RESAMPLING AND MULTIPLICITY IN COST-EFFECTIVENESS INFERENCE

Robert L. Obenchain

Statistical and Mathematical Sciences, Eli Lilly and Company

Abstract

We compare published methods for placing statistical confidence limits around Incremental Cost-Effectiveness Ratio (ICER) statistics and show that only a non-parametric, bootstrap approach using polar angles gives completely reasonable results when neither treatment has significant advantages in cost or effectiveness. We also discuss alternative ways to report analytical results using plots, confidence or tolerance limits, and quadrant acceptability fractions. Finally, we use simulation to study the multiplicity bias that can be introduced into ICER confidence limits when only the most favorable results are reported over several possible choices of numerator cost measure and denominator effectiveness measure.

Key Words: Incremental cost-effectiveness ratio, Bootstrap resampling, Polar coordinates, Confidence and tolerance intervals, Exchangeable variables, Multiplicity bias.

The author wishes to thank the members of the Cost-Effectiveness Inference Working Group (CEIWG), especially Cathy Melfi, John Cook and Magnus Tambour, for their comments on concepts and computing algorithms discussed in this paper.

This paper was presented at the ***Biopharmaceutical Applied Statistics Symposium*** *'98 (BASS V).*

Lilly Research Laboratories, Indianapolis, Indiana 46285-1850, (317) 276-3150 voice, (317) 276-3150 fax, ochain@lilly.com

1. INTRODUCTION

Cost Effectiveness Analysis (CEA) has been a hot topic in pharmacoeconomics and health care outcomes research for at least the last decade. The first 27 papers in my reference list represent key technical and expository works in this CEA literature that focus primarily on Incremental Cost-Effectiveness Ratio (ICER) statistics for comparing just two treatments. Our discussion here first reviews three very different types of econometric methods of revealing the uncertainty associated with ICER estimates.

We use a numerical example in section 2 to illustrate the wide variety of options one typically has in reporting findings about ICER uncertainty. This example also illustrates shortcomings in the Bonferroni box and Fieller's theorem approaches when neither treatment has significant cost or effectiveness advantages over the other.

Section 3 discusses converting ICER angle bootstrap confidence regions into non-parametric tolerance regions for ICER uncertainty. Section 4 uses simulation to study the multiplicity bias that is introduced into bootstrap ICER angle confidence limits when there are several possible choices of numerator cost measure and/or denominator effectiveness measure but only the most favorable analysis is reported. Finally, section 5 provides an overall summary of findings.

1.1 INCREMENTAL COST-EFFECTIVENESS RATIOS

To perform a 2-sample CEA, measurements on at least one (continuous) cost variable, C, as well as at least one (binary or continuous) treatment effectiveness indicator, E, need to be collected on each patient who receives one of the two treatments that are to be compared. In other words, a pair of measurements, (C_{Ti} , E_{Ti}), are recorded for each of the i = 1, …, N_T patients who received the <u>new</u> (or test) treatment, T. Similarly, a pair of measurements, (C_{Sj} , E_{Sj}), are recorded for each of the j = 1, …, N_S patients who received the <u>standard</u> treatment, S.

The "incremental" cost-effectiveness ratio, ICER, for treatment T relative to treatment S is defined, Black(1990), to be the difference in average per patient cost (treatment T minus treatment S) divided by the corresponding difference in effectiveness averages,

$$ICER = \frac{\overline{C}_T - \overline{C}_S}{\overline{E}_T - \overline{E}_S} = \frac{\Delta C}{\Delta E} \ . \qquad (1)$$

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

Note that, as displayed in Figure 1, the ICER statistic is nothing more than the <u>slope of the line</u> on the cost-effectiveness plane that connects the health economics study point, (ΔE, ΔC) =

($\overline{E}_T - \overline{E}_S, \overline{C}_T - \overline{C}_S$), with the origin, (0,0). In particular, note that the ICER is clearly not a sufficient statistic for CEA in the sense that the opposite outcome, resulting from a switch in the numerical sign of both the numerator cost difference and the denominator effectiveness difference, yields the same value for the ICER slope!

Insert Figure 1 About Here.

Three very different types of methodology for placing statistical confidence limits around Incremental Cost-Effectiveness Ratios (ICERs) are currently in active use. These three approaches are (i) "box" methods for combining cost-difference limits with effectiveness-difference limits, (ii) parametric methods for analysis of "ratio estimates," including Fieller's theorem, and (ii) non-parametric, bootstrap methods**.**

1.2 THE BOX APPROACH

In the so-called "box" approach discussed by Wakker and Klaassen(10) and Tambour and Zethraeus(25, 26), a rectangular confidence region is defined by combining separate intervals of the form $\left(\overline{E}_T - \overline{E}_S\right) \pm t \cdot s_E$ and $\left(\overline{C}_T - \overline{C}_S\right) \pm t \cdot s_C$, where s_E and s_C are estimates of the standard deviations of the between cohort differences in effectiveness and cost, respectively. The minimum overall confidence level can be maintained at 100(1–α)% by choosing t so that the confidence intervals for the cost difference and the effectiveness difference each have confidence level 100[1–(α/2)]%; this is a "Bonferroni adjustment," Miller(28), page 67. In other words, t is then chosen so that Pr(T > +t) = Pr(T < –t) = α / 4, where T follows a Student's t-distribution with N_T+N_S-2 degrees-of-freedom.

The primary advantage of this box approach is its simplicity. Furthermore, the calculations require only knowledge of sample means and variances, statistics that are commonly reported in health economics studies. One downside of the box approach, illustrated in Figure 2 below, is that the corresponding upper and lower confidence limits for the ICER slope tend to be quite wide (conservative.)

Insert Figure 2 About Here.

Major difficulties arise when the origin, (0,0), of the cost-effectiveness plane is covered by the box. The corresponding ICER confidence limits then range all the way from minus infinity to plus infinity! In other words, a confidence region derived via the Bonferroni "box" approach may place no restriction, whatsoever, on the ICER statistic.

1.3 FIELLER'S THEOREM

The Fieller(29) approach discussed in references (5, 6, 9, 11, 12, 13, 17 and 20) recognizes that the ICER statistic is a "ratio estimator" in the sense of Cochran(30) and, thus, is asymptotically normally distributed. However, the characteristic property of the Fieller approach is that it treats both the numerator and denominator between cohort differences, $\left(\overline{E}_T - \overline{E}_S\right)$ and $\left(\overline{C}_T - \overline{C}_S\right)$, as if they were a pair of correlated normal variables. Due to the Central Limit Theorem, each of these differences is usually very well approximated by a normal distribution when cohort sample sizes are large. Still, the Fieller approach does recognize (small sample) situations where the stochastic distribution of the ICER is actually highly skewed. Willan and O'Brien(5,12), O'Brien et al.(6) and Sacristan et al.(9) discuss many technical details. Chaudhary and Stearns(13) developed an extremely useful quadratic (closed) form solution for Fieller ICER confidence limits; older proposals involving Taylor's series approximations are thus no longer needed to perform calculations.

Unfortunately, as shown in Figure 3, Fieller confidence limits for an ICER correspond to a "bow-tie" shaped confidence region on the cost-effectiveness plane. This somewhat curious shape presents no real difficulties when the mean vector of between treatment differences, $(\Delta E, \Delta C) = \left(\overline{E}_T - \overline{E}_S, \overline{C}_T - \overline{C}_S\right)$, is highly significantly different from (0,0). For example, suppose that Figure 3 depicts the case where a highly significant, observed $(\Delta E, \Delta C)$ difference falls in the right half of the bow-tie region. The left half of this bow-tie region would then be highly unlikely to contain the unknown, true expected value of $(\Delta E, \Delta C)$. In other words, the left "half" could be simply ignored in this case.

Insert Figure 3 About Here.

One potential problem with the Fieller approach is that transformations of ICER statistics in the form of simple "scale changes" are sometimes appropriate. For example, scale changes occur in converting a numerator cost difference from one currency into another or in discounting charges relative to a different base year. Unfortunately, the Fieller approach is sensitive to these simple scale changes in the sense that the resulting ICER confidence interval is not "rescaling commutative." In other words, rescaling an ICER statistic by one multiplicative factor changes its upper and lower Fieller confidence limits by a different factor.

The main problem with Fieller limits is that they, like box limits, also tend to get very wide when $(\Delta E, \Delta C)$ is not highly significantly different from (0,0). In fact, the term inside the square root in Chaudhary and Stearns(13) formula 10, page 1450, then becomes negative, implying that only imaginary solutions exist. In actual practice, this again implies an ICER confidence region that spans 100% of the cost-effectiveness plane!

1.4 BOOTSTRAP RESAMPLING

Efron and Gong(31), Efron and Tibshirani(32,33) and Westfall and Young(34) describe bootstrap approaches to a wide variety of problems. For applications of bootstrap methodology specifically to ICER statistical inference, see references (4, 5, 11, 12, 13, 16, 17, 18, 20, 22 and 27.) Interestingly, all of these authors recommend resampling observed patient data pairs within treatment groups. In other words, each original pairing of a cost measurement with an effectiveness measurement for a single patient remains not only intact but also identified with the treatment that patient actually received. Each bootstrap ICER re-computation involves (a) resampling N_T data pairs, (C_{Ti} , E_{Ti}), with replacement from the N_T patients taking the new treatment, T, (b) resampling N_S data pairs, (C_{Sj} , E_{Sj}), with replacement from the N_S patients taking the standard treatment, S, and (c) calculating the resulting within treatment means, between treatment mean differences, $(\Delta E, \Delta C)$, and the ICER statistic as in equation (1). Each resulting pair of mean differences, $(\Delta E, \Delta C)$, represents a resampled point on the cost-effectiveness plane. Together, they yield a rather dramatic graphical display of the variability in 2-sample cost and effectiveness differences that result when a study is literally "redone" hundreds of times. This is illustrated in Figure 4, reproduced from Obenchain et al.(18).

