NBER WORKING PAPER SERIES

ECONOMIC FOUNDATIONS OF COST-EFFECTIVE ANALYSIS

Alan M. Garber

Charles E. Phelps

Working Paper No. 4164

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
September 1992

Department of Veterans Affairs Medical Center, Palo Alto, California, Department of Medicine, Stanford University, NBER; Departments of Political Science and of Economics, Department of Community and Preventative Medicine, University of Rochester, NBER respectively. This research was supported in part by grants AGO7651 from the National Institute on Aging, HL4629 from the National Heart, Lung, and Blood Institute, and HS5477 from the Agency for Health Care Policy and Research. Much of the work was performed while Garber was a Henry J. Kaiser Family Foundation Faculty Scholar in General Internal Medicine. Garber is an HSR&D Senior Research Associate of the Department of Veterans Affairs. We are grateful to Elisabeth Coutts for expert research assistance, and to Andrew Dick, Dana Goldman, and Alvin Mushlin for helpful comments. This paper is part of NBER's research program in Health Care. Any opinions expressed are those of the authors and not those of the National Bureau of Economic Research.

NBER Working Paper #4164
September 1992

ECONOMIC FOUNDATIONS OF COST-EFFECTIVENESS ANALYSIS

ABSTRACT

In order to address several controversies in the application of cost-effectiveness analysis, we investigate the principles underlying the technique and discuss the implications for the evaluation of medical interventions. Using a standard von Neumann-Morgenstern utility framework, we show how a cost-effectiveness criterion can be derived to guide resource allocation decisions. We investigate its relation to age, gender, income level, and risk aversion. Cost-effectiveness analysis can be a useful and powerful tool for resource allocation decisions, but in the presence of heterogeneous preferences and personal characteristics, a cost-effectiveness criterion that is applied at the population level is unlikely to yield pareto-optimal resource allocations.

Alan M. Garber
NBER
204 Junipero Serra Boulevard
Stanford, CA 94305-8091
and NBER

Charles E. Phelps
601 Elmwood Avenue
Box 644
Rochester, NY 14627
and NBER

I. Issues in Cost-Effectiveness Analysis

A substantial body of literature uses cost-effectiveness (CE) analysis as a way to rank (or at least to provide guidance about) the desirability of using alternative medical interventions. Despite its many similarities to the usual cost-benefit (CB) analysis practiced by economists, cost-effectiveness is widely used by practitioners who consider it "different" from CB analysis (Charles E. Phelps and Alvin I. Mushlin, 1991). Many physicians, and others who perform CE analysis, prefer it to CB analysis because it does not require placing a dollar value on a health outcome. Typically, CE analysis describes an intervention in terms of the ratio of incremental costs per unit of incremental benefit (i.e., marginal cost/marginal benefit), in contrast to the usual CB analysis approach of describing net benefits of a project in dollars. CE studies translate the output of a medical intervention into a common denominator, such as life years saved. As techniques have emerged to "quality-adjust" those life years (see George W. Torrance, 1986 for an excellent summary), the Quality Adjusted Life Year (QALY, pronounced kwa-lee) has become the common currency for sophisticated CE analyses. Typical decision analyses about the use of an intervention compare the cost per QALY to that for other commonly used medical interventions, arguing that the use of a new technique or technology can be justified if it has at least as favorable a cost per QALY as generally accepted interventions.

Despite the widespread use of CE analysis, we are unaware of any published formal justification of the technique on the basis of first principles. The intuitive appeal of the logic of CE (minimizing the cost of producing a given level of health, or correspondingly, maximizing the achievable level of health for a given budget) sounds like a familiar economic problem, and for the most part, practitioners have assumed that CE analysis could be a tool for utility maximization.

Yet, even within this broad level of agreement (unsupported by any formal proof of the conclusion), a number of thorny problems remain prominent in the CE literature, including: (1) Should cost-effectiveness estimates include "unrelated" future medical costs incurred during years of life "extended" by a current medical intervention? (2) Does the use of life-years (or variants thereof) as measures of effectiveness discriminate against older persons? (3) Can one find the proper "cutoff" CE ratio in ways other than looking at the CE ratios of other commonly used medical interventions? This last issue is particularly challenging, since the range of cost-effectiveness rates for common interventions is wide. Furthermore, health insurance alters the incentives for using medical care, leading to the widely held belief that our society uses too much of it (Mark V. Pauly, 1986, and references therein). Although one might conclude that equalization of cost-effectiveness ratios at the margin is necessary for Pareto optimality, there is still the question of the proper level. We discuss each of these issues briefly, then turn to our formal model of utility maximization to answer them.

Future Medical Costs. The issue of including future medical costs creates a vague discomfort for the authors of many CE studies. There is no controversy about including in a CE analysis those future health care costs (or savings) that are directly attributable to the intervention. But what about health expenditures that result simply from living longer? "If we extend life," some authors argue, "then we will have to spend more for the medical care of future diseases. Therefore these medical costs are a consequence of the current treatment, and should count as a relevant cost." As Milton C. Weinstein and Henry V. Fineberg (1980) say in their classic text on clinical decision analysis:

> Often ignored are the costs of medical care received during extended years of life. Credit given to control of blood

> pressure for reducing costs associated with treatment of strokes and myocardial infarctions must be balanced against the costs for other diseases incurred during the added years of life. (p. 36)

Similarly, in their book on economic evaluation of health care programs, Michael F. Drummond et al. (1987) note that:

> ...if hypertension therapy does extend the lives of people, there is nothing to say that they should have to be given cancer therapy at a later date. This is a decision that should be made on its own merits.... However, in the calculation of the life extension from instituting the hypertension screening programme, if it has been assumed that those developing cancer will have the benefits of life extension from the therapies available, then consistency would demand that the costs of cancer therapy be also included. (p. 80)

In a book about disease prevention, Louise L. Russell (1986) reached a wholly different conclusion, arguing that:

> ...if the purpose of the analysis is instead to determine whether the program is a good investment, only the costs of the preventive program should be counted. Added years of life involve added expenditures for food, clothes, and housing as well as medical care. None ...is relevant to deciding whether the program is a good investment... (pp. 35-6).

The handling of unrelated "future medical costs" is important because they can be large enough to raise the cost-effectiveness ratio substantially. The impact is greatest when the intervention primarily extends life, such as for vaccines against potentially fatal contagious diseases. Several studies have highlighted the sensitivity of cost-effectiveness estimates to the inclusion of future medical costs.[1]

---

[1] In a study of influenza vaccine (Office of Technology Assessment, 1981), the cost per healthy life year (QALY) depended on the age of the person, ranging from $258 (for children 1 - 3 years of age) to $23 (for persons 45 - 64 years of age), when future medical costs were omitted. By contrast, when future medical costs of extended life years were included, the

The Question of Age Bias. A wholly separate concern arises from the usual (but not universal) practice of describing the "benefit" of the intervention in terms of life years saved, increases in life expectancy, or quality-adjusted versions thereof. Do these methods of analysis intrinsically bias the results against older persons, for whom the potential increase in life expectancy from any intervention is of necessity limited? Jerry Avorn (1984) has asserted:

> ...[Cost-benefit and cost-effectiveness analysis] can have major shortcomings when applied to the care of several high-risk populations, particularly the elderly...As usually applied, these methods embody a set of hidden value assumptions that virtually guarantee an anti-geriatric bias to their purportedly objective data. (p. 1295)

How, if at all, are these formal methods biased, and against what criterion should bias be measured?

