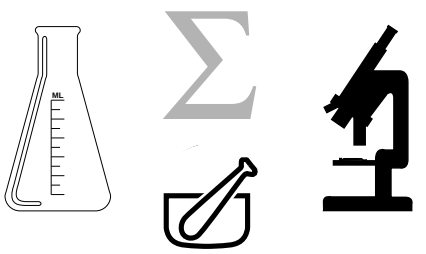
Biopharmaceutical Section



American Statistical Association

Biopharmaceutical Report

Volume 5, No. 2 Summer 1997

Chair: ***Gary L. Neidert*** **Co-Editors:** ***Curtis Wiltse and Bill Huster***

Issues and Algorithms in Cost-Effectiveness Inference

Robert L. Obenchain
Eli Lilly & Company

Introduction

Recent efforts to control health care expenditures in many countries around the world have tended to focus attention on the cost-effectiveness of alternative treatment methods, including pharmacotherapy. There appears to be an emerging consensus that, in addition to traditional randomized clinical trials, new forms of study design and analysis methodology are needed to address the wide array of questions now being asked by regulators, payers, health care providers, and patients. Drummond (1992) concluded "Maintenance of good methodological standards is, in the long run, the best policy both for pharmaceutical industry sponsors and economic analysts."

A number of organizations are in the process of developing guidelines for pharmacoeconomic practice [PhRMA (1994, 1995), A*P*O*R (1996), PCEHM (1996); see also the review by Genduso and Kotsanos (1996)]. As statisticians, our interests in these sorts of guidelines tend to focus on methods of accounting for uncertainty in cost-effectiveness inference. A wide variety of methodological advances have already appeared in published literature; see Drummond, Heyse, and Cook (1996) as well as the list of references at the end of this article. In reality, little is currently known about the relative advantages and disadvantages of these alternative approaches to statistical inference.

A primary purpose of this article is to encourage interested biopharmaceutical statisticians and econometricians from industry, academia, and government to actively participate in a new, standing Cost-Effectiveness Inference committee, cosponsored by ASA, PhRMA, and A*P*O*R. This working group is expected to have a technical focus, with initial emphasis limited to reviewing inference methods for the special case where just two alternative therapies are being compared using an incremental cost-effectiveness ratio (*ICER*) statistic. Thus, the remainder of this article introduces a few of the many unresolved issues surrounding *ICER* analyses in pharmacoeconomics.

As well as performing research on the advantages and disadvantages of alternative methods, our committee will also collect and distribute computational algorithms for cost-effectiveness inference. Current pharmacoeconomic guidelines stress "full disclosure" of methods and data, and distribution of algorithms (source code) would greatly expedite this disclosure process.

Committee members will cross-validate computer algorithms primarily by verifying that software (appropriately compiled or interpreted) produces correct output for a suite of benchmark numerical examples. Source code, test data, and executable modules for the algorithms we have reviewed will be posted to StatLib (http://lib.stat.cmu.edu/DOS/general) on the Internet so that they may be downloaded by any interested party. Researches will then have the options either to use the algorithms we have evaluated or else to cross-validate their own algorithms against our standards.

Incremental Cost-Effectiveness Ratios (ICERs)

In a two-sample cost-effectiveness analysis, the data consist of a (continuous) cost variable, C_{Ti} , and a treatment effectiveness indicator, E_{Ti} (which may be binary, with 0→no, 1→yes or continuous), for each of the $i = 1, \ldots, NT$ patients who received the new treatment, *T*. Similarly, a pair of (C_{Sj}, E_{Sj}) values would be collected for each of the $j = 1, \ldots, NS$ patients who received the standard treatment, *S*.

The "incremental" cost-effectiveness ratio, *ICER*, Black (1990), is the ratio defined as the difference in average per patient costs divided by the corresponding difference in effectiveness average.

Contents

$$ICER = \frac{\overline{C}_T - \overline{C}_S}{(\overline{E}_T - \overline{E}_S)} \ ,$$

where a T subscript denotes an average over patients on the new treatment while subscript S denotes the corresponding average for patients on the standard treatment.

Transformations of ICER statistics in the form of simple "scale changes" are sometimes needed. For example, this occurs in converting a numerator cost difference from one currency into another or in discounting charges relative to a different base year. Similarly, when one's effectiveness measure is binary (0 => ineffective, 1 => effective), one might re-express the denominator in **percentage points** as follows:

$$ICER = \frac{\overline{C}_T - \overline{C}_S}{100 \times (\overline{E}_T - \overline{E}_S)} \ .$$

Unfortunately, some inference methods are sensitive to these sorts of simple scale changes.

Two very different types of methodology for placing statistical confidence limits around Incremental Cost-Effectiveness Ratios (ICERs) appear to currently be in active use. These two approaches are (i) parametric methods for analysis of ratio estimates, including Fieller's theorem; and (ii) nonparametric, bootstrap methods**.**

Fieller's Theorem

Like older methods based upon a Taylor series approximation, the Fieller's theorem approach recognizes that the ICER is a "ratio estimator" in the sense of Cochran (1977) and, thus, is asymptotically normally distributed. The characteristic property of the Fieller approach is that it treats the numerator and denominator between cohort differences as if they were a pair of correlated normal variables. This allows the Fieller approach to recognize (small sample) situations where the stochastic distribution of the ICER is actually highly skewed.

Technical Note: Descriptions of the Fieller and Taylor series approaches are given in Willan and O'Brien (1994, 1996), O'Brien et al. (1994), Sacristan et al. (1995) and Chaudhary and Stearns (1996).

