ADVANCES IN COST-EFFECTIVENESS
ANALYSIS OF HEALTH INTERVENTIONS

Alan M. Garber

ABSTRACT

The growing application of cost-effectiveness (CE) analysis and controversies about its methods has led to a need to explore its welfare economic foundations. Examination of its welfare theoretic foundations can provide a rationale for selecting specific standards for the application of CE analysis while deepening our understanding of the implications of alternative methodological approaches.

In this paper, I explore conditions under which decision making based on CE analysis, carried out a specific way, leads to a distribution of resources that has desirable social welfare properties. The first section describes the basics of CE analysis and how it can be applied to aid decisions about the allocation of health resources. The paper then turns to the potential welfare economic foundations of CE analysis, and addresses specific issues in carrying out CE analysis, such as which costs to include, whose perspective matters in the analysis, and how health outcomes are measured. It demonstrates how a welfare economic foundation can help resolve ambiguities and uncertainties about the application of CE analysis. The paper also discusses the limitations of such an approach, which indeed reflect limitations of CE analysis as an analytic framework. Finally, it addresses unresolved issues such as the difficulties in using the results of CE analysis to make health policy at the societal or group level.

Alan M. Garber
Center for Health Policy and Center for Primary Care and Outcomes Research
179 Encina Commons
Stanford University
Stanford, California 94305-8006
and NBER
garber@stanford.edu

1. Introduction

This paper discusses the welfare economic foundations of cost-effectiveness (CE) analysis. Although it is not a comprehensive review of the techniques of CE analysis, the paper addresses application as well as theory because the welfare economic properties of decisions based on CE analysis necessarily depend upon the way that the method is applied. In fact, application has stimulated much of the interest in the theoretical foundations of CE analysis. As government officials, private insurers, health care providers, and others have begun to use CE analysis to inform decisions about the adoption and allocation of specific health interventions, they have revealed the need to improve and standardize its methods.

There is no doubt that CE analysis is potentially useful: by quantifying the tradeoffs between resources consumed and health outcomes achieved with the use of specific interventions, the technique can help physicians, health plans, insurers, government agencies, and individuals to prioritize services and to allocate health care resources. CE analysis aids such decisions by structuring comparisons among alternative interventions. Meaningful comparisons, in turn, require standardization. Without standardization, there can be no assurance that the results of a CE analysis of one set of interventions will be comparable to the results of a study of a different set of interventions. Thus the method must be valid, and it must be applied consistently. Perhaps the most important contribution of an examination of welfare economic foundations is that it can help ensure that any set of standards adopted for CE analysis will be logically consistent, valid, and credible.

Several efforts around the world have sought to move the field of CE analysis forward by strengthening the methodology and promoting standardization. Among these are various governmental guidelines (such as Australian pharmacoeconomic guidelines and those of Ontario), the European Community Concerted Action on the Harmonisation of the Methodology for Economic Evaluation of Health Technology (HARMET), and the Panel on Cost-Effectiveness in Health and Medicine. The last group, sponsored by the U.S. Department of Health and Human Services, issued a comprehensive report in 1996 detailing recommendations for the application of CE analysis (Gold, et al. (1996)). The report distinguished between recommendations that had a strong theoretical justification and those that had no firm theoretical grounding, but were made to

ensure uniformity, usually based in part upon ease of implementation and other practical considerations.

The advantages of methodological standardization in CE analysis are greatest when the standards are selected with both rigor and transparency. To the extent that standards are chosen arbitrarily, they merely ensure that diverse studies will use consistent — but potentially invalid and misleading — methods. To develop recommendations that could be justified from first principles, the Panel on Cost-Effectiveness in Health and Medicine drew upon recent work on the welfare economic foundations of CE analysis. Since CE analysis evolved largely outside the framework of welfare economics, an exploration of the welfare economic foundations neither recapitulates nor parallels the history of the development of the approach. Yet by relating CE analysis to theoretical foundations it is possible to illuminate the consequences of alternative methodological practices. For example, there has been a longstanding controversy about future costs of health care: Should costs that result solely from living longer, but otherwise are not directly influenced by an intervention, be attributed to that intervention? Some investigators, such as Weinstein and Stason (1977), have recommended always including such "unrelated" future costs of care while others, such as Russell (1986), have urged the opposite. Presumably one of these practices is incorrect, and the persistence of two distinct practices renders the results of different studies non-comparable. Other methodological controversies are no easier to resolve, such as whether to incorporate time costs as dollar costs (hence part of the numerator of the CE ratio), or as a reduction in the health outcome like years of life (in the denominator). In cases such as these, which are discussed below, methodological standardization offers the prospect of replacing a set of inconsistent practices with a single correct method.

An exploration of the welfare theoretic foundations for CE analysis can provide a rationale for selecting specific standards while deepening our understanding of the implications of alternative methodological approaches. However, few attempts to explore the theoretical foundations of CE analysis have been published. Both proponents and critics of CE analysis have been skeptical of the value of some of the traditional standards of welfare economics, at least when applied to health care. To many economists, the forms of market failure common in health care supply much of the rationale for applying a tool like CE analysis or CB analysis. But others are skeptical of the premises and conclusions of welfare economics more generally, and see CE analysis as a method to make policy decisions when market outcomes are unacceptable.

Some proponents of CE analysis have adopted an "extra-welfarist" perspective, arguing that there are fundamental justifications for pursuing CE analysis without reference to welfare economics (see Hurley (1999)). The assumptions and, some would argue, the values underlying this perspective can be more general than under the typical welfare economic perspective. Proponents of the extra-welfarist perspective claim that improvement of health is a primary goal of social policy, a goal whose value is self-evident and does not depend upon the maximization of individual utility functions. They do not necessarily accept the arguments of social welfare (e.g., the prominence of individual consumption of goods and services) that are typical in formulations proposed by economists, nor do they accept the typical assumptions made. For extra-welfarists, CE analysis offers a way for a social decision maker to learn how to obtain the greatest health effect from a specified expenditure, or to find the lowest-cost approach to achieve a given health effect. It is unnecessary to ask whether an allocation based on CE analysis leads to a potential Pareto improvement or a Pareto-optimal distribution.

Although this perspective makes it possible to analyze the optimal allocation of health resources without accepting the full range of welfare economic assumptions, it has other limitations. By eschewing any claim to justification on the basis of a more fundamental framework, the extra-welfarist perspective requires acceptance of the principle that maximizing quality-adjusted life years or another specific health outcome measure should be the goal of health care provision. Acceptance of a specific measure is much more problematic than accepting the general concept that improvement in health is a social good. Results from a study using QALYs as the health measure may differ from those that measure health in terms of longevity. Usually, the validity of the health outcome measure must be assumed rather than tested. The extra-welfarist approach can determine the best measure of health outcomes by appeal to political processes. But to the extent that it rejects market and personal valuations of health improvements, the extra-welfarist approach cannot appeal to a more fundamental set of principles to resolve whether one measure of health outcomes is more valid than another. Nor is it easy to use this approach to evaluate tradeoffs between health and other social goods, such as education, nutrition, or other aspects of well-being. Finally, it provides no direct mechanism for resolving certain economic issues — such as what constitutes a cost, and how cost should be measured.

In contrast to the extra-welfarist perspective, this paper uses a welfare economic framework to address questions of standardization. The fundamental question underlying our approach is

simple: does decision making based on CE analysis, carried out a specific way, lead to a distribution of resources that has desirable social welfare properties? In other words, does a ranking of alternative uses of health resources based on CE analysis lead to an allocation that improves welfare? The answer depends on the way that CE analysis is performed, the way the results are used, and the definition of *social welfare improvement*.

To economists familiar with cost-benefit (CB) analysis, these questions imply another: Why perform CE analysis, rather than CB analysis, whose economic foundations and social welfare implications are well known? In some circumstances they appear to give nearly equivalent results (Phelps and Mushlin (1991)). However, in principle, CB analysis is more general than CE analysis (Kenkel (1997)). Furthermore, CE and CB analysis grew from different historical traditions and have been adopted for different reasons. CB analysis requires placing dollar valuations on the outcomes of any program or intervention. In the context of health and medical care, making that valuation can be equivalent to placing a dollar value on a human life (or, more precisely, on changes in the probability distribution of the length or quality of human life). To many in the worlds of medicine and of public health, any attempt to place a value on a human life — even if it is usually a valuation of a small change in the probability of death or a change in the distribution of expected mortality, rather than an attempt to put a price on an identified individual's life (Schelling (1968)) — is anathema. Thus most "economic" evaluations in health care have applied CE analysis, which limits the analyst's responsibility to providing information about the efficiency with which alternative strategies achieve health effects. The often implicit task of placing monetary valuations on health outcomes falls upon decisionmakers and others who read the analyses.

The fundamental differences between the techniques may also reflect the contexts in which they developed. CB analysis was developed primarily to assist in making decisions about the provision of public goods. Although CE analysis has also been used to evaluate public health measures that are public goods or create externalities (e.g., vaccination programs), it is more often used for the evaluation of private goods and services. The reason to apply formal analysis in this context is that information in health care is imperfect and often asymmetric. Asymmetry is common because the producers of health care, consumers, and payers possess different amounts of information about the benefits, risks, costs, and other characteristics of health services. Although limited and asymmetric information is an issue in some contexts in which CB analysis has been applied, nonexcludability and nonrivalry in consumption are the forms of market failure chiefly

responsible for the popularity of CB analysis. CE analysis, in contrast, assists patients and their agents in making decisions about health care, which is generally a private good (with some notable exceptions, such as infectious disease control). Although the primary function of insurance is risk-spreading, health insurers reimburse for services used rather than making lump sum payments. Consequently, a health insurer should also assure that optimality is achieved in health care consumption by designing coverage and reimbursement so that the marginal utilities of health care dollars are equated across patients and interventions.

Information provided by CE analysis is important in two ways: First, health care is valued insofar as it improves health and well-being, not for intrinsic characteristics of the health services. The relationship between the use of a medical intervention and improved health outcomes may not be known to the individual patient or physician. CE analysis can reveal how much value the patient will obtain for a given expenditure on a health intervention. Second, as Pauly (1968) has noted, nearly all forms of health insurance are subject to moral hazard. Once an enrolled individual has a disease or other health condition, he or she would prefer to consume it to the point at which the marginal benefit equals the marginal cost to his or her patient. Because insurance lowers the patient's share to a small fraction of the full marginal cost (the fraction usually determined by a fixed usage fee, percentage copayment, or deductible), insurance ordinarily results in overconsumption. *Ex ante*, an individual would prefer actuarially fair insurance which guaranteed that care would be provided to the point at which marginal cost (insurance payment and copayment combined) equalled marginal benefit over insurance that was subject to moral hazard. Use of CE analysis to allocate care (usually based on coverage decisions) might help limit moral hazard by overcoming informational limitations.

In theory, the use of CE analysis to address moral hazard is straightforward. Consider a world of (near) perfect information. That is, effectiveness and costs of treatment are known, but information is not sufficiently inexpensive to enable insurers to monitor and overcome moral hazard. What would the ideal health insurance plan attempt to do? Risk-averse individuals desire insurance for the usual reasons. They might also want the insurer to act as their agent in deciding how much and what kinds of health care each should receive (or equivalently, the enrollees would commit to accept levels and types of care that met a net benefit criterion as long as the premiums were actuarially fair). Assume further that every potential subscriber to the insurance plan has the same *ex ante* probability of experiencing each possible stream of health outcomes, so that the

7

prospects of each are equal, as behind the Rawlsian veil of ignorance (Rawls (1971)). Under these circumstances, if the insurer could act as a perfect agent for the consumer, it would attempt to set the marginal benefits equal to the marginal costs of each intervention, but the marginal cost would be at the point of purchase of the intervention. That is, unless the insurer were a monopsonist, the cost would be the price paid (which in turn would be the sum of the insurer's payment and the copayment). This perspective adds insurer costs to the *patient perspective* that only includes out-of-pocket costs.

The same logic applies to a provider that acts as an insurer, such as a health maintenance organization. However, for services that the provider produces itself, the relevant price is the marginal cost defined over the suitable time horizon. A government program that intended to maximize the welfare of the citizens it serves would use a CE criterion on similar grounds. In each case, it would be optimal to equate the CE ratios of interventions used at the margin, using marginal costs that the program bears — that is, the prices that it actually pays.

To the extent that consensus about specific social welfare criteria is lacking, not everyone will be persuaded by an appeal to welfare economic foundations. Some writers have criticized the utilitarian viewpoint that they believe to be embedded in this approach. The justification for CE analysis on this basis is indeed rooted in the compensation principle (or Kaldor-Hicks criterion) of CB analysis (Hicks (1939), Kaldor (1939)). This principle states that we should undertake a project if and only if its net benefits are positive, since then those who gain from such a project gain by enough to compensate those who lose. If the losers are compensated, nobody is made worse off by the project, and someone is made better off. Thus the term *potential Pareto improvement* — the project *could* result in an actual Pareto improvement if the winners compensated the losers. Since a precisely compensating reallocation is unlikely to occur, this criterion is less compelling than Pareto improvement, since a project that produces positive net benefit would make people who shared the costs but not the benefits worse off.

The paper is organized as follows. The first section briefly describes the basics of CE analysis and how it can be applied to aid decisions about the allocation of health resources. The paper then turns to the potential welfare economic foundations of CE analysis, drawing heavily on my work with Charles Phelps. The paper then addresses specific issues in carrying out CE analysis, such as which costs to include, whose perspective matters in the analysis, and how health outcomes are measured. It demonstrates how a welfare economic foundation can help resolve ambiguities and

uncertainties about the application of CE analysis. The paper also discusses the limitations of such an approach, which indeed reflect limitations of CE analysis as an analytic framework. Finally, it addresses unresolved issues such as the difficulties in using the results of CE analysis to make health policy at the societal or group level.

