
3. Literature Review

Contents

	Page
Introduction.	39
Methodology and Review Articles: Conveying Principles and Practice to the Health Care Community.	40
Defining the Problem, Objectives, and Alternatives.	41
Assessing Production Relationships	43
Identifying, Measuring, and Valuing Benefits/Effectiveness and Costs	45
Benefits/Effectiveness.	45
costs * . \$ * * **.*....	48
Valuing Benefits and Costs Over Time: Discounting.	50
Addressing Problems of Uncertainty: Sensitivity Analysis	50
Examining Ethical Issues.	52
Presenting and Interpreting Findings	52
Linking Analysis to Policy Implementation	54
Conclusion.	54

INTRODUCTION

Application of cost-effectiveness analysis/cost-benefit analysis (CEA/CBA) to health care represents a fairly new endeavor. The novelty of CEA/CBA within medicine has both positive and negative reflections. On the one hand, the growth in the literature demonstrates considerable enthusiasm for the technique; perhaps more importantly, it may indicate increasing cost consciousness within the medical community. Further, it might signal an increased awareness of the idea that a comparison of costs and benefits has always been an implicit value of health care decisionmaking—and that to allow resource allocation decisions to continue to be made, perhaps unwittingly, without more explicit consideration of the costs in relation to benefits is not desirable. On the other hand, the enthusiasm for CEA/CBA is often undisciplined, perhaps reflecting the inevitable growing pains of any field of inquiry. The vast majority of literature contributions whose titles identify them as related to CEA/CBA have serious technical flaws or conceptual weaknesses in structure or interpretation.

Also included in the literature are many important exceptions to this general assessment. Several studies exhibit both the desired technical features and the potential to lend insight into important issues of health resource allocation. Through such examples, as well as direct “instruction” (e. g., articles that review methods, cited below), a small cadre of skilled practitioners of CEA/CBA seems to be providing the intellectual leadership to improve the general quality of the literature and advance the state of the art. This group includes both physicians and economists, and several of the recent exemplary studies have resulted from multidisciplinary collaboration (e.g., 472,575).

Two contextual aspects of the evaluation presented in this chapter warrant emphasis at the outset. First, many of the limitations of health

care CEA/CBAs are endemic to—and, more importantly, inherent in—almost all CEA/CBAs. For example, the inability of most health care CEA/CBAs to incorporate distributional considerations (177,179) is shared by CEA/CBAs on education, defense, energy, transportation, and so on. This chapter attempts to identify generic CEA/CBA problems and to distinguish them from problems that are specific to the health care literature. The chapter also distinguishes problems that are resolvable from those that are inherent in the process of analysis.

Second, literature reviews often restrict their attention to the most prominent articles and books in the literature, as is the case in the earlier reviews of health care CEA/CBA identified in appendix A. There is a logic to this approach: These publications reflect and indeed create the state of the art; because they are widely read, they have an influence on professional thinking and on future contributions to the literature disproportionate to their numbers. Nevertheless, such publications constitute only the most visible portion of the literature. The 10, 20, or 30 articles repeatedly cited in health care CEA/CBA reviews represent considerably less than a tenth of the publications that can be readily identified as part of this literature, often by explicit inclusion of the words “cost-benefit” or “cost-effectiveness” in their titles. A few such publications are clearly mislabeled; many others are on the right track but are so “poorly” handled that a CEA/CBA purist might exclude them from a CEA/CBA bibliography. Even though the less well-known publications have a readership and general influence far smaller than that of the more prominent contributions, however, they may constitute the principal exposure of many practicing health professionals to the language, concepts, and applications of CEA/CBA.

In order to capture the essence of what CEA/CBA means to health professionals, it is impor-

tant to critique the entirety of the literature. Thus, this review represents an attempt to integrate typical-practice and state of the art features of the literature. The basis of the review is an assessment of general tendencies in the literature as a whole, including the 90 percent that to a large extent has escaped attention in previous reviews. Common problems and deficiencies are frequently counterbalanced by reference to suc-

cessful attempts to address the deficiencies. Thus, while this review adopts a generally critical stance toward the literature, it acknowledges the many examples of technical proficiency in the practice of health care CEA/CBA. For a review that concentrates on the high-quality end of the spectrum, the reader is referred to the recent paper by Weinstein (569).

METHODOLOGY AND REVIEW ARTICLES: CONVEYING PRINCIPLES AND PRACTICE TO THE HEALTH CARE COMMUNITY

All good studies inform readers about CEA/CBA methods, either implicitly through its application (e.g., 122) or explicitly through discussion of methodological issues (e.g., 148). Recent books have served several functions: attempting to convey basic principles, break new methodological ground, and demonstrate the utility of CEA/CBA through specific applications (73, 516). Over the years, however, there has been only a handful of articles written solely to present or to evaluate the state of the art in health care CEA/CBA.

During the period studied (1966-78), the first two such articles were published in the first 2 years of the period. In 1966, Crystal and Brewster (722) wrote an introduction to CEA/CBA in the health field. In 1967, Klarman published the first of two prominent reviews he has written, this one appearing in the *American Journal of Public Health* (295). From then until 1972, no significant health care review or methodology contributions appeared in print, with the exception of a chapter by Grosse (241) in a book oriented toward students of economics and policy analysis. That chapter is particularly noteworthy for its review of CEA/CBA applications in the Department of Health, Education, and Welfare (HEW) during the author's tenure as an HEW official. Grosse conveyed much of the same material 2 years later in an article published in 1972 (240), though again the audience was not specifically health care professionals.

That same year, however, witnessed publication of a book that has become one of the health

care community's most widely read and frequently cited contributions, Cochrane's *Effectiveness and Efficiency: Random Reflections on Health Services* (97). This short book appears to have had a profound and sustained impact in turning the thoughts of health care professionals toward issues of resource scarcity and the link between efficiency and equity. It is at least possible that Cochrane's book played a significant role in the rapid growth in health care CEA/CBA that began in 1973.

The most often-cited review and discussion of health care CEA/CBA is Klarman's second article (291), which appeared in 1974. The following year, Dunlop (153) published a review that characterized the early literature, thus providing some interesting contrasts with current CEA/CBA practice. In 1975, the major portion of an issue of the *New England Journal of Medicine* (NEJM) was devoted to a discussion of CEA/CBA methodology and to several illustrations of its application. To many observers, this NEJM issue stands as a landmark in the evaluation of medical practice. It should be noted that none of the previous reviews had been published in a medical journal.

Two years later, another issue of NEJM offered readers a package of two articles and an editorial, including a discussion of CEA methodology (574), a sophisticated application of it (516), and an important, thoughtful treatment of the limitations of formal analysis, including the potential for a "tyranny of numbers" and associated disregard of equity considerations

(177). Many health services researchers consider this package, combined with Weinstein and Stason's book (575), to be a milestone in health care CEA/CBA.

A recent review was presented at an Urban Institute Conference on Medical Technology in December 1978 (569). This paper offered a state-of-the-art assessment of the literature and a review of "a nonrandom sample" of health care CEA/CBAs. The most noteworthy feature of

the paper is the author's discussion of remaining methodological issues. Although several of the issues have been of concern since the inception of formal CEA/CBA, others represent subtle, sophisticated problems, the existence of which is testimony to progress on more basic issues. Indeed, the paper serves as a vivid reminder simultaneously of the frustrating, seemingly intractable problems of CEA/CBA, and of the gradual yielding of some of them to sustained conceptual and empirical struggle.