ISSUE: Confidence intervals based upon Taylor series approximations are too narrow (anti-conservative) relative to the corresponding intervals from Fieller's theorem.

ISSUE: Fieller's theorem confidence intervals correspond to "bow tie" shaped confidence regions on the cost-effectiveness plane; see Figure 3. It follows that Fieller's theorem confidence intervals are themselves "too narrow" when the estimated mean of the joint

distribution of between cohort average differences, $\left(\overline{E}_T - \overline{E}_S , \overline{C}_T - \overline{C}_S\right)$, is not highly significantly different from (0,0).

ISSUE: The Fieller method of forming ICER confidence intervals is not "rescaling commutative." In other words, rescaling an ICER statistic by a multiplicative factor changes its upper and lower Fieller confidence limits by a different factor.

Bootstrap Analyses

Bootstrapping approaches resample (with replacement) from all of the observed data, as described in Drummond, Heyse and Cook (1996), Chaudhary and Stearns (1996), and Obenchain et al. (1997).

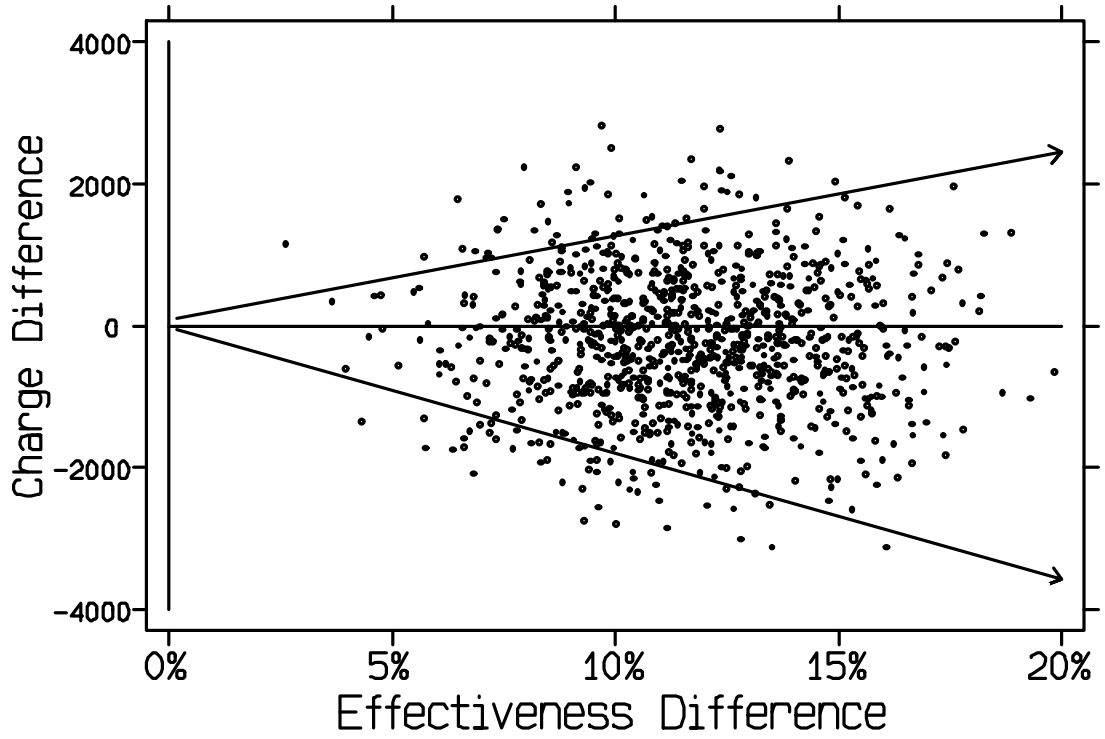
ISSUE: The observed data pairs, (C_{Ti}, E_{Ti}) or (C_{Sj}, E_{Sj}), for all patients in both treatment groups are needed to construct bootstrap confidence intervals. In other words, bootstrap intervals cannot be computed from the sorts of simple summary statistics (sample means, variances, and correlations) commonly reported in health economics studies and journal articles. On the other hand, bootstrap approaches to ICER confidence intervals do not need to make possibly unrealistic assumptions about parametric forms for stochastic distributions and, thus, offer great potential for increased realism, accuracy and robustness.

ISSUE: The bootstrap approach yields a rather dramatic graphical display of the variability in two-sample cost and effectiveness differences that result when an entire study is literally "redone" hundreds of times; see Figure 1, reproduced from Obenchain et al. (1997). Bootstrap analyses are thus actually much easier to explain and to appreciate than are the rather elaborate calculations and approximations used in parametric, "ratio estimator" approaches.

ISSUE: Like all methods based upon simulation or re-sampling, numerical values for bootstrap confidence limits can be sensitive to "parameters" such as the total number of replications performed and the initial seed value for the random number generator. Thus, to satisfy pharmacoeconomic "full disclosure" guidelines, these sorts of technical details usually need to be reported. Furthermore, "sensitivity" analyses should be performed to assure that bootstrap limits are not reported with inappropriate precision (too many decimal places.)

ISSUE: No implementation of ICER bootstrap analysis is currently available (mid 1997) in commercial statistical analysis software. On the other hand, algorithms needed to perform bootstrap analyses are all either straightforward or else published in statistical literature; see Efron and Gong (1983), Efron and Tibshirani (1986), L'Ecuyer (1988), O'Brien et al. (1994), and Westfall and Young (1992) for basic concepts. Obenchain (1997) provides examples of highly portable algorithms for ICER confidence limits using bootstrapping or Fieller's theorem.

Figure 1. A bootstrap analysis of ICER slopes.

