Introduction

In medical cost-effectiveness analysis (CEA), an incremental cost-effectiveness (C/E) ratio comparing a new treatment (T_1) to some alternative intervention (T_0) is typically defined as:

$$R \equiv (\mu_{C1} - \mu_{C0})/(\mu_{E1} - \mu_{E0}) \qquad (1)$$

where μ_{Ci} and μ_{Ei} represent the mean cost and mean health effect, respectively, of treatment T_i.* The numerator and denominator of this ratio are the incremental cost and incremental effectiveness, respectively, of the new intervention relative to its comparator. The ratio can be interpreted as the additional investment of resources needed for each additional unit of health improvement expected to result from investing in T_1 rather than T_0.

Because the true (population) means are not known, R is estimated using the "analogy" estimator[1]:

$$\hat{R} = (\bar{C}_1 - \bar{C}_0)/(\bar{E}_1 - \bar{E}_0) \qquad (2)$$

where $\bar{C}_i$ and $\bar{E}_i$ represent sample means for the cost and effect of intervention T_i. Due to uncertainty in these estimates, an important component of any CEA is an analysis of the uncertainty surrounding a C/E ratio estimate. While univariate sensitivity analysis and simple types of multivariate sensitivity analysis (such as best-case and worst-case scenarios) are now commonplace in the health economics literature, these methods suffer from a number of limitations.[2] Ideally, a sensitivity analysis should convey information regarding both the range of possible results and the probability of each possible outcome being realized. These estimates would need to incorporate information about the joint probability distribution of the key

---

* Some analysts make a distinction between cost-effectiveness analysis (CEA), in which outcomes are measured in "natural" units such as years of life saved, and cost-utility analysis (CUA), in which adjustments are made to reflect patient or community preferences over health

variables in an analysis – information that is absent from univariate and simple multivariate sensitivity analyses. To address this issue, analysts have recently begun to investigate new approaches to the analysis of uncertainty in CEA.

Most research in this area has focused on methods of analysis for studies in which the analyst has patient-level data on the costs and health effects of alternative interventions. The data may be gathered in either randomized controlled trials or observational studies, and CEAs based on this type of data are sometimes referred to as "stochastic CEAs." The wealth of data available to analysts in stochastic CEA presents the opportunity for more sophisticated methods of sensitivity analysis than are feasible in evaluations that rely on gross, population-level estimates of costs and effects. To date, discussions of methods for sensitivity analysis in stochastic CEA have focused primarily on the estimation of confidence intervals (CIs) around $\hat{R}$. As we will show, however, this approach suffers from important theoretical limitations; of particular concern is the fact that inherent ambiguities in the probability distribution of a C/E ratio estimator render any inference based on that distribution (including the construction of CIs) suspect.

As an alternative to the estimation of CIs around $\hat{R}$, we present a new framework for the analysis of uncertainty in economic evaluation. This method of analysis, which we term the "net health benefits" approach, offers several practical and theoretical advantages over the use of CIs for C/E ratios, is straightforward to apply, requires no more data than does C/E ratio analysis, and highlights some important principles in the theoretical underpinnings of CEA.

Estimating Confidence Intervals for C/E Ratios

Several methods have been presented for estimating CIs around $\hat{R}$, given sampled data on the costs and effects of an intervention and its comparator. The first method proposed for this purpose (the "delta method") estimates the variance of $\hat{R}$ using a second-order Taylor series approximation. A two-tailed $(1-\alpha)$ CI can then be constructed as:

---

states. In the present paper, the term CEA is used to encompass both of these analytic frameworks.

$$\hat{R} \pm z_{\alpha/2}\sqrt{\hat{\sigma}^2_{\hat{R}}} \qquad (3)$$

where z is the test statistic of the standard normal distribution and $\hat{\sigma}^2_{\hat{R}}$ is the estimated variance of the ratio; for a detailed description, see O'Brien et al.[3] However, the Taylor series approximation of variance does not generally work well for ratios,[4] and the assumption of a well-behaved parametric distribution for $\hat{R}$ is questionable. For example, if incremental costs and effects are distributed independent unit normal then $\hat{R}$ follows a Cauchy distribution (a t-distribution with one degree of freedom), which has no mean and infinite variance.[5 6]

Recently, analysts have begun to focus on alternative approaches, including non-parametric bootstrapping and the use of Fieller's Theorem, for constructing CIs. For an overview of these methods, which avoid some of the delta method's pitfalls, readers are referred to Chaudhary and Stearns.[7] However, while some progress has been made regarding the practical issue of *how* to construct CIs for C/E ratios, little attention has been devoted to questions of *why* (or *if*) these CIs are desirable and how they should be interpreted and used by decision makers. In the following section it is shown that the fundamental concept of using a CI to convey information about uncertainty around a C/E ratio suffers from important theoretical limitations — regardless of the estimation procedure employed.

Theoretical Problems in the Use of Confidence Intervals for C/E Ratios

THE DECISION RULES OF CEA

To illustrate the problems associated with constructing and interpreting CIs for $\hat{R}$, we must first briefly review the decision rules of CEA. Consider the ΔE-ΔC plane (Figure 1), in which the horizontal axis measures the incremental effectiveness and the vertical axis measures the incremental cost of T_1 compared to T_0. For each quadrant of the plane (note the quadrant

numbering convention we use), the theory of cost-effective resource allocation prescribes a choice between T_1 and T_0 as follows:

[INSERT FIGURE 1 HERE]

Quadrant ⇒	Decision Rule
I	$T_1 \succ T_0$ if and only if $\Delta C/\Delta E < \lambda$
II	$T_1 \succ T_0$
III	$T_1 \succ T_0$ if and only if $\Delta C/\Delta E > \lambda$
IV	$T_0 \succ T_1$

where $\succ$ is the preference relation (read "is preferred to") and λ is the threshold C/E ratio, indicated in Figure 1 by the slope of the dashed line. Thus, $T_1 \succ T_0$ for all points below the dashed line and $T_0 \succ T_1$ for all points above the line. The threshold ratio (λ) can be interpreted as the maximum amount that society would be willing to pay for an incremental gain in health or, equivalently, the minimum amount that society would be willing to accept in exchange for forgoing an incremental gain in health.[†] When considering the problem of allocating resources at the societal level, λ can be assumed to be exogenous to the evaluation of any individual intervention [this point will be addressed in detail in a later section of this paper]. With these decision rules in mind, let us now turn our attention to $\hat{R}$ and its probability distribution.

[†] Contingent valuation studies often find that the amount an individual would be willing to pay for a given reduction in risk to his or her health is less than the compensation that the same individual would require in order to willingly experience a risk increase of equal magnitude. However, it is not clear whether this disparity between willingness to pay and willingness to sell persists at the level of social choice.