How Should One Select the "Optimal" Cost-Effectiveness Ratio? A final problem emerges when one considers the common uses of CE analysis for decision making. Typically, practitioners of CE analysis calculate the incremental costs and incremental effectiveness (e.g., in QALYs) of an intervention, then they compare that ratio to those found for commonly used interventions. Table 1 provides estimates from studies of various interventions, updated to 1991 dollars through use of the medical CPI from the calculations in the original articles.

[Table 1 here -- CE Ratios from various interventions]

---

relevant costs per healthy life year increased by $1745 (for children 1 - 3) to $2084 (for adults age 45 - 64). Thus, the decision to include or exclude future medical costs changed the incremental cost effectiveness ratio by two orders of magnitude. A later analysis by Michael A. Riddiough et al. (1983) reached similar conclusions.

As should be clear from examination of this table, inferring the CE ratios of "common practices" provides little guidance regarding the *optimal* CE ratio--i.e., the willingness to pay for a health effect. Most practitioners of CE analysis discard interventions with CE values at the top range of a table such as this one, and conclude that interventions in the realm of $50,000 (or so) per QALY are "OK" but that more expensive technologies become more and more "out of bounds."[2]

Plan for Analysis. To address these problems, we first set up a simple model of expected utility maximization (Section II), from which we seek to answer the first question ("which costs to include"). We then generalize this model to explore a number of other issues associated with CE analysis, including whether or not the approach is internally consistent (Section III), and the nature of an optimal lifetime medical spending plan and the estimation of an optimal CE ratio (Section IV).

II. A Simple Model of Cost-effectiveness.

We begin with a simple three-period model in which the individual has income Y_i and medical care expenditures C_i in period i. Utility in each period is a function of income net of medical care expenditures. All individuals are alive in period 1 and survive to period 2 with probability P_2. Given survival of period 2, they survive into period 3 with probability P_3. In this model, C_1 affects P_2, but not P_3, and C_2 affects only P_3. Medical care affects utility only by altering the survival probabilities. Thus, the individual's Von Neumann-Morgenstern expected utility is:

---

[2] This leaves unresolved, of course, why interventions with relatively low marginal CE ratios are not expanded in scope at the expense of more-costly interventions, a shift of medical resources that would surely increase overall health absolutely (Phelps and Mushlin, 1991).

$$E(U) = U_1(Y_1 - C_1) + P_2(C_1)U_2(Y_2 - C_2) + P_2(C_1)P_3(C_2)U_3(Y_3) \quad . \qquad (1)$$

The only relevant choice variable is C_1, since C_2 is independent of C_1.[3] Following a similar analysis, we can solve for an optimal investment in C_2, which we denote as C_2^*, and corresponding outcome P_3^*. "Effectiveness" in this simple model is the increase in P_2 that results from investment in health care during period 1 (C_1), which in turn increases expected utility. Maximizing expected utility with respect to C_1 leads to an equation involving dP_2/dC_1, which (when inverted) provides the optimal incremental CE ratio. This ratio includes only current costs (not C_2 in this simple model), and comes directly from maximization of expected utility. Define $U_i' = dU_i/dY_i$. Then

$$\frac{dC_1}{dP_2} = \frac{U_2(Y_2 - C_2^*) + P_3^* U_3(Y_3)}{U_1'} \quad . \qquad (2)$$

Specifically, (2) says that an optimum is reached when the CE ratio equals the sum of future expected utility normalized by the marginal utility of income in period 1. This result generalizes to more periods, and allows discounting of future consumption, but this insight remains throughout such generalizations.

In a sense, this answers our first question. When one approaches the problem of defining an optimal CE ratio for medical resource decisions by using expected utility maximization principles, one can derive a CE "cutoff" for decision making that does not include future costs (C_2).

---

[3] Formally, $dC_2/dC_1 = 0$. This reflects the basic principle of dynamic programming that future decisions do not depend on past decisions, given the current state. This would not be true if either P_3, Y_2, or the function U_2 were a function of C_1; however, as we shall see subsequently, the substantive conclusions do not change when this assumption is relaxed.

How (if at all) Should One Include Future Costs?

What are the consequences of including unrelated future costs in the C/E analysis? Define expected total lifetime costs as $C^{tot} = C_1 + P_2(C_1) \bullet C_2$. Consider a CE ratio, dC^{tot}/dP_2, that includes future unrelated costs C_2 as well as the current costs included in dC_1/dP_2. From the definition of C^{tot}, we know immediately that $dC^{tot}/dP_2 = dC_1/dP_2 + C_2$. The optimization problem tells us that, if we wish to optimize using total costs, the optimal cutoff is the same as that in the previous problem plus C_2. One must only be consistent in practice: use the CE cutoff for decision making that corresponds to the cost accounting method one has chosen.

Are there reasons to prefer one approach over the other? The most important consideration is consistency, so that comparisons of the cost-effectiveness ratios of alternative interventions are meaningful. Insofar as it is difficult to measure unrelated future health expenditures, there is an advantage to omitting them from the analysis. However, if it is not possible to measure these costs, the assumption that they are truly unrelated to the intervention -- i.e., that C_2 is independent of C_1 -- cannot be tested. Thus, ordinarily there is no compelling reason to select one method over the other, but when it is known that an intervention has no direct impact on future costs, parsimony dictates excluding them.

Decision Making with Life Expectancy as the "Effectiveness" Measure.

The effectiveness measure of the preceding discussion is the probability of surviving a single period. We now show that the results also hold when we measure cost-effectiveness in more conventional terms, namely in terms of the cost per life year (or cost per year of life expectancy). For notational convenience, we suppress the dependence of the probability terms on prior health expenditures, and observe that in this simple model, life expectancy is given by:

$$LE = 1 + P_2 + P_2 P_3 \ , \tag{3}$$

so

$$\frac{dLE}{dP_2} = 1 + P_3 \ , \tag{4}$$

and

$$\frac{dC_1}{dLE} = \frac{U_2(Y_2 - C_2^*) + P_3^* U_3(Y_3)}{(1 + P_3^*) U_1'} \ . \tag{5}$$

The independence of future spending decisions (conditional on survival) from past spending decisions implies that here, too, the results will be equivalent. Since P_3^* is selected optimally by altering C_2 but independently from C_1, (5) only differs from (2) by a multiplicative constant. We generalize this result further below, but this very simple model of medical "effectiveness" provides the basic insight for much of what follows.

The above discussion counts gains in E(U) only from improvements in life expectancy, i.e., through the effects of medical interventions on survival probabilities P_i. Our more general framework allows changes in utility from quality of life improvements as well. A broad literature describes methods developed to measure quality of life and to assess the

value of quality improvements in terms of the increases in life expectancy that would provide equivalent increases in utility (Torrance, 1986 and references therein) in a strict Von Neumann - Morgenstern framework. Indeed, these approaches provide the basic framework for computing the quality adjustments in QALY measures.

III. Does Cost Effectiveness Provide an Internally Consistent Way to Maximize E(U)?

Common practice in CE analysis says that in order to maximize expected utility, one should adjust the intensity of all medical interventions so that they have a common CE ratio.[4] The intuition of the dictum derives from the idea that one should seek to equate the marginal benefit and marginal cost of all inputs in a productive process, as in other contexts. In this section we elaborate on our previous simple model of expected utility, incorporating more than one medical intervention, allowing those interventions to have differing effects on all future period survival probabilities, and introducing discounting and quality of life considerations. Using this model, we show that cost-effectiveness analysis provides a consistent criterion for selecting health interventions: the optimal CE cutoff is the same for all interventions, regardless of when they exert their effects. We thereby provide rigorous support for the common practice. We model this problem in discrete time using two interventions (**a** and **b**) available at constant cost w_a and w_b respectively, each with the ability to alter future quality of life and survival.

