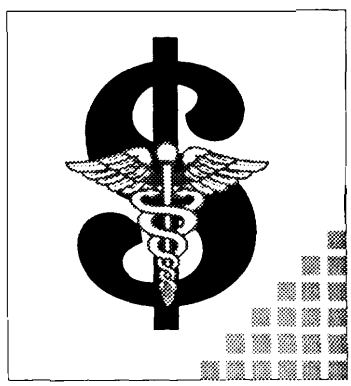
The State of Cost-Effectiveness Analysis

5

Information about the costs and effectiveness of interventions is crucial to any decision about how to make resource allocation decisions in health care, an intrinsic concern of health policy makers. Purchases and consumers of health care services are increasingly concerned with not just the effectiveness but the value of the care being provided

Where both the costs and the effectiveness of health interventions are components of decisions, the quality and validity of the decisions can be increased by considering those components explicitly (780). Cost-effectiveness analysis can improve decision-making by forcing a structuring of the decision process and providing a framework for identifying and considering as many of the relevant costs and benefits as is feasible (780).

Cost-effectiveness analysis and related techniques are increasingly commonplace in health care discussions and literature. With the greater use and acceptance of this technique, however, old issues relating to its validity and the quality of studies that employ it have gained new importance. and new issues have arisen. This chapter reviews some of the changes in how and why cost-effectiveness analyses are done, and the importance of those changes to policy makers and other users of these analyses.



ANALYTIC APPROACHES

The economic evaluation of health care alternatives comprises several related but distinct types of analyses. Three of these are

particularly relevant to resource allocation decisions and decisions about the relative value of alternative health interventions. [1]

Cost-benefit analyses are the oldest form of comparative economic evaluation. and they are frequently used in fields such as engineering and defense (box 5-1). In health care, cost-benefit analyses enumerate and compare the costs and benefits that arise as a consequence of applying an intervention (e.g., a medical technology or a public health program). Both costs and benefits are measured in dollars, enabling the analyst to compare a summary measure, such as the cost-benefit ratio or net cost (or savings), across any number of interventions (780).[2] Cost-benefit analysis is not a primary focus of this chapter, because its need to place dollar values on lives has resulted in disfavor among medical analysts, and it is relatively little used for the direct comparison of particular medical technologies. It is being applied in the analyses of some health programs, however, as discussed towards the end of this chapter.

Cost-effectiveness analysis (CEA), in contrast to cost-benefit analysis, does not convert units of health outcome into their "worth" in dollars. Rather, it calculates the cost per specified health effect of a technology or program--e. g., cost per lives saved, or cost per cases of' cancer avoided—and compares this cost-effectiveness ratio with ratios from other interventions (780). It is a crucial tool in the full assessment of medical technologies and services and is the primary focus of this chapter.

To be comparable in a CEA. all interventions must have their effects expressed as similar units, This is a problem when the goal is to compare technologies or services whose outcomes are not especially similar, such as prenatal care and stroke rehabilitation. Cost-utility analysis (CUA) is a variant of CEA that addresses this issue. In a CUA, the outcomes are expressed as uniform units of health that are presumed to have similar values across all conditions--"healthy days." "'health) years of life," or "quality-adjusted life years" (QALYs)—years of lifc saved by the technology, adjusted according to the quality of those lives (410).

Many analysts consider CUA and CEA to be separate entities.[3] In this chapter, however, cost-utility analysis is considered a subset of cost-effectiveness analysis that is especially powerful and that introduces some unique additional issues and concerns. The main reason for (his categorization here is pragmatic: both techniques are aimed at answering the basic question of the relative value of health interventions, and "cost-effectiveness" is the more inclusivc and familiar term. However. the North American research community also seems to be leaning towards considering CEA to be the umbrella term (867).

USES OF COST-EFFECTIVENESS ANALYSIS

The purposes to which cost-benefit and cost-cfl'activeness analyses are applied by health care decisionmakers are increasingly diverse. Some potential uses are still more anticipatory rhetoric

---

[1] Other examples of methods of economic evaluation of health care programs include cost-of **illness** analyses, descriptive studies that are aimed at enumerating all of the costs related to applying a particular technology, or set of services. to the population" of interest (e. g.. people with cancer) (410) and cost-minimization analyses, which assess the least-cost method of achieving a particular outcome (1 83). Neither 01 these methods assigns relative values to both costs and outcomes across alter-rui(ive interventions, and neither i s addressed further here.

[2] Because both costs and benefits can be considered either positive or negative (e.g., a positive cost is also a negative benefit), the cost-benefit ratio is susceptible to manipulation (e.g., by restating a cost as a negative benefit and moving it to the denominator of the ratio). For this reason, the use of a summary measure of net cost (or net savings) is often preferred (217).

[3] In this framework CEA is used to compare the relative costs of achieving a single, common effect in alternative ways, with the effect usually calculated in natural units (e.g., lives). CUA is considered by those analysts to be a method of comparing costs of achieving either a single or multiple effects, where the value of the effects are specified in terms of their worth of a particular level of health status (187,394).

BOX 5-1: The Origins of Cost-Effectiveness Analysis in Health Care

Attention to calculating the costs and benefits of public projects was a mandate of the U S Army Corps of Engineers, which was required by the River and Harbor Act of 1902 to assess the cost and benefits of river and harbor projects The Flood Control Act of 1936 took this concern a step further, requiring the U S Bureau of Reclamation to assess its water projects, stipulating that such projects should only be undertaken if the benefits should exceed the costs, As a result, cost-benefit analysis was applied to many water projects of the next *30* years (481) Other government entitles began using cost-benefit analysis to assist in program budgeting, and in the 1960s the Department of Defense began to employ the technique to evaluate alternative defense projects

Rice *(636)* was one of the first to employ methods for the economic evaluation of health care and health outcomes, in a 1966 paper on the cost of illness It rapidly became clear that this context raised considerable new issues and controversies Experts and stakeholder groups disagreed about what constituted appropriate outcome measures in such analyses, and valuing health and life in dollars—necessary for cost-benefit analysis—was controversial and, some maintained, unethical As a result the subsequent emphasis in health care tended to be on cost-effectiveness analysis, comparing health outcomes directly, rather than on cost-benefit analysts

During the late 1960s and early 1970s cost-benefit and cost-effectiveness analysts were applied to health subjects, but studies were not generally aimed at health declsionmakers who were making resource allocation decisions Neuhauser and Lewicki (560) helped change this situation when they analyzed 'the gain from the sixth stool guaiac " in a classic paper examining the marginal cost-effectiveness of a series of inexpensive tests for cancer screening.

In 1977 this initiative was followed by two milestone papers by Weinstein and Stason, which were published in the widely read /Vew *England Journal of Medicine* and accompanied by editorial discussion (720,906) These authors outlined the theoretical foundations and applications of cost-ef - fectiveness analysis of health care interventions and applied the technique to the allocation of resources for the medical management of hypertension Subsequent contributions to the health care cost-effectiveness literature in the 1970s emphasized the advantages of cost-effectiveness over cost-benefit analysis for analyzing health care topics (897) and highlighted the link between cost-effectiveness analysis and health technology assessment (244), These articles set the stage for a decade of growing interest and discussion of methodological issues and applications of the technique This literature, in turn, became the intellectual foundation for the increasingly widespread acceptance of cost-effectiveness analysis in every aspect of health care planning, management, and clinical decisionmaking that is coming to characterize the decade of the 1990s

SOURCE Adapted from B R Luce and K Simpson "Methods of Cost Effectiveness Analysis Areas of Consensus and Debate unpublished report prepared by Battelle Institute for the Pharmaceutical Manufacturers Association Washington DC Apr 22 1993

than reality, but they demonstrate the great variety of contexts in which CEA can be applied and the wide range of needs it can address. These uses include:

- manufacturers decisions about which lines of research on health care products to pursue,
- providers' decisions about purchasing new or replacement technologies.
- decisions by hospital or managed care formu - lary committees about which drugs to include on the list of pharmaceuticals that physicians may conveniently prescribe.

- payers' decisions about making changes in the way particular technologies are covered or paid for,
- inclusion in clinical practice guidelines, to help clinicians, patients, and payers judge the value of alternative care strategies,
- the design of insurance benefit packages under health reform, as policy makers attempt to judge which services should be covered,
- budgetary allocations within programs or agencies, and
- cross-agency resource allocation.

Although these uses differ considerably, in all cases decisionmakers can have a large stake in the results of the analyses they use. The validity and reliability of cost-ef'festiveness analyses, and differences in their usefulness that depend on the purpose to which they are put, are of widespread interest and concern.