2. Cost-Effectiveness Analysis for Decision Making

How useful and valid are the results of CE analysis if its purpose is to improve the well-being of a population by guiding the allocation of health care resources? Making this judgment requires choosing a benchmark for well-being and an explicit statement about how CE analysis can be used to achieve the welfare objectives. Major published recommendations for the use of CE analysis in guiding decisions state that it must be weighed with a variety of political, distributional and practical considerations. The information that CE analysis contributes is summarized by the *CE ratio.* The CE ratio is a cost per unit health effect achieved by using a particular health intervention. The CE ratio demonstrates which uses of health resources will provide health most efficiently; by first using interventions that have the lowest CE ratio, i.e., that produce the greatest effect from a specific expenditure, it is possible to obtain the greatest overall health effect from a limited budget for health care. Recent work on welfare foundations of CE analysis has used standard neoclassical welfare economic formulations to examine whether implementation of CE analysis in this way (i.e., using different interventions to the point that their incremental CE ratios are equal at the margin) leads to the same allocations as the ones that result from individual utility maximization subject to income constraints.

To explore these issues further requires knowing precisely what the CE ratio represents and how it is calculated. As one might expect, the closer the connection between the health outcome and individual welfare, the more plausible the claim that allocations based on CE criteria maximize welfare.

Several authoritative textbooks and reviews have described the general approach for performing a CE analysis; see for example, Drummond, et al. (1997), Gold, et al. (1996), Weinstein and Stason (1977). I briefly summarize the approach here.

First, the intervention to be studied, along with alternative interventions to which it is being compared, must be defined. One of the alternatives might be "doing nothing," or applying no specific intervention. This has been the principal alternative considered in many CE analyses. Yet a

CE analysis based on a comparison with this alternative is not always informative, since the comparison should be between relevant choices, such as two treatments or diagnostic approaches that clinicians or policymakers would consider to be the most promising. Little can be learned from a CE analysis that compares an intervention with placebo when placebo is not considered a reasonable option. The CE ratio for a comparison with placebo can be favorable even when the intervention in question is in every respect inferior to one or more commonly used alternatives. Several medications, for example, are both effective and cost-effective when used to treat adults with moderately elevated blood pressure. The relevant question for a new blood pressure medication is how it compares to another promising medication, or to others that are well-established, rather than how it compares to the abandoned approach of forgoing treatment.

After we choose the intervention and alternative to be studied, we must assemble several elements of the CE analysis to calculate the *incremental* (or marginal) CE ratio. Throughout this paper, the term CE ratio refers to the incremental CE ratio, unless otherwise specified. The term *incremental* is used rather than *marginal* to avoid confusion with the term *marginal cost*, which is usually the preferred measure of opportunity cost in CE analysis. *Incremental* refers to differences between two interventions; since the comparison does not always involve an infinitesimal change in costs and effectiveness, the term "marginal" can be misleading.

Let the subscripts *1* and *0* denote the intervention under study and the alternative to which it is compared, respectively. If C_1 and C_0 are the net present values of costs that result when the intervention and alternatives are used, and E_1 and E_0 their respective health outcomes, the incremental CE ratio is simply

$$CE \text{ ratio} = \frac{C_1 - C_0}{E_1 - E_0}. \tag{1}$$

This ratio, which is a cost per unit incremental health effect, is often used as a measure of value. The CE ratio of the intervention under study is compared to the CE ratios of other commonly used forms of medical care; if it is relatively low, the intervention under study is considered to be a good value. Note that the intervention and alternative can be two different intensities of the same treatment (e.g., dosage of a drug), and that the CE ratio can be defined as an infinitesimal charge. The continuously valued approach to the CE ratio underlies the analysis of Section 3.

The elements of the numerator of the CE ratio, or the incremental cost of the intervention, are discussed below. There is consensus that C_1 and C_0 should represent net present values, but the specific content of these numbers is controversial. Much of the literature has used formulations similar to that of Weinstein and Stason, who stated that net health care costs consist of "all direct medical and health care costs [including] costs of hospitalization, physician time, medications, laboratory services, counseling, and other ancillary services." In addition, the costs include those "associated with the adverse side effects of treatment," the (negative) costs from "savings in health care, rehabilitation and custodial costs due to the prevention or alleviation of disease," and "the costs of treating diseases that would not have occurred if the patient had not lived longer as a result of the original treatment" (Weinstein and Stason (1977), p. 718). Many studies have attempted to measure costs by including these categories. Some experts exclude those that arise solely from living longer, as previously noted. Others have included additional costs, such as "indirect" or "productivity" costs (i.e., time costs of treatment and/or disease, lost wages, and so on) and consumption expenditures. The Panel on Cost-Effectiveness in Health and Medicine recommended against including as costs the monetary value imputed for lost life years (i.e., lost earnings) and withheld endorsement of including future consumption expenditures, yet many CE studies have incorporated the imputed value of lost years of life in the cost measures.

The denominator of the CE ratio is calculated in an analogous manner; it represents the incremental health effects of using the intervention. Typical measures of health outcomes are either *years of life saved* or *quality-adjusted life years* (QALYs) saved. QALYs were introduced into the literature in the mid-1970s as a way to incorporate the benefits of treatment more fully than could be accommodated with earlier outcome measures. They are intended to serve as a comprehensive measure of health, or health-related well-being.   In many respects QALYs are analogous to life expectancy, but give credit to interventions that improve quality of life even when they do not affect survival.

Each year that an individual lives longer contributes an additional year to the life expectancy calculation. The amount that each additional year of life adds to QALYs, in contrast, is a preference weight or utility that takes a value between 0 and 1, varying with health status during the incremental year. Life years marred by functional limitations, pain, and other burdens associated with illness receive less weight than years in good health. Years when health is so bad that it is considered no better than death receive a preference weight of 0; in the usual formulation, death is

considered the worst possible health state. A preference weight of 1 corresponds to best health imaginable. Interventions can raise QALYs by lengthening life or improving its quality as reflected in the preference weight. Similarly, an intervention that lengthens life produces more QALYs if it maintains or improves quality of life than if it adds years of life that are impaired by significant morbidity. Both life expectancy and QALYs can be discounted; that is, less weight is given to years of life added in the more distant future.

QALY measurement is most easily understood by extending the measurement of life expectancy. Life expectancy is the sum of the probabilities that an individual will be alive at each age (denoted by i) in the future, up to the maximal life span, or

$$life\ expectancy\ =\ \sum_{i=current\ age}^{maximum\ age} F_i, \qquad (2)$$

where F_i is the probability that the person who is now at the "current age" will still be alive at age i; this discrete representation is most convenient for working with data such as life tables, but continuous time representations of life expectancy are also used.

Calculation of QALYs requires the information used to calculate life expectancy and the preference weights. Denote the preference weight for the health characterizing age i by q_i. Each such term is the expected value of quality adjustments for all possible states of health at age i. To illustrate the calculation, imagine that individuals alive at age 60 could be in one of only two possible states of health: perfect health, ($q_h = 1$), occurring with probability 0.5, or suffering from heart disease ($q_d = .8$), also occurring with probability 0.5. Then q_{60}, the expected value of the preference weight corresponding to being alive at age 60, is $(0.5 \times 1) + (0.5 \times 0.8) = 0.9$. After estimating the value of q_i for each age i, it is possible to calculate the expected number of QALYs, in the form of present value, according to the formula

$$QALY\ =\ \sum_{i=current\ age}^{maximum\ age} F_i \delta^i q_i\ , \qquad (3)$$

where δ is a time discount factor whose value is between 0 and 1. As in the formula for life expectancy, F_i is the probability that the person is still alive at age i. If $\delta=1$, two years of life in which $q_i = 0.33$ contribute the same number of QALYs as one year in which $q_i = 0.66$. If there is no

time discounting ($\delta = 1$) and if each year of life has perfect health, or quality adjustment is ignored ($q_i = 1$ for every value of i), then this formula simplifies to the formula for life expectancy.

The mechanical aspects of calculating QALYs are not difficult, but the measurement of the preference weights and the probabilities of alternative states of health is anything but straightforward. The specifics of QALY calculation necessarily account for much of the effort of CE analysis, since the outcome measure is critical to the interpretation of the results. As section 3 discusses, the outcome measure determines whether the application of CE analysis has desirable welfare-theoretic properties.

Time horizon

An intervention can alter both costs and health effects long after it is administered. For example, a mammogram uses resources at the time the test is conducted. But if it reveals an abnormality that leads to breast biopsy, mastectomy, and the prevention of morbidity and mortality from breast cancer, it alters the length of life, future morbidity, and future costs of health care. These long-term repercussions are relevant to any evaluation of screening with mammography, so the standard recommendation is that all future costs and health effects should be calculated or estimated in a CE analysis. Measuring these costs and health effects directly — without use of a model that extrapolates these numbers — would require observing until death a large number of women who underwent mammography, along with a number of women who did not have the test. For many treatments and diagnostic or screening strategies, such an approach would require decades of study, yet few randomized clinical trials last for more than five years. Strong beliefs in the credibility of direct clinical trial data, and skepticism about model-based extrapolations beyond the period of the trial, have led some investigators to calculate costs and outcomes for the period of the trial only. Thus, rather than estimate life expectancy or quality-adjusted life years, they calculate survival within the five years of a trial. Similarly, rather than estimate net present value of lifetime health care costs, they measure discounted costs during the period of the trial. Usually, when researchers adopt this approach, they do so in the belief that they have avoided making dubious assumptions needed to extrapolate events and costs that occur beyond the period for which they have valid and reliable data.

This practice is not endorsed by experts on CE analysis. There is no natural interpretation for life-years gained during a finite period of time, and the CE ratios that result from using different

time horizons, such as one year and five years, cannot be compared in any meaningful way. In fact, the resulting CE ratios can be understood best by interpreting them as special cases of standard CE ratios. In calculating a standard CE ratio, the time horizon is at least equal to the full span of life. The 5-year CE ratio is the same as a standard CE ratio calculated with an assumption that all individuals die at the end of five years. Thus, in the attempt to avoid the assumptions required for modeling long time horizons, researchers who truncate their analyses have made, perhaps unwittingly, the implausible alternative assumption that study subjects experience neither the costs nor the benefits of living beyond the period of study.

Although it seems intuitive that calculating the CE ratio based on a truncated time period should result in bias, it may not be possible to determine the sign of the bias *a priori*. The bias can only be calculated by making specific assumptions about the costs and health effects that occur after the period of observation. For example, suppose that the intervention in question lowers mortality rates during five years of observation. For individuals surviving the five years, subsequent survival experience and costs are the same for those treated with placebo as for those who received the intervention. Under these assumptions, both the gain in life-years and the increase in costs are greater for the intervention group than would be estimated on the basis of the truncated period of observation. The overall bias in the CE ratio depends upon the relative magnitudes of these omitted costs and health effects.

Average CE ratio

Some CE analyses report an *average CE ratio*, which is simply the ratio of C_1 to E_1. For comparisons among multiple alternatives, a similar practice is common: each intervention is compared to a single alternative. Both approaches are convenient because either they do not require a comparison treatment, or all treatments are compared to a single alternative, rather than to multiple alternatives. Both approaches, however, are misleading. The average CE ratio is equivalent to a standard (incremental) CE ratio in which the alternative is costless and results in immediate death. If such an alternative exists, it is rare for any but the most rapidly and uniformly fatal health conditions. The average CE ratio can deviate greatly from the incremental CE ratio when the intervention under study is a preventive service, which typically would be administered to a relatively healthy population. The members of such a population would be expected to have many years of good health and to generate substantial costs over their remaining lifetimes.
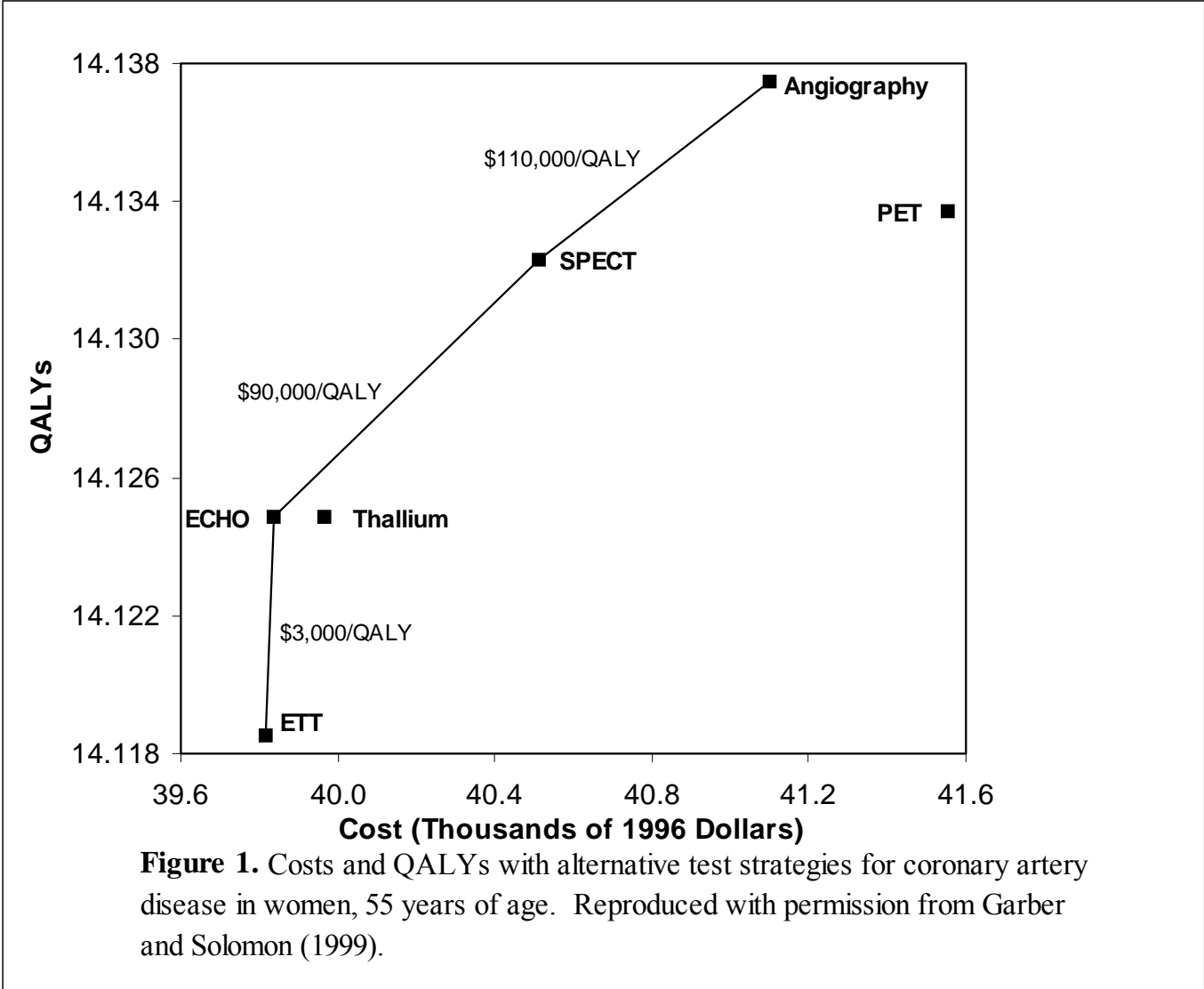
The average CE ratio will not, in general, lead to appropriate rankings of alternative health expenditures (see, for example, Karlsson and Johannesson (1996)), although occasionally it is possible to draw limited inferences about the value of the incremental CE ratio from the average CE ratio. The average CE ratio does not reliably indicate the way to achieve the greatest health benefit from a given expenditure. For example, an intervention that produces more favorable outcomes than one that has a lower average CE ratio could have an acceptable incremental CE ratio but might not be selected on the basis of the average CE ratio; alternatively, the average CE ratio might be considered "acceptable" when the incremental CE ratio was very high.

