Chapter 16
Health-Care Applications: From Hospitals to Physicians, from Productive Efficiency to Quality Frontiers

Jon A. Chilingerian and H. David Sherman

Abstract This chapter focuses on health-care applications of DEA. The paper begins with a brief history of health applications and discusses some of the models and the motivation behind the applications. Using DEA to develop quality frontiers in health services is offered as a new and promising direction. The paper concludes with an eight-step application procedure and list of do's and don'ts when applying DEA to health services.

Keywords Health Services Research • Physicians • Hospitals • HMOs • Frontier Analysis • Health-Care Management • DEA • Performance • Efficiency • Quality

16.1 Introduction

Throughout the world, health-care delivery systems have been under increasing pressure to improve performance: that is, to control health-care costs while guaranteeing high-quality services and better access to care. Improvements in health-care performance are important because they can boost the well-being, as well as the standard of living and the economic growth of any nation. The quest for high performance in health care has been a difficult and intractable problem historically. Efforts to reduce costs and improve service quality and access have been only marginally successful (Georgopoulos 1986; Newhouse 1994; Shortell and Kaluzny 2000).

Although no theoretically correct or precise measures of "health" exist, there is a great deal of interest in studying and understanding health-care costs, outcomes, and utilization. This interest in understanding health outcomes is associated with our continual desire to improve health care. While we may have no precise measures of

H.D. Sherman (✉)
College of Business Administration, Northeastern University, Boston, MA 02115, USA
e-mail: H.Sherman@NEU.edu

W.W. Cooper et al. (eds.), *Handbook on Data Envelopment Analysis*, International Series in Operations Research & Management Science 164,
DOI 10.1007/978-1-4419-6151-8_16,

health-care performance, the desire to improve health care can be seen in the unrelenting increase in the quantity of health products and services available to patients. Yet, even with these increases, health care seems to offer fewer perceived benefits to patients in relation to their perceived sacrifices (see Anderson et al. 2003; Bristol Royal Infirmary Inquiry Final Report 2002; Newhouse 2002).

As one economist has proclaimed, “Despite the lack of a summary measure of its efficiency, many seem convinced that the industry’s performance falls short” (Newhouse 2002: 14). For example, one medical center in England received special funding to become a Supra Regional Service center for pediatric cardiac surgical care (Bristol Royal Infirmary Inquiry Final Report 2002). This program received funding over a 14-year period, despite significantly higher mortality and morbidity rates and poor physician performance. Should a clinical manager have detected these poor practices sooner? In addition, what internal control systems are available to measure and evaluate individual physician performance?

In 2000, although the USA spent far more on health care per capita than any other country in the world, the number of physicians per 1,000 population, primary care visits per capita, acute beds per capita, hospital admissions per capita, and hospital days per capita were below the median of most other developed countries (Anderson et al. 2003). Are quality and productive efficiency of some national health-care systems really lower than that of many other industrialized nations? What accounts for performance differences?

Twenty years ago, research studies that questioned the patterns or cost of care rarely attempted to estimate the amount and sources of inefficiency and poor performance. Today that has changed. Several hundred productive efficiency studies have been conducted in countries such as Austria, Finland, the Netherlands, Norway, Spain, Sweden, the UK, and the USA. These studies have found evidence that technical inefficiency in these systems is significant. For example, in the USA, such inefficiencies result in billions of “wasted” US dollars (USD) each year. While other statistical techniques have also been used, Data Envelopment Analysis (DEA) has become the researchers’ method of choice for finding best practices and evaluating productive inefficiency.

While there are several techniques for estimating best practices such as stochastic frontier analysis, fixed effects regression models, and simple ratios, DEA has become a preferred methodology when evaluating health-care providers. DEA is a methodology that estimates the degree to which observed performance reaches its potential and/or indicates how well resources have been utilized (see Chap. 1 in this handbook). Benchmarked against actual behavior of decision-making units (DMUs), DEA finds a best practices frontier – i.e., the best attainable results observed in practice – rather than central tendencies. The distance of the DMUs to the frontier provides a measure of overall performance. Although criticized by some health economists (see Newhouse 1994), the late Harvey Leibenstein praised DEA as a primary method for measuring and partitioning X-inefficiency (Leibenstein and Maital 1992).

DEA offers many advantages when applied to the problem of evaluating the performance of health-care organizations. First, the models are nonparametric and do not require a functional form to be prescribed explicitly, i.e., linear, nonlinear, log-linear, and so on (Charnes et al. 1994). Second, unlike statistical regressions that average performance across many service providers, DEA estimates best practice by evaluating the performance behavior of each individual provider, comparing each provider with every other provider in the sample. The analysis identifies the amount of the performance deficiency and the source. Third, unlike regression and other statistical methods, DEA can handle multiple variables, so the analysis produces a single, overall measure of the best results observed.

Finally, to identify those providers who achieved the best results, DEA groups providers into homogeneous subgroups. Providers that lie on the frontier achieved the best possible results and are rated 100% efficient. Providers that do not lie on the surface underperformed, and their performance is measured by their distance from the frontier. The analysis not only provides a measure of their relative performance but also uncovers subgroups of providers similar in their behavior or similar in the focus of their attention to performance.

This chapter focuses on health-care applications of DEA. We begin with a brief history of health-care applications. Next, several health-care models and approaches are discussed, followed by new directions for future studies. We conclude with a summary of do's and don'ts when applying DEA to evaluate the performance of health-care organizations.

16.2 Brief Background and History

For health-care organizations, finding a single overall measure of performance has been difficult. Goals of most health-care services are multiple, conflicting, intangible, vague, and complex. Virtually, every study of efficiency could be criticized for failing to look at quality, clinical innovation, or the changing nature of the services (Newhouse 1994). The worldwide demand to provide health-care services at lower cost made this an ideal focus for DEA research, and health care (like banking) has had innumerable DEA applications.

The first application of DEA in health care began with H. David Sherman's Doctoral dissertation in 1981. W.W. "Bill" Cooper had been a professor at Harvard Business School in 1979, David Sherman was a doctoral candidate and he and Rajiv Banker happened to take Bill Cooper's seminar. Bill Cooper mentioned DEA as a new technique, and both Banker and Sherman used that technique in their dissertations. Working with the Massachusetts Rate Setting Commission to ensure relevance in the complex task of evaluating hospital performance, David Sherman applied DEA to evaluate the performance of medical and surgical departments in 15 hospitals. When Dr. Sherman became a professor at MIT in 1980, his doctoral student and research assistant, Jon Chilingerian helped him to compare DEA results with other statistical models and to look for interesting and novel health-care applications.

In 1983, Nunamaker published the first health application using DEA to study nursing services (Nunamaker 1983). In 1984, the second DEA paper (Sherman 1984) evaluated the medical and surgical departments of seven hospitals. By 1997, Hollingsworth et al. (1999) counted 91 DEA studies in health care. The health applications include the following: health districts, HMOs, mental health programs, hospitals, nursing homes, acute physicians, and primary care physicians.

DEA applications in health have been evolving over the last two decades. As access to bases and information technology have improved, so have the quality of the studies. In the next section, we review some of the DEA literature in health care starting with acute hospitals, nursing homes, other health organizations, and physicians.

16.2.1 Acute General Hospitals and Academic Medical Centers

Acute hospitals have received the most research attention using DEA. These hospital studies use DEA alone. They measure overall technical efficiency using the CCR model. Finally, they define outputs as patient days or discharges and do not measure clinical outcomes. A comprehensive review of efficiency studies in health care by Hollingsworth et al. (1999) found systematic differences among the average efficiency and range of DEA scores by ownership type and national hospital systems. For example, public hospitals had the highest mean efficiency scores (0.96) and not-for-profit hospitals had lower mean DEA scores (0.80). When comparing US studies with studies from other European countries, Hollingsworth et al. found a greater potential for improvement in the USA with an average efficiency score of 0.85, and a range of 0.60–0.98, in contrast to Europe with an average efficiency score of 0.91, and a range of 0.88–0.93.

Although comparing DEA scores among various hospital studies is useful for hypothesis generation, there are many limitations. First, these studies used different input and output measures during different time periods. Second, the distribution of DEA scores is so skewed, (given the huge spike of efficient units), that reliance on the usual measures of central tendency will be misleading. By excluding the efficient units, the average inefficiency score may be a more reasonable comparison. Third, the output measures in these studies are vastly different. Many did not use the same type of case-mix-adjusted data, and some studies used crude case-mix-adjusted data such as age-adjusted discharges, so the results are likely affected by unaccounted case-mix differences.

One recent study of 22 hospitals in the National Health Service in the UK used a four output, five input model (Kerr et al. 1999). The outputs were defined as the following: (1) Surgical inpatients and visits, (2) Medical inpatients and visits, Obstetrics/Gynecology patients and visits, Accidents and Emergency visits. Without knowing the complexity and severity of patients, raw measures of output will lead to distorted results. If Hospital A receives a lower DEA score because Hospital A admits more "fevers of unknown origin," and performs more combined liver–kidney transplants, hip replacements, and coronary bypass grafts and Hospital B has more tooth extractions, vaginal deliveries without complications, and circumcisions, it is an unfair comparison.

Acute hospitals are among the most complex organizations to manage. Periodically, there are random fluctuations and chaos is everywhere – the emergency room, the operating rooms, the intensive care units, and the like. Hospital studies are complex because of the amount of input and output information needed to describe the clinical activities and services and the patients' trajectories. Most of care programs are not really under the control of the hospital manager (see Chilingerian and Glavin 1994; Chilingerian and Sherman 1990). Therefore, traditional DEA hospital studies may not have been useful to practicing managers.

Most of the hospital studies have merely illustrated DEA as a methodology and demonstrated its potential. Unfortunately, they use very different hospital production models. Some combine patient days with discharges as outputs, and others separate the manager-controlled production process from the clinical-controlled process. Few studies have tested any clinical or organizational theory.

Evaluating acute hospitals requires a large and complex DEA model. To identify underperforming hospitals DEA makes nontestable assumptions such as no random fluctuations, no measurement errors, no omitted outputs and output homogeneity (see Newhouse 1994). If additional inputs would improve quality, omitting an output variable such as the quantity of case-mix-adjusted mortalities can distort DEA results. Perhaps bringing DEA inside the hospital to compare departments, care programs, care teams, diseases, specific procedures and physicians' practice patterns would be more useful for practice. One might conclude from this that hospital-level comparisons are not the best application for DEA.

Nevertheless, there have been innovative hospital-level studies potentially useful for policy makers. For example, dozens of DEA papers have focused on the association between hospital ownership and technical inefficiency studying several thousand hospitals as DMUs (see for example, Burgess and Wilson 1996). DEA studies have also focused on critical health policy issues such as the following: regional variations (Perez 1992), rural hospital closures (Ozcan and Lynch 1992), urban hospital closures (Lynch and Ozcan 1994), hospital consolidations (Luke and Ozcan 1995), and rural hospital performance (Ferrier Ferrier and Valdanis 1996). O'Neill (1998) has recently made an important methodological contribution by developing an interesting DEA performance measure that allows policy makers to make fair comparisons of teaching and nonteaching hospitals. The DEA work on hospital performance continues to be important and needs more development.

16.2.2 Nursing Homes

Sexton et al. and Nyman and Bricker published the first two DEA studies of nursing homes in 1989. Sexton et al. ran a model that relied on two output measures (Medicaid and Other) and six inputs that only included labor; consequently, the study had some limitations. Nyman and Bricker (1989) used DEA to study 195 for profit (FP) and not-for-profit (NFP) US nursing homes. Employing four categories of labor hours as inputs (e.g., nursing hours, social service hours, therapist hours, and other hours), and five outputs (skilled nursing patients (SNF), intermediate care patients (ICF), personal care patients, residential care patients, and limited care patients), they regressed the DEA scores in an ordinary least squares regression analysis and reported that NFP nursing homes were more efficient.

Another study in the USA by Nyman et al. (1990) investigated the technical efficiency of 296 nursing homes producing only intermediate care, no skilled nursing care, and relying on 11 labor inputs. They defined only one output, the quantity of intermediate care patients produced. Fizel and Nunnikhoven (1993) investigated the efficiency of US nursing home chains ignoring nonlabor inputs and focusing on two outputs: the quantity of intermediate and skilled nursing patients. They study found that chains were more efficient. Kooreman's 1994 study of 292 nursing homes in the Netherlands utilized six labor inputs and four case-mix-adjusted outputs: the quantity of physically disabled patients, quantity of psychogeriatrically disabled, quantity of full care, and quantity of day care patients. He found that 50% were operating efficiently, and the inefficient homes used 13% more labor inputs per unit of output, and quality and efficiency seemed to be going in the opposite direction.

A study of 461 nursing homes by Rosko et al., in 1995 employed five labor inputs, and two outputs, ICF patients and SNF patients. The study found that the variables associated with nursing home efficiency were managerial and environmental. Differences in efficiency were not associated with quality measures.

The nursing home studies are among the better applications of DEA. Although they encounter the same case-mix problems when modeling outputs, most of these studies regress the DEA scores to identify the variables associated with inefficiency. Since the outputs are often adjusted by payment types, or are crude patient types, many of these studies controlled for quality, patient characteristics, while exploring effects of ownership, operating environment and strategic choice on performance. These studies have found that managerial and environmental variables (the location of the home, nurse training, size of the homes, and wage rates) are strongly associated with the DEA scores, rather than quality of care, or patient mix.

16.2.3 Department Level, Team-Level, and General Health-Care Studies

In addition to acute care hospitals and nursing homes, DEA has been applied in a wide variety of health services and activities. For example, previous work has studied the productive efficiency of the following: Obstetrics units (Finkler and Wirtschafter 1993), pharmacies (Fare et al. 1994), intensive care units

(Puig-Junoy 1998), organ procurement programs (Ozcan et al. 1999), and dialysis centers (Ozgen and Ozcan 2002). All of these studies have used DEA to measure technical efficiency using basic models.

Ozcan et al.'s exploratory study conducted in 1999 on 64 organ procurement organizations employed four inputs (a capital proxy, FTE development labor, FTE Other Labor, and operating expenses), and two outputs (kidneys recovered, extra-renal organs recovered). The study found some evidence of scale efficiency not only did the larger programs produce 2.5 times more outputs, the average DEA efficiency scores were 95% for the large programs versus 79% for the smaller programs.

In 2000, a paper comparing the efficiency of 585 HMOs from 1985 to 1994 computed the estimates using data envelopment analysis, stochastic frontier analysis, and fixed effects regression modeling (Bryce et al. 2000). Unfortunately, they relied on a single output (total member-years of coverage) four input (hospital days, ambulatory visits, administrative expenses, and other expenses) model. They concluded that the three techniques identify different firms as more efficient.