DEFINING THE PROBLEM, OBJECTIVES, AND ALTERNATIVES

In a review of the early health care CBA literature, Dunlop (153) concluded that the most common use of CBA had been to analyze disease-specific programs of intervention. By beginning with disease problems, several early analyses had the opportunity to explore a wide range of alternative interventions. For example, the interest of LeSourd, et al. (327) in identifying efficient means of grappling with kidney disease led these investigators to compare the costs and benefits of a variety of programs ranging from prevention of disease to treatment of renal failure. Similarly, Acton (4) employed both CBA and CEA to examine several alternative prehospital programs for reducing deaths due to myocardial infarction. Even in studies of narrower problems—for example, the treatment of existing disease—comparison of alternatives characterized much of the early analyses. Thus, the focus of Klarman, et al. (298) *On* kidney disease treatment precluded consideration of prevention alternatives, but the authors examined all of the major therapeutic alternatives.

Most of the contributions to the early literature shared a focus on a problem and specific objectives that had a distinct health (or disease) starting point. At the extreme, HEW analysts used CEA and CBA to examine resource allocation across a wide variety of disease and accident control programs (240,241). More narrow problem definitions implied fewer and less disparate alternatives, but the health relevance of the objective was generally clear. Thus Weisbred's (577) examination of the costs and bene-

fits of medical research was restricted to the case of polio, but the analysis centered on the health consequences of polio research and consequent prevention of the disease.

In recent years, there have been numerous attempts to use CEA/CBA to analyze programs having clear health relevance (e.g., 122,472, 473,573), but two factors seem to be increasing the proportion of studies whose health relevance is implicit, tangential, or simply unclear. The first is a tendency to assume that certain programmatic outcomes are desirable, without questioning their ultimate health implications. Common examples are studies that conclude that certain screening efforts are "cost effective" because they are inexpensive, but that lack any exploration of the costs and health effectiveness of followup of the cases found (133). The second is a technical factor and reflects the current importance of cost containment as a health policy issue. Much analysis has moved from a focus on promoting health toward concern with achieving efficiency in the provision of existing health services, including particularly a group of intermediate medical services whose ultimate health impact cannot be discerned. In this case, the question becomes how to provide a medically accepted service most efficiently, without being able to evaluate definitively (if at all) its health implications.

Illustrative of this phenomenon is the CEA/CBA literature on computed tomography (CT) scanning, the single most studied technology of

the last 2 years examined in OTA's bibliographic search (1977-78). As observed in chapter 1, the CT scanner is the only expensive, equipment-embodied technology to have been the subject of considerable CEA/CBA attention, but it may be in the vanguard in this respect, rather than being an exception. Furthermore, the CT scanner exemplifies the difficulties involved in evaluating diagnostic procedures (360,558,559), an area that has been identified as deserving of much greater CEA/CBA effort (569). Despite sharing the same technology—the scanner—as a starting point, the authors of the numerous CEA/CBAs diverge significantly in their perceptions of the objectives of scanning and hence in their evaluations of its cost effectiveness. At one extreme, the diagnostic effectiveness of scanning is assumed, with no attempt to link diagnosis to either patient management or outcome; cost effectiveness is measured as the cost savings from using the CT scanner, as opposed to alternative techniques, to perform a given volume of diagnoses (211). At the other extreme, effectiveness is defined in terms of effects in disease management and patient outcome (28). The latter seems the socially most desirable concept of effectiveness, but the problems in its determination are substantial, and it misses additional benefits such as those associated with decreasing patients' uncertainty, directing short-term patient management, and contributing to greater medical understanding (2,32). Needless to say, the differing objectives result in widely varying assessments of the desirability of scanning.

Determining objectives for purposes of analysis is frequently regarded as a trivial exercise, but examples from the literature illustrate that it may require considerable thought and that the absence of such effort can damage the quality of analysis. A failure to appreciate the limits of a selected objective can mislead both analyst and decisionmaker. For example, when HEW analysts decided to compare the cost effectiveness of alternative disease control programs, they selected lives saved as the measure of effectiveness. This variable (and variants on it, such as life-years or quality-adjusted life years (QALYs) saved) is a common measure of health impact in CEAs, but it is not a comprehensive one. HEW

analysts recognized this by observing that an arthritis control program could never be justified on the basis of lives saved, but that the program ranked as one of the better investments when the benefits associated with disability avoided were taken into account in a CBA framework (240). Fairly few health care CEAs make explicit reference to the nature of the biases their effectiveness measures introduce. It may be that many such biases have not even occurred to the analysts.

Fein (177) has noted the tendency of health care CEA/CBAs to "relegate to a footnote" a variety of nonquantifiable benefits (e. g., caring, or reduction of pain). With a reference to both intangible benefits and distributional effects that cannot be valued, he noted that as "the numbers gain currency . . . a 'climate of opinion' is created: that which is measured is important and vice versa." This problem is common to analysis in virtually all areas, though its importance is probably greatest in the social welfare fields such as health, education, and justice. In some cases, though, benefits that are difficult to quantify or value have escaped even footnoting. For example, in an otherwise sophisticated CEA comparing hysterectomy and tubal ligation as sterilization alternatives, Deane and Ulene (134) ignored the preferences of the women involved. The authors carefully analyzed the direct costs of the procedures and indirect costs of complications and later disease, but the emotional reactions and feelings that might be expected to dominate many women's decisions received no consideration in their analysis. Problems of measuring and valuing intangible benefits pervade the health care CEA/CBA literature. They are exacerbated by the failure of many analysts to identify such benefits in specifying objectives.

In recent years, there appears to have been a narrowing of problem definition in health care CEA/CBA. Accompanying this has been a reduction in the number and scope of alternatives examined through CEA/CBAs. The extreme—an analysis of a single program or procedure, with the only "alternative" being its absence—has become reasonably common in the literature. Another development, exemplified by the analyses of Eddy (157,158) and Schoenbaum, et

a]. (472,473), represents an intermediate position between a single-program analysis and a comparison of numerous qualitatively diverse alternatives: Several analysts are using mathematical techniques to design or determine the optimal (i. e., most cost-effective) structure of a program by analyzing the effects of changes in several parameters and assumptions (e. g., com-

pliance rates, diagnostic accuracy, therapeutic effectiveness). In essence, such analysts are examining a large number of “programs” of a single type. Even though confining analysis to a single program type implies limitations, this approach holds the promise of making significant contributions to policy understanding and program development.

ASSESSING PRODUCTION RELATIONSHIPS

Technical aspects of analysis clearly differentiate high-quality analyses from the more typical contributions. Nowhere is this more evident than in the modeling of production relationships. A summary characterization of the difference is this: The better studies carefully consider and address production issues, whereas the typical contributions adopt a “black box” approach to production (that is, they observe existing programs’ inputs and outputs and ignore current inefficiencies and predictable future changes).

In part, this difference is legitimate, reflecting the diverse purposes of analyses. Many of the better studies have a prospective, or planning, intent. The studies are oriented toward predicting the costs and benefits or effectiveness of alternative future programs, so the analysts model idealized versions of these future programs, recognizing significant variations from current similar programs when such exist (e. g., scale, efficiency, relative costs, technological change). By contrast, many of the more typical analyses have a retrospective or evaluative purpose. In these the analysts wish to assess the performance of a program in terms of its realized costs and outcomes. Even for this type of assessment, however, it is usually important to examine the black box of production so as not to attribute to inputs outcomes that occur by chance.