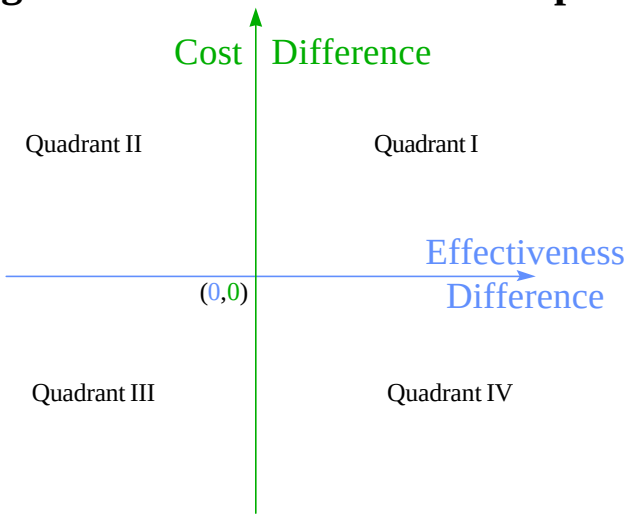


Technical Note: In addition to ICER rescaling commutivity, Chaudhary and Stearns (1996) discuss the closely related property of "transformation-respecting" methods. Bootstrap procedures yield confidence interval endpoints that change correctly and automatically under general monotone transformations as well as under simple rescalings.

ICER Interpretation

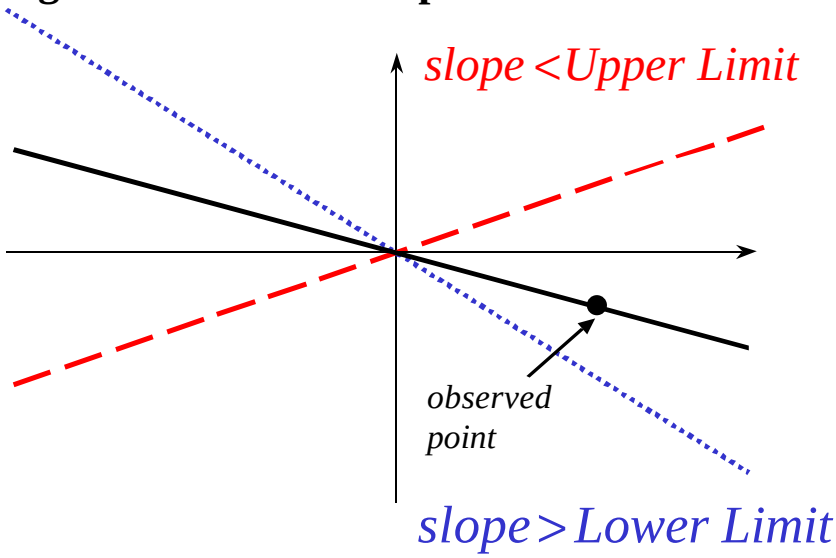
A health economics study comparing a new treatment, T, with a standard treatment, S, can be viewed as producing a single "point" on the **cost-effectiveness** plane of Black (1990), shown in Figure 2. The horizontal coordinate of this point is a between-cohort effectiveness difference, $\overline{E}_T - \overline{E}_S$, while the vertical coordinate is a measure of the corresponding cost difference, $\overline{C}_T - \overline{C}_S$. Here, we number the quadrants of the cost-effectiveness plane (I, II, III and IV) in the "standard" way, also shown in Figure 2.

Figure 2. The cost-effectiveness plane.



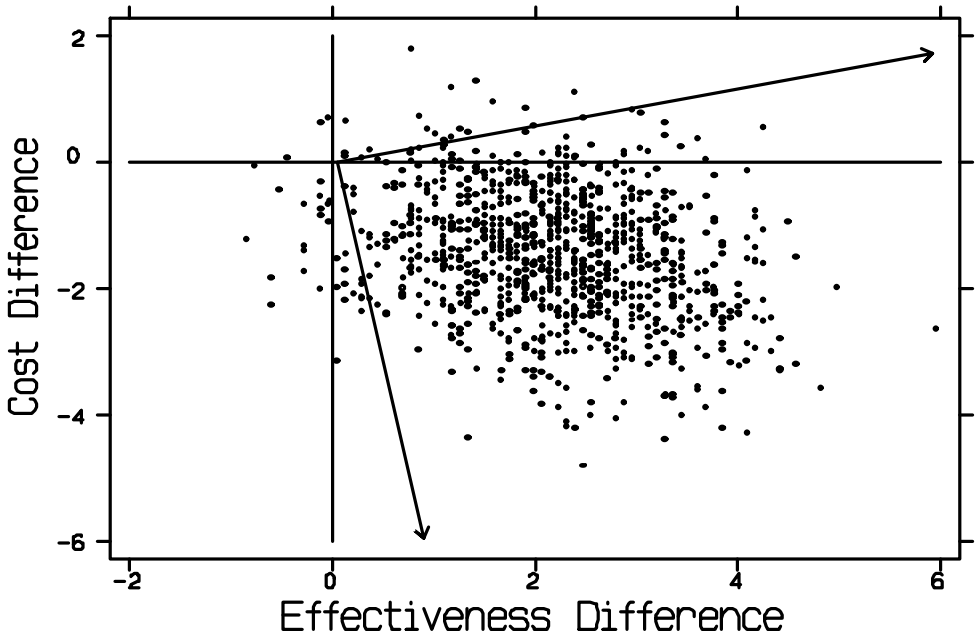
On this graphical display, the ICER is nothing more than the **slope of the line** connecting that health economics study point, $(\Delta E, \Delta C) = (\overline{E}_T - \overline{E}_S, \overline{C}_T - \overline{C}_S)$, with the origin, (0,0). And the corresponding Fieller confidence limits on this ICER slope form a "bow tie" shaped region. These concepts are illustrated in Figure 3.

