

# ICE preference maps: nonlinear generalizations of net benefit and acceptability

Robert L. Obenchain

## Abstract

The Net Benefit (NB) approach to Incremental Cost-Effectiveness (ICE) statistical inference uses a linear function (map) to assign a real valued, numerical preference score to every point on the 2-dimensional ICE plane. We argue that coherent ICE preferences satisfy four intuitive axioms and propose a 2-parameter family of maps that satisfy these axioms and provide highly realistic generalizations of NB. For example, nonlinear maps do not require that returns-to-scale be linear (constant) or that Willingness-To-Pay (WTP) and Willingness-To-Accept (WTA) are both equal to the shadow price of health,  $\lambda$ . In fact, all of our maps have the property that  $\lambda = (\text{WTP} \times \text{WTA})^{1/2}$ . With  $\lambda$  held fixed, this geometric mean relationship shows that WTA must decrease when WTP increases and vice versa. This relationship thus provides not only a polar angular measure of the size of “Bernie’s Kink,”  $\text{WTP} < \text{WTA}$ , but also the theoretical basis for Buckingham’s ALICE curve generalization of acceptability. Finally, we argue that uncertainty about economic preferences expressed by varying  $\lambda$  can totally swamp the statistical uncertainty in patient level data expressed by a wedge-shaped, bootstrap ICE confidence region that does not depend upon  $\lambda$  in the sense that it is equivariant under changes in  $\lambda$ .

## Keywords

ICE outcome pairs, Linear standardization, Preference map, Returns-to-Scale, Willingness-to-Pay/Accept

Financial support for this research was provided entirely by Eli Lilly and Company. The author’s independence in writing and publishing his research is ensured by Lilly’s published Principles of Medical Research. The author is a retired pensioner and a stockholder of Eli Lilly and Company.

## 1 Introduction

Cost-effectiveness analysis has been a hot topic in health services and outcomes research for at least the last decade. The simplest case, where two treatments are being compared head-to-head, is Incremental Cost-Effectiveness (ICE) inference. The tutorial of Briggs and Fenn (1998) provides a good introduction to and review of the extremely wide variety of methodologies that have been proposed to quantify ICE uncertainty. A synopsis of ICE statistical inference is that it addresses a complex 2-sample, 2-variable problem. Specifically, ICE inference examines

differences between two treatment groups using data on two types of outcomes (cost and effectiveness) that may be correlated.

As an aid to visualization in 2-dimensional, real Euclidean space, Black (1990) proposed that the incremental difference (new treatment minus standard treatment) in mean effectiveness,  $\Delta E$ , be plotted horizontally while the corresponding difference in mean cost,  $\Delta C$ , is plotted vertically. To depict uncertainty in a  $(\Delta E, \Delta C)$  point estimate (a bivariate statistic), its bootstrap distribution under resampling of patient outcome pairs with replacement within treatment groups is displayed as a scatter of points on the ICE plane; see Briggs and Fenn (1998, pp. 731–734) for a summary of ICE bootstrapping methodology.

The bivariate bootstrap distribution of uncertainty in  $(\Delta E, \Delta C)$  estimates yields a (univariate) confidence interval for the unknown true expected value of any scalar valued ICE summary statistic. Due to the inherently bivariate nature of the health and financial outcomes being compared, substantial challenges to the use and interpretation of the ICE ratio,  $\Delta C/\Delta E$ , have been noted (Chaudhary and Stearns 1996; Fan and Zhou 2007; Heitjan et al. 1999a, b; Laupacis et al. 1992). Although closely related to the ICE ratio, the Net Benefit (NB) approach (Laska et al. 1999; Stinnett 1999; Stinnett and Mullahy 1998) may have avoided similar criticism because it attempts to quantify overall incremental preference or utility. In NB, constant preference contours (indifference curves) are straight lines on the ICE plane with positive slope,  $\lambda$ . Within the North East (NE) quadrant, this slope can be interpreted as the willingness-to-pay (WTP) a higher cost in return for increased effectiveness; within the South West (SW) quadrant, this same slope is interpreted as the willingness-to-accept (WTA) a less effective treatment in return for lower cost. In other words, NB collects outcomes on the ICE plane into linear equivalence classes (straight lines of constant incremental preference) ordered by a scalar index, the NB.

Unfortunately, the assumption of uniform slope for iso-preference curves is unrealistic in the sense that linear utility has been consistently contradicted in empirical studies (O'Brien et al. 2002; O'Brien and Sculpher 2000; Willan et al. 2001). O'Brien et al. (2002) provide a good introduction to WTP and WTA concepts as well as a highly readable summary of relevant literature. Here we describe a logical foundation for ICE preference quantification (Obenchain 2000, 2001) including a family of two-parameter models that admit complex economic behaviors, such as non-equivalent WTP and WTA. Our axioms are quite general, and our arguments motivating them are both simple and intuitive.

The linear NB map does satisfy all four basic axioms. But our two-parameter family of simple “models” for preference variation across the entire ICE plane provides *nonlinear* generalizations of NB. For example, see Figs. 2a–2c of sub-section 3.6 and our discussions of their interpretation throughout this paper.

Section §2 introduces basic notation and demonstrates that all commonly considered transformations of the ICE plane, including both treatment re-labeling (new versus standard) and axis rescaling (changes in the shadow price of health), are simple linear transformations with a *fixed point* at the ICE origin. With  $\lambda$  held fixed, treatment differences in outcome are first standardized by expressing both  $(\Delta E, \Delta C)$  differences in identical units, either both in cost units

or else both in effectiveness units. Section §3 then discusses our four basic axioms expressed in *standardized units* and introduces our 2-parameter family of “signed-power” maps. Section §4 discusses the returns-to-scale properties of these maps. Section §5 defines how either WTP or WTA is related to the slope of the iso-preference contour that passes through any given point on the ICE plane. Section §6 then discusses how “Bernie’s Kink” (O’Brien et al. 2002; Willan et al. 2001),  $WTP < WTA$ , is related to the “Gray Areas” in the seminal work of Laupacis et al. (1992) and quantifies the angular size of this kink. Section §7 discusses important distinctions between “ALICE” curves and traditional, linear measures of acceptability using a numerical example. Section §8 then shows that economic uncertainty about shadow price,  $\lambda$ , can totally swamp any statistical uncertainty about the true location of  $(\Delta E, \Delta C)$  within the ICE plane. Finally, §9 discusses the advantages and practical implications of nonlinear preferences as well as the need for greater consensus on ICE vocabulary and methodology.

## 2 Basic ICE notation and terminology

### 2.1 ICE outcomes

An ICE outcome is represented by a pair of expected treatment differences, usually expressed in Cartesian coordinates as  $(\Delta E, \Delta C)$ . Here,  $\Delta E$  is a difference in average treatment effectiveness of the form “new” treatment minus “standard” treatment. The underlying effectiveness measurement needs to be defined in such a way that larger (more positive) values of  $\Delta E$  are unambiguously more favorable to the new treatment. The corresponding difference in average per-patient cost,  $\Delta C$ , must be such that smaller (more negative) values are unambiguously more favorable to the new treatment.

### 2.2 ICE transformations

Transformations of ICE outcome coordinates occur quite naturally. For example, interchanging the labels (new and standard) on the two treatments being compared would multiply both  $\Delta E$  and  $\Delta C$  by minus one.

Next, let  $\lambda$  denote society’s fixed “shadow price” for one unit of effectiveness. In other words,  $\lambda$  is a strictly positive *substitution rate* expressed in units of cost per unit of effectiveness. For any specified value of  $\lambda$ , a cost difference of  $y = \Delta C$  is re-expressed in effectiveness units by dividing it by  $\lambda$ . Alternatively, the corresponding effectiveness difference of  $x = \Delta E$  would be expressed in cost units by multiplying it by  $\lambda$ . An arbitrary ICE outcome  $(\Delta E, \Delta C)$  thus gets transformed into either  $(x, y) = (\lambda\Delta E, \Delta C)$  in cost units or else into  $(x, y) = (\Delta E, \Delta C/\lambda)$  in effectiveness units. Either choice represents a *standardized, canonical form* for expected, overall treatment differences that clearly depends upon choice of the fixed numerical value of  $\lambda$ .

### 3 ICE preference maps

Let  $P(x, y)$  denote a real-valued function, called an “ICE preference map,” that determines, as explained below, not only which of two treatments is preferred but also the strength of that preference. Specifically,  $P(x, y)$  can be visualized as a surface defined over the entire 2-dimensional Euclidean plane,  $(x, y)$ , of *standardized* ICE outcomes. Our three primary interpretation conventions (assumptions) for  $P(x, y)$  will be as follows:

$P(x, y) = 0$  means that the  $(x, y)$  pair of treatment differences correspond to no preference whatsoever, either for the new treatment over the standard treatment or vice-versa.

$P(x, y) > 0$  means that the treatment currently called new is preferred over the treatment currently called standard. Strictly positive  $P(x, y)$  values are at least ordinal measures of strength of preference for the new treatment over the standard treatment.

$P(x, y) < 0$  means that the treatment currently called standard is preferred over the treatment currently called new. The *absolute values* of negative  $P(x, y)$  values are at least ordinal measures of strength of preference for the standard treatment over the new treatment.

Table 1 lists our four axiomatic properties of ICE preference maps. When first examining this table, it may be helpful to note that the *linear* preference map,  $\mathbf{NB}(x, y) = x - y$  (Stinnett and Mullahy 1998), clearly satisfies all four of these axioms.

**Table 1. Four axioms of ICE preference**

<b>Indifference and Direction of Preference</b>	$P(x, y) = 0$ when $x = y$ , $P(x, y) > 0$ when $x > y$ , and $P(x, y) < 0$ when $x < y$ .
<b>Monotonicity</b>	$P(x, y) \geq P(x_0, y_0)$ for all $x \geq x_0$ and $y \leq y_0$
<b>Re-labeling</b>	$P(x, y) = -P(-x, -y)$
<b>Symmetry and Anti-symmetry</b>	$P(x, y) = P(-y, -x) = -P(y, x)$

Note that the re-labeling, symmetry and anti-symmetry axioms represent additional restrictions on ICE preferences only when  $x \neq y$ . After all, the  $P(x, x) = 0$  property of the first axiom renders the implications of all other axioms moot for all outcomes with  $x = y$ .

The next 5 sub-sections discuss the meaning and interpretation of these four necessary axioms. The final sub-section then introduces our 2-parameter family of preference maps sufficient to satisfy the axioms and provide realistic, nonlinear generalizations of NB.

### 3.1 ICE indifference and the direction of preference

When  $x = y$ , society receives exactly the difference in effectiveness for which it pays, no more and no less. Therefore, there is no compelling reason to prefer one treatment relative to the other, i.e.,  $P(x, y) = 0$ . If the new treatment is more effective, it is commensurably more costly. If the new treatment is less costly, it is commensurably less effective.

When  $x > y$ , the difference in effectiveness of the new treatment compared to the standard exceeds the cost difference between the treatments. Society thus receives a level of incremental effectiveness worth more than its incremental cost. Therefore, the new treatment is preferred over the standard, i.e.,  $P(x, y) > 0$ .

When  $x < y$ , the difference in cost is larger than the difference in effectiveness. Society thus receives a level of incremental effectiveness worth less than its incremental cost. The standard treatment is then preferred over the new treatment, i.e.,  $P(x, y) < 0$ .

This first axiom is by far the most restrictive of the four considered here. It dictates an infinite, linear interface of standardized slope +1 separating positive from negative ICE preferences. In original units, this is the line  $\Delta C/\Delta E = \lambda$  with slope determined by the shadow price of health, which ICE analysts may wish to deliberately vary to perform sensitivity analyses that turn out to be anything but subtle.

### 3.2 ICE monotonicity

Complete preference orderings of all outcomes of the ICE plane are subject to ongoing research and debate. However, a fundamental property of all sensible preference maps is that  $P(x, y) \geq P(x_0, y_0)$  for all  $x \geq x_0$  and all  $y \leq y_0$ . If the effectiveness of a new treatment is increased at the same time its ultimate cost is decreased, preference for that new product over a fixed standard treatment certainly cannot decrease. Remember that we are assuming that  $x$  has been defined so that larger (more positive) values of standardized effectiveness are more favorable to the treatment currently called new. Similarly,  $y$  must be defined so that smaller (more negative) values of standardized cost are more favorable to the treatment currently called new.