When the purpose is retrospective evaluation, the identification and measurement of experienced inputs and outcomes are appropriate. Often, however, it appears that authors who have a prospective planning objective in mind have not thought through the limitations of ex post evaluation. Implicit in their analyses is the assumption that existing programs are accurate

models of the alternative futures. Occasionally this may be reasonable, but the assumption is fraught with hazards. Common errors in the literature include:

- failure to account for scale effects, i.e., taking an existing program (e. g., a worksite hypertension screening and treatment program) and assuming that a national program intended to serve (say) 1,000 times as many people will require 1,000 times as many inputs (i. e., costs) and produce 1,000 times the output;
- failure to consider how environmental factors might alter program inputs and outcomes (e. g., assuming that the production function for an urban worksite hypertension screening and treatment program would serve as a valid model for planning a nationwide screening and treatment effort both inside and outside of work settings);
- ignoring predictable technical changes over time (e.g., assessing the “future” of CT scanning, assuming that the technology — and hence inputs and outcomes—will not change from what exists at present);
- ignoring predictable increases over time in the efficiency of operation of a technology or program, i.e., the “learning curve” phenomenon (e. g., assuming that the performance of program personnel will not improve over time as they gain familiarity and experience);
- ignoring likely shifts in the relative costs of inputs (e. g., the price—wage rate—of labor rising more rapidly than the price of equipment) and consequent changes in the mix of inputs used;

- failing to identify avoidable inefficiencies in the existing “model” program and thus inputting them to the structure and operation of future programs; such inefficiencies could reflect an inefficient input mix (e. g., having a high-priced technician perform a function which could be automated inexpensively) or an inappropriate use of technology (e. g., condemning CT scanning as exorbitantly expensive because it is used indiscriminately, rather than limiting uses to those which are medically and economically justifiable); and
- conversely, failing to anticipate that both the inputs and outcomes of a carefully monitored program in a major medical center may not be replicated as the program diffuses into general practice.

Most of these deficiencies of *ex post* evaluation for prospective planning become more severe the more novel the technology or program in question. A familiar, established, and successful program is more likely to represent a good model for planning purposes than is a new, or, especially, an experimental program. Yet a major role of a forward-looking CEA/CBA ought to be to assess the potential costs and benefits (effectiveness) of a program before it has diffused throughout the medical system.

CT serves as an excellent example of the great difficulties of undertaking useful analysis early to influence planning and decisionmaking. These difficulties span the spectrum of applications of CEA/CBA, but they are particularly severe in an area such as medicine in which technological change occurs rapidly and frequently. It was exceedingly difficult to perform an adequate analysis of CT scanning prior to its diffusion. Yet all of the studies in the literature relied on that early experience for data, and most of the early studies failed to anticipate changes that have already occurred, only a few years following publication of the studies. Furthermore, anticipated changes in radiological technology may make CT scanning technically obsolete within a few years, yet the nature and amount of relevant information are not adequate to incorporate this factor into an analysis intended

to assist planning. The CEA/CBA literature on the CT scanner does not address this issue.

Formal modeling is difficult, a simple fact that may account for the lack in much of the literature of imaginative, useful characterization of production relationships. At a minimum, modeling requires talent in disciplined conceptualization; frequently, it also necessitates application of specific mathematical or formal modeling skills. The latter, in particular, are not available in abundant supply. Medical education generally includes no consideration of such skills, and few analysts with appropriate training from other disciplinary backgrounds have devoted their attention to health care CEA/CBA issues. There are, of course, notable exceptions. By example, through methodological contributions, and by direct discussion of issues, numerous authors improve both the current and future state of the art of assessing production relationships. Review articles have communicated basic principles, improving the critical abilities of readers and, one would hope authors of future studies. Specific CEAs/CBAs in the literature have illustrated skillful conceptualization, use of mathematics, and formal modeling techniques (e.g., 22,122,134,157,159,479,480,516).

Mathematics and formal modeling can intimidate, impress, and confuse the uninitiated. In order to put the formalism into proper perspective, it is imperative that authors clarify the implications of both explicit and implicit assumptions in the modeling and emphasize the limitations of their studies. There is a significant risk that the uninitiated will be overly impressed with formalism, so the caveats should be more than simple disclaimers. Yet only a minority of health care CEAs/CBAs have taken this requirement seriously (e.g., 122,516).

High-quality analysis of production relationships does not require sophisticated modeling efforts. A few studies have exhibited both elegant conceptualization and structural simplicity. For example, in their analysis of the national swine flu immunization program, Schoenbaum, et al. (473) considered the effects of varying acceptance rates, probability of an epidem-

ic, and other factors in a manner that was technically sound and readily understandable. Particularly in the medical literature, which is read by an audience that generally is mathematically

unsophisticated, the clarity of studies such as this one probably serves to educate and to build interest in well thought out CEA/CBAS.

IDENTIFYING, MEASURING, AND VALUING BENEFITS/EFFECTIVENESS AND COSTS

Benefits/Effectiveness

A central concern of many health care CEA/CBAs, both conceptually and empirically, is adequately capturing the health consequences of programs. Only one such consequence lends itself to unassailable, objective measurement: reductions in mortality. Another common, if not universally accepted, measure of health improvement, however, is reduced days of morbidity or disability. Neither of these measures accounts for variations in the quality of the resulting days of less impaired health. Analysts have adopted a few **means of adjusting for** this quality factor, but to date there has been nothing approaching consensus on specific methods of adjustment.

Analysts' inability to quantify satisfactorily certain health benefits appears to be the primary reason for their exclusion from formal calculations. A second reason is the difficulty of identifying what it is that patients seek and receive from health care. As noted above, comprehensive analyses of the tangible costs and benefits of treatment alternatives have sometimes ignored the emotional or psychological motivations that may lead patients to prefer one treatment over another, as if the patients were merely inputs into a physical production function (134). In other words, patients' objectives and values are not limited to measurable physical health improvement, and if patients' objectives do not represent social concerns, the very reason for considering a health program is challenged.¹

¹There are instances in which the patients' objectives and values may be considered irrelevant, or at least secondary, to society's values. Care of the severely mentally ill patient represents an extreme example. Externalities and paternalism provide more common justifications. An example is a requirement that children receive certain immunizations prior to enrolling in school.

Obviously, the significance for an individual CEA/CBA of the inability to quantify certain benefits depends on the relative importance of those benefits in the program under consideration. Certain health problems present seemingly insurmountable barriers to the objective measurement of their benefits, yet their importance has prompted analysts to grapple with them in a CEA/CBA framework. Examples include mental retardation (99), mental illness (463), and care of the terminally ill. Without succeeding in quantifying the intangibles, the efforts of analysts to deal with problems such as these have contributed to an increased understanding of the nature of the problems and the associated programs.

For many health programs, the principal health benefits are the more tangible, or quantifiable, reductions in mortality, morbidity, and disability. Nevertheless, CEA/CBA assessment of benefits (effectiveness) is far from problem free. How does one measure and value benefits (effectiveness) in units that are commensurable with each other or with costs? Days of morbidity avoided, for example, are not directly comparable to days of mortality avoided.

In the health care CEA/CBA literature, there are three principal approaches to this problem: 1) accepting it as an unresolvable problem, and selecting a single (presumably dominant) outcome as the index of benefit or effectiveness; implicit in this approach is the assumption, or hope, that nonmeasured benefits vary proportionately and positively with the single outcome measure; 2) employing an index of health effects or of health status; and 3) adopting one of two methods of valuing major outcomes in monetary terms. The first two of these provide effectiveness measures for CEAs, while the third

yields the monetary benefit measurement needed for CBAs. A fourth approach, rarely found in the literature, would be to not force an aggregation of effects or benefits. A possible “array” method is discussed in chapter 4.

