Net Health Benefits:
A New Framework for the Analysis of Uncertainty in Cost–Effectiveness Analysis

AARON A. STINNETT, PhD, JOHN MULLAHY, PhD

In recent years, considerable attention has been devoted to the development of statistical methods for the analysis of uncertainty in cost–effectiveness (CE) analysis, with a focus on situations in which the analyst has patient-level data on the costs and health effects of alternative interventions. To date, discussions have focused almost exclusively on addressing the practical challenges involved in estimating confidence intervals for CE ratios. However, the general approach of using confidence intervals to convey information about uncertainty around CE ratio estimates suffers from theoretical limitations that render it inappropriate in many situations. The authors present an alternative framework for analyzing uncertainty in the economic evaluation of health interventions (the “net health benefits” approach) that is more broadly applicable and that avoids some problems of prior methods. This approach offers several practical and theoretical advantages over the analysis of CE ratios, is straightforward to apply, and highlights some important principles in the theoretical underpinnings of CE analysis. Key words: cost–effectiveness; net health benefits; uncertainty. (Med Decis Making 1998;18 suppl:S68–S80)

In medical cost–effectiveness analysis (CEA), an incremental cost–effectiveness (CE) ratio comparing a new treatment (T1) with an alternative intervention (T0) is typically defined as:

\[ R = \frac{\mu_{c1} - \mu_{c0}}{\mu_{e1} - \mu_{e0}} \]  

(1)

where \( \mu_{c1} \) and \( \mu_{e1} \) represent the mean cost and mean health effect, respectively, of treatment T1. 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 T1 rather than T0.

Because the true (population) means are not known, R is estimated using the “analogy” estimator:

\[ \hat{R} = \frac{\bar{C} - \bar{C}_0}{\bar{E} - \bar{E}_0} \]  

(2)

where \( \bar{C} \) and \( \bar{E} \) represent sample means for the cost and effect of intervention T1. Due to uncertainty in these estimates, an important component of any CEA is an analysis of the uncertainty surrounding a CE 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. 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 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.

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 states. In the present paper, the term CEA is used to encompass both of these analytic frameworks.
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 R. As we will show, however, this approach suffers from important theoretical limitations; of particular concern is the fact that ambiguities in the probability distribution of a CE ratio estimator render inference based on that distribution (including the construction of CIs) problematic when there is significant uncertainty regarding the sign of an intervention's incremental cost and/or incremental effectiveness.

As an alternative to the estimation of CIs around R, we present a new framework for the analysis of uncertainty in economic evaluation. This method of analysis, which we call the "net health benefits" approach, offers several practical and theoretical advantages over the use of CIs for CE ratios, is straightforward to apply, and highlights some important principles in the theoretical underpinnings of CEA.

**Estimating Confidence Intervals for CE Ratios**

Several methods have been presented for estimating CIs around 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 R using a second-order Taylor series approximation. A two-tailed $(1 - \alpha)$ CI can then be constructed as:

$$\hat{R} \pm z_{\alpha/2} \sqrt{\hat{\sigma}_R^2}$$

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

**Theoretical Problems in the Use of Confidence Intervals for CE Ratios**

THE DECISION RULES OF CEA

To illustrate the problems associated with constructing and interpreting CIs for 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 with $T_0$. For each quadrant of the plane (notice the quadrant-numbering convention we
AMBIGUITY BETWEEN AE-AC QUADRANTS IN THE DISTRIBUTION OF A CE RATIO ESTIMATOR

Notice that \( \hat{R} \) is positive in quadrants I and III and negative in quadrants II and IV of the \( \Delta E-\Delta C \) plane. If a negative ratio corresponds to quadrant II, then \( T_1 > T_0 \) but if it corresponds to quadrant IV then \( T_0 > T_1 \). Similarly, a positive ratio less than \( \lambda \) 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 \( \Delta E-\Delta C \) 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 \( \Delta E-\Delta C \) plane, this distribution and any inference based on it—including the construction of CIs—are ambiguous.

