explaining the greater use of the inpatient setting, the German per diem hospital product gave German hospitals incentives to maintain long LOS. In contrast, U.S. hospitals received a significant portion of their reimbursement as case rate (DRG) payments, as discussed earlier, which encouraged hospitals to manage LOS.

Screening: extent of screening and technology choice

Provider incentives and constraints. As discussed, both the U.S. and Germany performed mammographic and physical exam screening. The frequency of mammography in the U.S. was about 50-percent higher than that in Germany, while the frequency of breast physical exam was closer, with Germany

performing about 20-percent more than the U.S. Although this did not lead to meaningful differences in productive efficiency and the sum of resources consumed by screening was similar in the two countries, the higher frequency of mammography in the U.S. likely led to additional downstream resource consumption on biopsy. Thus, the differences are worth explaining.

For the U.S., we already discussed how both the emergence of incentives for physicians and hospitals to provide mammography and the relative lack of constraints on capital purchase led to its broad adoption. The incentives for breast physical exam were similar and could be acted upon without capital investment.

In Germany, ambulatory physicians were authorized by the regulatory bodies to administer a breast cancer screening program based on physical exam and were compensated essentially on an FFS basis. Thus, ambulatory physicians had an activity-increasing incentive created by regulation. In the face of this clear revenue incentive, it is a little surprising that the uptake of physical exam screening was only about 30 percent in Germany. One possible explanation for this relates to the reimbursement methodology utilized in the outpatient setting and its potential to limit the motivation of physicians in performing physical exam screening. As discussed in Chapter 2, German ambulatory physicians were compensated on a point system where different point levels were associated with various procedures. At the end of the payment period, physicians would submit their point totals to the local physician organization (KV); the points would be totaled over all physicians in the region; a point value would be established by dividing a prenegotiated fixed sum for reimbursement by the total points; and the individual physicians would receive payment corresponding to their total point accrual. Thus, each physician would have an incentive to perform as many procedures as possible and to submit the largest possible point total.

However, the point assignments to "low-tech" interventions like physical exam were lower than those of other procedures. Thus, the physician not only had an incentive to increase activity, but also had an incentive to optimize the procedures performed for the highest possible point value. Thus, for physicians who were busy, breast physical exam was not the best use of their time from the perspective of revenue generation. There is evidence to suggest that most German ambulatory physicians were busy; of all three countries, Germany had the highest number of outpatient visits per capita.[25]

Mammography of any kind in Germany required a referral, but its use for *screening* was not sanctioned officially as a reimbursable procedure by the regulators. Thus, women had to convince their physicians in the office setting to provide a referral for mammographic screening under the pretense that the

---

[25] Source: Sandier, 1989.

diagnostic was to be used to investigate a real or less-than-well-defined symptom.

System structure and regulation. Differences in physician incentives resulted from major differences in health care system structure and regulation. The difference in mammographic screening between the U.S. and Germany could be explained mostly by the regulatory constraint in Germany, even though it was partially circumvented. The differences in the frequencies of physical exam are less clear.

Competition among German ambulatory physicians may have helped to drive adoption of screening. Since physicians competed for patients and, therefore, wanted to maintain the satisfaction of existing patients, many would provide referrals upon request for this "illicit" screening. Thus, some incentive was created for providers to please the patient that carried virtually no downside risk since the referrals were not subject to great scrutiny. Given the unsanctioned nature of this behavior, however, it was not surprising that the uptake of mammography for screening purposes in Germany lagged that in the U.S.

Even today, Germany has yet to sanction a mammographic screening program. It is possible that insufficient pressure has been exerted on the regulators to adopt mammographic screening due to a relative disinterest among the population for the service. However, this seems unlikely due to the extent of estimated clandestine screening being performed. The reason may be related to the segregation of the hospital-based and ambulatory physician practices that contributes to the behavior observed. At the time of our analysis, most of the mammography equipment resided in the hospital setting; it may have been that the ambulatory physicians did not want to lose the revenue opportunity associated with sanctioned physical exam screening, which could be lessened if mammographic screening were introduced. However, this point of view must be tempered by the relatively low uptake of physical exam screening.

SUMMARY OF BREAST CANCER CASE RESULTS

Both the U.S. and the U.K. were more productive than Germany in the treatment of breast cancer. The sources of this difference were differences in provider behavior. Germany's relative productive efficiency was lowered by its longer LOS and greater use of the inpatient setting for biopsy and chemotherapy relative to the other two countries. These differences in provider behavior were caused by strong incentives for German hospitals and physicians to occupy hospital beds, which, in turn, were caused by the health care system structure and regulatory environment of Germany.

Comparing the U.S. with the U.K., however, yields an indeterminate conclusion on productive efficiency. Some differences in provider behavior – such as the

U.S.'s shorter LOS and use of all outpatient biopsy as opposed to the U.K.'s longer LOS and significant use of inpatient biopsy – caused the U.S. to be more productive relative to the U.K. Other differences – such as the U.K.'s limited mammographic screening program, use of FNA rather than surgical biopsy, use of the one-step biopsy/definitive treatment protocol, and lower staffing levels – caused the U.K. to be relatively more productive. These differences in provider behavior were caused by different physician and, to some extent, hospital incentives, as well as by physician supply constraints and capital controls in the U.K. U.S. physician and hospital competition for patients and payor competition for members encouraged adoption and coverage for mammographic screening and two-step biopsy to satisfy market demand. In the U.K., supply and budget constraints resulting from the global NHS budget restricted physicians and hospitals from investing in broad-based mammographic screening and led them to adopt FNA, but caused them to maintain the one-step biopsy/definitive treatment protocol.

Appendix 5A: Outcome measurement methodology

As described in Appendix 1E, the outcome measure we employed for breast cancer is based on the 5-year survival curves for each country. These survival curves came from cancer registries in each of the three countries.[26] This appendix describes the adjustments we made to those curves and how we arrived at the outcome statistic we used in our comparison of productive efficiency.

NORMALIZATION

The survival curves were normalized for age on the basis of the U.S. age distribution. We adjusted each country to have the same percentage of women under and equal to or over the age of 50. We did not adjust the curves to reflect different distributions of stage of disease and we only considered invasive cancers in our research. The primary reason for not adjusting for differences in the distribution of stages related to the presence of breast screening programs. Conceptually, effective screening programs allow cancers to be detected earlier, which likely leads to a shifting of stage distribution toward earlier stage cancers. By normalizing for stage, we would have removed this distribution shift and potentially masked any survival value it created. Since screening is a very significant part of breast cancer treatment, at least in the U.S., we decided to incorporate it into the input assessment. Thus, we were obliged not to remove its effect from the outcome side.

The presence of screening, however, introduces a set of biases in survival curves. Since we are calculating our outcome measure on the basis of these curves, we attempted to adjust the curves for these biases. This is discussed below.

ADJUSTMENT FOR SCREENING

Lead time bias is, perhaps, the most important bias that gets introduced to survival curves. Simple examples may be the best way to illustrate the point. Suppose an elderly woman is screened for breast cancer and is found to have a

---

[26] Sources: SEER; Thames Cancer Registry; Krebsregister – Saarland.

very small and slow-growing malignancy. If the malignancy remained undiscovered, she could die of causes other than breast cancer perhaps more than 5 years later. However, since her malignancy has been discovered, she becomes a part of the survival database and her "survival" biases the set as compared to a set that has no such cases. In fact, there is postmortem evidence of very early stage breast cancer in many undiagnosed women that potentially could otherwise have been identified and tracked by a comprehensive screening and tracking program.

An alternative scenario may be the case of two women, at time T_0, with identical cases of a slow-growing tumor. The first was caught early through mammographic screening, while the second was diagnosed a year later when she recognized some sort of symptom (e.g., a lump). For this hypothetical example, assume that the year delay in diagnosis (and presumably care) had no effect on the ultimate course of the disease and that both women died of the breast cancer 10 years after T_0. The screen-identified case would then have a survival of 10 years, while the symptomatically identified case would only be 9 years. This difference is an artifact of the screening program.

While there have been estimates for the amount of bias introduced through screening, there is no definitive way of removing it from survival statistics. We made a series of assumptions in order to adjust the survival curves. Our assumptions, in general, were conservative in that they tended to reduce the differences in the unadjusted survival curves, which, as seen in **Exhibit A-1**, are quite large between the U.S. and both Germany and the U.K.

Since there was virtually no screening going on in the U.K. during the time period of the analysis, the U.K.'s survival curve needed no adjustment. In Germany, we observed physical exam screening and "illicit" mammographic screening. [27] Since physical exam screening relies on detecting a symptom (e.g., a lump), it was unlikely to introduce much, if any, bias since the woman would likely detect the abnormality soon thereafter. In addition, as a conservative assumption, we ignored the illicit mammographic screening that was performed since we had no understanding of how effective it was in detecting cancers. Thus, we used the unadjusted survival curve in Germany as well. In the U.S., we ignored the physical exam screening and concentrated only on the mammographic screening. We estimated that about one-third of all cases were detected via screening and that these cases could have a lead time bias of about 3 years each. Thus, we adjusted the survival curve to reflect this estimate. See **Exhibit A-2** for the adjusted survival curves.

---

[27] "Illicit" refers to the fact that mammographic screening was not officially performed in the German public health system, yet there were many indications to the contrary.

LYs SAVED

To understand the LYs saved by treatment for breast cancer, the untreated survival curve must be known. With both the treated and untreated curves known, the LYs saved is simply the area between the two curves. However, we could not find, nor could we estimate, the untreated survival curve. Thus, we used the difference in area between a base country curve (i.e., the U.S.) and a comparison country curve as the basis for our LY-saved calculation (**Exhibit A-3**). This calculation can be interpreted as the incremental LYs saved (or lost) through treatment in one country versus the base country.

SUMMARY

Overall, the adjustment for screening and the method of outcome comparison results in differences that are quite small (i.e., within 10 percent). Given the precision of the methodology and based on the datasets we utilized, it is probably safest to conclude that there are no differences in the outcomes among the three countries. While we have used the values that we calculated via this methodology in the case report, the results would not change significantly if we were to use identical outcomes.

Appendix 5B: Sources

This list details data sources used in the breast cancer case study. We cover most of the main topics here, but this list is not exhaustive of all of the articles and government statistics that were employed throughout our work. In addition, we performed interviews with clinical and health care experts at a number of points during our study. Through these interviews, we collected qualitative and quantitative data on treatment patterns and checked our key assumptions and conclusions.

Below, we give the main sources used by topic.

OUTCOMES

U.S.

Surveillance Epidemiology and End Results (SEER) Public Use Database, National Cancer Institute.

U.K.

Thames Cancer Registry Database.

Germany

Krebsregister – Saarland.

INPUTS

Balch CM, et al. Clinical decision making in early stage breast cancer. Annals of Surgery 1993; 217: 207-225.

Lannin DR, et al. Difficulties in diagnosis of carcinoma of the breast in patients less than fifty years of age. Surgery, Gynecology and Obstetrics 1993; 177: 457-462.

Osteen RT, et al. Regional differences in surgical management of breast cancer. CA – A Cancer Journal for Clinicians 1992; 42.

Multiple interviews with clinicians, researchers, and educators in each country.

U.S.

Richards MS, et al. Breast biopsy rate and positivity in Rhode Island. American Journal of Public Health 1994; 84: 1310-1312.

SEER Public Use Database, National Cancer Institute.

U.K.

Foreman D, Rider L. Cancer in Yorkshire – Cancer Registry Special Report Series: 3. Breast Cancer, University of Leeds Print Service, May 1995.

McKinna JA, et al. The early diagnosis of breast cancer – A twenty-year experience at the Royal Marsden Hospital. European Journal of Cancer 1992; 28A: 911-916.

Sainsbury R, et al. Does it matter where you live? Treatment variation for breast cancer in Yorkshire. British Journal of Cancer 1995; 71: 1275-1278.

Thames Cancer Registry Database.

Germany

Arbeitsgruppe Zur Koordination Klinischer Krebsregister.

Großhadern Dataset, Tumorzentrum München.

Krebsregister – Saarland.

Mammakarzinome: Empfehlungen zur diagnostik, therapie und nachsorge. Tumorzentrum München, 1994.

ADDITIONAL INFORMATION

Cervical and Breast Screening in England, National Audit Office, February 1992.

European Journal of Cancer 1990; 26: 1167-1256.

Forrest P, et al. Breast Cancer Screening – Report to the Health Ministers of England, Wales, Scotland, and Northern Ireland, Department of Health and Social Security, 1986.

Health Care Financing Administration (HCFA) Bureau of Policy Development Office of Physician and Ambulatory Care Policy.
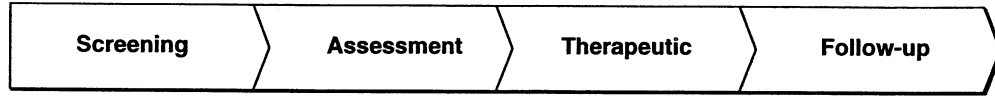
Medicare fee schedule database.

NHS Breast Screening Programme – Review 1993, NHS Breast Screening Programme, 1993.

Sandier S. Health Care Financing Review 1989.

Exhibit 1

PHASES OF BREAST CANCER
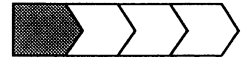## MANAGEMENT AND TREATMENT PROCESS

	Screening	Assessment	Therapeutic	Follow-up

	Screening	Assessment	Therapeutic	Follow-up
Purpose of phase	Detect potential malignancies as early as possible	Determine the nature of potential abnormalities in the breast	Intervene to remove or contain the cancer	Monitor for recurrence and treat as necessary
Comments/ rationale	In many cases survival is enhanced by early detection and treatment	Most breast abnormalities are not determined to be invasive cancer	For most cases (early stage cancer), there is a choice of equally effective surgical interventions	A small portion of those treated have recurrences, often within 2 years
Average percent of inputs consumed in each phase*	8%	44%	43%	5%

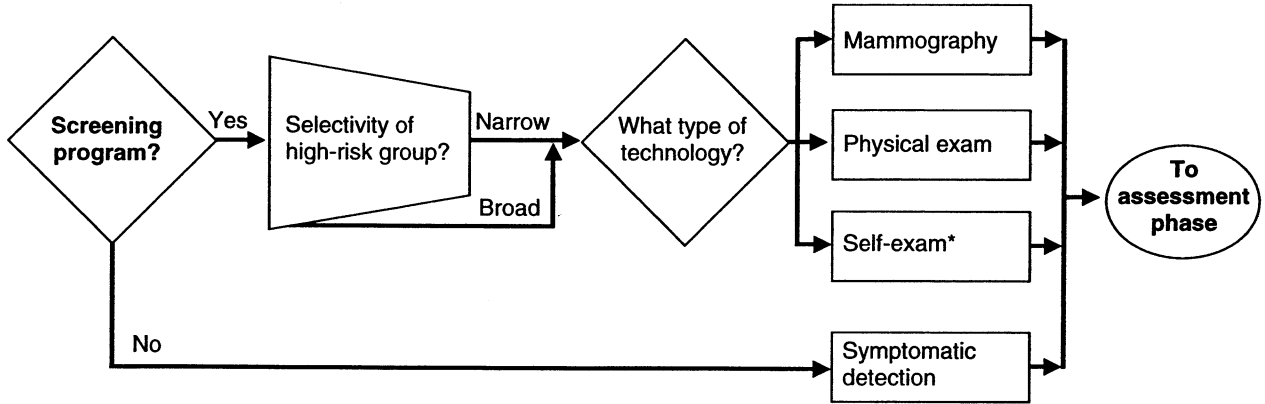* Average inputs consumed per phase in the U.S., U.K., and Germany

022 L 138266/5

Exhibit 2

OVERVIEW OF BREAST CANCER MANAGEMENT
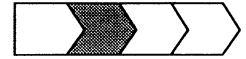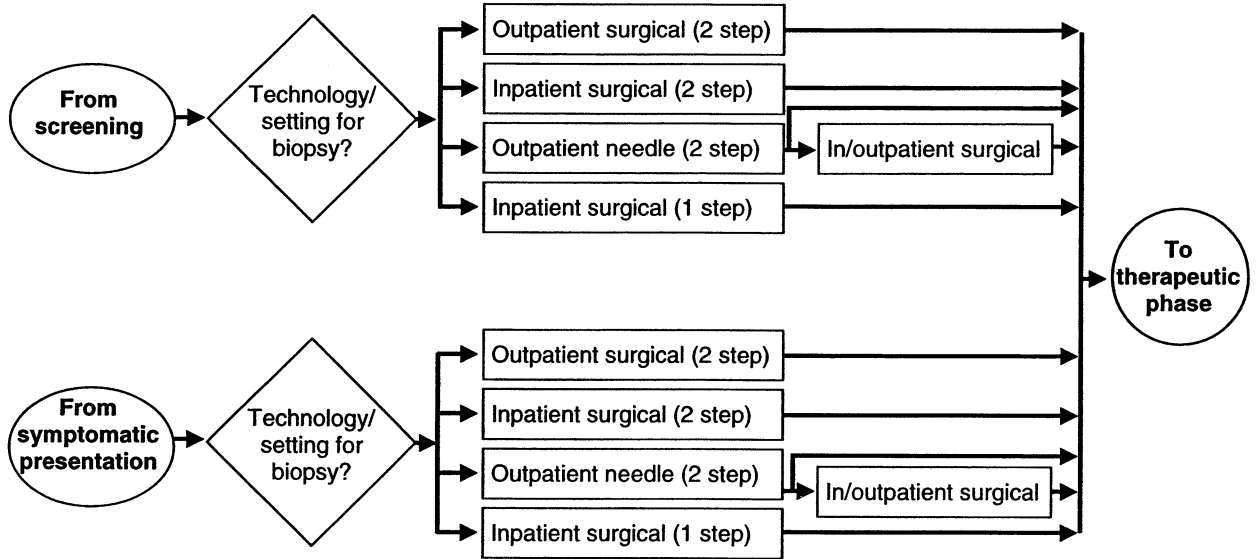# AND TREATMENT PROCESS
Screening phase

* Not considered in our economic analysis

022 L 138267/5

Exhibit 3

OVERVIEW OF BREAST CANCER
MANAGEMENT AND TREATMENT PROCESS
Assessment phase

022 L 138268/5

Exhibit 4

OVERVIEW OF BREAST CANCER
MANAGEMENT AND TREATMENT PROCESS
Therapeutic phase
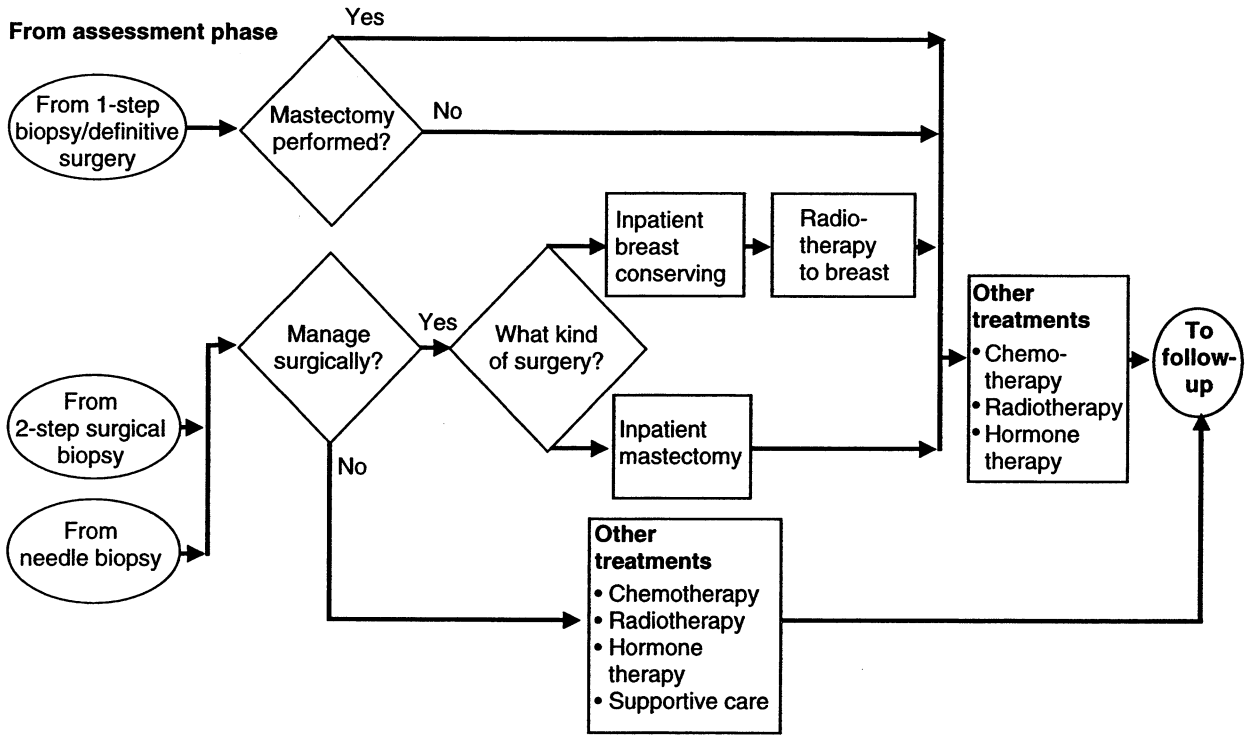
From assessment phase

022 L 138269/5

Exhibit 5

OVERVIEW OF BREAST CANCER
MANAGEMENT AND TREATMENT PROCESS
Follow-up phase

From therapeutic phase → What diagnostics to perform and how often? → Various imaging and diagnostic tests → Recurrence detected? —Yes→ What therapy? → • Surgery • Chemo-therapy • Radio-therapy → Successful? —No→ Death

Recurrence detected? —No→ (loop back to What diagnostics to perform and how often?)
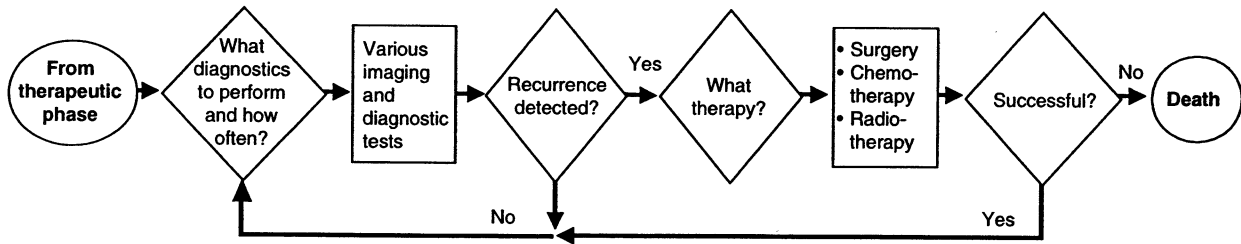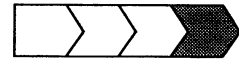
Successful? —Yes→ (loop back)

022 L 138270/5

Exhibit 6

INPUTS AND OUTCOMES PER CASE OF BREAST CANCER

Outcomes per case
Life years (LYs)

Inputs per case
Standardized input units

Inputs
Index, U.S. = 100%

Outcomes
Improvement in outcomes due to treatment
Index, U.S. = 100%

LYs

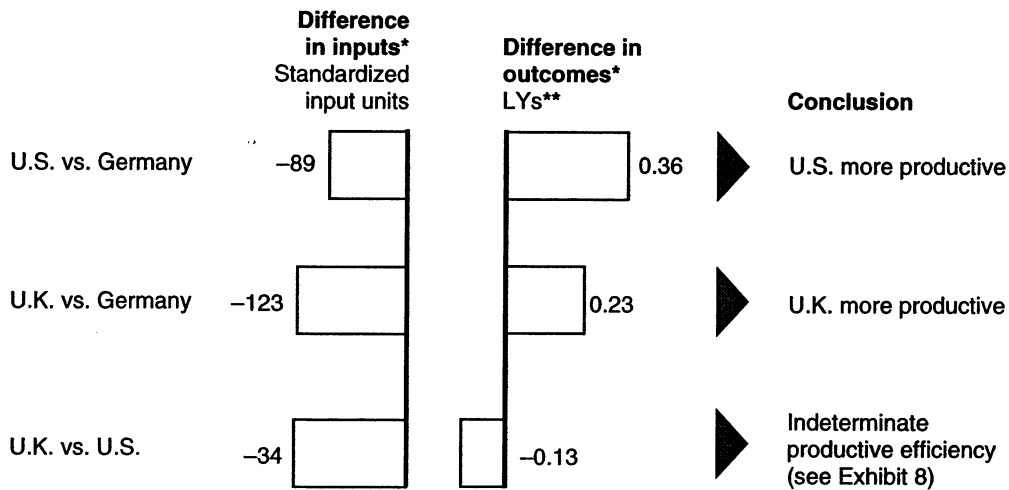|	Inputs	Outcomes	LYs
U.K.	85	63 / 97	0.23-3.73
U.S.	100	100	0.36-3.87
Germany	138	0 / 91	0-3.51
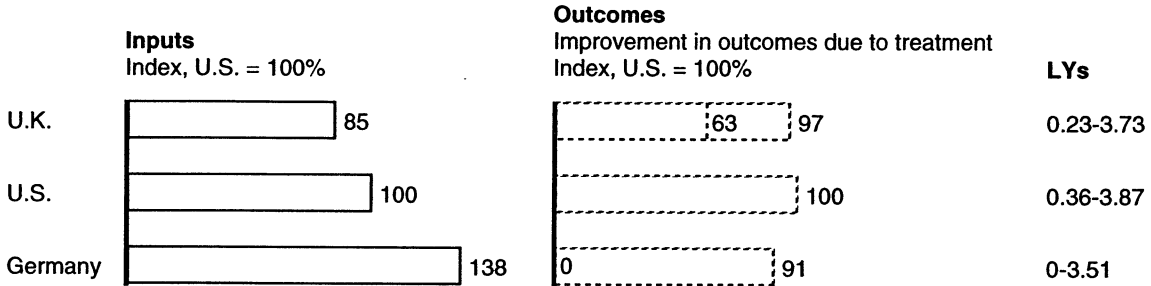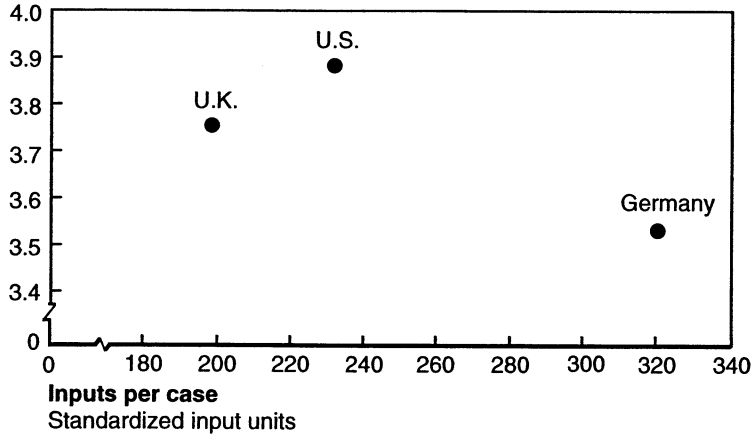
Note: Outcomes based on 5-year survival; baseline outcome with no treatment ranged from immediate death to lowest outcomes with treatment (Germany), thereby covering entire range of possibilities

---

Exhibit 7

ASSESSMENT OF RELATIVE PRODUCTIVE EFFICIENCY

	Difference in inputs* Standardized input units	Difference in outcomes* LYs**	Conclusion
U.S. vs. Germany	−89	0.36	U.S. more productive
U.K. vs. Germany	−123	0.23	U.K. more productive
U.K. vs. U.S.	−34	−0.13	Indeterminate productive efficiency (see Exhibit 8)

* Difference between number of units in first country to those in second country
** Calculated based on 5-year survival

022 ST 140055/5

Exhibit 8

ASSESSMENT OF RELATIVE
PRODUCTIVE EFFICIENCY: U.S. VS. U.K.

Countries compared	Disease	Difference in inputs Standardized input units	Difference in outcomes LYs	Average productivity advantage	Conclusion
U.K. vs. U.S.	Breast cancer	−34	−0.13	Can range from U.S. advantage of 35% to U.K. advantage of 13%*	• Productive efficiency indeterminate

Cost-effectiveness calculation

U.S. additional expenditure per LY
Dollars

U.S. prices   U.K. prices
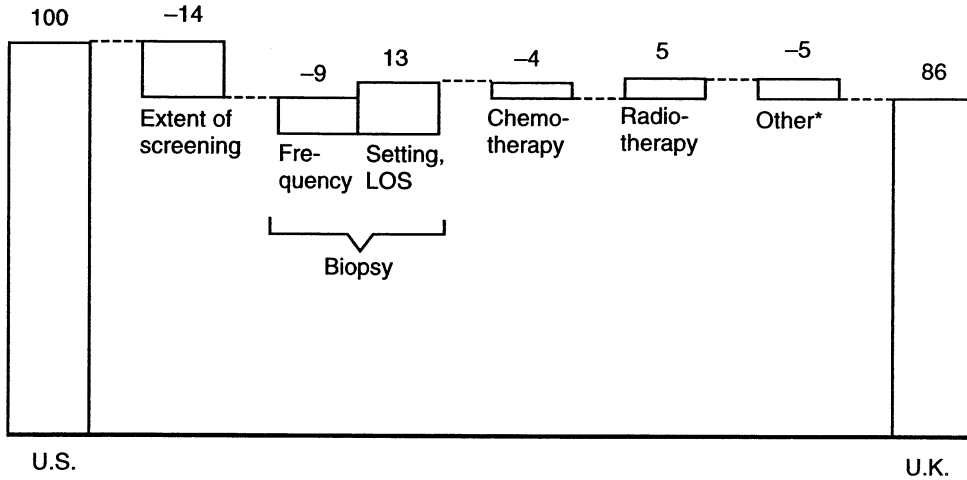
$32,000 $13,000

• U.S. had preferred input/outcome combination; U.K. could likely benefit from higher resource level

* These estimates cover the entire range of possible outcomes with no treatment: immediate death to lowest outcomes with treatment (which correspond to the outcomes with treatment in Germany)

Source: McKinsey analysis
022 AR 130773/5

Exhibit 9

SOURCES OF DIFFERENCE IN U.S. AND
U.K. INPUTS PER CASE – BREAST CANCER

Percent of difference


```
100        -14
                                        -4         5         -5
                              -9   13                                  86

        Extent of
        screening    Fre-    Setting,   Chemo-    Radio-    Other*
                     quency   LOS       therapy   therapy

                        Biopsy

U.S.                                                            U.K.
```
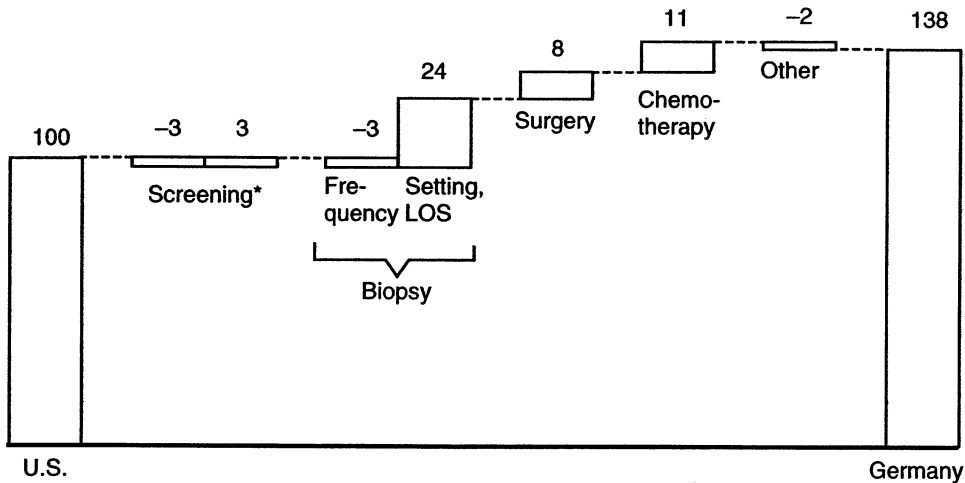
* Lower usage of diagnostics in both the assessment and follow-up treatment phases accounts for a significant portion of the "other" category

---

Exhibit 10

SOURCES DIFFERENCE IN U.S. AND
GERMANY INPUTS PER CASE – BREAST CANCER

Percent of difference ..


```
                                              11    -2    138
                                    8
                               24          Chemo-  Other
                                   Surgery therapy
100    -3   3     -3

     Screening*  Fre-   Setting,
                quency  LOS

                   Biopsy

U.S.                                               Germany
```

* In the U.S., lower input usage due to less mammography was offset by higher input usage due to more physical exams
022 SL 128974/5

Exhibit 11

IMPACT OF PROVIDER BEHAVIOR
DIFFERENCES ON PRODUCTIVE EFFICIENCY

● Major differentiating factor
◑ Secondary differentiating factor
○ Undifferentiating

Provider behaviors	U.S. vs. U.K.		U.S. vs. Germany	
Care triaging	●	U.S. broad-based screening program vs. no screening program in U.K.	○	
Treatment duration	◑	U.K. longer LOS than U.S.	●	Germany longer LOS than U.S.
Staffing levels	◑	U.K. lower staffing than U.S.	◑	Germany lower staffing than U.S.
Setting choice	◑	U.S. all outpatient biopsy vs. U.K. significant inpatient biopsy; U.K. 1-step biopsy/definitive treatment protocol vs. 2-step protocol in U.S.	●	Germany all inpatient biopsy vs. U.S. all outpatient biopsy; Germany greater portion of inpatient chemotherapy
Team-based approach	○		○	
Technology adoption	●	U.S. use of mammography for screening; U.K. use of fine needle aspiration (FNA) for biopsy vs. surgical in U.S.	○	

022 ST 140056/5

Exhibit 12

CAUSAL ANALYSIS: INCENTIVES AND CONSTRAINTS
U.S. vs. U.K.

● Major differentiating factor
◐ Secondary differentiating factor
○ Undifferentiating

Provider behaviors

	Extent of causal impact	Care triaging	Treatment duration	Staffing levels	Setting choice	Technology adoption
Incentives						
Physician	●	U.S. physicians had incentive to broadly screen patients			U.S. physicians had incentive to perform 2-step biopsy/definitive treatment protocol	U.S. physicians had incentive to adopt mammography and keep surgical biopsy vs. FNA, and use 2-step rather than 1-step biopsy/definitive treatment protocol
Hospital	◐	U.S. hospitals had incentive to broadly screen patients	U.S. hospitals had incentive to manage LOS	U.S. hospitals had incentive to keep LOS low, requiring higher staffing		U.S. hospitals had incentive to adopt mammography
Constraints						
Physician supply	◐	U.K. limited supply of physicians provided less time to screen patients			U.K. physician constraints led to maintenance of 1-step biopsy/definitive treatment protocol, keeping biopsy inpatient	U.K. adoption of FNA relieved constraints on physicians
Hospital supply	◐					U.K. adoption of FNA relieved constraints on hospital beds
Capital	◐	U.K. capital constraints precluded adoption of mammography for screening				U.K. capital constraints precluded broad adoption of mammography technology
Substitution	○					

Exhibit 13

CAUSAL ANALYSIS: SYSTEM STRUCTURE AND REGULATION
U.S. vs. U.K.