In 2002, Ozgen and Ozcan reported a study of 791 dialysis centers in the USA. Constructing a multiple output (quantity of dialysis treatments, dialysis training, and home visits) and multiple input model (including clinical providers by type, other staff, operating expenses, and the number of dialysis machines – a proxy for capital), the study evaluated pure technical efficiency assuming variable returns to scale. The study found that the wide variations in efficiency were associated with ownership status – for-profit dialysis centers were less inefficient than not-for-profit ones.

16.2.4 Physician-Level Studies

A new and interesting development in health care has been taking DEA down to the workshop-level, focusing on individual physician performance. The first application of DEA at the individual physician level was in 1989; the analysis identified the nontechnical factors associated with technical efficiency of surgeons and internists (Chilingerian 1989).

Physician studies have developed new conceptual models of clinical efficiency (Chilingerian and Sherman 1990), as well as using DEA to explore new areas such as the following: most productive scale size of physician panels (Chilingerian 1995); benchmarking primary care gatekeepers (Chilingerian and Sherman 1996); and preferred practice styles (Chilingerian and Sherman 1997; Ozcan 1998). Getting physicians to see how their practice behavior ranks in relation to their peers is a step toward changing the culture of medicine and offering insights into a theory of clinical production management (Chilingerian and Glavin 1994).

In 1989, Chilingerian used DEA to investigate the nontechnical factors associated with clinical efficiency of 12 acute hospital surgeons and 24 acute hospital internists. Since clinical efficiency assumes a constant quality outcome, the study classified each physician's patients into two outcomes: (1) satisfactory cases – i.e., patients discharged alive without morbidity; and (2) unsatisfactory cases – i.e., patients who experienced either morbidity, or who died in the hospital. Relying on a two output (low and high severity discharges with satisfactory outcomes), two input (ancillary service, and total length of stay) model, each of the 36 physicians were evaluated (Chilingerian 1989).

The study reported that 13 physicians were on the best practice frontier (i.e., 13 physicians produced good outcomes with fewer resources). After controlling for case-mix complexity and the severity mix of the patients, the DEA scores could not be explained by the type of patients treated. Moreover, younger physicians (age 40 and under) who belonged to a group practice HMO were more likely to practice efficiently.

In 1995, Chilingerian reported a 6-month follow-up study on the 36 acute care physicians (Chilingerian 1994). A Tobit analysis revealed that physicians affiliated with a group practice HMO were likely to be more efficient as well as physicians whose medical practices focused on a narrow range of diagnoses. Using DEA to investigate the most productive scale size of hospital-based physicians, the study reported that on average one high-severity patient utilized five times the resources of a low-severity patient.

In 1997, Chilingerian and Sherman evaluated the practice patterns of 326 primary care physicians in an HMO. They employed a seven output (gender and age-adjusted patients) eight input (acute hospital days, ambulatory surgery, office visits, subspecialty referrals, mental health visits, therapy units, tests, and emergency room visits) model. They demonstrated that the HMO's belief that generalists are more efficient than subspecialists was not supported. Using clinical directives to establish a cone ratio model, practice styles were identified that should have increased their office visits and reduced hospitals days. Thus, a DEA-based definition of efficiency identified relatively "efficient" physicians who could change their styles of practice by seeing more of their patients in their offices rather than in the hospital.

In 1998, Ozcan studied 160 primary care physicians' practice styles using DEA with weight restrictions to define preferred practice styles for a single medical condition. The study identified three outputs (low, medium, and high severity patients with otitis media, a disease that affects hearing) and five inputs (primary care visits, specialist visits, inpatient days, drugs, and lab tests). The study found 46 efficient and 114 less efficient practices. After defining preferred practice styles, even the efficient physicians could have a reduction in patient care costs of up to 24%.

16.2.5 Data Envelopment Analysis Versus Stochastic Frontier Analysis

Several health-care papers have compared DEA results with other techniques such as stochastic frontier analysis (SFA). These techniques use two vastly different optimization principles. SFA has one overall optimization across all the observations to arrive at the estimates of inefficiency. DEA runs a separate optimization for each hospital, thus allowing a better fit to each observation and a better basis for identifying sources of inefficiency for each hospital.

The differences in optimizing principles used in DEA and regression estimates suggest that one might be preferred over the other to help solve different problems. For example, SFR may be more helpful to understand the future behavior of the entire population of hospitals. DEA might be used when the policy problem centers on individual hospitals how specific inefficiency can be eliminated. One problem with comparing DEA to SFA is that SFA requires a specification such as a translog function with a single dependent variable. Forcing DEA to utilize the same dependent variable as a single output will lead to the conclusion that the results are similar, but not exactly the same (see Bryce et al. 2000).

16.2.6 Reviewer Comments on the Usefulness of DEA

As the life cycle of DEA matures, we have seen an increase in the health applications published. However, the skeptics remain. The most skeptical view on frontier estimation has come from Joseph P. Newhouse at the Harvard School of Public Health. In 1994, in the *Journal of Health Economics*, Newhouse, raises several fundamental questions about frontier studies (Newhouse 1994). For example, he asks if one could define a frontier with certainty what purpose would it serve? Can frontier studies be used to set reimbursement rates? Can frontier analysis create benchmarks?

Unlike kilowatt-hours, Newhouse also suggests that because the product in health care is neither homogeneous, nor unidimensional the output problem is serious. He also identifies three other problems with modeling medical care delivery as a production process. First, frontier studies often omit critical inputs such as physicians, contract nurses, capital inputs, students, and researchers. Employing case-mix measures such as Diagnostic Related Groups, or Ambulatory Cost Groups, can also hide severity within a diagnosis or illness. He points out that DEA assumes no random error or random fluctuations, which is often not the case in health care.

The problem of measurement errors, or unclear, ambiguous, and omitted inputs and outputs will haunt every DEA health-care application. The threshold question is this – how seriously do these issues affect the results? While Newhouse claims a serious distortion, the question is largely empirical. Researchers must put that issue to rest each time they conduct a DEA study. Do the results make sense?

To advance the field, every DEA health-care study should test the following hypothesis:

Ho: Variations in DEA scores (i.e., excess utilization of clinical resources) can be explained by complexity, severity, and type of illness.

Though researchers cannot prove that DEA scores are measuring efficiency, they can disconfirm the case-mix hypothesis. If DEA scores are not associated with severity, patient characteristics or poor quality the DEA results become very interesting. One analytic strategy for disconfirming the case mix/quality hypothesis is to regress the DEA scores against the best available measures of case mix and patient characteristics. We discuss how to use Tobit models to validate DEA scores in another section of this chapter.

16.2.7 Summary

One conclusion to be reached from all of the health-care studies is that even when it appears that a substantial amount or resources could be "saved" if every hospital, nursing, or physician were as good as those who use medical techniques the least, DEA scores must always be interpreted with care. Researchers should assume that there is unaccounted case mix until that issue is put to rest.

In summary, there are two obstacles to advancing applications of DEA to health services research. The first impediment is the confusion around modeling hospitals, physicians, nursing homes, and other health-care providers. The findings from DEA studies lose credibility when inputs and outputs are defined differently from study to study. A second impediment that interferes with the development of generalizations from DEA applications is the lack of stability in the results from different studies. A variable such as quality of care, associated with productive efficiency in one study, disappears in another study (see Kooreman 1994; Rosko et al. 1995). Rarely have two researchers studied the same problem and when they have, rarely have they employed the same categories of inputs and outputs.

One way to deal with these difficulties is to try to ascertain how much of the efficiency scores is explained by case mix. For example, studies that blend DEA with Tobit or other statistical models can sharpen the analysis of best practices. If adequate controls are included in the model, the challenge of output heterogeneity in health care can be laid to rest (see Rosko et al. 1995; Kooreman 1994; Chilingerian 1995).

This review suggests that DEA is *capable* of producing new knowledge advancing the science of health-care management. As Chap. 1 has argued:

> . . .DEA proves particularly adept at uncovering relationships that remain hidden from other methodologies. For instance, consider what one wants to mean by "efficiency", or more generally, what one wants to mean by saying that one DMU is more efficient than another DMU. This is accomplished in a straightforward manner by DEA without requiring expectations and variations with various types of models such as are required in linear and nonlinear regression models. . .. (See Chapter 1 of his handbook).

Before DEA becomes a primary tool to help policy makers and practicing clinical managers, researchers must shift from health care "illustrations of DEA" to advancing the field of performance improvement in the delivery of health services. If DEA is to reach its potential for offering new insights and making a difference to clinicians and patients, DEA research has to overcome the many pitfalls in health care. The remainder of this chapter identifies the issues, and offers some practical suggestions for dealing with them. We conclude with a new approach that investigates quality frontiers in health care.

16.3 Health-Care Models

Medical care production is different from manufacturing. A traditional factory physically transforms raw materials into finished products. The customer is absent, so there is no participation or co-production. If demand is higher than capacity, inventory can be stored. In manufacturing, quality can be built into design of a product. Since goods can be inspected and fixed at every workstation, quality can improve productive efficiency.

In health care after patients are admitted to a care facility (or visit a clinic) there are three major clinical processes: (1) investigation/diagnosis, (2) treatment/ therapy, and recovery. Health-care services are intangible "performances" that can only be experienced or used (Teboul 2002). They must be consumed immediately following production, and they cannot be stockpiled. Unused capacity – i.e., nursing care, empty beds, idle therapists, and the like – is a source of inefficiency. However, overutilization – i.e., unnecessary tests, X-rays, and surgeries, or unnecessary days in the intensive care unit or nursing home – is also a source of inefficiency. Once delivered, clinical services cannot be taken back and corrected patient perceptions are immediate. In the next section, efficiency in the health-care context is defined and discussed.

16.3.1 Clinical Efficiency Definitions

Clinical inefficiency in the provision of health-care services occurs when a provider uses a relatively excessive quantity of clinical inputs when compared with providers treating a similar case load and mix of patients. There are many categories of efficiency – technical, scale, allocative, and overall efficiency. Most studies in health care have measured the overall technical and scale components of clinical efficiency, which is represented by the "average productivity attainable at the most productive scale size. . .." (Banker, Charnes, and Cooper 1984, p. 1088). Most simply, technical inefficiency refers to the extent to which a DMU fails to produce maximum output from its chosen combination of factor inputs, and scale inefficiency refers to suboptimal activity levels.

Evaluating a health-care provider's clinical efficiency requires an ability to find "best practices" – i.e., the minimum set of inputs to produce a successfully treated patient. Technical inefficiency occurs when a provider uses a relatively excessive quantity of clinical resources (inputs) when compared with providers practicing with a similar size and mix of patients. Scale inefficiency occurs when a provider is operating at a suboptimal activity level – i.e., the unit is not diagnosing and/or treating the most productive quantity of patients of a given case mix. Hence, hospital providers will be considered 100% efficient if they cared for patients with fewer days of stay and ancillary services and at an efficient scale size. Primary care providers will be considered efficient if they cared for their panels of patients with fewer visits, ancillary tests, therapies, hospital days, drugs, and subspecialty consults.

Researchers can use a variety of DEA models to measure and explain overall technical and scale efficiency. The CCR model, initially proposed by Charnes et al. (1978) is considered a sensitive model for finding inefficiencies (Golany and Roll 1989). In 1984, Banker et al. added another very useful model (BCC model) for health-care studies. The BCC model can be used to separate technical from scale efficiency. Both models (if formulated as input-minimizing) can be used to explore some of the underlying reasons for inefficiency – e.g., to estimate divergence from most productive scale size and returns to scale (Banker et al. 1984). Consequently, DEA can yield theoretical insights into the managerial problems or decision choices that underlie efficient relationships: e.g., magnitude of slack, scale effects of certain outputs on the productivity of inputs, marginal rates of substitution and marginal rates of transformation and so on (Cooper et al. 2000b).

Once DEA rates a group of providers efficient and inefficient, researchers, managers and/or policy makers can use this information to benchmark best practice by constructing a theoretical production possibility set. Analysts or researchers could use the DEA linear programming formulations to estimate potential input savings (based on a proportional reduction of inputs). Analysts or researchers can use the ratios of the weights u_r and v_i to provide estimates of marginal rates of substitution and marginal rates of transformation of outputs, measured on a segment of the efficient frontier (Zhu 2000). Again, analysts or researchers could use the BCC model to evaluate returns to scale – i.e., in the case of physicians, the effects of a small versus large proportion of high severity cases. Furthermore, the production possibility set is used to estimate alternative outputs of low severity and high severity patients that can be offered clinical services having utilized a mix of clinical services (for a primary care example see Chilingerian 1995).

16.3.2 How to Model Health-Care Providers: Hospitals, Nursing Homes, Physicians

The threshold question for DEA is what type of production model to choose. Depending on the type of health-care organization, there are many ways of conceptualizing the inputs and outputs of production. Since the selection of inputs and outputs often drives the DEA results, it is important to develop a justification for selecting inputs and outputs. In the next section, we review DEA models used in various health applications. First, we differentiate clinical from managerial efficiency.

16.3.3 Managerial and Clinical Efficiency Models

In health care, technical efficiency is not always synonymous with managerial efficiency. On the one hand, technical efficiency in nursing homes, rehabilitation hospitals, and mental health facilities can be equated with managerial efficiency. On the other hand, medical care services especially in acute and primary care settings are fundamentally different in that there are two medical care production processes, and consequently types of technical and scale efficiency: managerial and clinical. Managerial efficiency requires practice management – i.e., achieving a maximum output from the resources allocated to each service department, given clinical technologies. Clinical efficiency requires patient management – i.e., physician decision-making that utilizes a minimal quantity of clinical resources to achieve a constant quality outcome, when caring for patients with similar diagnostic complexity and severity.

In the case of acute hospitals, the role of the manager is to set up and manage a decision-making organization whose basic function is to have clinical services ready for physicians and other clinical providers. For example, hospital managers must have admitting departments, dietary and diagnostic departments, operating rooms, and ICU and regular beds staffed and ready to go. Physicians make decisions to use these intermediate products and services, and managers make it all available. The major challenge facing hospital managers is to decide how much reserve capacity is reasonable given fluctuating patient admissions (daily and seasonal), and stochastic emergency events. Managerial efficiency can be equated with producing nursing care, diagnostic and therapeutic services, and treatment programs of satisfactory quality, using the least resources. Good practice management achieves managerial efficiency.