MECHANICS OF COST-EFFECTIVENESS ANALYSIS

A CEA typically proceeds through a series of steps, as follows (410,780):

1. **Identify interventions to be compared.** These may be different ways of providing the same technology (e.g., Pap smears), alternative medical treatments, or the variety of health programs across which society allocates its resources. The analyst specifies not only the exact technology involved but also the population it affects.
2. Identify costs of the interventions. These are the health care resources and other costs (e.g., travel and caregiver time) incurred when the interventions are used.
3. **Identify health and other impacts of the interventions.** This step requires identifying not only what health outcomes can occur in patients receiving the intervention, but often also identifying relevant impacts of the intervention on nonpatients as well (e.g.. on patients' employers and families).
4. Measure costs. After the analyst has identified what the costs are, he or she must apply "accepted economic procedures for attaching monetary values" to them (41 O). For example, if one resource cost identified is the need for a visit to a physician, the analyst must assign a dollar value to that visit+. g., the physician's charge for the visit, or the insurer payment for the visit.
5. Measure health and other outcomes. In CUA this measurement can be fairly complex, requiring the analyst to define a numerical health scale and using it to measure the outcomes. Thus, for example, if some of the outcomes identified as relevant are death, full health, and chronic disability, the analyst must assign values to these outcomes and the probabilities with which they occur.
6. **Examine uncertainties that underlie the analysis.** Rarely are many of the components of a cost-effectiveness model known with certainty. Usually they are either estimates, based on previously or concurrently collected data, with some confidence that the "real" number lies close to the estimate, or they are assumptions that are "best guesses" about the approximate size of the number. The effects of these uncertainties are usually examined through sensitivity analyses—"running the numbers" —again many times, with higher and lower estimates and alternative assumptions, to discover the range of possible results and the effects of different crucial assumptions on those results.
7. Present and interpret the results. The results of a cost-effectiveness analysis are usually presented as a single number whose magnitude can be compared across the alternative technologies analyzed—for example, the cost per life saved, or the cost per quality-adjusted life year. How those results are compared, however, can affect the conclusions a reader draws. A common form is to present the results as a marginal (or incremental) analysis, in which each result is compared with the previous one in stepwise fashion. This is the preferred method of presenting results inmost cases that involve analyzing alternative ways of reaching a common goal, because it allows the user to see what extra benefit is being gained for the extra costs of moving from one alternative to the next.

Questions about the relative cost-effectiveness of health interventions, like many other questions in health care studies. can be addressed either through retrospective or prospective studies. Most CEAs have traditionally been performed as retrospective analyses, using pre-existing data on the costs and effects of the different alternatives being compared, and a model created by the analyst that relates all of the components of the analysis. Models range from simple decision trees—schematic presentations of the choices the decisionmaker can follow and the consequences of each of those paths—to complex computer models that simulate the effects in great detail.

A cost-effectiveness analysis of Pap smear screening for elderly women (546) illustrates this common kind of CEA. In this analysis, the goal was to judge the effects and the costs to the Medicare program of covering this service, which was not previously a Medicare benefit. The population of interest comprised women age 65 and older. As a first step, the analysts chose to compare no screening with four other alternatives: one-time screening at age 65, and periodic screening beginning at age 65 and continuing thereafter every five years. every three years, or annually.

The main outcomes of interest in this analysis were the costs to the Medicare program and the health effects of screening for the women. These effects were measured in lives saved (i.e., deaths from cervical cancer averted). The model chosen for the analysis was a Markov model. a type of iterative decision model that represents the changes over time in the proportion of people who are in different health states (e.g., healthy, living with cancer, dead) based on assumptions about the probability with which people become ill, are cured using different treatments, or other factors.

In this analysis, running the model required estimates and assumptions of such factors as the total number of women in the population, the number who would choose to have Pap smear screens, the number of cancers detected by the screens, the cure rate for detected cancers, the death rate for cancers that are not cured. and the death rate for conditions other than cancer. It also required estimating the costs to Medicare of women undergoing screening, diagnosis, and treatment for cancer, At each iteration of the model. representing one year, the number of women alive and dead and the Medicare costs associated with prevention and treatment of cervical cancer were totaled. At the end of the iterations for that screening scenario (e.g., annual screening), the analyst summed the deaths per year and the total costs to Medicare.

Several of the model inputs required assumptions surrounded with a great deal of uncertainty. The sensitivity analysis. rerunning the model with slightly different assumptions. revealed that the overall costs and effects of screening varied considerably depending on the proportion of women assumed to develop cancer in the absence of screening, on the accuracy with which Pap smears detected cancers, and on the cure rates of treatments for detected cancers.

Finally. the analysts presented the results incrementally. Thus, it was possible to conclude that screening every five years was both less costly and more effective than one-time screening, under the baseline assumptions of the model. Screening every three years was slightly more expensive per life saved than screening every five years, but it saved many more lives. Screening every year saved the most lives of any scenario. but compared with the next-closest alternative (every three years) it was much more expensive for those additional lives that would be saved (546).

ISSUES IN THE USE OF COST-EFFECTIVENESS ANALYSES

The literature addressing cost-effectiveness and cost-benefit in health care is no longer small. Over 3,000 articles and letters on the topic were published from 1979 through 1990 alone, of' which nearly 2.000 were analyses of particular interven - tions (219). (The remainder were methodologic papers, reviews, letters, and other forms of' commentary.) The growth of this literature has continued at a rapid rate since 1966 and show's no signs of slowring (219,781). Entire journals are now

dedicated to cost and cost-effectiveness analyses (e.g.,Pharmacoeconomics).

As CEA gains in acceptance and application to health care issues, both in the United States and abroad, the degree to which the results of these analyses are reliable and valid for the purposes to which they are applied has become a matter of increasing practical concern. Four issues in particular that affect the usefulness of this tool to policy makers are addressed here:

1. **The comparability of analyses:** the extent to which different analyses of the same question give similar answers;
2. **The quality of analyses:** whether the analyses incorporate the basic principles that most analysts agree would tend to increase the validity of the results;
3. **The increasing use of prospective cost-effectiveness analyses,** in which clinical trials are designed that measure costs as well as health effects; and
4. The implications of cost-utility analysis, and the basic assumptions that underlie the results of CEAs that express health effects in terms of patient "utilities" and quality of life.

Comparability

The comparability of CEAs depends largely on the structure and assumptions of the analyses. As the number of analyses has proliferated, the approaches to analysis and the diversity of assumptions behind them have proliferated as well. As a result, CEAs in the current literature that appear to address similar questions often have results that in fact cannot easily be compared.

In a striking example, Brown and Fintor reviewed published CEAs of a single set of technologies: screening for the early detection of breast cancer (89). They identified 16 studies from 6 countries that assessed the cost per life- year saved or death averted. These 16 studies had wildly different results, mostly due to the different assumptions used in each cost-effectiveness model.

Four of the 16 studies reported breast cancer screening to actually save health care costs, while the other 12 found it to increase costs by differing amounts. Interestingly, those four studies all assumed that the lifetime costs of treating early-stage cancer were very different from treatment costs for later-stage disease, an assumption that Brown and Fintor traced to a single anecdotal report unconfirmed by any systematic study. The reviewers concluded that the results of those four studies were thus almost certainly invalid, and they turned to the other studies.

The cost-effectiveness ratios reported in these 12 studies differed enormously, ranging from $3,400 to $20,000 per life-year saved (89). When these results were adjusted to account for differences among studies in the assumption of the direct cost of screening (e.g., the cost of a mammogram), the range was even greater: $9,500 to $144,700 per life-year saved.

The reviewers then pointed out that the studies and their results were still not directly comparable for a variety of reasons. Two of the studies reporting the highest costs per life-year saved based their assumptions of the benefits of screening on evidence from short-term observational studies, and thus those analyses probably underestimated long-term benefits, which the observational studies were unable to detect. Two other studies were from countries with a relatively low prevalence of breast cancer, which would tend to make screening programs look less beneficial because they would detect fewer cases of cancer. And the studies differed in the exact technologies they were comparing; most compared mammography and/or clinical breast examination with no screening at all, while one study analyzed the cost-effectiveness of mammography compared to a pre-existing screening program using only clinical breast examination.