● Major differentiating factor
◐ Secondary differentiating factor
○ Undifferentiating

Provider incentives and constraints

	Extent of causal impact	Physician incentives	Hospital incentives	Physician supply	Hospital supply	Capital constraints
Health care system structure						
Product integration and pricing	◐	U.S. FFS vs. U.K. salary for specialist physicians encouraged broad-based mammographic screening and 2-step biopsy/definitive treatment protocol U.K. fixed salaries encouraged adoption of FNA and 1-step biopsy/definitive treatment protocol	U.K. annual hospital budget encouraged "bed-blocking" and therefore longer LOS; U.S. hospital case rates for Medicare encouraged shorter LOS			
Competitive intensity	●	U.S. physician competition for patients and payor competition for members encouraged adoption and coverage for mammography and 2-step biopsy/definitive treatment protocol	U.S. hospital competition for patients and payor competition for members encouraged adoption and coverage of mammography Some U.S. payor price-based competition put pressure on hospitals to manage LOS			
Regulation	●	U.S. physicians faced threat of malpractice suits	U.K. hospital budgets controlled and set through NHS global budget	U.K. physician supply strictly controlled	U.K. hospital supply strictly controlled	U.K. capital supply strictly controlled

Exhibit 14

CAUSAL ANALYSIS: INCENTIVES AND CONSTRAINTS
U.S. vs. Germany

● Major differentiating factor
◐ Secondary differentiating factor
○ Undifferentiating

	Extent of causal impact	Provider behaviors		
		Treatment duration	Staffing levels	Setting choice
Incentives				
Physician	◐	German physicians had incentive to keep LOS long to occupy hospital beds		Given hospital incentives, German specialist physicians had some incentive to keep procedures in inpatient setting
Hospital	●	U.S. hospitals had incentive to manage LOS; German hospitals had incentive to occupy hospital beds	U.S. hospitals had incentive to keep LOS low, requiring higher staffing	German hospitals had incentive to keep procedures in inpatient setting
Constraints				
Physician supply	○			
Hospital supply	●	High number of German hospital beds led to long LOS		
Capital	○			
Substitution	●			German hospital-based physicians not allowed to perform procedures in outpatient setting

Exhibit 15

CAUSAL ANALYSIS: SYSTEM STRUCTURE AND REGULATION
U.S. vs. Germany

● Major differentiating factor
◐ Secondary differentiating factor
○ Undifferentiating

	Extent of causal impact	Provider incentives and constraints			
		Physician incentives	Hospital incentives	Hospital supply	Substitution constraints
Health care system structure					
Product integration and pricing	●		German per diem vs. U.S. case rate product encouraged longer LOS		
Competitive intensity	◐		U.S. payor price-based competition for members and ability to negotiate differentially with individual hospitals encouraged hospital cost control; German payors had relatively restricted memberships along geographic or occupational lines and could not negotiate differentially with hospitals for price or other contract terms		
Regulation	●	German department chiefs could add private bed capacity if kept public bed capacity high	German hospitals faced capacity cuts if utilization below 85%	German hospital supply regulated through state government with political motives	German regulation enforced separation of inpatient and ambulatory care settings

Exhibit A-1
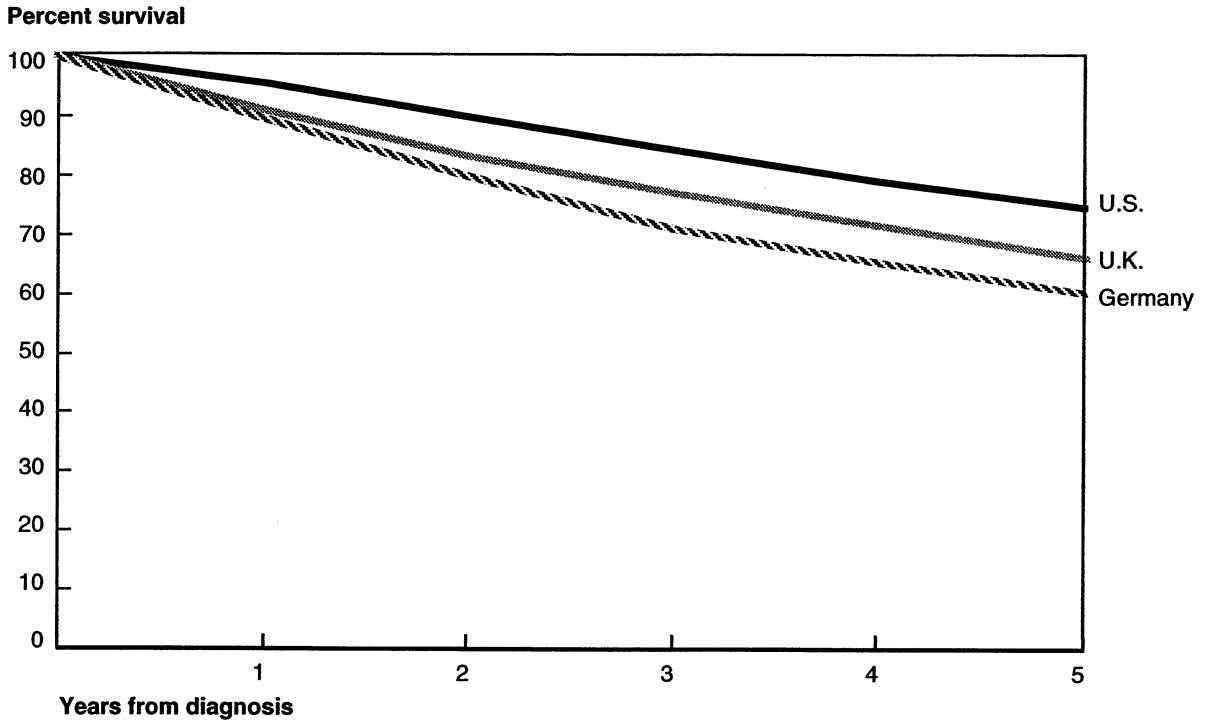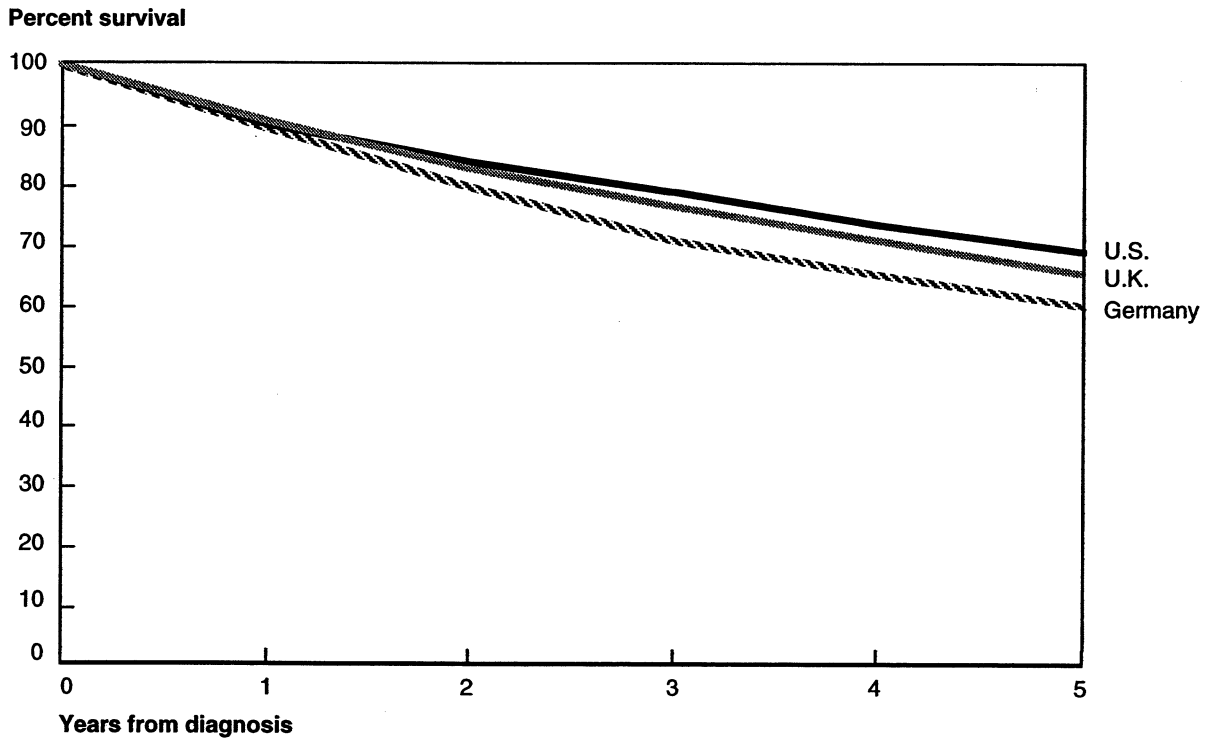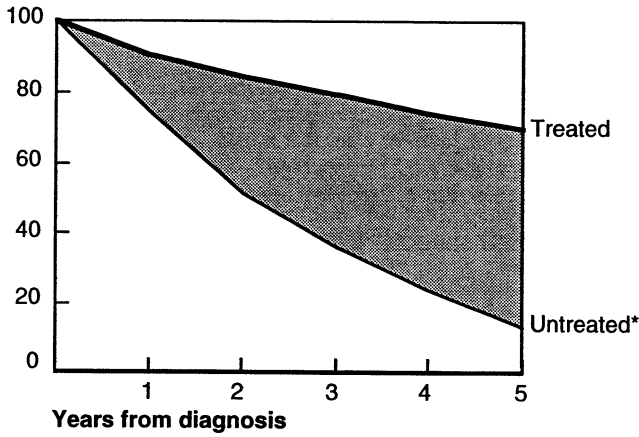**UNADJUSTED CUMULATIVE SURVIVAL CURVES**

Percent survival

Years from diagnosis

Exhibit A-2
ADJUSTED CUMULATIVE SURVIVAL CURVES

Percent survival

Years from diagnosis

022 ST 140074/5

COMPARISON OF OUTCOME METHODOLOGIES

Ideal comparison
Percent survival

Comparison without untreated baseline
Percent survival

Area between curves (shading) represents LYs saved through treatment
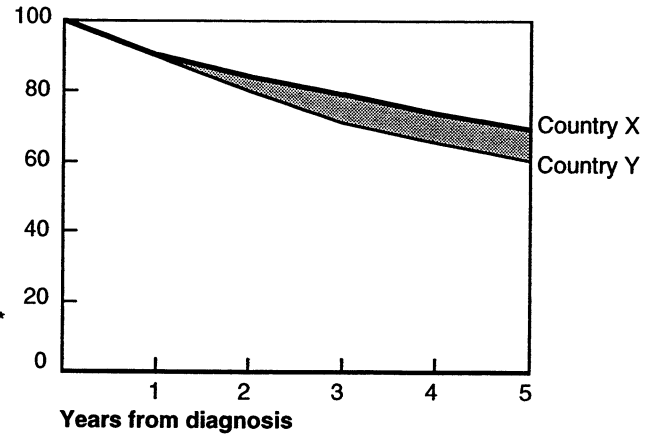
Area between curves (shading) represents incremental LYs saved or lost presumably due to treatment differences between two countries

* We could find no source for an untreated survival curve for breast cancer. This curve is purely to illustrate a methodology
022 ST 140076/5

Chapter 6: Lung cancer case

This chapter discusses the relative productive efficiency of Germany, the U.K., and the U.S. in the treatment of lung cancer.

We begin with an overview of the disease, a discussion of the productive efficiency measure used, and a description of the treatment process. After assessing the relative productive efficiency of these countries in the treatment of breast cancer, we analyze the provider behaviors driving these productive efficiency differences. Finally, we discuss how different health care system structures and regulatory environments affected provider incentives and constraints and, therefore, productive efficiency.

BRIEF DISEASE OVERVIEW

Lung cancer is a respiratory disorder that, for the majority of cases, is linked to cigarette smoking. It is the leading cause of cancer death in the U.S., U.K., and Germany.[1] In 1995, for the U.S. alone, lung cancer caused about 160,000 deaths.[2] The disease develops most often in scarred or chronically diseased lungs and is usually far advanced when 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 (SCLC) accounts for 20 to 25 percent of the cases and has a particularly poor prognosis, in part due to the rate of metastasis.[3] Non-small cell lung cancer (NSCLC) 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:

---

[1] Source: World Health Organization, 1987-1992.

[2] Source: Wingo et al., 1995.

[3] Metastases are tumors that form in parts of the body remote from the primary tumor and are the product of cancer spread originating from the primary site.

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.

DEFINITION OF PRODUCTIVE EFFICIENCY MEASURE

Timeframe of analysis

Our analysis of lung cancer focuses on treatment practices in roughly the mid to late 1980s. Because there are no national-level, population-based sources of comprehensive and detailed treatment and outcome information, we combined data from different subpopulations over the time period studied to build an aggregate picture of the course of lung cancer care and the associated outcomes in each country. By compiling information from many different sources, we believe that we created a reasonable picture of what was "typical" care for lung cancer patients in each country. (See Appendix 6B for a description of the major sources used.) The concept of typical care is, in a sense, an oversimplification since variation in treatment practices existed within each country. We did not determine that a single protocol was in effect for all patients in each country, but instead created more of a weighted average of observed treatments for the basis of comparison.

Although the analysis presented here relates to care in the 1980s, some aspects of care were in flux then and may have changed in the intervening years. Where we identified such changes, we have discussed them and their implications for the causal analysis of the underlying drivers of provider behavior and resulting productive efficiency.

Summary of disease management and treatment phases

We divided the management and treatment of lung cancer into three distinct phases (**Exhibit 1**): 1) diagnosis and staging; 2) curative care; and 3) 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 metastasis (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. Most cases are diagnosed at a point where the disease has spread from the primary tumor site, which results in an extremely poor prognosis. Consequently, a majority of lung cancer cases are treated with palliative care. 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). For the sake of brevity, we will refer to anticancer palliative care as palliative care and supportive palliative care as best supportive care (BSC).

Measurement of inputs

The input measure we used to compare resource consumption in treating lung cancer covers all the labor, capital, and supplies associated with the procedures performed in the three phases of management.[4] We excluded BSC in the palliative care phase, since no reliable data could be found as to the specific nature of BSC in each country. Thus, the inputs associated with BSC were not included in the comparison of total treatment resource consumption; we believe that this resource consumption was small and that differences among countries were likely insignificant.

Measurement of outcomes

As stated before, the prognosis for lung cancer is typically quite poor. The median survival is about a year, and only about 10 percent of cases survive through 5 years. It is generally believed that a 5-year survivor has a high likelihood of being cured of the disease. For these reasons, we chose an outcome measure of life years (LYs) saved based on the cumulative 5-year survival curve for the basis of our comparison of the relative effectiveness of lung cancer treatment care among the three countries.[5]

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 trade-off. The following example highlights this point. Suppose two patients with identical conditions were informed that they had terminal lung cancer and were offered palliative treatment that would extend their lives slightly, but at the expense of significantly reduced functional status. If one patient accepted palliative care and the other opted for BSC only, we would observe different outcomes based on patient preferences. Clearly, the outcome difference would not be a direct consequence of the structure and regulation of the health care system. Undoubtedly, a quality of life-adjusted

---

[4]   Refer to the discussion of input methodology in Appendices 1A through 1D.
[5]   Refer to the discussion of outcome methodology in Chapter 1.

outcome measure would handle this potential problem, but we were unable to use such a metric because the required data were unavailable. However, we believe it is reasonable to assume no significant differences in treatment preferences existed between countries and, therefore, believe our use of 5-year survival provides a reasonable basis for outcome comparison.[6]

We normalized the outcomes data in order to construct the 5-year aggregate survival curve. A discussion of this approach and other issues specific to the lung cancer outcome comparison can be found in Appendix 6A.

DESCRIPTION OF THE MANAGEMENT AND TREATMENT PROCESS

The following sections describe each of the three phases of the lung cancer management and treatment process in some detail and highlight the clinical and economic trade-offs implicit in the range of treatment options available in each phase. Later in this chapter, we discuss the specific practice patterns observed in each country and their implications for productive efficiency.

Diagnosis and staging phase

At its most basic level, this phase aimed to develop a treatment plan. The fundamental question providers and patients faced was whether to pursue a curative or palliative plan. Once this was decided, there were secondary questions around which specific protocol to employ.[7] The following two paragraphs create a context for understanding the major decisions in the diagnosis and staging phase of management.

About 90 percent of small cell patients have disease too advanced to warrant curative treatment. Of the remaining 10 percent, a very small portion would have been eligible for surgery; however, the majority of these cases were treated with aggressive chemotherapy. The overall prognosis for small cell carcinoma was, and continues to be, extremely poor. This, in part, is because small cell tumors apparently metastasize early after developing, making the disease systemic at the time of diagnosis and extremely difficult to eradicate. Because

---

[6] The availability of treatment options for terminal patients may affect the shape of the 5-year survival curve, but should not affect the percentage of cases that actually survive. This curve shape difference is due to the fact that in a resource constrained system, terminal patients are less likely to gain access to treatments like chemotherapy. Thus, they may die sooner, which changes the shape of the survival curve. These conditions may have been present in the U.K., so a small portion of the outcome difference between the U.K. and the U.S./Germany may be due to the availability of such life extending, but not life-saving, treatments.

[7] Diagnostic testing was also performed beyond the diagnostic and staging phase in the course of monitoring treatment effectiveness. For example, the U.S. would typically monitor the effectiveness of chemotherapy by assessing tumor shrinkage through a CT scan or X ray.

small cell carcinoma responds to chemotherapy, a chemotherapeutic protocol was commonly employed. Non-small cell carcinoma is thought to be quite different, relative to the speed at which it metastasizes. This, and the fact that it does not respond as well as small cell carcinoma to chemotherapy, made surgery the therapy of choice when the tumor was identified sufficiently early.

The surgical excision of the non-small cell tumor represented the best hope for curing the disease, and almost all survivors underwent a surgical procedure. However, surgery was no guarantee of cure, since a majority of surgical cases did not survive the disease. The outlook was even more grim for radiotherapy and chemotherapy, either alone or in combination. Almost no patients survived the disease with one or both of these treatments instead of surgery. Thus, the challenge for providers in dealing with both small and non-small cell carcinoma was to identify the cases that could most benefit from aggressive care and manage them under a curative intent protocol. The remaining group of patients then was managed under a palliative care protocol.

Ideally, information gathered in the diagnosis and staging phase of management would have allowed confirmation of the malignancy, identification of cell type, and determination of the extent of the spread of disease. This, along with general information concerning the overall condition of the patient, would be utilized in matching the patient to the appropriate treatment protocol. There was a range of diagnostic tools in use to gather this information, which included sputum cytology, X ray, various forms of biopsy, internal optical imaging (i.e., bronchoscopy and mediastinoscopy), computerized tomography (CT scan), ultrasound, and magnetic resonance imaging (MRI) (**Exhibit 2**). While the diagnosis of the cancer from the sputum or a biopsy sample may have been relatively straightforward, accurate staging often required significant and potentially expensive diagnostic work.[8] Thus, there was an immediate trade-off in terms of the extent of diagnostic work and the value of the information gained in selecting which treatment protocol to use.

The confirmation of the malignancy and identification of the cell type most often would have resulted from a microscopic examination of at least a sample of the sputum and, more likely, a tissue sample from a biopsy or brushing. CT scans were frequently performed early in the diagnostic workup to identify locations for biopsy and to assess the appropriateness of other procedures. Cell samples were retrieved through a number of procedures, including needle biopsy and

---

[8] Staging is a determination of the extent and location of cancer spread from its primary site in the lung. Local stage corresponds to a relatively contained tumor; regional stage in general indicates that the cancer has spread from the primary site to the mediastinal lymph nodes; distant stage is reserved for those cancers that have spread to other areas of the body such as the brain, bone, or liver. Note that due to the timeframe of our analysis and to the availability of the data, we utilized the local/regional/distant staging methodology instead of the American Joint Committee for Cancer (AJCC) methodology which is in common use today.

biopsy during the course of bronchoscopy or mediastinoscopy.[9] Needle biopsy and bronchoscopy often would not require an overnight hospital stay; however, mediastinoscopy would be performed under general anesthesia and typically required a short hospital stay.

The search for distant metastases to confirm distant stage cancer was accomplished through various imaging diagnostics. While not all of these were used under all circumstances, the set of diagnostics included X ray, CT scan, MRI, and ultrasound under some circumstances. These distant metastases were relatively easy to identify and almost always indicated a terminal prognosis. Therefore, patients found to have distant metastases normally received palliative care.

It was typically more difficult to determine whether a case should be classified as local or regional if distant metastases were not detected. Regional stage carcinomas are characterized by the spread of disease from a primary lung site into the lymph nodes of the mediastinum. Patients diagnosed with mediastinal lymph node metastases have a significantly reduced 5-year survival rate after surgical treatment as compared to those with localized surgically treated disease. Clinicians and institutions differ on their level of aggressiveness on NSCLC regional cases. On the conservative end of the spectrum, some surgeons will not consider resection for cases with any mediastinal lymph node involvement. The middle ground is defined by surgery on cases where the mediastinal involvement is limited to the ipsilateral nodes (i.e., the mediastinal nodes involved are on the same side of the chest as the primary tumor). The aggressive side of the spectrum is characterized by surgery for all cases of mediastinal involvement (i.e., including contralateral nodes) as long as all apparent disease can be removed. While surgery for SCLC is much less frequent than for NSCLC, it too is evaluated in light of mediastinal involvement. No matter what the approach, mediastinal involvement is an important piece of information in the determination of surgical candidacy and, therefore, curative or palliative care.

The most accurate method of staging was surgical-pathological staging, in which the chest was opened and the extent of cancer spread was assessed through a thorough visual, manual, and microscopic examination. However, this procedure consumed a large quantity of resources and caused significant morbidity. Thus, less intensive staging was typically employed to aid in identifying mediastinal involvement and to serve as the basis for determining surgical candidates. During the time period of our analysis, there were conflicting points of view regarding the value of CT scanning in assessing mediastinal involvement. There was both research concluding that CT scanning was better than a conventional chest X ray and research that concluded that it

---

[9] Biopsy during mediastinoscopy was relatively common, but probably would not be used in the primary diagnosis of lung cancer or determination of cell type. More likely, the results would be used to assess the stage of the disease.

was no better.[10] There was no consensus on what "image" constituted likely node involvement as researchers varied in their use of decision rules. For instance, a "likely node involvement" could be defined as any node with a maximum dimension exceeding 1.0 cm or 1.5 cm, etc. The eventual resolution of the issue came when it was determined that while CT scans offered clearer visualization of the mediastinal nodes compared to X ray, this advantage was offset by the fact that enlarged nodes did not always signify involvement and normal nodes sometimes carried micrometastases. However, CT scanning of the nodes was better than X ray in identifying those cases without node involvement (i.e., CT scans have a higher negative predictive value than X rays).[11]   Thus, if a conservative practice called for surgery only for local stage cases and X ray was employed instead of CT scanning to identify these surgical candidates, there was the potential to identify a suboptimal group for surgery. This is because a group of presumed local cases assessed using X-ray technology will have more true regional stage cases (i.e., false negatives) than a similar group assessed using CT scan technology. This greater mix of regional stage cases could, therefore, result in an overall difference in surgical outcome between those groups assessed with X rays versus CT scans.

During the time period of this analysis, CT scans were commonly used in the U.S. and Germany to supplement X rays in the diagnosis and staging of, and treatment planning for, lung cancer patients. In addition, mediastinoscopies were used to provide more information on the involvement of the mediastinum, typically after a CT scan showed an enlarged mediastinal lymph node.

All of these diagnostic procedures, along with basic assessments such as blood screens and physical exams, were in use during the time period of our analysis. However, as we discuss later, the observed extent of staging varied within and among countries. These differences influence the inputs consumed in the management and treatment of lung cancer since some of these procedures, like inpatient mediastinoscopy, utilize significant resources. In addition, it appears that the amount of information collected in the diagnosis and staging phase, as well as the consequences of the information on treatment choice, affected outcomes. While this is obviously a more difficult link to establish, compared with the obvious cause and effect relationship on the input side, we believe there is sufficient evidence to suggest a relationship.

Curative care phase

The definition we use to differentiate curative from palliative care is somewhat arbitrary in the sense that there would likely be a debate around what constitutes a curative care protocol. However, our definitions are neutral to the calculation

---

[10] Sources: Richey et al., 1984; McKenna et al., 1985.

[11] Source: McKenna et al., 1985.

of total resource consumption and for our purposes, therefore, need not be of concern. In general, all treatments administered to early stage cases and cases managed under a surgical protocol were considered in the category of curative care. [12]

In general, curative care for non-small cell cases often involved surgery, followed by chemotherapy, radiotherapy, or both. Nonsurgical cases were typically treated with intensive radiotherapy, chemotherapy, or both (**Exhibit 3**). Clinical research was just beginning to report the benefits of chemotherapy prior to surgery or intensive radiotherapy at the end of the time period we studied. We observed very little of this induction chemotherapy in the various data sources we utilized to compile an aggregate picture of care in each country.

Small cell cases were managed through the same three major interventions of surgery, chemotherapy, and radiotherapy. However, chemotherapy played a much more important role in the care protocols. Surgical management was rarely employed; and radiotherapy was secondary to chemotherapy in importance, although used almost as frequently.

The variations in management protocols in use led to differences in resource consumption. Not surprisingly, the high-consuming events were those administered in the inpatient setting. Of all inpatient care, radiotherapy consumed the most resources due to its long associated length of stay (LOS). However, only a small amount of radiotherapy was delivered in the inpatient setting and was typically reserved for those too weak to travel to and from the hospital especially when great travel distances were involved. Because a typical administration of radiotherapy doses (called "fractions") would last over a period of 4 to 5 weeks, the total resource requirement when delivered in the hospital made inpatient radiotherapy the single largest resource-consuming event utilized in lung cancer care.

All surgical procedures in each of the three countries were administered in the inpatient setting and had different associated hospital stays. Due to the length of these stays, which ranged up to almost 2 weeks, the resource consumption associated with surgery was quite large. Chemotherapy was administered in both the inpatient and outpatient settings. Some chemotherapeutic agents weakened the body's ability to fight infection and, therefore, warranted inpatient care. Given that a typical chemotherapy protocol called for one course of care per month for each of about 4 months, the total LOS was about 10 days. Total resource consumption for a typical inpatient chemotherapy protocol was large, yet the lowest of the three inpatient events discussed.

Radiotherapy and chemotherapy administered on an outpatient basis required considerably fewer resources compared with the inpatient versions of the

---

[12] There were a small number of early stage cases that only received supportive care. Most readers would not consider the care administered to these cases as curative in intent.

various protocols. As was the case in the inpatient setting, radiotherapy required more resources than chemotherapy. Other than the frequency differences of inpatient and outpatient radiotherapy and chemotherapy, protocols varied according to the number of fractions of radiotherapy and the number of courses of chemotherapy employed. Although the number of fractions used to deliver radiotherapy differed by country, the total levels of radiation specified in the major protocols (e.g., postoperative radiotherapy for local stage NSCLC) was relatively standardized. For instance, a radiation treatment of 3,000 rads could be delivered in over 6 weeks in 100-rad-daily fractions or over 4 weeks in 150-rad-daily fractions. Within a reasonable fraction range, there was no known difference in the efficacy of radiation therapy in eliminating cancerous cells. However, the higher fraction, lower session protocols potentially increased the side effects of radiation therapy. There is a clear economic impact inherent in the decision of which protocol to employ, given that the fewer the number of fractions, the lower the total input resources consumed. This is due to the fact that a session of radiotherapy consumes about the same amount of labor, capital, and supplies regardless of the amount of radiation delivered.

Palliative care phase

The purpose of palliative care is to provide treatment that, as much as possible, ameliorates the symptoms of lung cancer. Chemotherapy and radiotherapy are the primary interventions (**Exhibit 4**). Both therapies can shrink the size of an inoperable tumor or a metastastic mass that may be causing symptoms (e.g., a painful bone metastasis or a lung passage constricting mass). As in the curative treatment phase of the disease, these two interventions can be delivered in ways that create differences in input factor consumption. Furthermore, the frequencies by which the two therapies, along with BSC, are utilized contributes to differences in input consumption.

* * *

From a clinical perspective, the development of an appropriate treatment plan for lung cancer is dependent upon a range of factors. Cell type, whether small or non-small, influenced the extent to which surgery or chemotherapy was emphasized. Cancer staging, determined through the range of tests outlined above, informed treatment intent and the mix and order of treatment options offered. Finally, the patient's overall condition was an important input into the assessment of his or her ability to tolerate one or more of these debilitating treatments.

The overall resource consumption was dependent on a number of factors, including the frequency and type of treatments employed and the duration and setting of these treatments. There were differences in all of these factors across

the three countries, which introduced variances in the total amount of resources consumed and the resulting outcomes in the management of lung cancer.

ASSESSMENT OF RELATIVE PRODUCTIVE EFFICIENCY

U.S. versus Germany

Germany used 21-percent more inputs and achieved 12-percent worse outcomes than the U.S. in the treatment of lung cancer (**Exhibit 5**). With better outcomes and fewer inputs, the U.S. was clearly more productive than Germany in lung cancer treatment (**Exhibit 6**).

U.S. versus U.K.

The U.K. used 24-percent fewer inputs and achieved 58-percent worse outcomes than the U.S. in the treatment of lung cancer (**Exhibit 5**). As discussed in Chapter 1, when one country had both higher inputs and better outcomes, we assessed which nation had the higher productive efficiency by calculating average productivity and understanding the disease treatment process. Measuring average productivity requires comparison of each nation's outcomes *with* treatment to outcomes *without* treatment. Using 5-year survival curves to determine each nation's outcome with treatment and a baseline estimate of 3.8 months, the average productivity was determined, resulting in 82-percent higher average productivity in the U.S. than the U.K. (**Exhibit 6**).[13] As the disease treatment process in lung cancer did not appear to exhibit increasing returns at the positions of the U.K. and the U.S., we conclude that the U.S. was more productive than the U.K. in lung cancer treatment.

U.K. versus Germany

Germany used 59-percent more inputs and achieved 110-percent better outcomes than the U.K. in the treatment of lung cancer (**Exhibit 5**).[14] Based on 5-year survival, Germany had 33-percent higher average productivity than the U.K. (**Exhibit 6**). Because the disease treatment process in lung cancer did not appear

---

[13] Sources: Surveillance Epidemiology and End Results (SEER); 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 (i.e., patients who received only BSC); these results are the basis for our 3.8 month estimate. Due to their condition, 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.

[14] Calculation based on ratio of German results to those in the U.K.

to exhibit increasing returns, we conclude that Germany was more productive than the U.K. in lung cancer treatment.

MAJOR DRIVERS OF PRODUCTIVE EFFICIENCY DIFFERENCES IN TERMS OF PROVIDER BEHAVIOR DIFFERENCES

The productive efficiency differences observed were caused most directly by differences in provider behavior. In this section, we discuss the provider behaviors that were the major drivers of productive efficiency differences between the nations, focusing first on drivers of input differences.

Summary of drivers of input differences

U.S. versus U.K. Four behaviors influenced the overall differences in input consumption and outcomes between the U.S. and U.K. (**Exhibit 7**). The first was the frequency of diagnostic procedures. The other three related to the frequency of treatments and the associated LOS and staffing. Of the three treatments, radiotherapy was most important, followed by surgery and chemotherapy, in terms of introducing the most variance in overall resource consumption. The greater surgical frequency in the U.S. likely led to the U.S.'s better outcomes.

U.S. versus Germany. Three major differences influenced the overall differences in input consumption between the U.S. and Germany (**Exhibit 8**). These included surgery, radiotherapy, and chemotherapy practice variations, similar to the U.S. and U.K. differences.

Summary of drivers of productive efficiency differences

U.S. versus U.K. Of the six provider behavior categories introduced in our causal framework in Chapter 1, five led to offsetting differences in productive efficiency between the U.S. and the U.K.: care triaging, treatment duration, staffing levels, setting choice, and technology adoption (**Exhibit 9**). The U.S.'s productive efficiency relative to the U.K. was raised by its shorter hospital LOS for surgery and its substitution of outpatient for inpatient chemotherapy, as well as by its greater use of CT scans in diagnosis and staging. While higher staffing levels diminished the U.S.'s productive efficiency relative to the U.K., the net result of differences in treatment was higher productive efficiency in the U.S.

U.S. versus Germany. Again using our six categories, three major differences in provider behavior led to somewhat offsetting differences in productive efficiency between the U.S. and Germany: treatment duration, staffing levels, and setting choice (**Exhibit 9**). Germany's productive efficiency relative to the U.S. was lowered by its longer hospital LOS and its greater use of the inpatient setting for

chemotherapy. While Germany's productive efficiency relative to the U.S. was raised slightly by Germany's lower staffing levels, the net effect of provider treatment differences led to higher productive efficiency in the U.S.

Below, we discuss in greater detail the major provider treatment variations that led to differences in inputs and productive efficiency by each phase of the lung cancer management and treatment process.

Diagnosis and staging phase

Differences in the frequency and type of diagnostic testing had a significant effect on differences in overall input consumption and productive efficiency.[15] At the time of our analysis, the resources consumed during the diagnosis and staging management phase, on average, accounted for nearly 20 percent of total resources devoted to lung cancer care. In general, the U.K. performed fewer diagnostic tests per lung cancer patient than did the U.S. or Germany. The most important differences in behavior were in the areas of CT scans, endoscopic exams, and biopsy, where the U.K. appears to underinvest relative to the U.S. and Germany.

The importance of the CT scan in aiding biopsy and assessing the mediastinum has already been discussed. Only about 20 percent of cases in the U.K. were assessed with a CT, compared with 80 percent and close to 100 percent for the U.S. and Germany, respectively. Bronchoscopy with biopsy in the U.K. was performed in about 70 percent of cases. The U.K. percentage was also lower than in the U.S. and Germany, where 90 percent and close to 100 percent of cases, respectively, received bronchoscopy and biopsy. All countries performed a relatively low level of mediastinoscopy, with the U.S. at 15 percent and both the U.K. and Germany at approximately 10 percent. [16]

While all of these diagnostic activities consumed resources, mediastinoscopy was the most resource intensive because it was performed in an inpatient setting and typically required a 2-day LOS. This resource consumption was partially offset by the timing of the mediastinoscopies; many were performed immediately prior to surgical resection of the lung. Under these circumstances, the procedure would not add to the overall LOS, and the incremental resource consumption would be reduced to those resources directly associated with the procedure (e.g., extra surgical time).

---

[15] Sources: Edinburgh Lung Cancer Group, 1987; Humphrey et al., 1990; Scotland data, unpublished; clinician interviews. The frequencies of CT scanning, bronchoscopy, and mediastinoscopy are reported in these sources.

[16] Sources: Edinburgh Lung Cancer Group, 1987; Humphrey et al., 1990; Scotland data, unpublished; clinician interviews. We observed differing points of view as to the frequency of mediastinoscopy in Germany. The 10 percent represents our "best estimate" of these views.

In the U.S., the diagnosis and staging phase accounted for about 21 percent of all resources devoted to lung cancer. In comparison, the U.K. and Germany both utilized about 18 percent. Relative to the U.S., the U.K. and Germany consumed 8-percent fewer and 1-percent more resources, respectively.

Curative care phase

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 U.S. and U.K. followed with 22 percent and 13 percent, respectively. The average LOS in Germany and the U.K. was about 75-percent longer than that in the U.S. Finally, the staffing levels were highest in the U.S., followed by Germany and the U.K. with about the same hospital staffing content. Overall, surgery in the U.S. accounted for about 25 percent of all spending on lung cancer treatment. The U.K. was slightly higher at 26 percent, while Germany was highest at 33 percent. With these different approaches to surgical frequency, hospital staffing, and recovery LOS, we observed different levels of input usage and productive efficiency. Relative to the U.S., the U.K. and Germany consumed 25-percent fewer and 60-percent more resources, respectively, on lung cancer surgery.

The drivers of resource consumption differences associated with radiotherapy were similar to those driving differences in surgery. However, most radiotherapy was performed in the outpatient setting, which meant that the LOS and staffing factors were of less importance than for surgery. Providers in both the U.S. and Germany utilized radiotherapy with curative intent for about 20 percent of all lung cancer patients. The frequency in the U.K. was considerably lower, with about 7 percent of patients receiving curative intent therapy. We estimated that, in each country, about 10 percent of therapy was delivered in the inpatient setting. Due to differences in the radiotherapy protocols (i.e., different numbers of fractions employed), the U.K. LOS for inpatient radiotherapy were about one-third shorter than those for both the U.S. and Germany. The staffing levels were similar to those for surgery; the U.S. was highest, followed by Germany and the U.K. Overall, curative intent radiotherapy in the U.S. accounted for about 12 percent of all resource consumption on lung cancer treatment. In the U.K., the corresponding

level was 6 percent, while in Germany, it was about 8 percent. Relative to the U.S., the U.K. and Germany consumed about 60-percent fewer and 25-percent fewer resources, respectively, on curative intent radiotherapy.

Curative intent chemotherapy was a relatively minor part of the curative care management phase in that its frequency was quite low in all three countries. However, there were considerable differences in terms of the setting of care. Both the U.K. and Germany utilized the inpatient setting far more than did the U.S. This led the overall resource consumption for the U.K. and Germany to be about 120-percent and 90-percent greater, respectively, than that in the U.S. While this difference seems large, it should be considered in light of the low frequency of curative chemotherapy. Overall, curative intent chemotherapy in the U.S. accounted for about 2.5 percent of all spending on lung cancer treatment. The U.K. and Germany were higher at 6.5 percent and 3.5 percent, respectively.

Palliative care phase

The palliative care management phase, on average, accounted for about 40 percent of total resources devoted to lung cancer care. In general, the behavior toward palliative care paralleled the behavior toward curative intent care in each country, with only a few exceptions. Because the behaviors in this phase were so similar to the curative care phase, we will only address the exceptions.

One obvious exception is that little, if any, surgery was categorized as palliative in intent. The other major difference was in the frequency of treatment. For both radiotherapy and chemotherapy, the frequency of treatment was much higher than in the curative care phase because more cases were managed with palliative intent. In addition, the frequency of palliative chemotherapy in the U.K. was significantly lower than that of both the U.S. and Germany. However, because the U.K. utilized more inpatient chemotherapy than the U.S., the U.S. consumed the fewest resources on palliative chemotherapy. Refer to **Exhibit 10** for a summary of the drivers of resource consumption differences for both the curative and palliative care phases. Note that overall consumption differences are often the consequence of offsetting provider behaviors.