Comparison of multiple interventions to a single alternative is misleading for nearly the same reason, except that the "baseline" costs and outcomes are not zero, but instead are the costs and outcomes corresponding to the single comparator. It is easiest to understand why this is misleading by comparing it to the incremental approach.

Incremental CE ratio for multiple alternatives

It is possible to calculate a separate incremental CE ratio for every pair of alternative interventions. When many interventions are considered, the number of such pairs becomes large. However, because most of the incremental CE ratios are irrelevant, the analyst need not calculate all of them. Instead, to determine the incremental CE of a series of different combinations of technologies, the analyst should first rank each alternative by the health effect achieved — e.g., the number of QALYs (or life-years) it produces. Then the analyst should determine whether any interventions are *strictly dominated* (more expensive and less effective than at least one alternative intervention); if any are, they should be eliminated from further consideration. After eliminating all such alternatives, one should calculate the incremental CE ratios between each intervention and the next most expensive alternative. Subsequently, interventions that display *extended dominance* should also be eliminated, and the incremental CE ratios of all remaining alternatives calculated. Extended dominance is defined below.

Figure 1, from Garber and Solomon (1999), illustrates how incremental CE analysis can be applied when multiple alternatives are considered. It shows the costs and health effects of adopting each of several strategies for diagnosing coronary artery disease in 55 year-old women. The first five strategies are exercise treadmill testing (ETT); stress echocardiography (ECHO); planar thallium radionuclide imaging (Thallium); single photon emission computed tomography (SPECT); and positron emission tomography (PET). Each of these strategies starts with a noninvasive test for coronary disease. The "gold standard" test for coronary artery disease is cardiac catheterization with coronary angiography; the screening strategies that start with a noninvasive test proceed to catheterization if the test is abnormal. The final strategy shown in the figure (angiography) consists of initial testing with the gold standard test, so that the first test is considered definitive but riskier

Figure 1. Costs and QALYs with alternative test strategies for coronary artery disease in women, 55 years of age. Reproduced with permission from Garber and Solomon (1999).
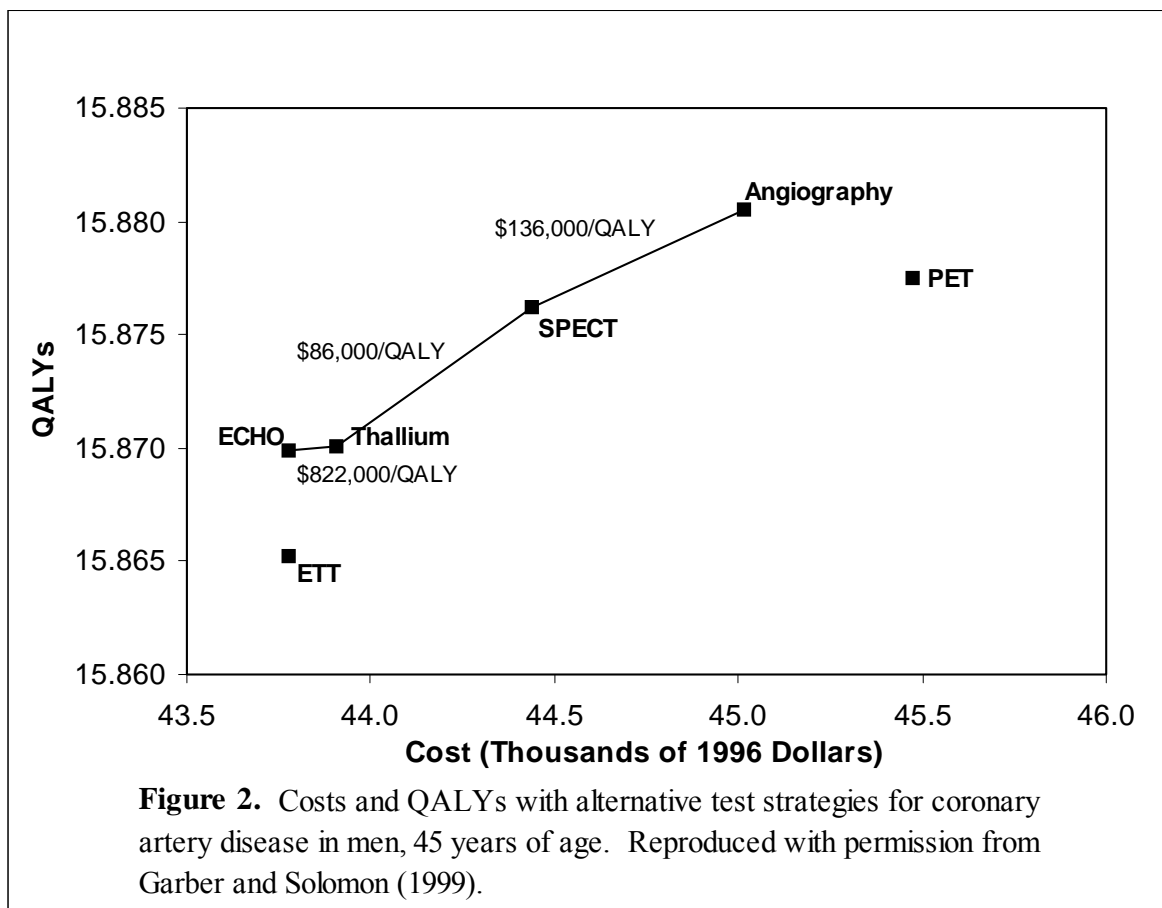
and more expensive than the other tests.

The costs and outcomes of each of the diagnostic strategies are calculated by modeling the consequences of alternative medical interventions that are pursued on the basis of the test results. For example, if a diagnostic test is positive and leads to the discovery of a severe form of coronary artery disease, it leads to surgical treatment, which in turn may prolong life substantially. A false positive test result has minimal adverse health effects, but leads to substantial expenditures for further testing that is, in retrospect, unnecessary.   Figure 1 is a compact representation of results from extensive modeling of alternative strategies that have large but often indirect and complex effects on both costs and health outcomes.

Because each point on the figure represents the overall costs and outcomes in QALYs that result from the use of each test, the incremental CE ratio between any pair of tests is the inverse of the slope of the line drawn between their corresponding points. A point that is above and to the left of another strictly dominates the alternative, i.e., has better outcomes and lower costs. In Figure 1, angiography eliminates PET scanning by strict dominance. Thallium is also eliminated by strict dominance because it produces slightly fewer QALYs than ECHO at greater cost. The incremental CE ratios are calculated for the remaining alternatives.

Figure 2 (also from Garber and Solomon), which shows similar results for 45 year-old men, illustrates extended dominance. For these subjects, unlike 55 year-old women, thallium is not eliminated by strict dominance, since no alternative intervention is both less expensive and more effective in these men. Extended dominance is a somewhat more subtle concept than strict dominance; it occurs whenever a linear combination of two alternatives strictly dominates a third (Keeney and Raiffa (1993), Johannesson and Weinstein (1993), Karlsson and Johannesson (1996)). Equivalently, the phenomenon occurs when any interventions have "higher incremental C/E ratios than a more effective option" (Siegel, et al. (1996)). Although no alternative is both less expensive and more effective than thallium, it is strictly dominated by at least one point on a line drawn between ECHO and SPECT, so it is eliminated by extended dominance.

Strict dominance and extended dominance are particularly important phenomena because they can identify interventions that should be eliminated from consideration, without making any

Figure 2. Costs and QALYs with alternative test strategies for coronary artery disease in men, 45 years of age. Reproduced with permission from Garber and Solomon (1999).

judgment about what a unit health effect is worth. Strict dominance cannot always be detected

without formal analysis, and extended dominance is even harder to discover, unless the analysis includes a systematic approach to incremental CE ratios.

A rational decision maker will never choose an option that can be eliminated under extended dominance, because a more expensive alternative would result in a lower or equivalent CE ratio. Suppose that there are three alternatives under consideration: A, B, and C. Both the costs and the outcomes associated with intervention C are greater than those of intervention B, which in turn are greater than those of intervention A. Thus none of the interventions strictly dominates any other. The (incremental) CE ratio of intervention B compared to A is $70,000/QALY, and the CE ratio of C compared to B is $10,000/QALY. If a decision maker would choose B over A, it implies that a gain of a QALY is worth at least $70,000 to him or her. If that is the case, then it must be true that it is worth an additional $10,000 to gain another QALY, so that C would be chosen over B. Thus alternative B is eliminated from consideration by extended dominance.

The CE ratios that result from comparing several interventions to a single alternative, rather than proceeding in this stepwise fashion, can be very different. Usually it is impossible to detect the presence of either strict or extended dominance from such an approach. In fact, the CE ratio produced this way may appear to be "reasonable" even though the intervention under consideration is strictly dominated by another! Suppose that there is an intervention A that generates lower costs than interventions B and C, as in Figure 3. We are interested in choosing among the three. If we calculate cost-effectiveness ratios of B compared to A and C compared to A, it is difficult to determine whether we should choose C over B. If the CE ratio of C compared to A is lower than the ratio of B compared to A, C could eliminate B by extended or strict dominance (points B^1 and B^2 in Figure 3, respectively) or, alternatively, B could have an "acceptable" CE ratio compared to B (point B^3). The only firm conclusion that can be drawn, without further information, is that B does not eliminate C by strict dominance.
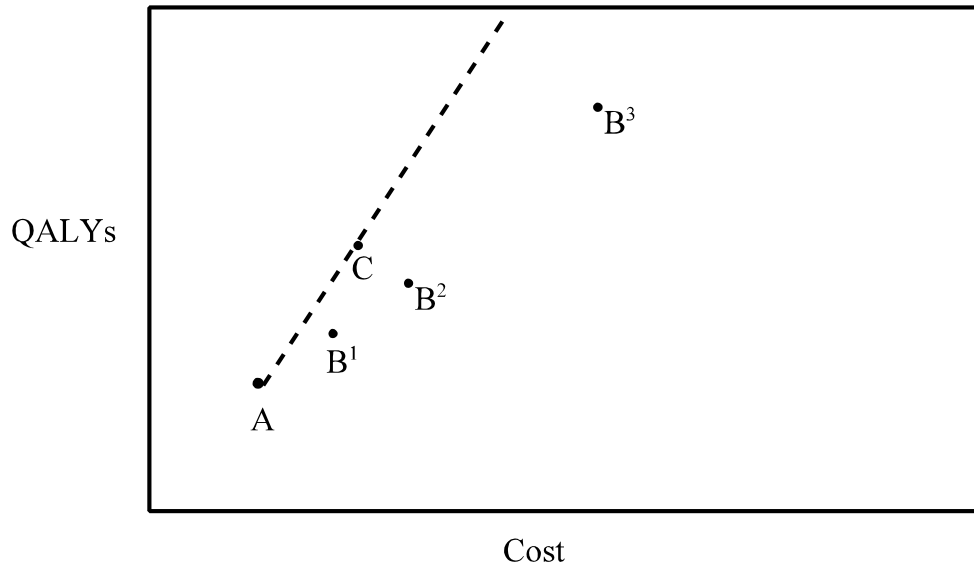
Figure 3. The consequences of comparing two interventions to a third. Intervention A is the lowest cost alternative; the incremental CE ratio of C compared to A is lower than the incremental CE ratio of B compared to A. Interventions B^1, B^2, and B^3 all have the same CE ratio compared to A. C eliminates B^1 by extended dominance and B^2 by strict dominance, while the CE ratio of B^3 compared to C could be "acceptable" (i.e., lower than a CE cutoff). Without further information, it is not possible to determine from the CE ratios of C compared to A and B compared to A which of these three conditions applies.

Calculation of the incremental CE ratio, then, consists of estimating the QALYs and the present value of costs under the intervention and under its alternatives. The use of the average CE ratio or comparison of several interventions with a single alternative is misleading.

20

Sensitivity analysis

Uncertainty characterizes several components of nearly every CE analysis. Estimates of health effects, whether measured in terms of life-years or quality-adjusted life years, often build upon models that incorporate data from multiple sources. Even if the data are derived primarily from a randomized clinical trial, extrapolations beyond the period of the trial require assumptions about disease course beyond the period of observation. And even if a trial is the sole source of all data used in a CE analysis, sampling variability makes estimates of effect sizes and costs uncertain.