AMBIGUITY BETWEEN ΔE-ΔC QUADRANTS IN THE DISTRIBUTION OF A C/E RATIO ESTIMATOR

Note that $\hat{R}$ is positive in quadrants I and III and negative in quadrants II and IV of the ΔE-ΔC plane. If a negative ratio corresponds to quadrant II then $T_1 \succ T_0$, but if it corresponds to quadrant IV then $T_0 \succ T_1$. Similarly, a positive ratio less than λ is favorable for T_1 in quadrant I but unfavorable for T_1 in quadrant III. Thus, $\hat{R}$ has no meaningful interpretation unless it is presented in the context of the quadrant of the ΔC-ΔE plane to which it corresponds. Because the probability distribution of $\hat{R}$ conveys no such contextual information even when there is non-negligible probability of the joint distribution of costs and effects extending to more than one quadrant of the ΔE-ΔC plane, this distribution and any inference based on it — including the construction of CIs — is ambiguous.

PROBLEMS WITH NEGATIVE C/E RATIOS

In some evaluations, the sign of either ΔE or (more likely) ΔC may be determined with high probability; when this is the case, the ambiguity discussed above is not an issue. Even in these situations, however, fundamental problems in the interpretation of negative C/E ratios present important complications for the construction and interpretation of CIs based on the distribution of $\hat{R}$. For example, suppose it is known with certainty that $\Delta C < 0$, so that the analysis is limited to quadrants II and III of the ΔE-ΔC plane; further suppose that there is non-negligible uncertainty with respect to the sign of ΔE, so that the distribution of $\hat{R}$ includes both positive and negative values. For the negative portion of $\hat{R}$'s distribution (corresponding to quadrant II), the new treatment is estimated to be both less costly and more effective than its comparator; thus, in this quadrant, a large magnitude is desirable in both the numerator and the denominator of the C/E ratio. However, these two desirable features drive $\hat{R}$ in opposite directions: large incremental health gains in the denominator drive the ratio closer to zero, but large incremental cost savings in the numerator drive the ratio toward negative infinity. The result is that the

negative portion of the probability distribution of $\hat{R}$ does not lend itself to meaningful interpretation. Therefore, unless the joint distribution of incremental costs and incremental health effects is limited to either quadrant I or quadrant III — which would in general seem tenuous to maintain *a priori* — any inference based on the distribution of $\hat{R}$ is problematic.

It should be noted that, despite the important conceptual problems with reporting negative C/E ratios, it is not uncommon for them to be reported in the health economics literature. For example, in a study published in the recent report of the Panel on Cost-Effectiveness in Health and Medicine as an example of an analysis performed in accordance with the Panel's recommendations,[‡] a cost-effectiveness ratio of –$13,000 per quality-adjusted life year (QALY) gained was reported for a strategy to fortify grain products with folic acid to prevent neural tube defects.[8] Interpreting this negative ratio, the authors of the study noted that the strategy "resulted in cost savings of about $13,000 accompanying every QALY gained." While this may be an accurate statement of the study's results, it is unclear what the reader is expected to conclude from the magnitude of the reported estimate. Is it better to save $13,000 per QALY gained than to save, say, $6,500 per QALY gained? The answer is, "it depends." If the former ratio has a larger magnitude due to greater cost savings in the numerator, then the answer is "yes," but if it has a larger magnitude due to lower incremental effectiveness in the denominator, then the answer is "no." From the magnitude of the ratio alone, absent information about the respective magnitudes of the numerator and denominator, one can draw no meaningful conclusions. Recent analyses have also reported negative values for the lower limits of CIs for C/E ratios (see, for example, Obenchain et al.[9]).

[‡] The Panel of Cost-Effectiveness in Health and Medicine was appointed by the U.S. Public Health Service to make recommendations for standardizing the methods employed in CEA.

Net Health Benefits

DEFINITION AND INTERPRETATION

In response to the problems associated with inference based on the distribution of a C/E ratio (including both the conceptual problems discussed above and further problems related to inference discussed in the following sections) we propose a new approach for the analysis of uncertainty in the economic evaluation of health interventions. We begin by defining the average net health benefit (NHB) of intervention T_i as:

$$\mu_{Ei} - \mu_{Ci}/\lambda \tag{4}$$

and the incremental NHB of T_1 compared to T_0 as:

$$\left(\mu_{E1} - \mu_{C1}/\lambda\right) - \left(\mu_{E0} - \mu_{C0}/\lambda\right) = \left(\mu_{E1} - \mu_{E0}\right) - \left(\mu_{C1} - \mu_{C0}\right)/\lambda \tag{5}$$

where all variables are as previously defined.

The first part of expression (4) is simply the health effect associated with intervention T_i. The second part of the expression represents the health gain that could have been attained by instead investing the resources consumed by T_i in a marginally cost-effective program ($R=\lambda$). Thus, the average NHB of an intervention is interpreted as the net benefit (measured in units of health) of investing resources in T_1 rather than investing those resources in a marginally cost-effective program. The interpretation of the incremental NHB of T_1 compared to T_0 — which is calculated as the difference between the average NHB of T_1 and the average NHB of T_0 — is similar. The first part of this expression on the right-hand side of equation (5) measures the incremental health effectiveness of new intervention T_1 relative to comparator T_0. The second part of the expression represents the health gain that could be attained by investing ΔC resources in a marginally cost-effective program ($R=\lambda$). Thus, this expression can be interpreted as the net benefit (measured in units of health) of investing resources in T_1 rather than implementing T_0 and

investing the "left over" ΔC resources in a marginally cost-effective program. For NHB>0, T_1 is deemed cost-effective and should be selected for implementation. For NHB<0, more health improvement could be attained by forgoing the intervention in question and investing resources elsewhere; T_1 is therefore deemed cost-ineffective and should not be selected for implementation.

The threshold ratio λ can also be interpreted as society's willingness to pay for an incremental gain in health. Following this interpretation, an intervention's average NHB compares the effectiveness of that intervention (μ_{Ei}) to the minimum health effect that society would demand in return for its investment of μ_{Ci}, which is equal to μ_{Ci}/λ. The incremental NHB of T_1 relative to T_0 compares the difference in the two programs' health effects ($\mu_{E1} - \mu_{E0}$) to the minimum difference in health effects society would demand in order to justify the additional expenditure required to implement T_1 rather than T_0; for an incremental cost of ($\mu_{C1} - \mu_{C0}$), the minimum acceptable health gain is equal to $(\mu_{C1} - \mu_{C0})/\lambda$.

In the remainder of this paper, we will assume that an incremental analysis comparing two or more interventions is being performed, so that all NHBs are understood to be incremental rather than average, unless otherwise stated.