Physicians are fundamentally different from other caregivers. Not only are they providers of medical care but they also enjoy the primary "decision rights" for patient care, with little interference from management. They are the DMUs that steer a patient through various phases of patient care – such as office visits, primary care and diagnostic services, hospital admissions, consultations with other physicians, surgeries, drugs, discharges, and follow-up visits. Physicians organize and direct the entire production process, drawing on the talents of a hidden network of providers – nurses, therapists, dietary specialists, and the like. The reason that physician practice patterns are of interest is that 80–90% of the health-care expenditures in every system are the result of physician decision-making. These are dominant and highly influential DMUs. Clinical inefficiency then, as it is used here, refers to physicians who utilize a relatively excessive quantity of

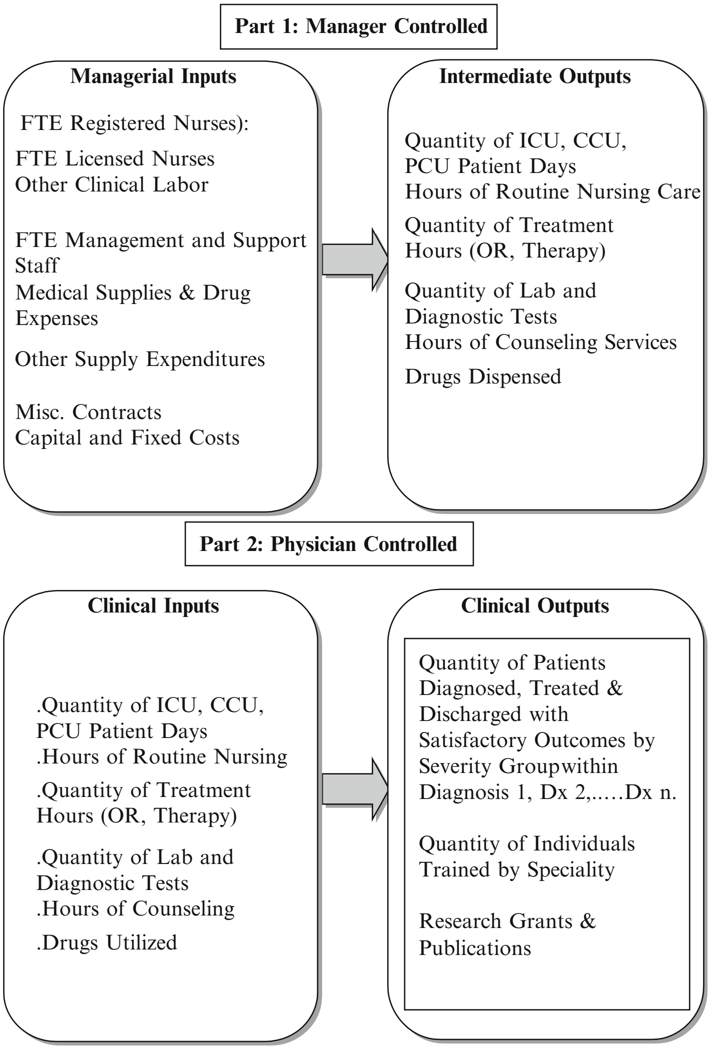


Fig. 16.1 Acute hospital as two-part DEA model (Chilingerian and Sherman 1990)

clinical resources, or inputs, when compared with physicians practicing with a similar size and mix of patients.

Figure 16.1 describes the medical service production system as a two-part service process: (1) a manager-controlled DMU, and (2) a physician-controlled DMU. In the diagram below, the intermediate outputs of the manager-controlled production process become the clinical inputs for the physician-controlled production process. A discharged patient is the final product, and the clinical inputs are the bundle of intermediate services that the patient received.

Hospital managers set up and manage the assets of the hospital. They control the labor, the medical supplies, and all expenditures related to nursing care, intensive care, emergency care, and ancillary services (such as lab tests, radiology, and other diagnostic services), pharmacy, dietary, as well as laundry, central supplies, billing, and other back-office functions. However, these departments (or functions) merely produce intermediate services that are available for utilization by physicians (see Chilingerian and Glavin 1994; Chilingerian and Sherman 1990; Fetter and Freeman 1986). Physician decision-making determines how efficiently these assets are utilized. Once a patient is admitted to hospital, physicians decide on the care program – i.e., the mix of diagnostic services and treatments, as well as the location and intensity of nursing care, and the trajectory of the patient. Physicians decide how and when to utilize nursing care, intensive care, emergency care, ancillary services, and other clinical inputs.

The productive efficiency of the hospital is complicated. A hospital can be clinically efficient, but not managerially efficient. A hospital can be managerially efficient, but not clinically efficient. More often, both parts of the production process are inefficient. If physicians over utilize hospital services the cost-per-patient day, cost-per-nursing hour, cost-per-test is reduced and the hospital appears to making the best use of its inputs.

To be efficient, clinical and nonclinical managers must perform two tasks very well. Clinical managers must manage physicians' decision-making (i.e., patient management) and nonclinical managers make the best use of all hospital assets by managing operations (i.e., practice management). Therefore, patient and practice management require an extraordinary amount of coordination and commitment to performance improvement.

16.3.3.1 Medical Center and Acute Hospital Models: Examples of Managerial Efficiency

Three examples of acute production models appear below. The first model (Burgess et al. 1998) includes five types of labor inputs and weighted beds as a proxy for capital, but excludes drugs, medical supplies and other operating expenses. The second model (Sexton et al. 1989) collapses nurses into one category, but adds physicians and residents and excludes beds. The third model collapses labor into one variable, includes other operating expenses and beds, but also adds a proxy measure of capital based on a count of the number of specialty and diagnostic services (Fig. 16.2–16.4).

The outputs are different in all three models. Conceptually if the inputs are costs, then the input/output ratios are cost per case, cost per procedure, cost per visit, or cost per nursing day. If the inputs are beds or FTE labor, then the input–output ratio is represented by labor utilized per admission, labor utilized per patient day, labor per surgery, and the like. Although mixing managerial inputs with clinical outputs is acceptable, managerial and clinical inefficiencies become indistinguishable.

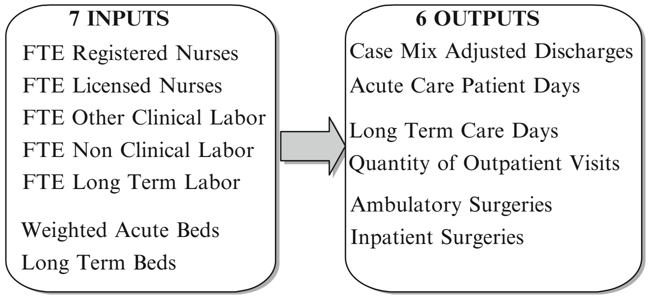


Fig. 16.2 Variables in general acute hospital model (Burgess et al. 1998)

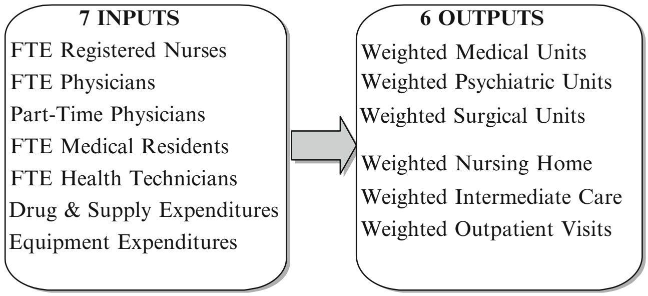


Fig. 16.3 Variables in a medical center study (Sexton et al. 1989)

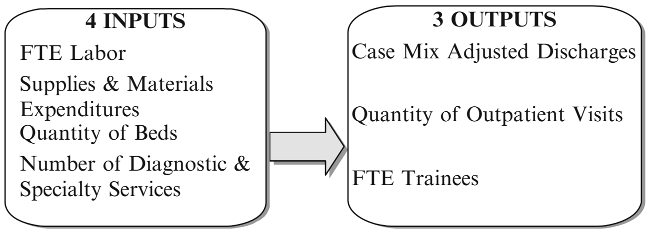


Fig. 16.4 Variables in urban hospital model (Ozcan and Luke 1992)

16.3.3.2 Nursing Homes: Another Example of Managerial Efficiency

Nursing home studies in the USA typically segment the outputs by sources of payments: quantity of residents supported by the state, or people without insurance. Figure 16.5 displays the inputs and outputs often used in DEA nursing home studies. The nine resource inputs are full-time equivalent (FTE) registered nurses, FTE licensed practical nurses, and FTE nurse aides, FTE other labor, and medical supplies and drugs, clinical and other supplies, and claimed fixed

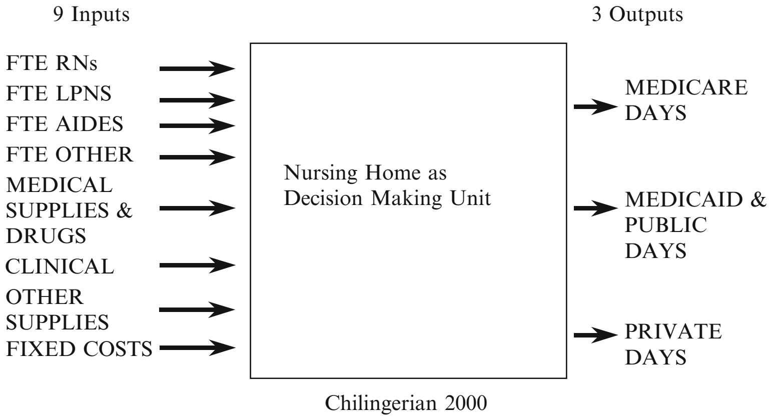


Fig. 16.5 Nursing home inputs and outputs (Chilingerian 2000)

costs (a proxy for capital). Since DEA can handle incommensurable data, the FTEs are in quantities, and the supplies, drugs, and fixed costs are measured by the amount of dollars spent. The outputs are the quantity of nursing home days produced during a given time period. In the Fig. 16.5, the outputs are the quantity of resident days broken into three payment classification groups: Medicare patients (a national program to pay for elderly care), Medicaid patients (a state program to pay for impoverished residents), and Private patients (residents without financial assistance).

Problems arise when the outputs are not homogeneous due to unaccounted case mix. For example, if the nursing home has skilled nursing beds, an Alzheimer unit, or a large proportion of patients older than 85, confused, requiring feeding, bathing, and toilet assistance, then the model is not measuring differences in pure productive efficiency. One solution is to collect information on patients' characteristics and regress the DEA scores against the patient, environmental, and managerial factors to be sure that the DEA scores are not due to case mix. For an example of this type of study, see Rosko et al. 1995.

16.3.3.3 Primary Care Physician Models: An Example of Clinical Efficiency

With growth of managed care, the primary care physician has emerged as an important force in the struggle for efficient and effective medical care. Since lab and radiology tests, prescription drugs, surgeries, and referrals to hospitals all require a physician's approval, physicians report cards or profiles have become a way to benchmark physician practice patterns. Managers could use DEA as a tool to profile and evaluate physicians.

Previous research has found that three patient variables drive managed care costs. They are patients' age, gender, and geographic location. Consequently, managed

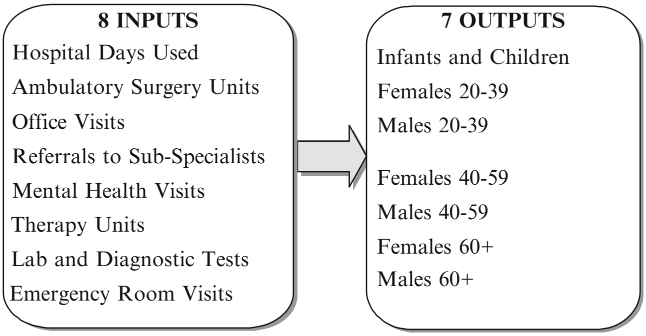


Fig. 16.6 Variables in a primary care physician study (Chilingerian and Sherman 1997)

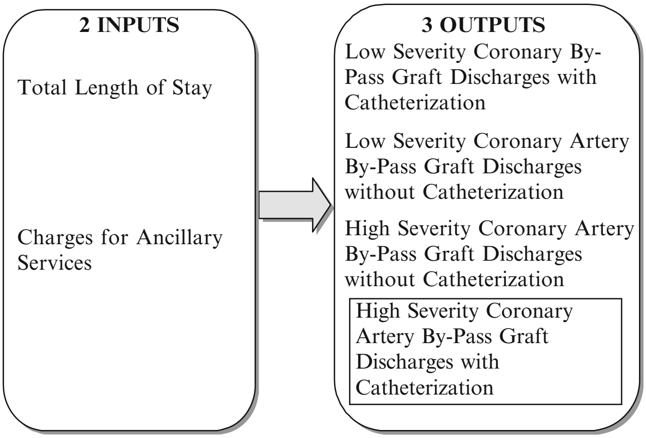


Fig. 16.7 Variables in a cardiac surgeon model of DRG 106 and 107 (Chilingerian et al. 2002)

care organizations set their budgets and prices based on these variables. The final product produced by physicians in managed care organizations is 1 year of comprehensive care for their patients. To care for patients, primary care physicians utilize office visits, hospital days, lab tests, and therapy units. Figure 16.6 is an example of how one large HMO conceptualized a physician DEA application.

16.3.4 Hospital Physician Models: Another Example of Clinical Efficiency

To study the practice behavior at the individual physician-level, analysts and researchers can use a variety of DEA models. For example, one recent study a study of 120 cardiac surgeons evaluated how efficiently they performed 30,000 coronary artery bypass grafts (CABG) on patients over a 2-year period (see Chilingerian et al. 2002). Figure 16.7 illustrates a two-input, four-output clinical production model used in that study. The two inputs are defined as

(1) the total length of stay (days) for the CABG cases handled and (2) the total ancillary and other charges (dollars) for the CABG cases handled. The ancillary and other charges input category includes ancillary, drug, equipment, and miscellaneous charges. The first input, length of stay, represents a measure of the duration of CABG admissions and the utilization of clinical inputs such as nursing care and support services. The second input, ancillary and other charges, represents a measure of the intensity of CABG admissions and the utilization of operating rooms, laboratory and radiological testing, drugs, and so on.

The four classes of clinical outputs represent completed CABG surgery cases. Since patients with more severe clinical conditions will likely require the use of more clinical inputs, the efficiency analysis must account for variations in case mix to be fair to surgeons or hospitals treating relatively sicker CABG patient populations. Accordingly, the outputs are defined by diagnostic category and severity level within diagnosis. In this example, a system of case mix classification called Diagnostic Related Groups (DRG) are used to segment outputs by complexity; moreover, a severity system called MEDSGRPS was used to further segment each DRG into low and high severity categories. The researchers treated DRG 106 and DRG 107 as separate clinical outputs because a CABG procedure with catheterization is more complicated and requires more clinical resources. As explained above, each DRG was further divided into low-severity and high-severity cases.

16.3.5 Profitability Models: A Nursing Home Example

Although "profit" is still a dirty word in health care, there is a need to do more performance studies that look at revenue and expenses, and investigate the factors affecting profitability. In these studies, the maximum profit includes actual profit, plus maximum overall inefficiency (see Cooper et al. 2000a). Figure 16.8 illustrates an example of a profitability model for a nursing home.