Insert Figure 4 About Here.

We do not always find ourselves in the simple situation of Figure 4 where all of the results generated in an ICER bootstrap analysis fall only on the right (more effective) side of the cost-effectiveness plane. Some bootstrap effectiveness differences, $\Delta E = \overline{E}_T - \overline{E}_S$, may turn out to be negative**,** rather than all positive. In fact, bootstrap resampled points may appear in all four quadrants of the cost-effectiveness plane when neither ΔE nor ΔC is significantly different from 0.

Using polar coordinates on the cost-effectiveness plane to analyze bootstrap results offers distinct advantages. Specifically, resampled $(\Delta E, \Delta C)$ points are then ordered (sorted) by polar angle, discarding information on radius. After all, if resampled points were ordered only by the numerical value of their ICER slope, results from different quadrants would be considered equivalent. Sorting results by polar angle allows one to keep track of the quadrant where each resampled $(\Delta E, \Delta C)$ falls. Knowledge of quadrant is essential for treating cases where neither the ΔE nor the ΔC in the original sample are significantly different from 0.

1.4.1 POLAR COORDINATES AND 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(3). One reasonable standardization is achieved by dividing each difference in treatment averages by the estimated standard deviation of the corresponding difference between just two patients. In other words, standardized effectiveness = **x** and cost = **y** coordinates are defined by

$$x = \frac{\left(\overline{E}_T - \overline{E}_S\right)}{\sqrt{s^2\left(E_T\right) + s^2\left(E_S\right)}} \quad \text{and} \quad y = \frac{\left(\overline{C}_T - \overline{C}_S\right)}{\sqrt{s^2\left(C_T\right) + s^2\left(C_S\right)}}, \tag{2}$$

where $s^2\left(E_T\right)$ denotes the sample standard deviation of the N_T effectiveness measurements from patients receiving the new treatment, etc. 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.

An alternative standardization to equation (2) would be to divide each mean difference by its own estimated standard error. For example, $\left(\overline{C}_T - \overline{C}_S\right)$ would then be divided by $\sqrt{s^2\left(C_T\right)/N_T + s^2\left(C_S\right)/N_S}$. This convention would be fine as long as sample sizes were equal, $N_T = N_S$. However, using this standardization when $N_T \neq N_S$ could provide an unfair advantage to one of the treatments, and might even encourage use of unequal sample sizes in cost-effectiveness analyses.

Figure 5, below, proposes a division of the cost-effectiveness plane into 8 Wedge-Shaped Sections of 5 different types; together, they span the full range of possible outcomes of health economic studies comparing a new treatment with a standard treatment. For example, the entire (–,+) quadrant is labeled "Highly Favorable" because the new treatment is both less costly and more effective than the standard treatment in this region. Similarly, the (+,–) quadrant is labeled "Highly Unfavorable" to the new treatment. Again, these are the two quadrants of the cost-effectiveness plane where the ICER slope is negative.

Insert Figure 5 About Here.

The ICER slope is positive in both the (+,+) quadrant and the (–,–) quadrant, and each of these quadrants is subdivided into 3 parts. Note that a numerically small but positive ICER slope is "Favorable" in the (+,+) quadrant yet "Unfavorable" in the (–,–) quadrant. Similarly, a large, positive ICER slope is "Favorable" in the (–,–) quadrant but "Unfavorable" in (+,+). The two segments of the fifth type are the "Gray Areas" in the middle of the (+,+) and (–,–) quadrants where the ICER slope is positive but neither very large nor very small.

In particular, note in Figure 5 the symmetry of each of the five types of cost-effectiveness segment about the standardized -45° line, $x + y = 0$. This observation suggests placing the $\theta = 0$ origin for measuring ICER angles along the -45° line within the (–,+) quadrant. Also, by convention, suppose that counter-clockwise rotations are said to represent increasing (positive) polar angles while clockwise rotations represent decreasing (negative) polar angles.

With these conventions, the ICER angle for any standardized point, (**x,y**), is defined by

$$\left|\theta\right| = \arctan\left(\left|x+y\right| / (x-y)\right) \textit{when } x \neq y \qquad (3)$$
$$= 90^\circ \textit{ when } x = y,$$

where $-180^o < \theta < 0^o$ when $x+y<0$, $\theta = 0^o$ when $-y = x > 0$, $0^o < \theta < 180^o$ when $x+y>0$, and $\theta = \pm 180^o$ when $-x = y > 0$. The corresponding standardized ICER slope is

$$s = y / x = \tan\left(\theta - 45^o\right). \qquad (4)$$

Note that the standardized ICER slope, s, is easily expressed as a function of the ICER angle, θ, but θ is not a one-to-one function of s alone because s is not sufficient to determine the ICER quadrant. Furthermore, since $\tan\left(-\theta - 45^o\right) = 1/\tan\left(\theta - 45^o\right)$, the standardized ICER slopes associated with ICER angles of θ and $-\theta$ are reciprocals of each other.

In non-parametric ICER analyses, it is essential to measure the angle subtended between pairs of ICER angle order statistics in a consistent way, say, always counter-clockwise. For example, with 1000 bootstrap replicates (numbered 1 to 1,000), the subtended angle between order statistic 25 with, say, $\theta = -6^o$ and order statistic 975 with, say, $\theta = +173^o$ would be $173^o - (-6^o) = 179^o$. Similarly, the subtended angle between order statistic 550 with $\theta = +20^o$ and order statistic 500 with $\theta = +10^o$ would be a counter-clockwise rotation of $+350^o$ rather than $10^o - 20^o = -10^o$, which would be a clockwise (negative) rotation. Like order statistics 25 and 975, order statistics 550 and 500 are also separated by 950 positions out of 1000. Thus either pair could be used to define a 95% confidence region. But, in the above example, the segment between order statistics 25 and 975 would subtend a much smaller polar angle (179^o rather than 350^o.)

Finally, note that it makes good sense to define **Contours of Constant Cost-Effectiveness** (of the new treatment **T** relative to the standard treatment **S**) as in Figure 6, below. Note that each such contour consists of a pair of line segments, joined at (**x,y**) = (0,0), and making equal angles, $\pm\theta$, with the standardized -45° line, $x + y = 0$.

Insert Figure 6 About Here.

Table 1, below, describes proposed terminology for the five distinct sections of the of the cost-effectiveness plane, listing the corresponding quadrant and range of both ICER slopes and angles.

Insert Table 1 About Here.

The 60^o and 120^o values proposed in Table 1 as boundaries between the "Favorable", "Grey Area" 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 the A^o angle could conceivably vary between therapeutic areas.

On the other hand, taking $A^o = 15^o$ does allow the "Gray Areas" to occupy exactly 1/3rd of the total cost-effectiveness plane, i.e. these two wedges have a combined angular-measure of 120^o out of 360^o. This leaves 1/3rd of the cost-effectiveness plane either "Green" for highly favorable(1/4th) or "Yellow" for favorable(1/12th) to the new treatment. The final 1/3rd of the cost-effectiveness plane is either "Red" for highly unfavorable(1/4th) or "Pink" for unfavorable(1/12th) to the new treatment.

1.4.2 ICER ANGLE CONFIDENCE LIMITS

Consider now the following two possible definitions for $100(1-\alpha)\%$ confidence limits based upon ICER angle order statistics.

Minimum Angle Definition: The bootstrap $100(1-\alpha)\%$ confidence region for cost-effectiveness is the wedge-shaped region subtending the smallest polar angle and yet containing $100(1-\alpha)\%$ of the resampled cost-effectiveness pairs.

Count Outward Definition: The bootstrap "central" $100(1-\alpha)\%$ confidence region for cost-effectiveness is the wedge-shaped region formed by including only the $50(1-\alpha)\%$ of ICER angles immediately clockwise plus the $50(1-\alpha)\%$ of ICER angles immediately counter-clockwise of the ICER angle observed for the original data (i.e. the ICER for the sample in which each patient appears exactly once.)

Only in very clear-cut situations, such as that of Figure 4 where all observed ICER angles are between -45^o and $+135^o$, does it make sense to use a "count inward" definition. In this formulation, one would exclude both the top $100(\alpha/2)\%$ of resampled cost-effectiveness pairs with largest (most positive) ICER angles as well as the bottom $100(\alpha/2)\%$ of resampled cost-effectiveness pairs with smallest (most negative) ICER angles. However, Chaudhary and Stearns(13), Briggs, Wonderling and Mooney(4) and Stinnet(16) have all pointed out that the resulting "count inward" intervals are biased. Both the mean and the median of the bootstrap distribution of the ICER slopes tends to deviate from the original ICER slope point estimate.

The "minimum angle" and the "count outward" definitions both offer great intuitive appeal. The strength of the "minimum angle" definition is that the resulting $100(1-\alpha)\%$ confidence region can never occupy more that $100(1-\alpha)\%$ of the cost-effectiveness plane in terms of angular measure; its weakness is that its definition depends heavily on the standardized (***x,y***) scaling of equation (2). The strength of the "count outward" approach is that it's definition (using ICER angle order statistics) is actually independent of choice of (***x,y***) scaling! As a result, the "count outward" definition probably should be generally preferred over the "minimum angle" definition

…except in truly pathological numerical examples where the "minimum angle" definition can yield a much, much smaller subtended polar angle.

1.4.3 DOWNSIDES TO ICER BOOTSTRAPPING

Note that the bootstrap approach to ICER inference requires access to patient level data pairings, (C_{Ti} , E_{Ti}) and (C_{Sj} , E_{Sj}). Bootstrap replicates cannot be constructed from the sorts of simple summary statistics (sample means, variances and correlations) commonly reported in health economics studies and journal articles.