Setting aside all of the studies whose results were incomparable for these various reasons, the reviewers were still faced with two studies that ap-

peared to be very comparable in their methods and circumstances, yet whose results differed by a factor of four.[4] Examining these two studies more closely, Brown and Fintor identified a number of ways in which the underlying assumptions of these analyses differed. Examples include:

- frequency of screening (annually vs. every two years),
- efficacy of screening (30 vs. 33 percent reduction in mortality),
- length of time the population was "followed" and benefits were calculated (30 vs. 98 years), and
- cost of a mammogram ($40 vs. $50).

When the two models in these studies were rerun using the same numbers for these and other differing assumptions, the model yielded nearly identical results.

The CEAs examined in Brown and Fintnor's study exemplify many of the strengths and the weaknesses in the CEA literature. On the one hand, the similarity in the results of the final two analyses when they were run using similar assumptions affirms the basic reliability of the CEA method. On the other hand, the diversity of results from the overall set of studies would bewilder most users and might lead them to discount even the better quality analyses.

Mason and colleagues undertook a somewhat similar exercise and suggested seven categories across which cost-effectiveness studies often differed so much as to make their results incomparable. They analyzed a previously published table that compared the cost per QALY of 21 different medical technologies, ranging from advice to stop smoking to heart transplantation. The cost per QALY in this table ranged from 220 to 126,290 British pounds (499). Mason et al. examined the individual studies whose results were used to create this table and concluded that displaying their results in a single table was misleading. because the studies themselves involved such very different assumptions and circumstances that it was invalid to compare their results directly (495). Categories of factors that differed among studies were:

1. **Year the** study was done. In some older studies, the comparisons in the cost-effectiveness analysis were among technologies that were sometimes outmoded, making the results irrel - evant or misleading.
2. Discount rates. Analysts usually "discount" the value of costs incurred in the future, to reflect the fact that having a dollar today is generally considered preferable to having that same dollar sometime in the future (even if there is no inflation).[5] CEAs differ not only in the exact discount rate they use (e.g.. 4 or 6 percent) but in whethcr they discount future health benefits as well as future health care costs.
3. Preference values for health states. To calculate a QALY, the analyst "adjusts" years of life in a study population according to the level of poor health in that population and how bad people think those disabilities are—the "quality" of their lives compared with perfect health. Different ways of measuring these values give somewhat different results. and the QALY in one study thus may not actually be the same as a QALY in anothcr study. (This difference did not arise in the studies reviewed by Brown arid Fintor. because those studies used only likes saved rather than quality-adjusted lives as the "health effect" endpoint.)
4. Range of costs **included.** Most CEAs measure at least the direct costs associated with the technology being studied: the costs of applying the technology, the costs associated with side effects. and so on. Some. how'ever, also consider indirect costs, such as the monetary value of

[4] The two studies that appeared to be the most comparable were not necessarily of better quality than all of the excluded studies; they were just the most similar in method.

[5] Benefits and costs accruing now are worth more than if they accrue in the future. The procedure for adjusting for the time value of resources or costs is referred to as discounting, in that benefits and costs that are incurred in the future are valued in current dollars (906).

the patients' and families' time. Including indirect costs can change the results considerably.

5. Comparisons **used in the analysis.** Some studies analyze the cost-effectiveness of the technology of interest compared with doing nothing. Others compare it with another technology, while still others consider the incremental cost per QALY of expanding the technology to other groups of patients, or to other circumstances. In the breast cancer screening example above, this was clearly a problem in comparing results across studies.
6. Location and context of the study. Comparing the results of studies done in different countries without accounting for differences in the relative cost of alternatives within each country can lead to invalid results. In an example of how important country-specific prices can be, Drummond and colleagues compared the results of identically performed CEAs of a drug in four countries. They concluded that the drug was the most cost-effective in the United States despite being more expensive in this country than in others, because alternative treatments (e.g., surgery) were also more expensive (184).
7. **Quality of the study.** This category encompasses a wide range of differences among CEAs regarding both the assumptions they use and the way the results are presented. One criterion defining a poor-quality study, for example, is a study that does not identify the sources for its assumptions (e.g., about disease prevalence or health outcomes). Indeed, many published studies do not even state what their assumptions are. Where those assumptions are not stated, later users of the analyses cannot detect how those assumptions might have affected the cost-effectiveness results.

Quality of Analyses

Differences in quality may affect the comparability of results, as described above. Of even greater concern, poor quality analyses may be invalid representations of the true relative costs and effects of the alternatives, possibly leading to worse rather than better decisionmaking, and very uneven quality of the analyses themselves.

Evidence suggests that the quality of CEAs is indeed cause for concern, even in the peer-reviewed literature. In an examination of this topic, Udvarhelyi and colleagues reviewed 77 articles on the cost-effectiveness of medical interventions that were published either in the late 1970s or the mid- 1980s and rated them according to six "fundamental principles" of cost-effectiveness and cost-benefit analysis (77 1). The principles they used were:

1. **An explicit statement of a perspective for the analysis should be provided. The** study should state whether, for example, it is the costs and benefits to patients, payers, or the health care system as a whole that is of concern.
2. **An explicit description of the benefits of the program or technology being studied should be provided.** This description should include assumptions about presumed benefits if the actual benefits are uncertain.
3. **Investigators should specify what types of costs were used or considered in the analysis.** The types of costs chosen should be linked to the perspective of the analysis (if the perspective is the hospital, for example, long-term morbidity costs might be excluded).
4. **If costs and benefits accrue during different periods, discounting should be used to adjust for the differential timing.** The reviewers did not require any particular discount rate, but they judged articles according to whether discounting was at least addressed whenever long time periods were involved.
5. **Sensitivity analyses should be done to test important assumptions.** As part of this principle, the reviewers stated that all assumptions of the cost-effectiveness model should be explicit, so that if the authors did not perform sensitivity analyses on important assumptions readers could be alerted to the possibility that the analysis' conclusions could change.
6. **A summary measure of cost effectiveness or cost benefit (e.g., the incremental cost-effec-**

tiveness ratio of each alternative analyzed) should be calculated. The reviewers specifically preferred the use of incremental cost-effectiveness ratios to average ratios.

Fewer than one-fourth of the analyses adhered to at least five of these six principles, and only half the articles adhered to more than three (771). The first principle was especially poorly followed: "Few studies (18 percent) explicitly stated what perspective the analyses used, leaving this to the readers' judgment. " Only 30 percent of authors used sensitivity analyses to test the vulnerability of their conclusions to key assumptions.

Equally troubling, the quality of analyses did not seem to be improving over time. The articles published in the later time period studied (1 985-87) were of no higher quality, overall, according to these criteria than articles published in 1978-80 (77 1). Yet the principles used in this review were not new ones (780), and previous reviews of the medical cost-effectiveness literature had also criticized the quality of papers published (18 1,894).

A recent review by Jefferson and Demichelli supports these general conclusions. In a very comprehensive review of the world literature on the cost-effectiveness of vaccines against hepatitis B, these authors found problems with both the comparability and the quality of studies (394). Of the 90 studies they examined, for example, only 37 contained a sensitivity y analysis, and only 36 clearly explained the time span over which benefits and costs were assumed to accrue. Of these, the time span used ranged from 4 to over 22 years. Similar-1 y, only one-fourth of the studies specified the discount rates they used, and the rates given ranged from 3 to 10 percent.

As the criteria used by Udvarhelyi and colleagues demonstrate, much of the concern about the quality of CEAs is not only that the appropriate structure and assumptions be used, but that these be explicitly stated. If the discount rate used in an analysis is not mentioned in the published account of the analysis, for example. it is both impossible to tell if an appropriate rate was used and impossible to compare the results of that analysis with others.

Explicitly stated structures and assumptions can enable later reviewers to redo analyses with common assumptions and examine the reliability of results, as Brown and Fintnor did in the case of breast cancer screening. But as CEAs become increasingly used in health care decisions. the inconsistency among analyses in their basic assumptions is itself of concern. End users of analyses may have neither the time nor. often, the training to examine and reconstruct individual analyses in detail.

Inconsistencies exist even among good quality analyses. One of the clearest and most long-standing examples of inconsistencies is in the use of discount rates, the analyst attempt to account for differences in how costs and benefits are valued in the future compared with the present. Virtually all analysts agree that costs that will not be incurred until far in the future should be discounted to reflect their true current value to decisionmakers in the present. Most also agree that health benefits received in the future should be discounted as well. What they do not agree on is exactly what that rate should be in either of these cases (399, 42 1,433,906).