Finally, it is also important to note that many cases in each country were managed under a BSC protocol. The frequency of this management was highest in the U.K., where about 35 percent of cases received BSC. The U.S. and Germany were 16 percent and 9 percent, respectively. As stated previously, we believe that the input consumption related to BSC was small. In addition, these frequencies represent instances where BSC was the *only* treatment received. The actual frequencies of supportive care were probably higher and more similar among countries. Thus, we do not believe that the differences in BSC usage introduced much, if any, difference in input consumption.

CAUSAL ANALYSIS OF PROVIDER BEHAVIOR DIFFERENCES: INCENTIVES AND CONSTRAINTS, SYSTEM STRUCTURE, AND REGULATION

The differences in provider behavior in the U.S., U.K., and Germany can be explained by the incentives and constraints providers faced in each country and by underlying differences in the health care system structure and regulation, which are described in detail in Chapter 2. Below, we explain how these differences led to the provider behavioral differences we observed in the treatment of lung cancer.

U.S. versus U.K.

As discussed, productive efficiency differences between the U.S. and U.K. were caused by differences in three areas of lung cancer treatment: diagnostic procedures, chemotherapy, and inpatient procedures. Diagnostic testing practices had a large impact on productive efficiency due to differences in care triaging (i.e., the extent of diagnostic testing) and the technology used (e.g., CT scans). Chemotherapy practices had a large impact on productive efficiency due to differences in the choice of care settings. Inpatient procedure practices influenced productive efficiency due to differences in LOS and staffing levels. In this section, we discuss each of these in turn. We also discuss differences in frequency of surgery, radiotherapy, and chemotherapy, which did not necessarily lead to differences in productive efficiency, but led to differences in outcomes and resource consumption.

Diagnostic testing: triaging and technology used

Provider incentives and constraints (Exhibit 11). The U.K. performed 40-percent fewer diagnostic tests for lung cancer than the U.S. This potentially lowered the U.K.'s relative productive efficiency by giving providers less information for staging, which, in turn, suggests that they would have been less able to identify optimal candidates for surgery.

In the U.S., both physicians and hospitals had incentives to assess patients and adopt CT scans to do so. Since there were no major regulatory barriers to the purchase of capital equipment, CT scans for diagnostic assessment could be easily purchased. The availability of capital and physician and hospital incentives, combined with an ample supply of physicians and hospitals, therefore led to broad use of CT scans for diagnostic purposes in lung cancer patients. In the U.K., however, physicians and hospitals did not have incentives to perform diagnostic testing and constraints on physician supply gave physicians less time to assess patients. In addition, capital constraints precluded the broad adoption of CT technology.

The lower rate of diagnostic testing in the U.K. was also the result of the more restricted referral process. Based on interviews with clinicians, we have reached the following understanding for how a typical lung cancer patient moved through the U.K. health care delivery system.[17] The National Health Service (NHS) stipulated that patients with nonemergency conditions enter the system by visiting a GP. For the typical lung cancer patient, the first interaction with providers would have been at this level. In fact, many patients were treated exclusively by GPs and were never referred to a hospital where they could be seen by a chest physician or surgeon. Once referred to a chest physician, the patient could be treated and sent back to the GP or referred to the thoracic surgical center in a regional hospital. For those patients which they referred, GPs sent an overwhelming majority on to chest physicians. Chest surgeons likewise received the great majority of their referrals from chest physicians (as did radiotherapists). This system of referrals served to limit access to the relatively limited supply of specialist physicians, hospital beds, and expensive capital equipment (most of which were only located in the regional center) and to protect these resources from potential overuse. Since relatively few patients were able to make it through these two levels of referral, the frequency of CT scanning, as well as surgery, was reduced. This referral process and access control created a substitution constraint for U.K. physicians; for example, the community could not have a clinic that provided CT scans that the patient could access directly or through a GP referral since this substitute was not possible in the U.K.

Due to this substitution constraint, some good candidates for surgery in the U.K. may have been prevented from reaching the regional surgical center before a CT scan was performed. Thus, the decision to exclude surgery from treatment for some potentially attractive surgical candidates might have been made without the added information afforded by a CT scan. In addition, some cases that reached the regional surgical center may have undergone surgery without having undergone scanning due to limited CT supply. This group had a raised likelihood of mediastinal involvement that went unnoticed because the scan was not performed. This mediastinal involvement would lead to a poorer prognosis and a lower likelihood of survival. Thus, the lower outcome in the U.K. may have resulted from a less effective triage of surgical candidates, as well as the U.K.'s lower surgical frequency, described below. While we have evidence that the surgical frequency was lower in the U.K., we have no direct evidence that the frequency of surgery on understaged cases was higher in the U.K. compared to the U.S.; instead, we have inferred this second explanation for differences in outcomes from the lower observed frequency of CT usage. As previously stated, lower CT usage in the assessment of mediastinal lymph node involvement, in combination with a

---

[17] Sources: Surveillance Epidemiology and End Results (SEER); Joslin and Rider, 1993. In 1992, an audit of referral patterns for lung cancer services was conducted. We utilized the results to supplement our interviews.

conservative practice of operating only on local stage cases, could have resulted in a less than optimal mix of surgical candidates and therefore in worse outcomes and lower productive efficiency in the U.K. This would result from a higher frequency of regional cases in the identified surgical candidates than would have occurred if CT node assessment were used in all cases. Thus, if more resources were shifted to CT scanning (and taken from longer hospital length of stay or more inpatient stays with chemotherapy), the U.K. could have likely identified a better mix of surgical candidates and achieved better outcomes and higher productive efficiency.

System structure and regulation (Exhibit 12). The sources of these differences in physician and hospital incentives, physician supply, and capital and substitution constraints were major differences in health care system structure and regulation. Most physician services in the U.S., including both specialists and primary care, were negotiated and compensated on a fee-for-service (FFS) basis by payors. U.S. physicians also faced the threat of malpractice suits, which may have encouraged them to be more comprehensive in testing. U.S. physicians thus had an incentive to provide diagnostic testing to all of their patients that they suspected had lung cancer, as they were compensated for these activities. In contrast, specialists in the U.K. were paid flat salaries, negotiated on an annual basis, providing no incentive to increase the use of diagnostic testing.

Differences in physician and hospital incentives were also driven by differences in the competitive intensity of the nations' health care systems. U.S. physicians' and hospitals' competition for patients encouraged them to adopt CT technology and perform diagnostic testing to respond to consumer demand. Likewise, U.S. payors' competition for members led them to reimburse this testing. In contrast, the majority of the U.K. population had to rely on the NHS for all their health care needs, creating no competitive payor pressure (although there might have been political pressure) in the U.K. system to respond to consumer demands for diagnostic testing.

Regulation also contributed to the more limited access that lung cancer patients had to diagnostic testing in the U.K. The supply of physicians in the U.K. was controlled by regulation and, because it was more limited than the supply of physicians in the U.S., could not as easily accommodate an increased level of services. The NHS also tightly controlled hospital budgets and reimbursed hospitals in a lump sum for all operating costs. This fixed reimbursement placed the burden of resource allocation on the hospitals and physicians; these two entities had to decide how much would be spent on what disease and, within a given disease, who was treated at what level. This higher level of hospital product integration, combined with a limited and fixed capital budget and restrictive capital allocation processes, encouraged medical rationing according to need and resulted in capital being less readily available for CT scanning equipment in the U.K. than it was in the U.S. These aspects of the U.K. system also precluded easy substitution of capital resources for other care treatment resources, such as longer hospital stays. The monopolistic position of the NHS

allowed it to enforce this strict resource supply control, as well as mandates for system usage, like the rules for referral that also contributed to the limited use of diagnostic testing.

The recent reforms of the NHS and the growth in the private market for health care may influence the behavior toward care allocation that we observed for lung cancer. It is difficult to imagine the frequencies of treatments rising dramatically in the NHS system, since the amount of total resources available for care has not increased significantly. However, the reforms do allow more autonomy (and, therefore, resource flexibility) through the creation of private trusts for hospitals and "GP fundholders." This may lead to earlier and more frequent use of CT scans in the course of lung cancer care – "funded" by care resources directed from other phases of treatment, such as surgical LOS or inpatient (versus outpatient) chemotherapy – which could potentially improve outcomes without raising overall costs. Furthermore, it is possible that the private market will "solve" the substitution problem, allowing the lung cancer patient to go out of the NHS to get a CT scan and access to other care that is restricted by the referral patterns, should they continue.

Chemotherapy: choice of setting

Provider incentives and constraints (Exhibit 11). The U.K.'s productive efficiency relative to the U.S. was also lowered by its greater proportion of inpatient chemotherapy. In the U.S., physicians administering chemotherapy were likely indifferent to the setting choice for chemotherapy as they received the same reimbursement for treatment in inpatient and outpatient settings. The decision to move to more outpatient chemotherapy was influenced by payors, as we discuss in the system structure section below.

The higher proportion of inpatient chemotherapy in the U.K. likely had two causes. First, since chemotherapy was not available in every hospital setting, patients may have had to travel sizable distances to reach a location with a qualified provider. If these distances were large enough and the patient was relatively frail, the chemotherapy may have been delivered in an inpatient setting. Second, U.K. hospitals may have been slow in establishing outpatient facilities where they could administer chemotherapy due to capital constraints that precluded equipping of outpatient facilities.

System structure and regulation (Exhibit 12). The sources of these differences in setting choice for chemotherapy were major differences in health care system structure and regulation. U.S. payors competed for members based on a number of factors, including price. As payors faced rising costs in care provision, they could apply pressure on providers to lower the unit cost of procedures, lower the frequency of procedure use, or both. Because the frequency issue involved clinical judgment and the unit cost issue often did not, payors applied pressure on the unit cost side. The market for care provision responded in a number of ways, one of which was to move procedures outside of the hospital to take

advantage of cost savings in the outpatient environment stemming from the avoided overnight stay.

In the U.K., physician and capital constraints were due to regulation, by which the NHS dictated and controlled levels of expenditure. As previously discussed, the lack of competition in both the care provision and health coverage markets enabled the NHS to enforce these constraints throughout the health care system.

Inpatient procedures: LOS and hospital staffing

Provider incentives and constraints (Exhibit 11). As discussed above, differences in LOS and staffing levels between the U.S. and the U.K. for inpatient procedures led to differences in relative productive efficiency. Patients undergoing inpatient procedures in the U.S. had a shorter average LOS than those in the U.K. due to differing hospital incentives to manage LOS. The shorter LOS in the U.S., in turn, caused U.S. hospitals to keep staffing high, as a shorter average LOS means patients were discharged earlier in the U.S. than in the U.K.; as U.S. hospitals thus had a higher proportion of their patients in acute conditions, they had to keep staffing higher to meet the needs of those patients.

One exception to this pattern was radiotherapy, as U.S. patients had a longer average LOS for radiotherapy than patients in the U.K. Since radiotherapy is delivered once a day, the total number of fractions corresponds to the total days of treatment. If the treatment is performed in the hospital, the number of fractions corresponds to the number of hospital-days required. We stated previously that the U.K. employed radiotherapy protocols of the same total radiation dosage, but with fewer fractions. Thus, the U.S. had a longer LOS per inpatient radiotherapy case due to the greater number of fractions employed.

In the U.S., delivering radiotherapy in an increased number of smaller doses was due to activity-increasing incentives for U.S. physicians and hospitals. In contrast, constraints on physicians, hospitals, and capital in the U.K. led to less fractioned (higher dosages but fewer total treatments) radiotherapy. By reducing the number of fractions delivered to patients, less provider time was required per total course of care, a single radiotherapy linear accelerator could support more patients over a given time period, and fewer bed-days would be consumed should the radiotherapy be administered as an inpatient procedure; in short, more patients, whether lung cancer or others requiring radiotherapy, could be treated with the same level of resources.

The advantage in relative productive efficiency the U.K. gained by using fewer fraction protocols was completely offset by the U.K.'s higher use of staffing for radiotherapy. Through clinician interviews we learned that both the staffing levels and treatment duration were higher in the U.K. than in the U.S. and Germany. Clinicians attributed both these differences to the much older equipment in use in the U.K. Presumably, the older equipment required more labor to operate and more setup time for each patient. Thus, the total labor

requirement for a fraction of outpatient radiotherapy was higher in the U.K. than it was in the U.S. This difference was due to U.K. capital constraints, which precluded the purchase of newer radiotherapy equipment.

System structure and regulation (Exhibit 12). The sources of these differences in hospital incentives were differences in health care system structure and regulation. At the time of our study, most U.S. payors individually contracted with hospitals for services. Payments stipulated by these contracts were generally of three basic types: FFS (about 45 to 50 percent), per case (about 35 to 40 percent), or per diem (about 15 percent). FFS payments paid hospitals a fixed amount per procedure or service. Per-case payments paid hospitals a fixed amount per admission, adjusted for the patient's diagnostic-related groups (DRG).[18] Under this system, the reimbursement was defined solely by the procedure or diagnosis, not the number of days the patient was in the hospital. While FFS reimbursement did not influence hospital decisions on LOS or staffing, DRG-based reimbursement gave hospitals a clear incentive to decrease costs by reducing LOS per patient, since the hospital itself assumed the risk for the cost of the patient's stay. Hospitals thus had an incentive to reduce LOS by streamlining admission and discharge processes and shortening pre- and postoperative time. As discussed above, this shorter LOS likely led to higher U.S. staffing levels, as the intensity of case per average hospital-day increased with decreasing LOS. The net effect of shorter LOS and higher staffing was lower input usage (and, thus, higher profit margins on a per-case basis) for the hospital.

The per diem contract, in which reimbursement was tied to the hospital-day rather than the entire admission, was not as common as FFS or per-case payment. Payors following this arrangement coordinated physician care and managed patient throughput (e.g., through a utilization review process). Thus, under a per diem contract (as with FFS contracts), the payor primarily assumed the risk for hospital stay (which could also be transferred to physicians via capitated contracts).

Regardless of the type of contractual arrangement, the risk-bearing entity (payor or provider) had a clear incentive to limit hospital stay in the U.S. Consequentially, the intensity of competition among U.S. payors in the health coverage market likely helped to accelerate the adoption of case rates; since payors competed for customers on the basis of price, among other things, they saw the adoption of case rate pricing as a way to control increasing hospital costs.

In contrast, the U.K. health care system structure and regulation created incentives for U.K. hospitals that were entirely different from those in the U.S. and were not aligned with reducing LOS. In the U.K., fixed hospital budgets, coupled with limited competition among hospitals for patients or payor funds,

---

[18] The Medicare program in the U.S. adopted the case rate (DRG) payments in 1983; many other payors subsequently adopted this payment system as well.

created no incentives for U.K. hospitals to manage LOS. As shorter LOS implied that patients would be turned over more rapidly (i.e., greater admission and discharge rates), caring for new patients and performing additional procedures was more costly than maintaining patients in recovery at the end of a longer LOS. In other words, cycling through more patient cases whose treatment required higher input usage would have raised *total* hospital costs without generating a corresponding increase in revenue. More efficient hospitals that utilized or exceeded their entire operating budgets could not easily be differentiated from inefficient hospitals that consumed their entire budgets. Thus, if the budget allocation process was not sensitive enough to detect superior performance, it could deprive operating funds from the more efficient hospitals and, in the worst case, reduce their operating budgets to curtail overspending.

Differences in system structure and regulation also contributed to the lower staffing levels in the U.K. U.S. hospitals' competition for patients may have led them to compete by offering high levels of service, promoting greater staffing intensity. In the U.K., however, hospitals did not compete for patients and had limited budgets with which to fund all their hospital needs, including staffing. In addition, staffing levels were also affected by the availability of physicians, which was regulated and limited by the NHS. There was thus much more pressure to lower staffing levels in the U.K. than the U.S.

The difference between the two countries in the approach to delivering radiotherapy can be explained through much the same logic as the differences in care triaging in diagnostic testing cited above. The practice of more fractionated radiotherapy in the U.S. was supported by the activity-increasing FFS and case rate reimbursements for physicians and hospitals. In the U.K., the same constraints on physicians, hospitals, and capital that we discussed previously led to lower fractions of radiotherapy treatment. By utilizing fewer fractions per patient, the U.K. could increase the throughput of radiotherapy equipment (i.e., increase the number of patients a radiotherapy center could support) without raising operating costs or purchasing additional equipment.

Surgery, radiotherapy, and chemotherapy: frequency

Provider incentives and constraints. The frequencies of surgery, chemotherapy, and radiotherapy were all lower in the U.K. than in the U.S. While these differences in the frequency of procedures did not necessarily impact relative productive efficiency, they were the most important drivers of input differences between the two countries.

Importantly, the lower frequency of surgery also led to the U.K.'s lower outcomes in lung cancer treatment. Although not precise, the research literature suggests that about 20 to 30+ percent of all non-small cell, and a marginal percentage of small cell, lung cancer patients could potentially benefit from surgery. However, it appears that the long-run survival for lung cancer under the best of circumstances is currently just over 10 percent. This implies that in order to achieve the 10-percent

survival, two to three times as many patients in the U.K. must receive surgery. The U.K. surgery rate of 13 percent was much lower than the suggested range of eligibility. As previously discussed, we have no reason to believe that providers in the U.K. were better than their U.S. counterparts in identifying the cases that would be cured through surgery, but rather, that they likely had incomplete staging.[19] Therefore, this lower level of surgery likely led to a lower overall survival rate for lung cancer patients as a whole.

The higher frequency of surgery, radiotherapy, and chemotherapy in the U.S. was caused in part by activity-increasing incentives for U.S. physicians and hospitals, as opposed to activity-neutral incentives for U.K. physicians and hospitals. In addition, constraints on physician and hospital supply in the U.K. and the complex referral process contributed to the lower frequency of treatments observed.

System structure and regulation (Exhibit 12). The sources of these differences in physician and hospital incentives and physician supply and capital constraints were major differences in health care system structure and regulation. As previously discussed, the U.S.'s FFS physician reimbursement policies created activity-increasing incentives for U.S. physicians, while the flat salaries for specialists in the U.K. did not. Likewise, the previously discussed lack of competition in the U.K. health care markets enabled the NHS to limit the frequency of these procedures, while competition in the U.S. did not.

U.S. versus Germany

Differences in productive efficiency between the U.S. and Germany were primarily caused by provider behavior differences in two areas of lung cancer treatment: choice of setting for chemotherapy and differences in inpatient LOS and staffing levels (see again **Exhibit 9**). In this section, we describe each of these in turn. Differences in surgical and other treatment frequencies are also discussed, as these led to differences in input consumption, but not necessarily to differences in productive efficiency. As we have already discussed the incentives and constraints, system structure, and regulations that led to U.S. provider behavior, our focus in this section will be on Germany.

<u>Chemotherapy: choice of setting</u>

Provider incentives and constraints (Exhibit 13). The major difference in chemotherapy practice between the U.S. and Germany impacting relative

---

[19] In other words, it is unlikely that the 10 percent of lung cancer patients that could be cured by surgery were fully contained in the 13 percent of lung cancer patients that were surgically treated in the U.K. It was more likely that these potential survivors would be contained in the 22 percent of lung cancer patients receiving surgery in the U.S.

productive efficiency was that Germany utilized the inpatient setting to a much greater degree for chemotherapy than did the U.S.

U.S. physicians had incentives to provide chemotherapy in the outpatient setting, as discussed in the U.S. versus U.K. comparison. German hospitals and physicians, on the other hand, had strong incentives to keep procedures inpatient to occupy hospital beds, which was made possible by the abundant supply of hospital beds in Germany. In addition, German regulation, discussed below, established substitution constraints that gave German hospital-based physicians strong incentives to keep procedures in the inpatient setting.

System structure and regulation (Exhibit 14). The sources of these differences in physician and hospital incentives, hospital supply, and substitution constraints were differences in health care system structure and regulation. In the U.S., some price-based payor competition for members encouraged payors to be cost-conscious; payors may have, therefore, encouraged surgeons to move chemotherapy to the outpatient setting to avoid the cost of a hospital stay.

In Germany, however, regulation created strong incentives for physicians and hospitals to provide chemotherapy in an inpatient setting. German hospitals were compensated on a per diem basis by the sickness funds, whereby they received additional reimbursement for each day of a patient's stay. Annual hospital budgeting was based on this per diem reimbursement, with the per diem level of the previous year being increased or decreased as necessary to compensate for inflation and for some of the increase or decrease in hospital utilization. Each hospital negotiated as a block with all payors for these per diem rates.

State authorities, as regulators of hospital capacity, had little incentive to reduce the use of local health care resources, as discussed in Chapter 2; in fact, they 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. Consequently, regulations generally stipulated very high occupancy rates, usually about 85 percent. If this target was not met, the hospital was at risk for being reviewed and having its capacity cut. The combination of this regulatory threat and the large supply of hospital beds created a strong incentive for hospitals, and, therefore, the physicians they employed, to keep chemotherapy in the inpatient setting in order to keep more beds occupied. Physicians had further incentive to keep the utilization of public beds high because hospitals typically specified that beds had to be used for public and private patients in a relatively set ratio; German hospital department chiefs could, therefore, add private bed capacity, and thus earn higher private patient fees, by increasing the use of public beds.

Limited competition in the German payor market also contributed to the continued use of chemotherapy in Germany. German sickness funds (payors) had relatively restricted memberships along geographic or occupational lines,

and retained their members for a relatively long period. Payors did not compete for members on price and could not negotiate differentially from other payors with each hospital; therefore, they had no real incentive or ability to pressure hospitals to manage costs or move chemotherapy to an outpatient setting.

Regulation in Germany also created incentives for administering chemotherapy in inpatient settings. Regulations forced physicians to choose between working in either the hospital or office setting and specified which services could be performed in each setting; most hospital-based physicians were not authorized to perform procedures in the outpatient setting.[20] If a hospital-based physician converted chemotherapy to an outpatient event, therefore, the physician and the hospital would lose the revenue associated with the procedure.

Inpatient procedures: LOS and hospital staffing

Provider incentives and constraints (Exhibit 13). As discussed above, the longer LOS for inpatient procedures in Germany, even offset by higher staffing levels in the U.S., led to lower productive efficiency for Germany relative to the U.S. in lung cancer treatment. Because we have already explained the incentives and constraints in the U.S. system that led to shorter LOS, we will focus on Germany in this section.

The longer stays in Germany resulted from the same incentives that caused chemotherapy to be delivered in the inpatient setting. As previously explained, German hospitals had strong incentives to keep hospital beds occupied and, therefore, to raise or maintain the relatively long duration of hospital stays. This longer LOS in Germany was made possible by the large number of German hospital beds. Longer stays in German hospitals were partially offset by the lower staffing levels for both nurses and physicians relative to the U.S. While hospitals in both countries had incentives to attract patients by maintaining high staffing levels, the significantly shorter patient stays in U.S. hospitals likely led to the need for higher levels of staffing, as previously discussed. German physicians, particularly department chiefs, also had incentive to fill public hospital beds, as discussed earlier.

System structure and regulation (Exhibit 14). Differences in hospital and physician incentives and hospital supply resulted from differences in the health care system structure and regulation. In addition to the system structure and regulation differences discussed above that led to incentives to fill German hospitals, the German per diem reimbursement schedule gave German hospitals incentives to maintain long LOS. In contrast, U.S. hospitals received a significant portion of their reimbursement as case rate (DRG) payments, as discussed above, which encouraged hospitals to manage LOS.

---

[20] Physicians in the German public system were licensed to practice in either the hospital or office-based setting. An exception allowed academic department heads to practice in both settings.

Surgery: frequency

Provider incentives and constraints. While there were differences in the frequencies of each of the three major treatments for lung cancer between the U.S. and Germany, the difference in surgical frequency was the most important. The frequency of lung cancer surgery was higher in Germany than in the U.S. While this difference in the frequency of surgery did not necessarily impact relative productive efficiency, it was the most important driver of input differences between the two countries.

Activity-increasing incentives influenced physicians and hospitals in both countries. In the comparison of the U.S. and U.K., we discussed these incentives for the U.S. The forces behind the incentives created in Germany were different from those in the U.S. and, potentially, more influential.

Because German hospitals were reimbursed entirely on a per diem basis, hospitals had an incentive to maintain or increase all hospital-based activities. As employees of the hospitals, physicians also had an activity-increasing incentive, albeit indirectly. While the reimbursement mechanism in Germany differed from the FFS and case rate mechanisms in the U.S., it had the same activity-increasing effect. Thus, while the incentives created by these reimbursement mechanisms were an important part of the explanation, we have to look beyond them to understand what caused greater surgical frequency in Germany. The answer may lie in differences in the supply of physicians and hospitals. On a per capita basis, Germany has more hospital beds and hospital based physicians than the U.S.[21] Furthermore, since per diem reimbursement was based on meeting a target hospital occupancy that was quite high, the prosperity of all hospitals depended on increasing to reach, or maintaining, a high level of patient volume. This increased the incentive for German physicians and hospitals to provide care and likely explained the higher levels compared with the U.S.

While less important than the frequency of surgery, the frequencies of chemotherapy and radiotherapy also differed between the two countries. Presumably because a greater portion of lung cancer patients in the U.S. were treated with palliative intent, the frequency of palliative chemotherapy was higher in the U.S. This is consistent with the higher rate of surgery (i.e., curative intent care) observed in Germany. The frequency difference in radiotherapy was insignificant and led to only a very small difference in resource consumption.

System structure and regulation. As previously discussed, per diem hospital reimbursement and regulation in Germany created strong incentives for hospitals, and the physicians they employed, to increase hospital utilization. Increasing the frequency of inpatient procedures, such as lung cancer surgery,

---

[21] Note that while German hospitals had lower levels of staffing per bed than the U.S., they had more total beds and hospital physicians per capita than the U.S.

was obviously one way to do this. In the U.S., the reimbursement methodology also created strong incentives to maintain high levels of care. The difference, therefore, was likely due to the regulatory environment. At one time, the U.S. government encouraged the construction of hospitals. Germany had a regulatory incentive to build capacity as well; however, because it was slower in shutting down hospital growth, the supply of hospital beds per capita in Germany outgrew that in the U.S. The regulatory oversight in Germany for hospital bed supply was in the hands of the local and state governments. Federal subsidies to the hospital were based on the total number of beds, giving local and state governments an incentive to increase hospital bed capacity, primarily as a way to create jobs. Thus, by responding to the incentive to reach target capacity, hospitals and providers pushed the frequency of lung cancer surgery beyond the levels observed in the U.S.

SUMMARY OF LUNG CANCER CASE RESULTS

In the treatment of lung cancer, the U.S. was the most productive, Germany was the next most productive, and the U.K. was the least productive. The sources of these differences in productive efficiency were differences in provider behavior. Comparing the U.S. with the U.K., the U.K.'s relative productive efficiency was reduced by its lower use of diagnostic testing (particularly CT scans for staging and surgical candidate identification), longer LOS, and greater proportion of inpatient chemotherapy. While the U.S.'s productive efficiency was slightly reduced by its higher staffing levels, the impact of this difference was minimal. These differences in provider behavior were caused by different physician and hospital incentives, as well as by physician supply, capital, and substitution constraints in the U.K. U.S. physician and hospital competition for patients in the U.S. encouraged adoption and coverage of CT scans for diagnostic testing, while U.S. payor price-based competition for members encouraged shorter LOS and outpatient chemotherapy. In the U.K., capital constraints resulting from the global NHS budget, as well as the complex and restricted referral process for surgery, led to less resources directed toward CT scanning in diagnosis and staging (versus toward other treatment aspects). This resulted in a less than optimal surgical mix and lower productive efficiency.

Comparing Germany with the U.S., Germany's relative productive efficiency was lowered by its longer LOS and greater proportion of inpatient chemotherapy than the U.S. While the U.S.'s productive efficiency was slightly reduced by its higher staffing levels, the impact of this difference was minimal. These differences in provider behavior were caused by strong incentives for German hospitals and physicians to occupy hospital beds, which, in turn, were caused by the health care system structure and regulatory environment of Germany.

Appendix 6A: Outcome measurement methodology

As described in **Appendix 1E**, the outcome measure we employed for lung cancer is based on the 5-year survival curves for each country. These survival curves came from cancer registries in each of the three countries. This appendix describes how we arrived at the outcome statistic we used in our comparison of productive efficiency.

NORMALIZATION

The survival curves were normalized for age, sex, cell type (i.e., small versus non-small), and stage. We adjusted each country to have the same distributions of these variable as the U.S. In adjusting for differences in the stage distribution among countries, we were conscious of an artificial bias that can get introduced into survival comparisons when there are differences in the distributions of stages that relate to differences in staging activities and the technologies employed.

While the U.S. and Germany appeared to employ similar levels of diagnostics during the diagnosis and staging management phase, the U.K. underinvested compared to both. Thus there was a concern that the different stage distribution observed in the U.K. data was at least partially due to different approaches to staging and not entirely due to real differences in the population with lung cancer. If this were the case, we may have inadvertently biased the U.K. aggregate 5-year survival curve by normalizing for stage. To check the extent of bias potentially introduced, we compared the outcome metric derived from the normalized 5-year survival curve with that from the un-normalized survival curve (i.e., the raw 5-year survival curve for all lung cancer cases taken together) and found an insignificant difference. Thus we can be relatively confident that no significant bias was introduced.

LYs SAVED

To understand the LYs saved by treatment for lung cancer, the untreated survival curve must be known. With both the treated and untreated curves known, the LYs saved is simply the area between the two curves (**Exhibit A-1**).

While we were unable to find untreated case survival data for all stages of disease and normalize it for all the variables for which we adjusted our survival curves, we were able to crudely estimate a range of curves that likely bracketed untreated survival. Thus, we were able to calculate the LYs saved by treatment in each of the three countries and make that the basis of our outcome comparison.

Appendix 6B: Sources

This list details data sources used in the lung cancer case study. We cover most of the main topics here, but this list is not exhaustive of all of the articles and government statistics that were employed throughout our work. In addition, we performed interviews with clinical and health care experts at a number of points during our study. Through these interviews, we collected qualitative and quantitative data on treatment patterns and checked our key assumptions and conclusions.

Below, we give the main sources used by topic.

OUTCOMES

U.S.

Surveillance Epidemiology and End Results (SEER) Public Use Database, National Cancer Institute.

U.K.

Thames Cancer Registry Database.

Germany

Krebsregister – Saarland.

INPUTS

U.S.

SEER Public Use Database, National Cancer Institute.

U.K.

Connolly CK et al. Investigation, treatment and prognosis of bronchial carcinoma in the Yorkshire region of England 1976-1983. British Journal of Cancer 1990; 61: 579-583.

Joslin C, Rider L. Cancer in Yorkshire – Cancer Registry Special Report Services: 1 Lung Cancer. University of Leeds Print Service, May 1993.

Stephens R, Gibson D. The impact of clinical trials on the treatment of lung cancer. Clinical Oncology 1993; 5: 211-219.

Thames Cancer Registry Database.

Germany

Arbeitsgruppe Zur Koordination Klinischer Krebsregister.

FREQUENCIES OF DIAGNOSTIC PROCEDURES

Edinburgh Lung Cancer Group. Patients presenting with lung cancer in South East Scotland. Thorax 1987; 42: 853-857.

Humphrey et al. National survey of the pattern of care for carcinoma of the lung. The Journal of Thoracic and Cardiovascular Surgery 1990; 100: 837-43.

Unpublished data from lung cancer research in Scotland.

ADDITIONAL INFORMATION

Feinstein AR et al. The Will Rogers phenomenon: Stage migration and new diagnostic techniques as a source of misleading statistics for survival in cancer. The New England Journal of Medicine 1985; 312: 1604-8.

McKenna et al. Roentgenographic evaluation of mediastinal nodes for preoperative assessment in lung cancer. Chest 1985; 88.

Richey et al. Thoracic CT scanning in the staging of bronchogenic carcinoma. Chest 1984; 85: 218-221.

Sandler A, Buzaid A. Lung cancer: A review of current therapeutic modalities. Lung 1992; 170: 249-265.

Tsang G, Watson D. The practice of cardiothoracic surgeons in the perioperative staging of non-small cell lung cancer. Thorax 1992; 47: 3-5.

Wingo PA et al. Cancer statistics 1995. CA – A Journal for Clinicians 1995; 45: 8-30.

World Health Organization: World Health Statistics Annuals, World Health Organization, (Geneva, Switzerland), 1987-1992.