Figure 3. An ICER slope and its Fieller limits.



We do not always find ourselves in the **simple situation** (depicted in Figure 1) where all of the results generated in an ICER bootstrap analysis fall within Quadrants I and IV of the cost-effectiveness plane. In other words, some bootstrap effectiveness differences, $\Delta E = \overline{E}_T - \overline{E}_S$, may turn out to be **negative,** rather than all positive. But we still want a cost-effectiveness confidence region that is "wedge" shaped, as in Figure 4, rather than "bow tie" shaped.

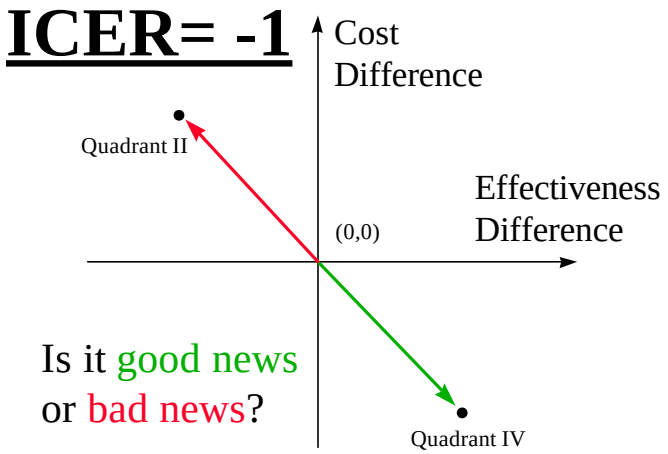
Figure 4. A bootstrap analysis with points in all 4 quadrants.



While the vast majority of bootstrap replicates in Figure 4 fall into Quadrant IV (900 of 1000) or Quadrant I (77 of 1000), a few outcomes fall into Quadrant III (20 of 1000) and even into Quadrant II (3 of 1000). In particular, notice also that negative values for the ICER slope arise from outcomes in either Quadrant IV or Quadrant II. But these two types of outcomes have diametrically opposite interpretations!

Specifically, consider the two health economic studies depicted in Figure 5, where the numerical value of the **ICER slope** is minus one, say, in both cases.

Figure 5. The ICER slope tells only half of the story!

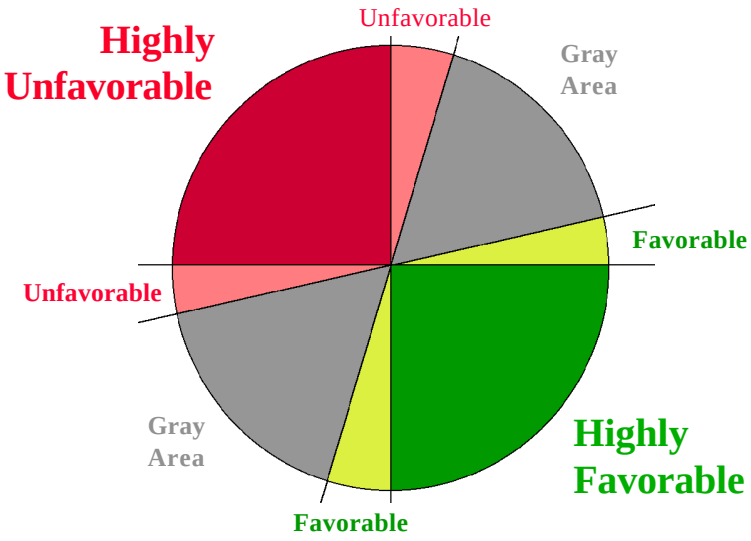


Note that a health economics study that produces a point in Quadrant IV suggests that the new treatment, T, completely dominates the standard treatment, S. After all, $\overline{E}_T - \overline{E}_S > 0$ and $\overline{C}_T - \overline{C}_S < 0$ together mean that T is both more effective and less costly than S. In sharp contrast, S completely dominates T when a health economics study produces a point in Quadrant II. This time $\overline{E}_T - \overline{E}_S < 0$ and $\overline{C}_T - \overline{C}_S > 0$, meaning that T is now less effective and more costly than S. And yet both of the hypothetical studies depicted in Figure 5 yielded the same numerical value (-1) for the ICER slope!

ISSUE: In some situations, the **ICER slope** is not, by itself, a **sufficient** statistic for making cost-effectiveness inferences.

Figure 6 divides the cost-effectiveness plane into Five Sections that span the full range of possible outcomes of health economic studies comparing a new treatment, T, with a standard treatment, S. For example, Quadrant IV is labeled "Highly Favorable," while Quadrant II is labeled "Highly Unfavorable." Again, these are the two quadrants of the cost-effectiveness plane where the ICER is negative.

Figure 6. Dividing up the cost-effectiveness plane.



The ICER slope is positive in Quadrants I and III, and these quadrants are divided into 3 parts each. A small, positive ICER slope is "Favorable" if the health economics study point falls in Quadrant I but "Unfavorable" when that point falls in Quadrant III. Similarly, a large, positive ICER slope is "Favorable" if the health economics study point falls in Quadrant III but "Unfavorable" when that point falls in Quadrant I. The fifth section is a "Gray Area" within Quadrants I and III where the ICER slope is positive but neither very big nor very small.