Since the performance measure, takes the form of Profit = Revenues−Expenses, an additive mode could be used (see early chapters in this handbook). The additive model shown below has several advantages.

$$\max z = \sum_{r=1}^{s} \mu_r y_{ro} - \sum_{i=1}^{m} v_i x_{io} + u_o,$$

subject to

$$\sum_{r=1}^{s} \mu_r y_{rj} - \sum_{i=1}^{m} v_i x_{ij} + u_o \leq 0,$$

$$\mu_r, v_i \geq 1.$$

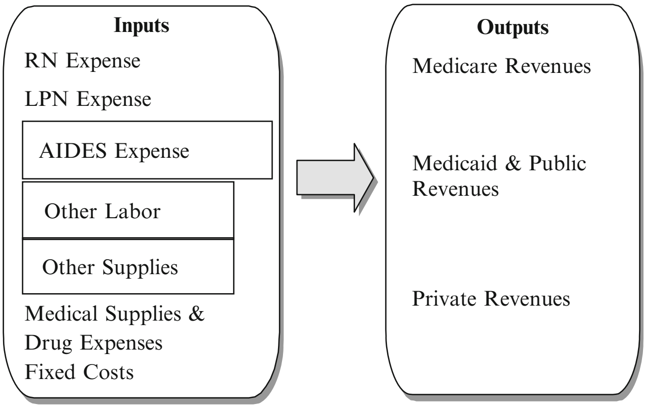


Fig. 16.8 Hypothetical example of nursing home inputs and outputs in the profitability model

One advantage of the additive model is that the objective function, max z, can be interpreted as maximizing profit, or maximizing an excess of revenue over expenses. The model contains an unconstrained variable, u_o, which forces the model to assume variable return to scale (see early chapters in this handbook). Given the advantages and disadvantages of having a great deal of capacity, this assumption works for a profitability model. The optimal value of v_r^* can provide insights with respect to the trade-offs associated with adding skilled nursing, or reducing other labor or fixed expenses.

Employing profitability models to evaluate health-care organizations is a promising area for research for two reasons. First, profit models make the evaluation of the complex delivery systems such as acute care hospitals more manageable. Second, this approach opens up the opportunity to use a multi-stage evaluation approach. With a three-stage approach, DEA could be used to investigate the underlying reasons for the profitability. During stage one, an average DEA score of the clinical efficiency of the each hospital's busiest physicians could be obtained by studying the individual physician level. For the second stage, a profitability model would be run, obtaining DEA scores for each hospital in the comparison set. Finally, the third stage would regress the hospital's profitability scores against the explanatory variables, including the average clinical efficiency of physicians.

The variables that influence health-care profit margins include the following: cost drivers, prices (relative to competitors), relative scale and capacity, the market share, quality of care, and technical delivery mode (Neuman et al. 1988). According to the literature, the main cost drivers are: the volume of patients, within diagnostic categories, adjusted by severity; the practice style and clinical behavior of the medical staff; and the degree of managerial and clinical efficiency (Chilingerian and Glavin 1994).

16.4 Special Issues for Health Applications

There are conceptual, methodological, and practical problems associated with evaluating health-care performance with DEA. First, conceptualizing clinical performance involves identifying appropriate inputs and outputs. Selecting inputs and outputs raises several questions – Which inputs and outputs should the unit be held accountable? What is the product of a health-care provider? Can outputs be defined while holding quality constant? Should intermediate and final products be evaluated separately?

Another conceptual challenge involves specifying the technical relationship among inputs. In analyzing clinical processes, it is possible to substitute inputs, ancillary services for routine care days, more primary care prevention for acute hospital services, and the like. Within the boundaries of current professional knowledge, there are varieties of best practices. Consequently, an evaluation model should distinguish best practices from alternative practice styles.

Methodological problems also exist around valid comparisons. Teaching hospitals, medical centers, and community hospitals are the same. Nursing homes and skilled nursing facilities may not be the same. These institutions have a different mix of patients and have different missions. Finally, the information needed to perform a good study may not be available. Much of the interesting DEA research has not relied on an easily accessible dataset, but required merging several data bases (see Burgess et al. 1998; Rosko et al. 1995).

There are problems about choice of inputs and outputs, and especially finding an "acceptable" concept of product/service. Which inputs and outputs should physicians be held accountable? In addition, there are other issues about measures and concepts. For example:

- Defining Models from Stakeholder Views
- Selection of Inputs and outputs
- Should inputs include environmental and organizational factors?
- Problems on the Best Practice Frontier
  - Are the input factors in medical services substitutable?
  - Are returns to scale constant or variable?
  - Do economies of scale and scope exist?

We discuss each of these issues in the next section.

16.4.1 Defining Models from Stakeholder Views

The models used in DEA depend on the type of payment system and the stakeholder perspective. Every health-care system faces different payment and financing schemes that determine the location of financial risk and, subsequently, the interests of the stakeholders. Examples of payment and finance systems include fee-for-service reimbursement, prospective payment, fixed budgets, and capitation. All of these payment systems have one thing in common – they are all seriously flawed. They can imperil quality and often have contributed to increased health-care costs.

There are also many stakeholders, each with different and sometimes conflicting interests. For example, there are institutional providers (hospitals, nursing homes, clinics, etc.) as well as individual providers (physicians, nurses, therapists, etc.), governments (local, regional, and national health authorities), third party payers (insurance companies, sick funds, etc.), and patients and their families. All stakeholders want better quality and better access for patients. Although stakeholders talk about more efficient care, they focus most of their attention on how to stem their rising costs.

In many countries, ambulatory physician services are reimbursed on a fee-for-service basis. Since there is no incentive to reduce patient care, from a provider, self-interest standpoint under an unrestricted fee-for-service the care philosophy becomes *more care = better care*, which can lead to over utilization of clinical services and rising health costs. Unless physician decision-making is managed by strong professional norms and benchmarking best practices, systems that create incentives to maximize patient visits, hospital days and procedures leads to rising costs and less efficient care.

Prepayment is the flip side of fee-for-service. With a prepayment system, there is a contractually determined, fixed price/payment for a defined set of services. For example, countries such as Germany, Italy, and the USA have prospective payment systems that reimburse the hospital for each patient admission based on diagnosis irrespective of the actual costs. Taking the hospital's perspective, to enhance revenue the hospital must keep the beds full (increase admissions), and minimize the hospital's cost-per-admission, which of course can lead to over utilization of clinical services and rising health costs.

Capitation and gate keeping is another payment scheme that places physicians' interests at risk. The idea behind this design is that by having the primary care physician (PCP) assume responsibility for all aspects of care for a panel of patients, the patients will enjoy a greater continuity of care while the HMO achieves a more efficient care process at a more consistent level of quality. The "gatekeeper" receives a fixed monthly payment for each patient (adjusted actuarially by age and gender) as an advance payment to provide all the primary care the patient needs. The incentives can be designed so that primary care physicians prosper if they keep their patients out of the hospital and away from specialists. However, the less care the PCPs provide, the greater is the financial gain. Failure to provide needed care in a timely manner to HMO patients may make these patients' treatments more expensive when their illnesses become more advanced. This type of payment scheme can lead to underserved patients and rising health costs.

The payment systems described above can create weak or even distorted motivating environments. DEA models should be defined from a particular stakeholder point-of-view; moreover, models should be selected to ensure that patients are neither underserved nor overserved.

16.4.2 Selecting Appropriate Health-Care Outputs and Inputs: The Greatest Challenge for DEA

Acute care hospitals and medical centers are complex medical care production processes, often bundling hundreds of intermediate products and services to care for each patient (Harris 1979). Major surgery might also require major anesthesia, blood bank services, hematology tests, pathology and cytology specimen, drugs, physical therapy, an intensive care bed, and time in routine care beds. Given the complexity, the greatest challenge for DEA health-care applications lies in conceptualizing and measuring the inputs and outputs (see Chilingerian and Sherman 1990; Newhouse 1994).

Dazzling new technologies, desires to improve efficiency, and new consumer attitudes have changed service capacities, practice behaviors and outcomes. The explosive growth in outpatient surgeries is a good example. In the USA, in 1984 only 400,000 outpatient surgeries were performed; by 2000 that number grew to 8.3 million (Lapetina and Armstrong 2002). The intensity and mix of outpatient and inpatient surgeries are continually changing. When outpatient surgeries shift from tooth extractions to heterogeneous procedures such as hernia repairs, cataract and knee surgeries, and noninvasive interventions, simple counts of the quantity of surgical outputs are misleading. The mix and type of surgeries must be taken into account as well as the resources and technological capabilities of the DMUs.

The first DEA paper published by Nunamaker (1983) used a one input–three-output model. The outputs used crude case-mix adjustments (pediatric days, routine days, and maternity days). The single input was an aggregate measure of inpatient routine costs, which will lead to unstable results unless the inputs among the comparison set homogeneous. We know that hospital salaries, education, experience and mix of the nurses and other staff, the vintage of the capital and equipment, the number of intensive beds, medical supplies and materials are not the same (see Lewin 1983).

To make the DEA results useful for practice, it makes sense to segment inputs into a few familiar managerial categories that make up a large percent of the expenditures. Since acute hospitals have high labor costs, it makes sense to distinguish the types of personnel: i.e., the number of full time equivalent (FTE) registered nurses, FTE licensed practical nurses, FTE nurses aids, FTE therapists, other FTE clinical, and general administrative staff. Selecting managerially relevant categories is necessary if analysts are to use DEA to identify the sources waste and inefficiency.

16.4.2.1 Take Two Aspirin and Call Me in the Morning

With respect to outputs, there are two requirements. First, health-care outputs cannot be adequately defined without measures of case-mix complexity and severity. Second, it makes no sense to evaluate the efficiency of a medical service that results in an adverse event: such as morbidity, mortality, readmissions, and the like. To construct the "best practices" production frontier, observed behavior is evaluated as clinically inefficient if it is possible to decrease any input without increasing any other input and without decreasing output; and if it is possible to increase any output without decreasing any other output and without increasing any input (see Chap. 1 in this handbook). A DMU will be characterized as perfectly efficient only when both criteria are satisfied.

If hospital discharges include both satisfactory and unsatisfactory outcomes, a hospital could be considered efficient if they produced more output per unit of input. If a patient die in the operating room and few clinical resources are utilized, the outcome falls short of the clinical objective. Clinical efficiency requires that outcomes be considered in the performance standard. Therefore, as a concept, clinical efficiency makes sense if and only if the clinical outputs achieve constant quality outcomes. Therefore, the best attainable position for a health-care organization is when a unit achieves maximum the outputs should guarantee constant quality outcomes. Unsatisfactory outcomes should be taken out of the DEA analysis and evaluated separately (see Chilingerian and Sherman 1990). Likewise, if a primary care physician uses fewer office visits, fewer hospitals days, and fewer tests because she or he are postponing care, they should not be on the best practice frontier.

16.4.2.2 Using DEA to Adjust Outputs for Patient Characteristics and Case mix

If we are to go beyond simple illustrations, a DEA study should provide some guarantee that the outputs are similar. For example, it has long been established that acute patient care cannot be measured by routine patient days, maternity days, and pediatric days alone because patient days are an intermediate output and a poor indicator for other services such as blood work, X-rays, drugs, intensive care days, physical therapy and the like. Therefore, it is important that both clinical and ancillary services be included. In contrast to acute hospitals, nursing home studies that define nursing days segmented by payer-mix and/or age-adjusted may be an acceptable proxy of final outputs.

To help guide policy makers or practitioners, researchers might consider the following four-part DEA analytic strategy. The first part would begin by running the some DEA models and the second part by regressing the DEA scores against the case-mix and patient characteristic variables using a censored regression model such as Tobit. If the goodness-of-fit test is significant, adjust each health-care provider outputs by multiplying them by the ratio of the original DEA score to the Tobit's predicted DEA score. Some providers operating with a less complex case mix will have their outputs lowered and other providers operating with a more complex mix will have their outputs increased. A second DEA model would be run during the third part and the last step would be regressing the "new" efficiency scores again to validate that they are unrelated to any control variables other than the critical managerial or policy variables of interest. Using DEA with this analytic strategy might provide some additional insights to policy makers or managers searching for answers to questions such as what is the impact of ownership structure or leadership on the productivity of the industry.

16.4.3 Should Environmental and Organizational Factors Be Used as Inputs?

Random variations in the economic environment (such as labor markets, accidents, epidemics, equipment failures, and weather) and organizational factors (such as leader behavior, employee know-how, and coordination techniques) can influence organizational performance. However, there are so critical inputs and outputs that must be included in health applications, that nondiscretionary variables should be omitted from the DEA model. For a strong health application, the clinical production model should only include resources that clinical decision-makers utilize and manage. Environmental (and other organizational) factors omitted from the DEA model should be included in a second stage model that investigates variables associated with the DEA scores.

Every researcher must decide on a reasonable conceptual model. To guide research, collaborating with practicing managers or policy makers to identify relevant environmental and explanatory factors can only strengthen a health application.

16.4.4 Problems on the Best Practice Frontier: A Physician Example[1]

As described in Chap. 1, of this book, DEA relies on the Pareto-Koopmans definition of efficiency. For example, to construct the "best practice" production frontier for primary care physicians, observed behavior is evaluated by using the following input–output criteria:

1. A physician is clinically inefficient if it is possible to utilize fewer clinical resources without increasing any other resources and without decreasing the number of patients cared for;
2. A physician is clinically inefficient if it is possible to care for more of a given segment of patients without decreasing the care given to any other segment of patients and without utilizing more clinical resources.

[1]The next section draws heavily on Chilingerian and Sherman (1995), Rosko (1990) and Charnes, Cooper and Rhodes (1978).

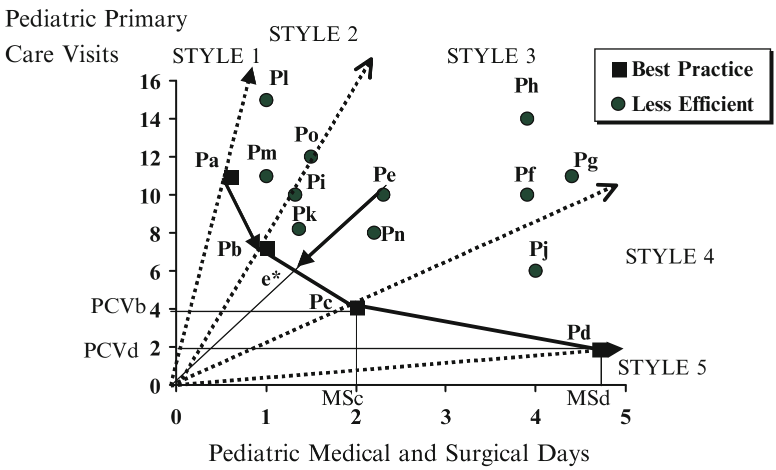


Fig. 16.9 Two-dimensional picture of pediatrician practice behavior

A physician can be characterized as clinically efficient only when both criteria are satisfied and when constant quality outcomes can be assumed.

