COST-EFFECTIVENESS ANALYSIS WHEN THE WILLINGNESS TO ACCEPT IS GREATER THAN THE WILLINGNESS TO PAY

COST-EFFECTIVENESS ANALYSIS WHEN THE WILLINGNESS TO ACCEPT IS GREATER THAN THE WILLINGNESS TO PAY

By

RINA A. LEYVA LOAYZA, B.Sc.

A Thesis

Submitted to the School of Graduate Studies

in Partial Fulfillment of the Requirements

for the Degree

Master of Science

McMaster University

MASTER OF SCIENCE (2001)   **McMaster University**
(Statistics)   **Hamilton, Ontario**

TITLE	**Cost-effectiveness analysis when the WTA is Greater than the WTP**
AUTHOR:	**Rina A. Leyva Loayza, B.Sc. (University of Guelph)**
SUPERVISOR:	**Dr. A.R. Willan**
NUMBER OF PAGES:	**xii, 75**

ABSTRACT

There are three approaches to health economic evaluation for comparing two therapies: cost minimization, incremental cost-effectiveness ratios (ICER), and incremental net benefit (INB). Of the three, the ICER method has long been the standard in the assessment of the cost-effectiveness analysis of a new treatment. However, due to concerns with interpretability and statistical inference inherent to the ICER statistic and its confidence intervals, authors have suggested the use of incremental net benefit (INB) approach as an alternative. The INB can be expressed either in units of effectiveness or costs. When analyzing data from a clinical trial, expressing incremental net benefit in units of cost allows the investigator to examine all three approaches in a single graph, complete with the corresponding statistical inferences. Furthermore, if costs and effectiveness are not censored, this can be achieved using common statistical procedures.

The standard INB analysis assumes that the willingness-to-accept (WTA) compensation for the loss of a unit of health benefit (at some cost saving) is the same as the willingness-to-pay (WTP) for it. Theoretical and empirical evidence suggest, however, that in health care the WTA is about twice the WTP. In this thesis we show that the method of INB analysis can be adapted to capture the WTA vs WTP disparity. Using the Bayesian theory, statistical procedures are provided for the cost-effectiveness analysis in the comparison of two arms of a randomized clinical trial that allows WTA and WTP to have different values. An example that adjusts the disparity between WTA and WTP is provided.

DEDICATION

I would like to dedicate this thesis to my parents Rina Loayza Delgado and Victor Leyva Vallejos and sisters Lilian Veronica and Victoria Luz Leyva Loayza. Thank you for your unconditional love, support and understanding.

DEDICATORIA

Esta tesis esta dedicada a mis padres Rina Loayza Delgado y Victor Leyva Vallejos y hemanas Lilian Veronica y Victoria Luz Leyva Loayza. Gracias por su apoyo incondicional, comprension, amor, y por ser la fuente de luz e inspiracion de mis logros y exitos.

TABLE OF CONTENTS

CHAPTER 1:   INTRODUCTION

CHAPTER 2:   ECONOMIC ANALYSIS METHODS

CHAPTER 3:   WILLINGNESS TO PAY AND WILLINGNESS TO ACCEPT

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

CAD$	Canadian dollars
CBA	Cost Benefit Analysis
CE	Cost-Effectiveness
CEA	Cost-Effectiveness Analysis
CI	Confidence Interval
CM	Cost Minimization
CMA	Cost Minimization Analysis
CV	Contingent Valuation
EORTC	European Organization for Research and Treatment of Cancer
HRPC	Hormone Resistant Prostate Cancer
ICE	Incremental Cost-Effectiveness
ICER(s)	Incremental Cost-Effectiveness Ratio(s)
INB	Incremental Net Benefit
IVF	*in vitro* Fertilization
QALW(s)	Quality-Adjusted Life-Week(s)
QALY(s)	Quality-Adjusted Life-Year(s)
QLQ	Quality of Life Questionnaire
RCT	Randomized Controlled Trial
S	Standard

T	Treatment
WTA	Willingness-to-Accept
WTP	Willingness-to-Pay

CHAPTER 1

INTRODUCTION

1.1 Opening Statement

The research presented in this thesis involves the introduction of a novel statistical approach for the incremental net benefit analysis when the willingness-to-pay and the willingness-to-accept values are different. Much of the literature reviewed in Chapter 2 of this thesis focuses on the different approaches for economic evaluation for comparing two therapies. Textbooks and guidelines on health economic evaluation typically distinguish three different types of evaluation methods for comparing two therapies: cost minimization analysis (CMA), incremental cost-effectiveness ratio (ICER) analysis, and incremental net benefit (INB) analysis (Willan and Lin, 2000). The relevance of this discussion to the developing literature on statistical inference in health economics is important.