Similar inconsistencies occur across other basic attributes. For example, there have been considerable strides in understanding and enabling the measurement of peoples' values for different health outcomes, but there is also still considerable disagreement over which measures are best. The diversity of the academic debate is reflected in the diversity of assumptions underlying different CUAs. QALYs are probably the most common health outcome measure, but measures such as the saved young life equivalent and healthy life expectancies have also been both proposed and used (567). And the analysts who calculate their results in terms of QALYs use a wide range of underlying methods to value those years, itself a source of considerable differences among studies (254,41 O, 568).

Prospective Analyses: Clinical-Economic Trials[6]

As interest in cost-effectiveness has grown on the part of purchasers and payers of health care, the cost implications of new technologies are sometimes evaluated even while the efficacy of the technology is itself still being investigated in a clinical trial (box 5-2). Both the economic and clinical data from the trial are then analyzed to provide information about the cost-effectiveness of the technology. These *clinical-economic trials* are frequently initiated early in the development of a technology or medical practice (e.g., prior to FDA approval), although some have been conducted later, after the technology gains wider use in routine clinical practice.

CEAs based on economic data collected in clinical trials are still relatively rare (5), but they appear to be increasingly attractive. To date, most such studies have been sponsored by private manufacturers, whose goals for doing them include:

- to ensure that data on the economic implications of the technology are available for marketing purposes,
- to facilitate establishment of a price that will provide adequate return on investment while maintaining the economic viability of the technology,
- to formulate priorities on those drug indications, among alternatives, that should be sought in the FDA approval process, and
- to minimize the time between a technology's development and its coverage by health insurers.

Although manufacturers clearly have their own interests in supporting clinical-economic trials, prospective collection of data on costs as well as effectiveness has certain broader attractions. First, clinical trials conducted for the FDA approval process provide an opportunity to collect economic data when most needed to plan and guide appropriate utilization of technology by health care providers. Second, in contrast to clinical trials, studies using secondary data often incorporate data from disparate and at times incompatible sources, making the results difficult to interpret or apply. Thus, early, relevant, rigorously collected economic data coupled with a strong experimental design could be especially valuable.

Despite their considerable theoretical benefits, clinical-economic trials have significant drawbacks as well. Clinical trial timing, protocols, settings, and the nature of cost measurements all place their own special constraints on the collection of relevant cost data and have important implications for the validity and general inability of the results of CEAs based on these trials.

Influence of Clinical Protocol on Resource Use

Clinical trial protocols themselves can influence resource utilization and costs. Protocols for clinical studies of a new technology often include tests to monitor study participants for serious or unknown side effects. For example, in the Women's Health Study conducted by NIH, in which hormonal therapy is being assessed, women will undergo frequent office followup. electrocardiogram, endometrial biopsy, and mammography to monitor safety of hormonal therapy. It is unlikely, however, that all of these tests will become components of the intervention in later routine clinical practice. Although such tests might be identified and their costs excluded in the data collection, monitoring can have more profound effects when an abnormal test induces further testing or treatment.

Another important component of many clinical trial protocols is "blinding" patients and physicians to treatment alternatives, to e1iminate biases in the perception of which treatment is preferred.

---

6 This section is based on N.R. Powe and R.I. Griffiths, "Economic Data and Analysis in Clinical Trials," paper prepared under contract to OTA, forthcoming, 1994.

BOX 5-2: Examples of Clinical-Economic Trials

The most prominent use of clinical-economic trials has been in the area of new pharmaceuticals Most such studies are sponsored by the manufacturers, although they are often performed by academic institutions or other private organizations at the medical centers where the clinical study IS being conducted Examples include studies of the costs and health benefits of granulocyte macrophage colony stimulating factor (GM-CSF) as adjuvant therapy in relapsed Hodgkins disease (313), short- versus long-course antibiotic treatment of spontaneous bacterial peritonitis (665), and low- versus high-osmolall{y radiographic contrast media in patients undergoing cardiac angiography (611)

Cltnlcal-economic trials of medical equipment have also been conducted, although much less frequently Among the few devices that have been the subject of this kind of study are home air-fluidized therapy for treating pressure sores (733), automated versus manual methods of syringe filling (4) and videopelviscopy versus laparotomy for ectopic pregnancy (45)

A few cllnlcal-economic studies have been performed by providers to demonstrate the cost-effectiveness of one type of treatment over another in order to provide]ustification for payment or to improve efficiency For example, a study conducted by the First Hill Orthopedic Clinic in Seattle demonstrated that in spite of the fact that patient length of stay was substantially more than that which formed the basis for Medicare payment, total hip arthroplasty in patients older than 80 was a cost-effective alternative to nurslng home placement (70)

Public agencies and private philanthropic organizations have also played a role in conducting or sponsoring clinical trials with an economic component For example, a study of outpatient management of burns using biobrane versus 1 percent silver sulfidiazine was sponsored in part by the World Health Organization (276) The National Center for Health Services Research (the predecessor of the Agency for Health Care Policy and Research) sponsored several clinical studies with an economic component, including an investigation of the costs and benefits of different long-term Immunosuppressive drugs in patients who had undergone kidney transplants (694). The National Cancer Institute has also recently showed some Interest in funding clinical trials with economic components (851)

SOURCE N R Powe and R I Griffiths "Economic Data and Analysis in Clinical Trials " paper prepared under contract to Off Ice of Tee, hnology Assessment forthcoming 1 994

But the lack of knowledge about the treatment a patient has received can influence providers' resource use. For example, physicians may be more aggressive in managing a complication if they cannot be certain which treatment was uscd, since knowledge of the treatment influences their judgments about the importance and likelihood of complicat ions.

Cost Variability and Sample Size

According to researchers who have conducted clinical-economic trials. a major problem in such trials is that the costs measured are generally much more variable than the effects (870). In addition. the distributions of costs are typically skewed, with either a few persons who use few services (or

who do not utilize services) or a few persons who utilize a large amount of resources and incur high costs (612). Because of the large variance in costs, a clinical trial usually needs more participants to obtain a statistically significant result for costs than are needed to make conclusions about effects with equal confidence. Some researchers have suggested that in fact costs often need not, and should not, be measured in clinical trials to the same level of statistical precision as health effects (1 86,572,573).

Cost Measurement

Some costs are simply difficult to measure, or difficult to measure consistently, during a clinical trial. Some hospitals and practice settings do not have the sophisticated data systems available to identify institutional or provider costs in detail. This may limit the ability of researchers to perform studies in some settings or limit the type of data that can be collected (e.g., requiring the researcher to rely on charges rather than cost data). Multicenter studies raise important issues of standardization across different accounting systems and different types of providers.

Some costs may not be measurable at all during the trial because the economic consequences of treatment choices may extend far beyond the time horizon of that clinical trial. For example, thrombolytic therapy (drugs to break up clots that block blood vessels) used to treat an acute myocardial infarction can cause a stroke (a clinical endpoint), and the patient could require long-term nursing care, the cost of which could extend for many years. If the clinical protocol stipulated that patient followup end at the event of a stroke, the full economic consequences of the treatment choice would not be obtained through primary data collection.

Generalizability

Economic data from a clinical trial may reflect "cost-efficacy" (2 18) rather than real-world cost-effectiveness, just as clinical trials may reflect clinical efficacy (whether the technology works under ideal or highly specified circumstances) rather than broader clinical effectiveness. Data from a study with strict selection criteria for patient enrollment performed at the best academic medical centers may be very different from data from a trial conducted in several community hospitals, which are more representative of the average U.S. hospital.

Often studies are performed at a single institution that is part of, or affiliated with, an academic medical center. Here, there are two possible problems that influence generalizability. First, medical practice may not be similar to that in nonteaching hospitals and other clinical settings. For example, physicians at teaching institutions may order more tests and consume more resources as a result of teaching or research activities. This may result in an overestimation of costs. Alternatively, if this more careful testing actually prevents complications of treatment, the data may lead the analysts to underestimate average long-run resource use.

Institutions that adopt technologies early may have the most experience in applying them. This experience could result in physician selection of more ideally suited patients to receive a treatment, or in better identification or management of side effects. This might not only translate into better outcomes than are realized in general medical practice (61 2) but also into more efficient use. In addition to differences in physician practice behavior, there can be differences in the cost of resources across institutions of difference sizes (economies of scale or scope), location (geographic variation in resource inputs) and organization characteristics (for-profit vs. not-for-profit institutions).