### 3.3 ICE re-labeling

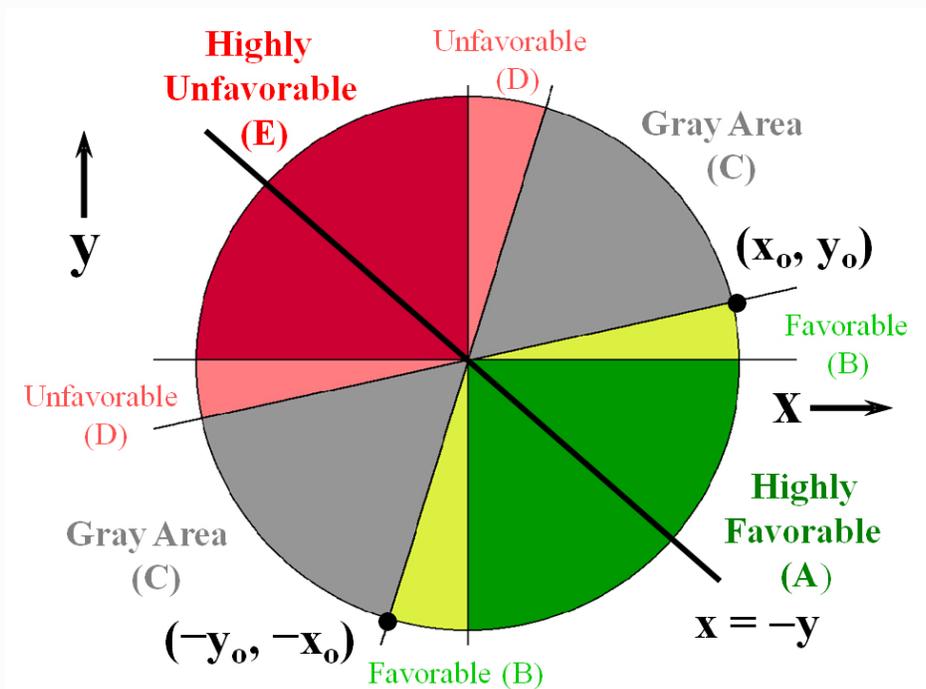
One meaning of  $P(x, y) = -P(-x, -y)$  is that, when reversing treatment labels (new and standard) on a single pair of treatments, the direction of preference is reversed while the strength of preference is preserved. The implications of this axiom are broader in the sense that this same preference equality must also hold when a fixed new treatment is either preferred or not preferred to a fixed standard by a fixed, specified amount. This axiom imposes a form of *fairness* or *even-handedness* upon head-to-head ICE treatment comparisons.

### 3.4 ICE symmetry

Axiom 4 can be expressed in two equivalent ways. Starting with the third axiom plus either form of the fourth axiom, the other form of the fourth axiom follows immediately by simple

algebra. For example, the re-labeling property,  $P(x, y) = -P(-x, -y)$ , can be combined with the symmetry property,  $P(x, y) = P(-y, -x)$ , to yield  $P(x, y) = -P(y, x)$ , which is the anti-symmetry property.

The ICE symmetry axiom requires that preferences for any pair of outcomes symmetrically located relative to the upper-left to lower-right standardized diagonal,  $x = -y$ , of the ICE plane must be *identical*,  $P(x, y) = P(-y, -x)$ . In other words, for any  $(x, y)$  outcome, the alternative outcome of  $(-y, -x)$  must yield the exact same strength of preference in the same direction (new over standard or vice versa). The “named” wedge-shaped segments of the ICE plane discussed in Laupacis et al. (1992) and illustrated here in Fig. 1 were apparently the first published depictions of ICE preference symmetry.



**Fig. 1** The wedge-shaped regions of the ICE Plane discussed in Laupacis et al. (1992) are symmetrically positioned relative to the  $x = -y$  diagonal,

Suppose now that  $(x_0, y_0)$  is any fixed point within the NE quadrant,  $0 < x_0$  and  $0 < y_0$ . As a result,  $(-y_0, -x_0)$  is then a fixed point within the SW quadrant, as illustrated in Fig. 1. Denoting the standardized ICE ratio (slope) corresponding to  $(x_0, y_0)$  by  $s_0 = y_0/x_0 > 0$ , it follows that the standardized slope corresponding to the outcome pair  $(x, y) = (-y_0, -x_0)$  is  $s = y/x = -x_0/-y_0 = +1/s_0 > 0$ . In other words, one immediate implication of the ICE symmetry axiom is that all  $(x_0, y_0)$  and  $(-y_0, -x_0)$  pairings with *equivalent preferences* have *standardized ICE ratios*,  $s = y/x$ , that are *numerical reciprocals or inverses*.

This inverse relationship has considerable intuitive appeal. Within the NE quadrant,  $s = y/x$  is a positive “loss over gain” ratio; the positive numerator represents an undesirable additional cost (loss) while the positive denominator represents a desirable increase in effectiveness (gain). Meanwhile, within the SW quadrant,  $s = y/x$  is a positive “gain over loss” ratio; the negative

numerator represents a desirable cost reduction (gain) while the negative denominator represents an undesirable reduction in effectiveness (loss).

In other words, numerically *small* and positive standardized ICE ratios are desirable within the NE quadrant where they represent loss/gain ratios, while numerically *large* and positive standardized ICE ratios are desirable within the SW quadrant (Heitjan et al. 1999a, b) where they represent gain/loss ratios. In Fig. 1, these two regions are represented by the wedges labeled “Favorable (B)” that are colored yellow-green. By assuring that outcomes within the NE and SW quadrants that yield *equivalent preferences* also yield standardized ICE ratios that are *numerical reciprocals* ( $y_o/x_o$  and  $-x_o/-y_o = x_o/y_o$ ), the ICE symmetry axiom simply formalizes their *basic intuition*.

Within the South East (SE) quadrant [dark green, labeled “Highly Favorable (A)” in Fig. 1],  $s = y/x$  is a negative “gain over gain” ratio; the negative numerator represents a desirable cost reduction while the positive denominator represents a desirable increase in effectiveness for the new treatment. All ICE ratios for outcome differences within the SE quadrant thus represent a distinct preference for the new treatment over the standard.

Finally, within the North West (NW) quadrant [dark red, labeled “Highly Unfavorable (E)” in Fig. 1],  $s = y/x$  is a negative “loss over loss” ratio; the positive numerator represents an undesirable added cost while the negative denominator represents an undesirable reduction in effectiveness for the new treatment. All NW quadrant outcome differences thus represent a distinct preference for the standard treatment over the new treatment.

Next, note that the *linear* preference map,  $NB(x, y) = x - y$ , possesses a purely *optional* property that is much stronger and more restrictive than  $P(x, y) = P(-y, -x)$ . This linear NB preference is constant everywhere on the straight line passing through the points  $(x, y)$  and  $(-y, -x)$ . Again, when one’s preference map is linear, preference is assumed constant on all straight lines ( $x - y = \text{constant}$ ) that are parallel to the lower-left to upper-right diagonal ( $x = y$ ) of the ICE plane.

Finally, note that ICE preference symmetry property,  $P(x, y) = P(-y, -x)$ , does impose an additional restriction besides reciprocal ICE ratios,  $y/x$  and  $-x/-y = x/y$ , for the corresponding outcomes pairings. Namely, all such outcome pairs clearly also have the same ICE radius,  $r = (x^2 + y^2)^{1/2}$ .

### 3.5 ICE anti-symmetry

As previously noted, the ICE anti-symmetry axiom can be viewed as following directly from the re-labeling and symmetry restrictions. In its own right, the anti-symmetry requirement that  $P(x, y) = -P(y, x)$  is quite intuitive. It requires symmetry in *strength* of preferences about the  $x = y$  diagonal. However, this property is called anti-symmetry here because the *direction* of preferences is *reversed* on the two different sides of the  $x = y$  diagonal. After all, when pairs of outcomes of the form  $(x, y)$  and  $(y, x)$  are not on the  $x = y$  diagonal, they are symmetrically located relative to this  $x = y$  diagonal.

### 3.6 Two-parameter ICE preference maps

To generalize the linear preference map  $\mathbf{NB}(\mathbf{x}, \mathbf{y}) = \mathbf{x} - \mathbf{y}$ , let us now consider the family of ICE preference maps of the form

$$P(\mathbf{x}, \mathbf{y}) \propto (\mathbf{x}^2 + \mathbf{y}^2)^{(\beta-\gamma)/2} \{\mathbf{x}-\mathbf{y}\}^\gamma, \quad (1)$$

where  $\propto$  means “is proportional to,”  $\beta$  and  $\gamma$  are strictly positive “power” parameters, and the special notation  $\{\mathbf{z}\}^\gamma$  denotes a “signed-power.” Specifically,  $\{\mathbf{z}\}^\gamma$  denotes the product of  $\text{sign}(\mathbf{z})$  [which is +1, 0 or -1] times the absolute value of  $\mathbf{z}$  raised to the power  $\gamma$ . Special care needs to be taken in Eq. 1 because non-integer powers of negative real numbers are generally imaginary; ICE preferences need to be expressed as *real numbers* that may, however, possibly convey only ordinal measures of strength.

It is straight-forward to verify that all ICE maps of form (1) satisfy axioms 1, 3 and 4 of Table 1. In the Appendix, we show that the following range restriction on the ratio of the  $\beta$  and  $\gamma$  power parameters,

$$1/\Omega \leq \gamma/\beta \leq \Omega \quad \text{for } \Omega = (1 + \sqrt{2})^2 \approx 5.828, \quad (2)$$

is necessary and sufficient for maps defined by Eq. 1 to also satisfy axiom 2, ICE monotonicity. The class of two-parameter ICE preference maps satisfying Eqs. 1 and 2 is our proposed “signed-power” family.

Figures 2a–2d depict ICE preference maps using equally spaced *indifference curves* (level curves, iso-preference contours) drawn using the `contourplot()` function within the “lattice” graphics package for R (The R Project for Statistical Computing 2007). Rather *round* nonlinear maps like the one in Fig. 2a result when  $\gamma < \beta$ ; the linear map,  $\mathbf{NB}(\mathbf{x}, \mathbf{y}) = \mathbf{x} - \mathbf{y}$ , of Fig. 2b results when  $\gamma = \beta = 1$ ; and *highly directional* nonlinear maps like those in Figs. 2c–2d result when  $\gamma > \beta$ . Note that the  $\gamma/\beta$  ratio of 0.25 for the map of Fig. 2a is well above the  $1/\Omega \approx 0.1716$  lower limit allowed under restriction (2), while  $\gamma/\beta = 4$  of Fig. 2c is well below the  $\Omega \approx 5.828$  upper limit for maps possessing ICE monotonicity.

## 4 Returns-to-scale

Suppose now that the observed treatment differences in cost,  $\mathbf{y}$ , and effectiveness,  $\mathbf{x}$ , are somehow both multiplied by a strictly positive and finite real valued factor  $f$ . In other words, the observed effectiveness difference of  $\mathbf{x}$  becomes  $f$  times  $\mathbf{x}$ , while the observed cost difference of  $\mathbf{y}$  becomes  $f$  times  $\mathbf{y}$ . The resulting new value of preference in Eq. 1 is then

$$P(f \cdot \mathbf{x}, f \cdot \mathbf{y}) \propto f^{(\beta-\gamma)+\gamma} \cdot [\mathbf{x}^2 + \mathbf{y}^2]^{(\beta-\gamma)/2} \cdot \{\mathbf{x}-\mathbf{y}\}^\gamma \propto f^\beta \cdot P(\mathbf{x}, \mathbf{y}). \quad (3)$$

In other words, for every map in our 2-parameter family, returns-to-scale depend solely upon the  $\beta$  power parameter associated with only the **ICE radius** term in Eq. 1. For more detail on ICE radius, see the first paragraph of the Appendix.

Specifically, returns-to-scale will be:

Decreasing	when $0 < \beta < 1$ ,
Constant (Linear)	when $\beta = 1$ , and
Increasing	when $1 < \beta < +\infty$ .

## 5 Willingness to Pay or Accept

WTP or WTA at any point on the ICE plane is assumed here to be determined by the iso-preference contour that passes through that given point. In fact, we define a (unitless) *standardized* “willingness” rate as being equal to the  $dy/dx$  slope of the tangent to the constant preference contour at point  $(x, y)$ . While this is fully consistent with NB analysis where these contours are straight lines of slope  $WTP = WTA = \lambda$ , variation in “willingness” will be shown to represent an important generalization of NB to encompass *non-linear ICE preference maps*.

Figure 2 (page 10) displays a variety of maps satisfying equations (1) and (2), and thus satisfy all 4 ICE preference axioms of Table 1. The “roundish” iso-preference contours of the map displayed in Fig. 2a will be dismissed as “unrealistic” in sub-section §6.1. Fig. 2b depicts  $WTP/\lambda = WTA/\lambda = +1$  within the NB map. The Figs. 2c-2d depict highly realistic maps that highlight the influence of ICE radius in determining preferences within the SE and NW quadrants of the ICE plane. The “eta” value, listed in all 4 “subtitles” of Fig. 2, denotes values for the  $\eta = \gamma/\beta$  ratio introduced next.

As shown in the Appendix, the *standardized willingness* rate,  $w$ , at  $(x, y)$  for all signed-power ICE preference maps of form (1) is

$$\begin{aligned}
 w(x, y) &= [\beta x^2 + (\gamma - \beta)xy + \gamma y^2] / [\gamma x^2 + (\gamma - \beta)xy + \beta y^2] \\
 &= [1 + (\eta - 1)s + \eta s^2] / [\eta + (\eta - 1)s + s^2],
 \end{aligned} \tag{4}$$

where  $s \equiv y/x$  is the standardized ICE ratio and  $\eta \equiv \gamma/\beta$  is the map “power-parameter ratio.” Clearly,  $\eta$  cannot be negative because  $\beta$  and  $\gamma$  are both strictly positive, but  $\eta$  values strictly  $< 1$  (as in Fig. 2a) are problematic.