We first give precise definition to the three measures of effectiveness most commonly used in CE analysis. The most general measure is

---

[4] We address later the question of how one might select that ratio.

QALYs. If P_j is the probability that a person alive the preceding period will be alive during period j, then the cumulative probability that a person is alive (the survivor function) at period i is

$$F_i = \prod_{j=1}^{i} P_j \ .$$

The expected number of QALYs can be written as

$$QALY = \sum_{i=1}^{N} F_i \delta^i k_i \ ,$$

where N is the maximum life span, $\delta = 1/(1+r)$ is a time discount factor, and the k_i terms represent quality adjustments. The value of k_i can range from 0 (for the worst state of health, usually assumed to be death or its equivalent) to 1 (corresponding to "perfect" health). Each such term is the expected value of quality adjustments for all possible states of health in period i; discounting considerations aside, two years of life in which $k_i = .5$ contribute the same number of QALYs as one year in which $k_i = 1$. The other two commonly used measures of effectiveness are special cases of QALYs. The simplest measure, life expectancy, sets $k_i = \delta = 1$ for all i; discounted life expectancy sets $k_i = 1$ for all i, but $\delta < 1$.

We now turn to the framework of utility maximization and relate it to the definition of QALYs. We posit von Neumann-Morgenstern utility maximization, and assume that lifetime expected utility as viewed from time 0, which we denote by E_0U, takes the form:

$$E_0 U = U_0(Y_0 - w_a a - w_b b) + \sum_{i=1}^{N} U_i(Y_i) F_i \ . \qquad (6)$$

Period-specific utility, U_i, takes the form $U_i = \upsilon\delta_i k^i$, where $\upsilon = U_0(Y)$ and Y is a constant (in real terms) across time periods, and k_i is a period-

specific multiplier interpreted as a quality adjustment above. In this form, the utility function and its argument, income, are constant over time, but period-specific utility can change by the multiplicative terms k_i and can be discounted. This assumption implies that expected utility can be rewritten as:

$$E_0 U = U_0(Y - w_a a - w_b b) + v \sum_{i=1}^{N} \left[\delta^i k_i \prod_{j=1}^{i} P_j \right] . \tag{7}$$

Thus the summation above is the number of QALYs remaining as of period 1. Define $dU_o(Y - w_a\mathbf{a} - w_b\mathbf{b})/dY = U_0'$. The dependence of U_0' on **a** and **b** will be suppressed notationally from here forward, but it is important to remember this relationship.

The two available medical interventions, **a** and **b**, can affect both the survival probabilities (the P_i terms) and the expected quality adjustments (the k_i terms) in future periods. Define $\partial P_i/\partial \mathbf{a} = \epsilon_i^a$, $\partial P_i/\partial \mathbf{b} = \epsilon_i^b$, $\partial k_i/\partial \mathbf{a} = \psi_i^a$, $\partial k_i/\partial \mathbf{b} = \psi_i^b$, and $\mathbf{V}_i = k_i \mathbf{F}_i$. Now optimize with respect to **a** and **b** in the usual fashion. Differentiation with respect to **a** yields the following condition:

$$\frac{\partial E_0 U}{\partial a} = -w_a U_0' + v \sum_{i=1}^{N} \delta^i \frac{\partial V_i}{\partial a} . \tag{8}$$

The change in expected utility consists of an expenditure-induced loss of period 0 utility and a gain in future expected utility, which can be due to changes in the survival distribution as well as changes in the quality adjustments $\mathbf{k}_i$. The derivative of each $\mathbf{V}_i$ term has the form which decomposes the change in period i's expected utility into a change in the quality factor expected during period i, weighted by the probability of

$$\frac{\partial V_i}{\partial a} = \psi_i^a F_i + k_i \frac{\partial F_i}{\partial a} , \tag{9}$$

being alive then, and the change in the survival probability, weighted by the expected quality. Rewriting equation 9, and substituting the definition of the survival probability, we have

$$\frac{\partial E_0 U}{\partial a} = -w_a U_0' + v \left\{ \sum_{i=1}^{N} \delta^i \prod_{j=1}^{i} P_j \left[\psi_i^a + k_i \sum_{k=1}^{i} \frac{\epsilon_k^a}{P_k} \right] \right\} \tag{10}$$

which we set to zero for optimality[5]. An equivalent expression arises for intervention **b**, replacing ϵ_i^a with ϵ_i^b and ψ_i^a with ψ_i^b.

In this form, utility is a function of (discounted) quality-adjusted life years, and the term in braces represents the incremental effect of **a** on QALYs, which we denote as $\partial Q/\partial \mathbf{a}$. This term plays a central role in the analysis that follows.

If utility is the same in every period (except for discounting), then (and only then) does the problem in expected utility maximization become equivalent to a problem in discounted life expectancy, since each period's utility is assumed to be proportional to the first period's. In standard models of lifetime consumption planning, optimization implies equating the <u>marginal</u> utility of income in each period, rather than <u>total</u> (i.e., the absolute level of) utility (Jack Hirshleifer, 1966; Isaac Ehrlich and Gary S.

---

[5] We assumed that there was no immediate effect of the intervention on quality of life (hence there is no k_0 term). If there was such an immediate effect, the marginal utility of expenditures on **a** in period 0 would consist of two terms, the negative one from the loss of income, and a positive term from the increase in k_0. This generalization does not affect any of the substantive conclusions that we draw from the analysis.

Becker, 1972). Only if $k_i = 1 \; \forall \; i$ would optimal income transfers equalize both marginal utility and income. Allowing for quality adjustment greatly relieves this restriction, since the quality adjustment is designed to account for differences in the level of utility across states of health and across ages. Differences in quality of life could arise from shifts in health, changes in the utility function with age, or changes in the values of other arguments of utility functions, such as exogenously determined consumption of complements or substitutes for consumer goods and services.

The above equations yield the simple result that optimal investment in **a** is defined by:

$$w_a = \frac{v}{U_0'} \frac{\partial Q}{\partial a} \; . \tag{11}$$

With this utility structure, the marginal benefit of medical care is simply the scaled utility (v/U_0') of the incremental QALYs derived from incremental **a**, and at the optimum, incremental benefit equals incremental cost (w_a). A comparable result holds for intervention **b**:

$$w_b = \frac{v}{U_0'} \frac{\partial Q}{\partial b} \; . \tag{12}$$

With these tools, we can return to the problem we visited above, namely to consider whether one can define a utility-maximizing program using CE ratios. This time, we have two interventions, rather than one,

$$\left[\frac{dC}{dQ}\right]_a = \frac{\dfrac{dC}{da}}{\dfrac{dQ}{da}} = \frac{\dfrac{\partial C}{\partial a} + z w_b}{\dfrac{\partial Q}{\partial a} + z \dfrac{\partial Q}{\partial b}} \; . \tag{13}$$

and current medical cost is $C = w_a a + w_b b$. If the CE method is internally consistent, the optimal CE cutoff for interventions **a** and **b** must be the same, even if they exert their health effects at different times. If the CE method is not consistent, it cannot be used to allocate resources efficiently. We also need to allow for substitution in production of health between **a** and **b**; define the marginal rate of substitution as $z = (db/da)$. By definition, $dC/da = w_a + zw_b$. Now, define the cost-effectiveness ratio for intervention **a** as:

Substituting the optimal values for $\partial Q/\partial a$ and $\partial Q/\partial b$ from (11) and (12) leads to an extremely simple but important result. At the optimum investment in intervention **a**,

$$\left[\frac{dC}{dQ}\right]_a = \frac{w_a + z w_b}{\left(w_a + z w_b\right)\left[\frac{U_0'}{\upsilon}\right]} = \frac{\upsilon}{U_0'} \quad . \tag{14}$$

At the optimum, the ratio of incremental costs to incremental QALYs from further investment in intervention **a** equals υ scaled by U_0'. Thus the optimal CE cutoff is the ratio of future period-specific utility υ to marginal utility in the base period.