Not all sources of variability are purely random. For example, the costs of an intervention — or of treatments for conditions it prevents — may vary from one setting to another. Thus, for reasons ranging from the usual stochastic nature of experimental information to (possibly non-random) variation in costs and health effects to uncertainty in model structure and specifications, point estimates of CE ratios should ordinarily be considered just that. The variation in possible values around those point estimates may be large.

For this reason, CE analyses are considered incomplete if they do not include some form of sensitivity analysis. *Sensitivity analysis* is an exercise that shows the effects of variation in uncertain parameters on the final results of the analysis (i.e., the CE ratio). Textbooks on CE analysis and decision analysis discuss methods of sensitivity analysis, and most commercial software for CE and decision analysis implements one- or two-way sensitivity analysis. In *one-way* sensitivity analysis, one uncertain parameter is varied at a time, with the values of all other parameters held constant. In *two-way* sensitivity analysis, two parameters are varied simultaneously. When more than two parameters are varied, the presentation of results of multi-way sensitivity analysis can be quite challenging, and creative approaches to graphical presentation are necessary (two-way sensitivity analysis requires three-dimensional plotting, with axes for each of the two parameters being varied and for the CE ratio).

The limitations of traditional sensitivity analysis are most apparent when it is important to display the effects of uncertainty in multiple parameters simultaneously. More powerful alternative approaches, although they are still under development, have been gaining in popularity in part because they are more suitable for complex models with multiple sources of uncertainty. Most are statistical approaches that involve calculating confidence regions around CE ratios and other outcome variables. Briggs and Sculpher's 1995 survey of sensitivity analysis in economic

21

evaluation noted that only one of the 121 CE analyses they reviewed had adopted a "probabilistic sensitivity analysis" approach, whereas 42 used "one way simple sensitivity analysis" and 15 used "multi way simple sensitivity analysis" (Briggs and Sculpher (1995)). Methods for calculating the range of uncertainty using a probabilistic approach range from the traditional delta method to newer simulation and resampling techniques, such as the bootstrap, which makes it possible to limit parametric assumptions (Mullahy and Manning (1994), O'Brien, et al. (1994), Briggs, et al. (1994), Wakker and Klaassen (1995)). But the computational burdens of such approaches remain formidable, and in many cases the statistical theory is not well developed or, like the delta method, require strict distributional assumptions. Furthermore, the patchwork of data used to develop many CE models limits the range of approaches that can be used to gauge the effects of uncertainty.

The welfare theoretical implications of uncertainty in the analysis are important, even if they are indirect. It is not unusual for the range of uncertainty to be great enough to be consistent with different orderings of effectiveness (and costs) of the interventions under consideration. Occasionally differences in costs among alternative interventions are known with a high degree of certainty, but ranges of estimated effectiveness overlap substantially. A common response to this situation is to assume that the effectiveness of each intervention is roughly equal, and to choose the lowest-cost alternative. However, the apparent equivalence of effectiveness may be a consequence either of similar true effectiveness, or of large but highly uncertain differences in effectiveness. In the latter case, further information might alter the ranking of alternatives.

Interpretation for medical decision making and health policy

After the CE ratios of non-dominated alternatives are calculated, there remains the task of choosing among them. If an intervention improves health at a cost of $80,000/QALY, should it be adopted? Cost-benefit analysis leads to specific recommendations because it places a monetary value on the benefits: any intervention that produces a net benefit generates a potential Pareto improvement. But CE analysis is often preferred precisely because it avoids monetary valuation of health benefits. The next section describes how it is possible to derive a "cutoff" CE ratio that leads to the same choices as a cost-benefit criterion. However, people who apply and use CE analyses and wish to avoid the valuation of health benefits implicit in such efforts often use an alternative approach based on *league tables*.

The term league table apparently originates from the tables of football team rankings published in European nations. League tables in CE analysis also display rankings. This approach compares the CE ratio of the intervention under study to those of other common medical interventions. By compiling a league table of (incremental) CE ratios of other health interventions, usually culled from the literature, one can demonstrate how the CE ratio of the intervention under study compares with those of the other interventions in the table. If the CE ratio is low, the intervention is termed a good value, while if the CE ratio is high, it is identified as a poor value relative to other accepted interventions. Thus the tabular comparison helps to establish whether the intervention should be used.

3. When Does CE Analysis Lead to Optimal Decisions?

The league table approach, however, has severe limitations as a guide to medical choices (Birch and Gafni (1994)). Several problems become apparent to readers of the studies that generated the numbers. For example, the various studies summarized in the table may not use comparable methodology; some of the CE ratios may be incremental, others average; assumptions underlying the cost estimates may differ greatly. Although league tables distinguish between interventions that are relatively good and relatively poor values, that judgment is highly dependent upon the specific alternatives displayed in each table. Unless there is a reason to believe that the interventions appearing in the league table were chosen by a process that maximizes value, we can hardly infer that standing in the league table establishes value in any absolute sense. Finally, even if we could infer whether the intervention was a relatively good or bad value, the league table approach does not establish how much should be spent. This observation leads us back to the question posed at the outset: when we apply the results of CE analysis to allocate health care, do we make optimal decisions? No discussion of the welfare economic foundations and welfare implications of applications of CE analysis is meaningful without consideration of how and why CE analysis is being used. For whom is CE analysis being conducted, and how will its results be used in allocation decisions?

The answers to these questions depend upon the *perspective* of the analysis. The approved practice, under most circumstances, is to adopt a societal perspective, in which we are seeking to make the best decision about health care allocation for a group of people. Often, however, this perspective is taken to mean something more specific: the analysis is intended to aid someone such

23

as a social planner — perhaps the health minister of a country with national health insurance or governmentally provided health care — who must decide which health services to provide or reimburse. The adoption of a societal perspective can give rise to ambiguities. For example, how should the government payer handle heterogeneous preferences, if it recognizes them at all?

The following discussion builds upon the presentation in Garber and Phelps (1997). In that paper, the perspective is that of a "perfect insurer," and CE analysis is treated as a tool to determine which services, in what quantities, the perfect insurer should reimburse. Suppose that there is no specific information to suggest that an individual's risk of various health events differs from the average for the insured population, that utility functions and other characteristics are homogeneous, and that the insurance is actuarially fair. Which services would the optimal policy cover? From this point of view, the usual marginal conditions apply, and CB criteria (i.e., measure benefits and costs accurately and cover those services at quantities that result in maximum net benefit) lead to expected utility maximization. Only those services whose expected benefits equal or exceed their expected costs, which will be included in the premium and copayments, will be covered.

The Garber-Phelps approach has two major characteristics: it uses first-order conditions to derive *cutoff* or *threshold* CE ratios, and it determines when various rules for conducting CE analysis allow the technique to be used to determine optimal health resource allocations. It is possible, for example, that ignoring certain categories of costs, such as earnings lost as a result of early death, would mean that decision rules based on CE analysis would no longer be reliable guides to welfare maximization, or that inappropriately including such costs would also lead to incorrect rankings of alternative health programs.

Garber and Phelps construct the health care allocation problem as a simple von Neumann-Morgenstern utility maximization; essentially, they ask whether the first order conditions can be expressed in a form that leads to a CE criterion. That is, they ask whether it is possible to identify a threshold CE ratio such that acceptance of all interventions whose CE ratio falls below the threshold and rejection of those with higher CE ratios would correspond to the allocation selected by direct utility maximization. In the Garber-Phelps model, the threshold CE ratio for an expenditure on a health intervention in the initial period is simply the ratio between the initial period utility and the marginal utility of income in that period. Fundamental to this approach is an assumption that the effectiveness measure is at least an affine transformation of utility. Embedded in the model is an additional assumption that period-specific income is fixed.

24

The general model is based on an expected utility function in which first period utility U_0 is a function of initial income Y_0 less expenditures on intervention **a**, whose unit price is w_a, and expenditures on intervention **b**, at unit price w_b. Subsequent period-specific utilities are given by the utility functions $U_i(Y_i)$ weighted by the probability that the individual will be alive in period i, F_i:

$$E_0 U = U_0(Y_0 - w_a \mathbf{a} - w_b \mathbf{b}) + \sum_{i=1}^{N} U_i(Y_i) F_i \; . \tag{4}$$

U_i can be written as $U_i = \upsilon \delta^i k_i$, where $\upsilon = U_0(Y)$. In this formulation, Y is constant over time, and k_i is a period-specific multiplier. Thus the summation term has the form of QALYs, in which the quality adjustment for period i is simply U_i; this corresponds to the common use of the term *"utilities"* to describe the quality adjustments. We denote the summation term by **Q**.

Interventions **a** and **b** can have effects on the probabilities of survival in the future via F_i and on the utilities via k_i. Both F_i and k_i, and their dependence on **a** and **b**, can have an arbitrary time pattern. Obtaining the first order conditions for the maximization of utility with respect to expenditures on **a** and **b** is straightforward (note that there can be corner solutions, since optimal expenditures might be zero for either or both interventions). Denote the marginal effect of intervention **a** on future period-specific mortality P_i by $\partial P_i / \partial \mathbf{a} = \varepsilon_i^a$, and let the marginal effect of **a** on period-specific quality adjustments k_i be denoted by $\partial k_i / \partial \mathbf{a} = \psi_i^a$. Using the relationship between conditional mortality and cumulative probability of survival

$$F_i = \prod_{j=1}^{i} P_j \; , \tag{5}$$

and differentiating expected utility with respect to intervention **a**, we have

$$\frac{\partial E_0 U}{\partial \mathbf{a}} = -w_a U_0' + \upsilon \left\{ \sum_{i=1}^{N} \delta^i \prod_{j=1}^{i} P_j \left(\psi_i^a + k_i \sum_{k=1}^{i} \frac{\varepsilon_k^a}{P_k} \right) \right\} , \tag{6}$$

which when equated to 0 gives the first order condition

$$w_a = \frac{\upsilon}{U_0'} \frac{\partial Q}{\partial \mathbf{a}} \; . \tag{7}$$

An analogous relationship results from maximization with respect to intervention **b**:

$$w_b = \frac{\upsilon}{U_0'} \frac{\partial Q}{\partial \boldsymbol{b}} \ .$$

(8)

The analysis then proceeds to show how the first order conditions can be translated into CE criteria, in which future unrelated costs of health care are either included or excluded.

First, consider obtaining the optimal cutoff CE ratio when unrelated future costs are ignored. Current medical costs are $C = w_a\boldsymbol{a} + w_b\boldsymbol{b}$. Let $\boldsymbol{z} = d\boldsymbol{b}/d\boldsymbol{a}$, the marginal rate of substitution between **b** and **a**. Differentiating C with respect to **a** and substituting **z** yields the relationship $dC/d\boldsymbol{a} = w_a + \boldsymbol{z}w_b$. Then the CE ratio for intervention **a** is

$$\left(\frac{dC}{dQ}\right)_a = \frac{\dfrac{dC}{d\boldsymbol{a}}}{\dfrac{dQ}{d\boldsymbol{a}}} = \frac{\dfrac{\partial C}{\partial \boldsymbol{a}} + zw_b}{\dfrac{\partial Q}{\partial \boldsymbol{a}} + z\dfrac{\partial Q}{\partial \boldsymbol{b}}} \ .$$

(9)　　　　　　　　(9)

Using the first order conditions to solve for the optimal values of $\partial Q/\partial\boldsymbol{a}$ and $\partial Q/\partial\boldsymbol{b}$ implies that, at the optimum investment in intervention **a**,

$$\left(\frac{dC}{dQ}\right)_a = \frac{w_a + zw_b}{\left(w_a + zw_b\right)\left(\dfrac{U_0'}{\upsilon}\right)} = \frac{\upsilon}{U_0'} \ .$$

(10)

According to this equation, the ratio of incremental costs to incremental QALYs from further investment in intervention **a** is proportional to the reciprocal of the marginal utility of consumption in the initial period, U_0'. Here, the term *incremental* is completely synonymous with *marginal*, since the CE condition is based on a comparison of an incremental expenditure on **a**, rather than on a comparison to a (discrete) alternative intervention. We can use an analogous procedure to obtain the optimal cutoff CE ratio for intervention **b**, yielding the result that, at optimal investment in **b**,

$$\left(\frac{dC}{dQ}\right)_b = \frac{\dfrac{w_a}{z} + w_b}{\left(\dfrac{w_a}{z} + w_b\right)\left(\dfrac{U_0'}{\upsilon}\right)} = \frac{\upsilon}{U_0'} \ .$$

(11)

Thus, when future costs are ignored, the first order conditions imply that a single optimal CE ratio applies to all interventions.

A similar analysis establishes the optimal CE ratio when future costs are included. In this case, the numerator of the CE ratio is the marginal cost of the intervention, including future health care costs. The lifetime costs are

$$C^{tot} = w_a \mathbf{a} + w_b \mathbf{b} + P_1 \delta c_1 + P_1 P_2 \delta^2 c_2 + \dots , \tag{12}$$

where c_i = total health expenditures in period i. Associated with the use of an intervention are costs of the intervention itself, induce change in expenditures for the other intervention, along with expenditures that result from living longer:

$$\frac{dC^{tot}}{d\mathbf{a}} = w_a + w_b \frac{d\mathbf{b}}{d\mathbf{a}} + \frac{1}{P_1}\left[\frac{\partial P_1}{\partial \mathbf{a}} + \frac{\partial P_1}{\partial \mathbf{b}}\frac{d\mathbf{b}}{d\mathbf{a}}\right]\left[\delta P_1 c_1 + \delta^2 P_1 P_2 c_2 + \dots\right] + \frac{1}{P_2}\left[\frac{\partial P_2}{\partial \mathbf{a}} + \frac{\partial P_2}{\partial \mathbf{b}}\cdot\frac{d\mathbf{b}}{d\mathbf{a}}\right]\left[P_1 P_2 \delta^2 c_2 + \dots\right] + \dots . \tag{13}$$

This expression can be rewritten

$$\frac{dC^{tot}}{d\mathbf{a}} = w_a + w_b \frac{d\mathbf{b}}{d\mathbf{a}} + \frac{\partial E}{\partial \mathbf{a}} + z\frac{\partial E}{\partial \mathbf{b}} = \frac{dC}{d\mathbf{a}} + \frac{\partial E}{\partial \mathbf{a}} + z\frac{\partial E}{\partial \mathbf{b}}, \tag{14}$$

where E = the net present value of expected health expenditures and as before $\mathbf{z} = d\mathbf{b}/d\mathbf{a}$.