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 pseudo-random number generator. To satisfy pharmacoeconomic "full disclosure" guidelines, these sorts of technical details need to be reported.

No implementation of ICER bootstrap analysis is currently available (early 1999) in commercial statistical analysis software. On the other hand, my freely distributed software, Obenchain(35), provides a 32-bit Microsoft Windows API interface, Petzold(36), to highly portable algorithms for ICER confidence limit calculations as well as Graphics Server(37) support for plotting windows.

Finally, "sensitivity" analyses need to be performed to assure that bootstrap limits are not reported with too many decimal places. For example, one could perform two separate and independent ICER bootstrap analyses with the same number of replications (default = 25,000) on the observed data. One could then report the results from the first analysis rounded to the number of digits confirmed by the second analysis.

2. REPORTING ICER UNCERTAINTY

Let us now consider a numerical example that illustrates the wide range of options one typically has in reporting the uncertainty associated with ICER statistics. I also picked this example to illustrate potential problems with the box and Fieller's theorem approaches that are easily treated by bootstrapping ICER angles.

Consider the following simulated data for $N_T = N_S = 50$ patients. I started by generating 4 independent groups of 50 independent and identically distributed normal Z-scores (mean zero and variance one.) The first 2 groups of values were designated measures of effectiveness and transformed by E = 10 + Z. The last two groups were designated measures of cost and transformed by C = 500 + 10 * Z. All numerical values were then truncated to two decimal places. Finally, the group of effectiveness measures with the larger mean value and the group of costs with the lesser mean value were paired and designated as the observations on the patients receiving the "new" treatment. [Since I used the "?normal" function in JMP® version 3.2.2 to generate the data, I do not know what initial seed value was used. But the numerical values I generated are saved in a file named RANDOM.DAT that is distributed with my free software, Obenchain(35).]

	Patients	Eff. Mean +/– Std. Dev.	Cost Mean +/– Std. Dev.
New	**50**	**10.21 +/– 0.972**	**499.51 +/– 9.47**
Standard	**50**	**10.02 +/– 0.905**	**501.72 +/– 9.03**

Since it is clear from the way the data were generated that true effectiveness and cost differences are zero ($\Delta E = \Delta C = 0$), it is certainly not surprising that these same differences are not judged statistically significant using the statistics from the above table.

The "box" 95% confidence region for (ΔE, ΔC) uses the $(1-\alpha/4) = 0.9875$ significance point of t = +2.276 for Student's t-distribution with 98 degrees of freedom and is of the form:

$$\mathbf{-0.230 < \Delta E < +0.625 \quad and \quad -5.137 < \Delta C < +0.711.}$$

Thus (0,0) is inside the box. Similarly, the term inside the square root of the Chaudhary and Stearns(13) formula is negative for these data! Thus both the box and Fieller approaches imply that the true value for the ICER slope could be anything ($-\infty < s < +\infty$) and the true value for the ICER angle could be anything ($-180^o < \theta \leq +180^o$.) Some readers may well think that this is the <u>exactly correct</u> conclusion to draw! But why settle for a 95% confidence region that occupies 100% of the cost-effectiveness plane when, as shown below, a segment subtending a polar angle of less than 180^o can perform the same "job?"

Bootstrap calculations, performed using the Obenchain(35) software, used a randomly selected seed of 26894. Only the first 1,000 out of 25,000 total replications are shown in Figure 7 to avoid excessive over-printing. The observed ICER slope was –11.14 and the observed ICER angle was -3.51^o. Resampled ICER angles ranged from -179.45^o to $+179.64^o$.

Insert Figure 7 About Here.

Figure 7 also displays the "count outward" 95% limits for ICER slopes and ICER angles based upon all 25,000 bootstrap replicates (not just the 1,000 outcomes actually displayed in the figure.) This sort of graphical display is my personal favorite way to see exactly what one's data are saying about ICER uncertainty. In fact, the only major downside of this sort of graphical display is that one cannot see exactly where the original ($\Delta E, \Delta C$) outcome fell. On the other hand, the count-outward definition does assure that the original ($\Delta E, \Delta C$) outcome (each patient appearing exactly once) falls at the median ICER angle within one's confidence wedge.

The "minimum angle" and "count outward" definitions give quite similar results in this example. The 95% confidence interval with "minimum angle" is defined by order statistic 621 of 25,000 (with angle = -89.13^o and slope = +10.16) and order statistic 24,371 of 25,000 (with angle = $+85.16^o$ and slope = +8.32); the minimum subtended polar angle is thus 174.29^o. By way of contrast, the 95% "count outward" confidence interval is defined by order statistic 586 of 25,000 (with angle = -91.55^o and slope = +9.34) and order statistic 24,336 of 25,000 (with angle = $+83.28^o$ and slope = +7.78); this very slightly larger polar angle still subtends only 174.83^o. Note that, for both 95% intervals, the lower and upper limits are separated by exactly $(0.95\times25{,}000) - 1 = 23{,}784$ ICER angle order statistics.

The most straight-forward way to express the results shown in Figure 7 in words uses ICER angles. Specifically, one would simply say "The 95% confidence interval extends over the ICER angle range from -91.55^o to $+83.28^o$." Geometrically inclined readers would then be able to immediately visualize something very much like Figure 7. For other readers, it would probably help to also say which parts of the areas named in Table 1 and Figure 5 lie within your confidence region. For example, here you might add "This interval starts approximately in the middle of the Grey Area within the (–,–) quadrant, extends counter-clockwise through the entire (–,+) quadrant, and ends approximately in the middle of the Grey Area within the (+,+) quadrant."

This example also illustrates the added awkwardness that results when attempting to express ICER uncertainty in words using only ICER slopes. One then has to say something like "The 95% confidence interval includes all positive ICER slopes greater than +9.34 in the (–,–) quadrant, all negative ICER slopes in the (–,+) quadrant, and all positive ICER slopes less than +7.78 in the (+,+) quadrant." Technically, a somewhat simplified statement like "the 95% confidence interval includes all ICER slopes greater than +9.34 or less than +7.78" is correct in the weak sense that all negative ICER slopes are certainly less than +7.78. But even this simplified statement would probably sound quite strange to most readers. Furthermore, the simplification fails to say that the lower-right-hand half of the cost-effectiveness plane is mostly inside the resulting interval (rather than all of the upper-left-hand half.)

Van Hout, Al, Gordon and Rutten(7) introduced the concept of an Acceptability Curve associated with positive ICER slopes. This curve is a plot of the function AC(s)= "integrated probability density over the cost-effectiveness plane under or to the right of the ICER = s line" versus s over the range $0 \le s < +\infty$. ICER bootstrapping leads to particularly simple estimates of the probability for any sub-region of the cost-effectiveness plane; one simply divides the number of bootstrap replicates that fall within the sub-region by the total number of replicates generated. A particularly simple yet frequently highly informative alternative to plotting an acceptability curve is to report Quadrant Acceptability Fractions. For the current example, the fractions of bootstrap replicates falling into each of the four quadrants of the cost-effectiveness plane were

(–,–) 0.13; (–,+) 0.76; (+,+) 0.10; (+,–) 0.01.

3. ICER ANGLE TOLERANCE INTERVALS

Suppose one wishes to define a wedge-shaped ICER angle region with stated confidence that it contains at least a stated percentage of the distribution of uncertainty in an observed ICER statistic. Guttman(38) and Quesenbury(39) discuss the basic theory of tolerance intervals with end points at order statistics, which turn out to be distribution free due to a probability-integral transformation argument. This theory extends immediately to angular order statistics if one assumes that the distribution of ICER angles is absolutely continuous around the full circle surrounding the origin of the cost-effectiveness plane.

Remember that we are cutting the circle at the $x + y = 0$ line in the (+,+) quadrant and defining ICER angles on the range $-180^o < \theta \leq +180^o$. Note that ICER angle order statistic number N of N within this half-open interval is counter-clockwise "adjacent" to order statistic number 1 of N in the exact same sense as order statistic number 1 of N is adjacent to order statistic number 2 of N. Technically, we used this "wrap-around" concept to define both minimum angle and count outward ICER angle confidence limits in section 1.4.2. Equivalently, note that N points divide a circle into only N "statistically equivalent" segments, instead of the N+1 "statistically equivalent" intervals into which N points divide a line. The end result is that, although ICER angle order statistics are numbered 1 through N, non-parametric ICER angle tolerance interval calculations proceed as if the sample size were N–1 points along a line rather than N points around a circle.

The end-points of ICER angle intervals yielding 100(1–α)% confidence of containing the unknown, true ICER angle use polar angle order statistics separated by one fewer than (1–α)N adjacent order statistics. Here we wish to consider end-points at order statistics separated by a few more than (1–α)N adjacent order statistics in order to have 100(1–α)% confidence that at least 100(1–α)% of the ICER angle uncertainty distribution lies between the resulting non-parametric limits. For example, when N = 25,000, incomplete beta function calculations reveal that the interval between order statistics 1 and 23,807 (which is 56 order statistics higher than 1 + 0.95 × 25,000 = 23,751) yields a tolerance interval with 95.008% confidence in at least 95% ICER uncertainty content. The interval between order statistics 2 and 23,808 (or 3 and 23,809, etc.) would have this same property.