Note that the optimal CE cutoff depends on total medical spending in the initial period. Recall that U_0' depends on income net of medical spending, i.e., $Y_0 - w_a a - w_b b$. As current health expenditures increase, the U_0' term in the denominator rises, making the optimal CE cutoff smaller, and hence a more stringent test for a medical intervention. We explore this phenomenon in Section IV.

An exactly parallel development shows that for intervention **b**, the same condition holds. Tracing through similar steps, we find that optimal investment implies

$$\left[\frac{dC}{dQ}\right]_b = \frac{\frac{w_a}{z} + w_b}{\left[\frac{w_a}{z} + w_b\right]\left[\frac{U_0'}{v}\right]} = \frac{v}{U_0'} \quad . \tag{15}$$

This proves the internal consistency of CE analysis, since the optimal CE cutoff, dC/dQ, is the same for both interventions. This result obviously generalizes to multiple interventions that exert their effects at different times. In the two-intervention model, intervention a might be a treatment for heart attacks, which has an immediate effect only, while b is a preventive intervention that has no immediate effects but diminishes mortality rates in the future.

Do the same results hold if unrelated future costs are included in the definition of the costs for the cost-effectiveness ratio? To answer, we need to define the present value of expected *total* costs of care, which are

$$C^{tot} = w_a a + w_b b + P_1 \delta c_1 + P_1 P_2 \delta^2 c_2 + \ldots \quad , \tag{16}$$

where c_i = total health expenditures in period i. The change in costs due to an intervention include both direct expenditures for the intervention, the change in expenditures for the other interventions, and the expenditures that result from living longer:

$$\frac{dC^{tot}}{da} = w_a + w_b \frac{db}{da} + \frac{1}{P_1}\left[\frac{\partial P_1}{\partial a} + \frac{\partial P_1}{\partial b}\frac{db}{da}\right]\left[\delta P_1 c_1 + \delta^2 P_1 P_2 c_2 + \ldots\right] + \frac{1}{P_2}\left[\frac{\partial P_2}{\partial a} + \frac{\partial P_2}{\partial b}\cdot\frac{db}{da}\right]\left[P_1 P_2 \delta^2 c_2 + \ldots\right] + \tag{17}$$

The above expression can also be written as

$$\frac{dC^{tot}}{da} = w_a + w_b \frac{db}{da} + \frac{\partial E}{\partial a} + z\frac{\partial E}{\partial b} = \frac{dC}{da} + \frac{\partial E}{\partial a} + z\frac{\partial E}{\partial b}, \tag{18}$$

where E = the present value of expected health expenditures.

When combined with the logic we used to demonstrate consistency when future costs are excluded, these results imply that

$$\left[\frac{dC^{tot}}{dQ}\right]_b = \left[\frac{dC}{dQ}\right]_b + \frac{\frac{1}{z}\left[\frac{\partial E}{\partial a}\right] + \frac{\partial E}{\partial b}}{\frac{1}{z}\left[\frac{\partial Q}{\partial a}\right] + \frac{\partial Q}{\partial b}}. \tag{19}$$

and, by similar reasoning,

$$\left[\frac{dC^{tot}}{dQ}\right]_a = \left[\frac{dC}{dQ}\right]_a + \frac{\frac{\partial E}{\partial a} + z\frac{\partial E}{\partial b}}{\frac{\partial Q}{\partial a} + z\frac{\partial Q}{\partial b}}. \tag{20}$$

The preceding analysis showed that the first terms on the right-hand-sides of these two equations are equal at the optimum. By multiplying the numerator and denominator of the second terms in (20) by z, it is seen that the second terms are also equal. Thus, at the optimum,

$$\left[\frac{dC^{tot}}{dQ}\right]_a^* = \left[\frac{dC^{tot}}{dQ}\right]_b^*, \tag{21}$$

implying that the CE ratio is also consistent when unrelated future costs are included.

Thus, if utility can be expressed in terms of QALYs, our model can be used to show that the optimal CE cutoff is the same for all medical interventions, so that CE methods are internally consistent. The result does not depend on our choice of including or excluding unrelated future costs.

In addition, we have shown that this optimal cutoff depends strictly on parameters of the utility function, specifically, v/U_0'. This result allows, at least in concept, inferences about the optimal cutoff for CE analysis that flow directly from the preference structure of consumers, rather than relying on the often-distorted and confusing inferences that one can draw from calculating CE ratios for observed medical practices (see e.g., Table 1). We examine this issue in greater detail below.

IV. Optimal Lifetime Medical Spending Program

This model also provides a framework for defining an optimal lifetime spending program for medical care and for exploring its relationship to the cost-effectiveness criterion. One health intervention can differ from another in many respects, including its marginal productivity in producing health (i.e., the size of the effects on survival probabilities and quality of life), the time course of its impact ("treatment" expenditures are ordinarily those for existing, symptomatic illnesses, and tend to have an immediate effect, while "preventive" care usually has the aim of preventing future disease), and its costs. Optimal expenditures on health care can also vary because of person-specific characteristics--factors that cause them either to have different optimal CE cutoffs or to have the same cutoffs, but different utility-maximizing expenditures. In this section, we explore the causes of variation in optimal CE cutoffs and in expenditures, emphasizing the personal factors responsible for variation in optimal expenditures.

We first note the implications of equations (14) and (15) for variation in the optimal CE cutoff. These equations say that the CE cutoff is just the ratio of the fixed component of future period-specific utility (v) to marginal utility in the "initial" period. A number of factors might cause this ratio to vary among individuals, such as variation in risk aversion or other characteristics of the utility function. Even if the utility function does not

differ among persons, the values of its arguments, such as income, may. The ratio of the level of utility to the marginal utility rises with income (or wealth). Furthermore, changes in health status that increase the utility of expenditures for goods and services designed to mitigate the effects of illness, such as arthritis-induced expenditures for mineral baths and pain relievers, tend to diminish the level of utility and to raise the marginal utility of income.

Even if different individuals have the same CE cutoff, there will be several reasons for them to have different optimal expenditures. For example, since advancing age is associated with a decrease in the number of potential years of life left (i.e., a decrease in annual survival probabilities), a life-saving intervention might not be capable of increasing life expectancy or QALYs by as great an amount at advanced ages as in youth. It also seems obvious that individuals with a high rate of time preference (low value of δ) would spend less on preventive care than those with a low rate of time preference.[6] Although these findings are true in general, we now explore them in detail by analyzing specific examples. We begin by specifying an intertemporal production function for health and a specific utility function. For expositional simplicity, we assume that medical spending only alters future probabilities of death, but these ideas readily generalize to the improvement of quality of life (Joseph S. Pliskin et al, 1980, John M. Miyamako and Stephen A. Eraker, 1985).

---

[6] Insofar as we use the correct value of δ for each person, people with the different values of δ can have the same CE cutoff but it will correspond to different amounts of care. If we use a single value of δ for a heterogenous population, the CE criterion may not lead to optimal expenditures, since individual utilities will not be functions of the population-level value of δ. The same is true of other parameters of the utility function.