By following the procedures used to obtain the optimal CE ratio when future costs are excluded, it is easy to show that

$$\left(\frac{dC^{tot}}{dQ}\right)_b = \left(\frac{dC^{tot}}{dQ}\right)_a = \frac{\upsilon}{U_0'} + \frac{\frac{1}{z}\left(\frac{\partial E}{\partial \mathbf{a}}\right) + \frac{\partial E}{\partial \mathbf{b}}}{\frac{1}{z}\left(\frac{\partial Q}{\partial \mathbf{a}}\right) + \frac{\partial Q}{\partial \mathbf{b}}}. \tag{15}$$

Thus, when unrelated future costs are included, the first order conditions imply a fixed optimal CE ratio that is the same for all interventions. The second ratio on the right-hand side of Eq. 15 is a constant when the future costs are unrelated, so the optimal CE ratio when future costs are included is equal to the optimal CE ratio when the future costs are excluded, plus a constant.

This result follows from a number of assumptions. A key one is the optimality of future health care expenditures. If the expenditures are not optimal, it will ordinarily be difficult to apply a CE criterion, since the quality adjustment terms for future years will need to reflect differential utility losses from varying distortions in health care consumption in future years. In addition, this analysis uses a strict definition of "unrelated" future expenditures: conditional on reaching a given age, a person's expenditures on health care do not change with an increase in the quantities of intervention **a** or **b** consumed. Thus the goods under study cannot be close substitutes or complements for other forms of health care (nor can there be changes in the rates of substitution between quality-enhancing and life-prolonging health care). The conditional independence assumption, which is intended to be an accurate representation of the term "unrelated" that often appears in the literature without precise definition, is strict. Even if it can seldom be satisfied exactly, it may be a reasonable approximation for some interventions, such as the treatment of a young accident victim with severe blood loss whose future expected pattern of health may be unaltered by the accident if he or she survives.

This approach does not justify the application of a fixed threshold CE ratio when the first-order conditions cannot be met (e.g., the quantities of **a** and **b** cannot be varied continuously) or when the second-order conditions for a maximum cannot be met. Garber and Phelps argue that the quantities of most health interventions are continuously variable more often than is usually apparent. For example, a screening test might at first seem to be an example of an unambiguously discrete-valued quantity; a woman either has a mammogram or she does not. It is not possible to undergo partial mammographic screening for breast cancer. Yet there are several margins over which the quantity of mammography can be varied, such as the frequency of screening. In addition, the definition of a "positive" test — i.e., one that will lead to further diagnostic evaluation — is often variable (for example, one or more radiologists interpreting a mammogram could estimate the probability that a cancer is present). A more permissive threshold for abnormality results in more true-positive and false-positive test results, usually leading to better health outcomes and higher costs. Variation along such margins can be used to achieve the first-order conditions. As Garber and Phelps note, application of the CE approach in general requires the marginal conditions to hold, because otherwise the use of a fixed CE ratio to be applied across all interventions, as implied by the comparisons in league tables, will be misleading. When the marginal conditions do not hold, optimal health resource allocation will not imply a fixed CE ratio across all interventions.

Restrictions in this model reflect an interpretation of QALYs in utility terms. More flexible utility functions and less restrictive assumptions, such as allowing for variable income and intertemporal reallocation of income and consumption, can change the results, as Meltzer (1997) reported. Extending the Garber-Phelps approach by allowing for borrowing and lending and explicitly distinguishing between health and non-health consumption, he reported that the first-order conditions could no longer be expressed solely in terms of a ratio between marginal costs of health interventions and marginal outcomes. His CE condition implied that "cost-effectiveness analysis must include the total change in future expenditures which results from a medical intervention, regardless of whether those expenditures are medical or non-medical . . . the cost-effectiveness ratio can be viewed as being the sum of a component related to current cost and a component related to future cost." Thus, according to Meltzer, not only "unrelated" future expenditures for health care, but also non-medical consumption expenditures, must be incorporated whenever the intervention under study prolongs life. His results pose a severe challenge for the routine practice of CE analysis, since the utility terms that the quality adjustments need to measure are even further removed from routine measurement of QALYs than under the Garber-Phelps model. Furthermore, the unavailability of accurate health and non-health consumption data has deterred most researchers from implementing any approach that incorporates the present value of non-health consumption as a health cost.

One way to interpret the results of these papers is that the decisions based on CEA can have favorable welfare economic properties, but only if both the costs and outcomes are measured properly. The outcome measure can serve as a basis for determining the first-order conditions only if it is a valid proxy for utility. Common practices in quality of life measurement, however, cast into doubt their ability to proxy overall utility. When developers of instruments for quality of life give respondents any information about what they should assume concerning the socioeconomic status and other factors that might change with a health state, the instructions usually say to consider only health-related aspects. Although rarely are versions of this instruction complete and explicit enough to define "health-related" precisely, their wording often implies that the respondent should ignore financial consequences of a health condition. A treatment that improves an aspect of utility — including utility from consumption expenditures — that is not measured by the effectiveness measure cannot be evaluated properly in this circumstance. But insofar as QALYs or similar

outcome measures are used, and are sufficiently broad to serve as a proxy for utility, it becomes much more plausible to represent utility maximization by a CE criterion.

Difficulties with interpreting existing QALY instruments as utility measures should not cast doubt on the theoretical appropriateness of CE analysis. The analysis can have stronger justification as a tool for welfare improvement if a better instrument is used. Furthermore, even CE analysis based on flawed measures of utility can provide a reasonable prioritization of alternative programs to improve health. In many circumstances, the alternative to CE analysis is a decision making process that devotes little attention to either the costs or health consequences of the various policy options. Insofar as it de-emphasizes or ignores considerations such as costs, it would be surprising if such an alternative would consistently prove to be a better guide to improvement of social welfare than even a flawed implementation of CE analysis.

4. Perspective and Cost Measurement

Despite its prominence as the numerator of the CE ratio, cost typically receives less space and research effort than effectiveness in CE analyses. This disparity may reflect the belief that measuring costs is relatively straightforward or that uncertainty about costs can be addressed adequately in the sensitivity analysis. Typically there are few direct data about the QALYs or life expectancy attributable to the use of a particular health intervention. Even when preference and cost data used for CE calculations are collected as part of a randomized "clinical-economic trial," outcomes must be modeled, as noted previously, because the duration of the trial is too short (typically five years or less) to measure directly the QALYs that result. (Direct measurement of QALYs requires following trial participants until they die.) Cost data, on the other hand, are considered to be relatively explicit and objective.

Estimated costs are usually (but not always) based on prices or, in the case of hospital services, accounting costs. In the U.S., both accounting- and price-based costs are problematic because both vary greatly. The price of a prescription drug purchased at a retail outlet in New York may differ greatly from the price charged by a hospital pharmacy in Los Angeles, which in turn differs from the price that a managed care organization pays a drug manufacturer. For complex services, such as a major operation, price variation may arise from variation in the definition of the service (not all cardiac valve replacement operations, for example, are the same), and from variation in the prices of factors such as nursing time, surgeon time, and hospital facilities. Although price

and accounting cost variation is both large and pervasive in some systems, it is not an insuperable problem for CE analysis. The judicious application of sensitivity analysis can mitigate problems arising from both variation and uncertainty in costs. Furthermore, in most applications, the uncertainty is greatest for costs incurred in the distant future. Such cost estimates require speculation about future health care practices and disease patterns, and thus compound uncertainty about the costs per unit of service. Discounting future costs at an interest rate of 3% or higher, however, means that different methods for measuring costs incurred in the distant future often produce similar present values. Consequently, many CE studies focus on estimation of effectiveness, which often requires indirect inference from results of disparate studies and the use of complex models.

Measurement of costs may nevertheless pose fundamental questions. The most basic is, what is the appropriate measure of cost for use in CE analysis? Should it be marginal cost, average cost, or neither? Many of the leading references on CE analysis say little about specific cost measures. For example, the aforementioned article by Weinstein and Stason (1977) enumerated categories of costs to include in direct medical and health care costs. But the article did not specify whether "costs" are prices in the service market, marginal costs of production, or average costs. In the presence of market imperfections — especially when fixed costs are significant — these alternative measures of cost can differ greatly. In a more detailed discussion of costs, the first edition of the textbook by Drummond, et al. (1987) stated that the costs should be "an estimate of the worth of the resources depleted by the programme" (p. 27) and subsequently discussed the various categories of costs (marginal, variable, average, and fixed costs), noting the reasons why different cost measures might be used. Their discussion suggests that the difference in total costs between two alternatives should be used as the measure of costs. Their discussion of how capital costs can be measured, however, stops short of recommending a specific measure to use if fixed or capital costs are large.

The treatment of fixed costs is only one of several controversies surrounding the measurement of costs in both CE and cost-benefit analysis. Experts debate whether only direct costs of the alternatives and of subsequent health care should be included, or whether productivity (indirect) costs (lost earnings or lost value of time) should also be included. They also debate how direct costs should be measured. What if, as is usual in health care, prices do not equal marginal costs? What is the appropriate measure of opportunity cost when markets are imperfect?

31

Should the societal perspective be the default?

Although there are not ready answers to all of these questions, they can be best addressed in the context of a specific perspective. Textbooks and review articles routinely emphasize the importance of selecting the perspective of the analysis (U.S. Congress Office of Technology Assessment (1980), Weinstein and Stason (1977)). Perspective determines whose costs are counted; the perspective of the patient, for example, is usually held to mean that only the costs that the patient bears directly — not the payments of an insurer or government program — matter. Since a typical American with indemnity health insurance bears 20% or less of the price of a covered health service, and in other health care systems the patient's share of costs is often negligible, an intervention that looks very cost-effective when only the patient's out-of-pocket costs are considered may seem like a poor value when the cost measure reflects total costs to the health care system. Opportunity costs, therefore, must be defined with reference to the perspective of the analysis.

The standard recommendation to conduct CE analysis from the societal perspective means that all costs, whether born by patients, insurers, or other parties, are included. Other perspectives may also be considered, but they are options to be contrasted with the societal perspective, not replacements for it. As in other perspectives, there should not be double-counting of costs (which in turn implies that pure, frictionless transfer payments are not counted as costs), nor in the societal perspective should any relevant costs be omitted. Consider an operation that costs $10,000, for which the insurer pays $8,000 and the patient pays a $2,000 copayment. A CE analysis conducted from the perspective of the patient would assign only a $2,000 cost to the intervention, one conducted from the insurer's perspective would assign $8,000, and one conducted from a societal perspective would assign the full cost of $10,000.

Critics of recommendations to make the societal perspective the default or principal perspective for CE analyses often note that analyses are conducted for a variety of reasons. Consumers and producers of CE analyses can be payers, pharmaceutical companies, providers, and purchasers of health care, so their cost perspectives may be relevant in many important and common situations. These criticisms of the use of the societal perspective are based on an assumption that a payer or government agency, for example, can ignore costs that it does not bear. Yet this assumption is not always realistic. Consider a private insurer; the *"payer's perspective,"* as usually conceived, includes reimbursements that the insurer pays but not the out-of-pocket payments of its

subscribers. If an insurer does not care about the well-being of its subscribers, so that it can ignore the costs the subscriber bears, then why does it care about maximizing each subscriber's QALYs, which are usually far more difficult to measure? If an insurer sells policies in a competitive market, the value of the policy will depend in part upon the out-of-pocket expenses and time costs that the patient bears. The belief that the insurer ignores costs to the patient overlooks an important fact: insurance programs that account for out-of-pocket expenses and time costs as well as payments by the insurer offer greater benefit to subscribers than do those that ignore such costs. In the face of informational limitations and other forms of market failure, a private insurer may not provide optimal levels and types of insurance coverage, but one that ignores costs borne by the subscriber is unlikely to survive long in the marketplace.

Government programs can also act as payers or as providers (as does Great Britain's National Health Service); the same consideration applies to them. Some government functionaries may consider only the costs that their agencies or programs bear. Implicit in such a strict *government perspective* is an assumption that the health benefits the agency provides are relevant, but monetary benefits and costs, unless directly borne by the agency, are not. Such a point of view, even if widely held by government officials, is at odds with the overt aim of such programs: to serve citizens. The beneficiaries of such programs care about the costs that they bear themselves, in addition to the health improvements that result from the services that they receive. Officials who hold a narrow governmental perspective might recommend extensive centralization of clinical services so that, for example, a diabetic might need to travel for several hours for a routine office visit. Surely the inconvenience and cost to the patient, if regularly ignored, would have repercussions for the official, the agency, and the government. The consequences might not be equally severe or immediate in every society or political system. Nevertheless, government agencies must be concerned about their budgets and the costs and benefits to the populations that they serve. Thus the societal perspective is informative even for payers, government agencies, and other entities that would seem to have an interest in a more limited range of costs.

The challenge of fixed costs

Implementation of the societal perspective can be difficult, especially when the production of a health intervention requires high fixed costs. The societal perspective usually implies that health services should be used to the point where marginal costs equal the value of the marginal

gain in health outcomes. But in the presence of significant fixed costs, price deviates substantially from marginal cost. Large investments for research and development are necessary before many drugs and medical devices can be marketed. Marginal costs of production fail to account for the substantial development investments that are characteristic of pharmaceuticals. Typical recommendations to use marginal costs in CE analysis differ strikingly from typical practice, which uses some measure of the sales price of medications. Price is often many multiples of the marginal cost of producing a drug, at least while the drug is still under patent protection. Many of the same issues arise in joint production and in other situations in which costing is ambiguous.