CONFIDENCE INTERVALS FOR NET HEALTH BENEFITS

The most straightforward estimator of NHB is the "analogy" estimator:

$$\hat{NHB} = \bar{E}_1 - \bar{E}_0 - (\bar{C}_1 - \bar{C}_0)/\lambda \qquad (6)$$

A parametric CI for $\hat{NHB}$ can be readily constructed. The variance of $\hat{NHB}$ is estimated as:

$$\hat{\sigma}^2_{NHB} = \sum_{i=0}^{1} \left(s^2_{Ei} + s^2_{Ci} \ \lambda^2 - 2 r_i s_{Ei} s_{Ci} / \lambda \right) / n_i \qquad (7)$$

where E_i and C_i are random variables representing the costs and effects of intervention T_i; s^2_{Ei} and s^2_{Ci} represent the sample variance of E_i and C_i, respectively; r_i is the sample correlation

coefficient for E_i and C_i; n_i is the number of observations for T_i; and λ is the threshold C/E ratio. Note that because interventions T_1 and T_0 are generally observed in different (independent) samples, it is reasonable to assume no correlation between E_1 and E_0 and between C_1 and C_0, but it is necessary to allow for nonzero correlations between the costs and effects of an individual intervention.

By the Central Limit Theorem, $\hat{NHB}$ is asymptotically normal, and a two-tailed $(1-\alpha)$ CI can be constructed as:

$$\hat{NHB} \pm z_{\alpha/2}\sqrt{\hat{\sigma}^2_{NHB}} \tag{8}$$

where z is the test statistic for the standard normal distribution. As a rule of thumb, the normal approximation is generally considered reasonable for sample sizes ≥ 30. If the joint probability distribution of costs and effects is multivariate normal, then — because NHB is linear in its arguments — $\hat{NHB}$ has an exact normal distribution even in finite samples.

Alternatively, a nonparametric CI for $\hat{NHB}$ can be constructed using bootstrapping techniques, based on the empirical joint distribution of costs and effects for an intervention and its comparator. The method, based on Efron and Tibshirani[10] and Davidson and MacKinnon[11] is as follows:

1. From the original sample of n_1 observations for T_1 and n_0 observations for T_0, use equation (6) to estimate the incremental NHB of T_1 relative to T_0.

2. Repeat the following steps (a-c) a large number of times; although there is no precise guide as to how many replicates are sufficient, a few thousand are often satisfactory. Each replicate is referred to as a "bootstrap" replicate, and each of these yields a bootstrap NHB estimate, which we will denote $\hat{NHB}_b$ (for b= 1 to B, where B represents the number of replicates).

a. From the observations for T_1, draw a random sample of size n_1 with replacement. Calculate the average cost and effectiveness for this bootstrap sample; these values will be denoted $\bar{C}_1^b$ and $\bar{E}_1^b$.

b. From the observations for T_0, draw a random sample of size n_0 with replacement. Calculate the average cost and effectiveness for this bootstrap sample; these values will be denoted $\bar{C}_0^b$ and $\bar{E}_0^b$.

c. For each bootstrap replicate, calculate an estimate of the incremental NHB of T_1 relative to T_0:

$$\hat{NHB}_b = (\bar{E}_1^b - \bar{E}_0^b) - (\bar{C}_1^b - \bar{C}_0^b)/\lambda \qquad (9)$$

3. Eliminate the $B^*(\alpha/2)$ lowest values and the $B^*(\alpha/2)$ highest values of NHB_b, where $(1-\alpha)$ is the desired confidence level. The lowest remaining value and the highest remaining value for NHB_b are the lower and upper bounds, respectively, of a nonparametric CI for NHB. (More complex methods that account for complications such as finite-sample asymmetries in the sampling distribution have been discussed by Efron and Tibshirani[10]; the simple approach presented here should be viewed as a starting point if inference is a central concern in the analysis.)

Note that NHB as described above represents the net health benefit per patient, or, more generally, per unit of observation. Alternatively, a decision maker may wish to consider the total net health benefit associated with a particular program: $NHB_{total} = N*NHB$, where N is the size of the target population. A $(1-\alpha)$ CI for NHB_{total} is simply N times the $(1-\alpha)$ CI for NHB.

Net Health Benefits versus C/E Ratios

INTERPRETABILITY

As discussed earlier, the interpretation of a C/E ratio estimate is ambiguous without information regarding the quadrant of the ΔE-ΔC plane to which the estimate corresponds, and the probability distribution of $\hat{R}$ conveys no such contextual information. Moreover, because the magnitude of a negative C/E ratio conveys no useful information, inference that is dependent on the negative portion of $\hat{R}$'s probability distribution is problematic. In contrast, NHB suffers

from no such ambiguities. A positive (negative) value for NHB is unambiguously favorable (unfavorable) for the new intervention being evaluated, and values of NHB become continuously more favorable as one moves upward from negative infinity. That is, for a rational decision maker whose objective is to allocate scarce health care resources efficiently, preferences are monotonic in net health benefits but not in cost-effectiveness ratios.

To make this distinction clear, note that a program's attractiveness is a monotonically decreasing function of incremental cost (holding effectiveness constant) and a monotonically increasing function of incremental effectiveness (holding costs constant). These characteristics of preferences are clearly reflected in NHB, which is a linearly increasing function of effectiveness ($\partial NHB/\partial \Delta E = 1$) and a linearly decreasing function of costs ($\partial NHB/\partial \Delta C = -1/\lambda$). In contrast, the direction of the change in cost-effectiveness ratio R associated with a change in ΔC depends on the sign of ΔE ($\partial R/\partial \Delta C = 1/\Delta E$), and the direction of the change in R associated with a change in ΔE depends inversely on the sign of ΔC ($\partial R/\partial \Delta E = -\Delta C/(\Delta E)^2$).

IMPLICATIONS FOR STATISTICAL INFERENCE

Largely because NHB is linear in costs and effects, statistical inference is far more straightforward when using NHB than when using C/E ratios.[12] For example, while the assumption of a well-behaved parametric distribution for $\hat{R}$ is questionable, $\widehat{NHB}$ is asymptotically normal under quite general assumptions even if the joint distribution of costs and effects is non-normal. Similarly, while the Taylor series approximation of variance does not sufficiently capture the nonlinearity of $\hat{R}$,[4] the sample estimate of the variance of NHB is straightforward to calculate and can be estimated without bias.

Moreover, note that because the sample mean is an unbiased estimate of the population mean, the sample ("analogy") estimate $\widehat{NHB}$ is an unbiased estimate of the true NHB. In contrast, the sample estimate $\hat{R}$ is a biased estimate of the true C/E ratio R; that is, $E(\hat{R}) \neq R$.[13] Because the bias approaches zero as the sample size tends toward infinity, $\hat{R}$ is a consistent estimator of R; thus, the bias may be negligible in studies with large sample sizes but is potentially important when sample sizes are small. An unbiased estimate of the C/E ratio can be

obtained using a bootstrap adjustment to $\hat{R}$, but caution is urged when considering this approach because the bootstrap estimate of bias is subject to sampling variability, so that adjusting for bias may increase the mean square error of the estimate.[10 13]

Another attractive feature of NHB resulting from its linearity in costs and effects is the fact that the mean of a distribution of NHB estimates is equal to the value of $\hat{NHB}$ evaluated at the mean estimates of effects and costs. In contrast, the mean of a distribution of C/E ratio estimates is not generally equal to the ratio of the mean estimate of incremental effectiveness to the mean estimate of incremental cost, and analysts have reached different conclusions regarding which of these estimates (the mean ratio or the ratio of means) should be used to estimate a C/E ratio under conditions of uncertainty.[14 15 16 17]

Note also that a bootstrap estimate of the variance of $\hat{NHB}$ would be expected to converge more quickly than similar estimates for C/E ratios. Cost-effectiveness ratios take on values approaching infinity (or negative infinity) when incremental effectiveness is close to zero, and the possibility of C/E ratio estimates of near-infinite magnitude can cause significant problems for the convergence of bootstrap estimates of the mean and variance of the distribution of $\hat{R}$.