Define $\mu_i(\mathbf{a}) = 1\text{-}P_i(\mathbf{a})$ as the age-specific probability of death in period i as a function of the level of **a** used in period 0. Let $C_0 = w_a\mathbf{a}_0 + w_b\mathbf{b}_0$ represent spending in period 0. Then production of health can be characterized as the relative mortality reduction that results from the expenditure on **a**:

$$\frac{\mu_i(a)}{\mu_i(0)} = \left[1 - \alpha\rho^i(1 - e^{-\phi a})\right] . \tag{22}$$

In this equation, $\mu_i(0)$ is the mortality rate when $\mathbf{a}=0$, α reflects the largest reduction in the risk of dying that an expenditure can provide, and ρ (ordinarily $0 \leq \rho \leq 1$) represents the persistence of the treatment's effect over time. The parameter ϕ scales the impact of **a** on the relative mortality rate. These relations imply that the marginal productivity of **a** in increasing the age-specific probability of survival is

$$\frac{dP_i(a)}{da} = (1-P_i(0))\alpha\rho^i\phi e^{-(\phi+1)a} . \tag{23}$$

If we think of disease-specific expenditures for preventive care applied to the general population, then α will be very small, because even elimination of a single disease cannot increase the probability of survival very much. For the prevention of a single disease, α cannot exceed the

one-period probability of dying of that disease, and usually will be far less.[7] Thus, for example, the widely publicized effort to get Americans to reduce fat consumption will have little effect on mortality. Under fairly optimistic assumptions, which include a reduction in mortality from certain forms of cancer as well as from heart disease, if Americans reduced fat consumption to 30% of calories, life expectancy for a 50 year-old man would increase by about 4 days. A 50 year-old woman would only live about two days longer (Warren S. Browner et al., 1991). Only for patients with life-threatening diseases is the potential improvement in life expectancy very large. Hence, with the exception of effective treatments applied to people who are already ill, we expect α to be small.

The parameterization in Equations 22 and 23 implies that the intervention's effectiveness will decay exponentially over time, at rate $1-\rho$. Unless they are used on an ongoing basis, the effectiveness of many preventive interventions for the control of such risk factors as hypertension and hypercholesterolemia declines with time. The protective effects of vaccines also diminish with time. They prevent infectious diseases by

---

[7] Overall mortality rates represent an unattainable upper bound on the value of α at any age. For a medical intervention's value of α to reach these levels, it would have to eliminate all causes of death during a given year.

Age Interval	All-Cause 5-year Mortality Rates
30-35	.007
35-40	.009
40-45	.012
45-50	.019
55-60	.047
60-65	.074

These figures are for both sexes and all races, U.S. Life Tables for 1987 (National Center for Health Statistics, 1990, p. 6).

stimulating the production of specific antibodies, whose levels gradually decline after the initial response to the vaccine. Effectiveness falls as the antibody levels drop; in this context, ρ might represent the proportion of the antibodies, or the rate of effectiveness, persisting from one year to the next.

In order to assess the implications of this model, we specify a separable utility function whose utility is convex in Y_i. For the period-specific utility, a convenient and commonly used functional form specifies $U = \beta(1 - e^{-\gamma Y})$, with corresponding $U' = \beta e^{-\gamma Y}$, $U'' = -\beta\gamma e^{-\gamma Y}$, absolute risk aversion $r = -U''/U' = \gamma$, and relative risk aversion $r^* = \gamma Y$ (see John W. Pratt, 1964, and Kenneth J. Arrow, 1974). The expressions for **a** and **b** contain the ratio U/U' as a central component. With this utility function, we can specify the ratio U/U' if we know the relative risk aversion measure $r^* = \gamma Y$. This function serves as the period-specific component expected utility (7). This functional form allows us to assess the impact of variation in risk aversion and in other parameters of the utility function on the optimal CE ratio.

To analyze the dependence of the CE ratio on the value of γ, the risk aversion parameter, we first recall that U_0' varies with total medical spending, since we evaluate it at $Y - w_a\mathbf{a} - w_b\mathbf{b}$. Thus, the optimal CE cutoff will depend both on the utility function and the degree of medical spending. (Of course, medical spending also depends on these same preferences.) Combining the utility function (7) and the production function (22), and maximizing with respect to **a**, gives the optimal spending on medical care and (from that) the optimal CE ratio.

Because optimal **a** cannot be readily determined analytically (equation 10, which gives the derivative of expected utility with respect to **a**, does not have a simple closed-form solution, because the equation contains an arbitrary number of survival probability terms that are each functions of **a**),

we used iterative techniques to solve for the optimal values of medical spending and the CE cutoff. Solutions were computed for a wide range of parameters for income, risk aversion levels (r*), discount rates (δ), maximal reduction in mortality rates (α), persistence of the medical intervention's effects (ρ), and gender. These simulations use actual mortality tables for U.S. citizens, specific to gender (but not race). For the base case, we selected the median annual per capita income in the United States ($18,000 in 1989), a maximal reduction in mortality of $\alpha = .3$ (as appropriate for an effective treatment of a quite dangerous disease), a persistence parameter of $\rho = .6$, and a discount rate of 5% ($\delta = .95$). The resulting optimal spending rates and CE cutoffs are shown in Table 2 for females; the patterns are quite similar for males, but the optimal spending is slightly higher at each age interval because of the higher age-specific risk of death for males. CE ratios are very similar for both genders, as we will show further below.

[Table 2 here -- base case results]

These results convey two major features of our model's behavior. First, optimal spending rises rapidly with age (as does the actual pattern of spending in the US and elsewhere), a consequence of the increase in mortality that accompanies aging; as illness risks increase, the demand for medical intervention rises. Second, and more subtle, the optimal CE ratio falls with increasing expenditure (and hence with age), since the foregone utility from not spending income on other goods increases as medical spending increases.

The corner solution at younger ages--zero medical spending--does not result from a low CE ratio (the young have higher CE cutoffs than the old), but rather reflects the infrequency of death at younger ages. The risk of dying is so small in the youngest age groups that it cannot be reduced much more by expenditures on health care, so spending the entire budget on other

goods and services provides the greatest utility. The other feature driving these results is that, as a person ages, annual mortality rates rise, so that any treatment affecting future mortality risks is "amortized" over a shorter and shorter period. These results enter our model through use of actual life tables.

A variety of sensitivity analyses (see Table 3) show that the optimal spending pattern behaves much as one might expect, and the optimal CE cutoff is remarkably stable over a wide range of production function parameters (α and ρ). Increasing either α or ρ increases the marginal productivity of medical spending and optimal spending on **a**. This spending increase reduces the income available for other goods and services, thus making the optimal CE cutoff slightly more stringent (CE falls). This same patterns occur at all age intervals, although optimal spending remains zero for younger persons over a broader range of the production parameters.[8]

[Table 3 here -- sensitivity to production parameters]

Higher values of α are possible for people who have potentially fatal illnesses (such as cancer) for which the treatment is reasonably effective. Our inclusion of an upper limit of 0.3 for α corresponds to such a case.

Varying the discount rate has effects on optimal spending and the CE cutoff as one might anticipate, although the effect interacts with age much more than in the case of the production function parameters. We vary the discount rate from 0 to .1 (δ = 1 to .91), reflecting values found in the literature (see, e.g., W. Kip Viscusi and Michael J. Moore, 1989, and Maureen L. Cropper et al., 1992, but see Victor R. Fuchs, 1982 for a larger estimate). Table 4 shows the results for males (the pattern is very similar for females, but optimal spending is zero for a greater range of age

---

[8] More detailed results are available from the authors upon request.

and values of δ and lower in general, given the lower mortality risk for females at any age). While the optimal spending depends importantly on δ, the optimal CE cutoff remains stable over values we have tested.