PROBLEMS WITH NEGATIVE CE RATIOS

In some evaluations, the sign of either \( \Delta E \) or (more likely) \( \Delta C \) 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 CE 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 \( \Delta E-\Delta C \) plane; further suppose that there is non-negligible uncertainty with respect to the sign of \( \Delta E \), 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 CE 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 CE 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, the 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.
fecteds. 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 CE ratios (see, for example, Obenchain et al.9).

Net Health Benefits

DEFINITION AND INTERPRETATION

In response to the problems associated with inference based on the distribution of a CE 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 T1 as:

$$NHB = \mu_{T1} - \mu_{C0}/\lambda$$

and the incremental NHB of T1 compared with T0 as:

$$\Delta NHB = (\mu_{T1} - \mu_{C1}/\lambda) - (\mu_{T0} - \mu_{C0}/\lambda)$$

where all variables are as previously defined.

The first part of equation 4 is simply the health effect associated with intervention T1. The second part of the expression represents the health gain (measured in units of health) of investing resources in T1 rather than in T0; it is the incremental health gain. The incremental NHB of T1 relative to T0 has the minimum difference in health effects that society would demand in return for its investment of $\mu_{C0}$, which is equal to $\mu_{C0}/\lambda$. The incremental NHB of T1 relative to T0 compares the difference in the two programs' health effects ($\mu_{C0}$ - $\mu_{C0}$) with the minimum difference in health effects that society would demand in order to justify the additional expenditure required to implement T1 rather than T0; for an incremental cost of $\mu_{C0}$, the minimum acceptable health gain is equal to $\mu_{C0}/\lambda$.

In the remainder of this article, we 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. The concept of an average NHB has been introduced here because we find it useful pedagogically to illustrate that, just as incremental cost is the difference between two average costs and incremental effectiveness is the difference between two average effectiveness measures, so also is incremental NHB the difference between two average NHBs. In contrast, an incremental CE ratio is not equal to the difference between two average CE ratios. In a later section (on the analysis of stochastic dominance), we show that this difference between NHB and CE ratios has important practical impli-
cations for analyses with more than two comparators.

CONFIDENCE INTERVALS FOR NET HEALTH BENEFITS

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

$$\text{NHB} = \bar{E}_1 - \bar{E}_0 - (\bar{C}_1 - \bar{C}_0)/\lambda$$  \hspace{1cm} (6)

A parametric CI for NHB can be readily constructed. The variance of NHB is estimated as:

$$\sigma_{\text{NHB}}^2 = \sum_{i=1}^{n} \left( s_{E_i}^2 + s_{C_i}^2/\lambda^2 - 2r_i s_{E_i} s_{C_i}/\lambda \right) n_i$$  \hspace{1cm} (7)

where $E_i$ and $C_i$ are random variables representing the costs and effects of intervention $T_i$; $s_{E_i}^2$ and $s_{C_i}^2$ represent the sample variances 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 $\lambda$ is the threshold CE ratio. Notice 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 non-zero correlations between the costs and effects of an individual intervention.

By the central limit theorem, NHB is asymptotically normal, and a two-tailed $(1-\alpha)$ CI can be constructed as:

$$\text{NHB} \pm z_{\alpha/2} \sqrt{\sigma_{\text{NHB}}}$$  \hspace{1cm} (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 $\geq 30$.

Alternatively, for those who prefer not to rely on the asymptotic normality implied by the central limit theorem, a nonparametric CI for 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 replications 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 denote $\text{NHB}_b$ (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 ($\bar{C}_1$) and effectiveness ($\bar{E}_1$) for this bootstrap sample.

   b. From the observations for $T_0$, draw a random sample of size $n_0$ with replacement. Calculate the average cost ($\bar{C}_0$) and effectiveness ($\bar{E}_0$) for this bootstrap sample.