A brief example with a small group of pediatricians illustrates how DEA can be used to define a best practice production frontier for primary care physicians. Consider 14 primary care pediatricians who cared for 1,000 female children 1–4 years old from a homogeneous socio-economic background. Figure 16.9 plots the amount of medical-surgical hospital days used and primary care visits for one full year, for the 1,000 children.

The Pareto-Koopmans efficiency criteria suggests that any pediatrician in Fig. 16.9 who was lower and to the left of another pediatrician was more successful because he or she used fewer clinical inputs to care for the "same" type of children. By floating a piecewise linear surface on top of the actual observations, Fig. 16.9 also plots a best practices production frontier (BPPF) for each of the 14 pediatricians. As shown in Fig. 16.9, pediatricians Pa, Pb, Pc, and Pd dominate Pe, Pf, Pg, Ph, Pi, Pj, Pk, Pl, and Pm. Accordingly, Pa, Pb, Pc, and Pd, physicians not dominated by any other physicians lie on a best practices production frontier and all designated as frontier points.

DEA defines a best practice frontier by constructing a set of piecewise linear curves, such that any point on the BPPF is a weighted sum of observed DMUs. DEA measures the relative efficiency of each physician by the relative distance of that physician from the frontier. Physicians on the frontier (points Pa, Pb, Pc, and Pd) have a DEA score of 1 and physicians off the frontier have a value between 0.0 and 1.0. For example, in Fig. 16.9, the efficiency score of Pe, a pediatrician less efficient than Pb and Pc, can be measured by constructing a line segment from the origin to point Pe. This line segment crosses the BPPF at point e'. The efficiency rating of Pe is equal to the ratio of the length of line segment Oe* to the length of OPe. Since the numerator Oe* is less than the denominator OPe, the efficiency rating of physician Pe will be less than 1.0.

The practice of pediatric primary care medicine is a very complicated production problem where there are many slightly different ways to mix inputs to care for a population of patients. To solve the problem, each physician adopts a "style of practice"[2] that can be represented by the mix of clinical inputs used. The geometry of convex cones provides a pictorial language for interpreting the practice style of physicians. In mathematics, a cone is a collection of points such that if a physician P is at any point in the collection, then μP is also in the collection for all real scalars $\mu \geq 0$ (Charnes et al. 1953). However, cones can also be interpreted as a linear partition of physician practice styles based on a set of linear constraints such as a range of substitution ratios.

For example, in Fig. 16.9 the ray that begins at the origin (0) and intersects the frontier at Pb and the ray that begins at the origin and intersects the frontier at Pc define the upper and lower boundaries of a cone or *practice style 3*. All points contained on the line that begins at the origin (0) and intersects the frontier at Pb have the same ratio of medicine and surgery days to primary care office visits. Therefore, a physician practicing *style 3* will have a ratio of visits to hospital days that falls between the line that begins at the origin (0) and intersects the frontier at Pb and the line that begins at the origin and intersects the frontier at Pc.[3]

So, each cone can also be thought of as a different practice style. For example, practice style 2 represents physicians who use relatively more visits than hospital days. In contrast, practice style 5 represents physicians who use relatively more medical-surgical days than primary care office visits. By modifying the standard DEA model, it can be used to locate cones or natural clustering of DMUs based on the optimal weights assigned by the linear program. The next section illustrates how the new DEA concept of assurance regions can be used to identify and investigate practice styles.

16.4.4.1 The Concept of a Preferred Practice Cone or Quality Assurance Region

The largest single expense item for an HMO is hospital days. Lower utilization rates for hospital days represent more desirable practice style. Physicians who succeed in substituting primary care visits and ambulatory surgeries for medical-surgical days will contribute to a more efficient delivery system. Understandably, most HMO managers would prefer their physicians to reduce hospital days and increase their primary care visits.

[2]Practice style, as it used here, refers to the treatment patterns defined by the specific mix of resource inputs used by a physician (or group of physicians) to care for a given mix of patients.

[3]Although the two-input single output problem can be solved graphically, multi-input multi-output problems require a mathematical formulation that can only be solved by using a linear programming model.

Microeconomics offers a way to quantify the clinical guideline that primary care physicians (PCPs) should trade-off medical surgical days (input X_2) for more primary care visits (input X_1) while maintaining a constant output rate. Economists refer to the change in medical-surgical days (ΔX_2) per unit change in primary care visits (ΔX_1) as the marginal rate of technical substitution – i.e., $-\mathrm{d}x_2/\mathrm{d}x_1$, which equals the marginal product of medical surgical days (ω_h) divided by the marginal product of primary care visits (ω_v).[4]

In Fig. 16.9 every physician on the BPPF practices an input minimizing style of care – i.e., their DEA efficiency score = 1. However, not all of these physicians on the BPPF are practicing minimum-cost care. In other words, the mix of inputs used by some of these physicians is not the most cost effective. Consider the physicians Pd in practice style 5 and Pc in practice style 4. Since style 4 cares for the same patients with fewer hospital days, cost conscious HMO managers would probably want their primary care physicians practicing to the left of style 5.

Now, consider the BPPF in Fig. 16.9. If the quantities of medical-surgical days are increased from MSc to MSd, the same panel of patients will be cared for with slightly fewer of the less expensive primary care visits and much more of the expensive medical and surgical days. This points out a weakness in using DEA to estimate the best practice production frontier. From an HMO's perspective, although the practice style of physician Pd represents an undesirable practice region, the standard DEA model would allow Pd to be on the BPPF.

Modified versions of additive DEA models allow managers to specify bounds on the ratio of inputs such as the ratio of hospital days to primary care office visits (see Cooper et al. 2000a). These bounds can be called a preferred practice cone (Chilingerian and Sherman 1997). This new development in DEA bounds the optimal weights (or marginal productivities) and narrows the set of technically efficient behaviors. By incorporating available managerial knowledge into a standard model, an assurance region truncates the BPPF frontier – i.e., tightens the production possibility set – and opens up the potential to find more inefficiency.

16.4.4.2 Constant Versus Variable Returns to Scale

One difficult issue with every health application is whether to use DEA models that assume variable or constant returns to scale. Researchers should address this question based on prior knowledge and logical inferences about the production context. While imaginative guesses are tolerable, it is unacceptable to pick a model to get "better looking" DEA results.

Although a health-care organization's production function may exhibit variable returns to scale, there are intuitive reasons for expecting that a physician's clinical

[4] In economics, holding constant all other inputs, the marginal product of an input is the addition to total output resulting from using the last unit of input. The ratio of the marginal products is the marginal rate of technical substitution, defined as $-dx_2/dx_1$.

production function exhibit constant returns to scale (Pauly 1980). Physicians are taught that similar patients with common conditions should be taken through the same clinical process. Thus if a surgeon operates on twice as many patients with simple inguinal hernias or performs twice as many coronary bypass grafts, they would expect to use twice as many clinical resources. Consequently, scaling up the quantity of patients should result in a doubling of the inputs.

Hospitals and physician practices both vary in size. In North America, the largest stand-alone hospitals have less than 1,500 beds. In Europe, there are medical centers with as many as 2,000 and 3,000 beds. What is the minimal efficient bed size or the most productive bed size? Some primary care physician practices vary from 1,200 to 4,000 patients. From a quality standpoint, what is the most optimal patient size? Questions about returns to scale should be addressed in future DEA research and would benefit from cross-cultural comparisons.

16.4.4.3 Scale and Scope Issues

Health-care providers and organizations are multiservice firms, offering many clinical services to provide convenient, one-stop shopping, to connect diagnostic services with treatments, etc. Is it more efficient to offer many services under one roof? If offering many services together is more efficient, then economies of scope exist.

On the contrary, the rise in “focused factories” in health care such as hernia clinics, heart clinics, hip replacement centers, and the like, is stirring a great deal of interest throughout the world. Are patients better off going to a super-focused clinic? For example, is a cataract hospital more efficient than a general hospital performing fewer cataract procedures in its operating rooms? Are the efficiency gains of a focus strategy large or small? Questions about scale versus scope should also be addressed in future DEA research.

16.4.5 Analyzing DEA Scores with Censored Regression Models

DEA’s greatest potential contribution to health care is in helping managers, researchers, and policy makers understand why some providers perform better or worse than others do. The question can be framed as follows – How much of the variations in performance are due to: (1) the characteristics of the patients, (2) the practice styles of physicians, (3) the microprocesses of care, (4) the managerial practices of the delivery systems, or (5) other factors in the environment? The following general model has been used in this type of health-care study:

DEA Score = f (ownership, competitive pressure, regulatory pressure, demand patterns, wage rates, patient characteristics, physician or provider practice characteristics, organizational setting, managerial practices, patient illness characteristics, and other control variables).

The DEA score depends on the selection of inputs and outputs. Hence every health application is obliged to disconfirm the hypothesis that DEA is not measuring efficiency, but is actually picking up differences in case mix or other nondiscretionary variables. The best way to validate or confirm variations in DEA scores is to regress the DEA scores against explanatory and control variables. But what type of regression models should be used?

If DEA scores are used in a two-stage regression analysis to explain efficiency, a model other than ordinary least squares (OLS) is required. Standard multiple regression assumes a normal and homoscedastic distribution of the disturbance and the dependent variable; however, in the case of a limited dependent variable the expected errors will not equal zero. Hence, standard regression will lead to a biased estimate (Maddala 1983). Logit models can be used if the DEA scores are converted to a binary variable – such as efficient/inefficient. However, converting the scores < 1 to a categorical variable results in the loss of valuable information; consequently logit is not recommended as a technique for exploring health-care problems with DEA.

Tobit models can also be used whenever there is a mass of observations at a limiting value. This works very well with DEA scores which contain both a limiting value (health-care providers whose DEA scores are clustered at 1) and some continuous parts (health-care providers whose DEA scores fall into a wide variation of strictly positive values < 1). No information is lost, and Tobit fits nicely with distribution of DEA scores as long as there are enough best practice providers. If, for example, in a sample of 200 providers less than 5 were on the frontier, a Tobit model would not be suitable.

In the econometrics literature, it is customary to refer to a distribution such as DEA as either a truncated or a censored normal distribution. There is, however, a basic distinction to be made between truncated and censored regressions. According to one source:

> The main difference arises from the fact that in the censored regression model the exogenous variables x_i are observed even for the observations for which $y_i > L_i$. In the truncated regression model, such observations are eliminated from the sample. (Maddala 1983:166)

Truncation occurs when there are no observations for either the dependent variable, y, or the explanatory variables, x. In contrast, a censored regression model has data on the explanatory variables, x, for all observations; however the values of the dependent variable are above (or below) a threshold are measured by a concentration of observations at a single value (Maddala 1983). The concentration of threshold values is often based on an actual measure of the dependent variable – i.e., zero arrests, zero expenditures – rather than an arbitrary value based on a lack of information.

DEA analysis does not exclude observations greater than 1, rather the analysis simply does not allow a DMU to be assigned a value greater than 1. Hence, Chilingerian (1995) has argued that DEA scores are best conceptualized as a censored, rather than a truncated distribution. The censored model would take the following form:

Efficiency score $=$ actual score if score < 1
Efficiency score $=$ 1 otherwise

There is a substantial literature on modeling data with dependent variables whose distributions are similar to DEA scores. For example, empirical studies with the number of arrests after release from prison as the dependent variable (Witte 1980) or the number of extra marital affairs as the dependent variables (Fair 1978) are among the best known examples in the published literature on the Tobit censored model. Each of these studies analyzes a dependent variable censored at a single value (zero arrests, zero marital affairs) for a significant fraction of observations. Just as the women in Fair's study can do no better than have zero extra-marital affairs, neither could a relatively efficient health-care provider be more efficient than 1. Thus, one could equate zero extramarital affairs or zero arrests to a "best practicing" hospital or provider.

A censored Tobit model fits a line that allows for the possibility of hypothetical scores > 1. The output can be interpreted as "adjusted" efficiency scores based on a set of explanatory variables strongly associated with efficiency. To understand why censored regression models make sense here, one must consider how DEA evaluates relative efficiency.

DEA scores reflect relative efficiency within similar peer groups (i.e., within a "cone of similar DMUs") without reference to relative efficiency among peer groups (i.e., cones). For example, an efficient provider scoring 1 in a peer group using a different mix of inputs (i.e., rates of substitution) may produce more costly care than a provider scoring 1 in a peer group using another mix of inputs (Chilingerian and Sherman 1990). Superior efficiency may not be reflected in the DEA scores because the constraints in the model do not allow a DMU to be assigned a value greater than 1. If DEA scores could be readjusted to compare efficiencies among peer groups, some physicians could have a score that is likely to be greater than 1. Despite the advantages to blending nonparametric DEA with censored regression models in practice, some conceptual problems do arise.

The main difficulty of using Tobit to regress efficiency scores is that DEA does not exactly fit the theory of a censored distribution. The theory of a censored distribution argues that due to an underlying stochastic choice mechanism or due to a defect in the sample data there are values above (or below) a threshold that are not observed for some observations (Maddala 1983). As mentioned above, DEA does not produce a concentration of ones due to a defect in the sample data, rather it is embedded in the mathematical formulation of the model.

A second difficulty of using Tobit is that it opens up the possibility of rank ordering superior efficiency among physicians on the frontier – in other words "hypothetical" scores > 1. In production economics, the idea that some DMUs with DEA scores of 1 may possibly have scores > 1 makes no sense. It suggests that some candidates for technical efficiency (perhaps due to random shifts such as luck, or measurement error) are actually less efficient.

Despite these drawbacks, blending DEA with Tobit model's estimates can be informative. Although DEA does not fit the theory of a censored regression, it easily fits the Tobit model and makes use of the properties of a censored regression in practice. For example, the output can be used to adjust efficiency scores based on factors strongly associated with efficiency.

Tobit may have the potential to sharpen a DEA analysis when expert information on input prices or exemplary DMUs is not available. Thus, in a complex area such as physician utilization behavior, Tobit could help researchers to understand the need to introduce boundary conditions for the DEA model's virtual multipliers.

The distribution of DEA scores is never normally distributed, and often skewed. Taking the reciprocal of the efficiency scalar, (1/DEA score), helps to normalize the DEA distribution (Chilingerian 1995).