The first approach—the most common one in the literature—is the easiest to accomplish and perhaps to understand. It is also, however, the least conceptually appealing, because of its unsatisfactory (often implicit) assumption that decreases in mortality, for example, correlate highly with decreases in such factors as morbidity, pain, and suffering. A prominent example, noted earlier, comes from the mid-1960’s HEW disease control program analysis in which “lives saved” served as the proxy for all health benefits in the CEA comparison of programs. As the analysts observed, “lives saved” as the effectiveness measure relegated arthritis to the bottom of the list of cost-effective programs. When the programs were compared by means of cost-benefit calculations, however, the ability to reduce arthritis-related morbidity and disability made the arthritis control program appear quite competitive with the programs that saved the most lives (240).

The single-measure index of effectiveness continues to dominate health care CEAs, but modifications point the way toward more refined measures of health benefits. “Lives saved” is a gross but important index of effectiveness for many health programs. “Life-years saved” adds an element of quality to the nature of deaths averted.² This measure has been employed in several CEAs. A further refinement involves adjusting the life-years to reflect the quality of those years. Klarman, et al. (298) provided an early example of quality adjustment in their CEA study of alternative renal disease treatments. They argued that a year of life with a well-functioning transplanted kidney was superior to a year of life on dialysis—given the time, inconvenience, and discomfort associated with the latter. Consequently, they arbitrarily valued

a year of life on dialysis as equal to 0.75 year with a transplanted kidney.

The idea of quality-adjusting provided the basis for Weinstein and Stason’s (575) use of an index of health effects in their study of hypertension screening and treatment programs. Their QALY involves adding changes in life expectancy to changes in quality-adjusted life expectancy resulting from reduction in morbidity, and subtracting changes in quality-adjusted life expectancy due to iatrogenic illness and treatment-induced side effects. Selection of appropriate weights remains arbitrary and hence a problem. An earlier attempt to develop a health status index produced weights (ranging from 0 to 1) that corresponded to a spectrum from death to complete health (726). Despite the conceptual appeal of such an index, the inherent weighting problems plus the so far inadequate empirical data base have led to very limited CEA/CBA application of this type of index (79).

The third approach to valuing benefits in commensurable units is to translate all quantifiable outcomes into monetary terms—benefit measurement for CBA. A common approach, monetary assessment of benefits is also the most controversial approach in the evaluation of health care programs. The principal issue, as discussed in chapter 2, is the valuation of human life. As indicated in chapter 2, the human capital approach employs a market measure of the value of life, whereas willingness-to-pay asks how people value their own lives, subject to their ability to back up their valuations with economic resources. Willingness-to-pay has considerable conceptual appeal, but to date no one has succeeded in developing techniques to produce consistent and meaningful estimates of willingness-to-pay (4). The human capital approach has its own conceptual attributes, but with its imputation that the worth of a life is determined solely by productivity, it has fallen into disfavor among many practitioners of health care CEA/CBA. Since the human capital approach is empirically more manageable and consistent, however, the vast majority of CBAs have employed this form of valuation. This is not to suggest, however, that the approach invariably has been applied correctly. Benefits

²“Life-years saved” is not clearly preferable to “lives saved.” Everyone would agree that more years saved per death averted is preferable to fewer (other things being equal); but is 10-years saved for one person preferable to 4 years saved for each of two people? The answer is inherently subjective.

should be measured as the costs of illness avoided. Some analysts have used existing cost-of-illness estimates as direct measures of benefit, without recognizing that many of the illnesses avoided would have occurred years into the future and hence that benefits should have been discounted. This has had the effect of inflating benefit estimates, in some cases considerably (309).

In any given study, beyond the choice of a basic approach to measuring benefits lies determination of the specific measure(s). In CEAs, the effectiveness measure is often reasonably obvious, with different analysts selecting similar measures, thereby facilitating cross-study comparisons. Treatment of kidney failure provides an example in which different analysts have selected the same measure of effectiveness—life-years saved—and despite a difference of 10 years in publication dates, their analyses have produced consistent results (298,513). For some topics, however, effectiveness measures are less obvious, with the result that different investigators have selected qualitatively distinctive measures and undertaken analyses that produced quite different and not directly comparable results. The problem seems especially relevant to the area of diagnosis (360). Its presence in the literature can be anticipated to grow if analysis of diagnostic procedures increases, as some observers believe it should (569). Resolution of the problem, if possible, may lie in imaginative efforts to translate diagnostic accuracy into effects on patient management and health outcome. Among the CT papers, only Baker and Way's (28) attempted to do this. Their scaling of effects involved arbitrary and subjective judgments, but Baker and Way's effort stands out as one of the few published attempts to bridge the diagnosis-health outcome gap.

The literature includes few examples of such efforts to grapple with challenging assessment problems. Nevertheless, other approaches have been adopted. A few studies identify and array noncomparable measures of effects, including rank-ordered ones (148). The argument underlying this approach is that if effects are important but cannot be measured in a common metric, decisionmakers will find it more useful and

less misleading to see them arrayed in an "unfinished" CEA (i.e., one lacking a "bottom line" cost-effectiveness ratio) than to have one or more of them dropped for the sake of calculating a "final" cost-effectiveness ratio. Despite its "incompleteness," the CEA by Doherty, et al. (148), for example, contributed information and structure which can facilitate understanding of a policy issue.

Consistent definition of effectiveness can vary across health care functions. For example, mental retardation illustrates a substantive health problem for which assessments of prevention v. treatment necessarily involve quite different, noncomparable measures of effectiveness. Prevention of retardation (e. g., through phenylketonuria (PKU) screening), is commonly valued in a cost-benefit framework for its ability to avoid expenses of institutionalization and other care by preventing the birth of retarded children; that is, the benefits of the program are future costs avoided (e.g., 78,79,482). By contrast, many of the desired effects of programs providing care for an existing group of the retarded are less tangible and less economically oriented; the "costs avoided" metric is clearly inadequate (99). Obviously, the prevention-treatment effectiveness distinction is by no means universally applicable. Analysts have successfully relied on a consistent outcome measure in comparisons of prevention and treatment alternatives for kidney disease (327), myocardial infarction (4,122), and others.

In addition to addressing all of the problems noted above, analysts must identify and acquire data needed to measure benefits. The quality of data is rarely examined carefully in health care CEA/CBAs, yet it is a common constraining factor across most studies. Obviously, assessment of the health outcomes of a variety of procedures depends on the availability of valid, reliable experimental or epidemiological data; such data do not always exist, and even when they do, they are not always accessible. Benefit valuation for CBA requires in addition that such data be translated into their economic implications. The methods for doing this are conceptually clear, and solid empirical assessment of the costs of illness has been performed (385,721,

747). Nevertheless, there are significant variations from one study to the next, implying that use of differing estimation procedures and data could compound spurious variations in benefit estimates introduced by the use of different health-outcome data sources. The issue of the reliability and validity of cost-of-illness estimates is currently under study in a contract funded by the National Center for Health Services Research. The National Institutes of Health is in the process of publishing a bibliography of some 2,000 references relating to cost-of-illness estimation.

Data for several health care CEA/CBAs have relied on subjective rankings, surveys, and the like (4,28,78,79). The issue of how valid and reliable such data can be has yet to be resolved, though several studies find considerable variation depending on how survey questions are phrased (4,550). Clearly, conceptual and empirical work on benefit assessment measures is a pressing need in health care CEA/CBA (569).

Analysts in the United States have access to more numerous and varied data sources than do investigators in many other countries. Compared to the ideal, however, even U.S. data sources exhibit serious deficiencies. Many surveys are plagued by poor methods, producing unreliable data that contribute to misleading analyses (565,740). The national data collection effort is hindered by considerable duplication of effort and inconsistency across data sets, both resulting in part from a lack of interagency coordination.³ The development of better organized and planned basic data collection is essential to improving the quality of health care CEA/CBA.