For the most part, the CE literature gives little guidance on this subject. There is widespread understanding that neither charges nor actual payments for health care are necessarily equal to costs of production, at least as defined in conventional economic terms (Finkler (1982)). The Panel on Cost-Effectiveness in Cost and Medicine, noting that cost should represent an opportunity cost, went well beyond most of the published CE literature in discussing in comprehensive terms what the alternative measures of cost are, and what measures are theoretically justifiable. The Panel generally urged that long-term marginal costs should be used as the basis for costs, but the specific recommendation depended on the question being asked. They recommended that "fixed costs . . . should be excluded . . . costs should not be included for inputs or outputs that are unaffected by changes in the intensity or frequency of an intervention." The panel then made the observation that in the long run there are few fixed costs.

In a discussion of R&D and "first-copy" costs, the report reiterated the recommendation, stating "if the technology has already been developed and the decision addresses the use of the intervention, such as dosage of a drug or frequency of a screening test, then the price should exclude R&D costs. Instead, the relevant costs are the incremental production, distribution, and provision costs." Thus, it suggested that the first-copy or fixed R&D costs should be ignored, implying that the CE analysis should use the marginal cost of the intervention even if the price paid (as for a drug) would often be substantially higher.

This approach might correspond to the outcome that we would seek from a cost-benefit analysis in which we attempted to maximize welfare by adding consumer and producer surplus. The usual teaching (that is, abstracting from the difficult problem of determining how to pay the fixed costs) is that the socially optimal level of consumption would be the point at which the marginal benefits equal the marginal costs (see Figure 4), which might be low for a drug.

In a static partial equilibrium analysis that level of consumption would be Pareto optimal, and the effects of changes in price would be purely distributional. As Figure 4 shows, the revenues to a monopolist under an allocation that used marginal costs for the CE criterion but required payment of monopoly prices would lead to larger revenues for the producer than under the conditions of monopolistic supply and competitive demand (price-taking purchasers).
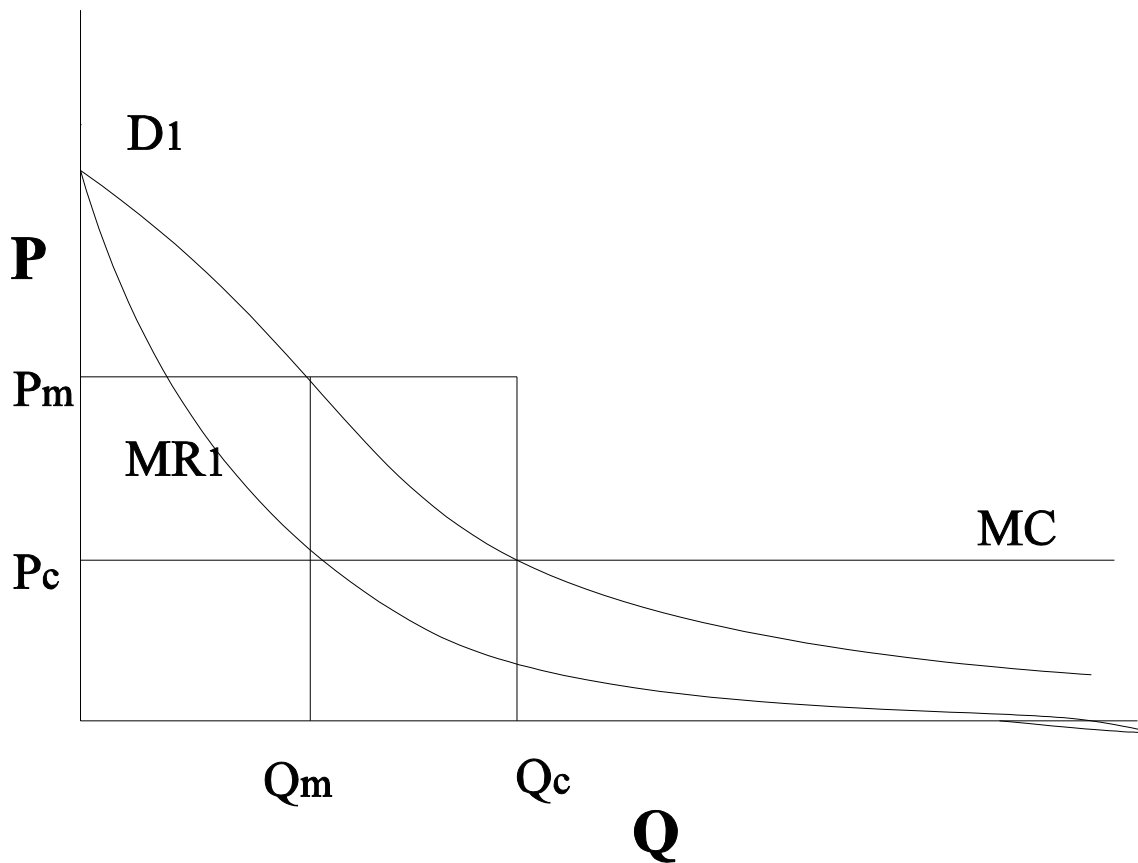

Figure 4. Monopolistic pricing and competitive quantities. The classic monopolist chooses the quantity to set marginal revenue to marginal cost, indicated by Q_m, and adopts the price corresponding to that quantity on the demand curve, P_m. Presumably implementation of a CE criterion with quantity set according to marginal cost pricing would result in the competitive quantity Q_c, but the price would be P_m rather than P_c. Monopoly revenues would therefore be P_m*Q_c rather than $P_m * Q_m$. If the purchasers are not price-takers the market behavior might correspond more closely to bilateral monopoly, so that the price paid might be less than P_m.

Despite the seeming clarity of their recommendation for excluding fixed costs, the Panel's discussion does not provide unambiguous guidance when fixed costs are substantial. The Panel seemed uncomfortable mandating that only this perspective on costs would be appropriate. Although the Panel did not state this explicitly, if a government agency or insurer announced that it would make

coverage or provision decisions based on decision rules that ignored fixed or first-copy costs, they would directly influence research and development decisions for future products and services by assuring high rewards to innovation. In other words, although fixed and first-copy costs for existing technologies have already been borne, investments in fixed costs are endogenous and dependent upon expected revenues, which in turn depend upon the rule for handling such costs in CE analysis.

Recognizing that the authors and readers of CE analysis are rarely concerned with producer's surplus and rents, the Panel's report leaves room for other perspectives:

> . . . For perspectives other than societal, the price paid by the decision maker for the good or service is the relevant one, inclusive of whatever return on investment in R&D or rent to patent- or copyright-holder has been incorporated in to the price. If a patient or insurance carrier pays a price for zidovudine (AZT) that reflects patent restrictions, for example, the relevant price for a CEA is the one paid, not the opportunity cost of the inputs that went into producing the actual units of AZT consumed . . .

Since the Panel generally endorsed the societal perspective, what justification can there be for this more limited perspective? Is this perspective appropriate when there are high fixed or first-copy costs?

This more limited perspective is used in most CE analyses of drugs, suggesting that few analysts consider the full societal perspective to be the appropriate one in this context. Few purchasers of health care would be interested in an analysis that evaluated CE of an intervention by assuming a cost much lower than the price at which they could obtain it. That may be why the Panel gave such explicit, and favorable, attention to a perspective that was not societal in the context of high fixed costs. But is the usual practice excessively narrow, ignoring benefits to the producers of interventions?

There is little question about the importance of this issue. New drugs and medical devices are almost always produced by monopolists (albeit sometimes competing with close substitutes), so the disparity between price and marginal cost is large. According to a comprehensive report on pharmaceutical R&D published by the Congressional Office of Technology Assessment in 1993, in

36

the U.S. the cost of bringing to market a drug whose R&D was initiated between 1970 and 1982 was about $194 million (U.S. Congress Office of Technology Assessment (1993)). This figure is open to debate, and industry sources claim the cost is $250 million or more. Nevertheless, there is no doubt that profits require charging more than marginal cost. Marginal costs — in particular, the costs of manufacturing additional units of a drug — are proprietary information, and are generally unknown.[1] However, because the original producer of a drug is usually believed to have the lowest manufacturing costs (since it is a large-scale producer), the prices of generic compounds after patent expiration give upper limits on the marginal costs, and these prices are often small fractions of the prices charged during the period of patent protection. Thus, the disparity between price and marginal cost is likely to be large for most drugs that are under patent protection. Although the same may be true of devices, they have been studied less and production costs may account for a larger share of their average costs.

By recognizing what CE analysis can do best, we can begin to reconcile the contradiction between the usual practice and the usual recommendation of adopting a societal perspective, i.e., one that includes all costs and ignores fixed costs. The technique is not particularly useful for determining the full social optimum, particularly in a dynamic context with large fixed costs, and it is rarely used for that purpose. Instead, *the relevant perspective in most cases is that of consumers and their agents*.

The perspective is essentially that of a perfect insurer, as defined in the Garber and Phelps paper.   Mark Pauly has argued that a similar perspective, that of a managed care organization, is often the best one to use in thinking about health care allocation decisions (Pauly (1995)). This perspective differs from a full societal perspective by ignoring producer surplus. Because the producer surplus is a real component of welfare, government or society should not ignore it. But the practical challenges that must be overcome to maximize the combined surplus by using CE analysis

---

[1]As part of a study for the Office of Technology Assessment, my colleagues and I attempted to determine the R&D costs and production costs for a very expensive drug (alglucerase) used to treat Gaucher disease, an uncommon genetic disorder. Although we were able to discuss the costs and view internal accounting documents from the company, it was very difficult to ascertain the manufacturing costs and the R&D costs. Production of alglucerase, which was made by chemical modification of an enzyme found in human placentas, was unusually expensive, but nevertheless we estimated that the price of the drug was about twice the marginal cost. The R&D costs born by the company were relatively small, since the drug was discovered by federal scientists and licensed to the company; see Garber, et al. (1992).

are considerable. For example, if "society" is a province of Canada and the intervention in question is a drug produced by an American company with investors from around the world, Canadians who give greater weight to benefits that accrue to other Canadians will not weigh the company's profits as highly. If the drug or other intervention cannot be obtained at marginal cost, and if health budgets are constrained, can there be any assurance that the attempted application of a CE criterion based on marginal cost will lead to an optimal distribution? A health plan or program that strictly applies the marginal cost concept will treat the costs of two drugs as if they are equal, if the marginal costs of production are similar, even if the price of one is ten times as great as the price of the other.

The attempt to invoke a full societal perspective raises both theoretical and practical difficulties. For example, if buyers purchase pharmaceuticals to the point at which marginal cost and marginal benefits are equal, but pay a monopoly price, monopoly profits should be substantially greater than under the conventional monopoly equilibrium (at which marginal revenue equals marginal cost; see Figure 4). Although the resulting allocation might be Pareto-optimal in a static world, it creates incentives that might cause distortions in investment decisions. The extraordinary profits would induce overinvestment in the development of new pharmaceuticals. Furthermore, as the preceding discussion noted, marginal costs (particularly for drugs still under patent) are usually unknowable, since they constitute proprietary information.

The approach that uses a full societal perspective, with marginal costs as the measure of the costs, implies the need for a nonmarket method of financing. Application of the CE threshold implies that the quantity of a drug purchased will be larger if the CE cost assigned to the intervention is marginal cost rather than the purchase price. To estimate the full optimum, the analyst would have to take into account distortions induced by the method of financing, such as deadweight losses due to income taxation for financing government health care programs. The behavioral change induced by tax incentives can be large, so that the cost of obtaining funds via taxation can greatly exceed the money raised. It is likely that the distortions induced by the modes of financing private health insurance are also large. The distortions introduced by the method of financing present a problem for any attempt to use a CE (or cost-benefit) framework to determine a full social optimum. The marginal cost criterion, with the implied increase in quantity consumed, will exacerbate the problem.

Distributional considerations

Distributional concerns about CE analysis are raised frequently; such concerns are also prominent in the most vociferous objections to application of CB analysis. Nearly every public program for health care is intended to mitigate inequalities in health, in part by ensuring that the poor have access to effective care. Thus, many discussions of the desirability of CE allocations consider distributional consequences. A strong emphasis on the magnitude of producer's surplus would be incongruous for those nations and groups with deep beliefs about the importance of distributive justice, especially insofar as the owners of companies that produce pharmaceuticals and other health care products are drawn from the upper ranks of the distribution of income and wealth.

Summary: costs and perspective

Fundamentally, the major issue in defining costs for CE analysis revolves around the definition of opportunity cost. Ordinarily, prices are reasonable proxies for costs. But numerous market imperfections imply that prices are not always good proxies for marginal costs of health care. Because the value of the cost estimate has implications for the adoption and scale of utilization of health interventions when CE analysis is used to aid decision making, these are not merely technical issues. In real-world situations in which the method is likely to be used, the attempt to implement a societal optimum by using nebulous marginal cost figures and purchasing goods and services as if the cost equaled the marginal cost may be unhelpful. Many of the controversies about costs disappear, or at least the problems are mitigated, when analysts present the form of consumer perspective suggested here, in which the premium and out-of-pocket costs of consumers purchasing idealized insurance are the basis for direct cost measurement. Producer benefits also matter, but CE analysis does not offer a comprehensive framework for evaluating them, particularly in a dynamic context. Thus, this perspective is both meaningful and understandable, and is the appropriate perspective for many government agencies, private payers, and providers making decisions about health care.

5. Measuring Outcomes

According to the preceding discussion, the welfare economic foundations of CE analysis rest upon the validity of the outcome measure as a representation of utility. This aim was not explicit in

the initial development of outcome measures for CE analyses in health care. Whether the purpose of the CE analysis is to maximize utility or to maximize a global measure of health-related quality of life, however, its credibility depends heavily on the comprehensiveness and relevance of the health outcome measure. A highly specific outcome or effectiveness measure like the yield of abnormal test results or the magnitude of the blood pressure response to an antihypertensive drug may be understandable, persuasive, and sensitive to the effects of the intervention under study. But such a measure cannot be used to compare a diverse set of health interventions to be administered to patients with different health conditions. Furthermore, despite occasional claims and implicit assumptions to the contrary, only rarely will such a measure capture all the potential benefits and harms of an intervention. Thus, a comprehensive and general measure of health outcomes is of fundamental importance, whether the analysis is to be justified by appeal to welfare economics or by simple appeal to the inherent plausibility of the health measure.