ISSUE: An unambiguous way to quantify cost-effectiveness, using **ICER angles**, is our next topic.

ICER Angles

The **scales** used along the horizontal (effectiveness-difference) and vertical (cost-difference) axes of the cost-effectiveness plane need to be **standardized** in order to define meaningful cost-effectiveness **angles,** Heyse and Cook (1992). One reasonable standardization is achieved by dividing each difference in treatment averages by the estimated standard deviation of a treatment difference between individual patients. In other words, we define standardized effectiveness = **x** and cost = **y** coordinates as follows:

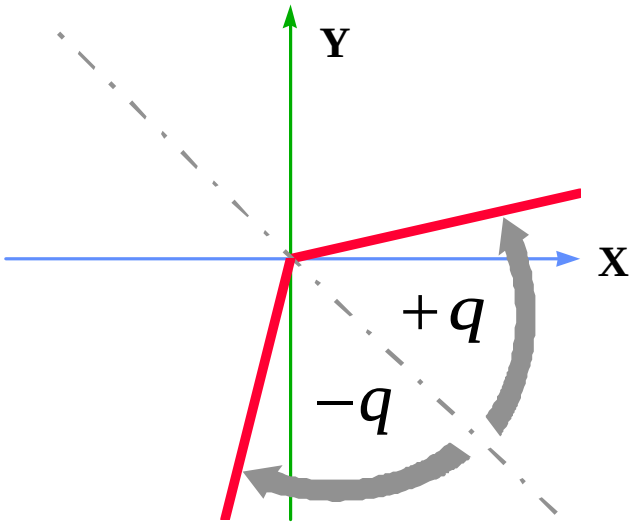
$$x = \frac{\left(\overline{E}_T - \overline{E}_S\right)}{\sqrt{Var\left(E_{Ti}\right) + Var\left(E_{Sj}\right)}} \quad \text{and} \quad y = \frac{\left(\overline{C}_T - \overline{C}_S\right)}{\sqrt{Var\left(C_{Ti}\right) + Var\left(C_{Sj}\right)}} \ .$$

Note, specifically, that the standardized **x** coordinate above is **unchanged** no matter what scaling (percentages, fractions, etc.) is used to measure effectiveness. Similarly, the standardized **y** coordinate above is **unchanged** no matter what monetary unit (dollars, yen, etc.) or base year is used to measure costs or charges. This type of scaling was used in Figure 4, above.

Technical Note: Standardized effectiveness = **x** and cost = **y** coordinates could also be defined using the estimated standard deviations of differences in treatment averages, $std.dev.\left(\overline{E}_T - \overline{E}_S\right)$ and $std.dev.\left(\overline{C}_T - \overline{C}_S\right)$, as their denominators. This alternative is easier to define and explain to laymen and would lead to identical ICER angles, at least when treatment cohort sample sizes are equal, NT=NS. Unfortunately, this alternative scaling also means that the resulting **x** and **y** coordinates would be expected to "grow" in size at a rate proportional to the square root of the sample size if additional patients were added to a study.

Returning to Figure 6, note the symmetry of each of the different types of cost-effectiveness region about the standardized -45° line, $x + y = 0$. Thus, one can define **Contours of Constant Cost-Effectiveness** (of the new treatment **T** relative to the standard treatment **S**) as in Figure 7. Note that each such contour consists of a pair of line segments, joined at (x,y)=(0,0), and making equal angles, $\pm q$, with the standardized -45° line, $x + y = 0$. The angles associated with the two segments of this contour are $+q$ and $-q$, respectively.

Figure 7. A contour of constant cost-effectiveness.



Relative to this same x + y = 0 line, the **ICER angle** for a standardized (**x,y**) point would be defined on $0^o \le |q| \le 180^o$ by…

$$|q| = \arctan\left(|x+y|/(x-y)\right)$$
$$when\ \ x \neq y$$
$$= 90^\circ\ \ when\ \ x = y.$$

Technical Note: In the above notation, the ICER slope can be written as

$$s = y/x = \tan\left(q - 45^o\right).$$

Thus the ICER slope, s, is easily expressed as a function of the ICER angle, q, but q is not a one-to-one function of s alone because s is not a sufficient statistic. Notice also that $\tan\left(-q - 45^o\right) = 1/\tan\left(q - 45^o\right)$, so that the ICER slopes associated with ICER angles of q and $-q$ are reciprocals of each other.

Table 1 lists proposed terminology for various ranges of values for ICER angles, ICER slopes, and the corresponding quadrants of the cost-effectiveness plane.

Table 1. Does a health economic study favor T over S?

Description	ICER Angle	ICER Slope	Cost-Effectiveness Quadrant
Highly Favorable	$0 \le \lvert q \rvert < 45^\circ$	Negative	IV
Favorable	$45^\circ \le \lvert q \rvert < 60^\circ$	Positive (extreme)	I or III
Mixed ("Gray Area")	$60^\circ \le \lvert q \rvert \le 120^\circ$	Positive (neither very large nor very small)	I or III
Unfavorable	$120^\circ < \lvert q \rvert \le 135^\circ$	Positive (extreme)	I or III
Highly Unfavorable	$135^\circ < \lvert q \rvert \le 180^\circ$	Negative	II

ISSUE: The 60^o and 120^o values proposed in Table 1 as boundaries between the "Favorable", "Mixed," and "Unfavorable" sections are really somewhat arbitrary. Values of the form $45^o + A^o$ and $135^o - A^o$ could just as easily have been used with, say, $A^o = 5^o$, 10^o or 20^o instead of $A^o = 15^o$. In fact, the numerical value considered most appropriate for A^o might vary between therapeutic areas.