Two remaining benefit assessment issues are noted here, with discussion deferred to later sections. One is the practice of discounting benefits, occasionally handled well in the literature, frequently ignored. The other is analysts' handling of distributional or equity concerns.

costs

Apart from the problems of measuring and valuating benefits/effectiveness, the literature reveals numerous examples of poor or inaccurate measurement and valuation of costs. Deficiencies associated with cost assessment are frequently more insidious than those associated with benefit assessment, because authors commonly devote less attraction to them. Since authors often do not discuss cost analysis problems, they fail to alert readers to them; furthermore, the analysts themselves in many cases seem unaware of the deficiencies of their approach, data sources, etc.

Costs are a reflection of resources consumed. Thus, many of the difficulties that have plagued cost assessment are perfectly analogous to those discussed above in the examination of analysts' handling of production relationships. Rather than repeat that discussion, this section simply notes several common problems: 1) often analysts have measured realized (ex post) costs in an analysis intended for prospective planning without allowing for learning, technical, and economic changes which seem likely to occur; 2) they have failed to distinguish the cost implications of running programs under optimal v. average conditions; 3) they have not always accounted for the differential valuation of costs occurring at different points of time (the discounting problem, discussed in the next section).

Just as there are problems unique to benefit assessment (e.g., valuation of lives saved), certain problems hinder cost analysis in particular. Chief among these in the health care literature is the use of inaccurate or inadequate proxies for true costs—a significant problem because of its pervasiveness and, evidently, the failure of many investigators even to be aware of it. A major source of inaccuracy is the use of market prices as measures of costs. The assumption that prices closely mirror true costs seems reasonable in some smoothly functioning markets, but market imperfections can distort the relationship between input prices and their true opportunity costs. This common problem reaches its extreme—and hence introduces the most significant distortions—in cost assessment in health

³This problem is discussed in some detail in *Selected Topics in Federal Health Statistics (740)*, a report prepared by OTA in June 1979.

care programs, particularly those associated with hospitals. In health care CEA/CBAs, the use of prices in lieu of true opportunity costs generally means adoption of published charges (e.g., from hospital billings or insurance charges) as the index of cost. Although occasionally analysts recognize that charges may not accurately mirror costs, very often the problem is not even acknowledged (92). The vast majority of health care CEA/CBAs employ charges uncritically, frequently introducing potentially large errors in the estimation of the true costs of the programs in question. At a minimum, analysts ought to explore the relationships between charges and actual market costs.

Inadequate cost assessment often results from failure to take into account costs which are real but hidden. For example, very few health care CEA/CBAs account explicitly for the costs of patients' time traveling to medical facilities and waiting for and receiving services. CBAs occasionally capture some of this by valuing lost productivity, but, most commonly, lost productivity measurement relates only to days of morbidity, disability, or mortality avoided, and not to hours involved in seeking and receiving care; and lost productivity is not the only time cost associated with health care services.⁴ This problem is exacerbated, however, by the fact that many employees are covered (e. g., through sick leave) for time off from work for medical visits. Thus, neither the physician nor the patient perceives the time as "lost," and analysts sensitive to the time-cost issues might overlook the fact that the time imposes real costs on society (e.g., physical productivity lost).

⁴That people value such lost time is demonstrated by the willingness of many individuals to accept significant charges from private physicians in lieu of waiting a long time in lower cost medical clinics. The waiting-time mechanism of rationing medical services is highly inefficient socially, producing a "deadweight loss," that is, patients lose their free time, and no one gains directly from that loss.

A second example of real costs that have escaped attention in health care CEA/CBAs is the value of volunteers' time mentioned in chapter 2.

An unresolved cost assessment issue is whether analysts ought to assume efficiency in program operation or build in "slack" for likely inefficiencies. The former is appropriate for evaluating the ideal, but the latter seems more likely to reflect what will come to pass should the program be implemented. This issue has received virtually no attention in the empirical literature. Common practice has been to measure resources used in programs, rather than to identify efficient resource use, but only a few studies suggest that the investigators have even contemplated the difference.

A technical cost issue of considerable importance derives directly from the discussion in an earlier section of the relative lack of attempts by analysts to distinguish marginal from average resource consumption. Most commonly, authors have used average total costs of existing programs to predict the costs of program expansion, modification, etc. When capital costs are substantial or marginal costs vary significantly, failure to distinguish marginal from average costs can produce, and often has produced, misleading cost estimates. Although some analysts have demonstrated sensitivity to the distinction, direct extrapolation from average costs dominates the health care CEA/CBA literature.

As in the case of benefit measurement, data availability and quality problems hinder effective cost analysis. For example, data on charges are relatively accessible, but many cost data (in particular, data needed to reflect opportunity costs) are not. The current interest in cost containment has promoted governmental efforts to acquire more and better cost information, but the acquisition and appropriate use of cost data will linger as a major problem in CEA/CBA for years to come.

VALUING BENEFITS AND COSTS OVER TIME: DISCOUNTING

The discounting of benefits and costs realized over time is one of the most technical features of CEA/CBA. It is also one of the principal sources of analytical weakness in the health care literature. Owing to the potentially profound influence of discounting on valuation of costs and benefits (effectiveness), the absence of discounting in numerous health care analyses severely discredits those analyses. In addition, the failure of some investigators to test the sensitivity of their findings to the value of the discount rate raises questions about the robustness of those findings. (Sensitivity analysis is discussed in the next section of this chapter.)

The literature is replete with examples of the total absence of discounting. In the studies that have employed discounting, the basic method is generally sound; that is, discounting has been done either technically well or not at all. However, only a minority of the studies that have used discounting have included examination of the effects of the value of the discount rate on the bottom line regarding net benefits or cost effectiveness. Yet, as several of these studies demonstrate, when significant realization of costs or benefits occurs well into the future, the discount rate selected and the method of discounting can play pivotal roles in determination of a program's value (513,572). For example, in her study comparing programs to treat or prevent myocardial infarction, Cretin (122) tested the sensitivity of her cost-effectiveness estimates to variations in the discount rate. The prevention program—screening of school-age children for high cholesterol—necessarily involved benefits deferred well into the future. With costs and benefits undiscounted, the net cost per year of life saved ranged from \$2,441 to \$2,855, depending on assumptions. Discounting at 5 per-

cent produced a cost per year of life saved of from \$9,353 to \$12,640. At 10 percent, discounting caused the figures to leap to \$66,660 to \$94,460. These estimates compared with a range of \$1,782 to \$6,100 per year of life saved by treatment alternatives, depending on the program and the discount rate. Cretin's article not only demonstrated the proper application of discounting, but it emphasized the dramatic effect that varying the discount rate can have on net cost estimation and hence on comparison of program alternatives.

A general CEA/CBA discounting question has received attention in the recent health care literature: Should effectiveness measures be discounted? Empirically, the question has been answered in the affirmative by Cretin (122), Stange and Sumner (513), and Weinstein and Stason (573), each of whom discounted effectiveness measures of mortality avoided in the future. The logic of discounting effectiveness is quite appealing, but the practice is fairly novel

⁵Some analysts have argued that society's interests in intergenerational equity, and in the future more generally, imply a social rate of time preference lower than the opportunity cost of capital, and hence that the former (lower) rate should be used to discount "benefits (effectiveness) and the latter (higher) to rate costs. This conceptual issue has not arisen in the empirical studies in the health care CEA/CBA literature. However, it is interesting to note that both individual and public decisions suggest that people often behave in quite the opposite manner. For example, conscious decisions to continue smoking imply a heavy discounting of the future relative to immediate gratification. At the societal level, public decisions to fund renal dialysis, rather than kidney disease research, screening, and prevention suggest a high social rate of time preference. Of course, this logic assumes that, as individuals and a collectivity, we can interpret the abstraction of a future death averted by prevention as the same "commodity" as postponement of the death of a visible individual. Obviously, we cannot do this. Nevertheless, behavior and decisions are far from being consistent with a low social rate of time preference.