Remembering that  $(x, y)$  denotes either  $(\lambda\Delta E, \Delta C)$  in cost units or  $(\Delta E, \Delta C/\lambda)$  in effectiveness units, it follows that  $w(x, y)$  represents either

[a] a non-negative value of  $WTP/\lambda$  when  $\Delta E, x, \Delta C$  and  $y$  are all positive (NE quadrant)

or else

[b] a non-negative value of  $WTA/\lambda$  when  $\Delta E, x, \Delta C$  and  $y$  are all negative (SW quadrant).

Since  $\beta$  and  $\gamma$  are unitless parameters and  $x$  and  $y$  are both measured here in the same units, it follows that  $\eta = \gamma/\beta$ ,  $s = y/x$  and  $w(x, y)$  are all *unitless quantities*.

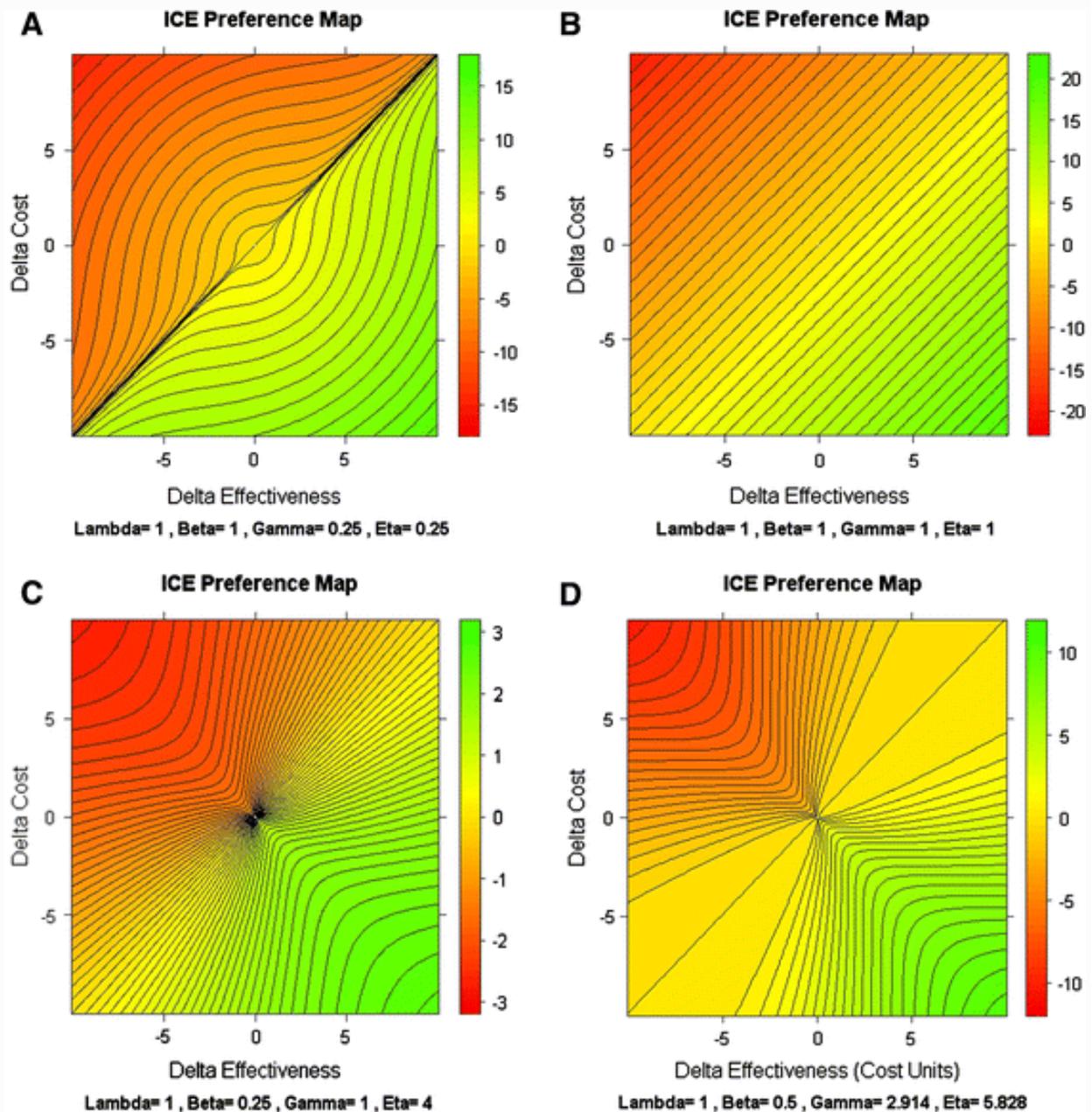


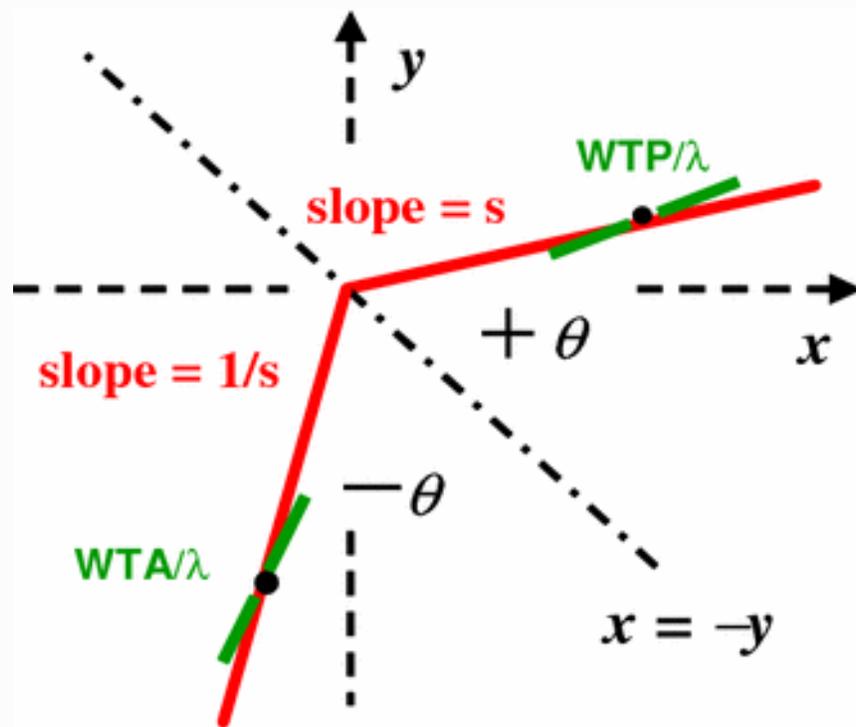
Fig. 2 ICE Preference Maps drawn using the `ICEepmap()` function within the "ICEinfer" R-package (Obenchain 2018)

**Four Special Cases:**

- (a) Level curves of the ICE preference map with  $\beta = 1.0$  and  $\gamma = 0.25$ .
- (b) Linear level curves used in **Net Benefit** analysis ( $\beta = \gamma = 1$ ).
- (c) Level curves of the ICE preference map with  $\beta = 0.25$  and  $\gamma = 1$ .
- (d) Level curves of the ICE preference map with  $\beta = 0.5$  and  $\gamma = 2.914$  ( $\eta = 5.828$ )

Note that the standardized **willingness rate**  $w(x, y)$  is generally different from the standardized ICE ratio,  $s = y/x$ , corresponding to that point; see Fig. 3 below. On the other hand, when  $\eta = 1$ , equation (4) degenerates to simply  $w(x, y) \equiv 1$  ...which is a single, fixed value for all points  $(x, y)$  and all directions  $s = y/x$ .

For any **fixed value of  $\eta$  different from 1**, the standardized willingness rate (4) varies only with  $s$ . In other words, standardized willingness is then constant everywhere along every straight-line trajectory,  $s$ , passing through the origin of the ICE plane ...except at the ICE origin itself. After all, neither  $w(0, 0)$  nor  $s = y/x$  are well defined at the ICE origin! Unlike ICE preferences, given by the  $P(x, y)$  contours shown in Fig. 2, standardized willingness clearly does not vary with ICE radius,  $r = (x^2 + y^2)^{1/2}$ , within the 2-parameter maps of Eq. 1.



**Fig. 3** A pair of **standardized “dual rays”** that (i) contain the same distributions of ICE preferences, (ii) correspond to equal absolute ICE polar angles relative to the  $x = -y$  diagonal, and (iii) have both standardized slopes ( $s = y/x$ ) and standardized **willingnesses** ( $w = WTP/\lambda$  or  $WTA/\lambda$ ) that are numerical reciprocals. The case depicted here corresponds to  $0 < s < w < 1 < 1/w < 1/s$  because  $\eta = \gamma/\beta > 1$  ...as in Fig. 2c.

It can be shown that the maximum and minimum values for  $w(x, y)$  of (4) when  $\eta \neq 1$  are  $[2 + \kappa(\eta^{1/2} - \eta^{-1/2})]/[2 - \kappa(\eta^{1/2} - \eta^{-1/2})]$  for  $\kappa = \pm 1$ ; both limits are non-negative when  $1/\Omega \leq \eta \leq \Omega$ , which is also the restriction (2) that assures ICE monotonicity. The corresponding standardized directions are  $s = (1 + \kappa\eta^{1/2})/(1 - \kappa\eta^{1/2})$ , which are both positive when  $\eta < 1$  as in Fig. 2a and both negative when  $\eta > 1$  as in Figs. 2c–2d. The maximum and minimum values for  $w(x, y)$  are  $+\infty$  and 0 only in the limiting case  $\eta = \Omega \approx 5.828$ , and the pair of directions pointing to these

limits then have (reciprocal) negative slopes of  $s = +1-\sqrt{2}$  and  $-1-\sqrt{2}$  (i.e., ICE Angles of  $\theta = \pm 22.5^\circ$  in Figs. 2d and 3).

The (un-standardized) ICE willingness rate takes on 3 different, simple forms in the following special cases when  $\eta \neq 1$ :

$$\begin{aligned} W(\Delta E, \Delta C) &= \lambda && \text{when } s = \pm 1 \text{ (} x = y \neq 0 \text{ or } x = -y \neq 0\text{),} \\ &= \lambda/\eta && \text{when } s = 0 \text{ (} y = 0 \text{ and } x \neq 0\text{),} \\ &= \lambda\eta && \text{as } s \text{ approaches } \pm\infty \text{ (} x = 0 \text{ and } y \neq 0\text{).} \end{aligned}$$

## 6 Symmetry, dual ICE rays and (WTP, WTA) pairings

The ICE symmetry axiom (Obenchain 2000, 2001) dictates a number of key relationships. Again, the major implication of  $P(x, y) = P(-y, -x)$  is perhaps that these symmetric outcomes have standardized slopes,  $s = y/x$ , that are numerical reciprocals. But Eq. 4 then implies that  $w(x, y)$  and  $w(-y, -x)$  are also numerical reciprocals for all ICE maps satisfying Eq. 1. The “dual rays” of Fig. 3 then consist of all standardized ICE outcome pairs of the form  $(f \cdot x, f \cdot y)$  and  $(-f \cdot y, -f \cdot x)$  with  $(x, y)$  fixed and ray slopes of  $s = y/x$  and  $1/s$ , respectively, as  $f$  increases from 0 to  $+\infty$ . Any such pair of dual rays also contains not only the *same distribution of preference strengths* (as a function of ICE radius) but also the *same direction of preference* (either always new over standard or always standard over new).

In other words, the *un-standardized*  $\Delta C/\Delta E$  ratios corresponding to any such pair of dual rays are  $\lambda s$  and  $\lambda/s$ , respectively, which are not reciprocals unless  $\lambda = 1$ . Similarly, the un-standardized willingness statistics of  $WTP = \lambda w$  or  $WTA = \lambda/w$  are the slopes of the iso-preference contours at all points where they cross one of these two rays. These *un-standardized* willingness slope pairings are also not reciprocals unless  $\lambda = 1$ . Note, furthermore, that these relationships *always hold*, for all choices of returns-to-scale,  $\beta$ , in (3) and all power parameter ratios,  $\eta = \gamma/\beta$ . Finally, choice of  $\lambda$  determines the orientation of the basic pattern of preference variation across the ICE plane and is clearly at least as important as choice of either  $\beta$  or  $\eta$ .

The single, most important implication of the ICE symmetry axiom appears to be that well-matched pairs of WTP and WTA values always have the property that WTP times WTA equals  $\lambda^2$ . In other words, for our signed-power ICE preference maps (or any differentiable ICE preference map satisfying the symmetry axiom), the following relationship will always hold:

$$\lambda = \sqrt{WTP \times WTA} \tag{5}$$

Equation 5 states that the shadow price of health is the **geometric mean** of all well-matched pairs of strictly positive WTP and WTA values. In other words, Eq. 5 shows that WTP and WTA can both vary simultaneously within a *fixed*, nonlinear ICE preference map corresponding to a single *fixed* value of  $\lambda$ . Specifically, relative to choice of shadow price,  $\lambda$ , choices for the values of  $\beta$  and  $\eta$  (or  $\gamma$ ) parameters are *clearly less important*.