Additional advantages of NHB over C/E ratios with regard to statistical inference are discussed in the following section, in the context of economic evaluations with multiple comparators.

Stochastic Analysis with Multiple Comparators

While several papers have discussed stochastic methods for analyzing uncertainty in CEA, these discussions have been limited to situations in which two interventions are being compared to each other. Because multiple mutually exclusive interventions are often available for consideration, it is important to consider methods for analyzing uncertainty in an economic evaluation with multiple comparators.

For the estimation of incremental C/E ratios, the presence of multiple comparators introduces significant complexity. In CEA, incremental C/E ratios are calculated after all dominated programs have been removed from consideration and the remaining programs have

been ranked in order of increasing cost. A program is dominated if it is both more costly and less effective than at least one of its comparators ("strong dominance"), or if it is both more costly and less effective than a convex combination of two of its comparators ("weak dominance," or "extended dominance," which only becomes a possibility when there are more than two programs being compared).[18 19]

When multiple comparators are being analyzed, the ranking of programs (including both the identification of dominated programs and the ranking of the undominated programs) may be uncertain. Thus, for a single intervention (call it T_a), stochastic analysis — performed, for example, using bootstrapping techniques — may indicate that the intervention has non-zero probabilities of being strongly dominated, of being weakly dominated, of dominating all of its comparators, and of being ranked between each possible pair of its comparators. In those bootstrap replications where the program is estimated either to be dominated or to dominate all of its comparators, no C/E ratio is calculated for that intervention. For those replications where the program is not dominated and does not dominate all other alternatives, the incremental C/E ratio of that intervention may be calculated relative to one program (T_b) in some cases and relative to another program (T_c, T_d, ...) in other cases.

Additional complications are also introduced. In analyses with only two mutually exclusive interventions, one may estimate the probability that the more costly intervention would be deemed cost-effective, conditional on some value of λ, as one indication of a program's attractiveness.[20] With multiple comparators, interpretation is less straightforward. For example, one might estimate the probability, conditional on a value of λ, that program T_a either is dominant over all other alternatives or has an incremental C/E ratio more favorable than λ. However, using this probability as an indication of the program's attractiveness would not take into account the fact that one or more of T_a's comparators might also have a C/E ratio less than λ; if one of the T_a's comparators (call it T_b) has a C/E ratio less than λ and is also more effective than T_a, then the theory of efficient resource allocation indicates that T_b should be selected for implementation rather than T_a. The analyst considering such a situation would need to simultaneously consider information about the probability of being (strongly or weakly) dominated, the ranking of the program (if not dominated) relative to the other undominated alternatives, the estimated cost-effectiveness of the intervention relative to the appropriate

comparator (which may vary as the ranking of programs by cost varies), and the cost-effectiveness of other, more effective comparators.

The NHB approach, however, can be readily extended to the stochastic evaluation of multiple comparators. Suppose that there are k programs being evaluated (T_i, i=1 to k). After B bootstrap replications have been performed for each intervention, the probability that intervention T_i is the most attractive option is simply calculated as the percentage of those replicates for which T_i is estimated to have the highest average NHB as defined in expression (4). Analysts may find this to be considerably simpler than what would be required to calculate the same probability using methods based on incremental C/E ratios. Using NHB methodology, one could also easily construct a CI for the difference between the average NHB for program T_i and the average NHB for the most attractive program from each bootstrap replication, to give an indication of how much program T_i tends to underperform compared to the optimal choice (the identity of which may vary across replicates); it is not clear how a similar measure could be constructed when basing an analysis on incremental C/E ratios.

The advantage of NHBs over C/E ratios in multiple-comparator analyses is attributable in large part to the fact that NHBs are separable while incremental C/E ratios are not separable. That is, the incremental NHB of T_i compared to T_0 is simply the difference between the average NHBs of the two interventions, so that these average NHB values play a meaningful role in incremental analyses making comparisons across interventions. In contrast, an incremental C/E ratio is not a separable function of average C/E ratios (μ_{Ci}/μ_{Ei}), and calculating interventions' average C/E ratios is not generally helpful when making comparisons across mutually exclusive alternatives; because of their limited usefulness and because they can be easily misinterpreted, the Panel on Cost-Effectiveness in Health and Medicine cautions against reporting average cost-effectiveness ratios.[21]

STOCHASTIC DOMINANCE

Another result of the separability of incremental NHBs vis-à-vis the inseparability of incremental C/E ratios is that the concept of stochastic dominance[22 23 24] is relevant in the analysis of NHBs but not in the analysis of C/E ratios. Stochastic dominance (which should not be

confused with the concepts of strong dominance and extended dominance discussed earlier in this paper) is a concept that is useful for making comparisons across mutually exclusive choices under conditions of uncertainty. In the context of NHBs, intervention T_1 dominates intervention T_0 through *first-order stochastic dominance* if and only if the following condition applies:

$$F(NHB) \leq G(NHB) \text{ for all values of NHB} \tag{10}$$

where F(NHB) is the cumulative distribution function of average NHB for treatment T_1 and G(NHB) is the cumulative distribution function of average NHB for treatment T_0, and where the weak inequality (≤) must be a strict inequality (<) for at least one value of NHB. Graphically, this means that the cumulative distribution function for T_1 must never lie above the cumulative distribution function for T_0, and must lie below it for at least one value of NHB; an example of first-order stochastic dominance of T_1 over T_0 is shown in Figure 2. It can be shown that if T_1 dominates T_0 through first-order stochastic dominance, any utility-maximizing decision maker whose utility is monotonically increasing in NHB ($\partial U/\partial NHB > 0$) should prefer T_1 to T_0.[22] This result holds regardless of the decision maker's risk posture.