On the other hand, taking $A^o = 15^o$ does allow the mixed ("Gray Area") regions to occupy exactly $1/3^{rd}$ of the total cost-effectiveness plane, leaving $1/3^{rd}$ either favorable ($1/12^{th}$) or highly favorable ($1/4^{th}$) as well as $1/3^{rd}$ either unfavorable ($1/12^{th}$) or highly unfavorable ($1/4^{th}$).

ICER Angle Bootstrap Confidence Regions

Definition One: The bootstrap $100(1-\alpha)\%$ confidence region for cost-effectiveness is the wedge-shaped region subtending the smallest total angle at the origin and yet containing $100(1-\alpha)\%$ of the simulated cost-effectiveness pairs.

Note, however, that this minimum subtended angle may be quite large (greater than 180^o or even 270^o) when a cost-effectiveness study provides only very "weak" information.

Technical Note: It is essential to measure the angle subtended between pairs of ICER angle order statistics in a consistent way [say, always clockwise.] For example, with 1000 bootstrap replicates (numbered 0 to 999) and a 95% confidence level, the subtended angle between order statistic 999 ($+173^o$) and order statistic 49 (-6^o) would be $173^o - (-6^o) = 179^o$. Similarly, the

subtended angle between order statistic 500 (+10°) and order statistic 550 (+20°) would be +350° rather than 10° - 20° = -10°.

Technical Note: Figure 4 displays a 95% confidence, bootstrap ICER interval of this minimum-subtended-angle form. The ICER angle point estimate from this study was 10.69°, while 1000 bootstrap replications produced ICER angles ranging from -144.83° to +145.60°, with mean=12.05° and median=11.92°. The minimum subtended angle was 96.49° and occurred between order statistic 981 (+60.68°) and order statistic 31 (-35.80°) out of 1000. This 95% confidence region thus lies almost entirely within the "highly favorable" and "favorable" sectors of the cost-effectiveness plane; only the last 0.68° laps over into the Quadrant I "Gray Area."

Definition Two: The bootstrap "central" $100(1-\alpha)\%$ confidence region for cost-effectiveness is the wedge-shaped region formed by excluding both the top $100(\alpha/2)\%$ of simulated cost-effectiveness pairs with largest (most positive) ICER angles as well as the bottom $100(\alpha/2)\%$ of simulated cost-effectiveness pairs with smallest (most negative) ICER angles

ISSUE: If any bootstrap replicates fall into Quadrant II (Highly Unfavorable), this second definition rather artificially divides them into two groups: $-180^\circ < \theta < -135^\circ$ and $+135^\circ < \theta \leq +180^\circ$, say. Note that an ICER angle of exactly $\pm 180^o$ could actually be placed into either group! This second definition becomes unambiguous when all bootstrap effectiveness differences turn out to be positive (Quadrants I and IV), as in Figure 1. Again, these are the special cases where ICER angles are restricted to the $-45^\circ < \theta < +135^\circ$ range, and ICER slopes turn out to be sufficient statistics for cost-effectiveness.

Technical Note: Figure 1 displayed a 90% confidence bootstrap interval of the "central" form. Without actually rescaling and redrawing the figure, we can gain cost-effectiveness insights by simply expressing those bootstrap results in terms of ICER angles. The ICER angle point estimate for this study was 42.41°, while 1000 bootstrap replications produced ICER angles ranging from +4.24° to +94.84° with mean=41.83° and median=41.74°. The "central" interval between order statistic 950 (+63.92°) and 50 (+18.89°) out of 1000 subtends an angle of 45.03°. By way of contrast, the region with minimum angle subtended at the origin (44.88°) occurs between order statistic 953 (+64.19°) and order statistic 53 (+19.30°) out of 1000. These 90% confidence regions also lie almost entirely within the "highly favorable" and "favorable" sectors of the cost-effectiveness plane.

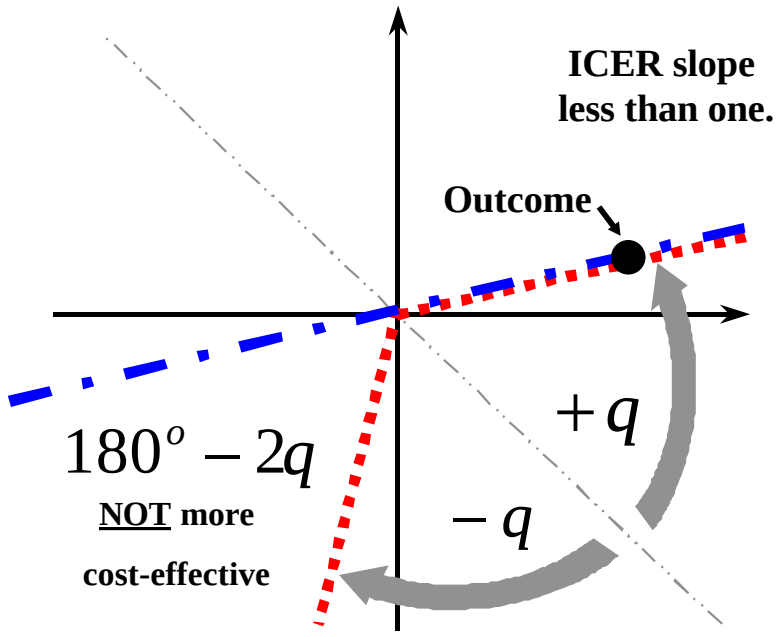
ISSUE: Chaudhary and Stearns (1996), Briggs, Wonderling, and Mooney (1996) and Stinnet (1996) point out that the central intervals of definition two are biased, i.e. the mean and median of the bootstrap distribution of the ICER slope can tend to deviate from the ICER slope point estimate. When is this bias large enough to cause concern, and how should one correct for it? Similarly, is a corresponding correction for minimum ICER angle confidence intervals needed?