Greene (1983) points out that for computational reasons, a convenient normalization in Tobit studies is to assume a censoring point at zero. To put a health-care application into this form, the DEA scores can be transformed with the formula:

Inefficiency score $=$ (1/DEA score) $-$ 1

Thus, the DEA score can become a dependent variable that takes the following form:

DEA Inefficiency score $= xB + u$ if efficiency score > 0
DEA Inefficiency Score $= 0$ otherwise

Once health-care providers' DEA scores have been transformed, Tobit becomes a very convenient and easy method to use for estimating efficiency. The slope coefficients of Tobit are interpreted as if they were an ordinary least squares regression. They represent the change in the dependent variable with respect to a one unit change in the independent variable, holding all else constant.

When using Tobit models they can tested with a log-likelihood ratio test. This statistic is calculated by $-2\log(\lambda)$, where $\log\lambda$ is the difference between the log of the maximized value of the likelihood function with all independent variables equal to zero, and the log of the maximized values of the likelihood function with the independent variables as observed in the regression. The log-likelihood ratio test has a chi square distribution, where the degrees of freedom are the number of explanatory variables in the regression.

16.5 New Directions: From Productive Efficiency Frontiers to Quality-Outcome Frontiers

Although most applications of DEA have been applied to estimations of technical efficiency and production frontiers, the methodology offers an empirical way to estimate quality frontiers. For example, any dimension of quality can be assessed by employing multiple indicators in a DEA model and comparing a provider against a composite unit projected onto a frontier.

Table 16.1 Nursing home outcomes: 100 patients traced for 6 months

Nursing home	Functional status Q_1	Functional status Q_3
N1	0.15	0.40
N2	0.15	0.14
N3	0.35	0.35
N4	0.30	0.15
N5	0.40	0.50
N6	0.20	0.30

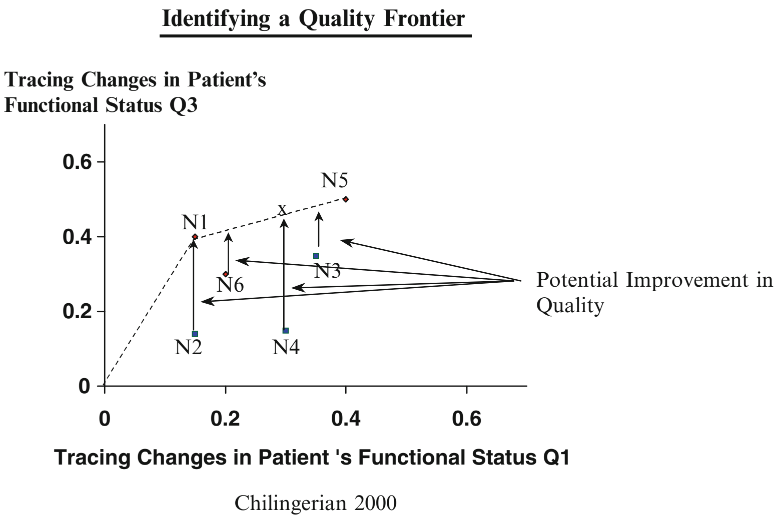


Fig. 16.10 Two-dimensional picture of a quality frontier

Although DEA could be applied to any type of quality evaluation, the following example is from a nursing home, demonstrating how to measure improvements in functional status. Drawing heavily on a paper by Chilingerian (2000), the following illustration explains how DEA works in estimating outcome frontiers. Consider a group of six nursing homes each with 100 patients whose bed mobility, eating, and toilet use have been traced for a period of 6 months.

Table 16.1 displays the overall functional status of the 100 patients in each of the six homes from quarter 1 to quarter 3. The functional status represents a vector of improvements or deteriorations of functional status in terms of the proportion of residents in quarter 1 completely independent and the changes to that group in quarter 3. That situation can be depicted graphically in Fig. 16.10 as a piecewise linear envelopment surface.

In Fig. 16.10, DEA identifies nursing homes that had the greatest improvement in functional quality over the 6-month period by allowing a line from the point of origin to connect the extreme observations in a "northwesterly" direction. Nursing homes N1 and N5 had the greatest improvement in the functional independence of their patient populations and the dotted lines connecting N1 and N5 represent the best practice frontier.

The vertical lines above N2, N6, N4, and N3 represent the potential improvements in quality outcomes if N2, N4, N6, and N3 had performed as well as a point on the line between N1 and N5. Note that since the initial functional status is a nondiscretionary variable, the potential improvement can only occur along the vertical line. Associated with each underperforming nursing home is an optimal comparison point on the frontier that is a convex combination of the nursing homes. For example, N4 could be projected onto the best practice frontier at point *X*. The performance measure is the linear distance from the frontier expressed as a percent such as 90, 84% and so on. To be rated 100%, the nursing homes must be on the best practice surface.

In this simple two-dimensional example, the nursing homes with the greatest improvements in functional status were identified. DEA is capable of assessing quality with dozens of indicators in *n*-dimensional space. For a complete mathematical explanation, see Chap. 1 in this book.

A model that could be used to introduce an outcome-quality frontier is called the additive model (described in Chap. 1). The additive to be used is based on the idea of subtracting the functional status of the patient at the outset from the status attained after the care process. The resulting measure is expressed as a change in functional status based on the (functional status achieved during quarter 3) minus (the functional status at quarter 1). The numerical differences from Q_1 to Q_3 can be interpreted as improvements, deteriorations, or no change in outcomes.

The optimal value $w_{o,}^{*}$ is a rating that measures the distance that a particular nursing home being rated lies from the frontier. A separate linear programming model is run for each nursing home (or unit) whose outcomes are to be assessed. The additive model is shown below:

$$
\begin{aligned}
&\text{(additive model)}\\
&\max_{u,v,uo}\ w_o = u^{\mathrm{T}}\mathrm{Y_o} - v^{\mathrm{T}}\mathrm{X_o} + \mathrm{u}_o\\
&\text{s.t.}\, u^{\mathrm{T}}\mathrm{Y} - v^{\mathrm{T}}\mathrm{X} + \mathrm{u}_o\vec{\mathbf{1}} \leq 0\\
&-u^{\mathrm{T}} \leq -\vec{\mathbf{1}}\\
&-v^{\mathrm{T}} \prec -\vec{\mathbf{1}}
\end{aligned}
$$

Here the vector Y represents the observed functional status variables at Q_3 and X represents the initial functional status variables observed at Q_1. The additive model subtracts the functional status variable at Q_3 from those at Q_1. The variables, u^{T} and v^{T} are the weights assigned by the linear program so $u^{\mathrm{T}}\mathrm{Y}$ is the weighted functional status at Q_3 and $v^{\mathrm{T}}\mathrm{X}$ is the weighted functional status at Q_1. The model is constrained so that every nursing home is included in the optimization such that the range of scores will be between 0 and 100%. The u^{T} and v^{T} are constrained to be nonnegative.

For the problem of developing a frontier measure of improvements or deterioration in functional status, the additive has number of advantages (see Cooper et al. 2000b). The model does not focus on a proportional reduction of inputs or an augmentation of outputs. It offers a global measure of a distance from a frontier, by giving an "equal focus on functional status before and after by maximizing the functional improvement between time periods (characterized as the difference between Q_1 and Q_3). Another advantage of the additive model is that it is translation invariant which means we can add a vector to the inputs and outputs and though we get a new dataset, the estimates of best practice and the outcome measures will be the same. This model is applied to a nursing home dataset to find a DEA outcome frontier and a DEA decision-making efficiency frontier.

16.5.1 A Field Test: Combining Outcome Frontiers and Efficiency Frontiers

To illustrate the ideas discussed above, the following DEA model will be used to find a quality-outcome frontier for 476 nursing homes in Massachusetts (the USA). We developed outcome measures for the nursing homes from the Management Minutes Questionnaire (MMQ). This is the case-mix reimbursement tool used in several states in the USA and in particular is used in the state of Massachusetts to pay nursing homes for the services they provide Medicaid residents. The MMQ collects information on the level of assistance that nursing home residents need from staff members to carry out activities of daily living such as dressing, eating, and moving about. Fries (1990) explains that the MMQ index is constructed for each resident based on a spectrum of resident characteristics, each with a specified weight. Values are supposed to correspond to actual nursing times so the total should correspond to total staffing needs for the resident. Weights are derived from expert opinion rather than statistical analysis, and total weights are adjusted with time values added for each of the items measured.

Changes in overall resident functioning (determined by measuring the change in MMQ scores over two quarters) were used as a proxy for quality of care. These variables depict the direction of functional status change (improvement, maintenance or decline) experienced by the residents during the last 6 months. Changes in a positive or static direction (improvement or maintenance) will be used as proxies for high-quality care (controlling for health status) and changes in a negative direction (decline) will be used as a proxy for a decrease in the quality of care.

In each of the nursing homes, the residents' functional status was evaluated. The proportion of residents who were independent in mobility and eating, and

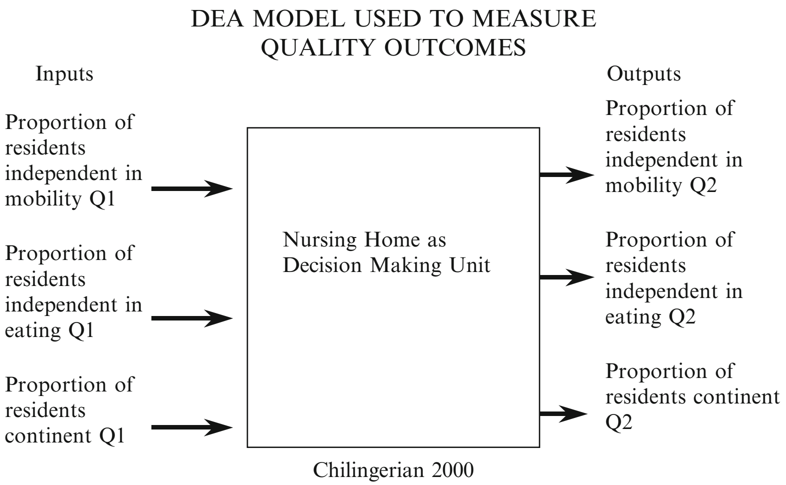


Fig. 16.11 Three-input–three-output model used to estimate quality

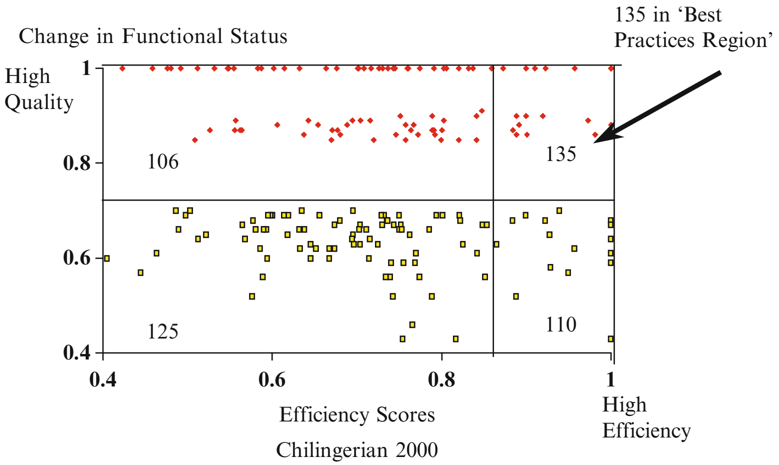


Fig. 16.12 Plot of change in functional status and productive efficiency (source: Chilingerian 2000)

continence were traced for 6 months from quarter 1 to quarter 3. If we consider how a patient's functional status changes over time, whether a nursing home is improving, maintaining, or declining, these changes become an outcome measurement tool. In particular, resident functional improvement was monitored by three activities of daily living: bed mobility, eating, and toilet use. The model used and the variables are displayed in Fig. 16.11 and 16.12.

The DEA model identified 41 or 9% of the nursing homes on the best practice quality frontier. The average quality score was 74%, which means that the functional status outcomes could potentially be improved by 26%.

Figure 16.12 plots the DEA measure of the changes in functional status against a DEA measure of the decision-making efficiency for the 476 nursing homes in the dataset. By partitioning this two-dimensional summary of nursing home performance at the means (81 and 74%), four categories of performers emerge: 135 with high quality outcome and high efficiency; 106 with high quality outcomes but lower efficiency; 110 with low quality outcomes but high efficiency; and 125 with lower quality outcomes and low efficiency. Each of these unique and exhaustive categories can be further analyzed to explore the factors associated with each category of performance. By studying the interaction of these two dimensions of performance, it may be possible to gain new insights into quality and efficiency.

16.6 A Health DEA Application Procedure: Eight Steps

Several studies have suggested analytic procedures for DEA studies (Golany and Roll 1989; Lewin and Minton 1986). Drawing heavily on these papers, this last section highlights how the researcher connects the empirical and conceptual domains with real world problems in health policy and management. The structure of DEA research can be seen as a sequence of the following eight applied research steps:

16.6.1 Step 1: Identification of Interesting Health-Care Problem and Research Objectives

Applied work always begins with a real-life policy and/or management problem. Interesting health policy and management research questions are needed to guide the data to be collected. If we take a close look at all of the DEA studies, many have been illustrations of uses of DEA based on available data. Therefore, the first step is to find questions of practical importance, and never let the available data drive the research. Identify what is new and interesting about the study.

Set challenging research goals – an interesting question is always preferable to a researchable question from an available dataset. The goal should always be to inform both theory and practice. The purpose of a good research question is to get to some answers.

16.6.2 Step 2: Conceptual Model of the Medical Care Production Process

There is no final word on how to frame a health-care production model. Each DEA study claims to capture some part of reality. Hence, the critical question in this step is – Does the production model make sense? Most likely, the problem has been studied before. So an obvious first step is to ask – How has it been modeled in previous work? What aspects were missing? There should always be a justification of the inputs and outputs selected. This can be based on the literature, prior knowledge, and/or expert knowledge (see O'Neill 1998; Chilingerian and Sherman 1997). The use of clinical experts is critical if the application is to become useful for practice.

A related question is choice of the DEA models. Because of strong intuitive appeal, the CCR (clinical and scale efficiency) and BCC (pure clinical efficiency) models have been used in most health-care studies. The multiplicative model (for Cobb–Douglas production functions) and the additive model hardly appear in the health-care literature. Researchers should consider the underlying production technology and take a fresh look at the choice of DEA models. If the justification for the final choice of a DEA model is sketchy, consider running more than one.

16.6.3 Step 3: Conceptual Map of Factors Influencing Care Production

Step three identifies the set of variables and some empirical measures based on several important questions. What is the theory of clinical production or successful performance? Do medical practices vary because some patients are sicker, poorer, or socially excluded? What explains best practices? For this step, researcher should identify the environmental and other factors out of the control of the managers, organizational design and managerial factors, provider and patient characteristics and case-mix variables (see Chilingerian and Glavin 1994). The goal is to build a conceptual map that identifies some of the obstacles, or explanatory variables associated with best practices from the literature and expert knowledge. If the theory is weak, try some simple maps or frames that raise theoretical issues.