[INSERT FIGURE 2 HERE]

When first-order stochastic dominance is not present, an analyst may also check for *second-order stochastic dominance* of T_1 over T_0, which is defined as the following condition:

$$\int_{-\infty}^{NHB} F(x)dx \leq \int_{-\infty}^{NHB} G(x)dx \quad \text{for all values of NHB} \tag{11}$$

where all variables are as previously defined, and where the weak inequality (≤) must be a strict inequality (<) for at least one value of NHB. Graphically, this means that the cumulative distribution function for T_1 might (or might not) lie above the cumulative distribution function for T_0 at some point, but the area under the cumulative distribution function for T_1 evaluated up to any NHB value is never greater than (and is at least sometimes less than) the area under the

cumulative distribution function for T_0 evaluated up to that same value of NHB. In the example shown in Figure 3, T_1 dominates T_0 if and only if the area between the two curves where F (solid line) lies below G (dotted line) is greater than the area between the two curves where F lies above G. It has been shown that if T_1 dominates T_0 through second-order stochastic dominance, any utility-maximizing decision maker whose utility is monotonically increasing in NHB and who is risk averse over NHB ($\partial U/\partial NHB > 0$ and $\partial^2 U/\partial NHB^2 < 0$) should prefer T_1 to T_0.[22]

[INSERT FIGURE 3 HERE]

Stochastic dominance is a powerful analytic tool because it allows one to identify cases in which a decision maker should unambiguously prefer one alternative over another despite the presence of uncertainty, with only very general assumptions required regarding the decision maker's utility function. This type of analysis can be readily performed in the NHB framework because preferences are monotonically increasing in NHB and because incremental NHBs are separable, so that it is meaningful to compare average NHBs across interventions. In contrast, the concept of stochastic dominance cannot be applied in analyses based on incremental C/E ratios, because preferences are not monotonic in R and monotonicity is a necessary condition for stochastic dominance.

Net Health Benefits and the Threshold C/E Ratio

REPORTING NET HEALTH BENEFITS AS A FUNCTION OF λ

One issue of concern in the use of NHBs is the fact that the societal threshold C/E ratio (λ) is not known.[25] Indeed, one might argue that uncertainty around λ limits the usefulness of the NHB approach. The prudent approach to addressing this issue is also a quite simple one, to wit: Carry out the analysis for a range of values for λ and report $\hat{NHB}$ as a function of λ. Indeed, because the threshold ratio plays such a prominent role in this type of analysis, we suggest that all empirical NHB estimates be reported using the notation NHB_λ to indicate the value of λ

corresponding to the estimate. A useful tool for presenting the results of NHB analyses and the sensitivity of these results to λ would be to plot both point NHB estimates and their CIs as a function of λ; an example of such a plot is shown in Figure 4.

[INSERT FIGURE 4 HERE]

Analysts may also make inferences regarding the probability that an intervention is estimated to be cost-effective relative to its comparator ($\hat{NHB}_\lambda>0$). For nonparametric estimation, the probability that $\hat{NHB}_\lambda$ is positive is calculated as the percentage of bootstrap replicate estimates that are positive, conditional on a given value of λ. This probability can then be plotted as a function of λ, producing a graph identical to the C/E acceptability curve presented by van Hout et al.[26] In addition, under the assumption that $\hat{NHB}$ is normal, one can use standard parametric techniques to perform the following one-sided hypothesis test of cost-effectiveness for any value of λ: H_0: $\hat{NHB}_\lambda=0$, versus H_a: $\hat{NHB}_\lambda>0$.

Note that the problem of uncertainty with respect to λ is not limited to the NHB approach; it is an equally important concern when using C/E ratios. One cannot generally assess whether or not a particular program is cost-effective without making some assumption regarding the value of λ. NHB analysis simply makes explicit that which C/E ratio analyses leave implicit. We would contend that the fact that NHB analysis *forces* explicit consideration of λ, and, hence, its unknown character, should be considered an advantage — *not* a drawback — of this approach.

THE THRESHOLD C/E RATIO AND OPPORTUNITY COSTS

Another attractive property of the NHB approach and its explicit consideration of λ is that it forces decision makers to confront the issue of opportunity costs. In CEA, a program with a C/E ratio above the threshold value ($R>\lambda$) may be recognized as being inefficient, or in some sense "expensive" relative to the benefit it offers, but the consequences associated with investing in such a program might not be clear to decision makers. It is important to bear in mind that, due to resource constraints, not all programs offering some potential for health improvement can be

implemented; therefore, investing in one program reduces the volume of resources available to invest in others. If $1/\lambda$ is interpreted as the shadow price of relaxing the health care budget constraint, this implies that there are investment opportunities offering C/E ratios of $R=\lambda$ that are currently being forgone.[27] That is, on the margin, a dollar invested in a program with $R>\lambda$ could instead be diverted to a program with $R=\lambda$, thereby yielding more health improvement without any additional net resource consumption. By quantifying these opportunity costs, the NHB approach confronts the decision maker with the fact that investing in a cost-ineffective program is not simply an unwise use of money in some vague sense — it is a forgone opportunity to achieve greater gains in people's health. When decision makers choose to invest in programs believed to have $R>\lambda$, explicitly reminding them of the human costs associated with these resource allocation decisions may motivate them to give further consideration to the fact that such investments implicitly assign a greater value to the health of some individuals (those targeted by the cost-ineffective program) than to the health of others (those who would benefit from marginally cost-effective programs currently being forgone).[28] While this information is implicit in the use of C/E ratios, NHBs offer the advantage of making the opportunity costs explicit.

TREATING λ AS A RANDOM VARIABLE

In this paper, we have suggested that NHB estimates should be calculated conditional on a value of λ, with sensitivity analyses performed to report $\widehat{NHB}$ as a function of λ. However, when considering how to reflect uncertainty around λ, one's first instinct might be to suggest estimating a probability distribution for λ and incorporating this distribution directly into the NHB analysis. When evaluating the merits of such an approach, it is important to consider the nature of the uncertainty around λ, and indeed the nature of λ itself.

Note that, in contrast to the "micro-level" clinical variables corresponding to individual interventions' respective costs and effects, λ is a "macro-level" public policy variable. Whereas fairly well-accepted methods have been developed for using the realized values of a random variable (such as costs or health effects) observed in a sample to make inferences regarding that

variable's distribution, it is not clear how one would obtain a sample of realized values for the public policy variable λ.

One approach might be to observe decisions about which health interventions are implemented and which are not, and to draw inferences about λ based on the relationship between programs' estimated C/E ratios and the levels at which those programs are implemented. It is this general approach that has led some researchers to conclude that the value of λ might be somewhere in the range of about $20,000 to $100,000 per quality-adjusted life year (QALY) gained. However, the usefulness of such an approach is limited by the fact that the concept of a threshold C/E ratio is generally cited for its normative appeal, and not for its descriptive accuracy. That is, while arguments based on welfare economics or optimization theory may suggest that a threshold value λ should play a meaningful role in attempts by a rational decision maker to allocate health care resources efficiently, it is well-known that human decisions often diverge greatly from the decisions suggested by these formal frameworks.[28] Thus, the degree to which one can make inferences about λ on the basis of observing actual resource allocation decisions is unclear.

Other possible approaches to estimating a probability distribution for λ would be to elicit the "expert judgments" of various individuals believed to have some information about λ in order to formulate some "prior" distribution on λ, or to undertake contingent valuation studies to estimate people's willingness to pay per unit of improved health. However, these approaches are complicated by a number of issues, such as the question of how to identify "experts" on this issue and the appropriate choice of methods for eliciting experts' judgments and individuals' willingness to pay values. While such complications certainly do not imply that these approaches have no value, they do suggest that substantial work may be required for developing and implementing these methods. This raises the question of whether analysts should wait until this issue has been resolved to perform NHB analyses, or whether it may be more reasonable for them to proceed by performing NHB analyses conditional on λ (as we have suggested in this paper) at least until some consensus emerges regarding a distribution for λ.