Underlying Assumptions of Cost-Utility Analysis

The great attraction of cost-utility analysis is that it incorporates into a single outcome measure both the quantity and quality of life, without needing to assign any particular dollar value to that life. Thus, in theory, it can be used to compare the cost-effectiveness of technologies as diverse as treatment for ectopic pregnancy, whose primary goal is

to save lives. with hearing screening, whose primary goal is to increase the quality of life.

Unlike other CEAs, cost-utility analyses quantify not only the costs per relevant outcome, but the value to be placed on that outcome. In doing so, they raise issues that do not arise, or are not of as great concern, in other cost-effectiveness analyses. These issues revolve around the implicit assumptions that an analyst makes when he or she presents results as a cost per QALY or some other related measure. Because these assumptions are implicit rather than explicit, it is especially vital that policy makers, and others using these results, be aware of them.

Measuring Utilities

A "utility" is a quantitative measure of the strength of preference for an outcome. The theory of utility measurement (the "expected utility theorem") states that in choosing among different options involving different outcomes (in this case, different health interventions involving different health outcomes), an individual should take the action that maximizes the expected utility of the outcome (41 O).

The most rudimentary method used by researchers to value outcomes is simply assigning arbitrary rates to a ranked list of outcomes. For example, death might be assigned a value of O, recovery with disability a value of 5, and full recovery a value of 10. A 1987 review of nearly 200 medical publications that included decision analyses found that a majority of analyses published up to that point valued outcomes either according to an arbitrary scale such as this or used even simpler measures, such as life expectancy or the simple occurrence of an event (i.e., the outcome measures of traditional CEA) (41 7).

For more complex measurements of utilities, one of the most commonly used is the rating scale, in which an individual rates each outcome by placing it at some point on a scale between two "anchors" (e.g., between O, representing death, and 100, representing perfect health). The Quality of Well-Being Scale, described briefly in chapter 4 as a tool for measuring health outcomes, uses such a rating scale to assign numerical values to health outcomes. Other methods of measuring utilities include techniques such as asking people how they would choose between one alternative that would lead to certain illness and another alternative in which one might either live in full health or die (the "standard gamble approach").

Academic debate continues regarding which valuation methods most accurately represent true utilities, as required by the theories that underlie decision analysis (762). Equally important to this debate, it appears that different measurement instruments administered in different ways can come up with substantially different results (254, 462,463,742). Methods that ask people how they would feel about having various disabilities or states of health for a short time (e.g., one year) may elicit different relative values for the different health states than a method that asks people to value those health states as if they were to have them for the rest of their lives. A method that asks about the (negative) value of losing an ability may yield different relative values than a method that asks about the value of gaining that same ability. There is some evidence that even a single measure can yield different utilities in different ethnic or geographic subpopulations (5 12).

The issue of different ways of measuring utilities has implications for both the comparability and the validity of CUAs. Two analyses of the same question may arrive at different answers using different measures of utilities. yet if both analyses express their results in QALYs, users maybe unable to pinpoint the reasons for the differences. The lack of agreement on the "best" measure of utilities means that measures vary considerably. Most importantly, from the user's perspective, there is no way to know how much this variability matters, There is no literature on the robustness of CUA results depending on the measures used, or on whether results are more sensitive to the measure used for some questions than for others.

Other. deeper issues of utility measurement are still open questions as well. For example, utility theory assumes that there is some simple underlying mathematical] relationship between how one

values short-term and longer lasting states of health (e.g., pain). But there are apparently no empirical studies to confirm whether this is true.

Whose Utilities To Measure

In the simplest case of an individual making a personal decision, the utilities in a decision analysis are simply those of the decisionmaker. The individual places his or her own values on the outcomes and calculates the overall values of the alternative possible decision paths accordingly.

For other decisions in health care, however, the decisionmaker is generally making the decision on behalf of a group, and the preferences of that group must be considered. Much of the recent research into documenting health outcome preferences has been conducted in the context of specific clinical conditions (e.g., cancer or renal disease). In these cases, the preferences measured are those of patients with the relevant condition. One current Patient Outcomes Research Team (PORT) effectiveness research project, for example, is surveying stroke patients to document their preferences for the various health states and outcomes associated with stroke and conditions to prevent and alleviate it (496). Thus, a CUA that compares two management strategies for stroke could be reasonably sure that the values of the relevant group of patients were taken into account.

The least consensus regarding whose preferences to include appears for decisions that cross the boundaries of specific patient groups, particularly those that involve resource allocation decisions. In a notable and very public recent example, the State of Oregon proposed basing its Medicaid benefits package on, among other factors, the relative value of providing various categories of health services (788). The calculation of these values included measurements of the public preferences of various health states that might be affected by the services. The group whose preferences were surveyed were a sample of the general public who owned telephones. One group of critics argued that the appropriate group to survey would have been low-income persons, who were most affected by the program. A second group of critics argued that for valuing services to persons with chronic conditions, the preferences of persons who had those conditions—not the values of the general public—should be used.

The issue of whose preferences should be measured is important because several studies have shown that patients' preferences can differ substantially from those of healthy persons, health care providers, and other groups (63,75a,387). For policy decisions that involve broad resource allocation, there is a strong intellectual argument that general public values are the relevant ones. The experience with Oregon's waiver proposal, however, suggests that the public itself is not entirely comfortable with the implications of this approach.

Distributional Considerations

The purpose of CUA in health care, like any CEA, is to improve decisionmaking that involves the allocation of health care resources. Unlike other CEAs, CUA actually incorporates values about health care outcomes into the quantitative part of the analysis. CUA's ability to assign numbers to the relative worth of very different health care activities is one of the attributes that makes this analytic tool attractive for policymaking. In doing so, however, it **introduces two dangers: decisionmakers may not fully understand the implications of the values that CUA incorporates; and they may not fully realize that although some values are quantified and incorporated into CUA results, other equally important values are not.**

The issue of whose utilities are represented in the CUA is one example of the importance to policymakers of understanding the assumptions behind it. Like any aggregate measure, the use of utilities in CUA implicitly assumes that it is average group values, and not individual values, that are important. Whereas the measured average utilities of a group (using a particular measurement instrument) tend to be quite consistent for any given question (763), measurements among individuals can vary enormously (667), and the value for

any one individual may be very different from the average value used to calculate QALYs.

This assumption probably makes sense for most policy decisions, because they involve the allocation of resources among groups. Nonetheless. it means that there are potentially large numbers of individuals who will disagree strongly with the decisions and their consequences, because the average group values do not represent their views. It is even conceivable that in some cases the number of individuals for whom average values are not representative will exceed those for whom they are.

Even if the disagreeing individuals are in the minority. their opinions may be of concern to public policy makers. For example, because CUA specifically assumes that quality as well as length of life matters, it places relatively less value on an extremely "low-quality" life (e.g., someone with only lower brainstem functioning) than on one of higher functioning. Individuals who strongly believe in the prolongation of life for life's sake, therefore. may sometimes find themselves at odds with the policy implications of some CUA results.

In addition to assumptions about whose values are important in distributing resources, the utility measurements underlying CUA tend to assume that relieving very severe distress that is very temporary is much less important than relieving more prolonged distress. CUA also assumes that "for a given degree of suffering, those whose illnesses happen to be cheaper to treat will be treated in preference to those whose treatments are more expensive" (623). Both of these assumptions are intended consequences of the principles of CUA as a method of allocating resources fairly. Nonetheless, there may be circumstances under which these assumptions are of concern. and they are certainly assumptions of which the users of CUA should be aware.

The most important point about the distributional assumptions of CUA, however, is what it does not assume. Because QALYs and other similar measures assume that all healthy lives are of the same value, they ignore social issues of distribution. The calculations in a CUA do not take into consideration, for example. whether benefits are being received by one subgroup (e.g., people in Nevada), while costs are borne by another subgroup (e.g., people in Connecticut). Nor does CUA account for the fact that society may *want* to value certain lives over others for certain purposes+. g., the sick over the healthy or the poor over the wealthy. Consequent] y. **these social decisions must be made outside the quantitative framework of the CUA.** There may be many times where, no matter how "valid" and high-quality the analytic results, policy makers will want to deliberate] y choose an intervention with a high cost-per-QALY over an intervention with a low cost-per-QALY, because the costlier intervention redresses social imbalances or achieves other social goals.

The importance of other social considerations is evident from a study. conducted in Norway, that tested whether health care priorities implied from conventional methods of deriving information on individual utilities actually conform to directly obtained priorities. The study is very small. based on a nonrandom sample of participants, and not easily generalizable, but it also raises some very unsettling questions about apply in: the assumptions of CUA in social decisionmaking.