ICER Angle Acceptability Curves

Van Hout, et al. (1994) introduced the concept of the **Acceptability Curve** associated with positive ICER slopes. This curve is a plot of the function AC(s)= "integrated (estimated) probability density over the cost-effectiveness plane under the ICER = s line" versus s over the range $0 \le s < +\infty$.. Furthermore, Van Hout et al. point out that AC(0) measures the "probability that the new therapy will save costs" [Quadrants III or IV] while AC($+\infty$) measures the "probability that the new therapy is effective" [Quadrants I or IV] …both relative to the standard therapy.

ISSUE: Because the ICER slope is not a sufficient statistic for cost-effectiveness, AC(s) probabilities for s > 0 and s < $+\infty$ apparently do not have simple interpretations. Specifically, the Van Hout et al. definition of outcomes "more" cost-effective than a given value, s, for the ICER slope is curious, at least to me. For example, Figure 8 depicts a study outcome in Quadrant I with an ICER slope in the 0 < s < 1 range, which corresponds to an ICER angle in the $45^o < q < 90^o$ range. Note that the Van Hout et al. region "below and/or to the right of a line of slope s" includes a wedge-shaped section of Quadrant III (subtending an angle of $180^o - 2q$) containing outcomes that do not strike me as actually being more cost-effective than the observed outcome.

Figure 8. Which outcomes are "more" cost-effective than a given ICER slope?

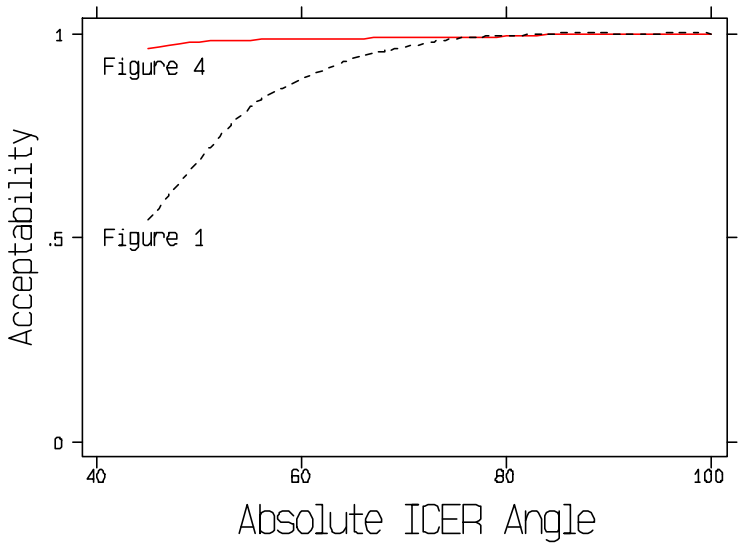


On the other hand, if the observed outcome in Figure 8 had fallen in Quadrant III (an ICER angle in the $-135^o < q < -90^o$ range), then the Van Hout et al. region would exclude a wedge-shaped section of Quadrant I (subtending an angle of $2|q| - 180^o$) containing outcomes that strike me as being more cost-effective than the observed outcome!

The ICER Angle Acceptability Curve is a plot of the function AC(t) = "integrated (estimated) probability density over the cost-effectiveness plane within the wedge-shaped segment $-t^o \le \theta \le +t^o$" versus t over the range $45^o \le t^o \le 135^o$. According to this reformulation using ICER angles, AC(45^o) measures the "probability that the new therapy saves costs **and** is more effective relative to the standard" [Quadrant IV.] Similarly, AC(135^o) measures the "probability that the new therapy saves costs **or** is more effective relative to the standard" [Quadrants I, III or IV.]

Figure 9 displays ICER Angle Acceptability Curves for the bootstrap examples of Figures 1 and 4. Note that bootstrap resampling provides **direct estimates** of the probabilities (integrated probability densities) of interest. Specifically, the estimated probability of any region of the cost-effectiveness plane is simply the number of bootstrap replicates that fall within that region divided by the total number of replicates generated. Note also that the figure displays only the ICER angle range from 45^o to 100^o, rather than all the way out to 135^o; acceptability probabilities already exceed 0.990 at $\theta = 100^o$ in these two studies, both of which highly favor the "new treatment" over the "standard."

Figure 9. Estimated ICER angle acceptability curves for figures 1 and 4.



Acknowledgement

I wish to thank John R. Cook of Merck Research Labs for detailed and extremely helpful comments on several versions of this manuscript.

Summary

Here, I have attempted to outline key issues in ICER cost-effectiveness inferences. Some issues question the philosophical basis for using ICER slopes while others simply question how statistics should be reported and interpreted in actual practice. These are open questions; my comments here should not be construed as any sort of consensus answers. For example, the practical advantages and disadvantages of ICER slopes relative to ICER angles is a subject worthy of much continued debate. In reality, these two types of measures actually complement each other. To satisfy the information needs of all participants in cost-effectiveness debates, it is probably most appropriate to report outcomes in terms of both ICER slopes and ICER angles.

Why not join us working on the **Cost-Effectiveness Inference** committee? Our charter is to help pave the way for development of consensus views by researching statistical issues in cost-effectiveness inference and setting standards for computer algorithms and validation of software.