In Chapter 3, the focus is on the different approaches to calculate the WTP and WTA values. Furthermore, the evidence that willingness-to-accept is higher than willingness-to-pay is presented and the statistical procedure to analyze such difference as part of a clinical trial is introduced. The modified INB technique is used to analyze the effects of different values of γ in a randomized controlled trial comparing two therapies. The example is presented in Chapter 4.

Finally, the advantages of using the modified INB approach for economic analysis in clinical trials are discussed in Chapter 5.

1.2 Overview of Health Economic Evaluation Methods

It is becoming increasingly common to conduct cost-effectiveness analyses of medical therapies prospectively as an integrated component of a randomized controlled trial (RCT). This has provoked considerable debate on how best to measure the cost-effectiveness of a new therapy (Treatment) relative to a standard and has motivated the development of new statistical methodology to quantify the uncertainty in cost-effectiveness analyses. Most of the attention has centred on making inference about the incremental cost-effectiveness ratio (ICER), the ratio of the mean difference in cost to the mean difference in effectiveness, using confidence intervals (Briggs et al., 1997; Chaudhary and Stearns, 1996; Heitjan, 2000; Manning et al., 1996; Mullahy and Manning, 1994; Mullahy, 1996; O'Brien et al., 1994; Polsky et al., 1997; Van Hout et al., 1994; Wakker and Klaassem, 1995; Willan and O'Brien, 1996; Willan and Lin, 2000).

There are several methods available in the literature for calculating confidence intervals for ICERs; however, no consensus has been reached as to which (if any) is the most appropriate method to use. Perhaps due to the fast moving nature of this field of research, recently published overviews of methods for handling uncertainty in economic evaluation have tended to focus on the Taylor's series method (Briggs and Fenn, 1998; Drummond et al., 1997; Manning et al., 1996), while the emerging evidence from Monte Carlo evaluation studies have identified Fieller's theorem and the non-parametric

bootstrapping as the most valid methods (Briggs et al., 1999; Polsky et al., 1997). However, there are statistical and conceptual problems that complicate the estimation of a confidence interval for the ratio between two random variables (Briggs and Fenn, 1998; Tambour et al., 1998; Willan and Lin, 2000). For example, the cost-effectiveness tradeoff represented by the ratio of two positive differences is not necessarily equivalent to the tradeoff represented by an equal ratio of negative differences. Moreover, the ranking of negative ICERs is ambiguous, and summarization to a single number may be misleading (Heitjan, 2000). A detailed discussion of this procedure and the problems associated with it is presented in Chapter 2.

An alternative approach to measure cost-effectiveness is incremental net benefit (INB), which is defined as the difference in effectiveness and the difference in costs on the same scale (Heitjan, 2000; Willan and Lin, 2000). Statistical inferences on INB are less problematic than inferences on the ICER because both costs and benefits are expressed in the same units (money), and it is written as a linear equation thus the confidence intervals for net benefits can be calculated using standard statistical approaches (Tambour et al., 1998) (See Chapter 3).

The value that society attaches to a unit of effectiveness has an important impact on the priority of the ranking of two therapies (Treatment versus Standard). This value is also important for an analysis of INB because this approach requires the specification of this value known as the willingness-to-pay (WTP) for a unit of effectiveness (denoted as λ); or at the very least, the analysis must be done as a function of λ so that readers can apply the WTP most appropriate to them. It is important to recognize that, although the

INB approach is more useful when compared to other evaluation techniques such as the ICER, a major disadvantage inherent in this framework is the difficulty of obtaining valid and reliable estimates of WTP (Gafni, 1991).

For a number of years economists conducting cost-benefit analysis (CBA) have carried out empirical studies to measure consumers' WTP for public program benefits that are not marketed–examples being environmental and health benefits (Mitchell and Carson, 1989; O'Brien and Gafni, 1996). The literature on economic evaluation in health care has shown that the method that is preferred for estimating money values is WTP or WTA survey techniques known as Contingent Valuation (CV) methods (Gafni, 1991). In a CV survey consumers are asked to consider a hypothetical scenario where they are asked for the maximum amount they are WTP to have the commodity, or the minimum amount they would be WTA in compensation to be deprived of it (O'Brien and Gafni, 1996; Olsen and Smith; 2001). Of these measures, the WTP method is the most widely used (Olsen and Smith, 2001). There are many techniques to elicit the WTP value; each approach has strengths and weaknesses with different measurement properties of precision and bias. A summary of the most common methods is presented in Chapter 3.