ADDRESSING PROBLEMS OF UNCERTAINTY: SENSITIVITY ANALYSIS

The discount rate is only one of numerous potentially significant influences on the magnitude of cost and benefit (effectiveness) estimates. As

noted in chapter 2, it is a rare study that can be carried from conception to empirical conclusion without the necessity of the analyst's making

assumptions to substitute for uncertainties, data unavailability, conceptual problems, and so on. Despite this, it is not common practice in health care CEA/CBAs to test the significance of assumptions. Frequently, analysts do not carefully distinguish between assumptions and sound empirical observations.

Of the possible uses of sensitivity analysis to address uncertainties, only one has been applied with any degree of frequency in the literature: the direct testing of findings to determine if they are sensitive to important assumptions. Even this most common application of sensitivity analysis has been used rather infrequently, and with a few notable exceptions, it has been used primarily for testing sensitivity to discount rates. The ability of sensitivity analysis to determine whether a major uncertainty precludes a definitive analysis does not appear to have encouraged analysts to tackle health care evaluation problems in which such uncertainties were obvious at the outset. Nor have analysts used measurable costs and benefits to establish minimum or maximum values for quantified variables in order for a program to appear worthwhile. However, a few studies have approximated such uses of analysis. Centerwall and Criqui's (86) assessment of thiamine fortification of alcoholic beverages allowed them to avoid valuing health benefits, since net cost savings were positive.

Cretin's (122) testing of the sensitivity of findings to variations in the discount rate illustrates art appropriate use of sensitivity analysis in its most common application. As Cretin's analysis demonstrated, program evaluation is highly sensitive to discounting when significant benefits (or costs) are deferred well into the future, a characteristic of many prevention programs. Discounting the costs of the cholesterol screening program by 10 percent instead of 5 percent increased costs per year of life by over 600 percent; for the treatment alternatives, the benefits of which are more immediate, however, the corresponding increase in costs was on the order of 50 percent. The potential for such dramatic differences explains why "responsible analysts usually offer the user of analysis a sensitivity

analysis with respect to the discount rate used" (569).

The authors of several of the more highly regarded studies have also tested the sensitivity of findings to other uncertainties. For many health care programs, patient acceptance or compliance is both a crucial variable and uncertainty, and hence a *worthy* candidate for sensitivity testing. The literature provides several examples. Schoenbaum, et al. (473) examined the effect of acceptance on the optimal structuring of the national swine flu immunization program. Eddy's (158) analysis of breast cancer screening also related such factors to program design. Weinstein and Stason's (573) study of hypertension control demonstrated how patient compliance can influence the outcomes of a CEA of disease management.

The literature offers only a few examples of sensitivity analysis applied to other cost and benefit (effectiveness) estimations, but those few are instructive. For example, LeSourd, et al. (327) found that the absolute magnitudes of individual benefit-cost ratios of kidney disease control alternatives were quite sensitive to variations in program size, target screening group, etc., but the relative rankings of the major programmatic alternatives (e.g., screening v. treatment, and within the latter, transplantation v. center dialysis v. home dialysis) were unaffected by the tested variations. In addition to testing sensitivity to the discount rate, Cretin (122) included high and low direct cost estimates for the screening program. The analysis demonstrated less sensitivity to the direct cost estimation than to discounting.

The use of sensitivity analysis reflects a more sophisticated appreciation of CEA/CBA than that which characterizes most of the existing health care literature. At one level, inclusion of thoughtful sensitivity analysis multiplies the number of figures in an analysis and can add considerable complexity to the presentation and interpretation of findings. However, both logic and empirical evidence indicate that the assumptions of an analysis can affect results significantly. Thus, both the credibility and useful-

ness of CEA/CBA would be increased by more frequent and judicious use of sensitivity analysis. With the exception of a handful of high-quality studies, the existing health care CEA/CBA literature lacks credibility in part because issues of sensitivity are addressed so rarely.

The use of sensitivity analysis carries with it a risk: A “solid” finding can dissolve under the

scrutiny of sensitivity analysis, and “nonresults” are less exciting and potentially less acceptable than definitive ones. Nevertheless, intellectually and from the perspective of the policymaker, accurate nonresults clearly are preferable to artificially precise ones. The bulk of the existing literature does not allow distinguishing between these.

EXAMINING ETHICAL ISSUES

As noted earlier, one of the limitations of CEA/CBA is its inability to handle distributional issues. Health care CEA/CBAs rarely have grappled seriously with distributional issues. Obviously, selection of a topic to study may be an implicit statement of concern with distributional issues, though it is expressed in terms of the objective of the analysis rather than as an analytical variable in the CEA/CBA (e.g., the relatively large CEA/CBA literature on mental illness or geriatric services). Health care CEA/CBA should not be singled out for its failure to incorporate distributional considerations successfully. This is the general state-of-the-art and perhaps reflects one of the inherent limitations of this form of analysis. But emphasis on this limitation is particularly important in the health care literature where a readership relatively unfamiliar with the technique may be unduly impressed by formalism and its derivative conclusions, failing to place those conclusions in their proper distributional context. Many health care CEA/CBAs identify this concern, but it seems buried by the analysis which follows and thus is frequently ignored (177).

The literature offers few examples of attempts to address the problem of differentially valuing

the costs and benefits accruing to different groups of people directly. Nevertheless, the equity concern most often debated in the literature—the valuation of life—clearly relates to this fundamental problem, for in CBA benefits are estimated according to one’s productivity (the human capital approach) or affluence (the willingness-to-pay approach). Less often recognized is that CEA effectiveness measures presumed to be “value-free” generally imply values. The trend of the literature away from CBA and toward CEA may reflect growing distaste for explicit valuation of life or the belief that both conceptual and empirical limitations make the effort a “quixotic quest for a value of life” (569). Of course, alternatively, or in addition, the growing preference for CEA may simply reflect the fact that CEA is easier to understand and perform.

The appropriate handling of distributional issues remains one of the least developed features of CEA/CBA in the health care literature and elsewhere. Even though both theoretical and empirical progress can be anticipated (569), the major problems of dealing with equity concerns seem unlikely to be resolved in the foreseeable future.

PRESENTING AND INTERPRETING FINDINGS

As emphasized throughout this report, two factors place a major responsibility on analysts to present and interpret their findings carefully and clearly: 1) technical limitations (inherent in analysis or in the abilities of particular analysts)

often seriously restrict the possibility of arriving at unequivocal, definitive conclusions; 2) the readership of health care CEA/CBA is generally unsophisticated about the techniques of this form of analysis, though this situation is chang-

ing. In addition, numerous readers will focus on, if not limit their attention to, the abstracts and conclusions of articles.

An overall assessment of the health care CEA/CBA literature suggests that relatively few analysts have addressed this responsibility successfully. Those few are generally the authors of the studies identified as technically high quality. Of course, the handling of the presentation and interpretation of findings is a characteristic against which the quality of analyses is judged, but it appears that a thoughtful, useful conclusion to an analysis tends to follow a technically and conceptually well-conceived study. Examples abound. The analysis by Schoenbaum, et al. (473) clearly identified factors that could influence the success and optimal structure of the national swine flu immunization program. Cretin's (122) concluding analysis and remarks clarified the crucial role of discounting and demonstrated the need to interpret the sensitivity analysis. Cretin purposefully and constructively made it impossible for the reader to conclude that there was an obvious "best" approach to reducing the toll of myocardial infarctions. Stason and Weinstein (516) discussed how compliance and a variety of other factors could affect their conclusions, though Fein (177) still found it necessary to emphasize limitations. Doherty, et al. (148) emphasized information organization and presentation in their assessment of health programs for the elderly; they refused to "reduce" their analysis to a "bottom line." The authors of all such studies seem to be motivated by "the philosophy that it is not so much the results of a [CBA/CEA] that are likely to have an impact on policy as the process of structuring information in a systematic framework that highlights the key uncertainties and the most important value tradeoffs" (569). This is inevitably reflected in these analysts' presentation and interpretation of their findings.