In assessing the merits of performing NHB analyses conditional on λ, one issue to consider is whether or not it is reasonable to treat λ as being statistically independent of the costs and effects observed in a particular study. Formally, if λ is statistically independent of ΔC and ΔE,

the joint distribution of $(\Delta C,\Delta E,\lambda)$ can be characterized as the product of two independent distributions: $\phi(\Delta C,\Delta E,\lambda)=\phi_1(\Delta C,\Delta E)\times\phi_2(\lambda)$. If it is indeed reasonable to treat λ as being statistically independent of ΔC and ΔE from a particular study — i.e., if it is reasonable to view λ as being decision maker-specific rather than intervention-specific — this implies that it is reasonable to estimate net health benefits conditional on λ, rather than directly incorporating into the analysis an estimate of $\phi_2(\lambda)$. Moreover, if $\phi_2(\lambda)$ is not dependent on $\phi_1(\Delta C,\Delta E)$, this indicates that estimates of $\phi_1(\Delta C,\Delta E)$ should not vary across studies. In this case, comparability across studies would be hindered if different analysts made different assumptions about $\phi_2(\lambda)$.

If evidence were to emerge suggesting that the assumption of independence is unreasonable — if, for the sake of argument, individuals indicated that their willingness to pay for incremental gains in health depended on the results of a particular ongoing clinical trial in such a way that the estimated distribution of society's willingness to pay for improved health would depend in a non-negligible manner on the results of the trial in question — one might reasonably question the degree to which such a result is consistent with a normative framework on which one might wish an economic evaluation to be based. For example, if λ is viewed (as it is in the constrained optimization view of CEA) as the reciprocal of the shadow price of relaxing the constraint on society's overall volume of health care resources, then λ should not be sensitive to the characteristics of any individual health care program unless that single program accounts for a large portion of society's health care resource consumption.[18]

On the basis of the above considerations, we conclude that conducting NHB analysis conditional on λ is preferable to the alternative of directly incorporating an estimated or postulated distribution for λ. If one does wish to estimate or postulate a distribution for λ and incorporate this into an estimate of the distribution of an intervention's $\widehat{\text{NHB}}$, this should be done in a supplementary analysis, where the primary analysis reports $\widehat{\text{NHB}}$ as a function of λ, as described earlier in this section.

Net Health Benefits and Cost-Benefit Analysis

The NHB framework for evaluation presents cost-effectiveness data in a format similar to that associated with cost-benefit analysis (CBA); in CBA, all costs and benefits are measured in monetary units, and programs with positive net benefits are prescribed for implementation. The NHB is analogous to the net benefit used in CBA, except that NHB is measured in units of health rather than money. Indeed, a measure identical to NHB was cited by Phelps and Mushlin as one piece of evidence for "the (near) equivalence of cost-effectiveness and cost-benefit analysis" (but was not suggested by them for practical application in economic evaluation).[29]

One particularly important distinction stands out between the NHB approach and CBA. To many analysts and decision makers, the most troubling feature of CBA is that it values changes in health based on people's willingness to pay for those changes, the result being that CBA-based allocations of health care resources are unlikely to be independent of the distribution of wealth. The ethical implications of such an approach to resource allocation are troubling to many (though certainly not all) people, and this is one of the principal reasons (in addition to the practical issue of how to estimate willingness-to-pay values) that most economic evaluations of health interventions are currently conducted in the framework of CEA rather than CBA. By retaining from CEA the principle (or assumption) that all QALYs are valued equally from the societal perspective, the NHB approach is less likely than CBA to meet with resistance on ethical grounds. In addition, by expressing results in units of health rather than money, the NHB approach may appeal to people's sense that the emphasis in discussions of health care resource allocation should be on human life and health rather than on money; to some, this simple difference in framing (though mathematically irrelevant) may signal information regarding the priorities of those involved in health policy research.

The above points notwithstanding, a strong case can still be made for using CBA. Whereas the theoretical foundations of CEA are still being debated,[30] [31] CBA is firmly grounded in the theory of welfare economics.[32] Moreover, while some are disturbed by what they regard as inherent ethical problems in the use of willingness-to-pay information for valuing health benefits, others have raised questions about the reasoning behind these concerns.[33] The NHB approach does not attempt to resolve this issue, which is beyond the scope of the present paper. Although

NHB is in some ways analogous to the outcome measure employed in CBA, the underlying assumptions of NHB analysis are, for better or for worse, those of CEA.

Discussion

The result of any economic evaluation could be expressed as a net health benefit rather than a C/E ratio. However, because C/E ratios are so commonly reported and widely accepted as an analytic tool for informing resource allocation decisions, we have focused here on the use of NHBs for those situations in which the analysis of C/E ratios is problematic. In particular, the NHB approach may be most helpful when analyzing the uncertainty around a C/E ratio estimate for which the joint distribution of incremental costs and incremental health effects extends significantly into more than one quadrant of the ΔE-ΔC plane, or when multiple comparators are being evaluated simultaneously. The use of NHBs in these situations permits one to obtain unambiguous, unbiased estimates of the distribution of possible consequences (measured in units of health) of the decision to invest or not invest in a particular program. Moreover, this information can be reasonably approximated using simple parametric statistical techniques (as long as the sample size is not forbiddingly small), and the results convey valuable information about the opportunity costs associated with health resource allocation decisions.