## 6.1 An additional “realism” restriction is still needed

Figure 3 portrays an accurate visualization of the numerical ordering between  $w$ ,  $s$  and their reciprocals (i.e.,  $0 < s < w < 1 < 1/w < 1/s$ ) only in cases where the  $\eta = \gamma/\beta$  ratio is  $> 1$  in Eq. 4, as in the “highly directional” nonlinear map depicted in Fig. 2c. Unfortunately,  $\eta < 1$  implies that  $0 < s < 1/w < 1 < w < 1/s$ , which yields  $WTA < \lambda < WTP$  below the  $x = y$  diagonal in the rather “round” maps ( $\eta < 1$ ) like Fig. 2a. This alternative ordering has apparently never been observed in empirical research on WTP and WTA (O’Brien et al. 2002; Willan et al. 2001).

In summary then, only the *nonlinear* ICE preference maps of form (1) with power parameter ratio,  $\eta = \gamma/\beta$ , confined to the finite interval of  $1 < \eta \leq \Omega = 3+2\sqrt{2}$  can be fully realistic. And only the limiting maps with  $\eta = \Omega$  allow willingness (standardized or un-standardized) to vary all of the way from 0 to  $+\infty$ .

## 6.2 The Laupacis view of ICE preferences corresponds to the $\beta = 0$ limit

Figure 1 depicts the limit of our signed-power family of ICE preference maps of Eq. 1 as the returns-to-scale parameter,  $\beta$ , approaches *zero* (while the  $\gamma$  parameter is held fixed at any finite value). In other words,  $\eta$  then approaches  $+\infty$  and the standardized willingness of Eq. 4 becomes  $w = (s + s^2)/(1 + s) = s$  in this limit. While failing to possess ICE monotonicity and allowing negative values for  $w$  in Eq. 4 within the SE and NW quadrants, these limiting maps still have iso-preference curves that correspond to pairs of dual rays with reciprocal slopes. They also *order preferences on ICE polar angle in the exact same way* that they are ordered on all of our  $\beta > 0$  maps for outcomes at the *same ICE radius*.

On the other hand, these limiting ( $\beta = 0$ ,  $\gamma > 0$ ) maps are not very realistic precisely because they yield *zero returns-to-scale*. In other words, they ignore ICE radius as a potential, partial determinant of preference ...especially within the SE and NW quadrants.

## 6.3 Symmetry and literature on (WTP, WTA) pairings

Preliminary standardization, in which  $(x, y)$  represents either  $(\lambda\Delta E, \Delta C)$  in cost units or  $(\Delta E, \Delta C/\lambda)$  in effectiveness units for a fixed numerical value of  $\lambda$ , is essential to be able to express not only standardized directions,  $s = y/x$ , but also standardized willingnesses,  $w$ , as unitless quantities. In turn, unitless quantities become absolutely essential when *reciprocals* are to be compared. After all, if the quantities being compared were not unitless, the statistic and its reciprocal would be expressed in different units, such as  $\$/QALY$  and  $QALY/\$$ , and clearly could not be meaningfully compared!

The unitless ratio,  $WTA/WTP > 1$ , has been proposed (O’Brien et al. 2002; Willan et al. 2001) as the primary measure of the size of the observed “kink” in consumer’s thresholds. Using relationship (5), functions of this  $WTA/WTP$  ratio can be given *simple geometric interpretations* consistent with the Laupacis et al. (1992) visualization of preferences. Specifically, the standardized slope ( $w = s$ ) of any WTP ray in Fig. 1 ( $\beta = 0$ ,  $\eta = +\infty$ ) is  $s = (WTP/WTA)^{1/2} < 1$ , while the standardized slope of the corresponding WTA ray is

$1/s = (WTA/WTP)^{1/2} > 1$ . Similarly, the angular size of the “Favorable (B)” wedges and “Unfavorable (D)” wedges are all equal, again as depicted in Fig. 1. The size of these wedges in degrees is given by multiplying  $\text{ArcTan}(s)$ , measured in radians, by 180 degrees and then dividing by  $\pi$ . Table 2 lists these measures for 10 of the 20 programs or products reported in Table 1 of O’Brien et al.(2002).

Note in Table 2 that the “Favorable (B)” and “Unfavorable (D)” wedges within the NE and SW quadrants are most narrow in the 7 environmental studies and most wide, approaching the maximum possible size of  $45^\circ$ , in some of the 4 safety and 7 experimental studies. For example, when this angle is only  $6^\circ$ , as in the “Trees in park” study, the corresponding Laupacis et al. (1992) “Gray Area (C)” wedges each measure  $78^\circ$ . At the other extreme of  $43^\circ$  for the “Favorable (B)” and “Unfavorable (D)” wedges in the “Job safety (VSL)” study, the “Gray Area (C)” wedges measure only  $4^\circ$  each.

Note that Eq. 5 does not actually establish numerical values for WTP, WTA,  $\eta$ ,  $w$  or  $s$  but only a relationship between WTP, WTA and the shadow price of health,  $\lambda$ . In other words,  $WTP = WTA = \lambda$  is always one possibility. This is the *only possibility* in the purely linear NB formulation ( $\beta = \gamma = \eta = 1$ ). In general, the cases where  $WTP = WTA = \lambda$  and  $w \equiv 1$  because  $\eta = 1$ , including the linear map shown in Fig. 2B, are the only visualizations in which society is willing to pay the full shadow price of health.

**Table 2** WTP-WTA statistics for empirical studies from Table 1 of O’Brien et al. (2002)

	WTA/WTP ratio	Standardized WTP slope	Standardized WTA slope	Angular size of the favorable (B) and unfavorable (D) wedges (degrees)	Angular size of the gray area (C) wedges (degrees)
<b><i>Environmental studies (7)</i></b>					
Elk hunting	Min = 3.2	0.56	1.79	29	32
Deer hunting	12.0	0.29	3.46	16	58
Woodland	25.2	0.20	5.03	11	68
Trees in park	Max = 89.4	0.11	9.46	6	78
<b><i>Health studies (2)</i></b>					
New drug	1.9	0.73	1.37	36	18
Injury (2 wk hospital)	6.4	0.39	2.53	22	46
<b><i>Safety studies (4)</i></b>					
Job (VSL)	Min = 1.1	0.94	1.07	43	4
Car	Max = 3.6	0.53	1.89	28	34
<b><i>Experimental studies (7)</i></b>					
Candy bar	Min = 1.3	0.89	1.12	42	6
Coffee mug	Max = 2.6	0.62	1.62	32	26

Another interesting and realistic possibility is  $WTP < \lambda < WTA$  as in Figs. 1 and 2c–2d as well as in Table 2. These are the cases where a bargain-seeking society or an individual may possibly be willing to pay only somewhat less than an established “fair” full price, possibly forcing providers to accept lower per-item profits while possibly also seeking higher volumes.

## 7 Nonlinear acceptability

The “acceptability curve” graph was originally proposed by Van Hout, Al, Gordon and Rutten (VAGR) in 1994 to portray ICE uncertainty. Given either (i) a parametric, bivariate distribution (normal, say) with mean  $(\Delta E, \Delta C)$  that has been fitted to some observed patient-level data or else (ii) a bootstrap resampling distribution of ICE uncertainty, the VAGR curve depicts the estimated “confidence level” associated with the region to the right or below a rotating *straight line through the ICE origin* that starts out horizontal (representing  $WTP = 0$ ) and rotates counter-clockwise by  $90^\circ$ , ending up being vertical (representing  $WTP = +\infty$ ). In 2004, Fenwick, O’Brien and Briggs (FOB) cataloged as many as 13 “special cases” yielding VAGR curves with quite different shapes, ranging from rather flat, to increasing, to decreasing, to distinctly non-monotone.

An acceptability curve that is always *monotone non-decreasing* results from the unpublished alternative definition of acceptability independently proposed by me in 2001 and by Professor Ken Buckingham of Otago University, New Zealand, in 2003. My freely distributed software (Obenchain 2005, 2007) uses Buckingham’s terminology for *Acceptability Levels In Cost Effectiveness* (ALICE) curves. For any *given and fixed* positive value of  $\lambda$ , the ALICE frontier is defined using a *pair of “kinked” dual ICE rays* (i.e., rays that remain symmetric relative to the  $x = -y$  diagonal while rotating so that their absolute ICE polar angle,  $|\theta|$ , is constantly increasing). Table 3 compares the VAGR and ALICE definitions within all four quadrants of the ICE plane.

**Table 3** VAGR and ALICE definitions of Acceptability

ICE quadrant	VAGR definition	ALICE definition
$\Delta C > 0, \Delta E > 0$ : NE quadrant	Acceptable if $\Delta C/\Delta E < \lambda s$	Acceptable if $\Delta C/\Delta E < \lambda s = WTP$
$\Delta C < 0, \Delta E > 0$ : SE quadrant	All outcomes are acceptable	All outcomes are acceptable
$\Delta C < 0, \Delta E < 0$ : SW quadrant	Acceptable if $\Delta C/\Delta E > \lambda s$ (see Note)	Acceptable if $\Delta C/\Delta E > \lambda/s = WTA$
$\Delta C > 0, \Delta E < 0$ : NW quadrant	No outcomes are acceptable	No outcomes are acceptable

*Note:* Decision Rule (4), page 176, of O’Brien et al. (2002) contains a typographical error; the SW quadrant rule is **not** identical to the NE quadrant rule,  $\Delta C/\Delta E < \lambda s$

In the notation of Table 3, the standardized ICE slope,  $s$ , is a unitless quantity that increases from 0 towards  $+\infty$ , while  $\lambda$  denotes a *given, fixed* value for the shadow price of health. Within the VAGR column of definitions, the product of  $(\lambda \text{ times } s)$  thus denotes a *variable* quantity corresponding to different *common values* for shadow price =  $WTP = WTA$  defining different *linear* VAGR thresholds for acceptability. Within the ALICE column of definitions,

$WTP = \lambda s$  increases with  $s$  within the NE quadrant while  $WTA = \lambda/s$  simultaneously decreases with  $s$  within the SW quadrant, defining a range of *kinked* ALICE thresholds for acceptability, clearly satisfying Eq. 5. This ALICE definition of acceptability agrees with the sum of double integrals in (Willan et al. 2001), page 3255. Technically, interest could even be restricted to the finite range  $0 \leq s \leq 1$  for ALICE curves because  $s > 1$  corresponds to  $WTA < \lambda < WTP$ , again an ordering that has apparently never been observed empirically.

Note in Table 3 that the VAGR and ALICE definitions of acceptability *differ only within the SW quadrant*. VAGR and ALICE curves thus contain the same basic information (displayed using different horizontal axes) whenever the ICE uncertainty distribution attributes zero credibility to the SW quadrant. At the other extreme, where 100% credibility is attributed to the SW quadrant, the VAGR and ALICE curves are again equivalent, but the VAGR curve would then be decreasing while the ALICE curve is increasing, as usual. In other cases, we will see that VAGR curves are non-monotone and biased.

Cases where the ICE uncertainty distribution attributes credibility not only to the SE quadrant but also to the most desirable parts of *both* the SW and NE quadrants are particularly important. We will now consider a numerical example of this “high uncertainty” type.

## 7.1 Comparison of VAGR and ALICE curves for a high uncertainty example

Obenchain et al. (2005) use an example where missing-data-imputation and sensitivity-analyses are needed to make meaningful cost-effectiveness comparisons. The data are from a registration trial (Goldstein et al. 2004) that compares duloxetine, a serotonin and norepinephrine reuptake inhibitor (SNRI), with paroxetine, a standard selective serotonin reuptake inhibitor (SSRI), for treatment of major depressive disorder (MDD). To keep things as simple as possible, we will describe here only head-to-head comparisons between the 91 patients randomized to duloxetine 80 mg/d (40 mg BID) and the 87 patients randomized to paroxetine 20 mg/d. The corresponding bootstrap ICE uncertainty distribution illustrated in (Obenchain et al. 2005), Fig. 2b, is displayed here in Fig. 4 using the “standardized” form explained in the next 3 paragraphs.

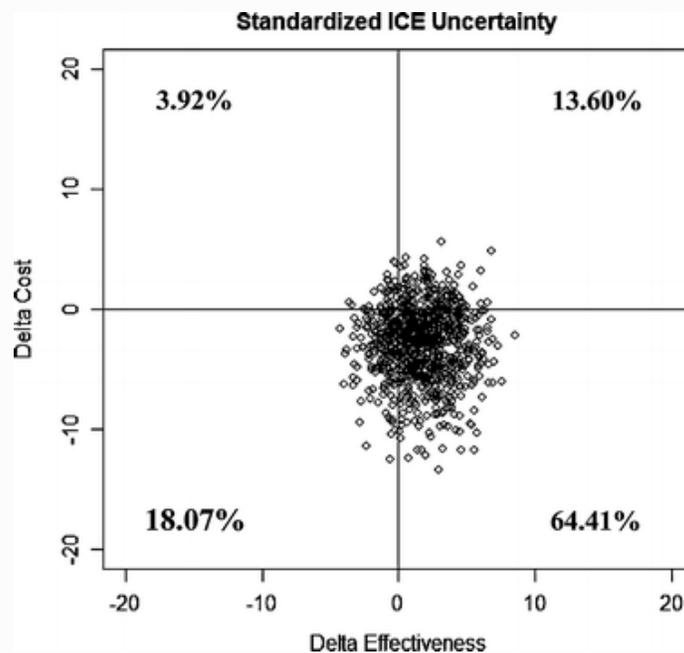
Patient self-reported health-care-utilization above and beyond that provided within study protocol were collected using the Resource Utilization Survey (Copley-Merriman et al. 1992) and quantified using published 1998 \$/unit costs (Schoenbaum et al. 2001) rounded to the nearest multiple of \$50.00. Values for \$/Week were then calculated by multiplying (total accumulated cost) for a patient by 7 and dividing by the (total days of cost accumulation) for that patient. For patients who discontinued from the study, this is Average-Value-Carried-Forward imputation.