Remember that the 95% "count outward" confidence interval for the example of section 3 was defined by order statistic 586 of 25,000 (with angle = -91.55^o and slope = +9.34) and order statistic 24,336 of 25,000 (with angle = $+83.28^o$ and slope = +7.78.) The corresponding 95% "count outward" tolerance interval is thus defined by order statistic 558 of 25,000 (with angle = -93.19^o and slope = +8.81) and order statistic 24,364 of 25,000 (with angle = $+84.91^o$ and slope = +8.24.) This tolerance interval subtends a polar angle of 178.10^o rather than the 174.83^o span of the corresponding 95% confidence interval. In other words, using a segment with polar angle increased by a little more than 3 degrees provides 95% confidence in coverage of at least 95% of the entire ICER uncertainty distribution rather than just 95% confidence in covering its unknown mean value.

4. POTENTIAL FOR MULTIPLICTY BIAS

Our last topic is the introduction of bias into ICER angle bootstrap confidence intervals when there are several different possible numerator cost measures and/or several different possible denominator effectiveness measures and only the most favorable analysis is actually reported. There are situations where subject matter experts would consider one of the several possible cost-effectiveness analyses more relevant than all of the others, but that would imply that the corresponding cost and effectiveness measures have a different joint distribution than that of all other pairings. That single analysis should be reported almost regardless of whether it was most favorable, least favorable or somewhere in between. So let us consider here only situations in which all potential variables are of exactly equal importance in the sense that they are exchangeable (or interchangeable) random variables. In other words, all cost measures are identically distributed, and the correlation between any two of them is ρ_{CC}. Similarly, all effectiveness measures are identically distributed, and the correlation between any two of them is ρ_{EE}. Finally, the common correlation of each cost measure with each effectiveness measure will be denoted by ρ_{CE}.

The above situation is easy to study using simulation if one makes the highly restrictive assumption that the joint distribution of all cost and effectiveness measures is multivariate normal. For example, with 3 cost measures and 3 effectiveness measures, the resulting overall 6×6 correlation matrix is block diagonal. The 6 off-diagonal correlations in the 3×3 cost block and the 3×3 effectiveness block are ρ_{CC} and ρ_{EE}, respectively. All 9 correlations in each 3×3 off-diagonal block are ρ_{CE}. Given numerical values for these 3 types of correlation, one then computes any 6x6 "square root" matrix, **W**, such that $\mathbf{WW}'$ equals the given 6x6 block-diagonal correlation matrix. Using well known algorithms, one can generate independent and identically distributed pseudo-normal variates (mean zero and variance one) and form them into a column vector, **z**. The column vector **Wz** then has the appropriate multivariate normal distribution (mean zero and variance one.)

Suppose now that one wishes to simulate a situation in which each of 9 possible cost-effectiveness analyses (from 3 possible numerators and 3 possible denominators) looks somewhat like that of Figure 4. The difficulty here is that the distribution of the total yearly cost measure studied in Obenchain et al.(18) was highly skewed in both the new treatment and the standard treatment samples, and the effectiveness measure on individual patients was binary (meet treatment guidelines for depression, yes=1 or no=0.) Since individual numerical values generated as outlined above are strictly multivariate normal, we cannot simply use the mean, variance and ρ_{CE} estimates from the Obenchain et al.(18) study and yet realistically expect individual bootstrap analyses to still look much like Figure 4. Instead, one will usually need to try out several different possible numerical values for ρ_{CC}, ρ_{EE} and all other simulation settings more-or-less by trial and error.

To give the reader some small feel for potential multiplicity bias in ICER bootstrapping, I will now outline a simulation study that is discussed in great detail in the documentation for my freely distributed simulation module, CE_MULT.EXE, Obenchain(40). I used 5,000 ICER angle bootstrap replications on each set of simulated data for 3 exchangeable cost and 3 exchangeable effectiveness variables with $\rho_{CC} = \rho_{EE} = 0.8$; $\rho_{CE} = 0.1$ on 100 patients in each treatment group. I then repeated this simulation scenario 100 times. As shown in Figure 8, the overall average upper count outward 95% confidence limit fell at an ICER angle of $+67^o$, which is within the "Gray" area of the (+,+) quadrant. But the corresponding average most favorable (most clockwise) out of 9 possible upper 95% limits fell at and ICER angle of only $+59^o$, which is within the "Favorable" area of the (+,+) quadrant.

Insert Figure 8 About Here.

Much more clockwise bias could have been introduced if exchangeable cost and effectiveness variables had not been taken to be so highly, positively correlated ($\rho_{CC} = \rho_{EE} = +0.8$.) My freely distributed simulation module, Obenchain(40), can be used to study any such scenario.

5. SUMMARY

Bootstrap approaches to ICER inference do not need to make unrealistic assumptions about parametric forms for stochastic distributions or cost-effectiveness correlations and, thus, offer great potential for increased simplicity, power, robustness and interpretability. In fact, bootstrap ICER analyses and their resulting graphical displays of uncertainty are actually easier to understand and appreciate than crude approximations or elaborate Fieller's theorem calculations.

The advantages of using standardized ICER angles over (un-standardized) ICER slopes in CEA tend to be theoretical and computational; some decision makers may not find polar coordinates helpful in visualizing ICER uncertainty. Still, as illustrated in the numerical examples, ICER slope and angle measures do tend to complement each other. Clarity of communication is enhanced when CEA results are reported both ways …and illustrated by a graphical display.

Only ICER angle bootstrap confidence intervals give visually reasonable results even when neither treatment (new or standard) has significant advantages in cost or effectiveness. The Bonferroni box and Fieller's theorem approaches completely fail in these cases in the sense that they imply that the true ICER angle could be "anything." Unfortunately, some decision makers may visualize the statement that "any ICER angle is possible" as implying approximately 25% confidence in each quadrant of the cost-effectiveness plane. The numerical example of section 2 showed that quadrant acceptability fraction estimates can be far from **(–,–) 0.25; (–,+) 0.25; (+,+) 0.25; (+,–) 0.25** in these cases.

By including a relatively small number of additional ICER angle order statistics within a wedge shaped region (using either the minimum angle or count outward definitions), an ICER angle confidence region can be converted into a non-parametric tolerance region for ICER uncertainty. Using a segment with a slightly increased polar angle thereby provides stated confidence in coverage of at least a stated minimum percentage of the entire ICER uncertainty distribution …rather than just a stated confidence in covering the unknown true ICER angle mean value.

Finally, we used simulation to study one simple scenario in which bias is introduced into bootstrap ICER angle confidence limits when only the most favorable result is reported out of 9 possible analyses using 3 exchangeable cost measures and 3 exchangeable effectiveness measures. Bias should be avoided in this sort of situation by, instead, pooling equal numbers of bootstrap replicates from each of the possible, different choices for numerator and denominator.

Freely distributed software, executable under Microsoft Windows® 95/NT/98, implements the techniques discussed here.

REFERENCES

1. Laska EM, Meisner MJ: Testing whether an identified treatment is best. *Biometrics* 45, 1139-1151, 1989.

2. Black WC: The CE plane: a graphic representation of cost-effectiveness. *Medical Decision Making* 10, 212-214, 1990.

3. Heyse JF, Cook JR: A new measure of cost-effectiveness in comparative clinical trials. *American Statistical Association*, Joint Statistical Meetings, 1992.

4. Briggs AH, Sculper MJ, Buxton MJ: Uncertainty in the economic evaluation of health care technologies: the role of sensitivity analysis. *Health Economics* 3, 95-104, 1994.

5. Willan AR, O'Brien BJ: Cost-effectiveness in clinical trials: from deterministic to stochastic models. *Proceedings of the Biopharmaceutical Section of the ASA*, 19-28. Toronto, Canada, 1994.

6. 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* 30, 231-243, 1994.

7. Van Hout BA, Al MJ, Gordon GS, Rutten FFH: Costs, effects and C/E ratios alongside a clinical trial. *Health Economics* 3, 309-319, 1994.

8. Heyse JF, Cook JR: Identification of the cost-effectiveness frontier: a statistical assessment. *International Conference on Health Policy Research*, 1995.

9. Sacristan JA, Day SJ, Navarro O, et al: Use of confidence intervals and sample size calculations in health economic studies. *Annals of Pharmacotherapy* 29, 719-725.

10. Wakker P, Klaassen MP: Confidence intervals for cost/effectiveness ratios. *Health Economics* 4, 373-381, 1995.

11. Drummond MF, Heyse JF, Cook JR: *Designing and Implementing Economic Evaluation in Health Care.* American Statistical Association, Continuing Education Program, Joint Statistical Meetings, Chicago, 1996.

12. Willan AR, O'Brien BJ: Confidence intervals for cost-effectiveness ratios: an application of Fieller's theorem. *Health Economics* 5, 297-305, 1996.

13. Chaudhary MA, Stearns SC: Estimating confidence intervals for cost-effectiveness ratios: an example from a randomized trial. *Statistics in Medicine* 15, 1447-1458, 1996.

14. Gold MR, Siegel JE, Russell LB, Weinstein MC, eds: *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, 1996.

15. Siegel JE, Weinstein MC, Russell LB, Gold MR: Panel on Cost-Effectiveness in Health and Medicine (PCEHM) recommendations for reporting cost-effectiveness analyses. *Journal of the American Medical Association* 276, 1339-1341, 1996.

16. Stinnett AA: Adjusting for bias in C/E ratio estimates. *Health Economics* 5, 470-472, 1996.

17. Polsky D, Glick H: Confidence intervals for cost effectiveness ratios: are the bootstrap and Fieller theorem methods dependably accurate in small samples?" *Methodologic Issues in Health Services Research, International Conference on Health Policy Research*, 1997.

18. Obenchain RL, Melfi CA, Croghan TW, Buesching DP: Bootstrap analyses of cost-effectiveness in antidepressant pharmacotherapy. *PharmacoEconomics* 17, 1200-1206, 1997.

19. Laska EM, Meisne MJ. and Siegel, C: Statistical inference for cost effectiveness ratios. *Health Economics* 6, 229-242, 1997.