By contrast, most health care CEA/CBAs seem oriented toward a "bottom line"—generally the estimation of a benefit-cost or cost-effectiveness ratio. Aside from questions of measurement underlying the cost and benefit (effectiveness) components of these ratios, even this basic "bottom line" has been technically misin-

terpreted in numerous studies. At the extreme, at least one article with a title beginning "Cost-Benefit Ratio . . ." does not contain a single cost-benefit ratio. Few analysts exhibit awareness of the deficiencies of a benefit-cost ratio as compared with a measure of net benefits. The benefit-cost ratio clearly dominates in empirical health care CBAs.

Cost-effectiveness ratios, and the words "cost effective," are employed even more uncritically than are benefit-cost ratios in CBAs. In many articles, "cost effective" refers to one of the two words but not both: That is, some authors have employed "cost effective" when they mean that a program or technology is effective, irrespective of cost; and other authors have used "cost effective" to connote "cheap," irrespective of effectiveness. There are several instances of purported CEAs in which only a single program or technology is examined and is then adduced as being cost effective, despite the absence of an alternative against which to compare it (133).

Subtleties of technical interpretation of CEA/CBA "bottom lines" largely have escaped attention in the health care literature. Only a few analyses demonstrated awareness that "cost effectiveness" of a use of a technology need not imply overall cost effectiveness. For example, in certain delivery settings, an automated electrocardiograph (EKG) may be more cost effective than a manually read EKG, but if the ease and availability of the former lead to excessive use, the national EKG bill might actually rise without necessarily contributing to improved health (18). Perhaps the most dramatic demonstration of the difference between average and marginal cost effectiveness was Neuhauser and Lewicki's (397) estimation that the cost per additional case of colon cancer found by repeated stool guaiacs rose from under \$1,200 for the first stool guaiac to \$47 million for the sixth!

Even when the use of a ratio or net benefit measure is technically correct, lost in such a number are the assumptions that underlie it and the intangible unmeasured costs and effectiveness that are excluded from it. A few studies have presented results in a manner that makes these factors clearer. The most common strategy has been careful discussion of how the "bot-

tom line” could be affected by such factors. Alternatively, some authors have presented ranges of results reflecting sensitivity to assumptions (122,327). A third approach, less commonly adopted, has been to step back from the bottom line and provide a tabular display of programs

and their (noncommensurable) effects. This approach does not yield a conclusion as to which of several competing programs is the “best,” but it does array alternative sets of consequences effectively and thereby might aid decisionmakers by clarifying tradeoffs (148).

LINKING ANALYSIS TO POLICY IMPLEMENTATION

In health care studies as elsewhere, the gap between CEA/CBA studies and policy formulation almost invariably has been bridged by a leap of faith that assumes a theoretically desirable program can be translated readily and directly into an operational one. Health care CEA/CBAs always have had a policy orientation, but the literature is nearly devoid of empirical attempts to make the adjustments (needed to reflect political and cost realities) proposed in the new literature linking analysis to policy implementation (336). Health care CEA/CBA perhaps should not be faulted for this lack; the implementation literature is simply too new.

Luft (336) in a contribution to the implementation literature, used two health care examples, development of freestanding surgicenters and use of work evaluation units “for objective testing of functional work capacity to supplement the usual information concerning the health status of patients who have recently had a myocardial infarction.” Through these examples, he demonstrated how role players’ differing interests can block implementation of socially desirable programs, and how analysts can use recognition of differing interests and influences in developing predictive CBAs. Empirical application of this important conceptual contribution might increase the realism and usefulness of CEA/CBAs.

Though not formally employing Luft’s approach, a few studies have observed how interests might be expected to block or inhibit implementation of socially desirable programs. In a recent study which concluded that in certain large delivery settings, automated electrocardiography may represent a cost-effective alternative to traditional manual readings, for example, the investigators suggested that diffusion of this technology might be inhibited by cardiologists, to whom it could represent a threat to reading fees and a change in referral patterns (18).

Another proposal that can be found in the literature is that analyses should build in a consideration of the sensitivity of basic findings to unanticipated cost overruns. We are not aware of any health care CEA/CBAs that have done this, but there are numerous examples of studies in which it might have been done. One example is the study by Schoenbaum, et al. (473). These analysts could not have anticipated the Guillain-Barre syndrome—and the resultant costs—which accompanied the national swine flu immunization program. The program experienced significant additional production and distribution costs, however, and the analysts might have explored the implications of unanticipated cost overruns. Analyses of future proposed public health programs might consider doing so.

CONCLUSION

The assessment in this chapter of the quality of the health care CEA/CBA literature has relied primarily on judgments of how the practice of analysis compares with a set of theoretical

standards. Two caveats related to this approach must be recognized. One is that words like “cost effective” have been used in the literature much more freely than they would have been had all

authors meant to adhere to strict CEA/CBA definitions. Nevertheless, since articles employing these words freely contribute to the health care community's perception of the meaning of terms and uses of analysis, it is appropriate to include them in a review of CEA/CBA literature.

The second caveat is that several of the standards of ideal (or idealized) analysis may be unattainable. If so, a review of the literature will necessarily have a critical tone. Many of the flaws of the health care CEA/CBA literature reflect inherent, or at least very common, analytical problems. Examples include difficulties incorporating distributional concerns into formal analysis and deficiencies of data accessibility, quality, and consistency. Some common CEA/CBA problems impose unusually severe burdens on health care studies. The difficult and often controversial valuation of less tangible costs and benefits, such as the saving of life and reduction of physical suffering and emotional distress, is often central to the health care analyst's chore. Even more basically, the estimation of production relationships seems particularly challenging in health care, where the difficulty of attributing health outcomes to health care inputs has led many scholars to rely for evaluation on intermediate (nonoutput) measures such as structure and process. Technical change occurs with such extraordinary rapidity that forward-looking health care CEA/CBAs are particularly handicapped. Furthermore, even some commonly accepted "second-best" CEA/CBA practices are hard to justify in health care CEA/CBA, one example being the use of market prices as measures of true opportunity costs.

Not all of the flaws in the health care literature are attributable to inherent difficulties. The relative novelty of CEA/CBA in health care seems to account for the exaggerated importance of several errors. Representative are the absence or mishandling of discounting and the presentation of purported CEAs that examine only one program (i. e., no alternatives) and conclude that it is cost effective. More significant is the tendency of investigators to use purely retrospective evaluation of existing programs to develop policy proposals for the fu-

ture, with little or no regard for the changes that will transform the structure and functioning of such programs. Many studies are plagued further by the "black box" approach to ascertaining production relationships: The identification of inputs and outputs without devoting sufficient attention to the efficiency of production, or even to basic questions of causation versus correlation.

By contrast, the best of health care CEA/CBA makes the novelty of the literature a source of encouragement. A handful of skilled analysts are breaking methodological and substantive ground, working on evaluative techniques, and producing informative, thought-provoking analyses. In recent years, investigators have demonstrated how analysis can yield insight into the nature of timely policy issues (473),⁷ contribute to efficient program planning (158), grapple with technical evaluation problems (573), and address inadequately studied technical aspects of medicine, such as diagnosis (360,559). Such works may presage a variety of interesting, useful developments in a field whose novelty provides a set of wide-open methodological and substantive opportunities.