In this study, when individuals were asked to prioritize between two hypothetical patients being admitted to a hospital, the great majority of respondents responded that priority should be assigned by order of admittance, regardless of differences in the likely health or disability of the patients after treatment (569). The author points out that this response differs from the implications of some American research (589), and acknowledges that cultural preferences may account for the difference. Nonetheless. the study challenges the concept that results from CUA are useful for deriving priorities for social policy. It suggests that more direct tests of the relationship between CUA implications and directly derived priorities is warranted before CUA results, even when combined with additional considerations of social issues, can be used to infer social preferences among alternatives with confidence.

FEDERAL COST-EFFECTIVENESS ACTIVITIES

The federal government sponsors a number of activities relating to CEA and development of related techniques, but in light of the growing interest in the technique among private and public policy makers, the level of this activity is surprisingly small. Much of it, in fact, relates to effectiveness research—the development of underlying tools and data (e.g., decision analysis and health measurement instruments)—rather than the sponsorship of specific analyses.

Agency for Health Care Policy and Research

The Agency for Health Care Policy and Research (AHCPR) inherited a long history of collecting and analyzing data on the cost of health care, a legacy of the health services research focus of its predecessor agency (the National Center for Health Services Research).[7] Most AHCPR activities related to CEA have been funding efforts to advance the underlying methods and tools, especially outcomes, quality of life, health status and patient utility measures, and methods of data acquisition (596). Sponsoring CEAs themselves has not been a particularly significant part of AHCPR's activities.

The PORT projects have been one of the most visible mechanisms for relevant research. All PORTS have undertaken some research related to measuring costs and health outcomes, and several are apparently conducting formal CEAs.

Another vehicle has been AHCPR's recent $14.5 million initiative to fund studies of the outcomes of existing pharmaceutical therapies (522).[8] Several of these projects are examining comparative effectiveness or costs, although only one study is a full CEA (on the cost-effectiveness of special pharmacist-based counseling for asthma patients, compared with usual care) (596).

AHCPR also funds a number of relevant, smaller independent studies. Again, a number of these are methodological in nature; examples include a multimedia-based method to assess patient preferences for different health states and treatments, and a "Longitudinal Comparisons of Measures for Health Outcomes" project. A few, however, specifically include analyses of the costs and effectiveness of particular interventions-e.g., a CEA of gallstone lithotripsy, and the cost-effectiveness of community-based care for elderly persons (817).

The 1992 act reauthorizing AHCPR (Public Law 102-410) made two significant changes that were intended to affect the level at which the agency sponsored CEAs or included CEAs done elsewhere in its own assessments. First, the act required that the individual technology assessments produced in the Office of Health Technology Assessment (OHTA) must include CEAs wherever valid data exist to support such analyses. Second, the act required that in producing clinical practice guidelines, AHCPR must consider the cost of alternative medical practices being addressed in the guideline. One consequence of these requirements is that to meet them, the agency must develop greater expertise in cost-effectiveness techniques. which might lead to greater interest in supporting research on the topic.

National Institutes of Health

On the whole, National Institutes of Health (NIH) appears to view itself as a source of new technologies and information on the efficacy of those technologies, with little reason to be involved in studies of the costs or cost-effectiveness of those technologies. At an agencywide level, the main departure from this stance is a small publication

[7] At the time it was abolished, the full name of this agency was the National Center for Health Services Research and Health Care Technology Assessment. Here and elsewhere, it is still referred to by the shorter title for simplicity's sake.

[8] The $14.5 million represents the costs, for all years, of the projects funded in the first year of this initiative.

whose goal is to show that NIH research can sometimes lead to reduced costs to the health care system (842).

As in many other ways, however, there is substantial variation among Institutes in the perceived relevance of this topic. The National Cancer Institute in particular has been active in both conducting analyses and developing and critiquing analytic methods. Its efforts include analyses of the cost-effectiveness of prostate cancer screening, the development of a detailed cost-effectiveness model for cancer screening generally, and a series of activities relating to the cost-effectiveness of mammography screening for breast cancer (846). NCI also recently sponsored a conference on economic evaluation during clinical trials (85 1).

The National Institute of Mental Health is notable for the number of studies it sponsors in which reimbursement is an issue+. g., the impact of differing reimbursement mechanisms for intensive case management. NIMH also sponsors a number of studies that examine the relative costs and effectiveness of alternative interventions, ranging from studies of pharmaceutical interventions to studies of community treatment programs (570,846).

At least two other Institutes, the National Institute on Aging and the National Heart, Lung. and Blood Institute, also have intramural and extramural experts in economic analyses with whom they can consult. Few other institutes, however, have more than three or four ongoing studies in which measuring costs and determining relative cost-effectiveness is a major focus (846).[9] For many of those cost-related studies that are being conducted, a motivating force for measuring costs appears to be the desire to show that the NIH -sponsored research will ultimately result in reduced treatment costs for the condition of interest.

NIH is sponsoring a few trials in which economic data collection is performed during the trial itself, although the funding for the economic evaluation has come from elsewhere (e. g., foundations and AHCPR) (61 2).

Centers for Disease Control and Prevention

Many of the pragmatic analyses and applications of CEA are done at the Centers for Disease Control and Prevention (CDC). in the context of increasing the “value for money" of the Nation’s public health programs and interventions. CDC has sponsored a few such studies off and on for at least a decade (e.g., 129,347). but the level of activity has expanded in the past two years with the agency's “prevention effectiveness” activity.

The CDC prevention effectiveness effort is explicitly designed to increase the incorporation of economic analyses into agency decisionmaking (749). In support of this effort, the agency is currently developing a basic, but fairly detailed, resource “how-to” manual on decision analysis, cost measurement, and CEA (831). The manual is being produced in-house and is expected to be published in 1994 (752).

CDC’s economic analysis activity is not large. but it is more extensive and more integrated with agency activities than similar activities at either AHCPR or NIH. It is also unusual in that a number of its economic evaluations are actually cost-benefit analyses, with benefits measured in dollars rather than the more common CEAs. Examples of ongoing and recent intramural analyses, which are spread throughout the various centers, include:

- a cost-benefit analysis of strategies to prevent nosocomial legionellosis (Legionnaire Disease) (836a),

[9] This observation i\ based on responses to a letter sent by OTA to the various NIH institutes during the course of this study (see appendix A).

- a cost-effectiveness analysis of folic acid food fortification and supplementation (836a),
- a review of economic evaluations of HIV 10 prevention and treatment programs (836a),
- the comparative costs and benefits of testing and counseling services for HIV-infected patients (353), and
- the cost-effectiveness of different strategies to prevent streptococcal infections in newborns (537).

In addition, CDC sponsors a number of extramural studies.

At least one explanation for CDC's relative interest in economic evaluation is that its programs face clearly defined limited resources, and they often have goals measurable through population-based health outcome measures (750). Cost-effectiveness and cost-benefit analyses thus can be directly useful to decisions that involve finding the best way to achieve a particular program's goals. Even so, its activities are not extensive.

Health Care Financing Administration

The Health Care Financing Administration's (HCFA) primary purpose is to administer the Medicare and Medicaid programs, providing health insurance for elderly, poor, and disabled persons. Its research activities have tended to focus around payment issues. It is clearly a major potential consumer of the research and assessments of effectiveness and cost-effectiveness done by others, particularly AHCPR. Its reliance on AHCPR for these evaluations results in somewhat of a conflict regarding cost issues, however, since on its part AHCPR does not want to be viewed as in HCFA's "pocket."

Consistent with its mission as a major payer of health care services, therefore, many of HCFA's studies of effectiveness and care patterns include a strong emphasis on measuring costs. Even more notable, of the 10 HCFA condition- or technology-oriented studies in its 1992 research status report that most obviously include both a cost and an effectiveness component, seven are on the topic of preventive services. These seven projects represent nearly $20 million in research (796).

The largest of these seven projects, which accounts for half of the $20 million, actually comprises a multiyear experiment being conducted at five different sites (and a cross-cutting evaluation of the experiment). At each of these sites, study participants (all of whom are elderly persons) are randomized to experimental and control groups in which the experimental group receives a comprehensive set of preventive services. Both the costs of providing these services and their effects on participants' health status are measured (796).