20. Polsky D, Glick H, Willke R, Shulman K: Confidence intervals for cost effectiveness ratios: a comparison of four methods. *Health Economics* 6, 243-252, 1997.

21. Obenchain RL: Issues and algorithms in cost-effectiveness inference. *Biopharmaceutical Reports* 5, 1-7, 1997.

22. Briggs AH, Wonderling DE, Mooney CZ: Pulling cost-effectiveness analysis up by its bootstraps: a non-parametric approach to confidence interval estimation. *Health Economics* 6, 327-340, 1997.

23. Sacristan JA, Obenchain RL: Reporting cost-effectiveness analyses with confidence. *Journal of the American Medical Association* 277, 375, 1997.

24. Stinnett AA, Mullahy J: Net health benefits: a new framework for the analysis of uncertainty in cost-effectiveness analysis. *Medical Decision Making* 18, Special Issue on Pharmacoeconomics: S68-S80, 1998.

25. Tambour M, Zethraeus N, Johannesson M: A note on confidence intervals in cost-effectiveness analysis. *International Journal of Technology Assessment* 14, 467-471, 1998.

26. Tambour M, Zethraeus N: Bootstrap confidence intervals for cost-effectiveness ratios: some simulation results. *Health Economics* 7, 143-147, 1998.

27. Obenchain RL: Multiplicity in cost-effectiveness inference. *Midwestern Biopharmaceutical Statistics Workshop (MBSW 20.)* Ball State University, Muncie, IN, May 1998.

28. Miller RG Jr: *Simultaneous Statistical Inference, 2nd Edition.* New York: Springer-Verlag, 1981.

29. Fieller EC: Some problems in interval estimation (with discussion.) *Journal of the Royal Statistical Society, Series B,* 16, 175-183, 1954.

30. Cochran WG: *Sampling Techniques.* New York: John Wiley, 1977.

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

32. Efron B, Tibshirani RJ: Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical Science* 1, 54-77, 1986.

33. Efron B, Tibshirani RJ: *An Introduction to the Bootstrap*. New York: Chapman and Hall, 1993.

34. Westfall PH, Young SS: *Resampling-Based Multiple Testing.* New York: John Wiley and Sons, 1992.

35. Obenchain RL: CEplane: calculation and graphical display of bootstrap and Fieller ICER confidence regions and quadrant acceptability fractions. Copyright © *Pharmaceutical Research and Manufacturers of America, PhRMA*, Washington DC. http://www.math.stat.edu/~indyasa/download.htm, 1998.

36. Petzold C: *Programming Windows 95*, 4th Edition. Redmond, Washington: Microsoft Press, 1996.

37. *Graphics Server API, The AutoGraph Library,* version 5.0. Kent, Washington: Pinnacle Publishing Inc., 1996.

38. Guttman, I: Tolerance Regions, Statistical. *The Encyclopedia of Statistical Sciences*, Kotz, Johnson and Read, eds. New York: John Wiley and Sons, Vol. 9, 272-287, 1988.

39. Quesenbury CP: Probability Integral Transformations. *The Encyclopedia of Statistical Sciences*, Kotz, Johnson and Read, eds. New York: John Wiley and Sons, Vol. 7, 225-231, 1986.

40. Obenchain RL: CE_mult: Simulating multiplicity bias in Incremental Cost-Effectiveness Ratios due to reporting only the most favorable outcome over several possible choices of cost or effectiveness measure. http://www.math.stat.edu/~indyasa/download.htm, 1998.

Figure 1. A negative ICER slope: new treatment less costly and more effective than the standard.

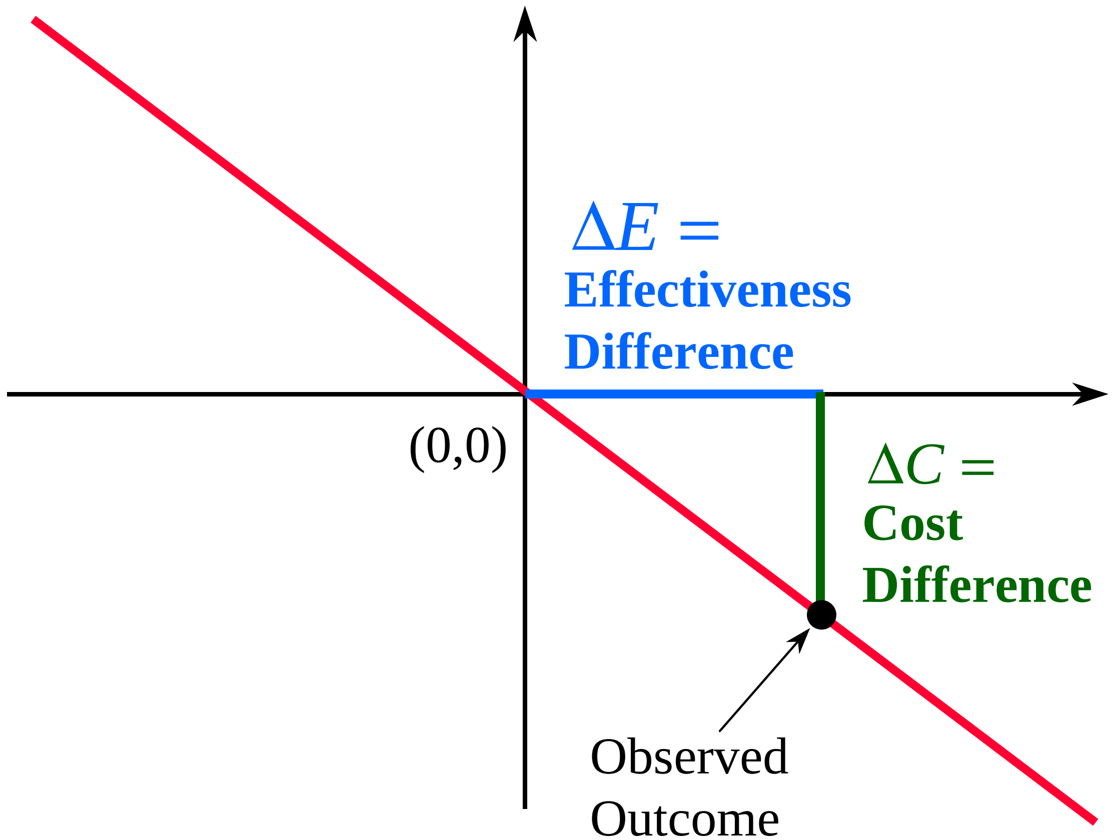


Figure 2. The "Box" Approach.

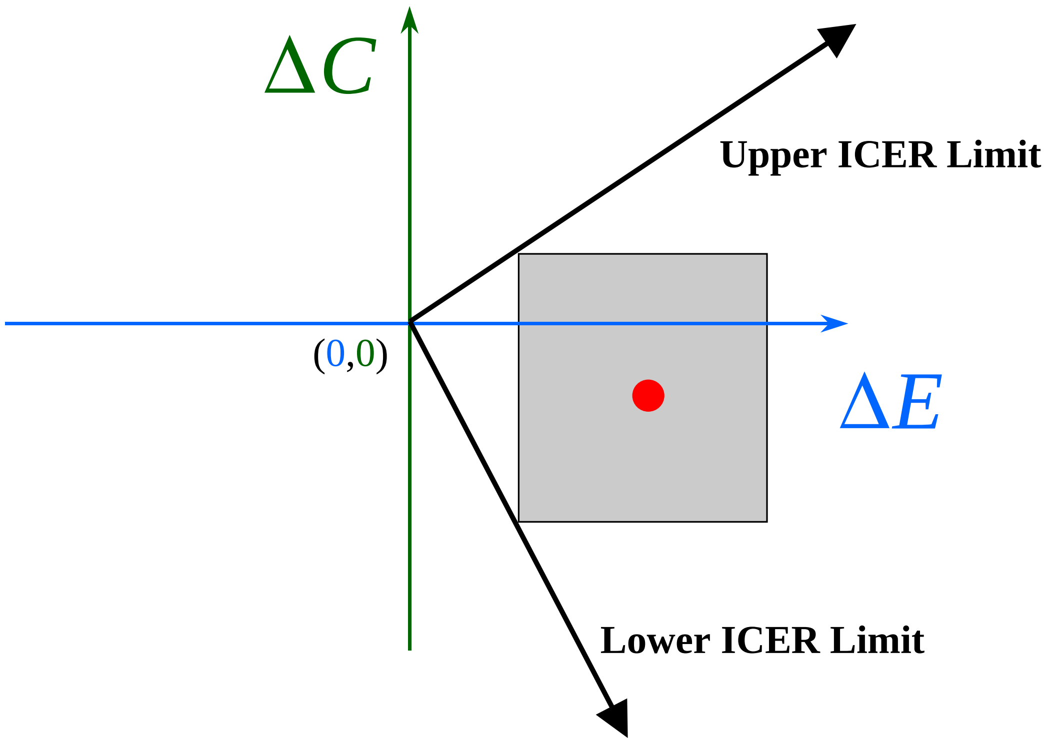


Figure 3. Upper and Lower Fieller ICER Limits.