16.6.4 Step 4: Selection of Factors

Now the researcher is ready to search for databases or collect the data. There is always some difficulty obtaining the variables for the study files: inputs, outputs, controls, explanatory variables, and the like. Sometimes physician, hospital, medical association, and insurance databases must be merged into one study file to link DEA with other variables.

16.6.5 Step 5: Analyze Factors Using Statistical Methods

DEA assumes that a model is assessing the efficiency of "comparable units," not product differences. Before running an efficiency analysis, if there is reason to believe that outputs are heterogeneous, it is recommended that peer groups be developed (Golany and Roll 1989). In health care, a variety of peer groups could be developed based on medical subspecialty (orthopedic surgeons versus cardiac surgeons), diagnostic complexity, and other product differences.

The mathematics of DEA assumes that there is an isotonic (order preserving) relation between inputs and outputs. An increase in an input should not result in a decrease in an output. Inputs should be correlated. Golany and Roll (1989) have argued that running a series of regressions (variable by variable) can help to reduce these problems. If there is multicollinearity among inputs (or among outputs), one remedy is to eliminate one or more inputs or outputs.

Finally, there has rarely been a DEA study that has not had to deal with the problem of zeroes. In any case, there are two crude ways of dealing with problem. One is to throw out all of the DMUs with missing values and reduce the number of DMUs. The other way is to substitute the zero with a very small number such as 0.001. In addition to this handbook, Charnes et al. (1991) offer better ways of dealing with the problem of zeros in the dataset.

16.6.6 Step 6: Run Several DEA Models

Are the results reasonable? If the dataset finds a threefold difference in costs among the units, something other than efficiency is being measured. If you are running different DEA models check for the stability of results.

Sometimes there are many self-referring units on the best practice frontier. One solution to this problem is to impose cone ratio conditions that reflect preferred practice styles (see Chilingerian and Sherman 1997; Ozcan 1998). As a rule of thumb, if the majority of the DMUs are showing up as 100% efficient and they are mostly self-referencing, there are two possible explanations. Perhaps the production technology is so complex that there are many slightly different ways of practicing. On the contrary, the free choice of weights is giving the providers the "benefit of the doubt" by hiding unacceptable practice styles that care for very few patients, and/or utilize a very high quantity of clinical resources.

16.6.7 Step 7: Analyze DEA Scores with Statistical Methods

The next step is to use the DEA results to test hypotheses about inefficiency. Blending DEA with various statistical methods has been all the rage in health-care studies. This has been a health trend. To convince the reader that the DEA scores are valid, every health-care study must test the hypothesis that there is unaccounted case mix. If the explanatory and control variables have no significant association with the DEA scores, something may be wrong with the production model.

16.6.8 Step 8: Share Results with Practitioners and Write It Up

Traditionally the DEA work in health care focused more on methodology than the issue of usefulness. To advance the field more quickly, research needs to spend time with clinicians and practitioners (who are not our students). Before writing up the results show the models and findings to real clinical managers and listen to their advice about performance effectiveness. This is a suggestion for everyone.

16.7 DEA Health Applications: Do's and Don'ts

Given the 20-year history we have with DEA health applications, there are several do's and don'ts. Based on our experience on reviewing and reading DEA papers – three are discussed.

16.7.1 Almost Never Include Physicians As a Labor Input

One of the concerns in DEA hospital studies is including physicians as an input in health care (Burgess et al. 1998; Chilingerian and Sherman 1990). In many health-care systems such as those of the USA, Belgium, Switzerland, there are academic medical centers that have salaried physicians reporting to a medical director in the hospital; there are also community general hospitals that have established cooperative, as opposed to hierarchical relations with physicians. Though physicians are granted "admitting privileges" and can use the hospital as a workshop to care for patients, some physicians may not be very active in admitting patients. Physicians can enjoy these privileges at several hospitals (Burns et al. 1994). Including a variable for the quantity of FTE physicians is legitimate input if they are a salaried such as in academic medical centers. However, in many countries community hospitals do not employ physicians so they should not be included as an input.

16.7.2 *Use Caution When Modeling Intermediate and Final Hospital Outputs*

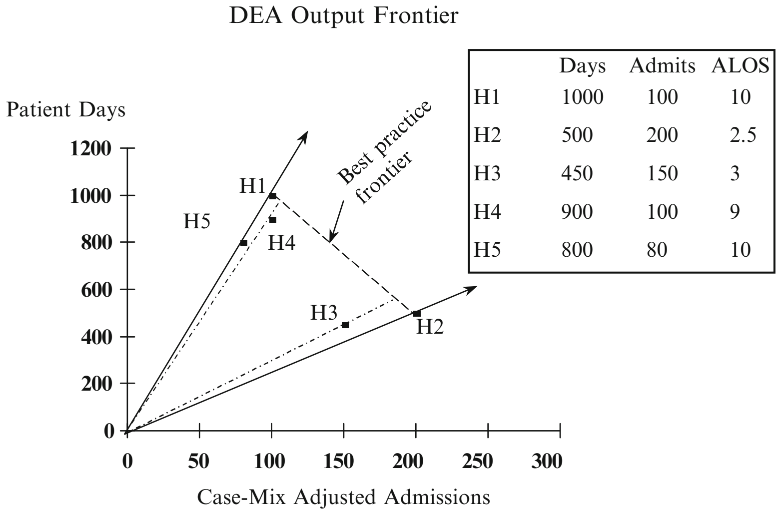
Although modeling hospitals with DEA can be complex, the DEA literature offers some exemplars for research purposes. While none of these papers are perfect, the models are reasonable given the available data. For example:

- Nunamaker (1983) used three outputs: age-adjusted days, routine days, maternity days.
- Sherman (1984) used three outputs: age-adjusted patient days, and nurses and interns trained as outputs.
- Banker et al. (1986) used three outputs: age-adjusted patient days
- Sexton et al. (1989) used six categories of workload-weighted units (for example, medical workload-weighted units (WWU), psychiatric WWU, surgical WWU, OPD WWU, etc.)
- Chilingerian and Sherman (1990) used two outputs: high severity cases of DRG 127 with satisfactory outcomes and low severity cases of DRG 127 with satisfactory outcomes.

Some of these papers focus on "inputs per patient day," while others do on "inputs per discharged case." Note that none of these papers mixed final products, i.e., discharges with intermediate products, i.e., patient days. If both discharges and patient days are the outputs, DEA will obtain results based on a composite of optimal – cost per day/cost per case/cost per visit. DEA will give misleading results.

There have been several papers published in the health-care literature that combine intermediate and final outputs in a single model. While there are arguments on both sides, there are conceptual problems with how DEA would evaluate efficiency using these models. Let us consider the model used by several researchers (see Burgess and Wilson 1996; Ferrier and Valdanis 1996) combining case-mix-adjusted discharges with total patient days. Assuming constant inputs, if a hospital maintains its case-mix-adjusted admissions, while increasing its patient days, its average length of stay (ALOS) is increasing. The DEA model, however, would rate the hospital with higher length of stay (LOS) as more efficient, though these hospitals would, by most managerial definitions, be less efficient because they could have discharged some patients sooner, and *utilized their capacity better* by admitting *more* patients. In a high fixed cost service, longer lengths of stay might suggest poorer quality (more morbidity, nosocomial infections, etc.) Since quality measures are not available, one cannot assume maximum output from inputs with constant quality outcomes.

Now, consider the alternative model with case-mix-adjusted days and teaching outputs, excluding discharges (see Sherman 1994). Assuming constant inputs, if a hospital maintains the number of interns being taught and increases its case-mix-adjusted patient days, the hospital is producing more outputs with fewer inputs – i.e., becoming more efficient. That is, DEA would rate that hospital



	Days	Admits	ALOS
H1	1000	100	10
H2	500	200	2.5
H3	450	150	3
H4	900	100	9
H5	800	80	10

Fig. 16.13 Two-dimensional picture of an output frontier that combines intermediate and final hospital outputs

more efficient. Although the ALOS is not known, the Sherman model is not rating hospitals with longer lengths of stay as better. This type of model works better conceptually and in practice.

To illustrate the point, consider the following data from a policy maker's perspective. In Fig. 16.13 five teaching hospitals, each with *similar inputs* and the same case mix, and the *same total ancillary charges*, *but with variable patient days and case-mix-adjusted admissions*. In the plot of the four hospitals, the DEA output frontier identifies the best practice frontier to be the segment connecting H1 and H2. H4, for example, is less efficient than H1 because it had fewer patient days, even though it discharged the patients sooner and presumably used its available capacity better.

H5 is inefficient and would become more efficient if H5 could move along the ray (represented by the solid lines) and toward H1. The efficiency rating is based on a radial measure of efficiency that maintains the rate of transformation. This measure forces H5 to increase days and admissions, and ignores the need to improve clinical efficiency – e.g., by reducing length of stay and increasing admissions. Do the definitions of efficiency underlying this model make sense?

H4 has the same admissions but fewer patient days; it has the same cost per case, but a higher cost per patient day than H1 and a lower ALOS. From a policy perspective, should not H4 be ranked at least as efficient as H1? DEA suggests that H4 becomes efficient by maintaining its ALOS and increasing days and admissions accordingly.

Now, when H2 is thought of as an extreme point, policy makers might consider H1 as the only frontier point. H2, 3, 4, and 5 all could improve their capacity utilizations (by treating more cases) and improve their clinical efficiency (by reducing lengths of stay). There are DEA models that define preferred practice regions (or cone ratios) to handle these situations.

Therefore, if the model includes *both* patient days and cases as DEA outputs (as the author's have done), DEA defines two type of efficiency – one based on cost per case, and another based on cost per admission. H1 is rated 100% because it produced more of the variable output – patient days. H2 is rated 100% efficient because it produced more admissions.

Hospitals with the same case mix and admissions, but higher lengths of stay would be rated as 100% efficient along side hospitals with the same case mix, patient days, and lower lengths of stay. This runs counter to what policy makers would want and would confuse everyone. There are two solutions. One is to use the managerial inputs with the clinical outputs (see Fig. 16.4). Another approach is to use a two-part production model: one DEA for practice management with managerial inputs and intermediate outputs, and another DEA model for patient management with clinical inputs and clinical outputs (see Fig. 16.1).

16.7.3 Do Check the Distribution of DEA Scores and Influence of Best Practice Providers on Reference Sets

DEA will yield some approximation of an efficiency score with most datasets. Consequently, every health-care application requires a careful check of the distribution of DEA scores. For example, whenever the range includes efficiency scores below 0.50, or whenever there are more than 50 or 60% of the DMUs on the frontier, there is a strong likelihood that something is wrong. The following common sense test appears to be helpful – Given the health-care context, do these results make sense? Though one can hardly imagine finding a hospital, nursing home, or physician operating at 17% efficiency and surviving another day, it is possible. Low scores always require a plausible explanation, which might be a noncompetitive environment, or the existence of subsidies. Newhouse points out that Zuckerman et al. (1994) found productive inefficiency to account for 14% of total hospital costs (1994). Should policy makers reduce hospital budgets by this amount? Would this force many hospitals out of business? Before anyone takes this findings too seriously, the inefficiencies have to be valid.

In health care, a high proportion of very low DEA scores are likely to be due to unaccounted case mix, heterogeneity of output measures, or returns to scale (Chilingerian 1995). Alternatively, a very high proportion of DMUs on the best practice frontier (40–50%+) could be the result of caring for many different types of patients, using slightly different styles of practice. It is important to examine the number of physicians appearing in only one reference set and to identify the most influential physicians – those physicians who appear in the most reference sets. When a priori information is available on best practices, it is possible to reduce the number of efficiency candidates in any given analysis. For example, a cone ratio DEA model allows for a meaningful upper and lower-bound restriction to be placed

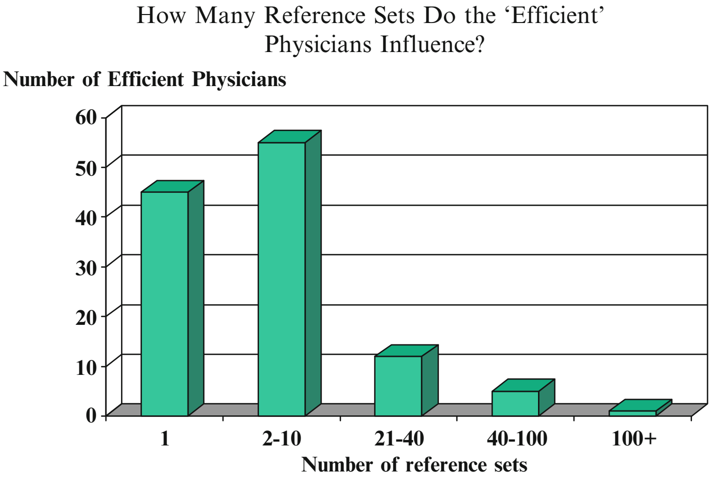


Fig. 16.14 Number of efficient physicians by the number of reference sets they influence

on each input virtual multiplier. The restriction reduces the number of efficient candidates, bringing the DEA measure of technical efficiency closer to a measure of overall efficiency. Imposing cone ratio models on the results will usually reduce the number of self-referencing DMUs on the frontier (see Chilingerian and Sherman 1997; Charnes et al. 1990). Running a second DEA model without the most influential observation will reveal how robust the DEA results.

For example, one DEA study of 326 primary care physicians conducted by Chilingerian and Sherman (1997) found 138 on the best practice frontier. In Fig. 16.14, 45 physicians appear in one reference set and are, therefore, self-referencing physicians. Expert opinion from the clinical director helped to identify a preferred practice region. The medical director's criterion for "best practice" was a primary care physician who (1) performed under budget, (2) utilized less than 369 medical/surgical days per 1,000 members, (3) had referral rates of less than 1.2 per member, and (4) provided at least 1,760 primary care office visits per 1,000 members. A second DEA model would be run setting the minimum and maximum marginal rates of substitution for referral rates, medical surgical hospital days, and primary care visits. The second DEA model reduced the number of efficient physicians from 138 to 85 and found most of the 45 self-referring physicians less efficient.

16.8 A Final Word

This chapter looks at some of the conceptual and methodological challenges associated with measuring and evaluating health-care performance with DEA. While DEA offers many advantages, the critics have raised fundamental questions such as the following: How useful is DEA? What purpose does a DEA study serve? (Newhouse 1994).