HCFA's emphasis on the cost-effectiveness of preventive services is deliberate. The original Medicare statute provided only for coverage for acute care. Services such as preventive care, outpatient drugs, and long-term care were specifically excluded. Congress has been interested in extending Medicare coverage to more preventive services, and over time a few (e.g., mammography screening) have been added. Interest in adding to this short list is manifested in legislation that requires HCFA to study technologies such as influenza vaccine and comprehensive screening. Because of concern over the continued increase in Medicare expenditures, preventive services are held to a higher standard—a demonstration that they are not only effective but, by some measure, cost-effective—than are new diagnostic and therapeutic services that need pass only administrative scrutiny to obtain coverage.

HCFA has long contemplated establishing regulations that would permit it to use cost-effectiveness as a criteria more explicitly in decisions regarding coverage for any technologies and services. The agency first considered drafting regulations on the topic in 1979 (708), and actually did publish proposed regulations a decade later, in 1989 (54 FR 4302). Agency anxiety about the re-

10 HIV is the human immunodeficiency virus, which causes AIDS (acquired immunodeficiency syndrome).

ception of the regulations, however. has delayed their being made final. At this point agency staff state that new proposed regulations would have to be published for public comment before they can go forward, an action unlikely to happen until the likely shape of health reform is clearer (708).

Office of Disease Prevention and Health Promotion

A significant effort to further cost-effectiveness techniques is being sponsored by the Office of Disease Prevention and Health Promotion (ODPHP), a small office under the Assistant Secretary for Health whose role is to coordinate and augment the Department of Health and Human Service's prevention activities. One of the major activities of this office was the establishment of a Preventive Services Task Force, which was charged with developing recommendations for clinical preventive services.

More recently, ODPHP has convened a Cost-Effectiveness Panel on Clinical Preventive Services to complement the work of the Task Force. Cost-effectiveness is an issue in determining both whether to recommend a screening test for a population and in the periodicity of screening. the frequency with which an individual should be screened for a particular disease. The Cost-Effectiveness Panel first met in 1993. It is expected to produce a report in 1995 that will attempt to advance the state of the literature on cost-effectiveness of clinical preventive services. address methodological issues and provide a framework for consistency in methodological approaches, and stimulate the application of CEAs.

Despite its 1 ink with the work of the Task Force, the recommendations of the Panel may not find their way immediately into the guidelines deliberated by the Task Force, due to hesitancy among Task Force members about the possible perception in their audience that recommendations would be too heavily influenced by cost (868). Nonetheless, the Task Force is including a short chapter on cost-effectiveness in its forthcoming update of its preventive services guidelines (697).

ODPHP has also convened an interagency discussion group to facilitate coordination and information flow across staff in the different agencies, such as NIH and CDC, that do at least some CEAs. [11] Persons from these agencies also attend the Cost-Effectiveness Panel meetings.

COST-EFFECTIVENESS ANALYSIS IN THE PRIVATE SECTOR

Recent Trends

A notable trend, associated with both the intellectual growth of effectiveness research and the growth in the private demand for health technology assessment' has been the increase in evidence on cost-effectiveness being produced by. or on behalf of, health product manufacturers, particularly pharmaceutical manufacturers. In 1992, for examplc, 27 out of 30 pharmaceutical firms participating in a survey said that they had begun to study "outcomes" (953). In a survey a year later, the companies reported that they were expanding ● 'outcomes research*' staff and doubling the number of economic studies (632). Pharmaceutical manufacturers are major clients of a number of technology assessment consulting firms.

Evidence of this increasing interest is apparent from the literature as well. Articles examining the economics of health care technologies have been appearing with increasing frequency since the mid- 1960s, but the number has mushroomed in the last several years, particularly for pharmaceuticals (215,475). In fact, 1992 saw the introduction of a new journal. *Pharmoeconomics*. devoted specifically to studies of the costs and cost-effectiveness of pharmaceuticals.

In addition to their use of academic consultants and private technology assessment consulting

[11] The full list of agencies and offices represented on the interagency discussion group comprises AHCPR, CDC, FDA, HCFA, Health Resources and Serv Ice\ Administration, NI H (Including several indivi dual institutes), ODPHP, Office of the Assistant Secretary for Health, Office of the Assistant Secretary for Planning and Evaluation, and Substance Abuse and Mental Health Sin-\ ices Administration.

firms, some pharmaceutical companies have entire in-house divisions that conduct economic analyses of their products. Glaxo Inc., for example, established its pharmacoeconomic research group, with a staff of 19, in 1988. Eli Lilly similar-1 y has a staff of over 20 persons and has been conducting cost-effectiveness studies for about 4 years (755).

The major impetus for assembling these resources and studies appears to be their perceived usefulness as a tool for marketing products to cost-conscious consumers. In OTA conversations with manufacturers and others involved in the field, the prevalence of controlled formularies at hospitals and in managed-care organizations was one of the most frequently mentioned reasons for the proliferation of cost-effectiveness research of pharmaceuticals (174,509,473a). A recent detailed study of technology evaluation by health care providers found hospital formulary committees to be relatively sophisticated in the kinds of evaluations of pharmaceuticals they performed, sometimes assessing not only information on patient outcomes and drug costs but total related hospital costs as well (474). In studies conducted by the American Society of Hospital Pharmacists, 58.4 percent of hospitals reported a well-controlled formulary system in 1989, up from 53.9 percent in 1985 (789). Of HMOs, 22 to 55 percent in a recent survey reported having some sort of formulary (789). And market observers clearly expect the trend towards more control over the roster of drugs that doctors can prescribe to continue (56).

In addition, manufacturers are at least sometimes employing cost-effectiveness research to make internal decisions regarding the allocation of research and development dollars and to decide whether to continue to pursue the development of a particular drug or class of drugs (473a,755). In some cases, CEAs of a drug are done as early as phase II clinical trials, while the safety and efficacy of the drug itself is still in question.

The interest in conducting cost-effectiveness studies is evident in the medical device manufacturing industry as well, although to a lesser extent. Some companies regularly conduct CEAs, particularly of high cost, highly visible or controversial devices (174,461 ,672). But others believe that rapid technology obsolescence in a device product line often makes cost-effectiveness data irrelevant (473a,755). In addition, hospitals do not generally consider comparative cost-effectiveness when making technology purchase decisions, limiting their evaluations to simpler financial assessments (474). Thus, the perceived importance of CEA seems to depend heavily on the type and anticipated lifespan of the device.

Issues

The increasing production and use of CEA in the private sector has several implications for federal regulators, planners, and health insurers. Most critical are the questions of whether those analyses will be of good quality; whether they will be comparable; and whether they will be either intentionally or unintentionally biased towards the result that the sponsor of the analysis (e.g., the product manufacturer) would find most favorable.

At the moment, there is no particular reason to believe that the quality of CEA performed by the private sector is necessarily any different from government or foundation research. Indeed. the same consulting firms and academics frequently contract with both private companies and public entities. Nonetheless, there is considerable concern that private sponsors might choose the enumeration of resources or costs in a CEA to increase the chances of obtaining a desired result (e.g., by including indirect costs only when doing so would increase the chances that the sponsor's product would appear to be more cost-effective). This possibility not only would bias results but might make studies across different sponsors incomparable. In addition, reviewers examining CEAs published in the medical literature could face substantial publication bias, since unfavorable privately sponsored studies might not be submitted for publication (particularly if they were performed in-house).

The concerns about framing studies to increase the probability of obtaining the desired result, and

the potential tendency to publish only favorable studies, are not entirely limited to the private sector. Some analysts point out that public organizations are not immune from these temptations, either.

One response to these issues has been increased attention, in the United States and elsewhere, towards making qua] it y standards for CEA more explicit and more standardized. The federal government is sponsoring one such effort in the area of prevention, the panel supported by ODPHP described above. A privately sponsored effort is also ongoing, based at the Leonard Davis Institute at the University of Pennsylvania (746a). A primary goal of that group is to devise ethical guidelines to minimize the potential for bias. The group held its first meeting in July 1993. Several international efforts to standardize CEA are described in box 5-3.

Because manufacturers who sponsor CEAs do so in great part to aid in marketing, this activity also affects the regulatory responsibilities of the Food and Drug Administration (FDA). The FDA oversees prescription drug advertising and labeling. A group at FDA meets weekly or semi-weekly to review cost-related claims in drug advertising and to examine what substantiation exists for the claims. When they see a questionable claim in advertising or detailing material, the y request supporting data from the manufacturer (539).

Whereas marketing claims regarding simple comparative cost are generally straightforward—FDA simpl y requires the source for the claim to be cited (e.g., a widely available list of drug average wholesale prices)--claims of cost-effectiveness arc considered claims about relative effectiveness and held to a more rigorous standard (539). FDA staff report that cost-effectiveness claims are on the increase (539). To address the need to evaluate these claims, the agency has been developing more expertise in the area and is developing a series of guidance documents to outline for companies what is necessary to support a cost-effectiveness claim (539).