Exhibit 1

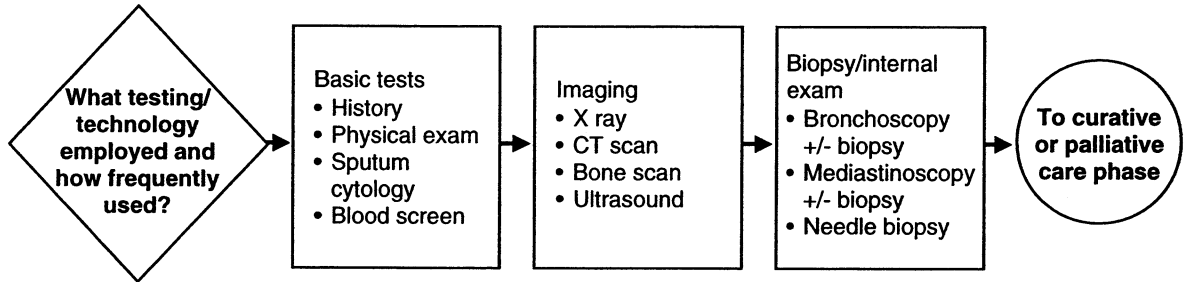**PHASES OF LUNG CANCER
MANAGEMENT AND TREATMENT PROCESS**

	Diagnosis and staging	**Curative care**	**Palliative care**
Purpose of phase	Confirm the existence of the cancer, its cell type, and the degree of metastases; plan treatment	Intervene to control the cancer with the intent of achieving remission	Intervene to lessen the symptoms of the cancer and/or its metastatic sites
Comments/ rationale	Treatment and prognosis vary greatly according to the type of cancer and the extent of its spread Accurate assessment increases the likelihood of matching the treatment to the condition	Some cases are caught sufficiently early in their progression to be cured by aggressive treatment	Most patients will die from lung cancer, but their symptoms can be alleviated to some extent
Average percent of inputs consumed in each phase*	19%	41%	40%

* Average inputs consumed per phase in the U.S., U.K., and Germany

022 B 127426/6

Exhibit 2

OVERVIEW OF LUNG CANCER
MANAGEMENT AND TREATMENT PROCESS
Diagnosis and staging phase*

What testing/ technology employed and how frequently used?
→
Basic tests
- History
- Physical exam
- Sputum cytology
- Blood screen
→
Imaging
- X ray
- CT scan
- Bone scan
- Ultrasound
→
Biopsy/internal exam
- Bronchoscopy +/- biopsy
- Mediastinoscopy +/- biopsy
- Needle biopsy
→
To curative or palliative care phase
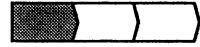
* Some relatively infrequent tests have been omitted

022 B 127427/6

Exhibit 3

OVERVIEW OF LUNG CANCER
MANAGEMENT AND TREATMENT PROCESS
Curative care phase
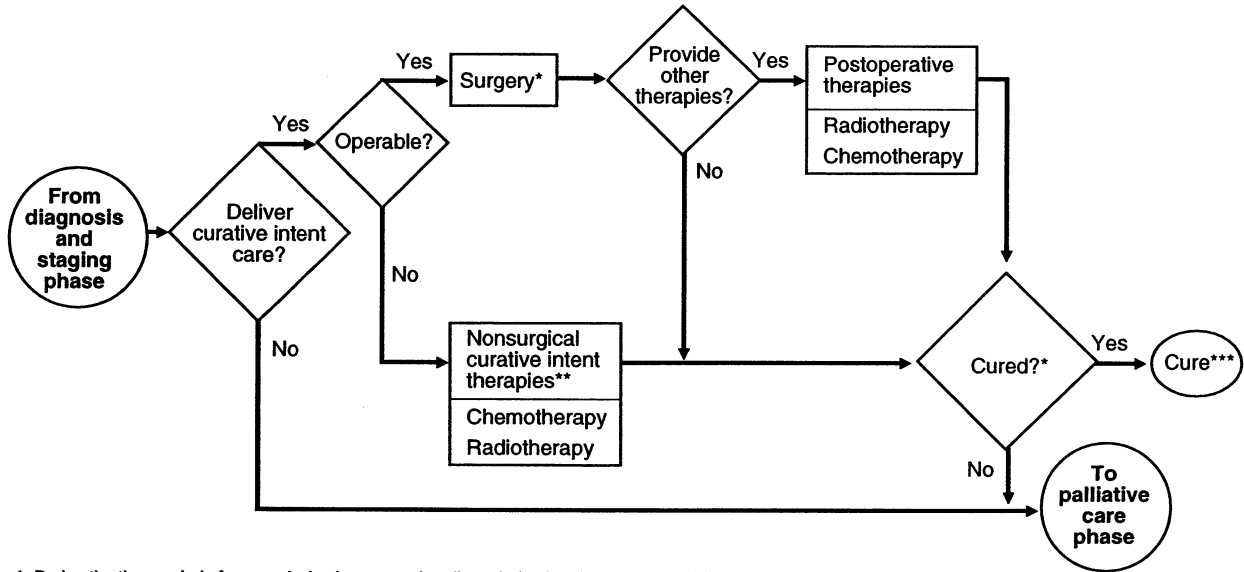
* During the time period of our analysis, there was virtually no induction (i.e., preoperative) therapy
** In a small number of cases, curative intent nonoperative care will be administered; some providers might consider all nonoperative care
   to be palliative in intent; our classification is arbitrary and has no bearing on our analysis
*** In this case, "cured" means that the cancer was removed or is in full remission; note that these cases can potentially relapse and receive
    additional treatment

022 ST 125850/6

Exhibit 4
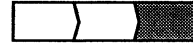
OVERVIEW OF LUNG CANCER
MANAGEMENT AND TREATMENT PROCESS
Palliative care phase

Exhibit 5
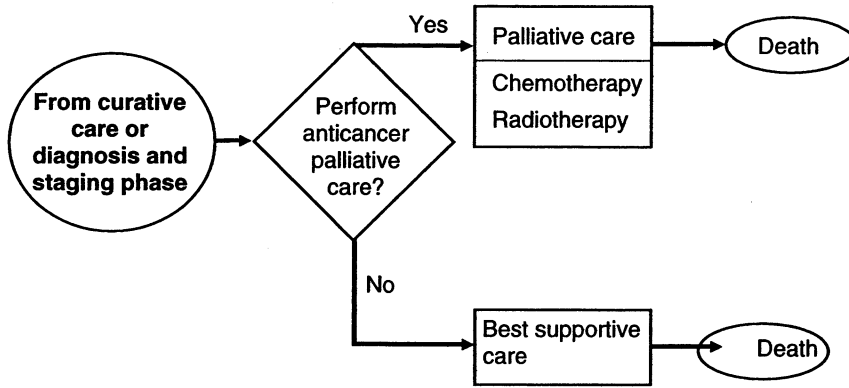
INPUTS AND OUTCOMES PER CASE OF LUNG CANCER

Outcomes per case
LYs

Inputs per case
Standardized input units

	Inputs Index, U.S. = 100%	**Outcomes** Improvement in outcomes due to treatment Index, U.S. = 100%	**LYs**
U.K.	76	42	0.22
U.S.	100	100	0.52
Germany	121	88	0.46
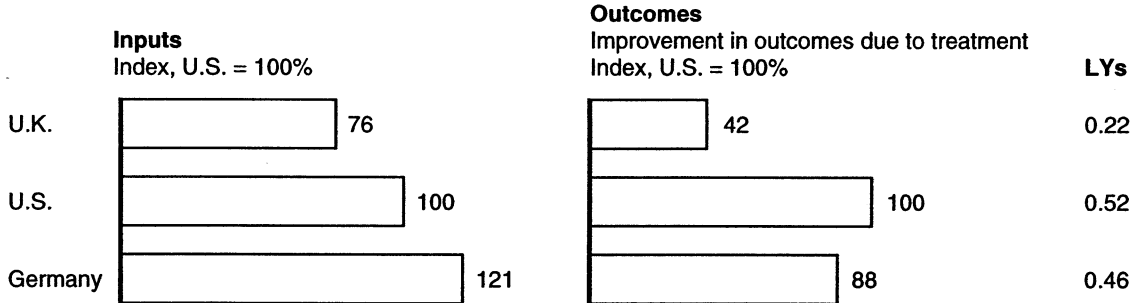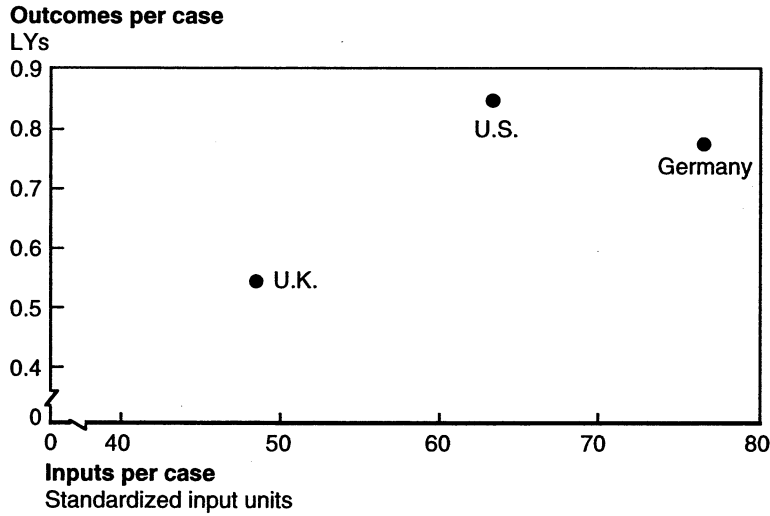
Note: Outcome measure calculated based on 5-year survival; baseline outcome with no treatment, estimated to be 3.8 months

Exhibit 6

ASSESSMENT OF RELATIVE PRODUCTIVE EFFICIENCY
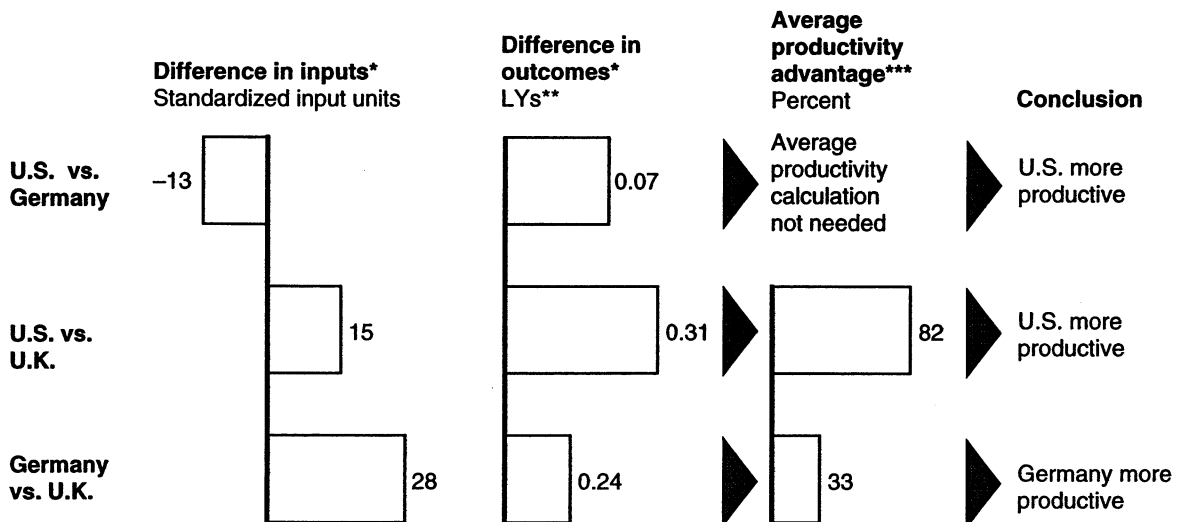
	Difference in inputs* Standardized input units	**Difference in outcomes*** LYs**	**Average productivity advantage**** Percent	**Conclusion**
U.S. vs. Germany	−13	0.07	Average productivity calculation not needed	U.S. more productive
U.S. vs. U.K.	15	0.31	82	U.S. more productive
Germany vs. U.K.	28	0.24	33	Germany more productive

* Difference between number of units in first country to those in second country
** Calculated based on 5-year survival
*** Average productivity advantage of first country over second

022 ST 140008/6

Exhibit 7

**SOURCES OF DIFFERENCE IN U.S. AND
U.K. INPUTS PER CASE – LUNG CANCER**

Percent of difference

* Other differences net to 0

Exhibit 8

**SOURCES OF DIFFERENCE IN U.S. AND
GERMANY INPUTS PER CASE – LUNG CANCER**

Percent of difference
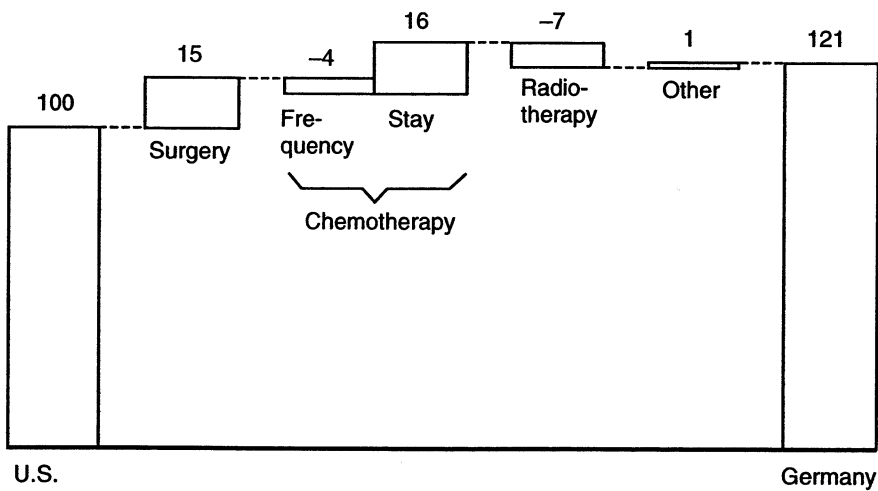
Exhibit 9

IMPACT OF PROVIDER BEHAVIOR
DIFFERENCES ON PRODUCTIVE EFFICIENCY

● Major differentiating factor
◑ Secondary differentiating factor
○ Undifferentiating

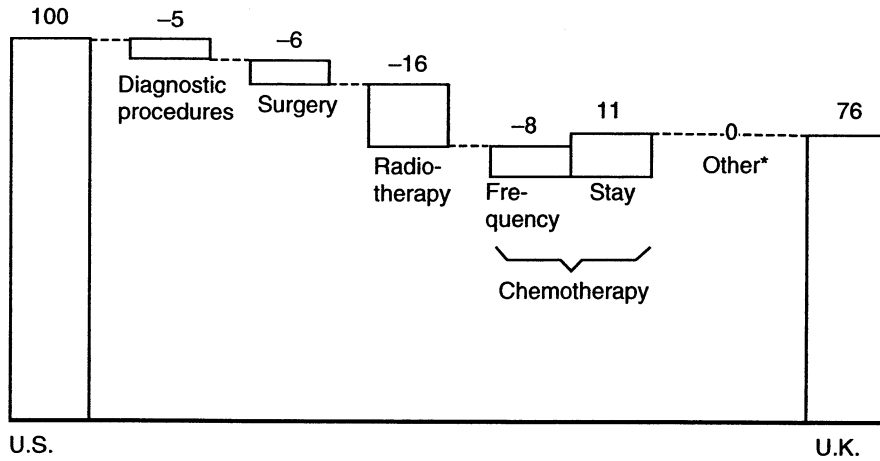| Provider behaviors		U.S. vs. U.K.		U.S. vs. Germany
Care triaging	◑	U.K. restricted referral process and limited CT use in diagnosis may have led to less effective staging	○	
Treatment duration	◑	U.K. longer surgical LOS than U.S.	●	Germany longer LOS than U.S.
Staffing levels	◑	U.K. lower staffing than U.S.	◑	Germany lower staffing than U.S.
Setting choice	◑	U.K. greater proportion of inpatient chemotherapy than U.S.	●	Germany greater proportion of inpatient chemotherapy than U.S.
Team-based approach	○		○	
Technology adoption	◑	U.K. less CT use for diagnosis and effective staging than U.S.	○	

Exhibit 10

DIFFERENCES IN PROVIDER BEHAVIOR IN LUNG CANCER MANAGEMENT/TREATMENT
Curative and palliative management phases

↑ Higher
↓ Lower
— About same time

Factors driving differences in behavior (related causality framework term)	Assessment relative to U.S. behavior (Curative in regular text, palliative in bold)					
	Surgery		Radiotherapy		Chemotherapy	
	U.K.	Germany	U.K.	Germany	U.K.	Germany
Frequency of treatment (care triaging)	↓ n/a*	↑ n/a	↓ **↓**	— **—**	— **↓**	— **↓**
Length of inpatient stay (treatment duration)	↑ n/a	↑ n/a	↓ **↓**	— **—**	— **—**	— **—**
Hospital staffing (staffing)	↓ n/a	↓ n/a	↓ **↓**	↓ **↓**	↓ **↓**	↓ **↓**
Frequency of inpatient care (setting)	n/a n/a	— n/a	— **—**	— **—**	↑ **↑**	↑ **↑**
Overall resource consumption	↓ n/a	n/a n/a	↓ **↓**	↓ **↓**	↑ **↑**	↑ **↑**

* Not applicable

022 MI 95034/6

Exhibit 11

CAUSAL ANALYSIS: INCENTIVES AND CONSTRAINTS
U.S. vs. U.K.

● Major differentiating factor
◐ Secondary differentiating factor
○ Undifferentiating

	Extent of causal impact	Provider behaviors				
Incentives		**Care triaging**	**Treatment duration**	**Staffing levels**	**Setting choice**	**Technology adoption**
Physician	○	U.S. physicians had incentive to assess patients				
Hospital	◐	U.S. hospitals had incentive to assess patients	U.S. hospitals had incentive to manage LOS	U.S. hospitals had incentive to keep LOS low, requiring higher staffing		U.S. hospitals had incentive to adopt CTs
Constraints						
Physician supply	◐	U.K. limited supply of physicians provided less time to assess patients				
Hospital supply	○					
Capital	◐				U.K. capital constraints precluded equipping of outpatient facilities	U.K. capital constants precluded adoption of CT technology
Substitution	◐	U.K. restricted referral and gatekeeper process led to less effective triaging				

Exhibit 12

CAUSAL ANALYSIS: SYSTEM STRUCTURE AND REGULATION
U.S. vs. U.K.

● Major differentiating factor
◐ Secondary differentiating factor
○ Undifferentiating

	Extent of causal impact	Provider incentives and constraints					
		Physician incentives	**Hospital incentives**	**Physician supply**	**Hospital supply**	**Capital constraints**	**Substitution constraints**
Health care system structure							
Product integration and pricing	○	U.S. FFS vs. U.K. salary for specialist physicians encouraged CT assessment	U.K. annual hospital budget encouraged "bed-blocking" and therefore longer LOS; U.S. hospital case rates for Medicare encouraged shorter LOS				U.K. salary and gate-keeper process led to substitution constraint
Competitive intensity	◐	U.S. physician competition for patients and payor competition for members encouraged adoption and coverage for CT scans	U.S. payor price-based competition for members put pressure on hospitals to manage LOS				
Regulation	●	U.S. physicians faced threat of malpractice suits	U.K. hospital budgets controlled and set through NHS global budget	U.K. physician supply strictly controlled	U.K. hospital supply strictly controlled	U.K. capital supply strictly controlled	

Exhibit 13

CAUSAL ANALYSIS: INCENTIVES AND CONSTRAINTS
U.S. vs. Germany

● Major differentiating factor
◐ Secondary differentiating factor
○ Undifferentiating

	Extent of causal impact	Provider behaviors		
		Treatment duration	Staffing levels	Setting choice
Incentives				
Physician	◐	German physicians had incentive to keep LOS long to occupy hospital beds		Given hospital incentives, German specialist physicians had some incentive to keep procedures in inpatient setting
Hospital	●	U.S. hospitals had incentive to manage LOS; German hospitals had incentive to occupy hospital beds	U.S. hospitals had incentive to keep LOS low, requiring higher staffing	German hospitals had incentive to keep procedures in inpatient setting
Constraints				
Physician supply	○			
Hospital supply	●	High number of German hospital beds led to long LOS		
Capital	○			
Substitution	◐			German hospital-based physicians not allowed to perform procedures in outpatient setting

Exhibit 14

# CAUSAL ANALYSIS: SYSTEM STRUCTURE AND REGULATION
U.S. vs. Germany

● Major differentiating factor
◐ Secondary differentiating factor
○ Undifferentiating

	Extent of causal impact	Provider incentives and constraints			
		Physician incentives	Hospital incentives	Hospital supply	Substitution constraints
Health care system structure					
Product integration and pricing	●		German per diem vs. U.S. case rate product encouraged longer LOS		
Competitive intensity	◐		U.S. payor price-based competition for members and ability to negotiate differentially with individual hospitals encouraged hospital cost control; German payors had relatively restricted memberships along geographic or occupational lines and could not negotiate differentially with hospitals for price or other contract terms		
Regulation	●	German department chiefs could add private bed capacity if kept public bed capacity high	German hospitals faced capacity cuts if utilization below 85%	German hospital supply regulated through state government with political motives	German regulation enforced separation of inpatient and ambulatory care settings

Exhibit A-1
OUTCOME METHODOLOGY

Percent survival

Years from diagnosis
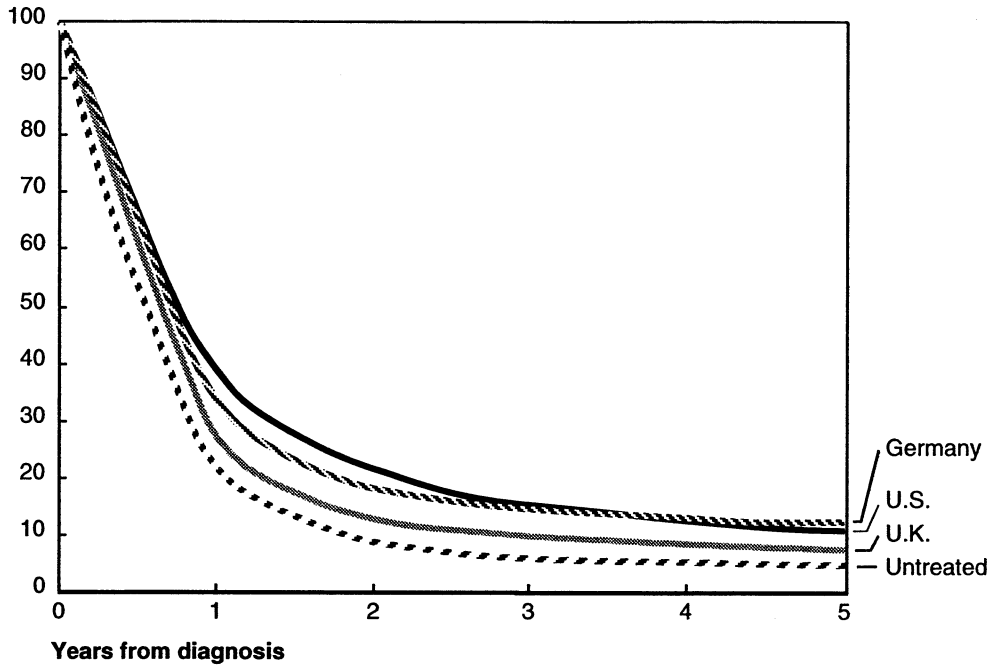
The area between the individual country's survival curve and the "untreated" estimated curve gives the LYs saved by treatment performed

022 ST 140084/6

Chapter 7: Cross-disease synthesis

Each disease case study (Chapters 3 through 6) has analyzed the U.S., the U.K., and German productive efficiency, treatment patterns, and contributing system characteristics. In this chapter, we synthesize the findings from these four case studies and draw general conclusions about the relative productive efficiency of the three countries' health care systems and about the major drivers of these differences.

Our major findings are as follows:

¶ **Relative productive efficiency.** No single country was more productive than the others in every disease. The U.S. was more productive than Germany, but had mixed results relative to the U.K.; Germany was generally less productive than the U.S., but had mixed results relative to the U.K.; and the U.K. had mixed results (**Exhibit 1**).

¶ **Provider behavior variations.** More frequent inpatient care and lengthier care were the most important drivers of Germany's lower productive efficiency relative to the U.S. In contrast, the speed and extent of technology adoption along with intensity in care triaging were the most important drivers of productive efficiency differences between the U.S. and the U.K.

¶ **Provider incentives and constraints.** These variations in provider behavior were fully consistent with the provider incentives and constraints embodied in the different health care systems. Germany's greater use of inpatient care resulted primarily from the strong incentives it created for hospitals and physicians to fill hospital beds, as well as from its greater supply of hospitals and physicians and the regulatory constraints between the inpatient and outpatient care settings. Fixed physician salaries and constraints on the supply of physicians, hospitals, and capital led the U.K. providers to be slower and more selective in their adoption of technology and more intense in triaging care.

¶ **Health care system structure.** The structure of markets for health coverage and care provision in each country directly influenced provider incentives and constraints and their resulting behavior and productive efficiency. The U.S.'s more integrated hospital case rate product and its greater competitive intensity led to higher productive efficiency relative to Germany in each disease. Although the U.S. had greater productive efficiency than the U.K. in lung cancer and

cholelithiasis, its greater competitive intensity and less integrated care led to lower productive efficiency in diabetes and indeterminate relative productive efficiency in breast cancer.

¶ **Regulation.** Particularly in Germany and the U.K., regulation played a strong role in shaping market structure and in creating influential supply constraints. Regulatory controls can provide significant overall benefits, but can be difficult to optimize over all diseases, sometimes compromising productive efficiency.

CHAPTER OVERVIEW

This chapter utilizes our framework for synthesizing *productive efficiency* results and causal factors introduced in Chapter 1 (**Exhibit 2**). The framework, which serves as an organizing structure for this chapter, describes four levels of causality driving productive efficiency differences in disease treatment: the most immediate driver, *provider behavior,* which refers to the specific treatment patterns observed; the economic *incentives and constraints* that influence provider behavior; the *structure* of the health care system; and the *regulations* that shape that system.

¶ **Productive efficiency.** The productive efficiency of disease treatment is a function of the inputs used and the outcomes achieved in treating a disease. (A detailed description of our methodology for assessing productive efficiency is provided in Chapter 1, as well as later in this chapter.)

¶ **Provider behavior.** In each of the case studies, we have outlined the "average" behaviors of providers in each country in treating the disease and the impact of these behaviors on relative inputs and outcomes. To identify patterns in these provider behaviors across cases, we have classified behaviors into six categories:

Provider behavior	Definition	Examples
Care triaging	Selection of patients for screening, diagnosis, or treatment; matching of treatment intensity to patient condition	• U.K. more intense triaging of diabetics for care • U.S. broad-based screening program
Treatment duration	Length of stay (LOS) for inpatient treatment, or duration of post-treatment recovery	• German long LOS for most inpatient care
Staffing levels	Number of physicians, nurses, and other staff per patient in general inpatient care	• U.S. higher staffing ratios at hospitals
Setting choice	Choice of inpatient versus outpatient treatment where option exists	• German use of inpatient biopsy for breast cancer versus U.S. outpatient biopsy
Team-based approach	Types and mix of providers and how they interact to deliver care	• U.K. use of dedicated diabetic clinics with multi-disciplinary care teams
Technology adoption	Use of more advanced equipment or care practice for screening, diagnosis, or treatment	• U.S. use of mammograms for breast cancer screening • Faster penetration of laparo-scopic surgery for chole-lithiasis in U.S. than in U.K.

Details of our categorization of the specific treatment variations observed across the disease cases are provided in **Exhibit 3.**

¶ **Provider incentives and constraints.** Within the range of accepted medical practice and knowledge, incentives and constraints strongly influence how providers treat patients. *Incentives*, economic and sometimes noneconomic, include both those that stem from the direct reimbursement mechanism and others as well, some of which may indirectly affect providers' incomes. Providers also face many "supply-side" *constraints*, which we have categorized as follows: controls on the number of physicians (overall and by specialty), on the number and size of hospitals, on capital expenditures by hospitals or physicians, and on providers' ability to substitute inputs or types of care (e.g., barriers between inpatient and outpatient services, "gatekeeper" or other physician referral restrictions).

¶ **Health care system structure.** Chapter 2 describes the significant differences across the three countries in the way the interactions among consumers (or their employers), payors, and providers are structured. These structural differences are an important driver of economic incentives and constraints for providers, and thus ultimately, of productive efficiency. As in Chapter 2, our causal analysis focuses on two specific aspects of the health coverage and care provision markets: the degree of *product integration and pricing mechanisms* in the various care "products" being bought and sold and *competitive intensity* .

¶ **Regulation.** The structure of the care provision and health coverage markets in each country along these two dimensions is strongly determined by government regulation, since regulation can set specific boundaries on the nature and form of competition or on the extent to which care is "bundled" or coordinated by payors and providers. Regulation may also explicitly or implicitly create supply constraints for providers.

* * *

Exhibit 4 summarizes the causal analyses for each disease case by national comparison and illustrates the relative importance of each factor in explaining the differences observed at each level of the causality framework for each disease comparison. The U.S. is used as the benchmark country because it was the most consistently productive nation across the bilateral comparisons.

In the remainder of this chapter, we discuss each level of causality in turn. Our objective is to summarize the most important differentiating factors at each level

and draw general conclusions about characteristics of provider behavior, system structure, and regulation that contribute to higher productive efficiency.

Caveats

In reviewing this chapter, three important caveats should be kept in mind:

1. **Timeframe**. Our analyses are based on the mid to late 1980s. We discuss the major changes all three countries have made since then in Chapter 2. Chapter 9 includes a more complete discussion of the changes and their implications for health care organizations and policymakers.

2. **Norms**. We have compared "average" treatment patterns and system characteristics for each country. Our study neither attempts to quantify nor to explain variations in behavior *within* each country. We do investigate the mix of behaviors when the causality analysis indicates the mix is important, e.g., case rates versus fee-for-service (FFS) in hospital reimbursement.

3. **Focus**. We focus only on productive efficiency in analyzing drivers of observed performance differences. Although there are other significant performance differences among the three countries (e.g., price levels and administrative costs which are discussed in Chapter 8), we do not explore the sources and drivers of these differences, nor do we assess the potential impact of actions to improve productive efficiency on these other performance dimensions.

DIFFERENCES IN PRODUCTIVE EFFICIENCY ACROSS COUNTRIES

As for each disease case, our synthesis of productive efficiency differences across countries begins with an analysis of differences in inputs and outcomes. We then use these results to explore the nations' relative productive efficiency in the treatment of these diseases. We conclude that, while no single country had the highest productive efficiency across all four diseases, the U.S. was more productive than Germany and had mixed results relative to the U.K., Germany was less productive than the U.S. but had mixed results compared with the U.K., and the U.K. had mixed results compared with both nations.

Differences in inputs and outcomes

The three countries applied significantly different levels of inputs to treating the four diseases studied; the pattern is consistent across diseases (**Exhibit 5**). In each case, the U.K. used the fewest inputs, Germany the most[1], and the U.S. was between the U.K. and Germany. The differences in input usage are quite large – on average the U.K. used about 23-percent fewer resources than the U.S. and Germany used about 39-percent more resources than the U.S.[2]

The consistent ordering of the three countries implies that, relative to the U.S., there are characteristics of the U.K. system that encouraged lower input usage and characteristics of the German system that encouraged higher input usage. (We discuss these system characteristics later in our analysis of the causal factors driving productive efficiency differences.) The magnitude of the input differences suggests that these system characteristics had a powerful influence on provider behavior and, thus, on relative input consumption.

Outcomes – defined in terms of the estimated benefit from the disease treatment process in each country – also varied across countries, but did not follow the simple pattern of input differences (**Exhibit 6**). The U.S. had the best outcomes for the cancers, the U.K. had the best outcomes for diabetes, and outcomes for cholelithiasis appear similar for the U.S. and Germany (with Germany having slightly better outcomes) but were worse in the U.K.

Implications for relative productive efficiency

One country was clearly more productive than another for a specific disease if it achieved equal or better outcomes with fewer inputs. In comparisons in which one country had both higher inputs and outcomes, it is not immediately clear which country was more productive. In these comparisons, we estimate relative productive efficiency by calculating the average productivity of both nations; if the country with the higher inputs and outcomes also had higher average productivity and there is no reason to believe that the treatment (production) process shows increasing returns with additional inputs at the per-case level, we conclude that it was more productive.

If the country with higher inputs and outcomes had lower average productivity, however, we can determine which country was more productive only with detailed knowledge of the treatment process. In those comparisons where we cannot make a productive efficiency determination, we conduct a cost-effectiveness analysis to assess which nation had the preferred input/outcome combination and comment on overall economic

---

[1] Excluding diabetes, where we did not include Germany in the study.

[2] Weighted average of input usage across the four disease case studies.

efficiency (**Exhibit 7**). (Refer to Chapter 1 and each disease case chapter for more detail on our methodology for estimating relative productive efficiency.)

Using this methodology, we assess the relative productive efficiency of disease treatment for the 10 cross-country comparisons included in our disease case studies (i.e., 4 diseases across 3 countries, with diabetes excluded for Germany).

¶ **Clear comparisons**. Out of the 10 possible cross-country comparisons, there were 4 cases of unambiguously higher productive efficiency in which 1 country dominated another by achieving better outcomes with fewer inputs (**Exhibit 8**):

- *The U.S. was more productive than Germany in the treatment of breast cancer.*

- *The U.S. was more productive than Germany in the treatment of lung cancer.*

- *The U.K. was more productive than the U.S. in the treatment of diabetes.*

- *The U.K. was more productive than Germany in the treatment of breast cancer.*

¶ **Estimated comparisons**. In four comparisons, the nation with higher inputs and outcomes had higher average productivity, and we could assume that the treatment process did not show increasing returns. In one additional comparison, the nation with higher inputs and outcomes had lower average productivity, but detailed knowledge of the production process allows us to draw a conclusion (**Exhibit 9**):

- *The U.S. was more productive than the U.K. in the treatment of cholelithiasis,* since the U.S. had higher inputs and outcomes and 73-percent higher average productivity than the U.K. in the treatment of this disease.

- *The U.S. was more productive than the U.K. in the treatment of lung cancer,* since the U.S. had higher inputs and outcomes and 82-percent higher average productivity than the U.K.

- *Germany was more productive than the U.K. in the treatment of cholelithiasis.* Germany had higher inputs and outcomes and 14-percent higher average productivity in the treatment of this disease.

- *Germany was more productive than the U.K. in the treatment of lung cancer,* since Germany had higher inputs and outcomes and 33-percent higher average productivity.

- *The U.S. was more productive than Germany in the treatment of cholelithiasis.* The U.S. had lower inputs and outcomes but 52-percent higher average productivity in the treatment of this disease. Given that the U.S. uses 72-percent fewer inputs per surgery with identical surgical outcomes – due to its much shorter hospital stays, shorter recovery periods, and broader adoption of the laparoscopic procedure – and that Germany's better outcomes on a per-case basis resulted solely from its higher surgical frequency, we conclude that the U.S. was more productive.

¶ **Indeterminate comparisons**. For one case comparison only, we cannot determine which country was more productive, but can only comment on the preferred input/outcome combination (**Exhibit 10**):

- *For breast cancer, it is unclear whether the U.S. or the U.K. was more productive.* The U.S. had both higher inputs and outcomes but could range from 35-percent higher to 13-percent lower average productivity over a range of potential baseline estimates for breast cancer survival without treatment; we therefore cannot comment on relative productive efficiency because it is likely that the treatment process shows diminishing returns. Examination of the two nations' treatment processes suggests that each country was employing more productive care approaches at different stages of treatment, with offsetting effects at the overall disease treatment level.

  However, we are able to conclude that the U.S. has the preferred input/outcome combination. In the U.S. prices, the U.S. spent an additional $32,000 per LY, which is below the $100,000 benchmark (described in Chapter 1 and Appendix 1E) and therefore can be considered cost-effective. In the U.K. prices, however, the U.S. spent only an additional $13,000 per LY; as health care expenditures under $30,000 are considered cost-effective (described in Chapter 1 and Appendix 1E), the U.K. could have benefited from higher inputs and expenditures in the treatment of this disease.

Summarizing our cross-country disease case comparisons, we conclude the following:

¶ *No single country had the highest productive efficiency across all four diseases.*

¶ *The U.S. was more productive than Germany and had mixed results relative to the U.K.* The U.S. was more productive than Germany in all three diseases compared and more productive than the U.K. in the treatment of cholelithiasis and lung cancer. In diabetes, the U.S. was

less productive than the U.K., and we could not determine which country was more productive in breast cancer.

¶ *Germany was less productive than the U.S. but had mixed results compared with the U.K.* Germany was less productive than the U.S. in breast cancer, lung cancer, and cholelithiasis. Comparing Germany with the U.K., Germany was more productive in cholelithiasis and lung cancer, and the U.K. was more productive in breast cancer.

¶ *The U.K. had mixed results, as it was more productive than the other nations in one disease each.* The U.K. was more productive than the U.S. in diabetes treatment, but less productive than the U.S. in cholelithiasis and lung cancer. Relative to Germany, the U.K. was more productive in the treatment of breast cancer, but less productive in cholelithiasis and lung cancer.

DIFFERENCES IN PROVIDER BEHAVIOR

These productive efficiency differences resulted from significant differences in the behavior of providers, particularly physicians, in managing and treating diseases. In the U.K., providers tended to focus treatment more narrowly on a selected subset of patients and to adopt technology more cautiously; the combination of these factors tended to drive the consistently lower U.K. input usage but led to higher productive efficiency only in some diseases. In Germany, providers used the inpatient care setting more often and had longer treatment lengths, particularly hospital stays, than the U.S.; these behaviors caused Germany's consistently higher input usage and lower productive efficiency.

Major differences between the U.S. and the U.K.

Of the six categories of provider behavior, two emerge as most important in driving productive efficiency differences between the U.S. and the U.K.: 1) slower and more selective technology adoption in the U.K.; and 2) more intense triaging of care in the U.K. In addition, the team-based care approach was important in driving the U.K.'s higher productive efficiency in diabetes. While the U.K. also had lower rates of surgery and other treatments across diseases, these differences contributed to its lower overall input usage, but not specifically to its relative productive efficiency (i.e., they placed the U.K. at a different point on the disease treatment production function rather than on a different production function).

1. Slower and more selective technology adoption in the U.K. The U.K. providers generally adopted new technologies later than the U.S. providers, used these technologies less frequently, and used them for more limited

purposes. For example, substitution of laparoscopic cholecystectomies for open cholecystectomies occurred later in the U.K. than in the U.S.; installation and use of CT scanners for lung cancer diagnosis and staging was also more limited in the U.K. than in the U.S.; and the U.S. adopted mammograms for broad-based screening of breast cancer, while the U.K. used them only for diagnosis. The one exception was the U.K.'s broader adoption of FNA for breast cancer biopsy, a technology that was not capital-intensive and reduced demand for surgical and hospital bed resources, but was less definitive (at that time) relative to surgical biopsy, the predominant choice in the U.S.

The U.K.'s slower adoption of technology tended to reduce input usage, but led to lower productive efficiency in some diseases and higher productive efficiency in others. In cholelithiasis, laparoscopic surgery reduced costs through shorter LOS and shorter recovery times; the slower U.K. adoption of this technology, therefore, increased input usage and lowered productive efficiency relative to the U.S. In lung cancer, the U.K.'s lower use of CT scans for lung cancer diagnosis and staging may have led to less effective use of surgical resources by presenting a less-than-optimal mix of surgical patients, thereby compromising surgical outcomes. In combination with the U.K.'s longer length of hospital stay after surgery and more hospital stays with chemotherapy, its lower use of CT scans led to lower productive efficiency. In breast cancer, in contrast, the U.K.'s adoption of mammography for diagnosis-only versus for broad-based screening as in the U.S., coupled with its adoption of FNA, appeared to reduce inputs and promote higher productive efficiency in some phases of the disease treatment process.

2. More intense triaging of care resources in the U.K. The U.K. providers were more intense in triaging care relative to the U.S., leading to lower input use but not necessarily to greater productive efficiency. For example, specialized diabetes care in dedicated clinics was only given to about one-third of the diabetics in the U.K., while another third received no additional routine physician care at all (beyond the usual general practitioner [GP] visits that occurred with the healthy population). By selectively providing intensive preventive and management care to the most vulnerable diabetics, the U.K. was able to use fewer resources than the U.S., yet achieve better outcomes by lowering complication rates. In breast cancer treatment, the lack of a broad-based screening program in the U.K. – regardless of technology employed – significantly reduced inputs throughout the treatment process without a corresponding outcome difference, likely increasing the U.K.'s productive efficiency relative to the U.S. In lung cancer, the U.K.'s more intense triaging of patients for surgery – through its complex and restricted referral processes – led to lower input use but potentially impeded effective staging, thereby lowering productive efficiency.

3. Team-based care approach of the U.K. providers in specialized clinics for diabetes. The use of multispecialty provider teams practicing in specialized diabetic clinics in the U.K. appears to have contributed to the lower U.K.

complication rates relative to the U.S. During the period of the study, such teams and clinics were used much less frequently in the U.S., at least for Type II diabetics. This team-based care approach seems to have been especially productive for diabetes, because it is a complex condition involving many organ systems and requiring many types of interventions for effective management. Team-based care approaches appeared to be less important for the other diseases studied because the conditions and treatments are more circumscribed.