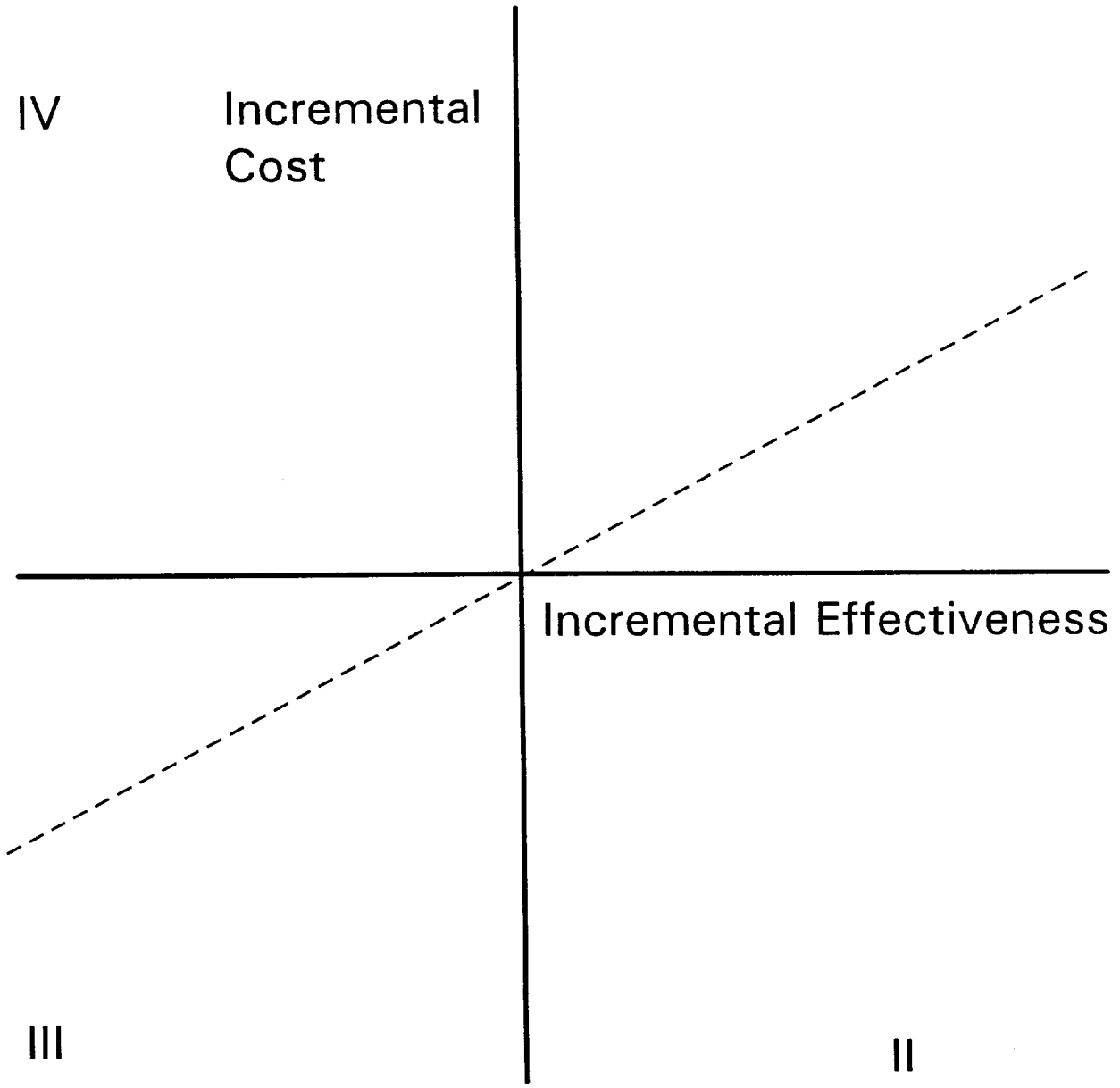

Figure 1. The ΔE-ΔC plane.

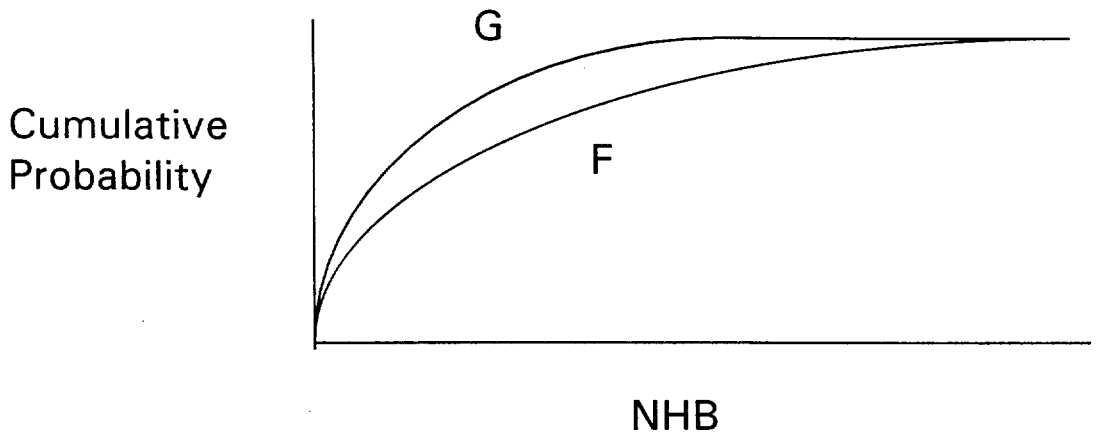

Figure 2. First-order stochastic dominance.

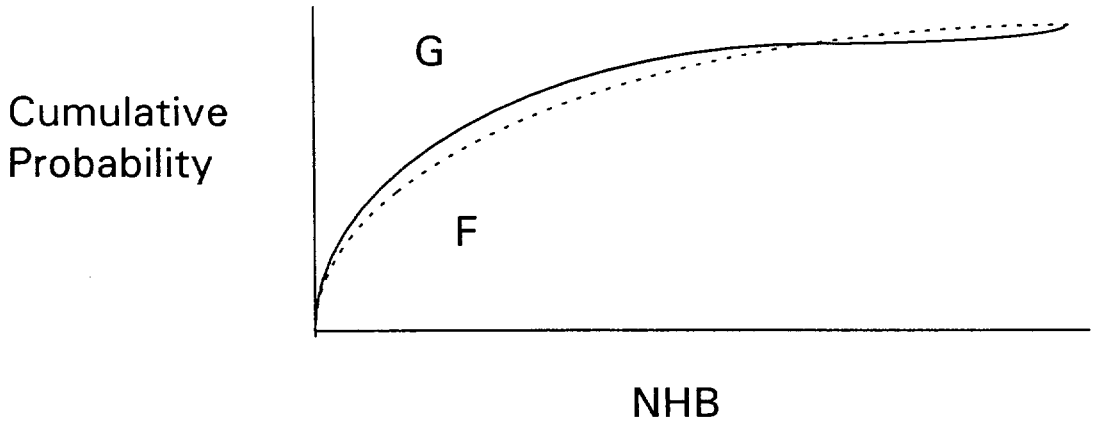

Figure 3. Second-order stochastic dominance.

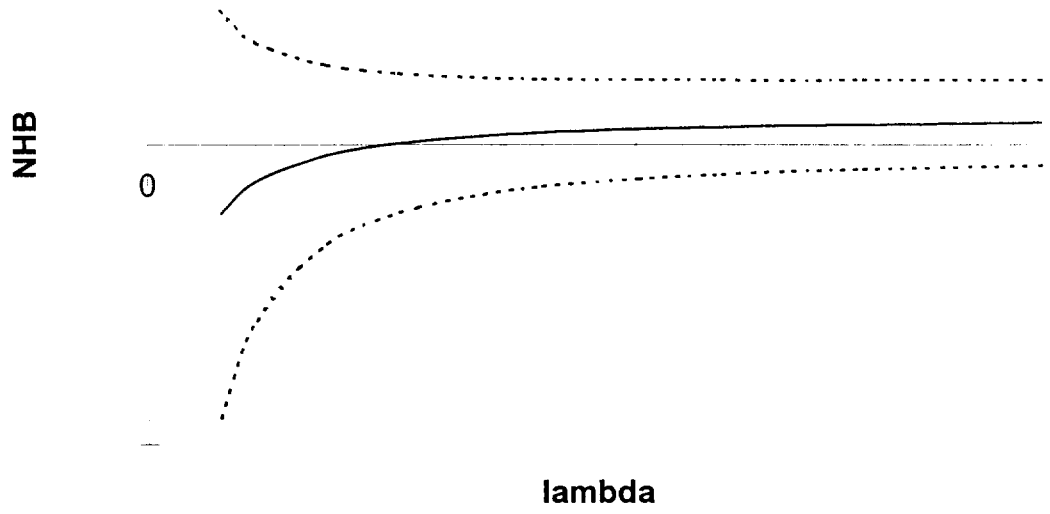

Fig. 4. Suggested format for presenting the results of an NHB analysis. NHB and the limits of a (1-α) CI are shown graphically as a function of the threshold C/E ratio, λ.