[Table 4 here -- sensitivity to discount rate]

Variability in rates of time preference may pose a special problem for most CE analyses, which are usually based on the assumption that the appropriate rate of time discount for the health benefits of an intervention is the same as the market rate of interest (Emmett B. Keeler and Shan Cretin, 1983). If rates of time preference are the same for all people, and if capital/savings markets are perfect, the market rate of interest equals the rate of time preference (approximately .02 to .03, see Robert J. Barro, 1987). But some estimates of rates of time preference are much higher than the usual values assumed for the real (or even nominal) rate of interest. Of course, there are multiple rates of interest to which one could refer, but the variability among them is evidence of capital market imperfections, perhaps explaining why market interest rates could fall short of average rates of time preference. Usual arguments about why the same discount rate should be used for health effects as for costs have less force, under the circumstances. Fortunately for social planning purposes, the optimal CE ratio does not vary importantly with the discount rate, mitigating concern about this issue.

These results confirm our earlier assertion: ordinarily a diminished planning horizon implies that preventive spending should decline with age. The same finding holds true if the effect of aging is captured instead in a greater mortality rate; for a given change in the survival probabilities, Equation 10 implies that the marginal utility of expenditures on **a** is negative if the levels of the survival probabilities (F_i terms) are small enough. In other words, if a person is unlikely to survive the current

period is small enough, he or she would rather increase current utility, spending money on current consumption, than modify a small probability of survival.

Why, then, do expenditures typically *rise* with age? Primarily because the benefit of treatment must be small when there is little disease to treat. In terms of our model, α of necessity must approach 0 as P approaches unity. Similar results hold for the quality adjustments; during ages when health is excellent and morbidity is minimal, medical expenditures are not likely to increase k_i significantly. Thus the CE criterion tends to promote large treatment expenditures (i.e., in which ρ may be small but α is large) at older ages and in persons who have diseases.

An important consideration is the role of the quality adjustments k_i. Thus far we have assumed that the CE analysis properly incorporates quality of life measures. Many CE analyses do not, implicitly assuming that $k_i = 1$ for all i. By omitting quality of life, they miss the effects of the intervention on future quality of life (i.e., implicitly assume $\psi_i^a = 0$), and they fail to discount properly years of life in which the expected level of utility is relatively low. Omitting the quality impact of treatment means that particular treatments will be undervalued, such as many forms of rehabilitative care and long-term care that are used most commonly in old age. On the other hand, failure to recognize that years of life extended at older ages are often characterized by worsened health status tends to bias expenditures in favor of the elderly. If one accepts the notion that quality of life falls as physical and mental disability increase (see, e.g., George W. Torrance 1987), then the usual pattern of declining physical function that accompanies aging implies that the pattern of multipliers k_i becomes smaller as a person grows older. If so, simplifying cost-effectiveness analyses by

assuming that $k_i = 1$ for all i, i.e., assuming that utility is a function of discounted life expectancy alone, will result in *overstating* the optimal spending for persons in their later years of life.

In all of the previous sensitivity analyses we have discussed, the optimal CE ratio remains fairly constant over a wide range of values of the parameters of the utility function and over a range of personal characteristics, although optimal spending varies considerably with age, the marginal productivity of medical care, and the discount rate. However, the optimal CE ratio is sensitive to two characteristics that vary among individuals -- income and risk aversion. These findings have important consequences for private and public allocation of medical care resources and for social planning of medical investments.

Figures 1a and 1b show how the optimal CE cutoff varies by income and the degree of risk aversion. We show the results for two income levels -- $18,000 and $29,000, corresponding to median per capita income and per-family income in 1989. We characterize risk aversion in terms of $r^* = \gamma Y$ in the specific utility function we employ. When researchers have estimated the degree of risk aversion using various methods, the estimates center on a relative risk aversion of about 2.0 (see Warren E. Weber, 1970 and 1975; Irwin E. Friend and Marshall E. Blume, 1975; Blume and Friend, 1975; Henry S. Farber, 1978; Frederick W. Siegel and James P. Hoban, 1982; Lars P. Hansen and Kenneth J. Singleton, 1983, Robert H. Litzenberger and Ehud I. Ronn, 1986; George A. Szpiro, 1986; Robert E. Hall, 1988; and Ricardo J. Caballero, 1991), with a range of about 1 to 4 (hence our choice of these parameters).[9] These figures also show the interaction of the effects of age and risk aversion: at higher degrees of risk

---

[9] Our thanks to Mark Machina for guiding us to these references.

aversion, the optimal CE ratio shifts more with age, and as people become less risk averse, age has a diminishing (and finally nearly zero) effect on the optimal CE ratio. Figures 1a and 1b also show the results previously mentioned, namely that the effects of gender matter only a very little in determining the optimal CE cutoff, entering this model solely through the effects of differential risks of mortality on the optimal spending program, and hence on the optimal CE cutoff.

[Figures 1a and 1b here -- income and r* effects]

V. Discussion.

Cost-effectiveness analysis has long been recognized as a convenient approach to guiding health care decisions. Its validity, however, has never been established with any rigor. We have shown that, within the framework of standard Von Neumann-Morgenstern utility maximization, cost-effectiveness analysis offers a valid criterion for choosing among health interventions. Surprisingly, the inclusion of unrelated future costs is without consequence so long as the practice is consistent. Although our analysis is based on a specific family of utility functions, the use of quality adjustments allows it to approximate a wide range of functional forms. The frequent use of life expectancy as the chief outcome variable in CE analysis is considerably more restrictive. With the quality adjustments, CE analysis can be a powerful and appropriate guide for decision making.

Insofar as the observed cost-effectiveness ratio of various medical interventions in common use varies by at least an order of magnitude, the usual practice of comparing the cost-effectiveness of a particular intervention with that for others offers little guidance for planning or resource allocation. Although it is clear that cost-effectiveness ratios should be equalized across interventions at the margin, the specific cutoff that should be used is ordinarily described as unknown. We have shown how

the optimal cutoff can be derived from the parameters of a flexible utility function, estimates of which are available from several sources. These estimates suggest CE cutoffs centering at around $35,000 for persons with an income of $18,000. The optimal CE cutoff, as discussed, varies greatly with income; over the range that we estimated, it is roughly double the annual income.

When effectiveness is measured in terms of life expectancy, the optimal CE ratio represents the same concept as the "willingness to pay" -- the amount an individual would pay to reduce a risk of death. Empirical work has shown that the "willingness to accept" -- the amount of money that individuals would require to voluntarily accept a risk of death from job causes -- is very high, on the order of $300,000 per year of life expectancy in jeopardy (Viscusi and Moore, 1989). The willingness to pay for a reduction in the risk of death may be quite different, and would ordinarily be substantially lower (W. Michael Hanemann, 1991).[10] Labor market data, therefore, provide an upper bound on the optimal cost-effectiveness ratio, although the actual appropriate cutoff for a given individual's utility function may be substantially lower.

As long as rates of time preference, attitudes toward risk, initial health endowment, and other aspects of utility vary, so will the CE ratio. It is for these and other reasons that the demand for health -- and health care -- varies. Whether this variation can be incorporated into policy is

---

[10] Hanemann explored public goods, of which environmental risk is an example. The obvious reason willingness to pay and willingness to accept can differ is the income effect, but it is usually negligible for environmental risks. The income effect can be much larger in health (falling ill with a serious disease is equivalent to a large loss in endowment). Furthermore, Hanemann showed that even when the income effect is small, willingness to pay and willingness to accept can differ greatly, as long as private goods are poor substitutes for the public good.

unclear, since the usual goal of cost-effectiveness analysis is to inform a population-level policy. It is seldom used to aid in decisions about specific cases.