Major differences between the U.S. and Germany

Two very different categories of provider behavior appear to have driven the relative productive efficiency differences observed between the U.S. and Germany: 1) substantially longer treatment duration, particularly hospital LOS; and 2) greater use of the inpatient care setting in Germany. In contrast to the factors driving the U.S. and the U.K. productive efficiency differences, these two treatment differences consistently resulted in higher input usage and lower productive efficiency in Germany relative to the U.S. While Germany also had higher surgical and other treatment frequencies compared with the U.S. across most diseases, these differences led to higher input usage, but not specifically to lower productive efficiency; more surgeries, for example, clearly placed Germany further out on the disease treatment production function but not necessarily on a different production function.

1. Substantially longer treatment duration in Germany. Germany's average LOS for hospital treatment was significantly longer than the U.S. LOS in all three cases, and was also longer than the U.K. LOS in most cases. Germany thus used significantly more inputs for each inpatient procedure than either the U.S. or the U.K. without any apparent outcome benefit. Although Germany's hospital staffing levels were relatively low, its long LOS more than offset its lower staffing, resulting in higher net input usage relative to the other countries.[3] In cholelithiasis, Germany also had a much longer recovery period for patients treated with both the laparoscopic and open surgical procedures, again without an apparent outcome benefit.

2. Greater use of the inpatient care setting in Germany. Compared with the U.S. providers, German providers used the inpatient setting more frequently than the generally less resource-intensive outpatient setting. In the cancer cases, this contributed to higher input usage and lower productive efficiency in Germany, since there was no apparent outcome benefit associated with this inpatient choice. For example, Germany used the inpatient setting for surgical biopsy in breast cancer whereas the U.S. used predominantly the outpatient setting. When coupled with Germany's longer hospital LOS, as described above, this difference caused Germany to use 30-percent more resources in

---

[3] Open cholecystectomy is a sole exception; although Germany's LOS is longer, its net input usage for this procedure is very similar to the U.S. and the U.K., largely because of lower staffing levels.

treating this disease. Similarly, Germany's use of the inpatient setting for chemotherapy in lung cancer, when coupled with Germany's longer hospital LOS, caused it to use 20-percent more inputs and to achieve lower productive efficiency relative to the U.S.

<p style="text-align:center">* * *</p>

Overall, provider behavior differed significantly and systematically across countries. Clear differences in the choices made by front-line care providers resulted in significant variations in relative productive efficiency, primarily by influencing input usage. It is somewhat surprising that these behavior differences were so large, given providers' similar clinical training and access to both technology and medical knowledge in the three countries. The consistency of many of these differences across diseases suggests that underlying system characteristics had a powerful influence on provider behavior.

DIFFERENCES IN PROVIDER INCENTIVES AND CONSTRAINTS, SYSTEM STRUCTURE, AND REGULATION

The disease cases show that providers responded predictably and consistently to the incentives and constraints created by the structure of their health care system, which in turn was influenced by regulation. While there are many ways to characterize health care system structure, we found two factors to be most important: product integration and pricing, and competitive intensity. Chapter 2 contains a more detailed description of the structure of each of the three health care systems at the time of our assessment, focusing on these two factors. In this synthesis, we highlight where and how these factors strongly influenced productive efficiency.

In the U.K., slower technology adoption and more intense care triaging was consistent with its relatively fixed salaries for physicians and fixed budgets for hospitals, and with its tight constraints on the supply of physicians, hospitals, and capital; these differences were shaped directly by the U.K. health care system's strong regulation, low competitive intensity, and higher product integration –particularly in diabetes. Germany's greater use of inpatient care resulted directly from incentives for hospitals and physicians to fill hospital beds and constraints on the substitution of outpatient care for inpatient care. These incentives and constraints, in turn, stemmed from the German health care system's regulation, lower competitive intensity, and per diem hospital product.

Major differences between the U.S. and the U.K.

Two differences in incentives and constraints between the U.S. and the U.K. were most important in driving provider behavior differences: 1) more activity- and technology-increasing incentives for physicians in the U.S.; and 2) tight constraints in the U.K. on the supply of capital, physicians, and hospitals. Differences in the level of competitive intensity and in the nature and degree of integration in the care provision products most directly influenced the physician incentive differences, whereas direct regulation in the U.K. system created the supply constraints.

1. More activity- and technology-increasing physician incentives in the U.S.
The U.S. specialist physicians had strong incentives to provide as much care to as many patients as possible and to quickly adopt technologies for which there was patient and/or payor demand. In contrast, the U.K. specialists had no economic incentives relative to the amount of care provided; to the extent that they had alternative income sources (e.g., private practice), 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 they freed up constrained care resources. These different physician incentives best explain the slower and more selective technology adoption and to some extent the more intense care triaging in the U.K. compared with the U.S. in the case studies. In particular, physician incentive differences promoted rapid adoption of the laparoscopic technology in cholelithiasis in the U.S. In breast cancer, physician incentives also help explain the U.S.'s widespread use of mammography, surgical biopsy rather than FNA, and the two-step biopsy/surgical treatment protocol versus the integrated one-step approach that dominated in the U.K. In addition, physician incentive differences influenced the U.S.'s broader screening program in breast cancer as well as the less differentiated and more fragmented care approaches in diabetes.

The sources of these different physician incentives were major differences in health care system structure, particularly in the degree of competitive intensity and in the integration and pricing of care provision products negotiated between payors and physicians. We briefly describe the differences between the U.S. and the U.K. systems below.

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

Although they faced some price-based competition and could have bundled and negotiated services in a variety of ways, the U.S. payors were not an effective force to counterbalance physician's activity- and technology-increasing incentives. The U.S. payors' inability to modify these incentives may be inherent

in the predominant form of health coverage products (i.e., indemnity FFS coverage) at the time, as well as in the lack of cost and outcome information and in the relative physician power. In indemnity coverage products, payors generally negotiated physician services on an FFS basis, with payment terms and levels set by locally prevailing practices. The U.S. physicians as a group were able to use local medical associations and specialty societies to promote changes in these locally prevailing practices, leading to increases in standards of care and thus to health insurance coverage for higher activity levels or new technology adoption.

In fact, the U.S. payors were often forced to adopt such coverage in their health insurance products in order to meet employer and consumer demands for new treatment approaches and thereby to compete effectively, as was the case in breast cancer screening and laparoscopic surgery for cholelithiasis. In addition, the U.S. payors tended to provide health coverage for 1-year terms and faced relatively high annual turnover in their member populations (i.e., 20 to 40 percent); this may have limited their willingness to make investments in preventive or education-oriented care that had a longer-term payback, except when they were responding to clear employer or consumer demands.

In the U.K., specialist physician services 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 (RHAs); however, specialists could also earn additional income from treating private patients on an FFS basis. GP services took the form of FFS-based contracts primarily, with rates negotiated on the basis of a complex formula through the NHS. Neither the U.K. specialists nor GPs competed in any meaningful way for NHS patients, given the tight physician supply, or for NHS contracts.

To some extent, physician incentives in the U.K. were also shaped by the structure and functions of the NHS. As the organizing force for health care in the U.K. and as the employer of many of the physicians, the NHS (through the RHAs and local district health authorities [DHAs] and the Family Practitioner Committees) was able to influence physicians through training, dissemination of information and guidelines, and, if necessary, through direct authority. The NHS also took a holistic, or systemwide, view in making care investments, as it provided health coverage for the lifetime of the entire population. For example, the NHS would often identify specific diseases for specialized care approaches (carve out), which enabled active management of care delivery, as for diabetes. The NHS, therefore, contributed to physicians' greater concern for cost-effectiveness in the U.K. and thereby to their greater willingness to adopt technology more slowly and selectively and to triage care more intensely. And by internalizing the interaction between payors and physicians, the U.K. may have been better able to apply these controls than the U.S. payors were able to do through arms-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.

Overall, the U.S.'s greater competitive intensity in both care provision and health coverage markets, along with its activity-based physician services, gave rise to stronger incentives to provide more care and to adopt technologies more quickly and broadly than in the U.K. The increased care provision and the early adoption of new technology was sometimes more productive, but other times less so (**Exhibit 11**).

¶ In diabetes, the combination was clearly less productive, as the disease benefited from the U.K.'s more integrated and actively managed care approach, as well as from the NHS's systemwide perspective in making preventive and other care investments.

¶ In cholelithiasis, the combination led to higher productive efficiency, as physicians adopted the more cost-effective laparoscopic technology more quickly to meet consumer demand and encountered high payor acceptance for this substitute.

¶ In breast cancer, the overall impact is difficult to assess; however, it appears that the combination led to a very broad-based mammographic screening program that was not productive in its entirety, in part due to market failure stemming from the consumer's lack of knowledge about the benefits of screening and from the fact that consumers were shielded from the cost of screening through insurance.

¶ In lung cancer, the combination led to more surgeries and possibly contributed to more effective care triaging, but productive efficiency in this disease was more influenced by the supply constraints imposed through the U.K. regulation, as discussed below.

2. Tight constraints on capital, physician, and hospital supply in the U.K. through regulation. The U.K. exercised strict controls over the number of physicians and the number and capacity of hospitals through the NHS budgeting process and regulations. In the U.S., the supply of physicians and hospitals was relatively unconstrained, although licensing requirements served as an entry barrier to some degree. These supply constraint differences contributed to the differences in the amount and intensity of care provided. The U.K. physician and hospital capacity constraints forced providers to be more intense in care triaging – as practiced by GPs in diabetes – or to substitute procedures that did not use constrained resources – such as FNA for breast cancer biopsy, which does not consume scarce hospital or surgeon resources. Triaging could either be explicit (e.g., through providers' decisions to limit care or resources) or implicit (e.g., through patient queuing).

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 aggregate system impact of a given technology; for example, the RHAs and DHAs could consider the extent to which more CT scans for lung cancer diagnosis and staging could impact overall systemwide usage and costs. These funding limits and allocation processes contributed to the slower adoption and narrower use of capital-intensive technology in the U.K. relative to the U.S., such as mammographic equipment, CT scans, and laparoscopic equipment. In addition, they may have precluded substitution of more capital-intensive resources, such as CT scans, for other care resources. It is possible, however, that the lower installation and use of CT scans for lung cancer diagnosis and staging was part of a rational policy to reduce CT use across the system; similarly, it is possible that the U.K.'s slower adoption of laparoscopy resulted from concerns about increasing the overall cholecystectomy rate.

In the U.S., individual hospitals and physicians made their own decisions on capital investment; 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 incorporate reimbursement for these technologies into their health coverage, as described earlier.

Additionally, restrictions on referral processes in the U.K. (specifically, the need to go through a GP gatekeeper before going to a district hospital and, in turn, another gatekeeper before going to a regional center) contributed to its more intense care triaging and lower use of care inputs. This reduction in overall input usage was accomplished in lung cancer through the restricted process by which patients received CT scans to support diagnosis and staging of the cancer, and eventually received surgery, if appropriate. In contrast, the U.S. had relatively few controls on referral processes at the time, with the exception of some managed care plans.

Overall, the U.K. supply constraints that resulted directly from regulation played a major role in driving the more intense care triaging and, to some extent, the slower technology adoption in the U.K. But similar to the differences in physician incentives, these constraints promoted more productive provider behavior in some diseases, such as diabetes, and less productive treatment in other diseases, such as lung cancer and cholelithiasis.

Major differences between the U.S. and Germany

Three main differences in incentives and constraints between the U.S. and Germany were most important in driving Germany's longer treatment lengths

and greater use of inpatient services: 1) stronger incentives for German hospitals to increase hospital LOS and occupancy, amplified by similar specialist physician incentives; 2) regulations on hospital supply in Germany that actually led to surplus capacity; and 3) constraints in Germany on the substitution of outpatient care for inpatient care. These differences resulted from the German health care system's lower competitive intensity and per diem hospital product, which in turn were heavily influenced by Germany's regulatory structure.

1. Incentives for German hospitals to increase hospital LOS and occupancy, amplified by physician incentives. German hospitals had strong incentives to increase their LOS, while the U.S. hospitals had some incentive to reduce LOS. These incentive differences led to Germany's significantly longer LOS for all inpatient procedures and to the lower productive efficiency observed in all three case study comparisons.

Differences in the hospital product that was negotiated with payors, in competitive intensity, and in potential regulatory threats caused these incentive differences. Specifically, 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 [DRG] system) and through a mixture of approaches from private insurers, including FFS, per diem, and case rate bases; case rates represented about 35 to 40 percent of an average U.S. hospital's total cost. While both German and U.S. hospitals competed aggressively for patients, only the U.S. hospitals faced any competition in their negotiations with payors; each German hospital, by law, negotiated with all payors as a block for annual per diem fees. The U.S. private payors also faced some price-based competition for members and therefore had both the incentive and ability to manage hospital costs and LOS, whereas regulation precluded German payors from competing on price and from bundling hospital care in any different or better way to manage LOS. Furthermore, German hospitals faced the threat of regulatory review and potential capacity cuts if their utilization fell below 85 percent; by maintaining high occupancy, hospitals avoided this threat.

In addition, incentives for specialist physicians in Germany reinforced these LOS-increasing hospital incentives. Employed by their hospitals and paid a flat salary, these physicians on the surface had no direct economic incentive to increase the amount of care provided; however, they had clear "noneconomic" incentives to further the interests of their employers – the hospitals – and therefore had a relatively strong incentive to increase the amount of inpatient care they provided. In addition, the German hospital department chief physicians had incentives to increase the workload of their hospitals, since they received FFS income from private patients in addition to their hospital salaries. Given that the department was allowed bed capacity for private patients in a relatively fixed ratio to its utilized public beds and that the workload of the hospital from publicly funded patients had an indirect but significant effect on

the chief's private income, incentives existed to increase utilization of public hospital beds.

Overall, Germany's per diem hospital product and lower competitive intensity, both of which stemmed from strong regulation, created very different hospital incentives, which in turn led to much longer inpatient stays and lower productive efficiency relative to the U.S.

2. Ineffective regulatory controls on hospital (and physician) supply in Germany, creating excess capacity. Hospital capacity in Germany was seemingly constrained, while the U.S. capacity was relatively unconstrained; yet Germany had more hospital beds per capita than the U.S. 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. The primary contributor to the German system's excess hospital capacity was the fact that 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 U.S., hospitals in Germany did not need to aggressively downsize or close in the face of falling demand. Furthermore, the regulations and system structure that increased hospital capacity in Germany also increased the number of hospital-based physicians.

3. Constraints on inpatient/outpatient substitution in Germany. The inpatient and ambulatory segments of care in Germany were strictly separated; they were 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, substitution of less expensive outpatient procedures for inpatient procedures did not occur in Germany to the extent it did in the U.S., 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 while German providers used an inpatient surgical biopsy; similarly, the U.S. replaced inpatient chemotherapy with outpatient chemotherapy more quickly than Germany.

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 U.S.

MAJOR CROSS-DISEASE CONCLUSIONS
FROM CAUSALITY ANALYSIS

While the causal factors driving productive efficiency vary across the four diseases and countries, several major themes emerge:

¶ Providers treated patients very differently across the three countries in ways that were fully consistent with their incentives and constraints, despite relatively similar clinical training and access to medical expertise and technology. These provider behavior variations – particularly regarding technology adoption, care triaging, team-based approaches, treatment duration, and care setting – led to large differences in resource use across diseases, significant outcome differences in some diseases, and substantial variations in productive efficiency.

¶ More stringent supply constraints resulting from a tight global budget reduce overall input usage, but do not necessarily lead to higher productive efficiency. As was seen in the U.K., the overall budget constraint led that health care system to use about 23-percent fewer resources than in the U.S. in the disease cases, but the U.K. was clearly more productive in only one disease – diabetes.

¶ More integrated care products at the disease level or phase of disease level (i.e., hospital episode or case) promote higher productive efficiency, particularly in some diseases. Care integration across multiple diseases, but only for one care component, such as hospital or specialist physician services (e.g., through the U.K. annual hospital budgets and fixed annual specialist salaries), may lead to less productive provider behaviors such as "bed-blocking" in the U.K. (which increased LOS) or possibly slower adoption of productive technologies.

¶ Higher competitive intensity among payors and physicians leads to greater responsiveness to demand and, thus, to faster and broader technology adoption and less intense care triaging, likely regardless of any global budget constraint. This response can sometimes be more productive (i.e., when the new technologies are cost-effective and/or when providers' incentives are aligned with productive efficiency), but other times, less productive (i.e., when new technologies have only marginal outcome benefits and providers' incentives are not aligned with productive efficiency, such as with FFS incentives).

¶ The nature and extent of regulation play a major role in driving productive efficiency of the health care system, not only by influencing the degree of care integration and competitive intensity in the system,

but also by creating supply and other constraints on providers. Such controls, whether on capital, on the substitution of alternative inputs, or on levels of hospital and physician supply, may provide significant overall benefits yet are difficult to optimize over all diseases. Thus, they can improve productive efficiency in some circumstances, while reducing it in others.

Less regulated systems will have fewer misdirected supply constraints, increased responsiveness to new technologies, and greater opportunities for experimentation and learning on the part of payors, providers, and other health care stakeholders to investigate which treatment approaches and technologies, incentives, and system structures work best by disease. However, it is unclear whether less regulated systems can capture the advantages of care integration at the disease level achieved by more centrally controlled systems such as the U.K., for chronic diseases in particular, given issues of adverse selection.

* * *

In summary, this synthesis of the four disease case studies shows that significant variations in care treatment approaches and provider behavior resulted in large productive efficiency differences across the three countries. These behavior differences were a direct result of differences in providers' incentives and constraints, which, in turn, were influenced by the characteristics of the health care system structure and government regulation.

Overall, no single country emerged as most productive across all diseases studied, and the relative benefits of competitive intensity, health care product integration, and regulation varied by disease. While these findings may appear confusing, they are actually quite consistent with, and reflective of, the complexity of health care economics. In fact, the mixed nature of our findings across diseases, as in the U.S.-U.K. comparison, illustrates and confirms some fundamental principles of health care economics, which have been discussed and debated widely in the literature. In particular, our findings confirm that while competition can sometimes be helpful in promoting productive efficiency, there is significant potential for unproductive market behavior in health care due primarily to information asymmetry among various system participants, including the consumer. The complexity of medical treatment processes and options, rapidly changing medical technology, and the presence of insurance for most consumers can make it difficult for competitive markets to promote high productive efficiency in some diseases, particularly those requiring more integrated care approaches.

In Chapter 8, we return to and attempt to shed light on the questions posed in Chapter 1 regarding the drivers of differences in overall health care spending

and aggregate outcomes for the three countries. We do this by combining our disease-level productive efficiency assessment with aggregate-level analyses. In this way, we comment on other important factors beyond productive efficiency, including administrative cost levels and relative input prices. In Chapter 9, we draw implications for policymakers and health care organizations, focusing on ways to improve productive efficiency.

Exhibit 1

RELATIVE PRODUCTIVE EFFICIENCY BY DISEASE CASE

| Disease | U.S. vs. U.K. | | | U.S. vs. Germany | | | Germany vs. U.K. | | |
	Inputs	Outcomes	More productive country	Inputs	Outcomes	More productive country	Inputs	Outcomes	More productive country
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		

022 EF 136209/7

Exhibit 2

FRAMEWORK FOR CAUSAL ANALYSIS

```
┌─────────────────────────────────┐
│ Productive efficiency           │
│ • Inputs                        │
│ • Outcomes                      │
└─────────────────────────────────┘
```

```
┌─────────────────────────────────┐
│ Provider behavior               │
│ • Care triaging                 │        Direct driver of productive efficiency
│ • Treatment duration            │
│ • Staffing levels               │
│ • Setting choice                │
│ • Team-based approach           │
│ • Technology adoption           │
└─────────────────────────────────┘
```

> **Vary by disease**

```
┌─────────────────────────────────┐
│ Provider incentives and constraints │
│ • Physician incentives          │
│ • Hospital incentives           │        Direct drivers of provider behavior
│ • Physician supply              │
│ • Hospital supply               │
│ • Capital constraints           │
│ • Substitution constraints      │
└─────────────────────────────────┘
```

```
┌─────────────────────────────────┐
│ Health care system structure    │        Economic interactions among
│ • Product integration and pricing │      participants in health care system
│ • Competitive intensity         │        that drive incentives and constraints
└─────────────────────────────────┘
```

> **Generally common across diseases**

```
┌─────────────────────────────────┐
│ Regulation                      │        The "rules and regulations" that
└─────────────────────────────────┘        shape interactions among participants
```
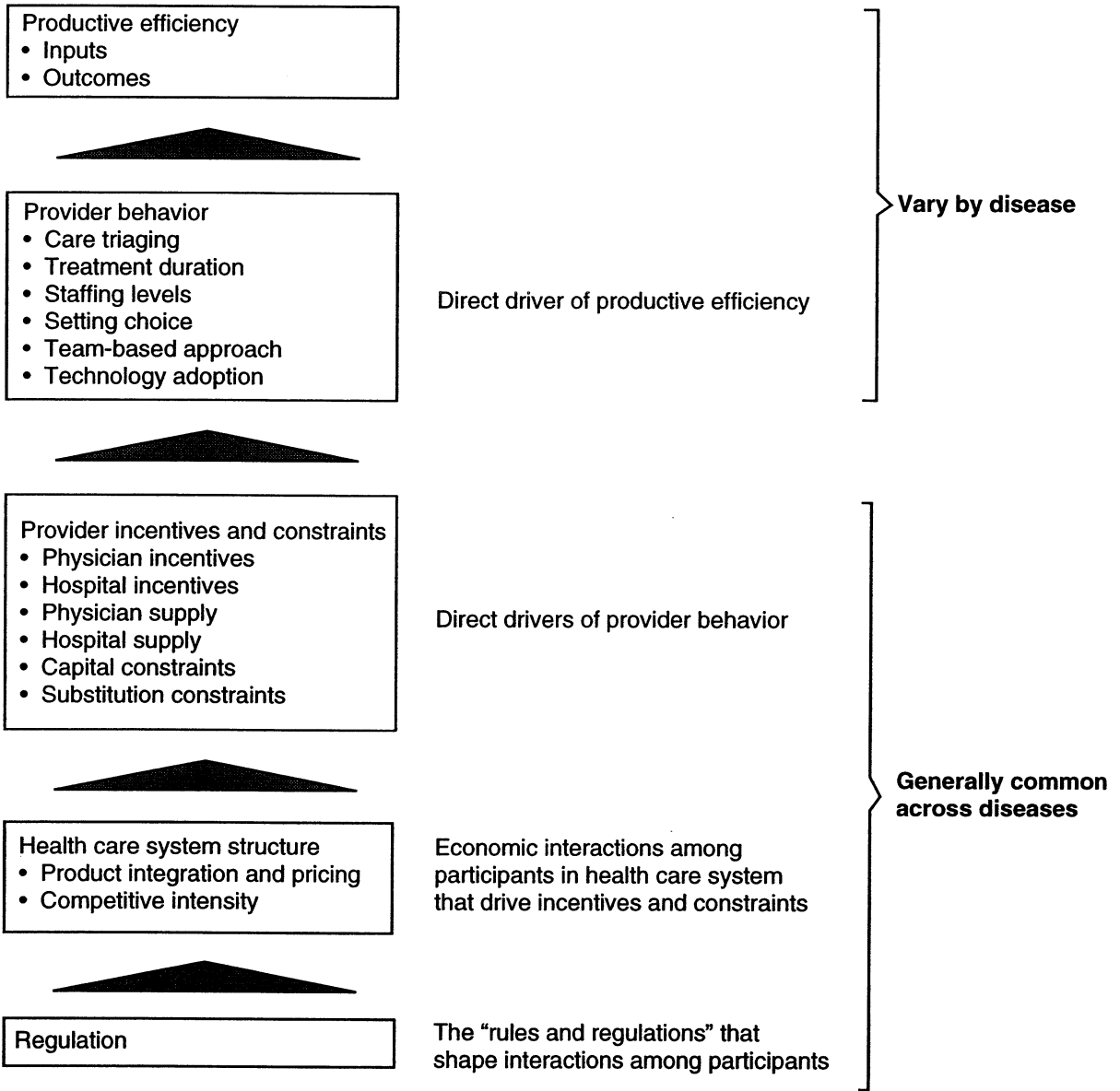
022 ST 140102/7

Exhibit 3

SPECIFIC DIFFERENCES IN PROVIDER BEHAVIOR IDENTIFIED IN CASE STUDIES

Provider behavior difference	U.S. vs. U.K.				U.S. vs. Germany		
	Diabetes	Choleli-thiasis	Breast cancer	Lung cancer	Choleli-thiasis	Breast cancer	Lung cancer
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 in U.K.; more 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	Slightly more frequent surgery in Germany	Broader screening in Germany; more lump-ectomy in Germany	More frequent surgery in Germany
Treatment duration	Longer LOS for compli-cations treatment in U.K.	Longer LOS in U.K.; longer patient recovery time in U.K.	Longer surgical LOS in U.K.	Longer surgical LOS in U.K.; shorter radio-therapy LOS in U.K.	Longer LOS in Germany; longer patient recovery time in Germany	Longer surgical LOS in Germany	Longer surgical LOS in Germany
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.	Lower hospital staffing in Germany	Lower hospital staffing in Germany	Lower hospital staffing in Germany
Setting choice			More inpatient biopsy in U.K.; maintenance of one-step biopsy/surgical treatment protocol	More inpatient chemotherapy in U.K. vs. outpatient in U.S.		Inpatient biopsy in Germany vs. out-patient in U.S.; more inpatient chemo-therapy in Germany	More inpatient chemo-therapy in Germany
Team-based approach	Multispecialty diabetic clinics in U.K.						
Tech-nology adoption		Later laparo-scopic adoption in U.K.	Use of fine needle aspiration (FNA) in U.K. vs. surgical biopsy; no mammo-graphic screening in U.K.	Less use of computerized tomography (CT) scan for staging in U.K.	Slightly later laparo-scopic adoption in Germany	More mammo-graphic screening in U.S.	

Exhibit 4

CAUSAL ANALYSIS OF PRODUCTIVE EFFICIENCY DIFFERENCES

● Major differentiating factor
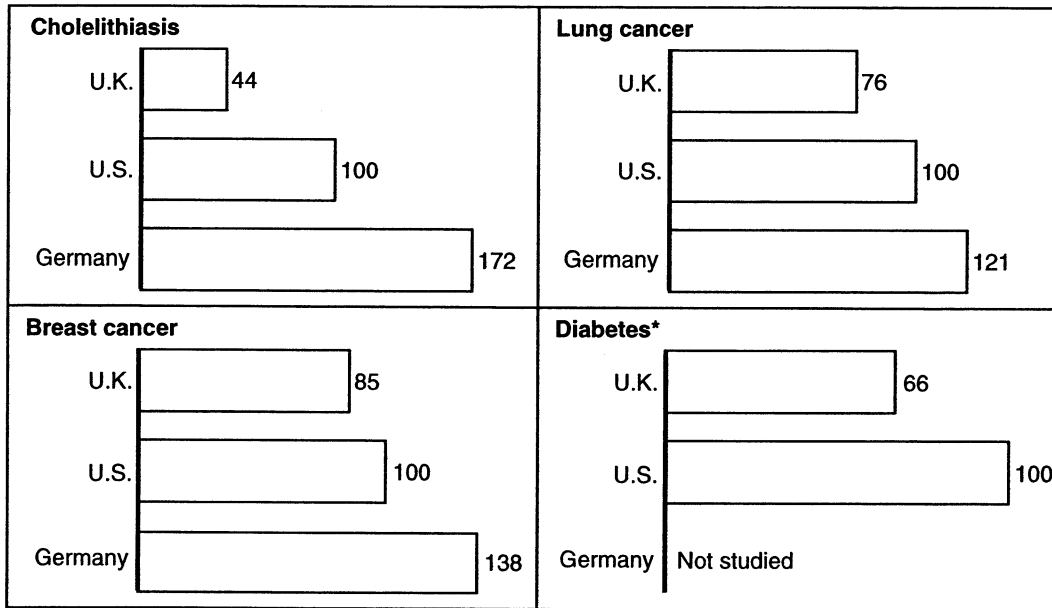◑ Secondary differentiating factor
○ Undifferentiating

	U.S. vs. U.K.					U.S. vs. Germany			
	Diabetes	Choleli-thiasis	Breast cancer	Lung cancer	Overall	Choleli-thiasis	Breast cancer	Lung cancer	Overall
Benchmark	U.K.	U.S.	Indeter-minate	U.S.		U.S.	U.S.	U.S.	
Productive efficiency									
• Inputs	●	●	●	●	●	●	●	●	●
• Outcomes	●	●	◑	●	●	◑	●	◑	◑
Provider behavior									
• Care triaging	●	○	●	◑	◑	○	○	◑	○
• Treatment duration	○	◑	◑	◑	◑	●	●	●	●
• Staffing levels	○	◑	◑	◑	◑	◑	◑	◑	◑
• Setting choice	○	○	◑	◑	◑	○	●	●	◑
• Team-based approach	●	○	○	○	○	○	○	○	○
• Technology adoption	○	●	●	◑	◑	◑	○	○	○
Provider incentives and constraints									
• Physician incentives	●	●	●	○	●	◑	◑	◑	◑
• Hospital incentives	○	◑	◑	◑	◑	●	●	●	●
• Physician supply	●	○	◑	◑	◑	○	○	○	○
• Hospital supply	○	◑	◑	○	◑	●	●	●	●
• Capital constraints	○	◑	●	◑	◑	○	○	○	○
• Substitution constraints	○	○	○	◑	○	○	●	●	◑
Health care system structure									
• Product integration and pricing	●	◑*	◑	○	◑	●	●	●	●
• Competitive intensity	●*	●	●*	◑	●	◑	◑	◑	◑
Regulation	●	●	●	●	●	●	●	●	●

* In these cases, a higher degree of the respective causal factor led to lower productive efficiency

022 EF 136244/7

Exhibit 5
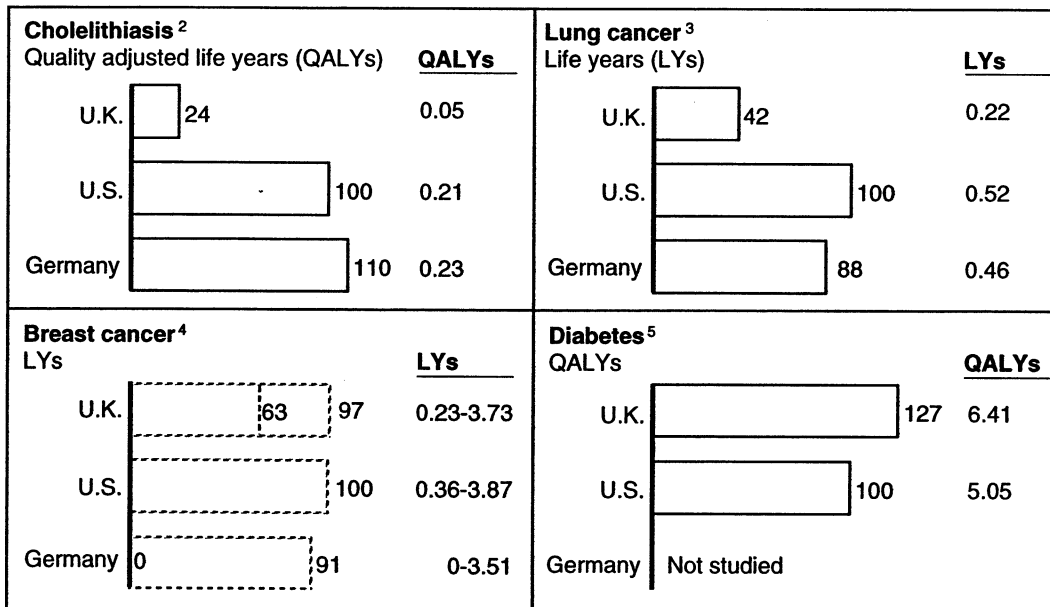
INPUTS BY DISEASE

Index, U.S. = 100% for each disease

Cholelithiasis		Lung cancer	
U.K.	44	U.K.	76
U.S.	100	U.S.	100
Germany	172	Germany	121

Breast cancer		Diabetes*	
U.K.	85	U.K.	66
U.S.	100	U.S.	100
Germany	138	Germany	Not studied

* Weighted average of Type I and Type II

Exhibit 6

OUTCOMES BY DISEASE

Improvement in outcomes due to treatment[1]

Index, U.S. = 100% for each disease

Cholelithiasis [2] Quality adjusted life years (QALYs)		QALYs	Lung cancer [3] Life years (LYs)		LYs
U.K.	24	0.05	U.K.	42	0.22
U.S.	100	0.21	U.S.	100	0.52
Germany	110	0.23	Germany	88	0.46

Breast cancer [4] LYs		LYs	Diabetes [5] QALYs		QALYs
U.K.	63 — 97	0.23-3.73	U.K.	127	6.41
U.S.	100	0.36-3.87	U.S.	100	5.05
Germany	0 — 91	0-3.51	Germany	Not studied	

1 Relative to baseline outcome with no treatment
2 Outcomes based on 30-day frequency of symptoms
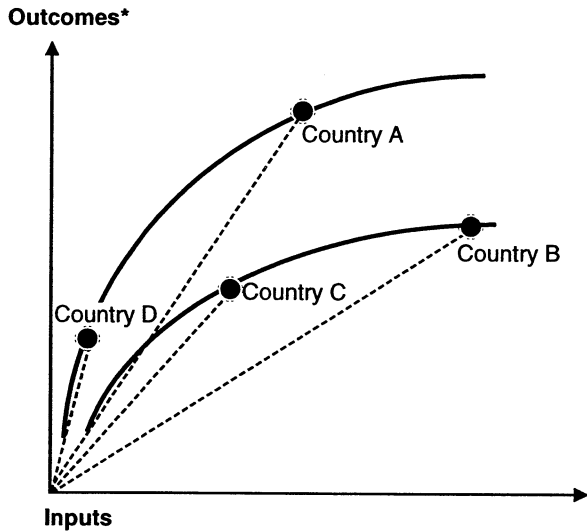3 Outcomes based on 5-year survival
4 Outcomes based on 5-year survival; baseline outcome with no treatment ranged from immediate death to lowest outcomes with treatment (Germany), thereby covering entire range of possibilities
5 Weighted average of Type I and Type II; baseline outcome with no treatment assumed Type I diabetics die within 1 year and Type II diabetics have same QALYs as lowest outcomes with treatment (U.S.)