Acknowledgments

The authors thank Milton Weinstein, David Paltiel, Denny Fryback, and Emmett Keeler for helpful comments and suggestions.

References

[1] Manski CF. Analog Estimation Models in Econometrics. London: Chapman and Hall, 1988.

[2] Manning WG, Fryback DG, and Weinstein MC. Reflecting uncertainty in cost-effectiveness analysis. In: Gold MR, Siegel JE, Russell LB, and Weinstein MC, eds. Cost-Effectiveness in Health and Medicine. New York: Oxford University Press, 1996.

[3] O'Brien BJ, Drummond MF, Labelle RJ, Willan A. In search of power and significance: issues in the design and analysis of stochastic cost-effectiveness studies in health care. Medical Care 1994;32:150-163.

[4] Mullahy J, Manning WG. Statistical Issues in Cost-Effectiveness Analyses. In: Sloan F, ed. Valuing Health Care: Costs, Benefits, and Effectiveness of Pharmaceuticals and Other Medical Technologies. Cambridge: Cambridge University Press, 1995.

[5] Wakker P, Klaassen MP. Confidence intervals for cost/effectiveness ratios, Health Economics 1995;4:373-381.

[6] Poirier DJ. Intermediate Statistics and Econometrics. Cambridge, Massachusetts: MIT Press, 1995.

[7] Chaudhary MA, Stearns SC. Estimating confidence intervals for cost-effectiveness ratios: an example from a randomized trial. Statistics in Medicine 1996;.

[8] Kelly AE, Haddix AC, Scanlon KS, Helmick CG, Mulinare J. Cost-effectiveness of strategies to prevent neural tube defects. In: Gold MR, Siegel JE, Russell LB, and Weinstein MC, eds. Cost-Effectiveness in Health and Medicine. New York: Oxford University Press, 1996.

[9] Obenchain RL, Melfi CA, Croghan TW, and Buesching DP. Bootstrap analyses of cost-effectiveness in antidepressant pharmacotherapy. PharmacoEconomics 1997;11:464-472.

[10] Efron B, Tibshirani RJ. An Introduction to the Bootstrap. London: Chapman & Hall, 1993.

[11] Davidson R, MacKinnon JG. Estimation and Inference in Econometrics. New York: Oxford University Press, 1993.

[12] Mullahy J. What you don't know can't hurt you? Statistical issues in designing medical technology evaluation standards. Medical Care 1996;34:DS124-DS135.

[13] Stinnett AA. Adjusting for bias in C/E ratio estimates. Health Economics 1996;5:470-472.

[14] Stinnett AA, Paltiel AD. Estimating C/E ratios under second-order uncertainty: the mean ratio versus the ratio of means. Medical Decision Making 1997; forthcoming.

[15] Matchar DB, Ancukiewicz M, Lipscomb J, Parmagiani G, Samsa G. Accounting for uncertainty in cost-effectiveness analysis illustrated by the stroke prevention policy model (abstract). Medical Decision Making 1995;15:420.

[16] Siegel C, Laska E, Meisner M. Statistical methods for cost-effectiveness analyses. Controlled Clinical Trials 1996;17:387-406.

[17] Mullahy J. Which cost-effectiveness ratio? Using cost and effectiveness data to value medical technologies. Unpublished manuscript 1997.

[18] Johannesson M, Weinstein MC. On the decision rules of cost-effectiveness analysis. Journal of Health Economics 1993;12:459-467.

[19] Cantor SB. Cost-effectiveness analysis, extended dominance, and ethics: a quantitative assessment. Medical Decision Making 1994;14:259-265.

[20] Van Hout BA, Malwenn JA, Gordon GS, Rutten FFH. Costs, effects, and C/E ratios alongside a clinical trial. Health Economics 1994;3:309-319.

[21] Siegel JE, Weinstein MC, Torrance GW. Reporting cost-effectiveness studies and results. In: Gold MR, Siegel JE, Russell LB, and Weinstein MC, eds. Cost-Effectiveness in Health and Medicine. New York: Oxford University Press, 1996.

[22] Hirshleifer J, Riley JG. The Analytics of Uncertainty and Information. Cambridge: Cambridge University Press, 1992.

[23] Levy H. Stochastic dominance. In: Eatwell J, Milgate M, Newman P, eds. The New Palgrave: Utility and Probability. New York: W.W. Norton & Company, 1990.

[24] Bawa,VS. Stochastic dominance: a research bibliography. Management Science 1982;28:698-712 .

[25] Weinstein, MC. From cost-effectiveness ratios to resource allocation: where to draw the line? In: Sloan FA, ed. Valuing Health Care: Costs, Benefits, and Effectiveness of Pharmaceuticals and Other Medical Technologies. New York: Cambridge University Press, 1995.

[26] Van Hout BA, Malwenn JA, Gordon GS, and Rutten FFH. Costs, effects, and C/E ratios alongside a clinical trial. Health Economics 1994;3,309-319.

[27] Stinnett AA, Paltiel AD. Mathematical programming for the efficient allocation of health care resources. Journal of Health Economics 1996;15:641-653.

[28] Paltiel AD, Stinnett AA. Making health policy decisions: Is human instinct rational? Is rational choice human? Chance 1996;9(2), 34-39.

[29] Phelps CE, Mushlin AI. On the (near) equivalence of cost-effectiveness and cost-benefit analyses. International Journal of Technology Assessment in Health Care 1991;7, 12-21.

[30] Garber AM, Phelps CE. Economic foundations of cost-effectiveness analysis. Journal of Health Economics 1997;16:1-31.

[31] Meltzer D. Accounting for future costs in medical cost-effectiveness analysis. Journal of Health Economics 1997;16:33-64.

[32] Mishan EJ. Cost-Benefit Analysis. New York: Praeger, 1976.

[33] Pauly MV. Valuing health care benefits in monetary terms. In: Sloan FA, ed. Valuing Health Care: Costs, Benefits, and Effectiveness of Pharmaceuticals and Other Medical Technologies. New York: Cambridge University Press, 1995.