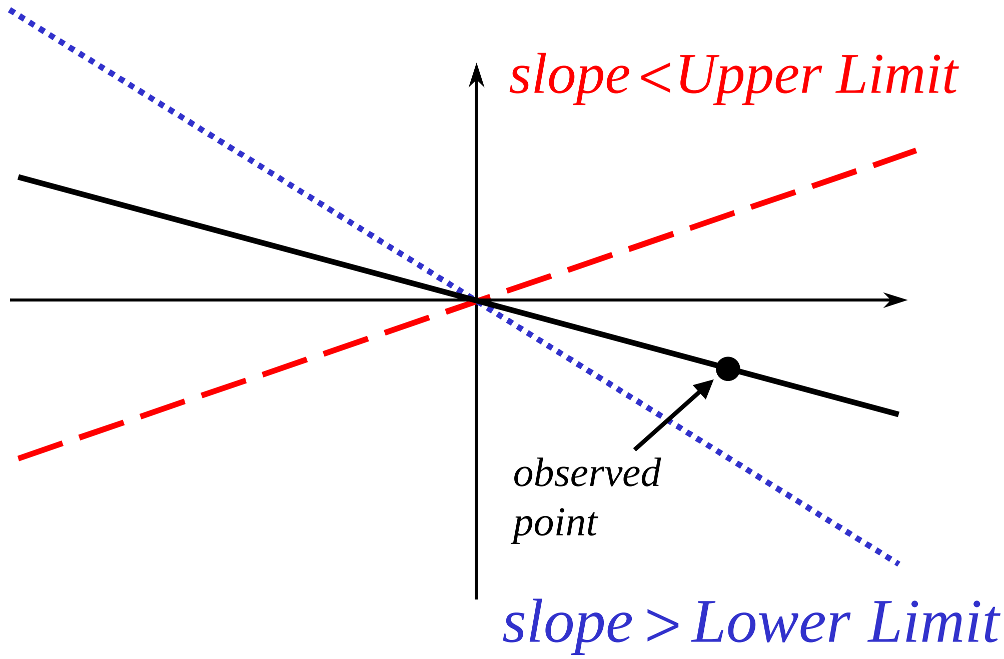


Figure 4. Upper and Lower Bootstrap ICER Limits.