Exhibit 7

ASSESSING PRODUCTIVE EFFICIENCY

Per case

Outcomes*

Inputs

Comparison 1: A vs. B

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

Comparison 2: A vs. C

A is more productive as 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

* Relative to baseline outcome with no treatment

Note: Solid lines represent treatment processes, or "production functions," while slope of each dotted line represents average productivity

022 ST 140125/7

Exhibit 8

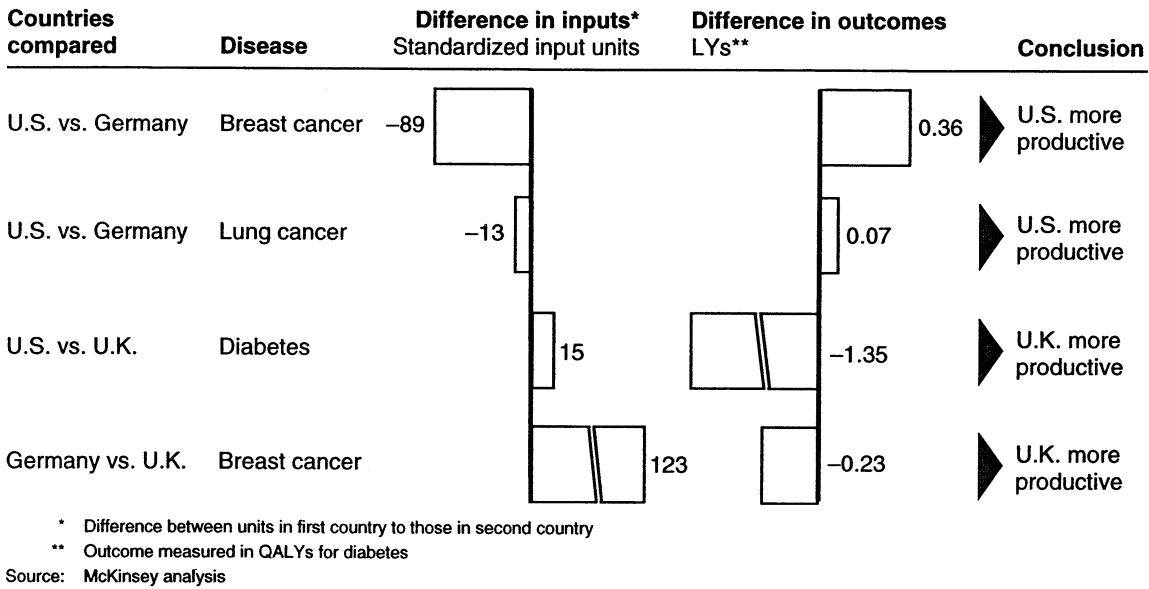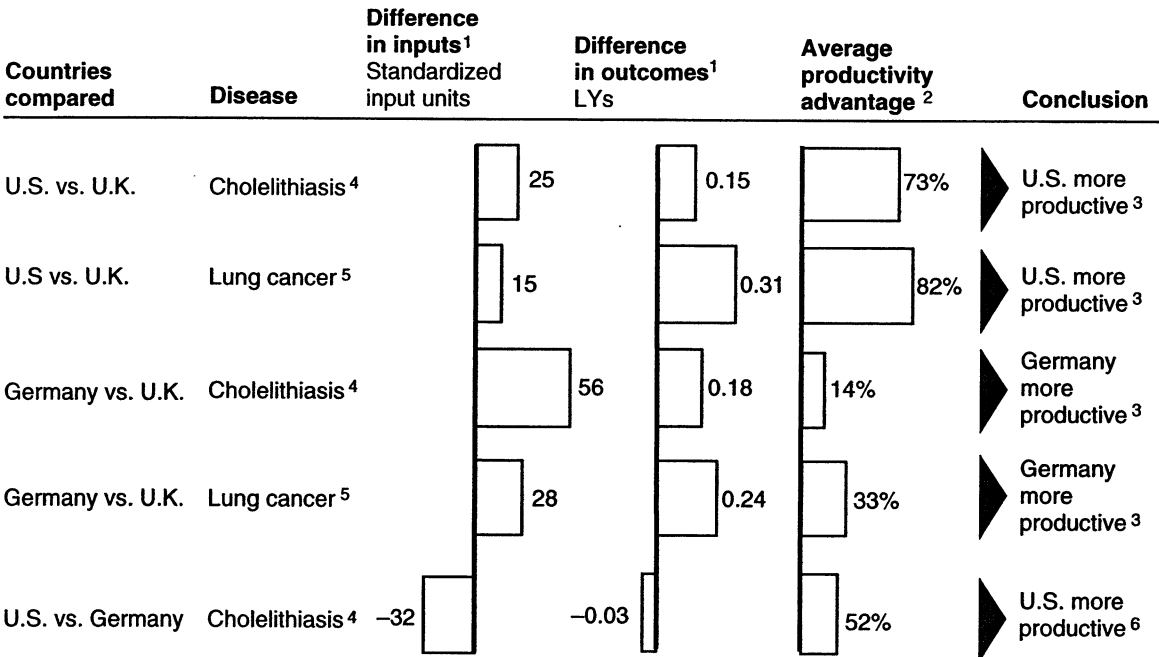# RELATIVE PRODUCTIVE EFFICIENCY BY DISEASE
Clear comparisons

Countries compared	Disease	Difference in inputs* Standardized input units	Difference in outcomes LYs**	Conclusion
U.S. vs. Germany	Breast cancer	−89	0.36	U.S. more productive
U.S. vs. Germany	Lung cancer	−13	0.07	U.S. more productive
U.S. vs. U.K.	Diabetes	15	−1.35	U.K. more productive
Germany vs. U.K.	Breast cancer	123	−0.23	U.K. more productive

    *   Difference between units in first country to those in second country
    ** Outcome measured in QALYs for diabetes
Source: McKinsey analysis

Exhibit 9

RELATIVE PRODUCTIVE EFFICIENCY BY DISEASE

Estimated comparisons

Countries compared	Disease	Difference in inputs[1] Standardized input units	Difference in outcomes[1] LYs	Average productivity advantage [2]	Conclusion
U.S. vs. U.K.	Cholelithiasis[4]	25	0.15	73%	U.S. more productive [3]
U.S vs. U.K.	Lung cancer[5]	15	0.31	82%	U.S. more productive [3]
Germany vs. U.K.	Cholelithiasis[4]	56	0.18	14%	Germany more productive [3]
Germany vs. U.K.	Lung cancer[5]	28	0.24	33%	Germany more productive [3]
U.S. vs. Germany	Cholelithiasis[4]	−32	−0.03	52%	U.S. more productive [6]

1 Difference between units in first country to those in second country
2 Average productivity advantage of first country over second; calculated by comparing simple ratios of absolute outcomes over absolute inputs for each country
3 Assuming treatment process does not show increasing returns with additional inputs
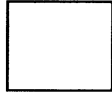4 Assuming symptoms every 30 days; outcome measure in QALYs
5 Survival without treatment estimated to be 3.8 months
6 Based on detailed knowledge of treatment process

Exhibit 10

RELATIVE PRODUCTIVE EFFICIENCY BY DISEASE

Indeterminate comparisons

Countries compared	Disease	Difference in inputs Standardized input units	Difference in outcomes* LYs	Average productivity advantage	Conclusion
U.K. vs. U.S.	Breast cancer	−34	−0.13	Can range from U.S. advantage of 35% to U.K. advantage of 13%**	• Productive efficiency indeterminate

Cost-effectiveness calculation

U.S. additional expenditure per LY
Dollars

U.S. prices	U.K. prices	
$32,000	$13,000	• U.S. had preferred input/outcome combination; U.K. could likely benefit from higher resource level

* Difference between units in first country to those in second country

** Depending on assumption for baseline outcome with no treatment and resulting range of improvement in outcomes provided in Exhibit 6

Source: McKinsey analysis
022 EF 136243/7

Exhibit 11

COMBINED EFFECTS OF COMPETITIVE INTENSITY
AND HEALTH CARE PRODUCT INTEGRATION BY DISEASE
Differences between the U.S. and the U.K.

Disease	Effects of greater health care product integration in U.K.	Effects of higher competitive intensity in U.S.	Impact on relative inputs and outcomes	Impact on relative productive efficiency
Diabetes	• Better triaging of diabetics for different care approaches • Better preventive care and aggressive management, as well as team-based approach, in U.K. diabetic clinics		• U.K. lower inputs and better outcomes	• U.K. more productive
Cholelithiasis	• Fewer surgeries	• More surgeries • More rapid adoption of laparoscopic procedure	• U.S. higher inputs and better outcomes	• U.S. more productive
Breast cancer	• No screening • Adoption of FNA for biopsy; maintenance of one-step biopsy/ surgical treatment protocol	• Broad-based mammographic screening program • Shorter surgical LOS and use of outpatient setting for biopsy	• U.S. higher inputs and somewhat better outcomes	• Indeterminate
Lung cancer	• Less surgery • More use of supportive care only	• Possibly greater use of CT scans for better staging and less restricted triaging process • Shorter surgical LOS and use of outpatient setting for chemotherapy	• U.S. higher inputs and better outcomes	• U.S. more productive

022 EF 136225/7

Chapter 8: Relationship of disease case study results to aggregate-level analyses

As discussed in Chapter 1, much of the prior work comparing health care system performance across countries has focused on aggregate-level (macro) analyses. These analyses showed that the U.S. spent considerably more than Germany and the U.K., and that Germany spent more than the U.K. (**Exhibit 1, left**), while all three countries achieved similar health outcomes as measured by life expectancy. These analyses have therefore produced many new questions: *What are the sources of the differences in spending among countries? Why do those differences appear to be unrelated to differences in overall life expectancy?*

In turn, our analyses of specific disease treatment processes in the U.S., the U.K., and Germany (Chapters 3 through 6) provide a detailed, micro-level view of health care productive efficiency differences and of the factors causing these differences. These analyses showed that Germany used considerably more inputs than the U.K. and the U.S., and that the U.S. used more inputs than the U.K. (**Exhibit 1, right**). Because our analysis has so far focused on only one component of health care spending and on only a small sample of diseases, our micro-level findings can neither resolve this apparent contradiction nor provide comprehensive or definitive answers to the questions raised on total health care spending.

In this chapter, therefore, we draw the link between our micro-level findings and the macro-level statistics on life expectancy and health care spending. We also expand the analysis of aggregate spending to determine the sources of health care spending differences across countries. These sources of differences explain a significant part, but not all, of the contradiction between aggregate spending and input usage.

HEALTH CARE PERFORMANCE ON THE AGGREGATE AND THE DISEASE CASE LEVEL

In the macro-level analyses referred to above, the performance of each country's health care system was measured by comparing life expectancies (measure of outcomes) to levels of health care spending (measure of inputs); while life expectancies were quite similar across countries, expenditures were vastly different. Our micro-level analysis, on the other hand, assesses relative productive efficiency in terms of (quality adjusted) life years (measure of outcomes) per

quantity of input usage at the disease case level. Macro- and micro-level analyses therefore differ both in terms of input and outcome measures.

While the correspondence between outcome measures is relatively close, the input measures are significantly different. Macro-level analyses have included total health care spending, while our analysis has focused on one component of spending only – namely the quantity of inputs. The relationship between these two factors, total health care spending and input usage, is depicted in **Exhibit 2** where health care spending is divided into two main components: medical spending and administrative costs. Medical spending, in turn, is the product of the quantity of inputs used – the measure in our micro-level analysis – and the price paid for those inputs. As will be discussed later, the three spending components – input quantities, input prices, and administrative costs – may in fact be interdependent. For example, administrative spending on "care management," including utilization review and quality assurance, is likely to increase productive efficiency and to reduce input quantities.

We draw the link between our micro-level findings and the macro-level statistics in two steps. In the first step, we translate our relative productive efficiency results to the macro level by reconciling our disease-specific input quantity and survival outcomes with aggregate figures on input quantity and life expectancy. We find a close correspondence. With comparable outcomes, Germany uses the most inputs, the U.S. fewer, and the U.K. the fewest. This finding has two implications. First, our relative productive efficiency results in general are roughly consistent with the macro-level results. Second, differences in health care spending in Germany, the U.K., and the U.S. are *not* caused by differences in relative productive efficiency, because the U.S. spends more (not less) than Germany despite being more productive. This finding therefore challenges the hypothesis that the cause of the U.S.'s poorer performance in managing health care spending is lower productive efficiency.

In our second step, we turn to the other components of aggregate health care spending. Our analysis of these factors suggests that higher spending in the U.S. relative to Germany is due to higher input price levels and higher administrative costs, which are only partially offset by lower input usage. Compared with the U.K., all three factors – higher input prices, higher input quantities, and higher administrative costs – contribute to higher spending in the U.S. Higher spending in Germany relative to the U.K. is due to its higher input use as well as higher input price levels.

STEP 1: ASSESSING PRODUCTIVE EFFICIENCY AT THE AGGREGATE AND THE DISEASE CASE LEVEL

To show that our disease-level productive efficiency results are roughly consistent with the aggregate-level results, we first look at inputs, then at

outcomes, and finally combine input and outcome measures to assess aggregate productive efficiency.

Inputs at the aggregate and the disease case level

While data limitations precluded direct study of input usage on a national level, proxies for the most important components exist.[1] 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 (**Exhibit 3**).[2] Germany used more of each of these inputs per capita than the U.S. The U.S. in turn used more than the U.K., with the exception of slightly higher consumption of drug prescriptions in the U.K. As described in the cross-disease synthesis (Chapter 7), we found this same pattern of relative input usage in each disease case – Germany highest, followed by the U.S., with the U.K. lowest. With consistent directional results from both disease-level and aggregate-level analyses, it seems clear that Germany used more medical inputs than the U.S. despite its lower spending level. 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 (**Exhibit 4**).

However, our case results show considerably higher input use in Germany than the level suggested by aggregate data. This could possibly be explained by higher disease incidence in the U.S., the inpatient focus of the sample of diseases studied, or data limitations. Each of these potential explanations is described below.

1. Higher disease incidence in the U.S. 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) (**Exhibit 5**). We did not address or compare incidences of various diseases in our disease-level analyses. Because we compared inputs on a per-case basis, we effectively normalized for any differences in incidence rates across countries. In the aggregate, however, input per capita is affected by differences in disease incidence among countries due to genetic differences in populations as well as to socioeconomic and environmental factors. For example, the incidence of lung cancer is driven largely by the frequency of smoking. As another example, incidence of Type II diabetes is affected by racial mix and obesity. Disease incidence is also generally a function of age distribution.

As **Exhibit 6** illustrates, incidence rates for breast and lung cancer are considerably lower in Germany. Thus, higher input usage per case relative to

---

[1]   Data on hospital supply and capital usage were not available.

[2]   Hospital personnel includes medical technicians and nurses (qualified and nonqualified).

the U.S. is slightly offset by the higher number of cases of lung and breast cancer in the U.S. In addition, as the two cancers have higher input usage per case than diabetes and cholelithiasis, this effect is magnified. Thus, different incidence rates can explain part of the inconsistency in magnitude between aggregate and disease-level input usage.

2. Inpatient focus of the diseases studied. Our disease cases addressed only a small portion of total health care spending; the four diseases studied covered only about 5 percent of total inpatient costs in the U.S. (**Exhibit 7**). For Germany, the sample was even smaller since we did not study diabetes. With such a small sample, it is possible that our disease cases may not entirely reflect average disease treatment in each country. In particular, 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 U.S. 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.

3. Data limitations. Data limitations on the aggregate level prevented us from studying all types of input usage. In particular, data did not include sufficient information on capital costs. Germany's supply of hospital capacity per capita far exceeded the U.S. supply, despite higher occupancy levels in Germany (**Exhibit 8**). However, the U.S. used more of some expensive technologies, such as computerized tomography (CT) scanners and magnetic resonance imaging (MRI) scanners (**Exhibit 9**). Capital inputs per bed might therefore have been higher in the U.S., offsetting Germany's greater bed capacity levels to some extent. However, because nonlabor inputs (capital and supplies) represented only about one-third of the total cost of inpatient care in each country and capital represented less than 10 percent (**Exhibit 10**), higher capital usage could therefore have had, at most, a small effect on total health care input use.

Outcomes at the aggregate and disease case level

While at the aggregate level, life expectancy outcomes showed no significant difference among the U.S., Germany, and the U.K., significant variations in treatment outcomes were observable at the disease case level.[3] In the disease

---

[3] Life expectancy in the U.S. is similar to that of Germany and the U.K. when the effects of infant mortality are removed. This adjusted measure of life expectancy is the most appropriate relative outcome measure because the inclusion of infant mortality effects may bias the results. Definitions of health status at birth vary widely among the U.S., Germany, and the U.K. For example, infants that may be considered stillborn (and thus do not contribute to infant mortality) in Germany and the U.K. may be included in the U.S. statistics, resulting in higher infant mortality and lower life expectancy at birth in the U.S.

case studies, outcomes were generally best in the U.S. and worst in the U.K. There are a few exceptions: in diabetes, the U.S. exhibited worse outcomes than the U.K., while in breast cancer Germany's outcomes were worse than those observed in the U.K.[4]

Since we studied only four diseases, our findings do not allow us to state that the U.S. had better treatment outcomes overall (measured across all diseases) relative to Germany and the U.K. However, in order for the U.S. to have *worse* treatment outcomes overall compared with the U.K., the majority of diseases would have to be sufficiently similar to diabetes and sufficiently different from lung cancer, breast cancer, and cholelithiasis to reverse the pattern of outcomes observed at the disease case level. Therefore, it seems unlikely that the U.S. had systematically worse treatment outcomes at the disease case level, especially given our causal analysis, which suggests strongly that the incentives of the U.S. system generally encouraged better outcomes. More likely, the U.S. had generally better treatment outcomes at the disease case level than the U.K. When comparing Germany and the U.K., however, it is unclear which outcomes were better, since German outcomes were better for cholelithiasis and lung cancer but worse for breast cancer.

There are several reasons that might explain why potentially differing outcomes at the disease case level did not translate into differing life expectancy at the aggregate level. Most importantly, life expectancy is not only a function of disease treatment, but also of overall disease incidence, availability of care, and several other factors that are difficult to quantify, including socioeconomic variables and lifestyle choices (**Exhibit 11**). Differences in any of these factors could easily offset any outcome advantages in disease treatment. In addition, life expectancy rates reflect health outcomes for the entire nation across the full spectrum of diseases. Finally, in two of our diseases – diabetes and cholelithiasis – our outcome measure, quality adjusted life years (QALYs), has little direct relationship to life expectancy.

Relative productive efficiency

We now combine our analysis of aggregate inputs and outputs to assess aggregate productive efficiency. Aggregate input usage in Germany was highest, followed by the U.S., and then the U.K. Given similar life expectancies, we conclude that the U.S. had higher overall productive efficiency relative to Germany at the aggregate level, and lower overall productive efficiency relative to the U.K.

These perspectives on overall productive efficiency at the aggregate level are partially consistent with the disease case results. We concluded that the U.S. was more productive than Germany across all diseases, more productive than the U.K. in cholelithiasis and lung cancer, but less productive than the U.K. in

---

[4] Germany was not included in the disease case comparison of diabetes.

diabetes. (Results for breast cancer were indeterminate when comparing the U.S. with the U.K.) Thus, although we cannot state that the U.S. was more productive across all diseases (as only four diseases were studied), our results do suggest that the U.S. was more productive overall relative to Germany. The disease cases did not allow us to conclude whether the U.S. or the U.K. was more productive.

Germany, in turn, was more productive than the U.K. in two of three diseases studied, while the aggregate analysis suggests it was less productive overall. Because the aggregate analysis uses only a very coarse outcome measure (life expectancy), and because we have conflicting evidence from only three cases, we cannot draw a conclusion.

<center>* * *</center>

Thus, the source of the U.S.'s higher spending and its apparently similar aggregate outcomes relative to the U.K. and Germany does not appear to be lower productive efficiency of the health care system. Similarly, Germany's higher spending relative to the U.K. and apparently similar aggregate outcomes do not appear to be explained by lower productive efficiency at the disease case level. We therefore turn to the other components of health care spending.

STEP 2: EXPLAINING DIFFERENT AGGREGATE SPENDING LEVELS BY EXAMINING OTHER COMPONENTS

In addition to the quantity of medical inputs, prices of medical inputs and administrative costs are the two other important components of total health care spending (see again **Exhibit 2**). We analyzed these two components in three major areas of health care spending: hospitals, physicians, and pharmaceuticals. These three spending areas account for about 70 percent of total health care spending in each of the three countries (**Exhibit 12**). In terms of the differences in health care spending, the three areas account for 73 percent of the spending difference between the U.S. and the U.K., 77 percent of the difference between the U.S. and Germany, and 62 percent of the difference between Germany and the U.K. (**Exhibit 13**). Other spending areas, such as nursing home and dental care, were not considered because they represented a relatively small portion of total health care spending and were not directly related to our study of disease case level productive efficiency.

Relative input prices

The prices of many medical inputs appear to have been higher in the U.S. than in either Germany or the U.K. **Exhibit 14** shows average input prices in the three countries for physicians, nurses, and prescriptions.[5] The most striking differences in input prices 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 U.K. The U.S.'s higher physician incomes reflect both a higher wage premium for physicians in the U.S. relative to other professional workers and somewhat higher average wages in the U.S. (**Exhibits 15 and 16**).

---

[5] These prices are converted to U.S. dollars at GDP PPP 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. A more detailed study of input prices was not possible due to data limitations. In particular, data on hospital supply prices were not available.

This pattern of higher input prices in the U.S. is not surprising given the structure of the three health care systems. Both Germany and the U.K. were more centrally administered systems compared with the U.S. Their governments and agencies may therefore have acted more like monopsony buyers of medical services and used their market power to drive down prices.[6] While the U.S. had some elements of monopsony purchasing (mostly through Medicare), many input prices in the U.S. were set in markets without dominant buyers but with strong sellers, particularly physicians. We can therefore hypothesize that 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, it is possible that differences in relative provider skill or experience levels contributed to observed price differences, which in turn could have contributed to different productive efficiency levels. Furthermore, physician incomes in the U.S. reflect to some extent the significant education costs borne directly by the physician. However, a comprehensive analysis of pricing levels, their causes, and their potential effect on productive efficiency was outside the scope of our study.[7]

Recent changes in the U.S., however, suggest that the balance of power between payors and providers may be shifting and driving down some U.S. input prices. For example, physician salaries have recently begun to decline, at least for surgeons and specialists, particularly in markets with high concentrations of managed care (**Exhibits 17 and 18**).

Relative administrative spending

A frequent criticism of the U.S. health care system is that it spends an inordinate amount on administrative costs relative to other countries. This hypothesis is difficult to test, because of both definitional questions about the nature of administrative costs and problems with finding comparable data on administrative spending across countries.

Administrative spending includes four distinct, but difficult to disaggregate, cost categories:

1. Payor, provider, and government agency costs for administration of the insurance and provider reimbursement system

2. Provider costs associated with management and administration of their health care facilities and practices

---

[6] A monopsony has the ability to reduce input prices since it faces little or no competition from other buyers in its input markets.

[7] For example, one interesting question is whether "artificially" low prices in the U.K. and Germany might have some long-term effect on the supply or quality of medical inputs, and thus on effectiveness of disease treatment, or may have in fact increased demand (i.e., supply-induced demand). These issues were outside the scope of this study.

3. Payor costs for sales and marketing of health coverage products to purchasers and members

4. Payor and provider costs for care management, including utilization review and quality assurance.

None of these cost categories can be considered purely administrative, as they may contribute directly or indirectly to the productive efficiency of the health care system. This observation is particularly true with the last category of administrative cost, as care management is designed to increase productive efficiency in the form of higher quality or lower costs of health care. To a lesser extent, this also holds true for the first three categories. For example, basic provider management functions, such as scheduling and ordering supplies, result in overhead costs that may reduce input costs and quantities. The existence of separate entities for administering health insurance (i.e., payors and government agencies) and providing health care services results in additional administrative costs for all parties involved to handle the reimbursement process, but may exert pressure that reduces supply-induced demand. In addition, for payors to remain competitive in a market-driven health care system that appears to promote higher productive efficiency, additional costs must be incurred in the form of sales and marketing expenditures.

Because these four administrative cost categories are associated with different health care system participants across countries, and because sufficient data on each category are not readily available, international comparisons of administrative spending are problematic at best. Nevertheless, the available literature suggests that administrative costs in the U.S. were indeed higher than those in Germany and the U.K.[8]

The best data available are for administrative costs associated with payors and hospitals. Payor-associated administrative costs were considerably higher in the U.S. than in Germany (**Exhibit 19**). While no directly comparable figure exists for the U.K. because of its nationalized health insurance, we used the administrative cost of the National Health Service (NHS) and the total operating costs of the Department of Health relative to the total NHS budget as an estimate for the share of insurance overhead in total spending.[9] This share was much smaller

---

[8] Sources: Woolhandler and Himmelstein, 1991; Department of Health, 1995; Medicare cost data, 1990; Federal Ministry of Health; NHS, 1995; German Hospital Society; Rohn-Klinikum annual report, 1990. Health insurers' administrative costs were reported for the U.S. in 1987, the U.K. in 1990, and Germany in 1990. Hospital administrative costs were determined for the U.S. in 1990, the U.K. in 1993, and Germany in 1990.

[9] The "payor" costs of the U.K. system include the expenditures to oversee and administer health care in the U.K. These costs are borne by the Department of Health and the NHS. For the NHS part of payor costs, no disaggregated information is available. We therefore included the total cost of the NHS Department of Administration. This figure overestimates the NHS expenditures to administer the health care system but leaves out the NHS expenditures to oversee the system.

than either the U.S.'s or Germany's. For hospitals, administrative costs were also considerably higher in the U.S. than in the U.K. and Germany (**Exhibit 20**).

No comprehensive data exist comparing all administrative costs across the U.S., the U.K., and Germany. However, one study compared total administrative spending between the U.S. and Canada (**Exhibit 21**).[10] The U.S.-Canada study is particularly relevant to our U.S.-U.K. comparison since Canada has a single payor system similar in some respects to the U.K.'s. This study showed that total administrative spending in the U.S. in 1987 was about twice the Canadian level in percentage terms, accounting for as much as 22 percent of health care spending and, therefore, was almost three times higher in dollars per capita.[11] Since 1987, the share of administrative costs in the U.S. has evidently risen, so this estimate is likely to be understated.[12] Thus, this comparison suggests a pattern of higher administrative costs in the U.S.

We combined this information on payors' and hospitals' administrative costs with the U.S.-Canada comparison study to estimate total administrative spending. Recalling that the U.S. administrative costs were 22 percent in 1987 and rising, we extrapolated that they were about 24 percent in 1990. We also obtained an estimate by summing the percentages for health insurers' and hospital administrative costs provided in **Exhibits 19 and 20**. We arrive again at 24 percent in the U.S., at 13 percent in Germany, and at 16 percent in the U.K.[13]

Several factors may have contributed to the U.S.'s higher administrative costs. 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. In addition, the U.S.'s higher administrative costs 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 U.S.'s higher productive efficiency observed in the disease cases.

---

[10] Source: Woolhandler and Himmelstein, 1991. This study compared administrative overhead for hospitals, physician practices, nursing homes, and insurers.

[11] We have adjusted Woolhandler and Himmelstein's estimates slightly, using a hospital administrative overhead percentage from 1990 Medicare cost reports, and an insurance administrative overhead percentage from a McKinsey analysis of 24 insurers.

[12] Source: Woolhandler and Himmelstein, 1991. No strictly comparable and comprehensive data are available for more recent years. The ratio of administrative staff to medical staff rose steadily between 1983 and 1990. In addition, insurance overhead, hospital administrative costs, and physician overhead rose steadily between 1983 and 1987. We extrapolate this trend to 1990.

[13] These figures may be a slight overestimate since the hospital administrative cost percentage appears to be slightly greater than the percentage for all health care services.

**Net effect of relative input levels,
input prices, and administrative costs**

By examining relative input quantities, input prices, and administrative costs, we can estimate the impact of each factor in explaining differences in total spending per capita among the three countries. Using input levels, price levels, and administrative costs of the country with higher expenditures as a base, we estimated each factor's relative contribution to the per capita spending differences among countries (**Exhibit 22**).[14] In comparing the U.S. with Germany, Germany's lower prices and administrative costs relative to the U.S. appeared to more than offset its higher input quantities in 1990, leading to lower overall spending in Germany. All three factors appeared to be low in the U.K. relative to both the U.S. and Germany in 1990, leading to even lower per capita spending. In all comparisons, the unexplained residual (which represented differences in spending that were unaccounted for by differences in input quantities, input prices, and administrative costs) was relatively small.

As discussed earlier, physicians, pharmaceuticals, and hospitals were responsible for a large portion of the total health care spending gap among countries. **Exhibit 23** estimates the relative magnitude of price and volume differences in each of these three categories, as well as administrative costs, in explaining the overall per capita spending gap among countries in 1990. When comparing the U.S. with the U.K. and Germany, higher prices for physicians and inpatient acute care and higher administrative costs appeared to be the most important factors contributing to higher spending in the U.S. When comparing Germany with the U.K., the higher use of inpatient services appeared to be the most important factor contributing to Germany's higher spending.

* * *

In summary, the combination of our disease-level productive efficiency analyses with aggregate-level analyses suggests that higher overall spending levels in the U.S. relative to the U.K. and Germany, and in Germany relative to the U.K., were not due to lower productive efficiency. Higher aggregate spending in the U.S. resulted from higher relative input prices and administrative costs rather than from lower productive efficiency.

Although we can observe different levels of medical input prices and administrative costs associated with different system structures, assessing the causes of higher administrative costs and input prices was not the focus of this

---

[14] The relative input levels were determined using information on the amount of labor (physicians, nurses, and medical technicians), supplies (pharmaceuticals), and capital consumed in each country, assuming a breakdown of 70 percent labor, 20 percent supplies, and 10 percent capital. Relative prices were determined using a price index reported by the OECD, while administrative costs were estimated as 24 percent in the U.S., 13 percent in Germany, and 16 percent in the U.K.
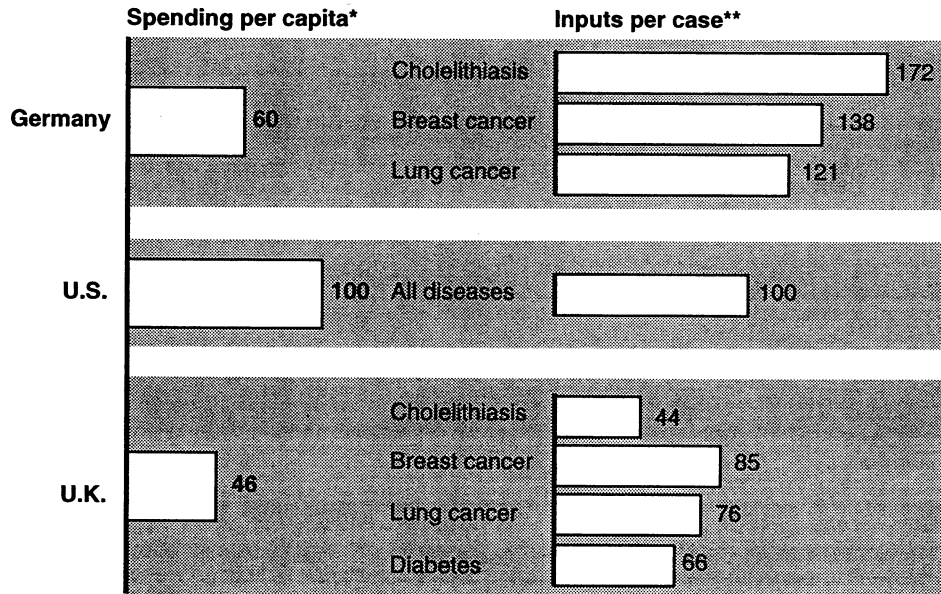
effort. There are, however, many potential explanations. As discussed, it is possible that the U.S.'s higher administrative spending could have resulted from the relative fragmentation of providers and payors in the U.S. markets, from the complexity of the U.S. insurance and reimbursement system, or from higher levels of care management. Higher input prices could have resulted from the market position of the U.S. physicians and other providers in some markets and medical specialties, or could reflect potentially higher skill levels on the part of the U.S. physicians, which could have contributed to the U.S.'s higher productive efficiency. From available data, we cannot determine which of these factors, if any, played a role, nor can we determine whether a truly competitive market could achieve the lower input price and administrative cost levels of the U.K. and German systems; to do so would require further study.

These findings raise important questions for policymakers and health care organizations in each of the countries studied regarding how to improve overall system performance. Although productive efficiency, relative input prices, and administrative cost levels are not completely independent factors, the most appropriate and effective actions for improving each are likely to be quite different. In Chapter 9, we comment on potential implications of these findings for policymakers and health care organizations, focusing on actions that can improve productive efficiency.

Exhibit 1

SPENDING DIFFERENCES VS. INPUT DIFFERENCES

Index: U.S. = 100%

Spending per capita*   **Inputs per case****

	Spending per capita	Disease	Inputs per case
Germany	60	Cholelithiasis	172
		Breast cancer	138
		Lung cancer	121
U.S.	100	All diseases	100
U.K.	46	Cholelithiasis	44
		Breast cancer	85
		Lung cancer	76
		Diabetes	66

* 1990 spending per capita converted to U.S. dollars using 1990 gross domestic product (GDP) purchasing power
 parity (PPP) and then indexed with U.S. spending = 100
** Indexed to standardized input units

Source: BASYS; McKinsey analysis

Exhibit 2

RELATIONSHIP BETWEEN AGGREGATE
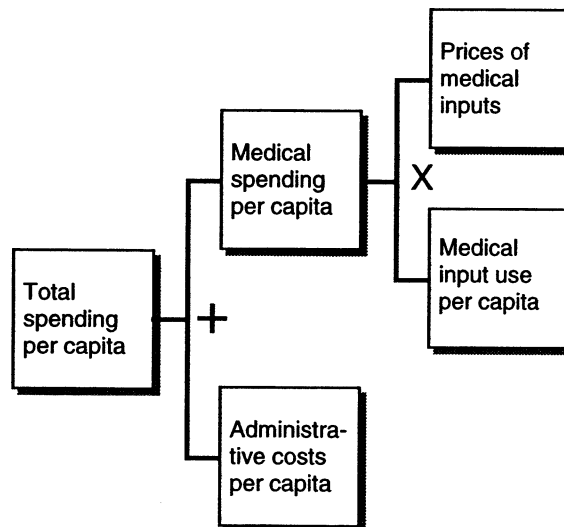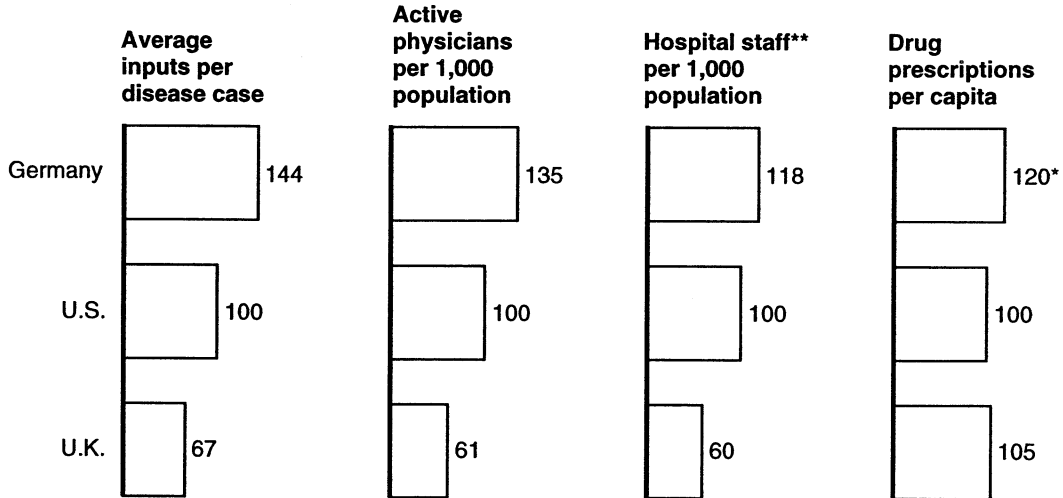## SPENDING AND DISEASE-LEVEL INPUTS

```
                                            ┌─────────────┐
                                            │ Prices of   │
                                            │ medical     │
                                            │ inputs      │
                          ┌──────────┐      └─────────────┘
                          │ Medical  │  X
                          │ spending │
                          │ per capita│     ┌─────────────┐
          ┌──────────┐    └──────────┘      │ Medical     │
          │ Total    │  +                   │ input use   │
          │ spending │                      │ per capita  │
          │ per capita│                     └─────────────┘
          └──────────┘    ┌──────────┐
                          │ Administra-│
                          │ tive costs │
                          │ per capita │
                          └──────────┘
```

Exhibit 3
MEDICAL INPUTS BY COUNTRY
1990
Index: U.S. = 100%

	Average inputs per disease case	Active physicians per 1,000 population	Hospital staff** per 1,000 population	Drug prescriptions per capita
Germany	144	135	118	120*
U.S.	100	100	100	100
U.K.	67	61	60	105

* Excluding over-the-counter drugs for which prescriptions are written
** Hospital staff includes nurses (qualified and nonqualified) and medical technicians for 1989. U.K. data is for England only

Source: OECD; German Health Care System; National Health Service (NHS); Bureau of Labor Statistics; HPSS-U.K.; American Hospital Association (AHA); Statistisches Bundesamt; McKinsey analysis

---

Exhibit 4
AGGREGATE INPUT DIFFERENCES VS. DISEASE CASE INPUT DIFFERENCES
Index: U.S. = 100%

Aggregate input usage*

Germany	116
	+16%
U.S.	100
	−25%
U.K.	75

Case study input usage

Cholelithiasis	172
Lung cancer	121
Breast cancer	138
	+39%**
All diseases	100
	−23%**
Diabetes	66
Cholelithiasis	44
Lung cancer	76
Breast cancer	85
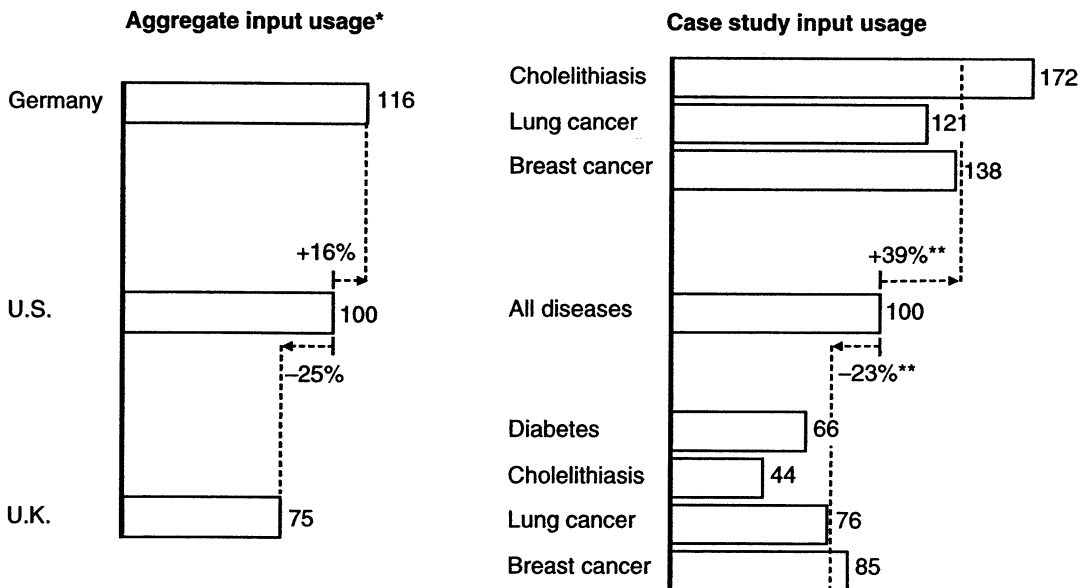
* Weighted average of inputs for labor (70%), prescriptions (20%), and capital (10%) for each of 3 countries
** Weighted average of input usage across the diseases
022 P 134529/8