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

$$\text{NHB}_b = (\bar{E}_1 - \bar{E}_0) - (\bar{C}_1 - \bar{C}_0)/\lambda$$  \hspace{1cm} (9)

3. Eliminate the $B^{*(1-\alpha/2)}$ lowest values and the $B^{*(1-\alpha/2)}$ highest values of $\text{NHB}_b$, where $(1-\alpha)$ is the desired confidence level. The lowest remaining value and the highest remaining value for $\text{NHB}_b$ are the lower and upper bounds, respectively, of a nonparametric CI for NHB.

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: $\text{NHB}_{\text{total}} = N \times \text{NHB}$, where $N$ is the size of the target population. A $(1 - \alpha)$ CI for $\text{NHB}_{\text{total}}$ is simply $N$ times the $(1 - \alpha)$ CI for NHB. In contrast, the point estimate and CI for a CE ratio do not reflect the magnitude of a program’s effects, and there is no measure in CE ratio analysis that is analogous to the total NHB measure.

Net Health Benefits versus CE Ratios

INTERPRETABILITY

As discussed earlier, the interpretation of a CE ratio estimate is ambiguous without information regarding the quadrant of the $\Delta E - \Delta C$ 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 CE ratio conveys no useful information, inference that is dependent on the negative portion of $\hat{R}$’s probability distribution is problematic. In contrast, the NHB suffers from no such ambiguities. A positive (negative) value for the 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 the NHB, which is a linearly increasing function of effectiveness \( \frac{\partial \text{NHB}}{\partial \text{AE}} = 1 \) and a linearly decreasing function of costs \( \frac{\partial \text{NHB}}{\partial \text{AC}} = -1/\lambda \). In contrast, the direction of the change in cost–effectiveness ratio \( R \) associated with a change in \( \Delta C \) depends on the sign of \( \Delta E \) \( \frac{\partial R}{\partial \Delta C} = 1/\Delta E \), and the direction of the change in \( R \) associated with a change in \( \Delta E \) depends inversely on the sign of \( \Delta C \) \( \frac{\partial R}{\partial \Delta E} = -\Delta C/(\Delta E)^2 \).

**IMPLICATIONS FOR STATISTICAL INFERENCE**

Largely because the NHB is linear in costs and effects, statistical inference is far more straightforward when using the NHB than when using CE ratios.\(^{12}\) For example, while the assumption of a well-behaved parametric distribution for \( \hat{R} \) is questionable, 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 \( \hat{R} \), the same 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 NHB is an unbiased estimate of the true NHB. In contrast, the sample estimate \( \hat{R} \) is a biased estimate of the true CE 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 CE 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 NHB evaluated at the mean estimates of effects and costs. In contrast, the mean of a distribution of CE ratio estimates is not generally equal to the ratio of the mean estimate of incremental cost to the mean estimate of incremental effectiveness, 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 CE ratio under conditions of uncertainty.\(^{14-17} \)

Note also that a bootstrap estimate of the variance of NHB would be expected to converge more quickly than similar estimates for CE ratios. Cost–effectiveness ratios take on values approaching infinity (or negative infinity) when incremental effectiveness is close to zero, and the possibility of CE 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 CE 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 investigators have discussed stochastic methods for analyzing uncertainty in CEA, these discussions have been limited to situations in which two interventions are being compared with 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 CE ratios, the presence of multiple comparators introduces significant complexity. In CEA, incremental CE 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 becomes a possibility only when more than two programs are being compared).\(^{18,19} \) After dominated programs have been removed from consideration, the undominated programs are ranked from lowest to highest cost, and an incremental CE ratio is calculated for each of the undominated programs on this rank list (except for the least costly of the undominated alternatives, for which no CE ratio is calculated); a program’s incremental CE ratio is calculated relative to the program immediately preceding it on the list. For a more detailed description of this CEA algorithm, see Johannesson and Weinstein\(^ {18} \) or the Comparative Benefits Modeling Project.\(^ {20} \)

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 $T_i$), 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 CE ratio is calculated for that intervention. For those replications where the program is not dominated and does not dominate all other alternatives, the incremental CE ratio of that intervention may be calculated relative to one program ($T_{i0}$) in some cases and relative to another program ($T_{i0}, T_{i1}, \ldots$) in other cases, due to uncertainty in the ordering of the rank list of undominated programs.