CONCLUSIONS

Analyses that consider the costs and health effects of an intervention in a structured fashion can improve decisionmaking, and greater use of CEA in this context appears to be on the horizon. Although formal CEAs are still not in abundance. the number of analyses on health care topics has grown considerable y over the past decade. Similarly, CEAs are still not routinely applied to most health care decisions, but the sponsorship, the use, and the interest in these analyses has been increasing rapidly.

As the use of CEA increases, attention to the validity and comparability of analyses becomes doubly important. Despite the greater number of analyses being produced now than in the past, it is not clear that the overall quality of analyses has improved. Furthermore, **inconsistencies among analyses in the approaches and assumptions they use will confuse policy makers and hinder the practical use of CEA. U.S. and international efforts to address this issue deserve attention and support.**

The greatest methodological change in the field in recent years has been the increasing prominence and sophistication of cost-utility analysis. The great attractions of CUA are that it incorporates a broader range of relevant health outcomes than simply lives saved, cases of disease averted, or other single measures; and that it quantifies these outcomes so that interventions with greatly differing purposes can be compared directly. Consequently. it has immense potential appeal to policy makers who are making allocation decisions across broad areas of health care.

The greatest danger of CUA is that it incorporates so much into a single measure, such as the QALY, that policy makers may be unaware or less attentive to the many issues that CUA does not address. CUA, for example, does not allow for the fact that society is not always indifferent to which groups benefit and which do not: an intervention that looks the most positive when

BOX 5-3: Standardizing Cost-Effectiveness Analyses: International Activities

The uneven quality and Inconsistencies across cost-effectiveness analyses (CEAs) have led to a call for greater attention to standardizing cost-effectiveness methods (1 85) In fact, an international movement is underway to begin to do so in at least four countries and communities

AUSTRALIA. Under Australia's health insurance system, an independent committee makes recommendations to the government regarding coverage for new drugs In 1987, the committee was directed to begin considering costs as well as effectiveness in its decisions In 1990, the committee issued draft guidelines to encourage (and, ultimately, mandate) manufacturers to present data on the cost-effectiveness of their products when applying to have their products covered. To Increase the comparability of these cost-effectiveness data, the guidelines set some explicit standards that analyses must meet to be acceptable These include

- standard values to be used for certain health care units (e.g , physicians visits),
- discouraging the use of redirect costs in the analysis (e g., lost work productivity),
- permitting the use of some direct nonmedical costs (e. g , costs of home help) as well as medical costs,
- encouraging the use of outcome measures such as Improvements in functioning, and
- encouraging the use of cost-utility analysis where possible

Manufacturers are to submit the results as incremental cost-effectiveness comparisons of the new drug with either the alternate drug most widely used in Australia, or to the alternate drug the new pharmaceutical would replace (39, 182),

CANADA. The provincial government of Ontario and the Canadian national government are in the process of developing guidelines similar to those in Australia. The Canadian guidelines, however, are not regulations Rather, they are an effort to develop "consistent and uniformly understood principles, definitions, and methodologies for the conduct and evaluation of economic analyses. " These suggested standards are intended to assist drug manufacturers in meeting the growing demand on the part of provincial drug programs for cost-effectiveness information (1 60) The authors stress that the guidelines are to be flexibly applied. Some of the recommendations of the draft guidelines are.

- all "relevant" costs should be included (indirect as well as direct costs where appropriate),
- the analysis should state clearly its perspective (e.g., societal vs. insurer), and two separate analyses from different perspectives may be appropriate,
- outcomes are to be expressed first in natural units (e g. myocardial Infarctions avoided) and also in alternate units such as benefits (dollars) or utility (using, for instance, quality-adjusted life years (QALYs)),

measured by cost per QALY may in fact not always be the best allocation of social resources when concerns such as distributive justice are taken into account. Nor does it address the question of whose values should matter the most for particular decisions: it treats all values as the social average. The issue of whether a particular demographic minority's values should matter more than average public values, or the subpopulation from whom the values in the analysis were derived, in a particular case is a decision that is ultimately a political one.

BOX 5-3 continued: Standardizing Cost-Effectiveness Analyses: International Activities

- an explanation of equity assumptions should be included (e g , whether QALYs gained by all Individuals were considered equal), and
- the Incremental cost-effectiveness of one drug relative to another is to be expressed as a ratio of the cost to the outcomes (180,578a)

SPAIN. In recognition of a growing desire to use CEAs in resource allocation decisions in Spain, a group of researchers, working with the government, have proposed a set of standardized approaches on seven topics The topics covered by their working document are (659)

- selection of alternatives being compared,
- the Inclusion of direct vs Indirect and intangible costs
- the analysis,
- methods for valuing costs,
- measures of health effects used,
- time horizon and discounting,
- treatmeny of risk and uncertainty,
- range of sensltivity analyses, and
- presentation of the results

The goal of this group is maximum standardization, particularly in areas such as the range of assumptions tested in sensitivity analyses and the discount rates used. The effort was not complete as of 1993 but agreement on standardization for many of the topics was well underway (657)

THE EUROPEAN UNION (EU). Another group of researchers, including the head of the Spanish group, have proposed an undertaking similar to the Spanish one but on a larger scale This effort is only just underway, and funding (by the EU) was still tentative as of 1993 The goal of the joint European effort is an ambitious one: "to develop and propose a unified methodology for the economic evaluation of health technologies to be adopted by EU regulatory agencies, national administrations and European multinational companies operating in the health care field with the eventual aim of harmonizing regulatory practices across EU countries (658) The proponents of the proposal argue that standardizing methods across countries will benefit those carrying out studies and will improve the transferability of results of studies conducted in one country to other countries where decisionmakers face similar issues

SOURCE Off Ice of Technology Assessment 1994, based on sources as shown Full citations are at the end of the report

CUA also incorporates a number of implicit assumptions of which its users should be aware. Most fundamentally, it assumes that the values elicited from people about health in surveys translate into valid representations of their preferences for different interventions or resource allocations. The one direct test of this assumption suggests that it may be flawed (569).

The other very significant change in cost-effec - tivencss methodology is the growing practice of conducting CEAs simultaneously with early clin ical trials of a new treatment efficacy and safety. This practice raises questions about the sample size needed for the economic component of these clinical-economic trials, and whether such studies may be biased towards finding no difference in costs between treatments. More fundamentally, these trials raise familiar issues of generalizability: the cost results derived from an efficacy trial

may not be applicable outside of the trial, in ordinary practice.

Despite these potential concerns, cost-effectiveness studies and related activity in the private sector has boomed. Spurred by the need to deal with an increasingly sophisticated cadre of managed-care administrators who are very cost-conscious, private industry has begun putting significant resources into efforts to show that its products are not only clinically effective but cost-effective. The pharmaceutical industry in particular has become very active in sponsoring CEAs of its new products. **To the extent that the results of these analyses are used in marketing claims, both purchasers (e.g., government and private insurance programs) and the FDA will need to become increasingly sophisticated at evaluating such claims.**

Given the growing level of interest among private and public policy makers alike in CEA, the **federal government's level of activity related to CEA is surprisingly weak.** Only in the area of preventive services is there any substantial investment. Although there is some overlap in activities that warrants close communication and coordination among agencies, ODPHP is well positioned to play this role and to support it with methodological work, as it is doing. AHCPR also supports some relevant methodological research, but in general CEA related to treatment and long-term management has been given relatively little attention by federal agencies.

Although there is still no uniform agreement among policy makers or the public about what role information regarding the cost-effectiveness of treatments should play in insurance coverage decisions, it is possible that there may be more agreement on this point in the near future. At present, federal agencies are not well positioned to support this research, through either their in-house expertise or their current sponsorship of methodological studies.

Perhaps the most ambitious endeavor regarding the use of cost-effectiveness information in health policymaking to date has been Oregon's attempt to use this information as a foundation for creating an entire health benefits program for the state's Medicaid program. The state's attempt to rank all primary and acute care services according to their importance, costs, and effectiveness gained national attention and spawned a furious debate over the ethics of the process (735). The ultimate reliance of Oregon on the opinions and judgments of its appointed commissioners to value services, and the lack of solid data to assist them in making their decisions (788), was a blunt reminder that the use of CEA in health policy decisionmaking has limitations.