Measures of effectiveness in this study were derived from blinded, clinical assessments on the Hamilton Depression Rating Scale (Hamilton 1967). With missing values imputed via MMRM models (Goldstein et al. 2004), the measure of overall effectiveness described here will be “integrated” decrease in HAMD-17 score from baseline to endpoint, which is a (signed) area-under-the-curve measure. (with larger values being more favorable to the new treatment over the standard.)

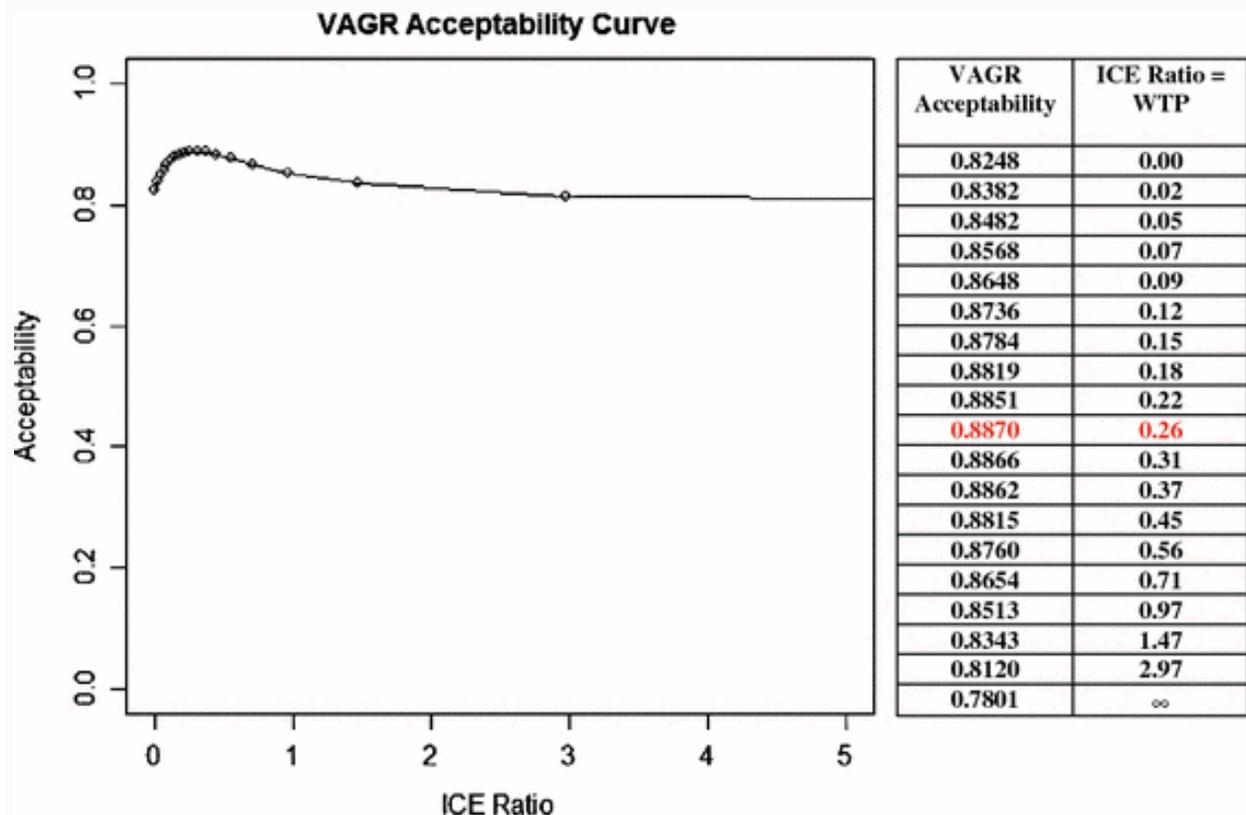
This example is anything but “ideal” for illustrating a situation where one particular value for the shadow price of health,  $\lambda$ , has been well established. Specifically, ICE ratios are expressed here in units of “\$ per Week per unit of Integrated-Decrease-from-Baseline in HAMD-17 score over an 8 week Acute Treatment period for MDD,” hereafter denoted by \$/Week/IDBAT. This scale certainly has no known relationship with any “standard” measure such as \$/QALY. Instead, we have simply assumed here that  $\lambda = \$0.26/\text{Week}/\text{IDBAT}$  and quantified both axes in Fig. 4 in “cost units” of \$/Week. As we will see later (Figs. 5 and 6), there is a weak sense in which this particular choice of  $\lambda$  is most favorable to the “new” treatment.

Fig. 4 shows that our chosen example illustrates a common situation. Relative to the standard treatment (paroxetine), the new treatment (duloxetine) here could represent what is known (somewhat derisively) as a “me too” treatment for MDD. Specifically, the bootstrap distribution of uncertainty here completely covers the ICE origin and lends considerable credibility to at least 3 of the 4 ICE quadrants ...at least when the \$/Week difference in medication *acquisition cost* is zero, as assumed here.

Our objectives in exploring this particular case-study example are two-fold. We intend to convince you not only (i) that the new treatment is at least somewhat cost-effective relative to the standard treatment in cases like that depicted in Fig. 4 but also that (ii) traditional VAGR acceptability curves are biased towards their average value in these critical “high uncertainty” cases. In particular, the all-important lower values of VAGR acceptability are biased upwards.



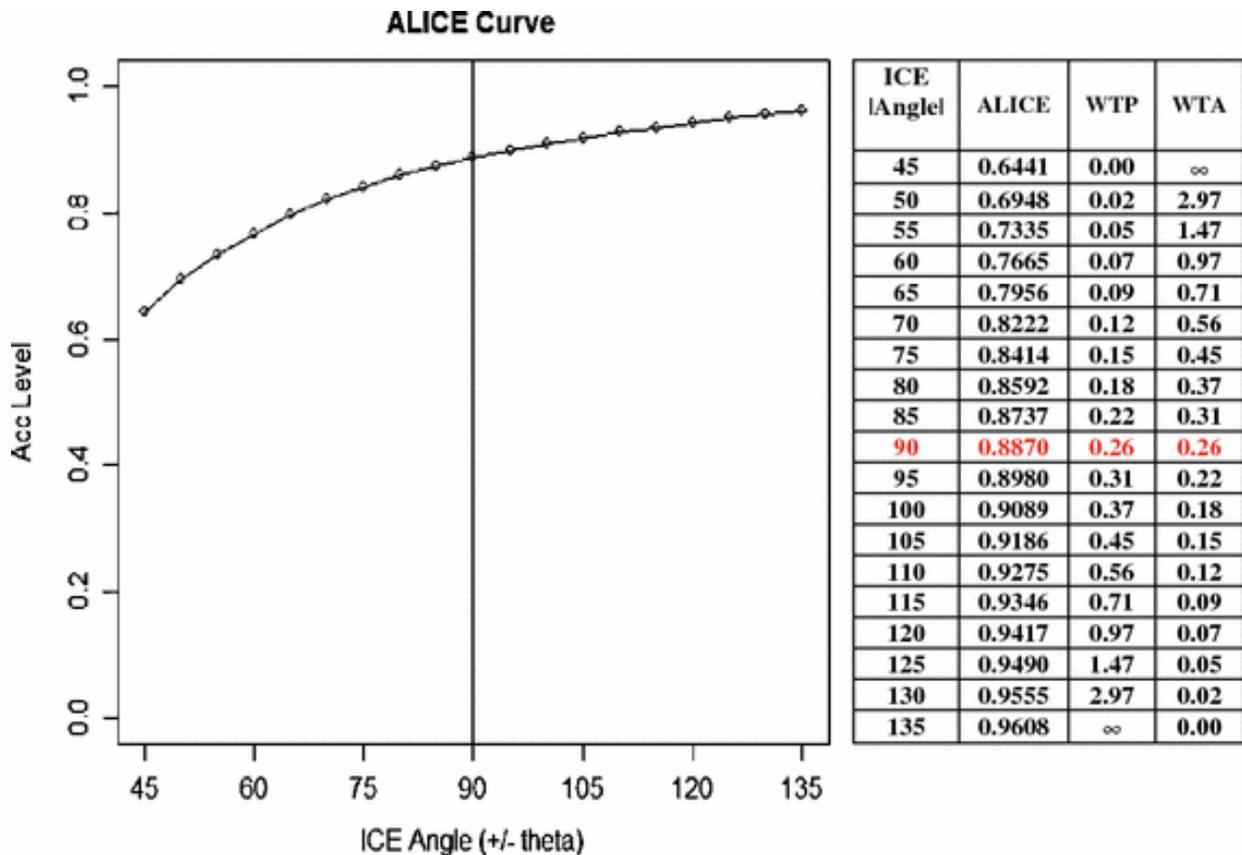
**Fig. 4** First 1,000 out of 25,000 replications quantifying the bootstrap distribution of ICE uncertainty for the “DulxParx” numerical example distributed with my **ICEinfer** R-package (random number generator seed = 12345). Both axes are standardized here to units of \$/Week using an assumed value for the shadow price of health of  $\lambda = 0.26$ . ICE Quadrant Confidence Levels (Obenchain 1999, 2005, 2018) are also displayed.



**Fig. 5** The VAGR Acceptability curve (Fenwick et al. 2004; Van Hout et al 1994) for the numerical example of Fig. 4 is non-monotone. This curve starts high (at 82.48%) because the two “less costly” ICE quadrants have confidence levels of 64.41% (SE) and 18.07% (SW) in Fig. 4. The VAGR curve reaches its maximum of 88.70% for WTP = 0.26, then ultimately drops to 78.01% in the limit as the ICE Ratio approaches  $+\infty$  (not displayed) because the confidence level of the NE quadrant (more effective BUT more costly) is only 13.60% in Fig. 4.

Figure 5 displays the non-monotone VAGR acceptability curve for our high uncertainty example that corresponds to a relatively wide range (from 0 to 5) for the unknown common value of  $WTP = \lambda = WTA$ . Only one numerical value within this wide range of alternative values for  $\lambda$  can correspond to the “true” shadow price of health.

Unfortunately, it is not clear how VAGR or FOB themselves would interpret the information provided by Fig. 5. Because acceptability is always rather high ( $>0.80$ ) over the finite range displayed here and the curve is also rather flat (max – min  $< 0.09$ ), outcomes researchers might conclude that (i) choice of  $\lambda$  is relatively unimportant here and/or that (ii) the odds that the new treatment is more cost-effective than standard are at least 4:1 (i.e.  $0.80/0.20 = 4$ ).



**Fig. 6** As is always the case, the ALICE curve for the numerical example of Fig. 4 is monotone nondecreasing. It starts at the SE quadrant confidence level of 64.41% because this (less costly AND more effective) region is below and to the right of the “kinked” boundary formed by  $WTP = 0$  and  $WTA = +\infty$ . The ALICE curve then increases to 88.70% as WTP increases and WTA decreases until  $WTP = WTA = \lambda = 0.26$ , which corresponds to the only ALICE frontier that is linear (i.e., formed by the pair of rays with ICE Angles of  $\theta = \pm 90^\circ$  as in Fig. 3) rather than “kinked.” The ALICE curve ultimately increases to 96.08%, which is the total confidence of the three (less costly OR more effective) ICE Quadrants lying below or to the right of the “kinked” boundary with  $WTP = +\infty$  (upward vertical ray,  $\theta = +135^\circ$ , in Fig. 3) and  $WTA = 0$  (horizontal ray to the left,  $\theta = -135^\circ$ , in Fig. 3)

Figure 6 displays the corresponding, monotone ALICE curve for this high uncertainty example. By covering only the *finite range* for absolute ICE angles of  $45^\circ \leq |\theta| \leq 135^\circ$ , the full *infinite range* of  $0 \leq s \leq +\infty$  is easily visualized in Fig. 6. Furthermore, the values of the ICE Ratio = WTP displayed in Fig. 5 are now seen to be equally spaced along the horizontal axis of Fig. 6. Finally, Fig. 6 assumes that  $\lambda = 0.26$  \$/Week/IDBAT is the fixed, most relevant value for the true shadow price of health and also allows an overall acceptability level to be determined for all possible budget constraints of the form  $WTP = \lambda s$  with  $s < 1$  plus, by symmetry,  $WTA = \lambda/s$ .

Different choices for  $\lambda$  would yield different ALICE curves. However, all such alternative ALICE curves for a given set of data would have the same starting and ending points at  $|\theta| = 45^\circ$

( $s = 0$ ) and  $|\theta| = 135^\circ$  ( $s = +\infty$ ). Namely, the smallest ALICE value (0.6441 here) will always be the estimated confidence that the new treatment is both “less costly AND more effective” than standard, while the largest ALICE value (0.9608 here) will always be the estimated confidence that the new treatment is either “less costly OR more effective” than standard. These limits correspond to the two key ICE quadrant confidence levels needed to quantify *statistical dominance* (Obenchain et al. 2005).