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 $\lambda$, as one indication of a program's attractiveness. With multiple comparators, interpretation is less straightforward. For example, one might estimate the probability, conditional on a value of $\lambda$, that program $T_i$ is either dominant over all other alternatives or has an incremental CE ratio more favorable than $\lambda$. However, using this probability as an indication of the program’s attractiveness would not take into account the fact that one or more of $T_i$’s comparators might also have a CE ratio less than $\lambda$; if one of the $T_i$’s comparators (call it $T_{i0}$) has a CE ratio less than $\lambda$ and is also more effective than $T_i$, then the theory of efficient resource allocation indicates that $T_{i0}$ should be selected for implementation rather than $T_i$. 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 equation 4. Analysts may find this to be considerably simpler than what would be required to calculate the same probability using methods based on incremental CE ratios. Using the NHB method, 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 with 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 CE ratios.

The advantage of NHBs over CE ratios in multiple-comparator analyses is attributable in large part to the fact that NHBs are separable while incremental CE ratios are not separable. That is, the incremental NHB of $T_i$ compared with $T_{i0}$ 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 CE ratio is not a separable function of average CE ratios ($\mu_{CE}/\mu_{T_{i0}}$), and calculating interventions’ average CE 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.

STOCHASTIC DOMINANCE

Another result of the separability of incremental NHBs vis-à-vis the inseparability of incremental CE ratios is that the concept of stochastic dominance is relevant in the analysis of NHBs but not in the analysis of CE ratios. Stochastic dominance (which should not be confused with the concepts of strong dominance and extended dominance discussed earlier) is a concept that is useful for making comparisons across mutually exclusive choices under conditions of uncertainty. The analysis of stochastic dominance is complementary to the other types of analysis (such as CI estimation and hypothesis testing) that have been discussed in earlier sections.

†† In a strictly classical statistical framework, it is meaningless to discuss the probability of an intervention's being dominated, because the quantities of interest are fixed, non-random parameters. Bayesian statistics allows one to discuss probability distributions for parameters and functions of parameters. The probabilities discussed here would correspond to Bayesian posterior probabilities under the assumption of improper or noninformative prior probability distributions.
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:

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

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 ($\leq$) must be a strict inequality ($<$) for at least one value of the 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 the 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 ($\frac{dU}{dNHB} > 0$) should prefer $T_1$ to $T_0$. This result holds regardless of the decision maker's risk posture.

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:

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

where all variables are as previously defined, and where the weak inequality ($\leq$) 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 $F$ (dotted line) lies below $G$ (solid line) is greater than the area between the two curves where $F$ lies above $G$. Note that the two curves may cross more than once. 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 ($\frac{dU}{dNHB} > 0$ and $\frac{d^2U}{dNHB^2} < 0$) should prefer $T_1$ to $T_0$.

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 CE ratios, because preferences are not monotonically increasing in R and monotonicity is a necessary condition for stochastic dominance.

### Net Health Benefits and Fieller Limits for CE Ratios

Readers familiar with the recent literature on methods for estimating CIs for CE ratios may notice a surface similarity between NHB analysis and the
estimation of "Fieller intervals" for CE ratios.23,29 The common link between Fieller's method and parametric NHB analysis is that both take advantage of the asymptotic (multivariate) normality of ΔC and ΔE and make use of a linear combination of these quantities that is also asymptotically normal. Fieller's method is based on the normality of ΔC − R*ΔE, and parametric NHB analysis is based on the normality of ΔE − ΔC/λ. However, that is where the similarity between the two approaches ends. Unlike NHB analysis, in which nonparametric bootstrapping may be used, Fieller's method is inextricably bound to the assumption of multivariate normality, and we are unaware of any formal investigation into the robustness of Fieller's method with regard to departures from normality. In addition, Fieller's method sometimes produces imaginary results (i.e., results that are a multiple of \( \sqrt{-1} \)); indeed, one of the two applications of this method reported by Willan and O'Brien29 produced imaginary results. NHB analysis, in contrast, does not suffer from this type of problem.