The optimal CE ratio, essentially a measure of the demand for quality-adjusted life years, varies across individuals much as the demand for any good or service varies. It exhibits only slight variation with age or gender, but is quite sensitive to income and attitudes towards risk (risk aversion). The variability of the optimal CE ratio across persons leads to a fundamental tension in using it to guide the allocation of health care resources: insurers and policy makers may wish to equate CE ratios across interventions and across populations, yet the members of the population have very different optimal CE ratios. If a single CE ratio is applied to all interventions and to all individuals in a group, for some of them the marginal benefit will fall much lower than the marginal cost, and for others, just the opposite. Undoubtedly, this individual variability in demand is responsible in part for the persistence of a pluralistic health care system in the U.S., which is inefficient in many other ways. CE analysis applied at the population level may give the most efficient egalitarian distribution of health resources, but it is not likely to be Pareto optimal.

REFERENCES

Arrow, Kenneth J., *Essays in the Theory of Risk Bearing*, Amsterdam: North-Holland, 1974.

Avorn, Jerry. "Benefit and Cost Analysis in Geriatric Care: Turning Age Discrimination into Health Policy," *New England Journal of Medicine*, 1984, *310*, 1294-1301.

Barro, Robert J., *Macroeconomics*, (2nd ed.), New York: Wiley & Sons, Inc., 1987.

Benninga, Simon, and Protopapadakis, Aris. "The Stock Market Premium, Production, and Relative Risk Aversion," *American Economic Review*, 1991, *81*, 591-599.

Blume, Marshall E., and Friend, Irwin E. "The Asset Structure of Individual Portfolios and Some Implications for Utility Functions," *Journal of Finance*, 1975, *30*, 585-603.

Boyle, M.H., Torrance, G.W., Sinclair, J.C., and Horwood, S.P. "Economic Evaluation of Neonatal Intensive Care of Very-Low-Birth-Weight Infants," *New England Journal of Medicine*, 1983, *308*, 1330-1337.

Browner, Warren S., Westenhouse, Janice, and Tice, Jeffrey A. "What if Americans Ate Less Fat? A Quantitative Estimate of the Effect on Mortality," *Journal of the American Medical Association*, 1991, *265*, 3285-3291.

Caballero, Ricardo J. "Earnings Uncertainty and Aggregate Wealth Accumulation," *American Economic Review*, 1991, *81*, 859-871.

Cohn, Richard A., Lewellen, Wilbur G., Lease, Ronald C., and Schlarbaum, Gary G. "Individual Investor Risk Aversion and Investment Portfolio Composition," *Journal of Finance*, 1975, *30*, 605-620.

Cropper, Maureen L., Aydede, Sema K., and Portney, Paul R. "Rates of Time Preference for Saving Lives," *American Economic Review*, 1992, *82*, 469-472.

Cummings, Steven R., Rubin, Susan M., and Oster, Gerry. "The Cost-Effectiveness of Counseling Smokers to Quit," *Journal of the American Medical Association*, 1989, *261*, 75-79.

Drummond, Michael F., Stoddart, Greg L., and Torrance, George W., *Methods for the Economic Evaluation of Health Care Programmes*, Oxford: Oxford University Press, 1987.

Eddy, David M. "Screening for Breast Cancer," *Annals of Internal Medicine*, 1989, *111*, 389-399.

Eddy, David M. "Screening for Cervical Cancer," *Annals of Internal Medicine*, 1990, *113*, 214-226.

Ehrlich, Isaac, and Becker, Gary S. "Market Insurance, Self-Insurance, and Self-Protection," *Journal of Political Economy*, 1972, *80*, 623-648.

Farber, Henry S. "Individual Preferences and Union Wage Determination," *Journal of Political Economy*, 1978, *86*, 923-942.

Friend, Irwin E., and Blume, Marshall E. "The Demand for Risky Assets," *American Economic Review*, 1975, *65*, 900-923.

Fuchs, Victor R. Time preference and health: an exploratory study. In: Fuchs VR, ed. Economic Aspects of Health. Chicago: University of Chicago Press, 1982: 93-120.

Goldman, Lee, Weinstein, Milton C., Goldman, Paula A., and Williams, Lawrence W. "Cost-Effectiveness of HMG-CoA Reductase Inhibition for Primary and Secondary Prevention of Coronary Heart Disease," *Journal of the American Medical Association*, 1991, *265*, 1145-1151.

Grossman, Sanford J., and Shiller, Robert J. "The Determinants of the Variability of Stock Market Prices," *American Economic Review*, 1981, *71*, 222-227.

Hall, Robert E. "Intertemporal Substitution," *Journal of Political Economy*, 1988, *96*, 339-357.

Hanemann, W. Michael. "Willingness to Pay and Willingness to Accept: How Much Can They Differ?," *American Economic Review*, 1991, *81*, 635-647.

Hansen, Lars Peter, and Singleton, Kenneth J. "Stochastic Consumption, Risk Aversion, and the Temporal Behavior of Asset Returns," *Journal of Political Economy*, 1983, *91*, 249-265.

Hirshleifer, Jack. "Investment Under Uncertainty: Applications of the State-Preference Approach," *Quarterly Journal of Economics*, 1966, *80*, 252-277.

Keeler, Emmett B., and Cretin, Shan. "Discounting of Life-Saving and Other Nonmonetary Effects," *Management Science*, 1983, *29*, 300-306.

Littenberg, Benjamin, Garber, Alan M., and Sox, Harold C., Jr. "Screening for Hypertension," *Annals of Internal Medicine*, 1990, *112*, 192-202.

Litzenberger, Robert H., and Ronn, Ehud I. "A Utility-Based Model of Common Stock Price Movements," *Journal of Finance*, 1986, *15*, 67-92.

Miyamako, John M., and Eraker, Stephen A. "Parameter Estimates for a QALY Utility Model," *Medical Decision Making*, 1985, *5*, 191-213.

National Center for Health Statistics, *Vital Statistics of the United States, 1987; Vol. II, Sec. 6 Life Tables.*, Washington, D.C.: Public Health Service, 1990.

Office of Technology Assessment U.S. Congress, *Cost Effectiveness of Influenza Vaccination*, Washington, D.C.: U.S. Government Printing Office, 1981.

Pauly, Mark V. "Taxation, Health Insurance, and Market Failure in the Medical Economy," *Journal of Economic Literature*, 1986, *24*, 629-675.

Phelps, Charles E., and Mushlin, Alvin I. "On the (Near) Equivalence of Cost Effectiveness and Cost Benefit Analysis," *International Journal of Technology Assessment in Health Care*, 1991, *7*, 12-21.

Pliskin, Joseph S., Shepherd, Donald S., and Weinstein, Milton C. "Utility Functions for Life Years and Health Status," *Management Science*, 1980, *28*, 206-224.

Pratt, John W. "Risk Aversion in the Small and in the Large," *Econometrica*, 1964, *32*, 122-136.

Riddiough, Michael A., Sisk, Jane E., and Bell, J.C. "Influenza Vaccination: Cost-Effectiveness and Public Policy," *Journal of the American Medical Association*, 1983, *249*, 189-95.

Russell, Louise L., *Is Prevention Better than Cure?*, Washington, D.C.: The Brookings Institution, 1986.

Siegel, Frederick W., and Hoban, James P. "Relative Risk Aversion Revisited," *Review of Economics and Statistics*, 1982, *64*, 481-487.

Sisk, Jane E., and Riegelman, Richard K. "Cost Effectiveness of Vaccination Against Pneumococcal Pneumonia: An Update," *Annals of Internal Medicine*, 1986, *104*, 79-86.