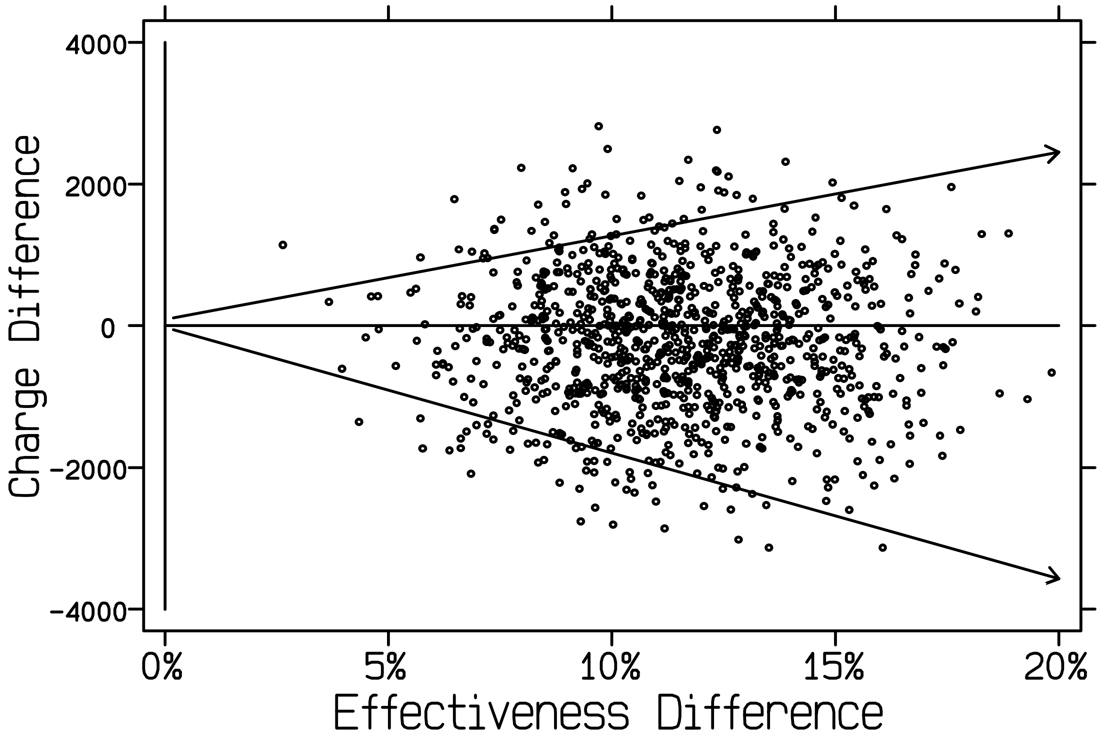


Figure 5. The Five Cost-Effectiveness Regions

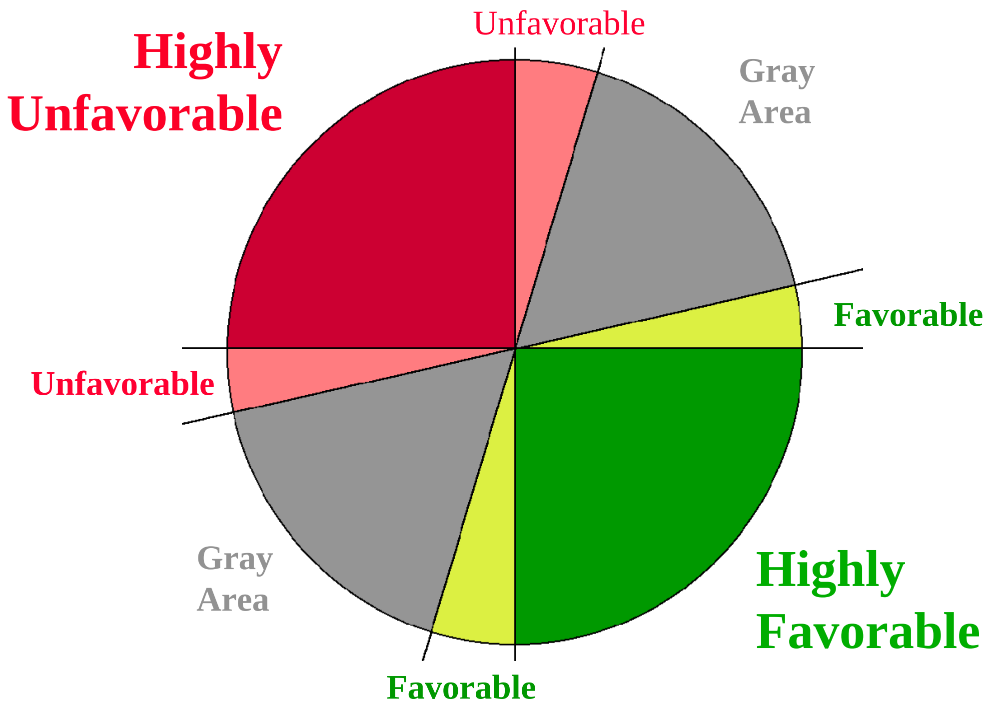


Figure 6. A Contour of Constant Cost-Effectiveness

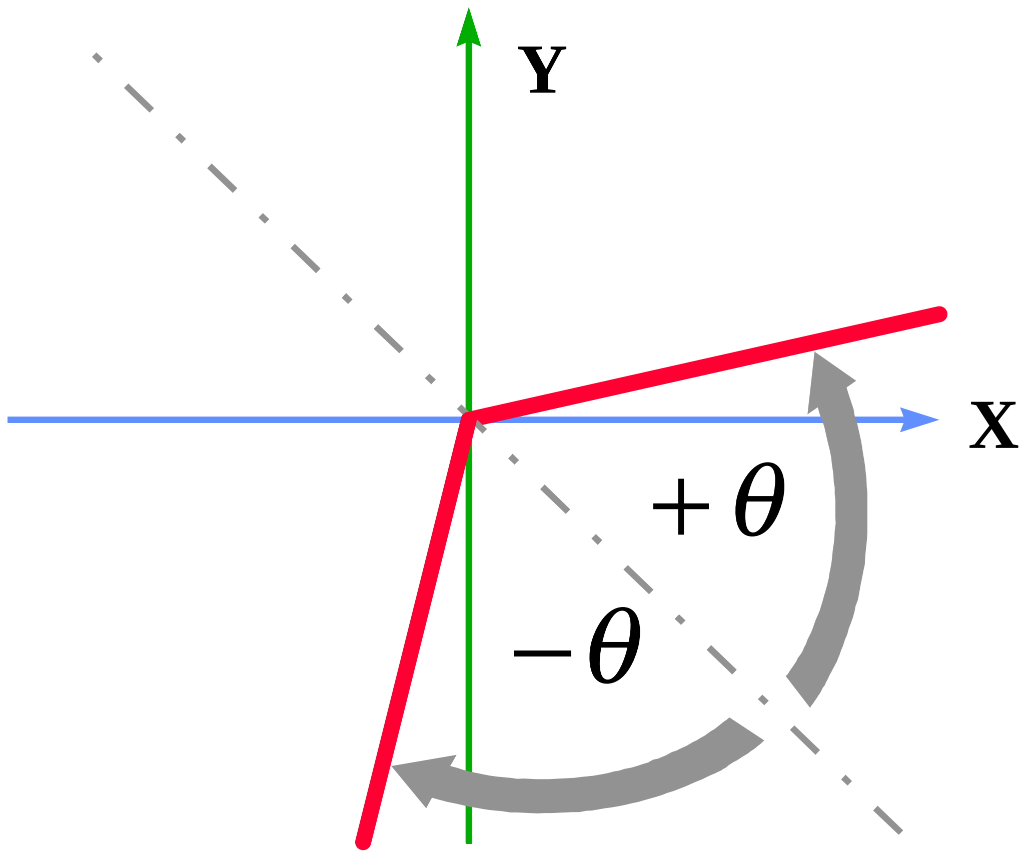


Figure 7. 1,000 Bootstrap Resampled Outcomes

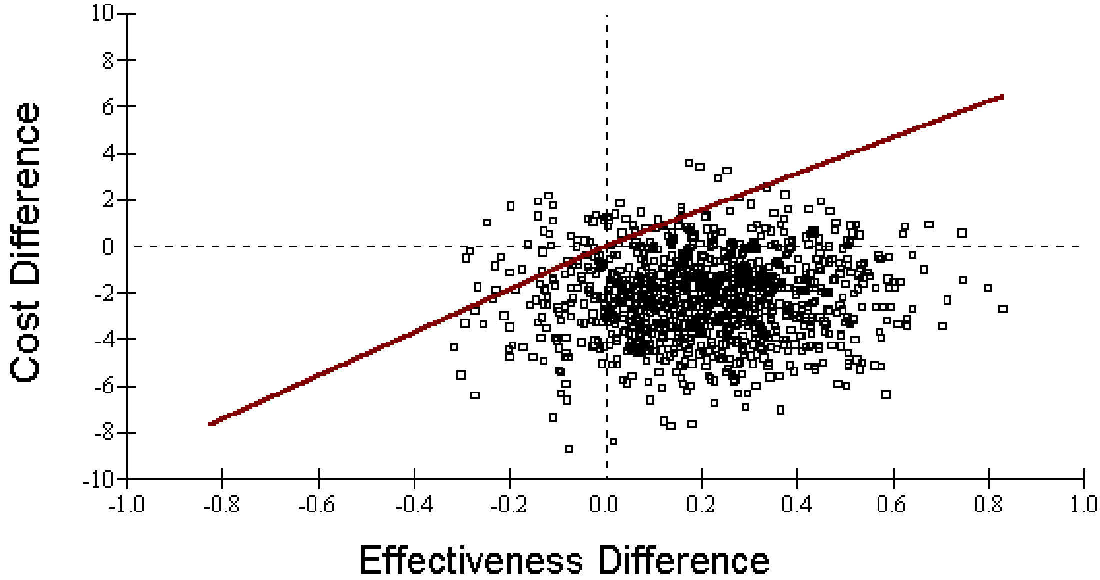


Figure 8. Multiplicity bias in ICER angle Bootstrap confidence intervals.

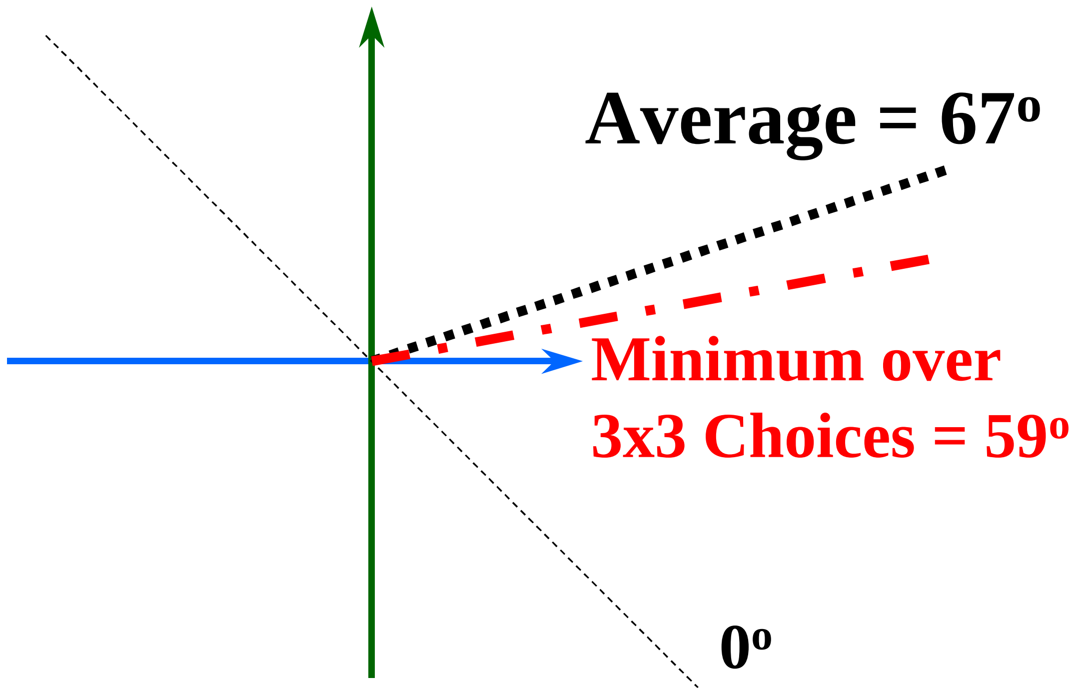


Table 1. Does a health economic study favor T over S?

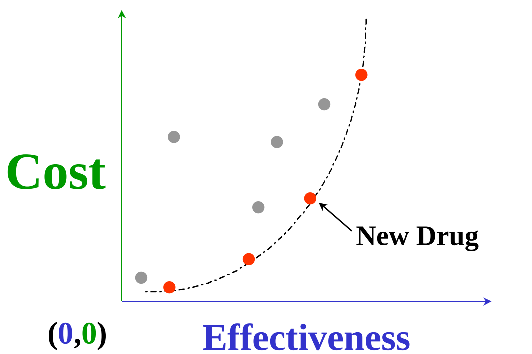
Description	**ICER Angle**	**ICER Slope**	**Cost-Effectiveness Quadrant**
Highly Favorable (Green Quadrant)	$0^\circ \leq \lvert\theta\rvert < 45^\circ$	Negative	(–,+)
Favorable (Yellow Areas)	$45^\circ \leq \lvert\theta\rvert < 60^\circ$	Positive (small or large extremes)	(+,+) or (–,–)
Mixed (Gray Areas)	$60^\circ \leq \lvert\theta\rvert \leq 120^\circ$	Positive (but neither very large nor very small)	(+,+) or (–,–)
Unfavorable (Pink Areas)	$120^\circ < \lvert\theta\rvert \leq 135^\circ$	Positive (small or large extremes)	(+,+) or (–,–)
Highly Unfavorable (Red Quadrant)	$135^\circ < \lvert\theta\rvert \leq 180^\circ$	Negative	(+,–)

Resampling and Multiplicity in Cost-Effectiveness Studies

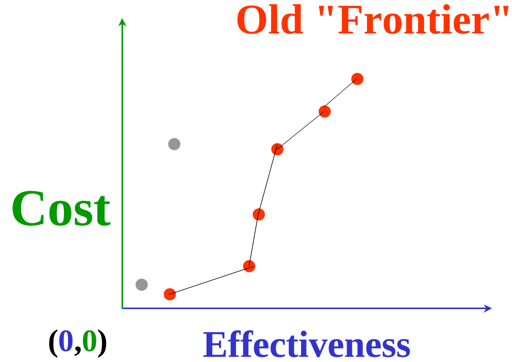
Bob Obenchain, Ph.D.
Lilly Research Laboratories

Accounting for Uncertainty & Bias in Cost-Effectiveness Statistical Inference

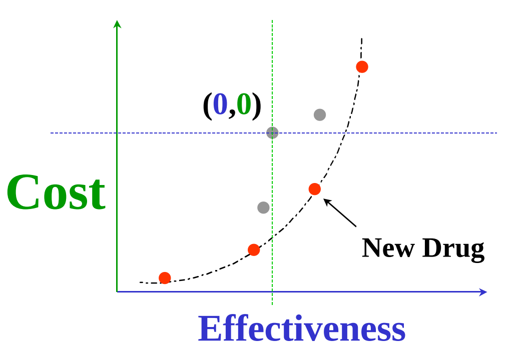
Conceptual Cost-Effectiveness "Frontier"



Old "Frontier"



New "Origin" for Comparison of 2 Treatments...

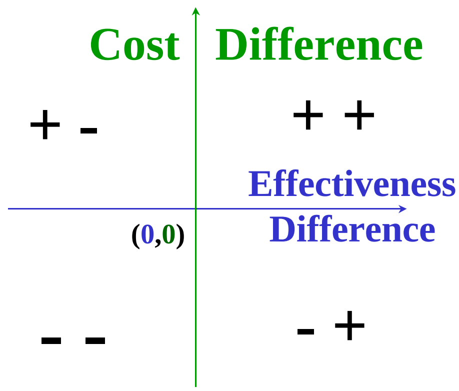


Cost and Effectiveness Average Differences...

$$\Delta C = \overline{C}_T - \overline{C}_S$$

$$\Delta E = \overline{E}_T - \overline{E}_S$$

- T = new treatment
- S = standard treatment

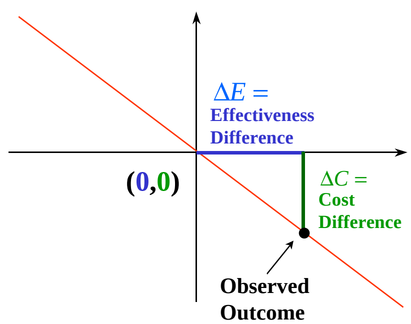


Incremental Cost-Effectiveness Ratio...

$$ICER = \frac{\Delta C}{\Delta E} = \frac{\overline{C}_T - \overline{C}_S}{\overline{E}_T - \overline{E}_S}$$

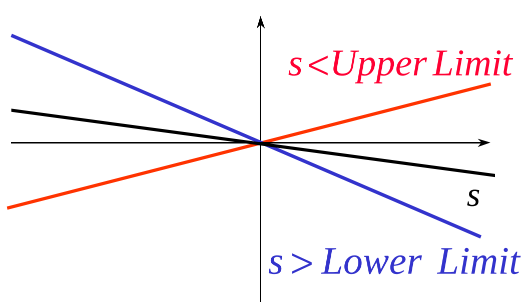
...a difference in mean charges divided by the corresponding difference in mean effectiveness

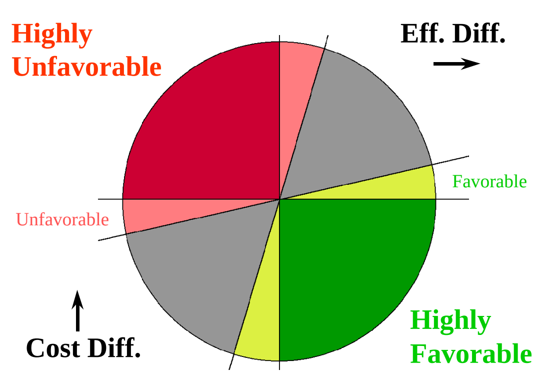
ICER = Slope of line through origin.