Moreover, no method has been presented for applying Fieller's method in studies with multiple comparators, and it is not at all clear that the method could be extended for application in such studies. It should also be noted that Fieller's method can, and sometimes does, result in negative confidence limits for a CE ratio. In some cases, the negative limit corresponds to the intervention being dominated, and in some cases it corresponds to the intervention being dominant; in neither case has any meaningful interpretation been offered for the magnitude of the negative confidence limit. Willan and O'Brien29 suggest handling the problem of negative limits by constructing one-sided intervals for CE ratios when one of the limits is estimated to be negative. In NHB analysis on the other hand, both the sign and the magnitude of the results always have meaningful interpretations. Based on these contrasting properties, we conclude that the differences between NHB analysis and Fieller's method outweigh the surface similarity between the two methods.

### Net Health Benefits and the Threshold CE Ratio

**REPORTING NET HEALTH BENEFITS AS A FUNCTION OF \( \lambda \)**

One issue of concern in the use of NHBs is the fact that the societal threshold CE ratio (\( \lambda \)) is not known. Indeed, one might argue that uncertainty around \( \lambda \) 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 \( \lambda \) and report NHB as a function of \( \lambda \). 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\( _{\lambda} \) to indicate the value of \( \lambda \) corresponding to the estimate. A useful tool for presenting the results of NHB analyses and the sensitivity of these results to \( \lambda \) would be to plot both point NHB estimates and their CIs as a function of \( \lambda \); an example of such a plot is shown in figure 4. Using such a figure, a decision maker may assess both the point estimates of programs' net benefits (measured in terms of health) and the uncertainty surrounding those estimates, over whatever range of \( \lambda \) that decision maker considers relevant.

Analysis may also make inferences regarding the probability that an intervention is estimated to be cost-effective relative to its comparator (NHB\( _{\lambda} > 0 \)). For nonparametric estimation, the probability that NHB\( _{\lambda} \) is positive is calculated as the percentage of bootstrap replicate estimates that are positive, conditional on a given value of \( \lambda \). This probability can then be plotted as a function of \( \lambda \), producing a graph identical to the CE acceptability curve presented by van Hout et al.31 In addition, under the assumption that NHB is normal, one can use standard parametric techniques to perform the following one-sided hypothesis test of cost-effectiveness for any value of \( \lambda \): H\( _{0} \): NHB\( _{\lambda} = 0 \), versus H\( _{1} \): NHB\( _{\lambda} > 0 \). If the null hypothesis is not rejected at the (1 − \( \alpha \)) confidence level, this means there would be a probability of at least \( \alpha \) of observing a value as large as NHB\( _{\lambda} \), even if the null hypothesis of NHB = 0 were true. In choosing what weight to place on the results of such a hypothesis test, a decision maker should...
give consideration to such issues as the power of the trial to detect significant differences in NHB and the value of \( \lambda \) used in performing the hypothesis test.

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

THE THRESHOLD CE RATIO AND OPPORTUNITY COSTS

Another attractive property of the NHB approach and its explicit consideration of \( \lambda \) is that it forces decision makers to confront the issue of opportunity costs. In CEA, a program with a CE 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 CE ratios of \( R = \lambda \) that are currently being forgone. 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). While this information is implicit in the use of CE ratios, NHBs offer the advantage of making the opportunity costs explicit.

TREATING \( \lambda \) AS A RANDOM VARIABLE

In this paper, we have suggested that NHB estimates be calculated conditional on a value of \( \lambda \), with sensitivity analyses performed to report NHB as a function of \( \lambda \). However, when considering how to reflect uncertainty around \( \lambda \), one's first instinct might be to suggest estimating a probability distribution for \( \lambda \) 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 \( \lambda \), and indeed the nature of \( \lambda \) itself.