Exhibit 5

RELATIONSHIP BETWEEN AGGREGATE
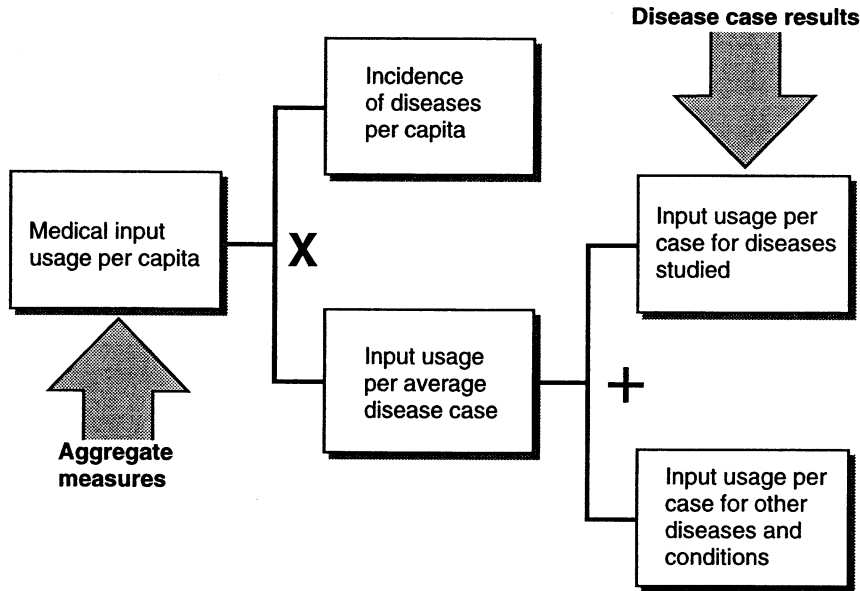## AND DISEASE CASE INPUTS

Disease case results

```
                    ┌─────────────┐
                    │ Incidence   │
                    │ of diseases │
                    │ per capita  │
                    └─────────────┘
┌─────────────┐                        ┌──────────────────┐
│ Medical input│    X                  │ Input usage per  │
│ usage per    │                       │ case for diseases│
│ per capita   │                       │ studied          │
└─────────────┘                        └──────────────────┘
                    ┌─────────────┐
   Aggregate        │ Input usage │         +
   measures         │ per average │
                    │ disease case│        ┌──────────────────┐
                    └─────────────┘        │ Input usage per  │
                                           │ case for other   │
                                           │ diseases and     │
                                           │ conditions       │
                                           └──────────────────┘
```

Exhibit 6

DISEASE INCIDENCE

1980

Diabetes[1] New cases per 100,000		**Cholelithiasis**[2] Cases per 1,000	
U.K.	100	U.K.	110
U.S.	236	U.S.	110
Germany	Not studied	Germany	110
Breast cancer[3] New cases per 100,000		**Lung cancer**[4] New cases per 100,000	
U.K.	76	U.K.	82
U.S.	87	U.S.	82
Germany	56	Germany	60

1 1993 prevalence of diabetes in the U.S. (3.03%) was also higher than that in the U.K. (2.44%); U.S. data for 1978
2 Prevalence for gallstones is reported as average of several studies conducted during the 1980s; this prevalence ranged from 7-15%
   (average of 11%) in the total population; while the prevalence depended on age and gender, it did not appear to vary significantly from
   one country to another
3 Age-adjusted incidence for white females
4 Age-adjusted incidence for white males; incident rates for white females were much higher in the U.S. than in the U.K. or Germany due
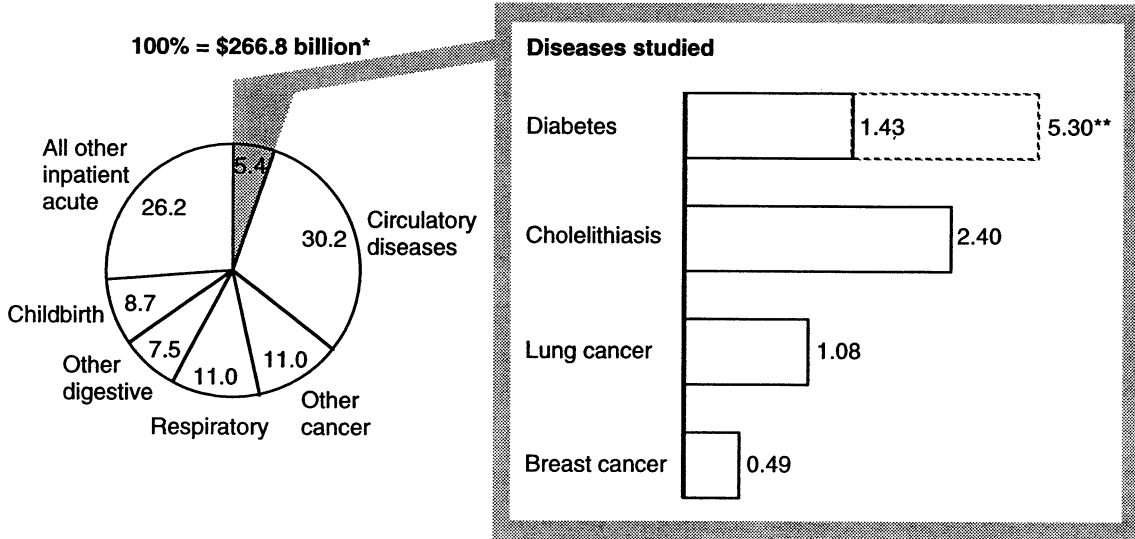   to a historically higher number of smokers among women

Source: National Cancer Institute (NCI); *European Journal of Cancer*, McKinsey analysis

Exhibit 7

INPATIENT SPENDING ADDRESSED BY DISEASE CASES – U.S.
1992
Percent

100% = $266.8 billion*

All other inpatient acute 26.2

5.4

Circulatory diseases 30.2

Childbirth 8.7

Other digestive 7.5

11.0

Respiratory

11.0

Other cancer

Diseases studied

Diabetes	1.43	5.30**
Cholelithiasis	2.40	
Lung cancer	1.08	
Breast cancer	0.49	

* Billed charges *not* including discounts
** Low estimate based on diabetes-coded discharge; high estimate based on total cost of diabetes, including added patient days in other diseases

Source: National Bill for Disease Hospital Cost and Utilization Project; American Diabetes Association

---

Exhibit 8

ACUTE CARE BEDS* PER 1,000 POPULATION
1985-92

Germany
U.S.
U.K.

1985 1986 1987 1988 1989 1990 1991 1992
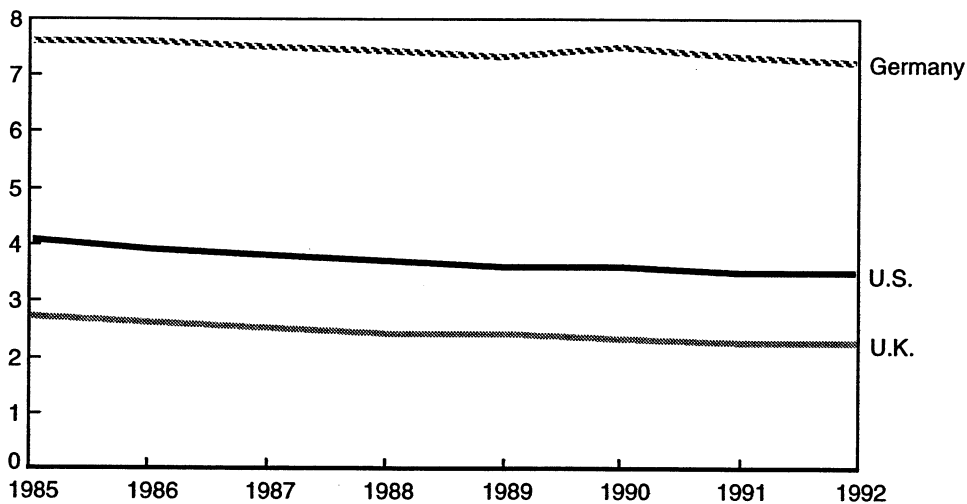
* Acute care hospitals or wards are defined by the OECD as those where length of stay (LOS) is less than 30 days for major medical (e.g., not psychiatric) services. Beds counted are those "maintained and ready for service." U.S. excludes federal hospitals; U.K. is NHS only; and LOS is less than 30 days for 97% of NHS beds counted

Source: OECD; AHA; NHS; Statistisches Bundesamt

022 ST 140080/8

Exhibit 9
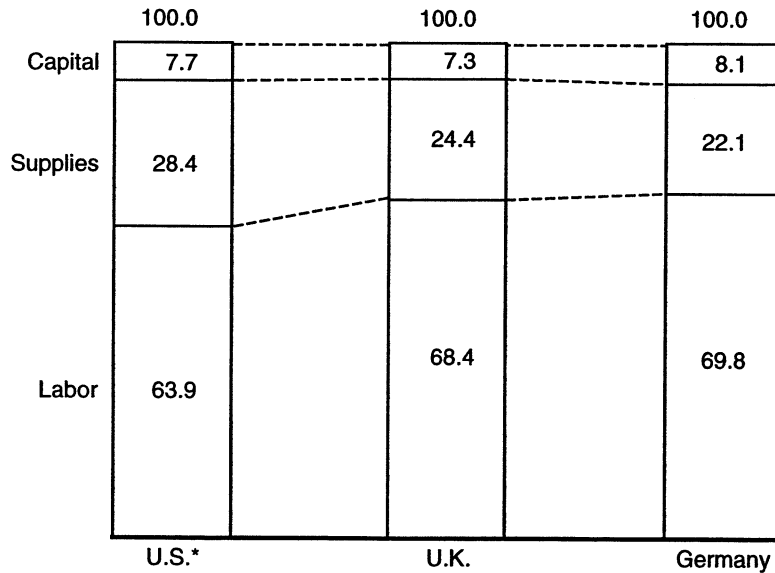
CT AND MRI SCANNERS PER MILLION POPULATION
1990

CT scanners

U.S.	26.8
Germany	12.2
U.K.	4.3

MRI scanners

	8.4
	2.3
	0.9
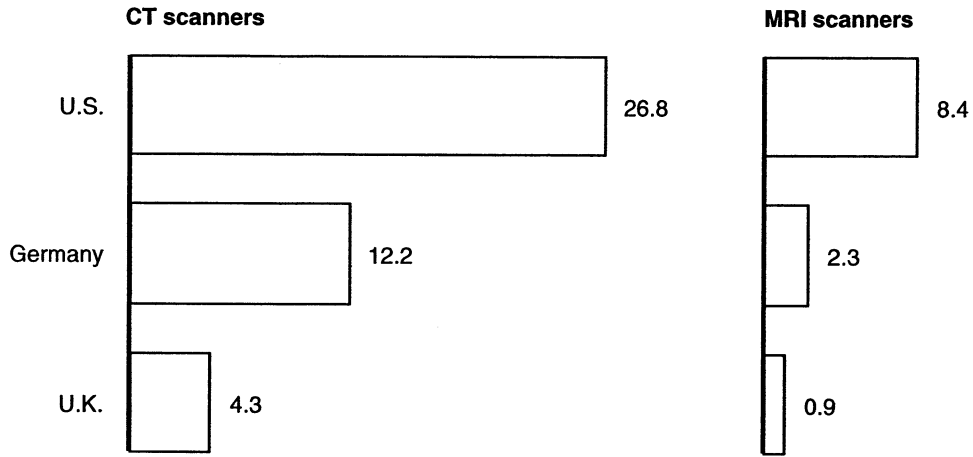
Source: Office of Technology Assessment

Exhibit 10

COST STRUCTURE OF INPATIENT CARE
1989
Percent

	100.0	100.0	100.0
Capital	7.7	7.3	8.1
Supplies	28.4	24.4	22.1
Labor	63.9	68.4	69.8
	U.S.*	U.K.	Germany

* U.S. data for 1990. Labor includes physicians contracted, but not employed, by hospital

Source: AHA; McKinsey; Deutsche Krankenhaus Gesellschaft (DKG)

022 ST 125659/8

Exhibit 11

**RELATIONSHIP BETWEEN LIFE EXPECTANCY
AND DISEASE-LEVEL OUTCOMES**
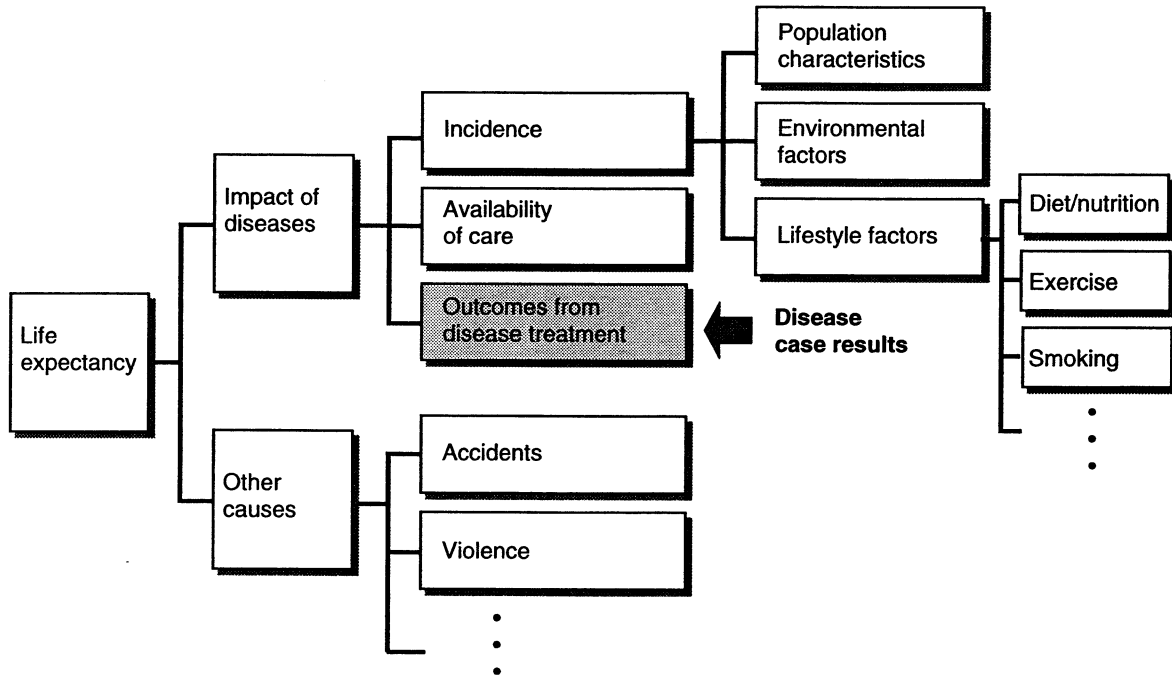
Exhibit 12

HEALTH CARE SPENDING BY CATEGORY

Percent, 1990

100% = $2,439

	U.S.	Germany	U.K.
Other*	29	32	30
Pharmaceuticals	8	16	14
Physicians	22	17	13
Hospitals	41	35	43

$2,439 $1,473 $1,113

Total of hospitals, physicians, pharmaceuticals   71%   68%   70%
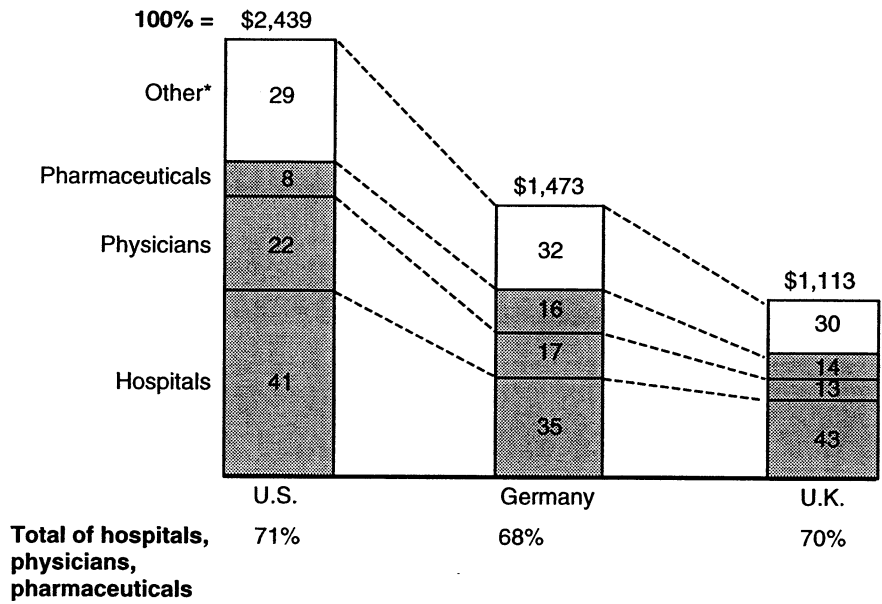
* Includes dentists, nursing home care, medical devices, and other items

Note: Figures reported here differ from those reported by the OECD; BASYS estimates do not include capital costs or administrative costs by benefit insurers; OECD figures for the U.K. may be underestimated

Source: BASYS

022 CL 132259/8

Exhibit 13

SOURCES OF DIFFERENCE IN HEALTH CARE SPENDING BY CATEGORY
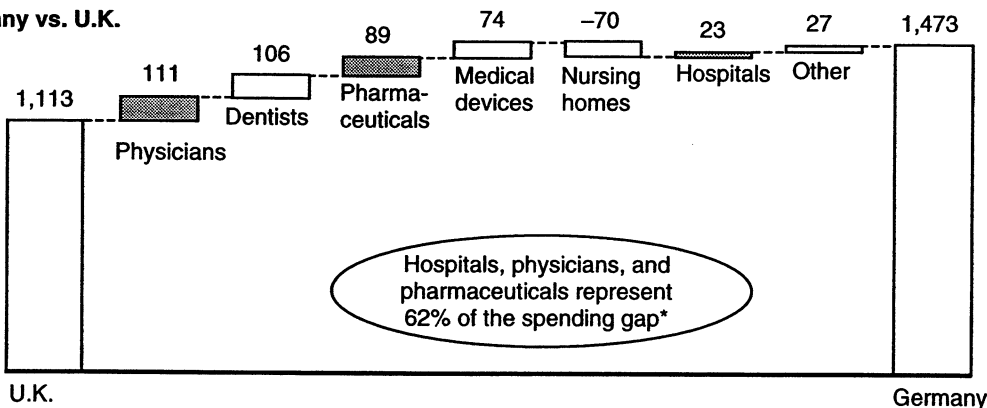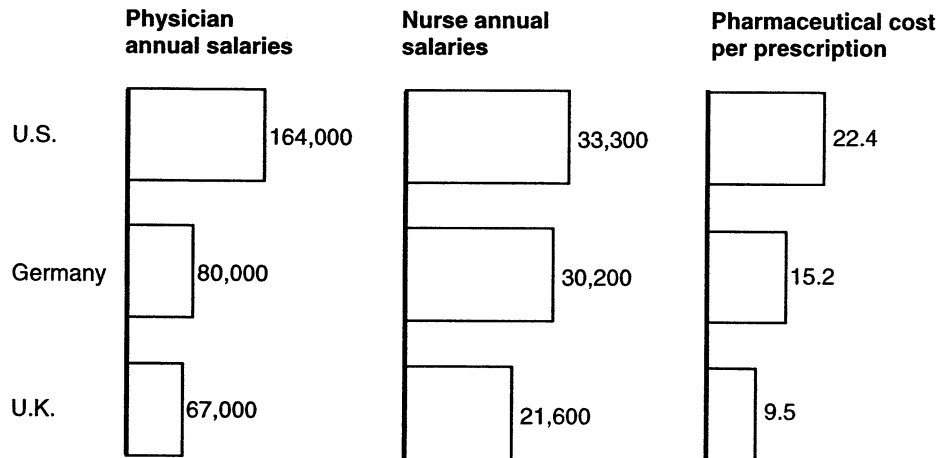
U.S. dollars per capita, 1990, PPP

U.S. vs. Germany

Hospitals, physicians, and pharmaceuticals represent 77% of the spending gap*

Germany — U.S.

U.S. vs. U.K.

Hospitals, physicians, and pharmaceuticals represent 73% of the spending gap*

U.K. — U.S.

Germany vs. U.K.

Hospitals, physicians, and pharmaceuticals represent 62% of the spending gap*
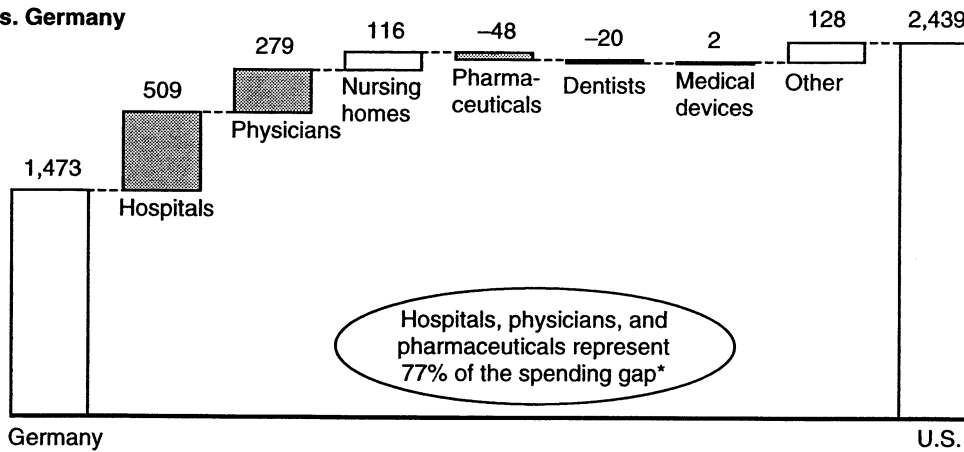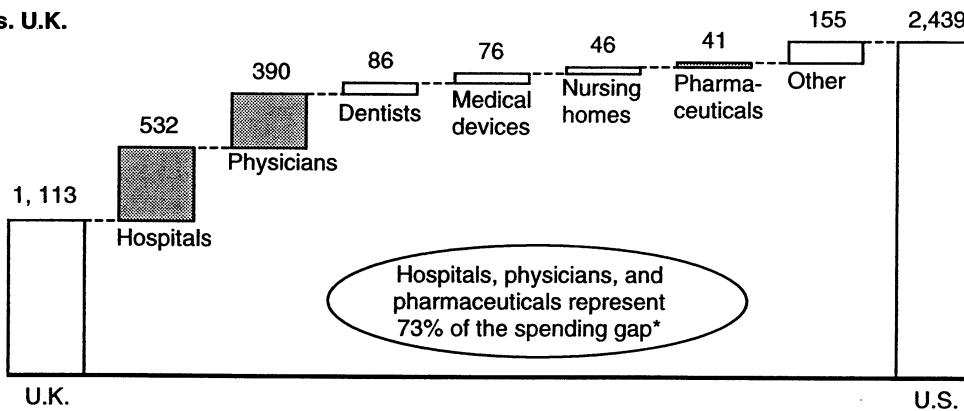
U.K. — Germany

* Spending on hospitals, physicians, and pharmaceuticals was $740 per capita higher in the U.S. than in Germany (accounting for 77% of the total spending difference of $966 per capita), $963 per capita higher in the U.S. than in the U.K. (accounting for 73% of the total spending difference of $1,326 per capita), and $223 per capita higher in Germany than in the U.K. (accounting for 62% of the total spending difference of $360 per capita)

Note: Figures reported here differ from those reported by the OECD; BASYS estimates do not include capital costs or administrative costs by benefit insurers; OECD figures for the U.K. may be underestimated

Source: BASYS

022 EF 136216/8

Exhibit 14

INPUT PRICES ACROSS THE U.S., GERMANY, AND THE U.K.

U.S. dollars, 1990, PPP

	Physician annual salaries	**Nurse annual salaries**	**Pharmaceutical cost per prescription**
U.S.	164,000	33,300	22.4
Germany	80,000	30,200	15.2
U.K.	67,000	21,600	9.5

Exhibit 15

PHYSICIAN INCOMES RELATIVE TO PROFESSIONAL WAGES

U.S.
Germany
U.K.

Professional wage* (1990 U.S. $ at PPP)

54,300

46,700

33,800

Physician incomes (1990 U.S. $ at PPP)

164,000

80,000

67,000

X

Physician income multiple of professional wage

3.0

1.7

2.0
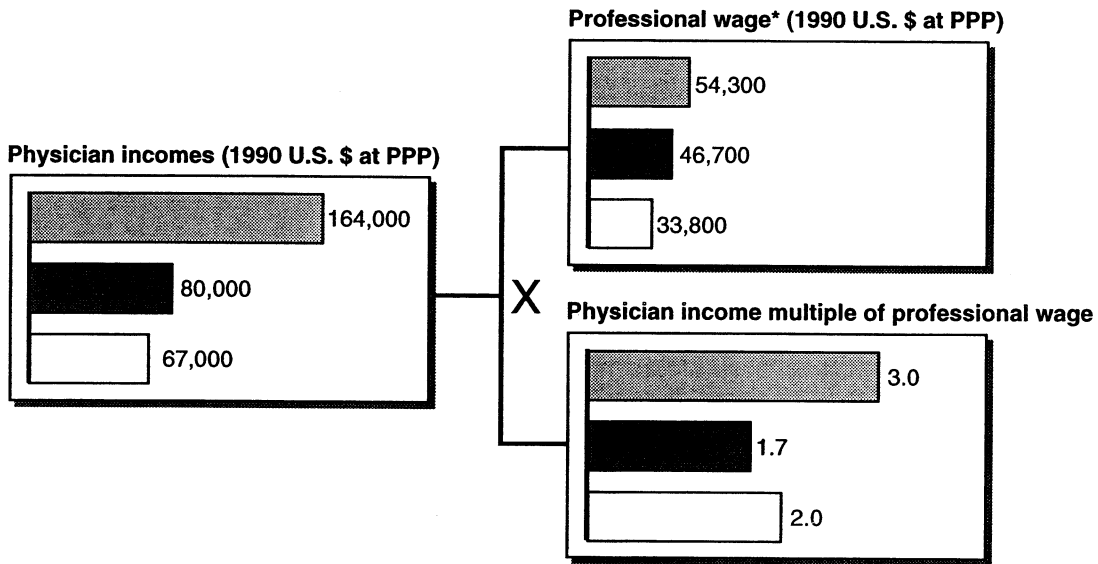
* Average salary of several professions requiring an advanced graduate degree

Source: U.S. Statistical Abstract; IDS; Federal Office of Statistics; McKinsey analysis

---

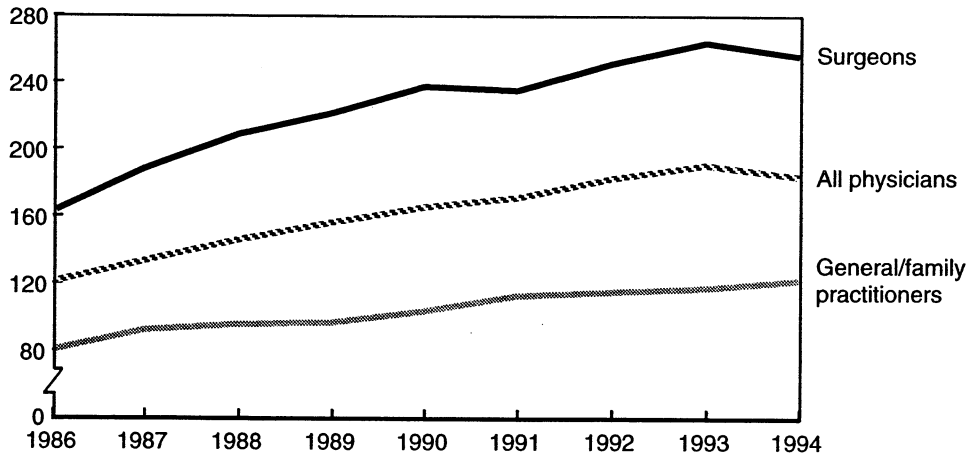Exhibit 16

PHYSICIAN INCOMES RELATIVE TO AVERAGE WAGES

U.S.
Germany
U.K.

Average wage (1990 U.S. $ at PPP)

31,600

24,800

22,400

Physician incomes (1990 U.S. $ at PPP)

164,000

80,000

67,000

X

Physician income multiple of average wage

5.2

3.2

3.0
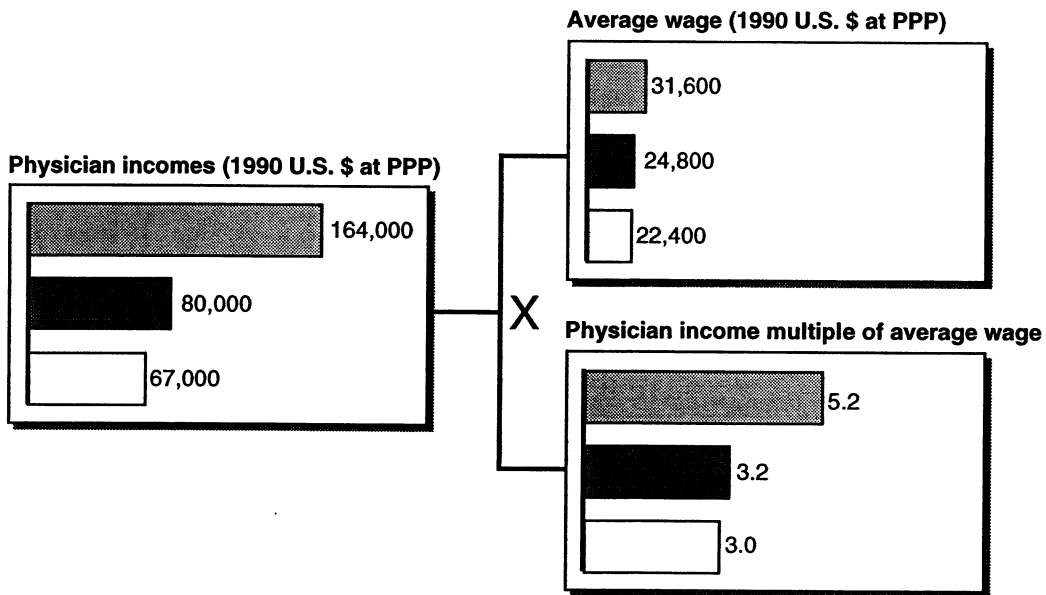
Exhibit 17
PHYSICIAN SALARIES IN THE U.S.
$ Thousands*/year

* Real dollars
Note: Includes active, nonfederal, patient care physicians excluding residents; declines in 1994 salaries occurred for all specialties
(internal medicine, surgery, pediatrics, obstetrics/gynecology, radiology, psychiatry, anesthesiology, and pathology) with the
exception of general/family practice
Source: American Medical Association (AMA)

Exhibit 18
EMPLOYED STAFF PHYSICIAN SALARIES FOR GROUP
## PRACTICES, HOSPITALS, AND HMOs IN THE U.S.
$ Thousands*/year
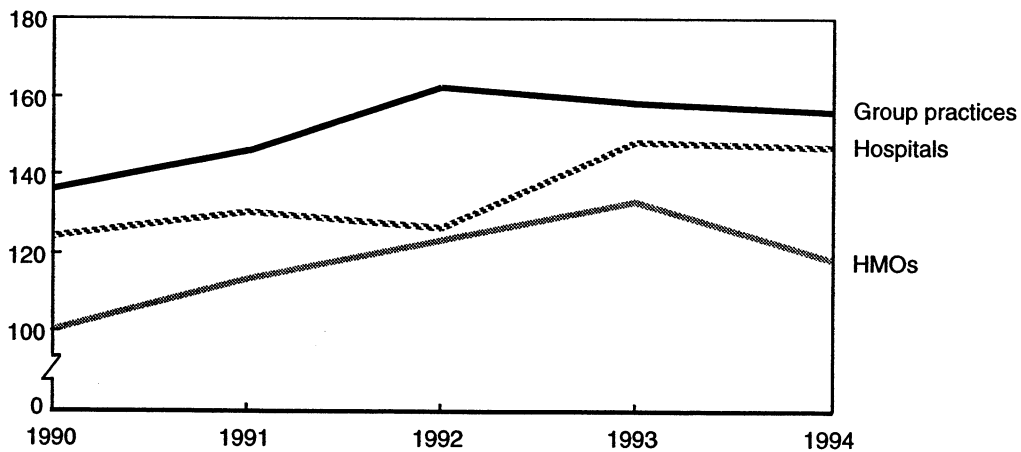
* Real dollars
Note: Based on a survey of 121 health care organizations and 9,733 staff physicians; it is important to note that salary differences may
be due in part to differences in the mix of specialists and primary care physicians; the survey also noted that the salaries of family
practitioners are rising ($114,800 in 1994 compared to $109,400 in 1993) while the salaries of most medical specialties remained
flat. This may have been due in part to the increasing emphasis on primary care physicians due to managed care
Source: William M. Mercer
022 EF 136219/8

Exhibit 19

HEALTH INSURERS' ADMINISTRATIVE COSTS
1990

Percent of revenues

U.S.	5.1[1]
U.K.	2.0[2]
Germany	4.6[3]

**Dollars per capita[4]**

137
19
70
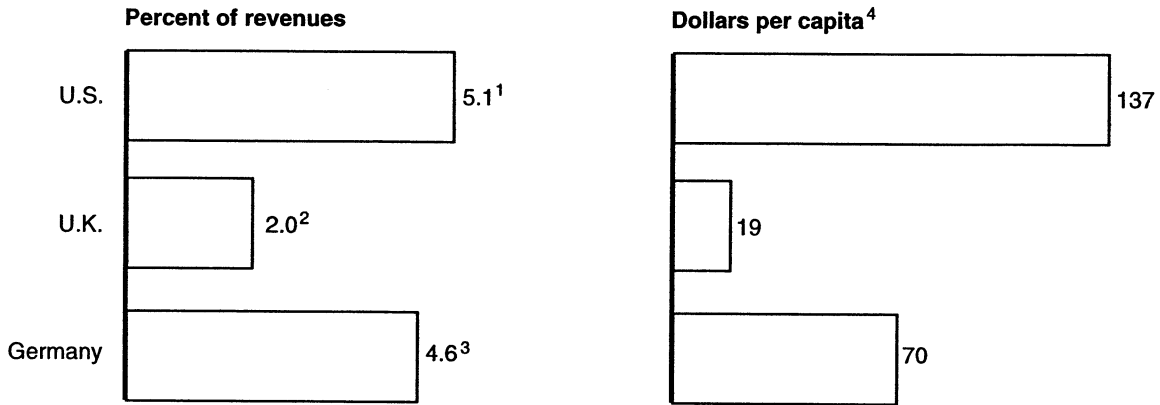
1 Includes health insurers' costs for private insurers and government sponsored health care (Medicare and Medicaid) for 1987
2 The health insurers' administrative costs for the U.K. were calculated by using the ratio of the total operating budget of the NHS Department of Administration and the Department of Health to the total health care expenditure by the NHS
3 Administrative costs for Germany are reported for 1995
4 Calculations based on 1990 health care spending per capita as reported by the OECD (BASYS spending figures do not include administrative costs of insurers)

Source: *The New England Journal of Medicine;* 1995 U.K. Department of Health annual report; Federal Ministry of Health; OECD; McKinsey analysis

---

Exhibit 20

HOSPITAL ADMINISTRATIVE COSTS
1990

Percent of total hospital costs

U.S.	18.4
U.K.*	14.3
Germany**	8.3

Dollars per capita*
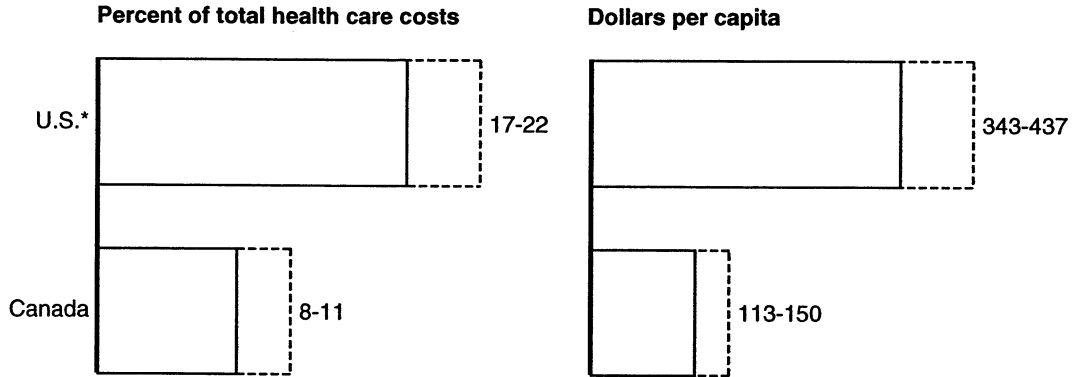
188
70
41
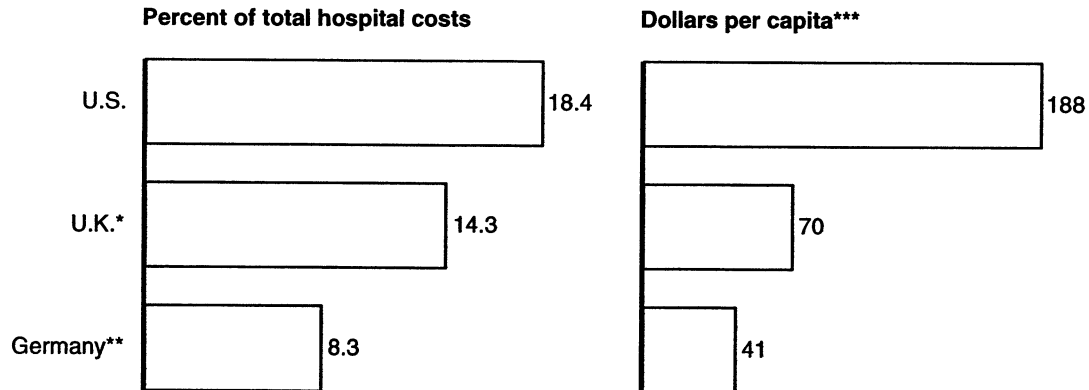
* U.K. data is given for 1993 and calculated assuming that the average administrative staff salary is equivalent to the average hospital staff salary

** For Germany, high estimate of 10.6% and low estimate of 6.0% is averaged to obtain 8.3%. High estimate assumes 12.4% of staff costs are administrative, as given by the 1994 Rohn-Klinikum annual report. Low estimate of 6.0% calculated from German Hospital Society data. These costs may be underestimated due to several reasons: first, there is economic incentive to underestimate these costs, as hospitals with high administrative costs will have budgets cut; second, administrative duties (such as collecting co-payments and compiling statistical data) for sickness funds may not have been accounted for; finally, administrative work completed by doctors is not accounted for

*** Calculated by multiplying the hospital administrative cost percentage by the hospital spending per capita as reported by BASYS

Source: Medicare reports; 1995 NHS Trust; German Hospital Society; Rohn-Klinikum annual report (1994); BASYS; McKinsey analysis

022 EF 136220/8

Exhibit 21

COMPARISON OF U.S. ADMINISTRATIVE
COSTS TO ANOTHER SINGLE-PAYOR SYSTEM
1987

Percent of total health care costs   **Dollars per capita**

U.S.*   17-22   343-437

Canada   8-11   113-150

Exhibit 22

SOURCES OF DIFFERENCE IN HEALTH CARE SPENDING

U.S. dollars per capita, 1990, PPP

U.S. vs. Germany

1,473   −390   737   360   259   2,439

Inputs used (U.S. ~15% lower)

Prices of inputs (U.S. ~40% higher)*

Administrative overhead (estimate U.S. 24%; Germany 13%)

Other

Germany   U.S.