Note that  $WTP = 0.26$  \$/Week/IDBAT yields the *maximum* VAGR acceptability (of 0.8870) in Fig. 5. Thus, it follows that no ALICE curve for any alternative value of  $\lambda$  (different from 0.26 \$/Week/IDBAT) can yield a larger acceptability level than the value (of 0.8870) displayed in Fig. 6 at  $|\theta| = 90^\circ$  ( $s = 1$ ). We are definitely not recommending that the numerical value of  $\lambda$  used to define ALICE levels be routinely chosen in this way. After all, this particular choice of  $\lambda$  is again, in a weak sense, most favorable to the new treatment!

Rather, our point here is simply that VARG acceptability curves, by using only alternative *linear frontiers* ( $WTP = \lambda = WTA$ ), are badly biased in all high uncertainty cases where the VARG curve ends up being flat or non-monotone. ALICE curves are then much less biased (upwards or downwards) because they use realistic *kinked frontiers*. Even the ALICE curve that is biased upwards *as much as possible*, as in Fig. 6, still suggests that administrative budget constraints (that reduce WTP and, when fair, also increase WTA) can *drastically decrease the overall acceptability level* of new over standard. This reduction is from 0.8870 at  $|\theta| = 90^\circ$  ( $s = 1$ ) to 0.6441 at  $|\theta| = 45^\circ$  ( $s = 0$ ) in Fig. 6, which is a reduction in confidence of 0.243. After all, for two exactly *equivalent treatments*, the VAGR acceptability is expected to always be 0.50 for all values of WTP. The corresponding ALICE level would then also be expected to equal 0.50 at  $s = 1$  ( $|\theta| = 90^\circ$ ), but it would be *expected to drop* to 0.25 at  $s = 0$  ( $|\theta| = 45^\circ$ ) as well as *to rise* to 0.75 at  $s = +\infty$  ( $|\theta| = 135^\circ$ ), at least when cost and effectiveness differences are uncorrelated.

In all cases where the ICE bootstrap uncertainty distribution lends credibility to only one quadrant or to at most two adjacent quadrants, the information contained in VAGR and ALICE curves will really be equivalent (even when the VAGR curve is decreasing due to increases in WTA). In these relatively simple (lower uncertainty) cases, VAGR acceptability is not really biased relative to the corresponding ALICE level.

## 7.2 Advantages and disadvantages of VAGR and ALICE curves

ALICE curves always concentrate attention upon only the *uncertainty within the available data* supporting an ICE policy decision rather than upon any uncertainty about  $\lambda$  itself. Whenever a VAGR curve is non-monotone, it is actually also depicting additional uncertainty about  $\lambda$  itself. When a VAGR curve is monotone, Table 3 shows that it can be reinterpreted as corresponding to a fixed value of  $\lambda$ . For example, when a VAGR curve is non-decreasing, it can be reinterpreted as displaying the uncertainty associated with values of WTP less than any value of ICE Ratio =  $\lambda$  within the plotting range. When a VAGR curve is non-increasing, it can be reinterpreted as displaying the uncertainty associated with values of WTA larger than any value of ICE Ratio =  $\lambda$  within the plotting range.

On the other hand, it is difficult to appreciate how a VAGR acceptability curve or an ALICE curve could be viewed as being a better graphical summary of ICE uncertainty than the bootstrap scatter itself! Given a scatter of bootstrap ICE uncertainty outcomes,  $(\Delta E, \Delta C)$ , it will sometimes be fairly easy to visualize the corresponding VAGR or ALICE curve(s). The inverse problem of visualizing a bootstrap uncertainty scatter from its VAGR or ALICE curve is much more difficult. Specifically, all information about ICE radius (and thus returns-to-scale) has been discarded in the VARG and ALICE formulations.

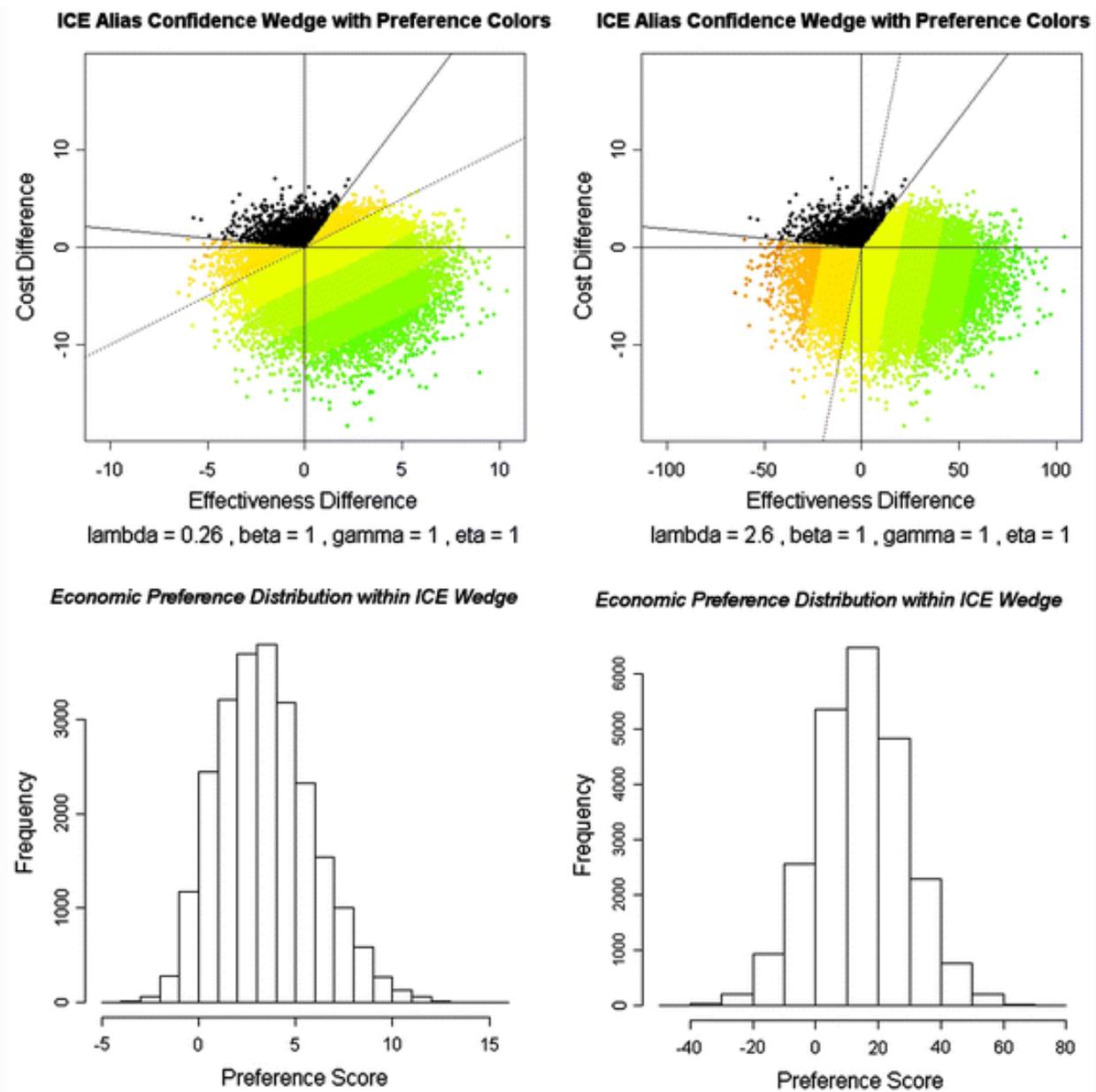
## 8 Statistical uncertainty and economic preference variation

To this point, we have concentrated upon the more “desirable” features of ICE preference maps. It’s important to note that these positive aspects of preference quantification hold within a context where  $\lambda$  can be held fixed, as in Eq. 5, while both WTP and WTA are allowed to vary. In contexts where  $\lambda$  itself is deliberately varied, the strong implications of our first ICE preference axiom lead immediately to disturbing contradictions about economic preferences that totally dominate or “swamp” any statistical uncertainty observable from patient level outcomes data.

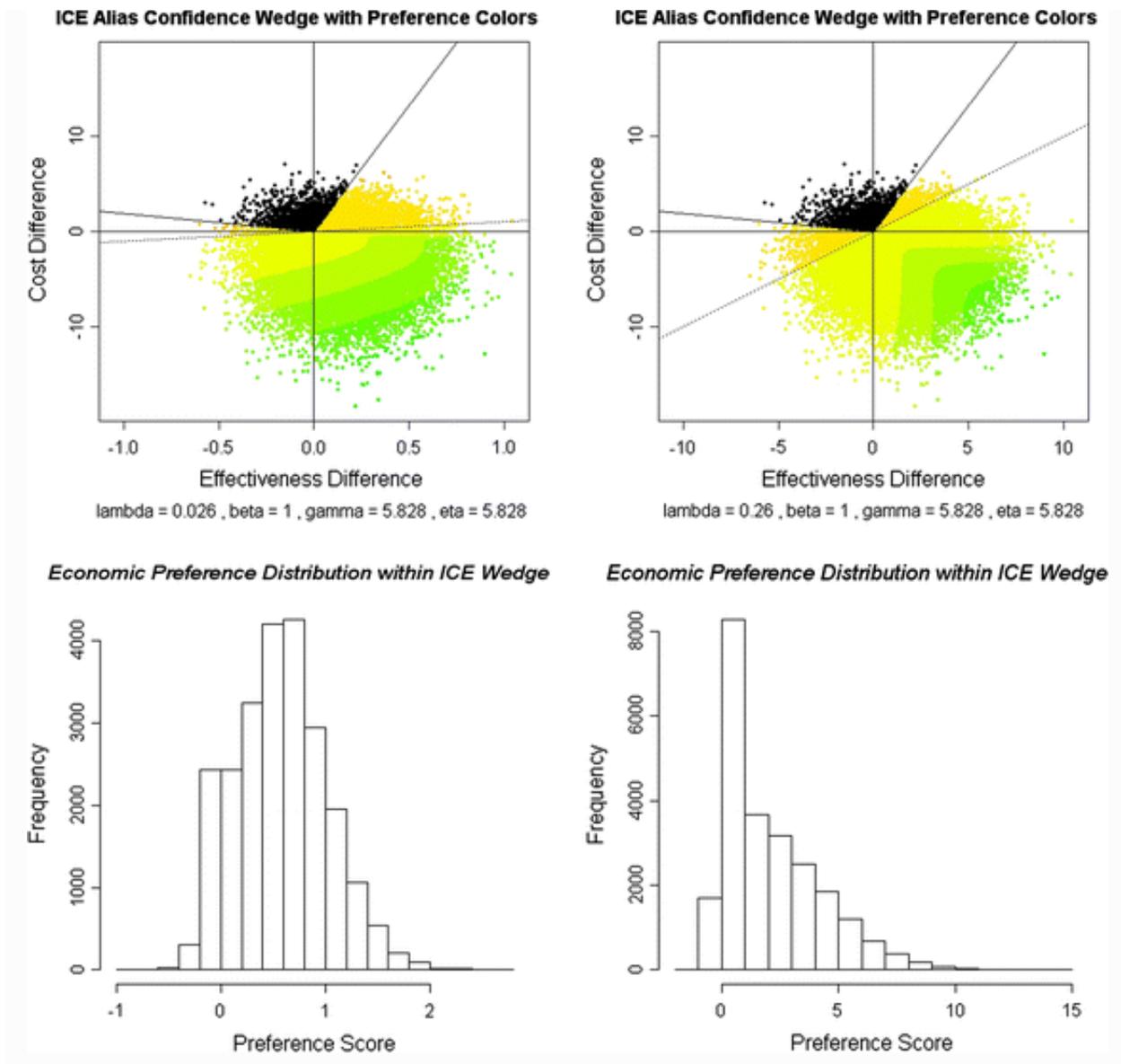
There is a well defined sense in which  $\lambda$  is little more than a nuisance parameter in ICE statistical inference. The wedge-shaped “count outwards” bootstrap confidence region (Obenchain 1999) is independent of choice of  $\lambda$  in sense that it is “equivariant” (commutative) under changes in  $\lambda$ . Indeed, the entire bootstrap distribution of ICE uncertainty itself has this equivariance property. When patient level outcomes are transformed using alternative values for  $\lambda$ , the key practical implication of equivariance is that every  $(x, y)$  point within the bootstrap ICE uncertainty scatter always remains either inside, on or else outside of the ICE *ray confidence limits* computed via the “count outwards” algorithm (Obenchain 1999). This point is dramatically illustrated in Figs. 7 and 8 for the high-uncertainty Dulx-Parx example introduced in §7.1.

Unfortunately, Figs. 7 and 8 also illustrate, by coloring bootstrap ICE outcomes within the equivariant confidence wedge with alternative preferences, that deliberately varying  $\lambda$  leads to incoherent ICE evaluations.