It is for these reasons that QALYs are most frequently recommended as the outcome measure for CE analysis. More general alternatives, like healthy-year equivalents (HYEs) have attractive theoretical properties (Gafni and Birch (1997), Mehrez and Gafni (1989), Mehrez and Gafni (1993)) but have not gained widespread acceptance, probably because they are perceived as difficult to implement (Johannesson, et al. (1993), Gold, et al. (1996)). The following brief discussion emphasizes measurement of the preference weights q_i that appear in Eq. 3.

Steps to measuring QALYs

Three components are needed to calculate an individual's utility at a point in time. First is the definition of the health state in question, which might be a particular disease with specific symptom severity; second is the utility attached to that health state, and third is the probability that the individual will be in that health state. By summing the products of the utilities of each possible health state and their probabilities, it is possible to obtain the expected utility (or QALY contribution) corresponding to the time period in question. This formulation has the advantage of breaking the task of calculating QALYs into manageable components: description of the health state; assessment of utilities toward the health state; and estimating the probability of the health state.

Defining and describing the health state are fundamental to modeling effectiveness. The CE analysis must include each state of health that the intervention might affect, either by preventing or

treating illness, or by causing side-effects. Thus, if the intervention under study is surgery for the treatment of coronary artery disease, important health states to model include the presence and severity of angina pectoris, heart attacks, and other symptoms of heart disease or complications of the procedure (or, for that matter, of any alternatives to which it is compared). The scope of available data and analytical tractability limit the number of health states that can be modeled. Many analyses use Markov modeling and related techniques to describe the progression over time of the probabilities of various health states, and if too many health states are included, there may be few or no transitions between infrequently occurring health states, precluding reliable estimation of some of the parameters of the model.

Dolan's chapter discusses how preference assessment is performed to estimate the utilities or quality weights specific to each health state. A critical issue for preference assessment is whether the respondent — the person whose preferences are being assessed — is asked to place a value on his own current or recent state of health, or is instead asked to place a value on a hypothetical state of health. For example, the preference questions could be directed toward people known to have a particular health state, such as moderately symptomatic coronary artery disease, and they could be asked how their current state of health compares to an ideal state of health. The alternative is to provide a description of a hypothetical state of health and to ask respondents to imagine themselves in that health state and to rate it.

There are several difficulties with rating one's own health state. First, the preferences of people experiencing a state of (usually chronic) ill health may differ systematically from the preferences of the general population. In the face of a disparity, there is no strict consensus about whose preferences should be used. The Panel on Cost-Effectiveness in Health and Medicine argued that when societal (i.e. governmental) resources are used to pay for health care, the preferences should be those of the general population rather than those of individuals with a health condition (Gold, et al. (1996)). Furthermore, it is difficult to study large samples of individuals who have a specific health condition, especially if the condition is uncommon. It is also possible that the disutility associated with a health state may reflect co-existing health conditions or risk factors that predispose to the disease rather than the disease itself. For example, high blood pressure is an asymptomatic condition that increases the risk of heart disease and stroke. People with high blood pressure rate their own health as relatively poor, even when they have not suffered any complications. Because treatment lowers the blood pressure but does not remedy associated health

conditions, it does not improve quality of life as greatly as would be predicted from a model in which preferences are obtained from people with the disease and treatment is assumed to restore them to perfect health.

The validity of the alternative approach, *rating hypothetical health states*, is highly dependent on the accuracy and completeness of the description of the hypothetical state(s). The health state description is not critical for a state of health that most respondents have experienced, such as the symptoms of a viral upper respiratory infection or mild low back pain. But for a health state that few respondents have experienced themselves or vicariously through a relative or friend, nearly all the information that the respondent can bring to bear on the question must be provided in the description. This requirement can be an advantage, since it is easier to control the impression that naive respondents have of the health state than the impressions of experienced respondents. But it also means that small and seemingly inconsequential changes in the presentation of the health state can greatly influence the utilities assigned to it. To enhance the reproducibility and validity of ratings of hypothetical states, it is essential to pay close attention to the wording and general design of such elicitations.

Estimating survival and probabilities of health states

Even for interventions that do not alter the length of life, it is usually necessary to describe patterns of survival since these patterns determine the changes in QALYs that result from use of the interventions. Many treatments, of course, are designed to prevent death, so estimation of survival effects, or the survival probabilities in Eqs. 2 and 3, is a key component of most CE analyses.

Approaches to measuring survival probabilities vary greatly. Survival estimates nearly always require an element of modeling, since experimental data (from a randomized trial) are usually limited to brief (less than five years) follow-up periods. To estimate the effect on life expectancy, it is necessary to combine such data with observational data about longer-term outcomes in typical practice settings.

The techniques for estimating the pattern of survival associated with an intervention vary. One study of a treatment for heart attacks shows how clinical trial and observational data can be combined to estimate long-term outcomes. Researchers from the GUSTO trial, a study of tissue-type plasminogen activator (t-PA), a drug used to dissolve the blood clots that can cause obstructions in the coronary arteries and precipitate heart attacks, sought to determine the long-term

survival benefit by supplementing direct clinical trial data, obtained during an average of 12 months of follow-up, with a model of survival based on an observational database (the Duke Cardiovascular Registry), and a parametric survival function for extrapolating beyond the 14 years of data represented in the observational database. Figure 5 displays the resulting survival curve. Published CE analyses have used a variety of other methods. Some analyses used life table data for either the general population or, where available, for patients who have a specific health condition, and applied a relative risk reduction as estimated in a clinical trial, imposing the assumption that the relative risk reduction is constant across different populations and ages.

By generalizing the methods for estimating survival, one can also estimate probabilities that various states of health will occur in the future, under either the treatment or the intervention. Usually Markov-like modeling offers the most convenient approach to estimating future probabilities of health states. One such approach estimates first the probability that an individual receiving the intervention is alive, say, two years in the future, then uses data from clinical trials or other sources to estimate the probability that, if alive, the patient will be in a symptomatic state of ill-health, and the probability that he or she will be in excellent health. Typically availability of data on rates of adverse events (such as onset or progression of disease, death rates, and morbidity), rather than technical issues (such as the formal structure of the model to depict disease advancement), limits the estimation of probabilities of health states.

Preference assessment

The remaining step in calculating QALYs is to assign utilities, or preference weights, to each of the health states. Several reviews describe and compare alternative methods for preference assessment, and Dolan (1999) discusses the topic extensively in this volume. Dolan reviews a wide range of issues in assessing preferences and in their interpretation from the point of view of QALY calculation. As his discussion of the methodological issues in assigning utilities to health states implies, preference assessment is sometimes a source of considerable uncertainty in CE analyses. The most reproducible methods of preference assessment, such as the visual analog scale, are not derived from von Neumann-Morgenstern utility theory. Methods that are more firmly grounded in utility theory, such as the standard gamble, are neither perfectly general nor easy for respondents to understand.

43

Since the validity of CE analysis as a guide to welfare maximization rests upon the validity of QALYs as a measure of utility, the conditions that preference assessment needs to meet are stringent. Usually discussions of quality of life for use in CE analysis emphasize that the measurement should be of *health-related* quality of life. Well known preference-weighted health status indices used to attach utilities to health states — such as the Health Utilities Index of Torrance and colleagues, the Quality of Well-Being scale developed by Kaplan and colleagues, and the Rosser scale — omit mention of non-health consumption and financial status (for an extended discussion of these and other scales, see
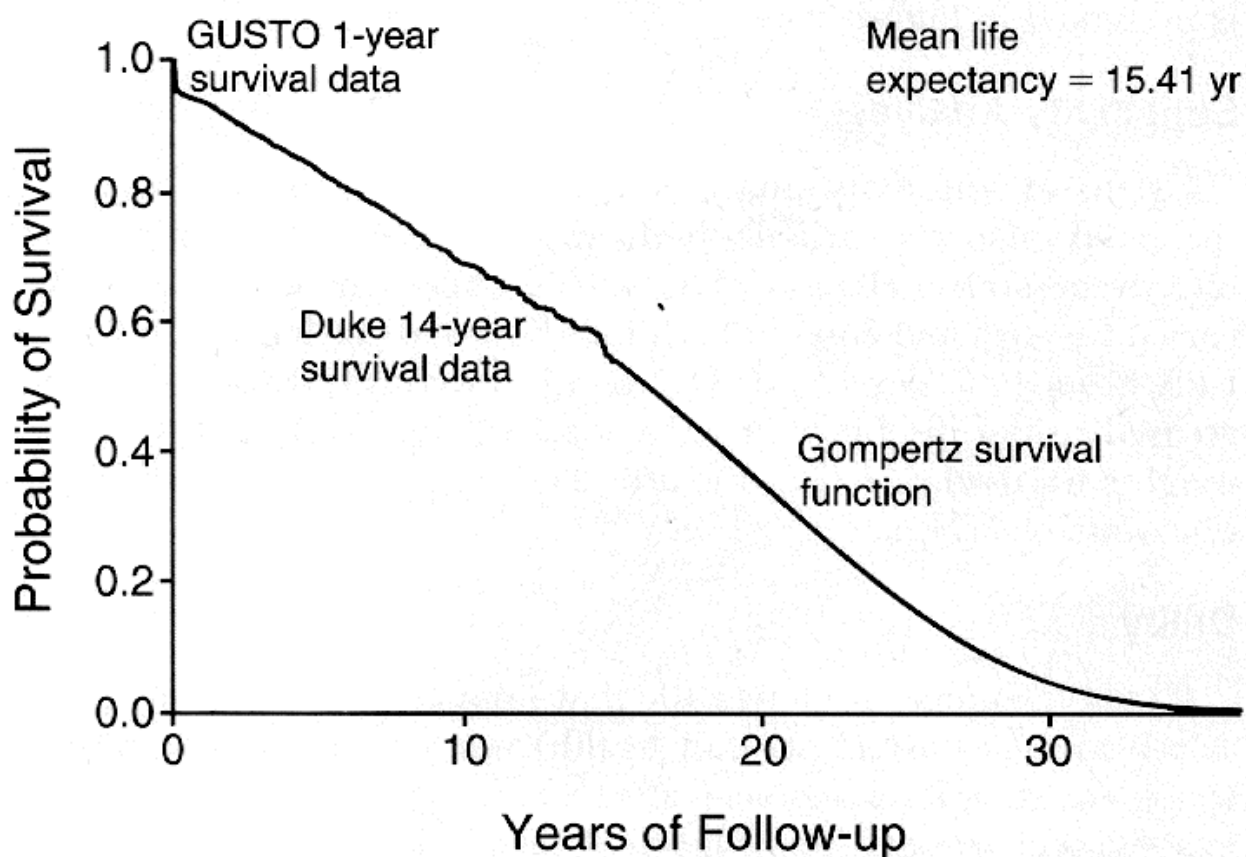

Figure 5. Probability of survival among patients treated with t-PA. A survival function of this type was used to estimate life expectancy for each treatment group. The curve consists of three parts: the survival pattern in the first year after treatment in the GUSTO study, data for an additional 14 years on survivors of myocardial infarction in the Duke Cardiovascular Disease Database, and a Gompertz parametric survival function adjusted to agree with the empirical survival data at the 10-year and 15-year follow-up points.

Source: Mark, et al. (1995). Reproduced by permission. Copyright © 1995 Massachusetts Medical Society. All rights reserved.

the book by Patrick and Erickson (1993)). According to some experts, respondents should be asked to ignore effects of states of ill health on income and other financial repercussions. Yet the plausibility of QALYs as measures of utility depends on the ability to represent fully the changes in well-being that occur with the adoption of an intervention, and often these changes will not be limited to those that are primarily health related. Such concerns may be of little importance if the only financial consequence is loss of earned income, which ordinarily would be incorporated into the numerator of the CE ratio. But if a health state causes alteration of non-health consumption, which is not reflected in the preference assessment procedure (e.g., development of severe arthritis may necessitate changes in clothing, furniture, and use of various non-health services), the adverse effects of the health state will be underestimated.

Preference heterogeneity and its consequences for CE analysis

Perhaps the greatest practical challenge to the use of QALYs to represent utilities is the variation in preferences that is all but certain to occur in the context of specific health limitations. Just as demand for any good or service varies, so do preferences for states of health. A well-known study of treatment of benign prostatic hyperplasia, which causes a variety of urinary symptoms, demonstrated that variation in attitudes toward specific health limitations can dramatically alter the value of treatment. The most common surgical treatment of prostatic disease is transurethral resection of the prostate, an operation that can be highly effective at relieving the excessive urinary frequency and nocturia (awakening at night to void) and other symptoms that men with prostatic obstruction experience.   The operation, however, can cause incontinence, impotence, and other side-effects, some of them permanent. Men who are candidates for surgery vary greatly in their relative preferences for the symptoms of prostatic hyperplasia and the side-effects of the operation, so that the expected quality of life is greater with surgery for some and with nonsurgical management for others (Barry, et al. (1988), Fowler, et al. (1988)).

Without even considering costs, then, the "best" treatment varies when preferences vary. When CE is used as a criterion for determining the allocation of interventions, preference variation often poses more significant problems. It is possible that every patient who is a candidate for treatment with a particular intervention will gain QALYs from it. But the intervention is much more cost-effective in those patients who experience the greatest disutility from the disease being

treated, and who lose little utility from the side-effects of treatment. Other patients who have identical health characteristics may experience little disutility from the disease and more from the treatment. It is very hard for any health care delivery or financing system to distinguish these two types of patients, both of whom would desire the intervention. Although individual clinical decisions can take such heterogeneity into account, even in the physician's office the necessary information, and the ability to use it, may be limited.