Illustrative of recent methodological developments of considerable promise is the growing analytical comprehensiveness of CEAs and the trend away from comparing direct program costs with single-outcome measures of effectiveness (e.g., "lives saved"). Recent efforts to incorporate indirect costs and develop more inclusive indexes of effectiveness (e. g., QALYs) have begun to transfer a major virtue of CBA—its comprehensiveness—to CEA, while greatly reducing the accompanying problem of explicitly valuing noneconomic health benefits. Several studies demonstrate comprehensive cost accounting, with both positive costs and "negative costs"—indirect economic benefits—aggregated

⁷The analysis of the national swine flu immunization program (473) was conceived, in part, as an experiment to see whether a formal analysis, relying heavily on concurrence of expert opinion (through use of a Delphi), could be accomplished quickly—prior to a policy decision—and still produce useful information. Despite its limitations—failure to anticipate social, legal, and medical problems and their economic sequelae—the analysis served to inform and put issues into perspective for much of the health care community.

on the cost side of the CEA equation. The remaining noneconomic values constitute the programs' effectiveness. In some instances, the remaining effectiveness measure is a simple single outcome—sterilization, for example (134)—while in others it is a more complex index, such as QALYs in hypertension control (573). In still others, effectiveness measurement or valuation is made irrelevant by the fact that complete cost accounting indicates a positive net benefit before “remaining effectiveness” is taken into account (86). The narrowing of the gap between CBA and CEA is made vividly clear by this last case. It is also interesting to note that Cretin (122) called her study a CBA, yet she did not place a dollar value on years of life saved and she presented results in terms of costs per added years of life—a typical CEA “bottom line.” One might be tempted to dismiss this as a case of mislabeling, but in fact the growing economic sophistication and comprehensiveness of CEAs have introduced a healthy terminological ambiguity.

OTA's assessment of the quality of the literature has relied on a comparison of practice to a set of theoretical standards. Nevertheless, there are other bases for assessment of quality. For example, if one believes that quality is best reflected in the validity and reliability of results, one might seek internal or external measures of validity and reliability. An example of an internal measure is comparison of findings across studies of the same topic. To be sure, one must be wary of one study's replicating the method of earlier studies, or of use of the same data sources leading to a shared bias (i. e., consistent but not valid results). In the absence of a shared bias, however, consistency of results is suggestive of meaningful findings.

The literature does provide a few cases of multiple analyses of a single subject. Studies of renal disease treatment offer an excellent example. Two contemporary analyses ranked treatment alternatives in the same order—transplantation being most cost effective in one study (298) and cost beneficial in the other (327), followed in both studies by home dialysis, and last, center dialysis. These results were confirmed in a study published 10 years later using

more recent data (513). Similarly, three separate studies of PKU screening concluded that this is a socially desirable medical practice (78,517,553). By contrast, analyses of CT scanning have produced widely discrepant findings, reflecting differences between head and body scanning, technical changes (realized and anticipated) over the time period covered by the studies, and differences in investigators' perspectives as to what constitutes effectiveness in scanning or, more generally, in diagnosis (2,28,211). Although a systematic comparison of analyses on single subjects was not attempted in this review, that might prove to be an enlightening approach to evaluating the literature.

Assessment of the quality of individual contributions to the literature has received primary attention in this chapter. Chapter 1 and appendix A examined the overall composition of the literature, but “quality” judgments were limited to observation of the conspicuous absence of certain substantive concerns, such as important disease problems (e.g., diabetes) and medical techniques (e.g., a large number of diagnostic techniques other than screening). Here it should be noted that an interesting indication of the overall composition of the literature is the mix of CEA/CBAs with positive and negative findings. If some medical practices are socially and economically desirable and others undesirable (or of questionable desirability), one might expect a “balanced” literature to include a **good** mix of positive and negative findings. A lack of balance certainly need not reflect poorly done individual studies. Rather, it might result from analysts' having a systematic bias in favor of studying desirable or undesirable programs. For example, if CEA/CBA were applied primarily to analyzing programs whose worth has been challenged, one might anticipate a preponderance of negative findings in the health care CEA/CBA literature. A preponderance of positive findings could follow from medical professionals' analyzing (or commissioning analyses of) projects whose diffusion into practice they favor. Dominance of either positive or negative findings might reflect systematic underestimation or overestimation of either benefits or costs. For example, as discussed above, few

analyses include a realistic assessment of the costs of implementing a policy and of the **possible dilution of benefits that may follow**. These factors should produce overly optimistic results, i.e., they introduce a distinct bias toward positive findings. On the other hand, many health care programs are characterized by important intangible benefits, the value of which frequently is not incorporated into analysis. This factor introduces a bias toward unduly negative findings.

The reading of the literature suggests a dominance of studies having positive findings. To be sure, there are notable exceptions, with some analyses producing distinctly negative findings (28,365,397,570) and surprisingly few deriving equivocal results (11).⁷ Also, there may be a shift taking place, with movement from the positive toward the negative. This could reflect the general questioning of medical technology and growth of cost consciousness, both of which emerged strongly in the 1970's.

Chapter 1 noted that this report was restricted to considering personal health care services. In concluding this review of the literature, it seems appropriate to observe that the community of health care CEA/CBA analysts seems to have established a similar boundary. Unless policymakers and analysts remain cognizant of the existence of that border and its implications, this limitation can mislead technical aspects of analysis and, more importantly, reinforce narrower views of health resource allocation. A prominent example of a technical problem is

the recent emphasis on measuring "net health-care cost" in CEAs (574). The socially relevant concept should perhaps be net social cost, in which net health-care cost is but one important component.

One of the drawbacks that can arise from limiting policy analysis to medical care parochialism is a failure to explore the possibility of cost-effective alternatives to personal health services. In the effort to reduce mortality and disability due to motor vehicle accidents, how might highway safety efforts—technical (e. g., safer road surfaces and shoulder barriers), legal (e. g., increased law enforcement), etc.—compare with improved emergency medical services? To reduce hypertension-related mortality and morbidity, what is the appropriate mix of medical interventions and community health education on risk avoidance? There is a paucity of comparative analyses crossing the medical-nonmedical, or personal health-public health, border. A noteworthy exception is comparison of community water fluoridation with a variety of individual treatment approaches to preventing dental caries. Noting this paucity is not meant to reflect adversely on either existing or future individual contributions to the health care literature—the quantity and importance of analyses of specific medical problems and technologies is sure to grow, a development to be desired. Rather, it is to suggest that policymakers, health planners, and individual health practitioners would benefit from the widening of perspective that "border-crossing" analyses could offer.⁸

⁷It can be argued that it takes a strong constitution to present equivocal findings. There is a common perception that the publication market prefers "definitive" to ambiguous findings. This is reflected in the CEA/CBA literature in which equivocal results seem to be much more rare than probabilities would lead one to expect. Although OTA's literature search discovered a preponderance of studies with positive findings, there appeared to be many more studies with negative than with ambiguous findings. Those with ambiguous findings tend to be competent analyses; their ambiguity often reflecting allowance for variation in uncertain parameters (e.g., 122).

⁸A budgetary pragmatist might argue that medical and nonmedical resource allocations are bureaucratically independent, with border-crossing reallocations extremely unlikely, and hence that border-crossing analysis is not worthwhile. Although this may be true in the short run, relative resource allocations do change over time and might be responsive to analytical input. Clearly, this is occurring at present in the new Federal Government prevention initiatives within HEW (743). More to the point, however, is OTA's finding that one important strength of analysis is its ability to affect thinking about problems—perspective and not the making of explicit resource allocation decisions.