Because the “count outwards” confidence wedge (Obenchain 1999) displays equivariance under changes in  $\lambda$ , we contend that it quantifies only the *statistical uncertainty* within the two samples of patient level cost and effectiveness data about where the unknown true  $(\Delta E, \Delta C)$  outcome falls on the ICE plane. We do not wish to imply that sensitivity analyses concerning choice of  $\lambda$  should not be performed. However, we do think that health services researchers need to be much more aware of the extent of the distinct trauma injected into evaluations of economic preference by deliberately varying the numerical shadow price. After all, *economic uncertainty* about  $\lambda$  is quite separate from the quantifiable *statistical uncertainty* derived from patient level measurements of effectiveness and cost.



**Fig. 7** Note that the bootstrap “count outwards” wedge-shaped 95% confidence region for the high-uncertainty Dulx-Parx example has a central polar angle of about  $237^\circ$ . This region is equivariant under changes in Lambda; it appears to be essentially unchanged between the left-hand and right-hand panels above, where Lambda increases by a factor of ten but the scale along the horizontal axis (effectiveness in cost units) also increases by a factor of ten. The two bootstrap scatters depicted here are not exactly identical because different initial random number “seeds” were used. Finally, note that when the confidence wedges are colored with linear NB preferences ( $\beta = \gamma = \eta = 1$ ) additional economic variation has been injected and cannot be ignored. Note the resulting pair of relatively “contradictory” NB preference distribution histograms depicted in the two lower panels



**Fig. 8** Here are two more visualizations of the bootstrap “count outwards” wedge-shaped 95% confidence region for the high-uncertainty Dulx-Parx example of Fig. 7. We again see “equivariance” as  $\lambda$  increases from 0.026 (left side) to 0.26 (right side) because the x-axis scaling again increases by a factor of 10. This time the confidence wedges are colored using highly-directional nonlinear economic preference maps ( $\beta = 1$ ,  $\gamma = \eta = 5.828$ ), and we again see that changing  $\lambda$  can inject additional economic variation that cannot be ignored. The two corresponding relatively “contradictory” preference distribution histograms are again depicted in the two lower panels

Because all ICE preference maps depend fundamentally upon choice of  $\lambda$ , and possibly also upon choice of returns-to-scale,  $\beta$ , and preference variation (shape) parameters,  $\gamma$  or  $\eta$ , we maintain that ICE preference maps are much more useful in *interpreting* the meaning of an equivariant ICE confidence region, as in Figs. 7 and 8, than in *defining* any such region.

## 9 Conclusions

There are two senses in which our efforts to consider nonlinear ICE preferences that are more realistic than linear NB have failed to circumvent the basic shortcomings of all ICE maps. First, all ICE maps attempt to reduce the effective dimensionality of the ICE decision space from two dimensions, (effectiveness, cost), to only one dimension ...a scalar (possibly ordinal) measure of overall preference. Proponents of linear NB argue that confidence intervals for  $(x - y)$  differences are easier to construct and interpret than confidence intervals for standardized ICE slopes,  $s = y/x$ , or (unstandardized) ICE Ratios,  $\lambda \times s$ . The reality is simply that NB confidence intervals are based upon overly simplistic and clearly unrealistic assumptions (like  $WTP = \lambda = WTA$ ) that also make them much easier to *misinterpret*.

Second, due to the first axiom of ICE preferences, all ICE maps (linear or nonlinear) are much too sensitive to choice of  $\lambda$  to allow outcomes researchers to entertain a wide range of mutually exclusive and contradictory alternative values for the shadow price of health. While Figs. 7 and 8 illustrate that this practice can always inject unwanted incoherence, the effect is even more obvious in those situations where the central polar ICE angle of the wedge is less than  $90^\circ$  and the wedge lies mostly within the NE or SW quadrant.

On the other hand, our two-parameter, nonlinear ICE maps do illustrate some profound new relationships between basic ICE concepts.

### 9.1 Practical implications of the “link” function

As illustrated in Table 3, the “link” function,  $\lambda = (WTP \times WTA)^{1/2}$ , justifies and quantifies the “kink,”  $0 \leq WTP < WTA$ , in empirically observed consumer’s thresholds. Perhaps even more importantly, this link provides an *objective way to determine*  $\lambda$ .

Specifically, health services researchers could simply elicit pairs of matched WTP and WTA values that may vary from patient to patient within a specified disease state. These researchers would then look for across-patient consistency in the geometric means of these WTP and WTA estimates. In particular, there is no need to express effectiveness in QALYs when eliciting WTP and WTA values to determine  $\lambda$  in this way. Researchers will need to use natural, disease specific units, which simplifies the elicitation process and should improve empirical accuracy. Like traditional views of  $\lambda$  itself, the concept of a QALY is actually quite complex (Johnson 2005). In fact, the link function may well provide the very definition of a “fair” shadow price; all other definitions apparently suffer serious vagaries and shortcomings (Gafni and Birch 2006).

When it comes to actual practical applications of ICE inference, the link function allows both WTP and WTA to vary with  $\lambda$  held fixed, as in the ICE ray frontiers defining ALICE curves. Importantly, the link function dictates that WTA must increase if WTP is (arbitrarily) decreased ...an implication that is *profoundly different* from the relatively naïve linear NB perspective where  $WTA = \lambda = WTP$  is assumed.

**Example:** Assume that  $\lambda$  really is the often mentioned value of \$50,000/QALY but that local government authorities or local payers in some particular region insist that \$10,000/QALY is the maximum additional cost that they can possibly agree to pay. This is a simple budget constraint that *does nothing to change the full, fair shadow price of health*, yet it does reduce the local WTP to  $\lambda/5$ . It would then be absolutely unfair and arbitrary to assume that that this sort of budget (maximum cost) restriction should also reduce WTA to  $\lambda/5$ . Instead, the  $\lambda = (\text{WTP} \times \text{WTA})^{1/2}$  shows that the corresponding “fair” value of WTA would thereby *increase* to  $5\lambda = \$250,000/\text{QALY}$ . In other words, only treatments that reduce both cost and effectiveness by at least a net ICE ratio of \$250,000/QALY have as high preferences to society as the desirable treatments that increase both cost and effectiveness by less than a net ICE ratio of \$10,000/QALY.

Finally, no single nonlinear ICE map, implied by explicit numerical choices for the  $\beta$  and  $\gamma$  power parameters in Eq. 1, really needs to be singled out for preferred use. All *symmetric, differentiable maps* necessarily satisfy the same geometric-mean relationship,  $\lambda = (\text{WTP} \times \text{WTA})^{1/2}$ , at all standardized outcome points  $(x, y)$  and  $(-y, -x)$ .

## 9.2 Practical implications for ICE confidence (or tolerance) regions/intervals

In the “intervals or surfaces?” terminology of Briggs and Fenn (1998), ICE inference methods clearly need to be based upon 2-dimensional confidence regions (surfaces) rather than upon an infinite family of at least partially self-contradictory confidence intervals for overall preference that result from deliberately varying  $\lambda$  (Laska et al. 1999; Stinnett 1999; Stinnett and Mullahy 1998). There is no current consensus about how geometrically simple (easy to define in written text) the boundary of an “ideal” ICE confidence region needs to be.

What is clear is that wedge-shaped confidence regions (Briggs and Fenn 1998; Chaudhary and Stearns 1996; Cook and Heyse 2000; Obenchain 1997, 1999; Willan et al. 2001) have the potential to focus attention upon meaningful sub-regions of the ICE plane and to suggest clear preference-based actions. Specifically, the counter-clockwise (upper?) and clockwise (lower?) limiting ICE rays defining a wedge-shaped confidence region also define a confidence interval for the ICE ratio ( $s = y/x$  or  $\Delta C/\Delta E$ ).

When applying Fieller’s theorem (Chaudhary and Stearns 1996), the computed limits may be imaginary for high levels of confidence, implying that the ICE ratio could then be any positive or negative numerical value. For example, the highest confidence level for which Fieller limits exist in the high-uncertainty Dulx-Parx example is 76%, with a counter-clockwise upper limit of slope zero and a clockwise lower limit also of slope zero. In fact, the equivariant bootstrap ICE confidence region with 76% confidence also has a central polar angle of almost  $180^\circ$  and consists of essentially the entire SE and SW quadrants, i.e., all positive or negative ICE ratios.

When the computed Fieller limits are real, they are the slopes of a pair of straight lines through the ICE origin, and the analyst still must determine which two of the resulting four ICE rays define the confidence wedge of interest. The correct choice becomes clear once the rays are plotted on the ICE plane along with the observed ICE outcome pair,  $(x, y)$  or  $(\Delta E, \Delta C)$ ; the correct pair of ICE rays then consists of the two rays closest to the observed ICE outcome point, counter-clockwise and clockwise, respectively.

When applying bootstrap methods (Briggs and Fenn 1998; Cook and Heyse 2000; Obenchain 1997, 1999; Willan et al. 2001), wedge-shaped confidence regions for all levels of confidence always exist. On the other hand, if a 95% confidence wedge were to occupy anywhere near 95% of the full ICE plane in terms of polar angular measure (i.e.,  $0.95 \times 360^\circ = 342^\circ$ ), that region would certainly not be very meaningful or interesting. For the high uncertainty Dulx-Parx example displayed in Figs. 7 and 8, the 95% confidence wedge subtends an ICE polar angle of  $237^\circ$ , which is only 65.8% of  $360^\circ$  and thus is at least somewhat restrictive and informative.

Furthermore, the minimum and maximum values of ICE radius observed for bootstrap outcomes falling strictly within a wedge-shaped ICE confidence region are easily computed. These ICE radii are expressed in the same units as both  $x$  and  $y$  (either effectiveness units or cost units). In high-uncertainty cases, the minimum observed ICE radius will be essentially zero, but the maximum will always be finite. In any case, a (strictly bounded) “wiper-blade” shaped ICE confidence region with a lower nominal confidence percentage than the original wedge-shaped region can be defined by counting “inward” a specified number of ICE *radius order statistics*, thereby decreasing the maximum ICE radius and/or increasing the minimum ICE radius.

Alternatively, a bootstrap ICE confidence wedge can be converted into an interesting ICE tolerance wedge or an ICE ratio tolerance interval (Obenchain 1999) by simply including a few additional ICE angle order statistics within the wedge. For example, suppose that 25,000 bootstrap replications are computed (default value in Obenchain 2005, 2007) and that ICE angle order statistics are sorted around a full circle centered at the ICE origin. Any set of  $0.95 \times 25\text{ K} = 23,750$  consecutive such order statistics then constitutes a 95% confidence wedge in the high-uncertainty cases where the ICE origin falls within the convex hull of the bootstrap ICE uncertainty scatter. The equivariant bootstrap confidence wedge for 25 K replications always results from counting outwards (counter-clockwise and clockwise, respectively) 11,875 consecutive ICE angle order statistics from the observed ICE ratio (i.e., the sample which uses the observed outcomes for each patient exactly once). The corresponding equivariant bootstrap tolerance wedge for 25 K replications that includes *at least 95% of the entire ICE uncertainty distribution* with 95% confidence then results from counting outwards 11,904 consecutive ICE angle order statistics from the observed ICE ratio (Obenchain 1999).

In view of the above non-parametric results, we contend that “count outwards” bootstrap ICE angle confidence (or tolerance) regions provide robust answers. Furthermore, our experience is that these bootstrap confidence limits are generally in very good agreement with Fieller limits in medium-to-large sample situations when Fieller’s theorem does yield real-valued limits (Chaudhary and Stearns 1996). Unfortunately, simulation studies published before 1999 (Briggs and Fenn 1998) and since (Fan and Zhou 2007) tend to focus on questionable computational algorithms and report somewhat contradictory results. This is clearly an area where much greater consensus is badly needed.

### 9.3 Pragmatic choice of ICE preference map

Suppose the question is: “Are there a few distinctive forms of ICE maps that tend to characterize the full spectrum of fundamentally different preference patterns?” There might then be as few as only *three saliently different types* of ICE preference maps. The Laupacis et al. (1992) map of Fig. 1 (with  $\beta = 0$ ) and the linear NB map (Stinnett and Mullahy 1998) of Fig. 2b (with  $\beta = \gamma = 1$ ) clearly represent two simple and historic forms. Rather than consider a range of values for the  $\eta = \gamma/\beta$  shape parameter of realistic, nonlinear ICE preferences, it makes sense to concentrate upon the extreme maps with shape  $\eta = \Omega \approx 5.828$  of (2) that satisfy the monotonicity axiom and allow willingness, (4), to be any non-negative value, as in Figs. 2d and 8. This third extreme

form of ICE preference map still allows  $\beta$  to be  $<1$ ,  $=1$  or  $>1$  to determine decreasing, constant or increasing returns-to-scale, (3).

The realistic, nonlinear ICE preference maps proposed here *encompass the entire ICE plane* rather than focus attention upon any particular sub-region. It can be quite confusing and counter-productive to, instead, use different basic terminology (Kent et al. 2004) within the NE and SW quadrants. Similarly, the FOB suggestion (Fenwick et al. 2004) to divide the ICE plane up into many sub-regions is tedious and counter-productive.