Sox, Harold C. Jr., Littenberg, Benjamin, and Garber, Alan M. "The Role of Exercise Testing in Screening for Coronary Artery Disease," *Annals of Internal Medicine*, 1989, *110*, 456-469.

Szpiro, George A. "Measuring Risk Aversion: An Alternative Approach," *Review of Economics and Statistics*, 1986, *68*, 156-159.

Thaler, Richard, and Rosen, Sherwin. The value of saving a life: evidence from the labor market. In: Terleckyj NE, ed. Household Production and Consumption. New York: National Bureau of Economic Research, Inc., 1976: 265-298. Studies in Income and Wealth, No. 40;

Torrance, George W. "Measurement of Health State Utilities for Economic Appraisal," *Journal of Health Economics*, 1986, *5*, 1-30.

Torrance, George W. "Utility Approach to Measuring Health-Related Quality of Life," *Journal of Chronic Diseases*, 1987, *40*, 593-600.

Viscusi, W. Kip, and Moore, Michael J. "Rates of Time Preference and Valuations of the Duration of Life," *Journal of Political Economy*, 1989, *38*, 297-317.

Weber, Warren E. "The Effects of Interest Rates on Aggregate Consumption," *American Economic Review*, 1970, *60*, 591-600.

Weber, Warren E. "Interest Rates, Inflation and Consumer Expenditures," *American Economic Review*, 1975, *65*, 843-858.

Weinstein, Milton C., and Fineberg, Harvey V., *Clinical Decision Analysis*, Philadelphia: W.B. Saunders Company, 1980.

Weinstein, Milton C. "Economic Assessment of Medical Practices and Technologies," *Medical Decision Making*, 1981, *1*, 309-330.

</ref_section>

Table 1. Estimated cost-effectiveness of commonly used medical interventions. (All interventions compared to "usual care" unless otherwise noted).

Intervention	Cost/life-year ($1989)
Low-dose lovastatin for high cholesterol[a]	
Male heart attack survivors, age 55-64, cholesterol level $\geq$ 250	1,600
Male heart attack survivors, age 55-64, cholesterol level < 250	1,700
Female nonsmokers, age 35-44	1,500,000
Female hypertensive nonsmokers, age 35-44	710,000
Exercise electrocardiogram as screening test[b]	
40 year-old males	92,200
40 year-old females	248,500
Hypertension screening[c]	
40 year-old males	20,400
40 year-old females	31,300
Breast cancer screening[d]	
Annual breast examination, females age 55-65	11,300
Annual breast examination and mammography, females age 55-65	30,400
Physician advice about smoking cessation[e]	
1% quit rate, males age 45-50	2,800
Pap smear starting at age 20, continuing to 74[f]	
Every three years, compared to not screening	17,800
Every two years, compared to every three years	351,717
Coronary Artery Bypass Graft[g]	
Left main coronary artery disease	6,500
Single vessel disease with moderate angina	65,300
Neonatal Intensive Care Units[h]	
Infants 1000-1500 grams	8,100
Infants 500-999 grams	57,200

[a]Goldman, 1991; [b]Sox, 1989; [c]Littenberg, 1990; [d]Eddy, 1989; [e]Cummings, 1989; [f]Eddy, 1990; [g]Weinstein, 1981; [h]Boyle, 1983

Table 2
Base Case Optimal Spending and CE Cutoff by Age
Women, Income = $18,000 ($1989)

Age	Optimal Spending	Optimal CE Ratio
30	$ 0	$36,870
40	0	36,870
50	300	35,950
60	1,010	33,890
70	1,480	32,600

Table 3. Optimal spending and sensitivity of CE ratio to production function, for 60 year-old men and women.

	Women		Men	
	Spending	CE Ratio	Spending	CE Ratio
Maximum Risk Reduction α				
0.05	$0	$36,870	$0	$36,870
0.10	0	36,870	350	35,800
0.15	310	35,940	770	34,590
0.20	600	35,080	1,060	33,750
0.25	830	34,420	1,290	33,111
*0.30	1,010	33,890	1,480	32,600
Persistence of effect (ρ)				
0.20	$310	$35,920	$790	$34,530
0.40	600	35,060	1,090	33,710
*0.60	1,010	33,890	1,480	32,600
0.80	1,690	32,040	2,130	30,880
0.90	2,250	30,570	2,660	29,530

* = base case

Table 4
Optimal Spending and CE Ratio
Men, Income = $18,000

Age	$\delta = 1$		$\delta = .98$		$\delta = .95$		$\delta = .91$	
	Spending	CE	Spending	CE	Spending	CE	Spending	CE
30	$780	$34,560	$ 320	$ 35,890	$ 0	$36,870	$0	$36,870
40	1000	33,930	620	35,020	150	36,410	0	36,870
50	1140	33,520	850	34,340	480	35,430	0	36,870
60	2000	31,210	1780	31,790	1480	32,600	1060	33,760
70	2260	30,540	2110	30,930	1900	31,470	1590	32,290

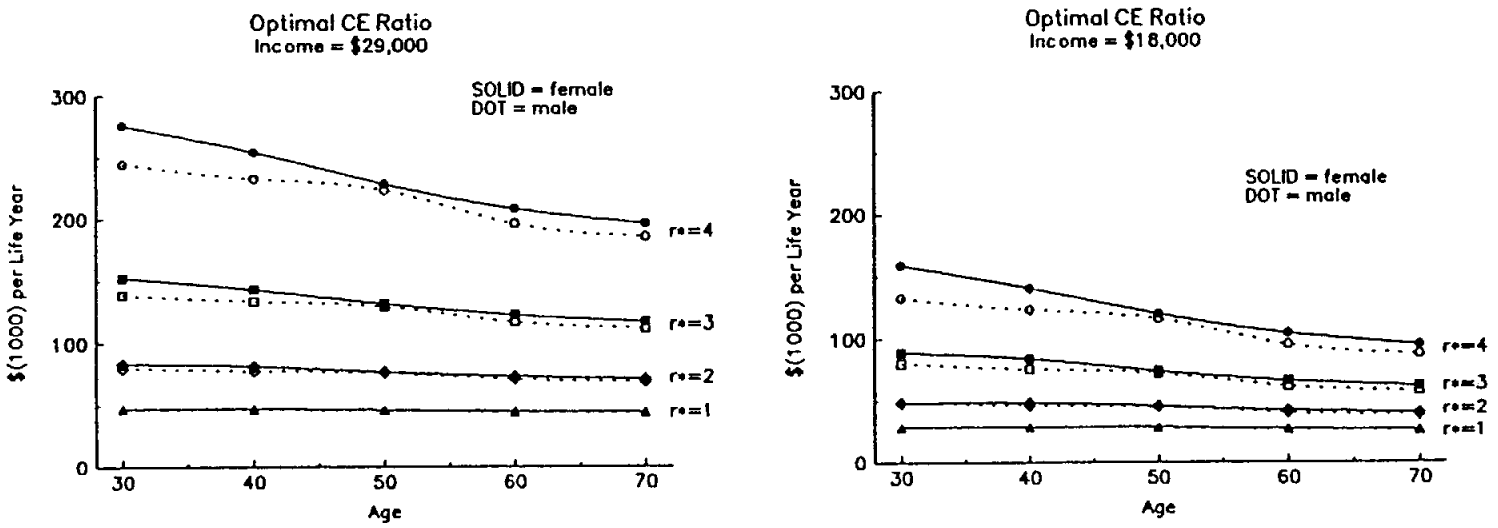

Figure 1. Optimal cost-effectiveness ratio for women and men, by age and relative risk aversion (R^*).