In theory it should not matter whether one poses the question of a person's WTP for introducing a program benefit, or their WTA monetary compensation to remove it (i.e. WTP = WTA). However, in practice there has been a wide and reproducible disparity between measured WTP and WTA values (Kolstad and Guzman, 1998; Mueser and Dow, 1997; O'Brien et al., 2001). According to a recent meta-analysis of published studies, the WTA to WTP ratio is approximately seven for environmental studies and approximately

two in the one health study where WTP and WTA were both measured (O'Brien et al., 2001).

There are several competing theories as to why WTA>WTP (See Chapter 3), one of which is proposed by Hanemann (1991) who argues that a lack of substitute commodities for a removed program will inflate WTA. However, most researchers seem to agree with the theory proposed by Kahneman and Tversky (1979) who argue in favor of a psychological theory of endowment effect. In this theory the loss of utility associated with the loss of something weights heavier than the utility associated with gaining the same benefit (Fox and Tversky, 1995; Hanemann, 1991; Mueser and Dow, 1997).

The reasons as to why WTA>WTP continues to be debated (Morrison, 1998), but what is important are the practical implications of this disparity for the cost-effectiveness analysis when comparing therapies. In an INB analysis, the gold standard is to assume that the WTA compensation for a loss of a unit of health benefit is the same as the WTP for it. Graphically, this equality is represented as a straight line through the origin of the cost-effectiveness plane. However, empirical and theoretical evidence suggests, that in health care there is a difference in that WTA is about twice the WTP (O'Brien et al., 2001). If this difference is taken into account in the cost-effectiveness (CE) plane, it produces a "kink" in the line that represents the accept-reject threshold for cost-effectiveness. This is an important fact that should be considered in an INB analysis because only then is the acceptable level of hypothetical compensation required to withdraw health benefits greater than the willingness-to-pay to achieve the same health benefits. The implications of this disparity in the economic analysis of clinical trials are

the focus of this thesis. In this work we introduce a modified version of the INB approach that allows the values for the willingness-to-pay and the willingness-to-accept to be different.

Since the primary goal of economic evaluation is to inform decision makers about the relationship between costs and benefits of new and existing therapies, we use the adapted INB statistical procedure to analyze the effects of the differences between WTA and WTP in the cost-effectiveness analysis of comparing two treatments for symptomatic hormone resistant prostate cancer (HRPC) (Bloomfield et al., 1998; Tannock et al., 1996).

1.3 The Cost-Effectiveness Plane and Randomized Clinical Trials

Randomized clinical trials are frequently used to estimate the effect of a new therapy (Treatment) relative to the effect of the current standard therapy (Willan and Lin, 2000). When data on both costs and effects are collected prospectively, changes in costs associated with a therapy can also be estimated (Polsky et al., 1997).

An important aspect in cost-effectiveness analysis is the graphical representation of the difference of costs and effects and the results of the ICERs or INB analysis in a plane. The CE plane (Anderson et al., 1986; Black, 1990; Briggs and Fenn, 1998) is often employed to show how decisions can be related to both costs and effects. In the CE plane presented in Figure 1.1, the horizontal axis represents the difference in effect between the standard and new treatment, and the vertical axis represents the difference in costs. The plane is divided into four quadrants indicating four possible situations in relation to the

additional costs and additional health outcome effects of a new treatment compared to the standard (Drummond et al., 2000). One possibility is the treatment maybe more effective and less costly (quadrant II), in which case it is said to dominate the standard therapy. Another case is the treatment maybe less effective and more costly (quadrant IV), in which case it is said to be dominated by the standard therapy. Finally, the treatment maybe more costly and more effective (quadrant I) or it maybe less costly and less effective (quadrant III). In these two cases the decision is no longer obvious. In practice the impact of many interventions falls into quadrant I. That is, they add to cost but increase effectiveness, certainly when compared with no intervention (Drummond et al., 2000), but many lie into quadrant II since increasing effectiveness can often lead to a reduction in cost.

For the cases in quadrants I and II, if it is possible to define some maximum acceptable value for the ICER, the 'ceiling' value of the ratio can be used to judge whether the treatment in question is cost-effective (Briggs and Fenn, 1998). The ceiling value of the ICER can be represented by the slope of the line on the CE plane of Figure 1.1. If the incremental costs and effects lie to the right of this line on the CE plane then the treatment is considered cost-effective, while points to the left of this line represent cost-ineffective interventions (Briggs and Fenn, 1998; Polsky et al., 1997).