Better comparisons of alternative methodologies for accessing uncertainty in ICE estimates and better vocabulary for generally communicating uncertainty are clearly needed. To yield realistic and reliable results, bootstrap computing algorithms actually need to be somewhat sophisticated. While use of polar coordinates in ICE inference can be a great help in arriving at the “right” answer, many outcomes researchers and most policy makers are clearly uninterested in this high level of detail. People expect results expressed in Cartesian coordinates or as ratios. Realistic approaches to ICE treatment comparisons must ultimately address the unprecedented challenge of making a truly bivariate (2-dimensional) inference problem meaningful to non-technical audiences.

## Notes

### Acknowledgements

I wish to thank my Lilly colleagues, particularly Gerhardt Pohl and Joe Johnston, for their invaluable comments on numerous versions of these materials. I also wish to thank Ken Buckingham for personal communications about his give/get (loss/gain) motivation for ICE symmetry communicated here in §3.4.

### Open Access

This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

## Appendix: Results derived from the partial derivatives of $(\beta, \gamma)$ ICE preference maps relative to their standardized $(x, y)$ coordinates

The 2-parameter ICE preference maps of Eq. 1 are similar to familiar Cobb-Douglas utility functions. Specifically, the amount of one “good” (to be raised to the positive power  $\beta$ ) is determined by the non-negative ICE radius,  $r = (x^2 + y^2)^{1/2}$ , expressed in the same units as  $x$  and  $y$ , while the amount of the other “good” (to be raised to the positive power  $\gamma$ ) is given by  $(x - y)/r = \sqrt{2} \cos(\theta)$ , which is proportional to the unitless ICE direction cosine. Note that  $\theta$  is well defined when  $r > 0$ , as illustrated in Fig. 3, and  $\cos(\theta) \equiv 0$  when  $r = 0$ . The key distinction in

our ICE preference formulation here is that a "direction cosine" may be negative and has only a finite range,  $-1 \leq \cos(\theta) \leq +1$ . Note that  $\cos(\theta) = 0$  if and only if  $x = y$ . Furthermore,  $\cos(\theta)$  and the associated ICE preference are both strictly positive only below the  $x = y$  standardized diagonal. Similarly, an ICE preference and  $\cos(\theta)$  are both strictly negative only above the  $x = y$  diagonal ...as dictated by the anti-symmetry axiom,  $P(x, y) = -P(y, x)$ .

The ICE monotonicity property states that preference,  $P(x, y)$ , cannot be lower at any point,  $(x, y)$ , within any SE-quadrant-shaped sub-region of the ICE plane that has its NW corner at  $(x_0, y_0)$ . Here we compute the partial derivatives of  $P(x, y)$  of Eq. 1 with respect to  $x$  and  $y$ , displaying expressions that can be verified using, say, Mathematica® software (Wolfram Research Inc. 2000). Specifically, we now show that the restriction producing ICE preference monotonicity also implies that  $w(x, y) \geq 0$  in Eq. 4.

For all outcomes other than  $x = 0$  (or  $y = 0$ ), we will use the conventions that (i)  $d|x|/dx = \text{sign}(x) = x/|x|$ , (ii)  $d[\text{sign}(x)]/dx = 0$ , and (iii)  $[\text{sign}(x)]^2 = +1$ . It then follows from Eq. 1 that

$$dP(x,y)/dx \propto +[x^2+y^2]^{(\beta-\gamma)/2-1}|x-y|^{\gamma-1} \times [\beta x^2 + (\gamma-\beta)xy + \gamma y^2], \quad (6)$$

while

$$dP(x,y)/dy \propto -[x^2+y^2]^{(\beta-\gamma)/2-1}|x-y|^{\gamma-1} \times [\gamma x^2 + (\gamma-\beta)xy + \beta y^2] \quad (7)$$

when  $(x, y) \neq (0, 0)$ , where  $\beta$  and  $\gamma$  are both strictly positive.

In Eqs. 6 and 7, note that the  $|x-y|^{\gamma-1}$  factors assure that the derivatives of  $P(x, y)$  will be zero when  $x = y$  and  $\gamma > 1$ . The derivatives of  $P(x, y)$  cannot change sign due to these  $|x-y|^{\gamma-1}$  factors but can change sign due to the  $xy$  cross-terms within the final [ ] factors of Eqs. 6 and 7. These final [ ] terms will be zero whenever  $y/x$  or  $x/y$  is of the special form

$$\frac{\beta-\gamma \pm \sqrt{\beta^2-6\beta\gamma+\gamma^2}}{2\gamma} \quad \text{or} \quad \frac{\beta-\gamma \pm \sqrt{\beta^2-6\beta\gamma+\gamma^2}}{2\beta} . \quad (8)$$

On the other hand, note that there cannot be *real valued solutions* of the form given in Eq. 8 when  $\beta^2 - 6\beta\gamma + \gamma^2$  is *negative*. Thus it follows that ICE monotonicity holds within the two-parameter family of Eq. 1 under the simple restriction that  $1/\Omega \leq \gamma/\beta \leq \Omega$  of Eq. 2, where  $1/\Omega = (1-\sqrt{2})^2 = 3-2\sqrt{2} \approx 0.1716$  and  $\Omega = (1+\sqrt{2})^2 = 3+2\sqrt{2} \approx 5.828$  are the two (reciprocal) solutions of the simple quadratic equation  $\Omega^2 - 6\Omega + 1 = 0$ . Finally, since WTP and WTA are defined to equal  $\lambda$  times  $(-dP/dx)/[dP/dy]$ , the standardized willingness defined in Eq. 4 cannot be negative under restriction (2). Quod Erat Demonstrandum.

## References

1. Black, W.C.: The CE plane: a graphic representation of cost-effectiveness. *Med. Decis. Making* 10, 212–214 (1990) [PubMedCrossRefGoogle Scholar](#)
2. Briggs, A.H., Fenn, P.: Confidence intervals or surfaces? Uncertainty on the cost-effectiveness plane (Student Corner.). *Health Econ.* 7, 723–740 (1998) [PubMedCrossRefGoogle Scholar](#)

3. Chaudhary, M.A., Stearns, S.C.: Estimating confidence intervals for cost-effectiveness ratios: an example from a randomized trial. *Stat. Med.* **15**, 1447–1458 (1996) [PubMedCrossRef](#) [Google Scholar](#)
4. Cook, J.R., Heyse, J.F.: Use of an angular transformation for ratio estimation in cost-effectiveness analysis. *Stat. Med.* **19**, 2989–3003 (2000) [PubMedCrossRef](#) [Google Scholar](#)
5. Copley-Merriman, C., Egbonu-Davis, L., Kotsanos, J.G., Conforti, P., Franson, T., Gordon, G.: Clinical economics: a method for prospective health resource data collection. *Pharmacoeconomics* **1**(5), 370–376 (1992) [PubMedCrossRef](#) [Google Scholar](#)
6. Fan, M.Y., Zhou, X.H.: A simulation study to compare methods for constructing confidence intervals for the incremental cost-effectiveness ratio. *Health Serv. Outcomes Res. Method* **2007**; doi: [10.1007/s10742-006-0017-9](https://doi.org/10.1007/s10742-006-0017-9)
7. Fenwick, E., O'Brien, B.J., Briggs, A.H.: Cost-effectiveness acceptability curves—facts, fallacies and frequently asked questions. *Health Econ.* **13**: 405–415 (2004) [PubMedCrossRef](#) [Google Scholar](#)
8. Gafni, A., Birch S.: Incremental cost-effectiveness ratios (ICERs): the silence of the lambda. *Soc. Sci. Med.* **62**, 2091–2100 (2006) [PubMedCrossRef](#) [Google Scholar](#)
9. Goldstein, D.J., Lu, Y., Detke, M.J., Wiltse, C., Mallincrodt, C., Demitrack, M.A.: Duloxetine in the treatment of depression—a double-blind, placebo-controlled comparison with paroxetine. *J. Clin. Psychopharmacol.* **24**, 389–399 (2004) [PubMedCrossRef](#) [Google Scholar](#)
10. Hamilton, M.: Development of a rating scale for primary depressive illness. *Brit. J. Soc. Clin. Psychol.* **6**, 278–296 (1967) [Google Scholar](#)
11. Heitjan, D.F., Moskowitz, A.J., Whang, W.: Problems with interval estimates of the incremental cost-effectiveness ratio. *Med. Decis. Making* **19**, 9–15 (1999a) [PubMedCrossRef](#) [Google Scholar](#)
12. Heitjan, D.F., Moskowitz, A.J., Whang, W.: Bayesian estimation of cost-effectiveness ratios from clinical trials. *Health Econ.* **8**, 191–201 (1999b) [PubMedCrossRef](#) [Google Scholar](#)
13. Johnson, F.R.: Einstein on willingness to pay per QALY: is there a better way? [editorial.] *Med. Decis. Making* **25**, 607–608 (2005) [PubMedCrossRef](#) [Google Scholar](#)
14. Kent, D.M., Fendrick, A.M., Langa, K.M.: New and dis-improved: on the evaluation and use of less effective, less expensive medical interventions. *Med. Decis. Making* **24**, 281–286 (2004) [PubMedCrossRef](#) [Google Scholar](#)
15. Laska, E.M., Meisner, M., Siegel, C., Stinnett, A.A.: Ratio-based and net benefit-based approaches to health care resource allocation: proofs of optimality and equivalence. *Health Econ.* **8**, 171–174 (1999) [PubMedCrossRef](#) [Google Scholar](#)
16. Laupacis, A., Feeny, D., Detsky, A.S., Tugwell P.X.: How attractive does a new technology have to be to warrant adoption and utilization? Tentative guidelines for using clinical and economic evaluations. *Can. Med. Assoc. J.* **146**(4), 473–481 (1992) [Google Scholar](#)
17. O'Brien, B.J., Gertsen, K., Willan, A.R., Faulkner L.A.: Is there a kink in consumers' threshold value for cost-effectiveness in health care? *Health Econ.* **11**(2), 175–180 (2002) [PubMedCross Ref](#) [Google Scholar](#)

18. O'Brien, B.J., Sculpher, M.J.: Building uncertainty into cost-effectiveness rankings: portfolio risk-return tradeoffs and implications for decision rules. *Med. Care* **38**, 460–468 (2000) [PubMedCrossRef](#) [Google Scholar](#)
19. Obenchain, R.L.: Issues and algorithms in cost-effectiveness inference. *Biophar. Rep.* **5**(2), 1–7, American Statistical Association, Washington (1997) [Google Scholar](#)
20. Obenchain, R.L.: ICEplane: a Microsoft Windows application for calculation and graphical display of bootstrap ICE confidence and tolerance regions. Copyright © Pharmaceutical Research and Manufacturers of America (PhRMA.) Version 2005.11. **(Requires 32-bit hardware.)** <http://localcontrolstatistics.org/other/ICE2005B.EXE> (1997–2005)
21. Obenchain, R.L.: Resampling and multiplicity in cost-effectiveness inference. *J. Biopharm. Stat.* **9**(4), 563–582 (1999) [PubMedCrossRef](#) [Google Scholar](#)
22. Obenchain, R.L.: The key role of “symmetry” in cost-effectiveness analyses. In: Proc. Biopharmaceutical Section, pp 15–17. Joint Statistical Meetings. American Statistical Association, Washington (2000)
23. Obenchain, R.L.: Incremental cost-effectiveness (ICE) preference maps. In: Proc. Biopharmaceutical Section, 10 pp. Joint Statistical Meetings. American Statistical Association, Washington, CD-ROM only. (2001)
24. Obenchain, R.L., Robinson, R.L., Swindle, R.W.: Cost-effectiveness inferences from bootstrap quadrant confidence levels: three degrees of dominance. *J. Biopharm. Stat.* **15**(3), 419–436 (2005)
25. Obenchain, R.L.: ICEinfer: an R-package of functions for computation and graphical display of ICE maps, bootstrap confidence regions and acceptability curves. Version 1.1 <https://CRAN.R-project.org/package=ICEinfer> (June 2018)
26. Schoenbaum, M., Unutzer, J., Sherbourne, C., Duan, N., Rubenstein, L.V., Miranda, J., Meredith L.S., Carney, M.F., Wells, K.: Cost-effectiveness of practice-initiated quality improvement for depression: results of a randomized controlled trial. *JAMA* **286**(11), 1325–1330 (2001)
27. Stinnett, A.A.: Is it really so bad to be unambiguously inefficient? The role of dominance in stochastic cost-effectiveness analysis. *Med. Decis. Making* **19**, 102–103 (1999)
28. Stinnett, A.A., Mullahy, J. Net health benefits: a new framework for the analysis of uncertainty in cost-effectiveness analysis. *Med. Decis. Making*, Special Issue on Pharmacoeconomics **18**, S68–S80 (1998)
29. The **R** Project for Statistical Computing. (Deepayan Sarkar’s Lattice Graphics package uses Paul Murrell’s Grid Graphics engine.) <http://www.r-project.org>
30. Van Hout, B.A., Al, M.J., Gordon, G.S., Rutten, F.F.H.: Costs, effects and C/E ratios alongside a clinical trial. *Health Econ.* **3**, 309–319 (1994)
31. Willan, A.R., O'Brien, B.J., Leyva, R.A.: Cost-effectiveness analysis when the WTA is greater than the WTP. *Stat. Med.* **20**(21), 3251–3259 (2001)
32. Wolfram Research, Inc., Mathematica®, Version 4.1, Champaign, IL (2000)