To express interest or get additional information about our committee, please direct email to ochain@lilly.com, call (317) 276-3150, or write to me at Health Services and Policy Research, Eli Lilly Corporate Center, Indianapolis, IN 46285-1850, USA.

References

A*P*O*R Consensus Development Committee. (1996) **Economic Evaluation of Pharmaceutical Therapy: The Health Care Providers Perspective.** Draft A*P*O*R Consensus Guidance, Working Paper. Association for Pharmacoeconomics and Outcomes Research. Princeton, NJ.

Black, W. C. (1990) "The CE plane: a graphic representation of cost-effectiveness." **Medical Decision Making** 10, 212-214.

Briggs, A. H., Wonderling, D. E. and Mooney, C. Z. (1996). "Pulling cost-effectiveness analysis up by its bootstraps: a non-parametric approach to confidence interval estimation." **Fifth European Workshop on Econometrics and Health Economics**, Barcelona.

Chaudhary, M. A. and Stearns, S. C. (1996). "Estimating confidence intervals for cost-effectiveness ratios: an example from a randomized trial." **Statistics in Medicine** 15, 1447-1458.

Cochran, W. G. (1977). **Sampling Techniques.** New York: John Wiley.

Drummond, M. F. (1992) "Economic evaluation of pharmaceuticals, science or marketing?" **PharmacoEconomics** 1, 8-13.

Drummond, M. F., Heyse, J. F. and Cook, J. R. (1996) **Designing and Implementing Economic Evaluation in Health Care**. American Statistical Association, Continuing Education Program, Joint Statistical Meetings, Chicago.

Efron, B. and Gong, G. (1983). "A leisurely look at the bootstrap, jackknife and cross- **The American Statistician** 37, 36-48.

Efron, B. and Tibshirani, R. J. (1986). **"**Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy." **Statistical Science** 1, 54-77.

Efron B. and Tibshirani, R. J. (1993). **An Introduction to the Bootstrap**. New York: Chapman and Hall.

Fieller, E. C. (1954). "Some problems in interval estimation." **Journal of the Royal Statistical Society, Series B** 16, 175-183.

Genduso, L. and Kotsanos, J. G. (1996). "Review of health economic guidelines in the form of regulations, principles, policies, and positions." **Drug Information Journal** 30, 1003-1016.

Gold, M. R., Siegel, J. E., Russell, L. B., Weinstein, M. C., eds. (1996). **Cost-Effectiveness in Health and Medicine. [**Panel on Cost-Effectiveness in Health and Medicine, PCEHM, of the U.S. Public Health Service.] New York, NY: Oxford University Press.

Heyse, J. F. and Cook, J. R. (1992). "A new measure of cost-effectiveness in comparative **American Statistical Association**, Joint Statistical Meetings.

Obenchain, R. L. (1997). "ICERconf: highly portable C-language source code for confidence intervals on incremental cost-effectiveness ratio slopes and angles." Copyright © Pharmaceutical Research and Manufacturers of America, PHRMA, Washington, D.C. Download from the **StatLib** web site at URL **http://lib.stat.cmu.edu/DOS/general**, file ICER9703.EXE.

Obenchain, R. L., Melfi, C. A., Croghan, T. W. and Buesching, D. P. (1997). "Bootstrap analyses of cost-effectiveness in antidepressant pharmacotherapy." **PharmacoEconomics** 11: 464-472.

Obenchain, R. L. and Sacristan, J. A. (1997). In reply to "The Negative Side of Cost- **Journal of the American Medical Association** 277, June.

O'Brien, B. J., Drummond, M. F., Labelle, R. J. and Willan, A. (1994). "In search of power and significance: issues in the design and analysis of stochastic cost-effectiveness studies in health care." **Medical Care** 30, 231-243.

Pharmaceutical Research and Manufacturers of America. (1995) **Methodological and Conduct Principles for Pharmacoeconomic Research.** Pharmaceutical Research and Manufacturers of America, PhRMA. Washington, D. C.

Sacristan, J. A., Day, S. J., Navarro O., et al. (1995). "Use of confidence intervals and sample size calculations in health economic studies." **Annals of Pharmacotherapy** 29, 719-725.

Sacristan, J. A. and Obenchain, R. L. (1997). "Reporting Cost-Effectiveness Analyses with **Journal of the American Medical Association** 277, 375.

Siegel, J. E., Weinstein, M. C, Russell, L. B., Gold, M.R. [for the Panel on Cost-Effectiveness in Health and Medicine, PCEHM.] (1996). "Recommendations for Reporting Cost- **Journal of the American Medical Association** 276, 1339-1341.

Stinnett, A. (1996) "Adjusting for bias in C/E ratio estimates." **Health Economics** 5, 470-472.

Townsend, R., Clemens, K., Luscombe, F., Mauskopf, J, Osterhaus, J. and Bobula, J. (1994) **Report of the Workgroup on Principles of Pharmacoeconomic Research.** Health Outcomes Workgroup, Pharmaceutical Research and Manufacturers of America (PhRMA), Washington, D. C.

Van Hout, B. A., Al, M. J., Gordon, G. S. and Rutten, F. F. H. (1994). "Costs, Effects and C/E Ratios Alongside a Clinical Trial." **Health Economics** 3, 309-319.

Willan, A. R. and O'Brien, B. J. (1994). "Cost-effectiveness in clinical trials: from deterministic **Proceedings of the Biopharmaceutical Section of the ASA**, 19-28. Toronto, Canada.

Willan, A. R. and O'Brien, B. J. (1996). "Confidence intervals for cost-effectiveness ratios: an application of Fieller's Theorem." **Health Economics** 5, 297-305.