Note that, in contrast to the "micro-level" clinical variables corresponding to individual interventions' respective costs and effects, \( \lambda \) 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 \( \lambda \).

One approach might be to observe decisions about which health interventions are implemented and which are not, and to draw inferences about \( \lambda \) based on the relationships between programs' estimated CE 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 \( \lambda \) 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 CE 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 \( \lambda \) 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. Thus, the degree to which one can make inferences about \( \lambda \) on the basis of observing actual resource allocation decisions is unclear.

Other possible approaches to estimating a probability distribution for \( \lambda \) would be to elicit the "expert judgments" of various individuals believed to have some information about \( \lambda \) in order to formulate some "prior" distribution of \( \lambda \), or to undertake contingent valuation studies to estimate people's willingness to pay per unit of improved health.

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

THE THRESHOLD CE RATIO AND OPPORTUNITY COSTS

Another attractive property of the NHB approach and its explicit consideration of \( \lambda \) is that it forces decision makers to confront the issue of opportunity costs. In CEA, a program with a CE 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 CE ratios of \( R = \lambda \) that are currently being forgone. 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).
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 $\lambda$ (as we suggest in this paper), at least until some consensus emerges regarding a distribution for $\lambda$.

In assessing the merits of performing NHB analyses conditional on $\lambda$, one issue to consider is whether or not it is reasonable to treat $\lambda$ as being statistically independent of the costs and effects observed in a particular study. Formally, if $\lambda$ is statistically independent of $\Delta C$ and $\Delta E$, 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 $\lambda$ as being statistically independent of $\Delta C$ and $\Delta E$ from a particular study—i.e., if it is reasonable to view $\lambda$ as being decision maker-specific rather than intervention-specific—this implies that it is reasonable to estimate net health benefits conditional on $\lambda$, rather than directly incorporating into the analysis an estimate of $\phi_1(\lambda)$. Moreover, if $\phi_2(\lambda)$ is not dependent on $\phi_1(\Delta C, \Delta E)$, this indicates that estimates of $\phi_2(\lambda)$ 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 $\lambda$ 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 $\lambda$ 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 $\lambda$ is preferable to the alternative of directly incorporating an estimated or postulated distribution for $\lambda$. If one does wish to estimate or postulate a distribution for $\lambda$ and incorporate it into an estimate of the distribution of an intervention's NHB, this should be done in a supplementary analysis, where the primary analysis reports NHB as a function of $\lambda$, 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 the NHB is measured in units of health rather than money, and that the NHB method as presented here assumes that society's willingness to pay per unit of health benefit is constant across disease categories, patient populations, and therapeutic classes, whereas CBA prescribes that analyses be based on willingness-to-pay values elicited for valuation of the precise program being evaluated.\(^{\dagger\dagger}\) 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).\(^{34}\)

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 principal (or assumption) that all QALYs are val-

\(^{\dagger\dagger}\)It would be feasible, however, for analysts to perform "NHB-type analyses" in which this assumption is relaxed or dropped by using program-specific willingness-to-pay estimates. In this case, the analysis would be exactly identical to CBA except that the results would be reported in units of health rather than money.
ised equally from the societal perspective—reflected by the use of the same value or range of values for λ in studies relating to different patient populations, different disease categories, and different therapeutic classes—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. It is for this reason of priority-signaling that we have chosen to use a summary measure denominated in units of health rather than money.

The above points notwithstanding, a strong case can still be made for using CBA. Whereas the theoretical foundations of CEA are still being debated, CBA is firmly grounded in the theory of welfare economics. 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. The NHB approach does not attempt to resolve this issue, which is beyond the scope of the present paper. Although the 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 CE ratio. However, because CE 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 CE ratios is problematic. In particular, the NHB approach may be most helpful when analyzing the uncertainty around a CE 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 (so long as the sample size is not forbid-

The authors thank Milton Weinstein, David Paltiel, Dennis Fryback, Emmett Keeler, and two anonymous referees for helpful comments and suggestions.

References