QALY measurement and the application of CE analysis

Technical issues in QALY measurement raise questions about the reliability and validity of QALYs, as usually calculated, as measures of utility. One message from the literature that uses weights based on preferences rather than statistical weights or simple sums to measure quality of life is that comprehensive measures of utility are difficult for study subjects to understand. The reproducibility of such measures, particularly when the underlying preference assessment technique is as complicated as the standard gamble, is often disappointing. The limitations of such measures are partly responsible for the popularity of quality of life measures that are not preference weighted (such as the Rand Corporation's SF-36 scale) or that are not even global measures of quality of life (such as disease-specific quality of life scales). Although these alternative measures offer apparent practical advantages, rarely can they be considered reasonable proxy measures of utility. The major conceptual problem with the preference assessment measures as usually applied is that they do not allow the state of being to be construed broadly enough, a problem that is far worse for disease-specific measures. Measures that are not preference weighted lack the interval scaling properties required for the tradeoff between length and quality of life implicit in QALYs.

The practical problems are particularly great when the benefit from a health intervention is small. Consider, for example, a medicated lotion that relieves the itch of a rash that appears on the arms and back. Even if the lotion completely relieves the rash as soon as it is applied, it will be extremely difficult to assess utilities for the relief of the rash using standard preference assessment techniques. All of the techniques require a tradeoff between a risk of death and symptom relief, but if the symptoms are mild or their duration is brief enough, it is difficult for respondents to estimate the risk of death (or for the time-tradeoff method, the reduction in the length of life) that they would tolerate for an improvement in the symptoms. For this intervention and others that produce small or brief improvements in quality of life, the willingness-to-pay approach used in CB analysis would likely offer a much more suitable approach to valuation.

An ideal measure of health outcomes would be less restrictive than QALYs, abandoning the additive separability embedded in the functional form and the (usually) constant rate of time preference, but preference assessment instruments capable of supporting more general models would impose upon respondents even greater cognitive burdens than current methods. Research on these methods remains active, in some cases reflecting the great interest of governments in applying CE analysis to health care decisions more extensively. As utility measurement improves, claims that the results of CE analysis can be applied to maximize social welfare can be made with greater confidence. Furthermore, although the QALY is not perfectly general as a measure of well-being, it is likely to be a close approximation to more general measures and to represent an acceptable tradeoff between conceptual validity and feasibility. Unlike many competing measures of quality of life, such as the statistically-weighted quality of life indices, QALYs are conceptually appropriate and have the potential to approach the theoretical ideal when preference assessment techniques are developed further.

6. Recommendations

A fundamental but often unstated characteristic of any CE analysis is its purpose. Is that purpose to enable an insurer, a health plan, or a government agency to decide whether to cover a specific intervention? Is it to help a consumer decide which form of treatment to receive? Is it to help a manager make decisions about large investments in health care infrastructure? Is it to help a formulary committee choose which of several drugs should be available in a hospital pharmacy? Or is it to help a decision maker determine the allocation of health care that will achieve a suitably defined social optimum, regardless of who that decision maker is?

Most experts in CE analysis argue that, unless there are compelling reasons to do otherwise, CE analyses should be conducted from the societal perspective. Under this perspective, all costs and all benefits are relevant, but usually analysts assume that the health benefits accrue entirely to the individual receiving care. Exceptions are sometimes made in other circumstances, such as when there are significant externalities. For example, family members may provide care or other people may bear a cost when an individual is injured or ill. Even in the absence of externalities, though, an attempt to use CE analysis to determine a full societal optimum, while laudable, in important circumstances may stretch the technique to the breaking point. Even for a circumscribed measure of optimality like the Kaldor-Hicks criterion (i.e, potential Pareto improvement), such determinations

47

may be difficult for products characterized by economies of scale and by other failures of the assumptions of perfectly competitive markets. How and whether to include the preference of producers in a CE analysis are certain to be controversial, particularly when the profits accrue in a population markedly different from the one that is being treated. Profits are certainly a component of overall welfare, and to remove them from the CE analysis is not the same as saying that they are unimportant. CE analysis does not provide a comprehensive framework for including them.

As common practice dictates, and the abilities of the technique mandate, most CE analyses should be conducted from a consumer-oriented perspective, but not from the one that is generally described as the consumer's or patient's perspective. Rather, the most robust perspective is that of an insurer acting as a perfect agent for its enrollees. Specifically, it assumes that the members of the defined population are behind a "veil of ignorance," having no particular information to distinguish their risk of developing any disease or health condition or desire to utilize services from the average for the defined population. The insurer charges an actuarially fair premium, and has no costs other than the payment of benefits. There are no informational failures of consequence, other than symmetric uncertainty, in the sense that neither the insurer nor any individual has more or less information than others.

Perhaps the most difficult challenge for the implementation of CE analysis is the technique's application in heterogeneous populations. The optimality properties of the CE approach are based upon the application of an individual's specific CE ratio cutoff to decisions about care. For that individual, any intervention whose CE ratio is below the cutoff is welfare-enhancing (i.e., passes a CB criterion), whereas any with a greater CE ratio does not. But for many reasons — income, risk preferences, and various other attitudes and values — CE cutoffs vary greatly across individuals. Many, if not most, CE analyses are used to inform decisions made at a group level yet implicitly apply a single cutoff. Decisions based on a single cutoff cannot claim to have the same optimality properties in a heterogeneous population. The cutoff will be greater than the actual cutoff for some people, and less than the actual cutoff for others. Furthermore, the optimal single cutoff for a heterogeneous population would not necessarily correspond to the average valuation.

The preceding discussion suggests that the welfare implications of the application of CE analysis are clearest when strong conditions are met. The research challenges include better measurement – for example of health outcomes, preferences, and costs – and further investigation into the implications of using CE analysis when ideal conditions do not apply. The measurement of

preferences is an area of ongoing research, and it would be helpful to compare the results of analyses that use QALYs with those that use either simpler measures of health outcomes (e.g., life expectancy) or more comprehensive measures (e.g., healthy year equivalents). Further investigation of the theoretical issues would help to clarify the meaning and generalizability of the results of CE analyses. For example, what are the welfare implications of prioritization based on CE ratios when some health services are subsidized but a number of substitutes for them are not? What are the implications of inter-individual variation in rates of time preference? What are the welfare gains from using individual rather than uniform CE cutoffs in heterogeneous populations? Under what circumstances are simple CE analyses accurate guides to welfare maximization?

CE analysis can be a useful aid to decision making in health care. In specific circumstances it can be quite powerful. Yet its grounding in welfare economics has often been implicit, and an explicit examination of how one can use a CE criterion to achieve a potential Pareto improvement demonstrates that the necessary conditions are exacting. Nevertheless, of widely accepted, existing methods for incorporating economic considerations in the prioritization and allocation of health care, CE analysis is probably the most rigorous. Exploration of its welfare economic foundations has the additional advantage of helping to resolve ambiguities in matters such as the measurement of costs, and can help to inform the development of new instruments for measuring quality-of-life effects. CE analysis is not a perfect tool, but in many situations, it may be good enough.

REFERENCES

Barry, M. J., A. G. Mulley, Jr., F. J. Fowler and J. W. Wennberg (1988), "Watchful waiting vs immediate transurethral resection for symptomatic prostatism. The importance of patients' preferences," Journal of the American Medical Association 259: 3010-3017.

Birch, S. and A. Gafni (1994), "Cost-effectiveness ratios: in a league of their own," Health Policy 28: 133-141.

Briggs, A. and M. Sculpher (1995), "Sensitivity analysis in economic evaluation: a review of published studies," Health Economics 4: 355-371.

Briggs, A., M. Sculpher and M. Buxton (1994), "Uncertainty in the economic evaluation of health care technologies: the role of sensitivity analysis," Health Economics 3: 95-104.

Dolan, P. (1999), "The Measurement of Health-Related Quality of Life for Use in Resource Allocation Decisions in Health Care," in J. P. Newhouse and A. J. Culyer, eds., Handbook of Health Economics (North Holland, Amsterdam) (in press).

Drummond, M. F., B. O'Brien, G. L. Stoddart and G. W. Torrance (1997), Methods for the Economic Evaluation of Health Care Programmes (Oxford University Press, New York).

Drummond, M. F., G. L. Stoddart and G. W. Torrance (1987), Methods for the Economic Evaluation of Health Care Programmes (Oxford University Press, Oxford).

Finkler, S. A. (1982), "The distinction between cost and charges," Annals of Internal Medicine 96: 102-109.

Fowler, F. J., Jr., J. E. Wennberg, R. P. Timothy, M. J. Barry, A. G. Mulley, Jr. and D. Hanley (1988), "Symptom status and quality of life following prostatectomy," Journal of the American Medical Association 259: 3018-3022.

Gafni, A. and S. Birch (1997), "QALYs and HYEs (healthy years equivalent). Spotting the differences," Journal of Health Economics 16: 601-608.

Garber, A. M., A. E. Clarke, D. P. Goldman and M. E. Gluck (1992), Federal and Private Roles in the Development and Provision of Alglucerase Therapy for Gaucher Disease (Congress of the United States, Office of Technology Assessment, Washington, D.C.).

Garber, A. M. and C. E. Phelps (1997), "Economic foundations of cost-effectiveness analysis," Journal of Health Economics 16: 1-31.

Garber, A. M. and N. A. Solomon (1999), "Cost-effectiveness of alternative test strategies for the diagnosis of coronary artery disease," Annals of Internal Medicine 130:719-728.

Gold, M. R., J. E. Siegel, L. B. Russell and M. C. Weinstein, eds. (1996), Cost-Effectiveness in Health and Medicine (Oxford University Press, New York).

Hicks, J. R. (1939), "Foundations of welfare economics," Economic Journal 49: 696-712.

Hurley, J. (1999), "The normative economics of the health sector," in J. P. Newhouse and A. J. Culyer, eds., Handbook of Health Economics (North Holland, Amsterdam) (in press).

Johannesson, M., J. S. Pliskin and M. C. Weinstein (1993), "Are healthy-years equivalents an improvement over quality-adjusted life years?" Medical Decision Making 13: 281-286.

Johannesson, M. and M. C. Weinstein (1993), "On the decision rules of cost-effectiveness analysis," Journal of Health Economics 12: 459-467.

Kaldor, N. (1939), "Welfare propositions of economics and interpersonal comparisons of utility," Economic Journal 49: 549-552.

Karlsson, G. and M. Johannesson (1996), "The decision rules of cost-effectiveness analysis," PharmacoEconomics 9: 113-120.

Keeney, R. L. and H. Raiffa (1993), Decisions with Multiple Objectives (Cambridge University Press, Cambridge).

Kenkel, D. S. (1997), "On valuing morbidity, cost-effectiveness analysis, and being rude," Journal of Health Economics 16: 749-757.

Luce, B. R., W. G. Manning, J. E. Siegel and J. Lipscomb (1996), "Estimating costs in cost-effectiveness analysis," in: M. R. Gold, J. E. Siegel, L. B. Russell and M. C. Weinstein, eds., Cost-Effectiveness in Health and Medicine (Oxford University Press, New York) 176-213.

Mark, D. B., M. A. Hlatky, R. M. Califf, C. D. Naylor, K. L. Lee, P. W. Armstrong, G. Barbash, H. White, M. L. Simoons and C. L. Nelson (1995), "Cost effectiveness of thrombolytic therapy with tissue plasminogen activator as compared with streptokinase for acute myocardial infarction," New England Journal of Medicine 332: 1418-1424.

Mehrez, A. and A. Gafni (1989), "Quality-adjusted life years, utility theory, and healthy-years equivalents," Medical Decision Making 9: 142-149.

Mehrez, A. and A. Gafni (1993), "Healthy-years equivalents versus quality-adjusted life years: in pursuit of progress," Medical Decision Making 13: 287-292.

Meltzer, D. (1997), "Accounting for future costs in medical cost-effectiveness analysis," Journal of Health Economics 16: 33-64.

Mullahy, J. and W. G. Manning (1994), "Statistical issues in cost-effectiveness analysis," in: F. Sloan, ed., Valuing Health Care: Costs, Benefits and Effectiveness of Pharmaceuticals and Other Medical Technologies (Cambridge University Press, New York) .

O'Brien, B. J., M. F. Drummond, R. J. Labelle and A. Willan (1994), "In search of power and significance: issues in the design and analysis of stochastic cost-effectiveness studies in health care," Medical Care 32: 150-163.

Patrick, D. L. and P. Erickson (1993), Health Status and Health Policy: Quality of Life in Health Care Evaluation and Resource Allocation (Oxford University Press, New York).

Pauly, M. V. (1968), "The economics of moral hazard: comment," American Economic Review 58: 531-537.

Pauly, M. V. (1995), "Valuing health care benefits in money terms," in: F. Sloan, ed., Valuing Health Care: Costs, Benefits, and Effectiveness of Medical Technologies (Cambridge University Press, Cambridge) 99-124.

Phelps, C. E. and A. I. Mushlin (1991), "On the (near) equivalence of cost effectiveness and cost benefit analysis," International Journal of Technology Assessment in Health Care 7: 12-21.

Rawls, J. (1971), A Theory of Justice (Harvard University Press, Cambridge, MA).

Russell, L. L. (1986), Is Prevention Better than Cure? (The Brookings Institution, Washington, D.C.).

Schelling, T. C. (1968), "The life you save may be your own," Problems in Public Expenditure Analysis (The Brookings Institution, Washington, D.C.) 127-159.

Siegel, J. E., M. C. Weinstein and G. W. Torrance (1996), "Reporting cost-effectiveness studies and results," in: M. R. Gold, J. E. Siegel, L. B. Russell and M. C. Weinstein, eds., Cost-Effectiveness in Health and Medicine (Oxford University Press, New York) 276-303.

U.S. Congress Office of Technology Assessment (1980), The Implications of Cost-Effectiveness Analysis of Medical Technology (U.S. Government Printing Office, Washington, D.C.).

U.S. Congress Office of Technology Assessment (1993), Pharmaceutical R&D: Costs, Risks and Rewards, OTA-H-522 (U.S. Government Printing Office, Washington, D.C.).

Wakker, P. and M. P. Klaassen (1995), "Confidence intervals for cost/effectiveness ratios," Health Economics 4: 373-381.

Weinstein, M. C. and W. B. Stason (1977), "Foundations of cost-effectiveness analysis for health and medical practices," New England Journal of Medicine 296: 716-721.