U.S. vs. U.K.

437   −185   2,439

686

388

1,113

Inputs used (U.S. ~30% higher)

Prices of inputs (U.S. ~75% higher)*

Administrative overhead (estimate U.S. 24%; U.K. 16%)

Other

U.K.   U.S.

Germany vs. U.K.

226   65   −324

393   1,473

1,113

Inputs used (Germany ~55% higher)

Prices of inputs (Germany ~25% higher)*

Administrative overhead (estimate Germany 13%; U.K. 16%)

Other
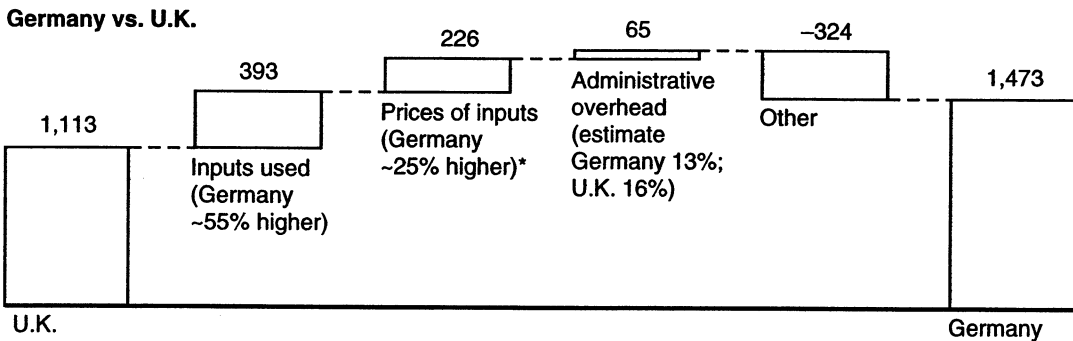
U.K.   Germany

* OECD price index used, which was directionally consistent with relative prices obtained using physician salaries and prices of pharmaceuticals and inpatient hospital stays
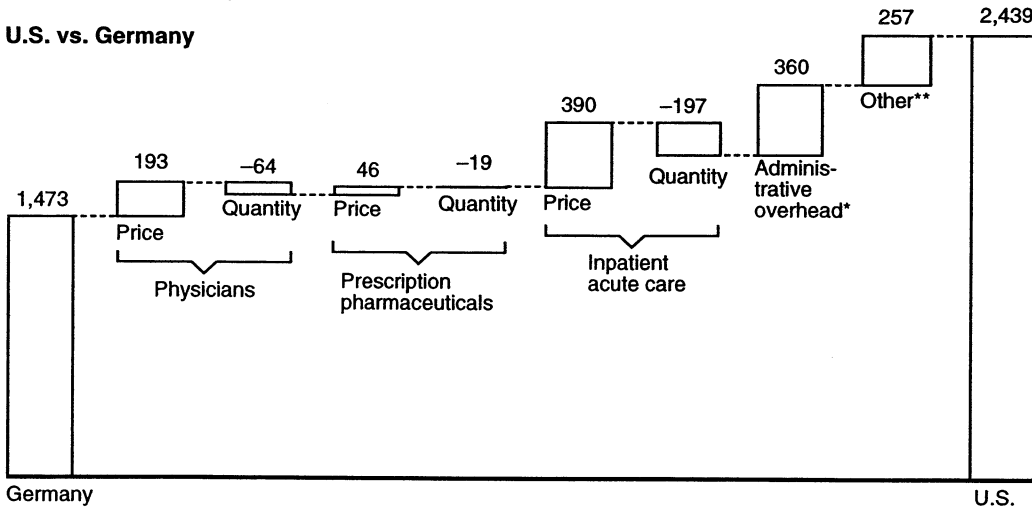
Note: Input levels, price levels, and administrative costs of the country with the higher spending level were used as a base for determining the expected spending of the second country and the corresponding spending gap in the above calculation. Relative input levels were determined using information on the amount of labor (physicians, nurses, and medical technicians), supplies (pharmaceuticals), and capital consumed in each country, assuming a breakdown of 70% labor, 20% supplies, and 10% capital; relative prices were determined using a price reported by the OECD; administrative costs were estimated as 24% in the U.S., 13% in Germany, and 16% in the U.K. based on available literature; education costs are not included

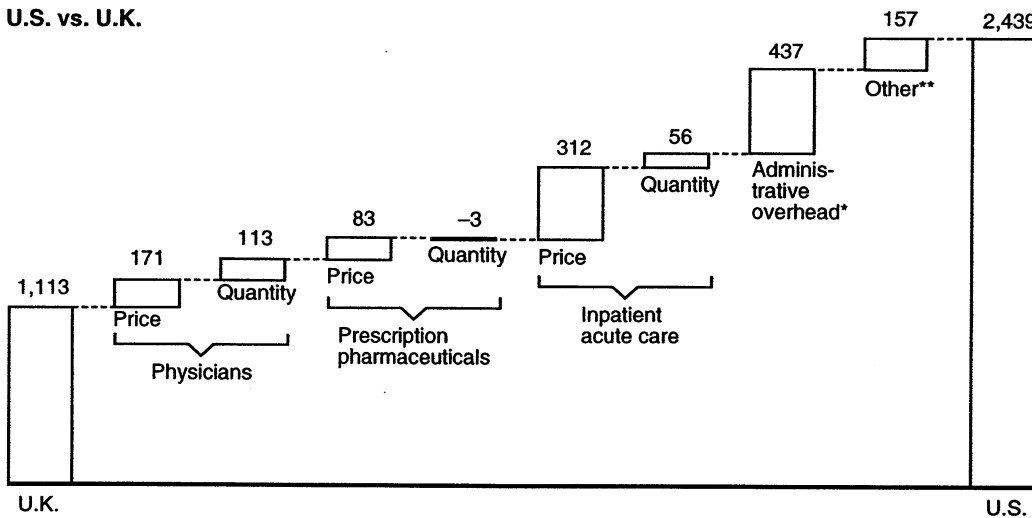Source: BASYS; McKinsey analysis

022 EF 136232/8

Exhibit 23

SPENDING DIFFERENCES DUE TO PHYSICIANS, PHARMACEUTICALS, AND HOSPITALS

U.S. dollars per capita, 1990, PPP

U.S. vs. Germany

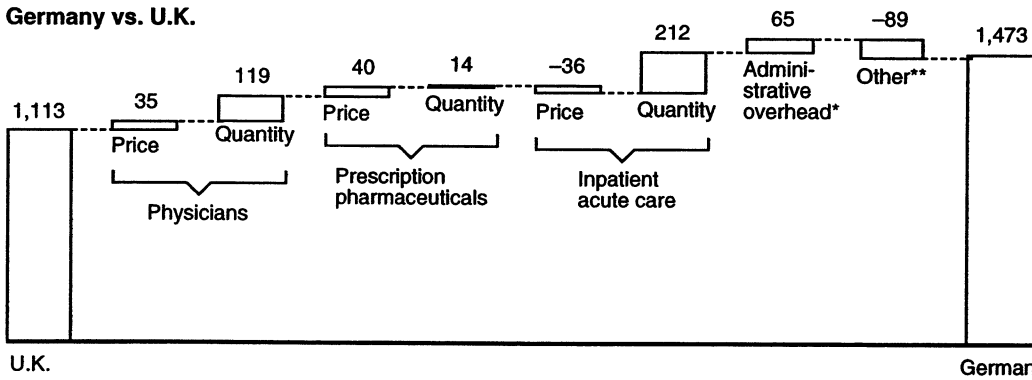Germany: 1,473
Price 193 — Quantity −64 (Physicians)
Price 46 — Quantity −19 (Prescription pharmaceuticals)
Price 390 — Quantity −197 (Inpatient acute care)
Administrative overhead* 360
Other** 257
U.S.: 2,439

U.S. vs. U.K.

U.K.: 1,113
Price 171 — Quantity 113 (Physicians)
Price 83 — Quantity −3 (Prescription pharmaceuticals)
Price 312 — Quantity 56 (Inpatient acute care)
Administrative overhead* 437
Other** 157
U.S.: 2,439

Germany vs. U.K.

U.K.: 1,113
Price 35 — Quantity 119 (Physicians)
Price 40 — Quantity 14 (Prescription pharmaceuticals)
Price −36 — Quantity 212 (Inpatient acute care)
Administrative overhead* 65
Other** −89
Germany: 1,473

* Administrative overhead estimated as 24% in the U.S., 13% in Germany, and 16% in the U.K. based on available literature
** "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

Note: Data sources are different than those for Exhibit 22, explaining variance in total price and quantity differences
Source: AHA; AMA; OECD; BASYS; German Health Care System; NERA; DKG; NHS; Statistisches Bundesamt; McKinsey analysis
022 EF 136223/8

Chapter 9: Implications for policymakers and health care organizations

As outlined in Chapter 1, policymakers and health care organizations need to improve their health care system performance – and specifically, their productive efficiency – in today's increasingly cost-conscious health care environments. Policymakers are struggling with how to structure and regulate their health care markets to improve performance while meeting social objectives. And health care organizations around the globe are struggling to strengthen their positions in their changing health care environments, as well as searching for new opportunities to create value within existing or new markets. Policymakers and health care organizations have pursued significant changes since the time of our assessment, and many are continuing their search for new ideas.

This health care productive efficiency assessment offers findings that inform the strategies of policymakers and health care organizations. The cross-disease synthesis and aggregate analyses highlight the striking variation in overall health care spending and disease-level productive efficiency among the U.S., the U.K., and Germany. While in some cases the system structure and regulation are precluding improvements at the individual health care organization level, most organizations have some freedom to take action. Nevertheless, policymakers play a critical role in stimulating performance improvement in their system through influencing and regulating the market environments in which local health care organizations function.

This chapter summarizes the performance improvement opportunities for policymakers and health care organizations, focusing on productive efficiency.

CRITICAL FIRST STEP: DEFINE THE PROBLEMS AND OPPORTUNITIES

The critical first step for policymakers and health care organizations in any improvement effort is to clearly identify the specific problem or opportunity in precise and appropriate terms – whether productive efficiency (and within this, input levels versus outcomes), relative input prices, or administrative costs. Many tend to combine and confuse these performance dimensions and, therefore, do not take a sufficiently targeted approach in reforming their systems.

For example, each of the systems studied faced different challenges and opportunities in managing their overall health care performance at the time of our assessment:

¶ Germany could have improved productive efficiency, or more specifically, better managed input levels, through directly regulating case rate hospital payments, or possibly allowing a competitive payor-hospital market to determine the form and level of hospital payment. Productive efficiency could have also been improved through eliminating regulatory barriers between the inpatient and outpatient care settings.

¶ The U.K. could have improved productive efficiency in some diseases by adopting some technologies faster and by triaging care more effectively, potentially through a more competitive approach. Alternatively, these issues could have been addressed by better central evaluation and adoption through administrative fiat.

¶ The U.S. could have addressed its lower productive efficiency in diabetes (and any similar diseases) by trying to achieve greater care integration through market mechanisms or selective regulation. The U.S. also needed to understand the causes and implications of its higher input prices and administrative costs.

But while policymakers and health care organizations must clearly identify the specific problems or opportunities in their systems, they must also take a holistic view. Productive efficiency, relative input levels and prices, and administrative cost levels are not independent factors; all seem to have some link to system structure (**Exhibit 1**). Policymakers and health care organizations must therefore consider the potential effects of actions to improve productive efficiency on the level of input prices and administrative costs, and carefully choose whether and how to intervene.

As we saw at the time of our assessment, the U.S. was more productive than Germany, but its higher administrative costs and relative input prices more than offset this efficiency advantage from an aggregate spending-level perspective. Further study is required to determine whether productive efficiency gains stemming from greater competitive intensity or care integration can be captured without triggering some increase in administrative costs and/or input price levels. It is plausible that a well-functioning competitive market, whether achieved through market forces or regulation, could keep administrative costs and input price levels in check.

In the remainder of this chapter, we first discuss general principles for policymakers and health care organizations that want to improve their productive efficiency. We conclude by summarizing major changes in each country since the time of our assessment and their expected impact on productive efficiency and overall health care system performance.

IMPLICATIONS FOR IMPROVING
PRODUCTIVE EFFICIENCY

Our assessment suggests three broad principles for improving productive efficiency on the part of policymakers and health care organizations:

¶ Recognize and leverage the power of economic incentives in influencing treatment decisions and productive efficiency

¶ Allow markets to define health care products broadly and ensure that there are not regulatory barriers to providing more integrated care services

¶ Allow for experimentation and flexibility in the system on the part of health care organizations and providers, given that different system structures seem to work better for different diseases and that medical knowledge and technology are rapidly evolving.

How policymakers and health care organizations can best put these principles into practice will vary significantly, depending on several factors including starting positions, performance pressures, degrees of freedom, and political or social constraints. Below we outline their potential application at the policymaker and health care organization levels.

For policymakers

Because health care system structure and regulation are strongly influenced by policymakers, we focus here on policies that could promote greater system productive efficiency. We recognize that policymakers and voters may be concerned about possible trade-offs between achieving higher levels of productive efficiency and fulfilling social or other objectives, such as equity in health care access or financing and/or specific welfare goals; of course, each country will decide these societal trade-offs for itself. While policies to improve productive efficiency could have negative near-term consequences for social or other objectives, the high cost of such policies should not be underestimated. As discussed in Chapter 1, many countries will find themselves unable to fulfill their social objectives and at the same time maintain their economic vitality without finding ways to improve their health care system's productive efficiency.

The regulation and resulting structure of the health care system – particularly restrictions on the health coverage and care provision markets – are important in creating performance incentives and constraints for individual health care organizations and providers. Policymakers seeking to promote more productive health care systems should consider the following actions:

¶ *Recognize and leverage the power of economic incentives in influencing provider behavior and fostering more productive treatment approaches.* As we saw in our case studies, providers – including hospitals and physicians – respond predictably and systematically to their economic incentives, within the boundaries of acceptable medical practice. We found no evidence that information on the best available medical practice was not widely accessible in all three countries; all things being equal, providers with activity-increasing incentives tend to drive up input usage, even when the likely benefits of more care seem small.

The potentially best approach to creating the right provider incentives is to let the market work, assuming sufficient payor pressure on providers and provider flexibility. However, if the system structure is such that policymakers directly set provider reimbursement (i.e., in Germany and the U.K. in some areas), rather than let the market determine it (i.e., in the U.S.), policymakers should ensure that economic incentives are aligned with productive efficiency for both hospitals and physicians, avoiding regulated fee-for-service (FFS) or per diem approaches. Providers are in the best position to judge the relative benefits of more care, and should at least have activity-neutral incentives, if not direct incentives, to carefully consider the relative cost-effectiveness of alternative treatment approaches.

¶ *Be cautious and selective with direct constraints on hospital, physician, or capital supply, explicitly avoiding and removing those barriers that hinder greater care integration by providers.* As was seen in Germany, supply constraints tend to be a blunt instrument and can often backfire, particularly those that restrict care integration across the inpatient and outpatient settings and those that – due to political decision making processes or other factors – result in higher levels of hospital and physician capacity. In the U.K., we also saw that capital constraints that slowed adoption of productive new technologies (such as laparoscopic cholecystectomy and computerized tomography [CT] scanners for lung cancer) led to lower productive efficiency in some diseases.

¶ *Actively promote care integration for selected chronic or other diseases, possibly through selective regulation or carve-out approaches.* In general, more care integration at the disease level or phase of disease level (e.g., hospital episode) is better than less. Policymakers should therefore avoid directly defining or encouraging a fragmented approach for major diseases through regulated FFS schedules or restrictions on payor's ability to negotiate or coordinate care. Although well-functioning, competitive markets should lead to efficient levels of care integration for most diseases, our case studies suggested that there may be some diseases for which the market alone will not evolve to an effective integration level: specifically, it was unclear whether the competitive U.S. system would evolve to the U.K.'s level of productive efficiency for

diabetes given issues of member and patient churn and adverse selection. In these chronic diseases for which a multi-year time horizon is required for productive efficiency, policymakers may want and need to carve out or directly regulate treatment for these diseases.

¶ *Potentially foster competitive intensity in health care markets through deregulation of key interactions (e.g., payor-provider) and/or creation of internal markets in more centrally controlled systems.* The case studies suggest that higher competitive intensity among payors and providers can lead to higher productive efficiency when new technologies are cost-effective and when the level of care integration in the care provision markets is relatively high. As was seen in the U.K. during the time of our assessment, a less competitive system is less responsive to consumer and payor demands, leading to slower adoption of some productive technologies. But higher competitive intensity can also lead to lower productive efficiency, such as when the care product is quite fragmented or when provider and consumer incentives are not aligned with productive efficiency.

In addition, as was noted earlier, the relationship between higher competitive intensity and levels of administrative costs and input prices is not well understood. In seeking to manage overall health care spending, policymakers must also assess the potential impact of increased fragmentation among payors and/or providers and the potential increase in overall system administrative costs. While in a well-functioning, competitive market such costs should be bid down over time, it is possible that administrative expenses could rise substantially in a more fragmented, competitive system – particularly one that has just recently privatized.

¶ *Allow significant flexibility and freedom in the system for experimentation and innovation on the part of health care organizations and providers.* Medical technology is constantly evolving, as are ways of delivering and managing care across most diseases. Too much rigidity or central control in the system will likely hinder productive efficiency over the long run, as it can stifle innovation and improvement where it needs to occur – at the health care organization and individual provider level. While it is possible that a more centrally controlled system could achieve faster and broader adoption of more productive treatment innovations and new technologies that have been invented elsewhere, such a system may always be playing "catch up" to more flexible, market-based systems.

For health care organizations

Our disease case study results show that most of the productive efficiency differences among the three countries studied can be explained by largely controllable care treatment decisions at the individual provider or health care organization level. While these decisions are greatly influenced by the overall health care system structure and by the economic incentives and supply constraints that arise from this structure, providers and managers of health care organizations can nevertheless take steps to improve within all systems.

The existence of large productive efficiency differences that providers and other health care organizations can influence presents these organizations with significant opportunities and threats – *opportunities* because less productive organizations have substantial room to improve and more productive organizations have an advantage they can exploit to grow in their own country/system or in other countries, and *threats* because less productive organizations are at risk of becoming the competitive victims of other organizations (local or global) if they themselves do not improve, particularly as their health care systems demand better performance. To respond, health care organizations should consider several actions:

¶ *Align providers' economic incentives with productive efficiency at the individual and group provider level for major diseases, encouraging the adoption of productive treatment approaches and technology.* As observed in the case studies, providers respond predictably and systematically to their economic incentives, within the boundaries of acceptable medical practice. Pure FFS incentives tend to drive up input use, all else being equal, and can therefore negatively impact productive efficiency (particularly in the absence of supply constraints). Payment mechanisms that share some financial risk for the cost of care with providers (such as capitation or case rate payments) appear to better align providers' incentives, and thus their behavior, with productive efficiency. Such financial risk sharing is particularly important with specialist physicians, given the specialists' more direct role relative to the primary care physicians' in driving treatment choices in major diseases. However, designing the right incentives in an overall system is not easy, given the complexity of disease treatment processes, the different requirements by disease, and the fact that disease treatment processes are constantly evolving; much more thought is required than just capitating a group of providers.

¶ *At the individual caregiver level, put greater focus on better care triaging – which treatments, for which people, at what stage of the disease progression – as well as on new technology evaluation.* As was observed in the case studies, more explicit consideration of these treatment and technology decisions can reduce input usage by eliminating less necessary care, and can improve outcomes by focusing attention on cases that can

benefit most from interventions. New technologies can also improve productive efficiency through improving outcomes, reducing inputs, or both.

¶ *Integrate elements of care delivery in alternative forms, or "packages," by disease.* Those organizations that understand and manage their system at the disease level – rather than only at the functional or component level (e.g., hospital, ambulatory physician services, pharmacy) – will likely be more productive. Specifically, health care organizations should aim to coordinate care across the care delivery system for major diseases, including as many of the relevant care inputs as possible for managing and treating a disease.

While such coordination may not be necessary for all diseases, those that involve multiple care inputs or providers, have complex treatment options, or involve long time lags between care interventions and their consequences will benefit from more integrated care approaches. Diabetes and other chronic conditions, as well as some cancers, are examples of diseases that would likely benefit from such approaches. Our experience with major payor and provider organizations, as well as literature available in the public domain, would suggest that several other diseases also lend themselves to a more integrated care approach.

Based on our case studies, it does not seem to matter whether the payors or providers perform this disease-specific care integration role; however, providers may have an advantage given their relatively superior information regarding treatment options and potential trade-offs.

¶ *Modify disease-specific care approaches over time as technology and the state of medical knowledge evolves.* Given the rapid changes in technology and evolving state of medical knowledge regarding clinical best practices, the disease-specific approaches around which a health care system is managed must be allowed to change over time. For example, changes were underway in the treatment of cholelithiasis at the time of our case study, with the introduction of the more cost-effective laparoscopic procedure. Similarly, breast cancer treatment was undergoing significant change in the late 1980s with the emergence of fine needle aspiration (FNA) for biopsies, the more widespread adoption of breast-conserving surgery (lumpectomy), and increasing awareness and data on the relative merits of mammographic screening. In these cases, it was important to allow the disease treatment process to evolve within payor and provider organizations to promote productive efficiency.

¶ *Provide the opportunity for payors and providers within the organization to purchase care provision services from each other in a variety of forms, including disease-specific packages.* The case studies showed that when

providers had broad collective responsibilities for a disease for defined patient populations rather than for individual procedures or care components only, higher productive efficiency resulted in some diseases (e.g., diabetes). In other diseases, more narrow care bundles were sufficient for higher productive efficiency. Flexibility within the system to purchase services in a variety of forms – depending on the disease and local market conditions – therefore seems important in promoting productive efficiency at the overall system level.

¶ *Systematically measure and monitor productive efficiency and treatment variations across providers in the system, while ensuring access to the best available information on medical practice by disease.* Ongoing measurement, feedback, and information to providers are critical to stimulating changes in care treatment patterns and innovation. While we did not see major differences in the level and use of information in our disease case studies, the rapid changes in medical practice, the wide variations observed in care treatment patterns, and our work with leading health care organizations suggest that systematic measurement and feedback is critical to fostering higher productive efficiency in most diseases.

Those organizations that are successful in achieving high productive efficiency can strengthen their position in current systems as their health care environment evolves and performance pressures intensify. But while most health care services will continue to be locally consumed and produced, organizations will likely need to benchmark themselves against global – not just local – best practices in their search for performance improvement. Increasingly, the best performing organizations can and should consider exploiting global expansion opportunities.

Although significant cultural and regulatory barriers can make it difficult for foreign health care organizations – particularly payors and providers – to successfully enter some global markets, barriers appear to be lessening in many parts of the world, particularly in Europe. Pharmaceutical and medical device/supply firms have faced a competitive global market for many years; it is likely only a matter of time before health care payors and provider organizations face similar global threats and opportunities. In recent years, foreign payor and provider organizations – particularly U.S. firms such as United Healthcare and Columbia/HCA – have directly established operations or formed skill-based alliances or contractual relationships in other global markets. Examples include United Healthcare's provision of managed care expertise to the major German sickness fund AOK and its joint venture health maintenance organization (HMO) in South Africa, as well as Columbia/HCA's ownership and operation of hospitals in the U.K. and Switzerland.

RECENT CHANGES IN HEALTH CARE
SYSTEMS AND THEIR IMPACT

Since the late 1980s, the health care systems in each country have changed significantly, in directions consistent with higher competitive intensity and greater care integration (**Exhibit 2**). These changes have resulted from initiatives taken by policymakers as well as health care organizations. While the impact of these changes on productive efficiency and other dimensions of health care system performance – including input prices and administrative costs – have yet to be determined, results to date appear consistent with the principles outlined in this chapter.

The U.S. changes

In the U.S., there is evidence that the largely market-based system is leading to greater competitive intensity and higher care integration at least in some diseases, without any significant regulatory changes. More integrated managed care products such as HMOs and preferred provider organizations (PPOs) have emerged as a result of greater employer demands for better value and the increasingly competitive health coverage and care provision markets. In addition, both payors and providers have created disease carve-out products in such areas as cancer care and even diabetes. Specialized clinics and more aggressive management for diabetic care, including emphasis on self-care, have emerged as a result of actions on the part of integrated provider systems, managed care payors, and manufacturers of diabetic supplies. 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 specialist physician compensation and to actual price reductions for health coverage in some markets. The effects on administrative costs are unclear, however. While 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, etc.) have increased. These changes, however, may have improved productive efficiency.

The U.K. changes

In the U.K., the 1991 reforms 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 National Health Service (NHS) largely intact. More decentralized health authorities were given the responsibility of

purchasing services from somewhat competing providers; general practitioners (GPs) were allowed to become "fundholders" and thereby assume and manage the financial risk of a broader set of care provision services (e.g., drugs, outpatient care, diagnostic tests, 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. However, the overall budget and many other supply constraints remain, and efforts to encourage the use of nonpublic financing sources have met with little success.

While the system changes have, not surprisingly, increased administrative costs, their productive efficiency impact is 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.[1] And there is some evidence that adoption of technology has quickened (e.g., a targeted breast cancer screening program based on mammography was established; adoption of laparoscopic technology for cholecystectomy has reached close to 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 GP fundholders can now encourage and achieve more rapid incremental improvements in health care delivery through exerting more direct pressure on local specialists and hospitals. While some supply and capital constraints remain for hospitals and their associated specialists, and competition has been limited to date outside the major metropolitan areas, we would expect some improvement in the U.K. system's productive efficiency over time, at least in the diseases studied.

German changes

In Germany, major reforms have been made in the 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 (e.g., by disease or case) have been left intact. While regulated case rate payments for hospitals have been introduced to substitute for per diem payments, 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-imposed limitations in what they can implement. While

---

[1] Source: Whitney CR. Rising health costs threaten generous benefits in Europe. The New York Times, August 6, 1996.

additional reforms are under discussion for 1997, most are focused on managing hospital costs through, for example, the introduction of a regional- or state-level hospital budget.

It is unlikely that recent changes in the German system will do much to improve productive efficiency, unless they eventually lead to (because of payor pressure) 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 adoption of case rate hospital payments across the board.
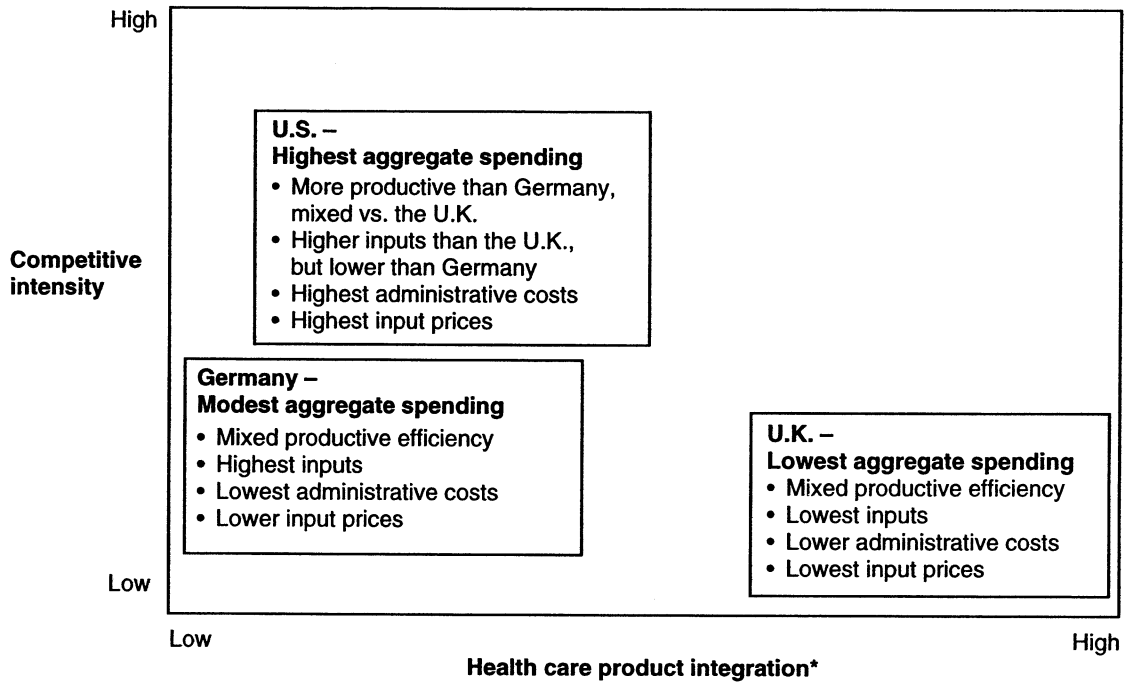
* * *

Overall, this assessment has provided findings regarding the magnitude of productive efficiency differences across countries and the major drivers and sources of these differences – findings difficult to obtain through the aggregate-level analyses that have dominated prior cross-country health care comparisons. In addition, it has shed light on the other sources of vastly different spending levels among three industrialized countries – specifically, different input levels and prices and administrative costs.

Policymakers and most health care organizations in each of the countries studied are actively pursuing performance improvements. In the U.S. and the U.K., recent changes appear to be moving these health care systems toward higher productive efficiency, with each system adopting the more productive characteristics of the other. Given the questionable impact of the German reforms, it is therefore likely that Germany's productive efficiency gap relative to the U.S. and possibly to the U.K. is widening.

Exhibit 1

RELATIONSHIP BETWEEN HEALTH CARE SYSTEM
STRUCTURE AND OVERALL PERFORMANCE
1990

High

**Competitive
intensity**

**U.S. –
Highest aggregate spending**
- More productive than Germany,
  mixed vs. the U.K.
- Higher inputs than the U.K.,
  but lower than Germany
- Highest administrative costs
- Highest input prices

**Germany –
Modest aggregate spending**
- Mixed productive efficiency
- Highest inputs
- Lowest administrative costs
- Lower input prices

**U.K. –
Lowest aggregate spending**
- Mixed productive efficiency
- Lowest inputs
- Lower administrative costs
- Lowest input prices

Low

Low   High
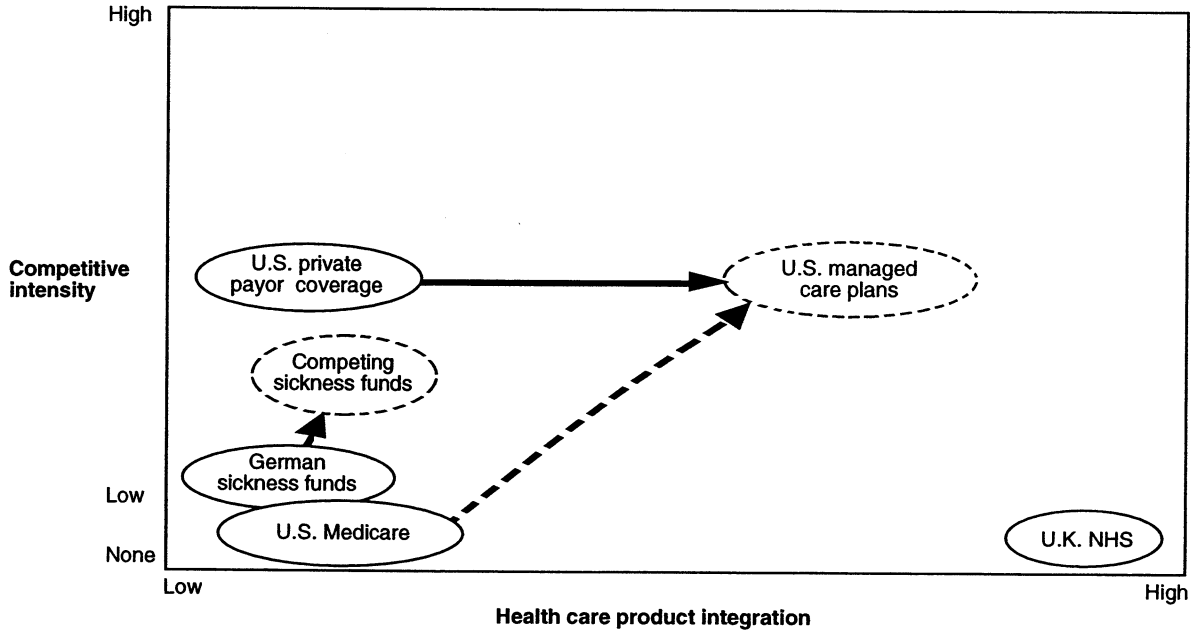
Health care product integration*

* See Chapter 2 of this report for a more complete definition of this term and characterization of the three health care systems at
   the time of our assessment
022 P 134726/9

Exhibit 2

RECENT CHANGES TO HEALTH CARE MARKET STRUCTURES

◯ Mid to late 1980s
⬭ Current

Health coverage market

High

Competitive
intensity

Low

None

U.S. private payor coverage → U.S. managed care plans

Competing sickness funds

German sickness funds

U.S. Medicare

U.K. NHS

Low — Health care product integration — High

Care provision markets: between payors and providers

High

Competitive
intensity

Low

None

U.S. hospital and physician services (wide range of provider products)

U.S. hospital and physician services

U.K. GP fundholder services

Hospital trust services

U.K. specialist services

U.K. GP services

U.S. hospital services for Medicare

German hospital services

U.K. specialist services

German specialist services

U.K. hospital services

Low — Health care product integration — High

022 EF 136258/9