There are several purposes for conducting DEA studies. One is to develop better descriptions and analyses about practice patterns and styles. Insights into most optimal caseloads, effects of severity on scale inefficiency, or other sources of inefficiency are helpful to health-care managers.

A second purpose for undertaking a health application with DEA is to identify best practices to create insights and new ideas that explain the successes and failures of clinical providers for policy makers. Examining why some provider firms succeed while others fail and identifying the sources of performance improvement remains an important societal problem, especially if the sources of failure are rooted in the payment and financing systems. Finally, once best practices are identified, health-care mangers need help in finding ways to reduce waste in the utilization of clinical resources to achieve health goals. This strategic objective ranks high on many health managers' agenda.

While the DEA toolbox has a potential to serve these goals, the work has been more illustration than practicable and theory developing. The body of DEA research that has accumulated is substantial. Is there anyway to order and account for the panorama of information? Is there an acceptable general DEA model or subset of models for health and hospital studies? Have DEA applications improved the delivery of care for patients? At this time, the answer to all of these questions is "No." There appears to be neither a taxonomy nor a general single model capable of handling all of the issues that critics can raise. Most distressing is the fact that very few DEA applications in health care have documented any significant improvements to quality, efficiency, or access.

Since the early 1990s, the pace of DEA research has been rapid. Has the information been mined? The problems inherent in deepening understanding of health-care operations and performance today are more intricate than those encountered when DEA research began "exploring and testing" DEA applications to understand health-care performance problems.

The main problem has been the lack of rigor in developing models. Simple measurement errors are less of a problem than selecting patently different input and output variables. To prevent the critics from challenging the models and to help advance policy and management, models should be based on good theoretical ideas. Otherwise, there is a stalemate in the theoretical development of health applications.

This chapter has uncovered research and managerial opportunities and challenges and noted that there is a great deal of work that remains to be done. Health care has a goldmine of clinical information and medical records that document cost, quality, and access. On the one hand, every single diagnosis and illness could be studied with DEA at the physician/patient level. On the other hand, in a single general hospital, there are the more than 5,000 distinct products and services – no single DEA model will ever be able to analyze all of those activities.

In 1996, Seiford asked the experts to nominate "novel" DEA application. Four health-care papers were included among the handful of innovative studies. Every one of these DEA studies focused on individual physicians as DMUs.

This remains a very promising area, since physicians are not only an important provider of care but also the principal decision-makers and entrepreneurs for the care programs and clinical DMUs.

A few years ago Scott (1979) proposed a multi-output-multi-input measure of clinical efficiency that divided the amount of improvement for each type of patient by the cost of each hospitalization (input). Measuring the change in severity of illness from admission to discharge would make a DEA efficiency measure clinically relevant by capturing a more complete picture of a physician's clinical performance relative to the resources used.

Finally, Farrell (1957) has argued that activity analysis can be used to evaluate the productive efficiency of various economic levels ranging from small workshops to an entire industry. Is DEA the right tool to motivate a general model of performance (quality, access, and productive efficiency) in health care? Clearly, DEA is proving to be an effective approach to evaluate individual efficiency as well as organizational efficiency; consequently, DEA could be used to evaluate quality and efficiency at all four levels of the health are industry:

1. The individual patient's experience
2. Individual physician level
3. The department or organizational level
4. The entire industry level

As information systems become more integrated and more available, and as DEA evolves from mere illustration to real health services research, future studies could attempt to connect individual patient changes in health status to physician efficiency, to department efficiency, to overall hospital efficiency, and then connect hospital efficiency to the efficiency of the entire hospital industry.

Acknowledgments We are very grateful to Dianne Chilingerian and W.W. Cooper for their encouragement and thoughtful comments. Part of the material for this chapter was adapted from Chilingerian, J. A., 1995, Evaluating physician efficiency in hospitals: A multivariate analysis of best practices, European Journal of Operational Research 80, 548–574; Chilingerian, J.A. 2000, Evaluating quality outcomes against best practice: a new frontier. The Quality Imperative: Measurement and Management of Quality. Imperial College Press, London, England; and Chilingerian, J. A., and H. David Sherman, 1997, DEA and primary care physician report cards: Deriving preferred practice cones from managed care service concepts and operating strategies, Annals of Operations Research 73, 35–66.

References

Anderson GF, Reinhardt UE, Hussey PS, Petrosyan V. It's the prices, stupid: why the United States is so different from other countries. Health Aff. 2003;22(3):89–105.

Banker RD. Estimating most productive scale size using data envelopment analysis. Eur J Oper Res. 1984;17:35–44.

Banker RD, Conrad R, Strauss R. A comparative application of data envelopment analysis and translog methods: an illustrative study of hospital production. Manage Sci. 1986;32(1):30–44.

Banker R, Charnes A, Cooper WW. Some models for estimating technical and scale inefficiencies in data envelopment analysis. Manage Sci. 1984;30:1078–92.

Banker RD, Morey RC. Efficiency analysis for exogenously fixed inputs and outputs. Oper Res. 1986a;34(4):513–21.

Banker RD, Morey RC. The use of categorical variables in data envelopment analysis. Manage Sci. 1986b;32(12):1613–27.

Bristol Royal Infirmary Inquiry Final Report, The Report of the public inquiry into the children's heart surgery at the Bristol Royal Infirmary, Learning from Bristol. Presented to Parliament by the Secretary of State for Health by command of her Majesty; 2002; London, England: Crown.

Bryce CL, Engberg JB, Wholey DR. Comparing the agreement among alternative models in evaluating HMO efficiency. Health Serv Res. 2000;35:509–28.

Burgess JF, Wilson PW. Hospital ownership and technical efficiency. Manage Sci. 1996;42 (1):110–23.

Burgess J, James F, Wilson PW. Variation in inefficiency among U.S. hospitals. INFOR. 1998;36:84–102.

Burns LR, Chilingerian JA, Wholey DR. The effect of physician practice organization on efficient utilization of hospital resources. Health Serv Res. 1994;29(5):583–603.

Caves RE. Industrial efficiency in six nations. Cambridge: MIT Press; 1992.

Charnes A, Cooper WW, Henderson A. An introduction to linear programming. New York: Wiley; 1953.

Charnes A, Cooper WW, Rhodes E. Measuring the efficiency of decision-making units. Eur J Oper Res. 1978;3:429–44.

Charnes A, Cooper WW, Thrall R. A structure for classifying and characterizing efficiencies in data envelopment analysis. J Prod Anal. 1991;2:197–237.

Charnes A, Cooper WW, Lewin AY, Seiford LM. Basic DEA models, data envelopment analysis: theory, methodology, and application. Boston: Kluwer Academic Publishers; 1994.

Charnes A, Cooper WW, Sun DB, Huang ZM. Polyhedral cone-ratio DEA models with an illustrative application to large commercial banks. J Econom. 1990;46:73–91.

Chilingerian JA. Investigating non-medical factors associated with the technical efficiency of physicians in the provision of hospital services: a pilot study. Best Paper Proceedings, Annual Meeting of the Academy of Management; 1989; Washington, D.C

Chilingerian JA. Evaluating physician efficiency in hospitals: a multivariate analysis of best practices. Eur J Oper Res. 1995;80:548–74.

Chilingerian JA, Glavin M. Temporary firms in community hospitals: elements of a managerial theory of clinical efficiency. Med Care Res Rev. 1994;51(3):289–335.

Chilingerian JA, Sherman HD. Managing physician efficiency and effectiveness in providing hospital services. Health Serv Manage Res. 1990;3(1):3–15.

Chilingerian JA, David Sherman H. Benchmarking physician practice patterns with DEA: a multi-stage approach for cost containment. Ann Oper Res. 1996;67:83–116.

Chilingerian JA, David Sherman H. DEA and primary care physician report cards: deriving preferred practice cones from managed care service concepts and operating strategies. Ann Oper Res. 1997;73:35–66.

Chilingerian JA. Evaluating quality outcomes against best practice: a new frontier. The quality imperative: measurement and management of quality. London: Imperial College Press; 2000.

Chilingerian JA, Glavin M, Bhalotra S. Using DEA to profile cardiac surgeon efficiency. Draft of Technical Report to AHRQ; 2002; Heller School, Brandeis University.

Cooper WW, Park KS, Pastor JT. Marginal rates and elasticities of substitution with additive models in DEA. J Prod Anal. 2000a;13:105–23.

Cooper WW, Seiford LM, Tone K. Data envelopment analysis: a comprehensive text with models, applications, references and DEA-solver software. Boston: Kluwer Academic Publishers; 2000b.

Fair R. A theory of extramarital affairs. J Polit Econ. 1978;86:45–61.

Fare R, Grosskopf S, Lindgren B, Roos P. Productivity developments in Swedish pharmacies: a malmquist output index approach, data envelopment analysis: theory, methodology, and applications. Boston: Kluwer Academic Publishers; 1994.

Farrell MJ. The measurement of productive efficiency. J R Stat Soc Ser A. 1957;120:253–81.

Ferrier GD, Valdanis V. Rural hospital performance and its correlates. J Prod Anal. 1996;7:63–80.

Fetter RB, Freeman J. Diagnostic related groups: product line management with in hospitals. Acad Manage Rev. 1986;11(1):41–54.

Finkler MD, Wirtschafter DD. Cost-effectiveness and data envelopment analysis. Health Care Manage Rev. 1993;18(3):81–9.

Fizel JL, Nunnikhoven TS. Technical efficiency of nursing home chains. Appl Econ. 1993;25:49–55.

Fries BE. Comparing case-mix systems for nursing home payment. Health Care Financ Rev. 1990;11:103–14.

Georgopoulos BS. Organizational problem solving effectiveness: a comparative study of hospital emergency services. San Francisco: Jossey-Bass Publishers; 1986.

Golany B, Roll Y. An application procedure for DEA. Omega. 1989;17(3):237–50.

Greene WH. Econometric analysis. 2nd ed. New York: Macmillan Publishing Company; 1983.

Harris JE. The internal organization of hospitals: some economic implications. Bell J Econ. 1977;8:467–82.

Harris JE. Regulation and internal control in hospitals. Bull N Y Acad Med. 1979;55(1):88–103.

Hao S, Pegels CC. Evaluating relative efficiencies of veteran's affairs medical centers using data envelopment, ratio, and multiple regression analysis. J Med Syst. 1994;18:55–67.

Hollingsworth B, Dawson PJ, Maniadakis N. Efficiency measurement of health care: a review of non-parametric methods and applications. Health Care Manage Sci. 1999;2(3):161–72.

Kerr C, Glass JC, McCallion GM, McKillop DG. Best-practice measures of resource utilization for hospitals: a useful complement in performance assessment. Public Adm. 1999;77(3):639–50.

Kooreman P. Nursing home care in The Netherlands: a non[parametric efficiency analysis. J Health Econ. 1994;13:301–16.

Lapetina EM, Armstrong EM. Preventing errors in the outpatient setting. Health Aff. 2002;21 (4):26–39.

Leibenstein H, Maital S. Empirical estimation and partitioning of x-inefficiency: a data envelopment approach. Am J Econ. 1992;82(2):428–33.

Lewin AY, Minton JW. Determining organizational effectiveness: another look and an agenda for research. Manage Sci. 1986;32(5):514–38.

Lynch J, Ozcan Y. Hospital closure: an efficiency analysis. Hosp Health Serv Adm. 1994;39 (2):205–12.

Luke RD, Ozcan YA. Local markets and systems: hospital consolidations in metropolitan areas. Health Serv Res. 1995;30(4):555–76.

Maddala GS. Limited dependent and qualitative variables in econometrics. New York: Cambridge University Press; 1983.

Morey D, Ozcan Y. Med Care. 1995;5:531–52.

Newhouse JP. Frontier estimation: how useful a tool for health economics? J Health Econ. 1994;13:317–22.

Newhouse J. Why is there a quality chasm? Health Aff. 2002;21(4):13–25.

Neuman BR, Suver JD, Zelman WN. Financial management: concepts and applications for health care providers. Maryland: National Health Publishing; 1988.

Nunamaker T. Measuring routine nursing service efficiency: a comparison of cost per day and data envelopment analysis models. Health Serv Res. 1983;18(2 Pt 1):183–205.

Nyman JA, Bricker DL. Profit incentives and technical efficiency in the production of nursing home care. Rev Econ Stat. 1989;56:586–94.

Nyman JA, Bricker DL, Link D. Technical efficiency in nursing homes. Med Care. 1990;28:541–51.

Ozcan Y. Physician benchmarking: measuring variation in practice behavior in treatment of otitis media. Health Care Manage Sci. 1998;1:5–18.
Ozcan Y, Lynch J. Rural hospital closures: an inquiry into efficiency. Adv Health Econ Health Serv Res. 1992;13:205–24.
Ozcan YA, Begun JW, McKinney MM. Benchmarking organ procurement organizations: a national study. Health Serv Res. 1999;34:855.
Ozcan YA, Luke RD. A national study of the efficiency of hospitals in urban markets. Health Serv Res. 1992;27:719–39.
Ozgen H, Ozcan YA. A national study of efficiency for dialysis centers: an examination of market competition and facility characteristics for production of multiple dialysis outputs. Health Serv Res. 2002;37:711–22.
Perez ED. Regional variations in VAMC's operating efficiency. J Med Syst. 1992;16(5):207–13.
Puig-Junoy J. Technical efficiency in the clinical management of critically ill patients. Health Econ. 1998;7:263–77.
Rosko MD. Measuring technical efficiency in health care organizations. J Med Syst. 1990;14 (5):307–22.
Rosko MD, Chilingerian JA, Zinn JS, Aaronson WE. The effects of ownership, operating environment, and strategic choices on nursing home efficiency. Med Care. 1995;33 (10):1001–21.
Scott RW. Measuring outputs in hospitals, measuring and interpreting productivity. Nat Res Counc. 1979;255–75.
Seiford L. Data envelopment analysis: the satt6e of the art. J Prod Anal. 1996;7(2/3):99–138.
Sexton T, Leiken A, Nolan A, Liss S, Hogan A, Silkman R. Evaluating managerial efficiency of veterans administration medical centers using data envelopment analysis. Med Care. 1989;27 (12):1175–88.
Sherman HD. Hospital efficiency measurement and evaluation. Med Care. 1984;22(10):922–8.
Shortell S, Kaluzny A. Health care management. 4th ed. New York: Delmar Publishers Inc.; 2000.
Teboul J. Le temps des services: une nouvelle approche de management. Paris: Editions d'Organisation; 2002.
Witte A. Estimating an economic model of crime with individual data. Q J Econ. 1980;94:57–84.
Zhu J. Further discussion on linear production functions and DEA. Eur J Oper Res. 2000;127:611–8.
Zuckerman S, Hadley J, Iezzoni L. Measuring hospital efficiency with frontier cost functions. J Health Econ. 1994;13:255–80.