- **A “Ratio-Estimate”**
- **Cauchy Distribution under normal-theory with zero expected differences and no correlation between C & E.**
- **Otherwise, Fieller’s Theorem yields a confidence interval.**

Fieller’s Theorem Confidence Limits form a “Bow-Tie” Shaped Region





The ICER slope in not a *Sufficient Statistic* for Cost-Effectiveness Inference!

Thus it is "better" to think in terms of…

ICER angles

Choice of **UNITS** along the two axes of the **Cost-Effectiveness** plane can be critical in defining coordinates...

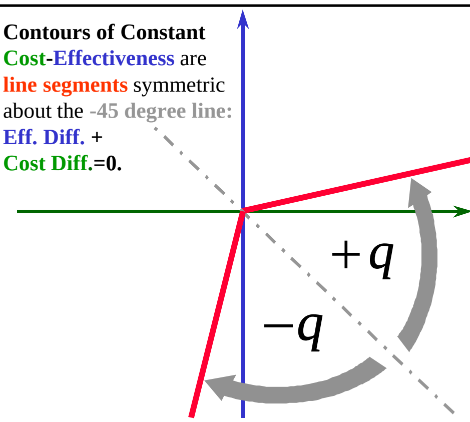
- Polar (radius and angle) rather than Cartesian co-ordinates
- Axis rescaling can transform any point strictly inside a quadrant into any other such point.

Standardize Scaling (units) along axes by dividing each coordinate by the estimated **standard deviation** of individual patient differences...

$$x = \left(\bar{E}_T - \bar{E}_S\right) / \sqrt{Var\left(E_{Ti}\right) + Var\left(E_{Sj}\right)}$$

$$y = \left(\bar{C}_T - \bar{C}_S\right) / \sqrt{Var\left(C_{Ti}\right) + Var\left(C_{Sj}\right)}$$

Contours of Constant Cost-Effectiveness are **line segments** symmetric about the **-45 degree line: Eff. Diff. + Cost Diff.=0.**

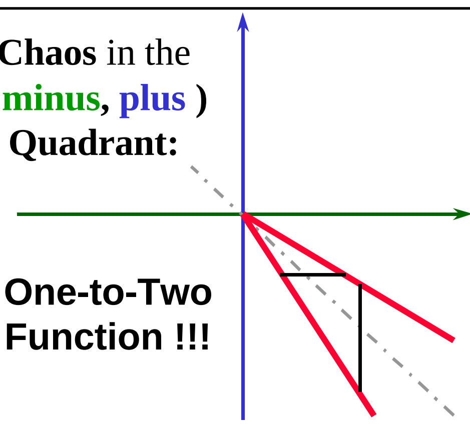


ICER Angle Table

Description	ICER Angle	ICER Slope	Cost-Effectiveness Quadrant
Highly Favorable	$0 \le \lvert q \rvert < 45^\circ$	Negative	(–,+)
Favorable	$45^\circ \le \lvert q \rvert < 60^\circ$	Positive (extreme)	(+,+) or (–,–)
Mixed ("Gray Area")	$60^\circ \le \lvert q \rvert \le 120^\circ$	Positive (neither very large nor very small)	(+,+) or (–,–)
Unfavorable	120° [illegible] $\lvert q \rvert < 135^\circ$	Positive (extreme)	(+,+) or (–,–)
Highly Unfavorable	$135^\circ < \lvert q \rvert \le 180^\circ$	Negative	(+,–)

Chaos in the **(minus, plus) Quadrant:**

One-to-Two Function !!!



Thus, while **negative** ICER slopes are definitely not **"meaningless,"** they certainly can cause **confusion.**

The <u>bottom line</u> is that **ANY** ICER angle in the **(–,+)** quadrant is **extremely good news** for the **NEW** treatment!!!

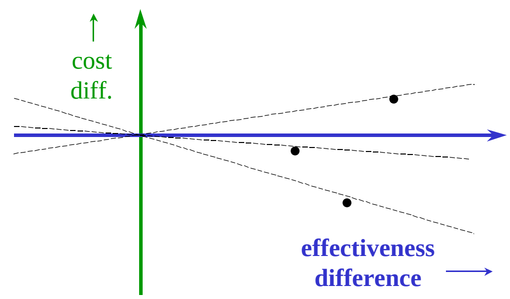
Confidence Regions...

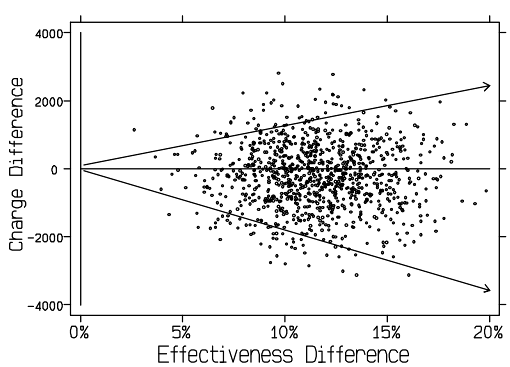
- **"Box Method" (without or with Bonferroni adjustment on overall confidence level)**
- **Normal Theory Ellipsoid**
- **Wedge Shaped Region with Limits expressed as ICER angles (or slopes)**

Bootstrap Analyses...

- repeated random sampling of patients, **with replacement**, within each of the two cohorts. I.E. resample using only **observed** cost and effectiveness **data pairs.**
- literally "repeat" a study over-and-over-again to see how variable results are.

Each Bootstrap Resample generates an ICER angle (or slope) estimate:





Definition One:

The bootstrap 95% confidence region is the pie-shaped segment subtending the smallest total angle at the origin and yet containing 95% of the simulated cost-effectiveness pairs.

Note: minimum subtended angle may be greater than 180°.

Definition Two:

The bootstrap "central" 95% confidence region is the union of two pie-shaped segments, each containing 47.5% of the simulated cost-effectiveness pairs, measured clockwise and counter-clockwise, respectively, from the observed ICER angle.

I.E. "count outward" from observed point

"Minimum Angle" Definition

…great intuitive appeal but highly dependent upon axis scaling.

"Count-Outward" Definition

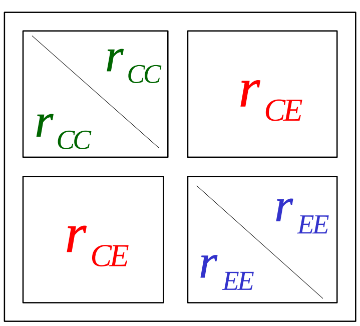
… same order-statistics for all scalings but reporting standardized angles helps in visualization.

And Now

Multiplicity...

What if there are several possible choices for both the numerator cost variable and the denominator effectiveness variable, all possible pairings are analyzed, and only the most favorable result is reported?

Correlation Patterns within Treatments for Interchangeable Cost & Effectiveness Variables



When alternative measures ARE NOT interchangeable, reporting only one analysis may be appropriate.

On the other hand, reporting only one analysis certainly lacks FAIR BALANCE whenever a wide range of outcomes could result.

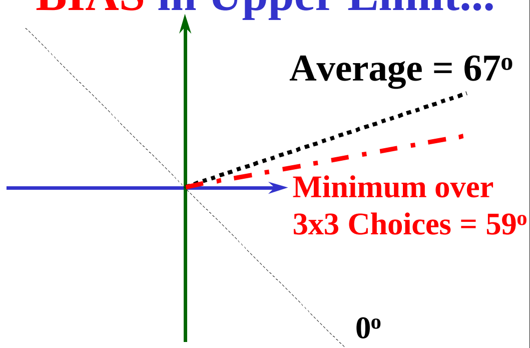
But, when alternative measures are **interchangeable**, reporting **only the most favorable** bootstrap analysis definitely introduces **BIAS.**

Simulation Scenario...

```
5000     number of replications
100      number of patients per treat group
3        number of cost & effect variables
0        random number seed (0 => use clock)
95       central confidence level
...
```

$$r_{CC} = r_{EE} = 0.8;\ r_{CE} = 0.1$$

BIAS in Upper Limit...

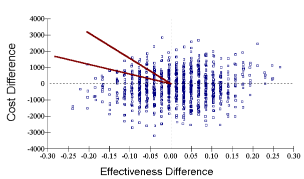


SUMMARY...

Bootstrapping provides extremely relevant information about uncertainty in cost-effectiveness analyses.

But considerable bias could still be introduced by reporting only the most favorable result over several possible choices of variables.

Reporting Options...



Show the Bootstrap scatter !!!

...or Confidence Levels of all 4 Quadrants

```
(-,-) 26.8%
(-,+) 28.8%
(+,+) 26.0%
(+,-) 18.4%
```

Free Software from URL www.math.iupui.edu/~indyasa/download.htm

CEP_9806.EXE: Bootstrap & Fieller's theorem MS Windows application.

CEM_9805.EXE: Simulate bias due to a multiplicity of interchangeable cost and/or effectiveness measures.