ΔCost

IV   **I**

Treatment less effective and more   Treatment more effective and more
costly (Reject)   costly

R E J E C T

A C C E P T

−   + **ΔEffectiveness**

Treatment less effective and less   Treatment more effective and less
costly   costly (Accept)

III   **II**
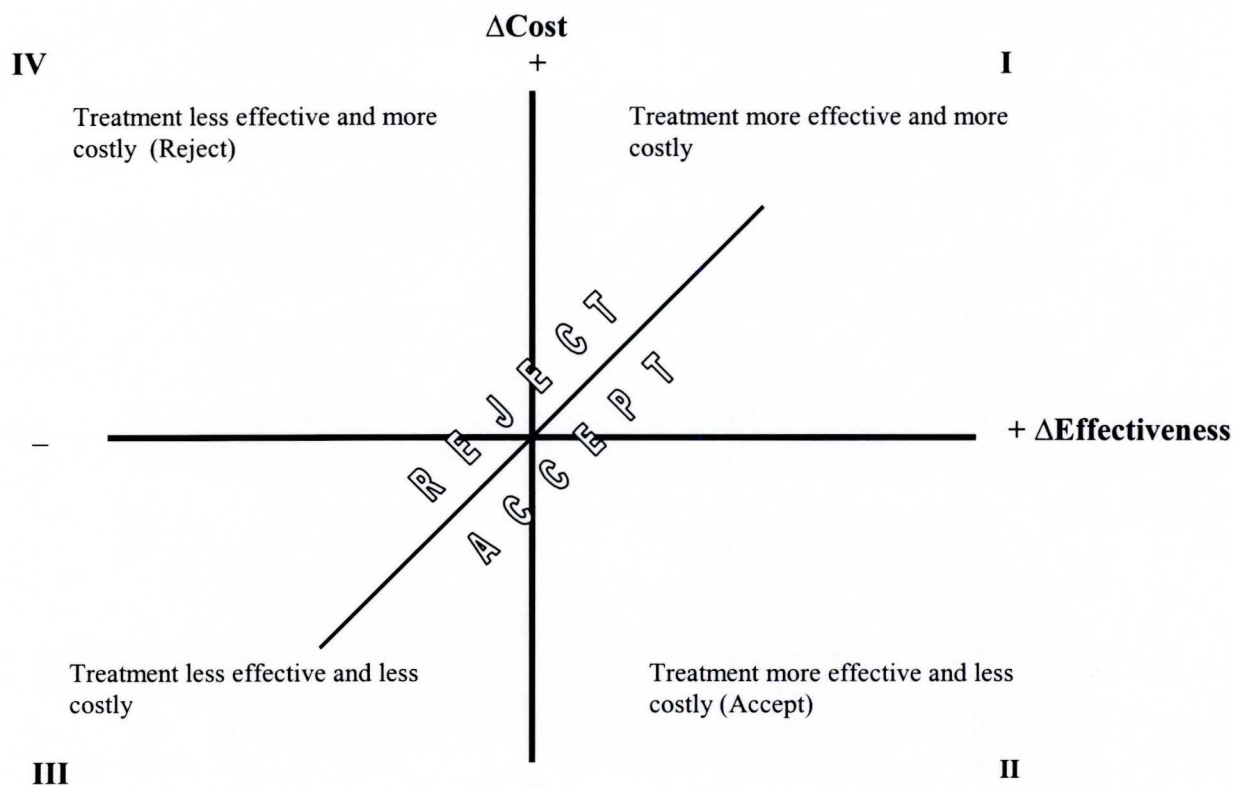
Figure 1.1: Schematic representation of the cost-effectiveness (CE) plane. In the diagram the horizontal axis represents the difference in effects (Δ_e), and the vertical axis represents the difference in cost (Δ_c). In quadrants II or IV the choice between the therapies is clear. In quadrant II the treatment is both more effective and less costly than the standard. That is, it dominates the standard therapy. In quadrant IV the opposite is true. In quadrants I and III the choice depends on the ceiling value or ICER. In the case of INB, it depends on the maximum one is willing to pay for a unit of effectiveness. The slope of the line gives the cost-effectiveness ratio or the willingness-to-pay value (λ). (Graph adapted from Black, 1990)

CHAPTER 2

ECONOMIC ANALYSIS METHODS

2.1 Cost-Effectiveness Analysis (CEA)

2.1.1 Definition

As costs have become more important in health care decision making, the number of economic evaluations of medical therapies has grown (Commonwealth of Australia, 1992; Glick, 1995; Leaf, 1989; Ontario Ministry of Health, 1991; Polsky et al., 1997). In addition to evidence on the effectiveness and safety of new therapies, there is an increasing demand from health care policy makers for data on the cost-effectiveness or value for money (Willan and O'Brien, 1996) of these medical therapies.

The most widely used technique of economic appraisal in health care is cost-effectiveness analysis (CEA) (Willan and O'Brien, 1996). Cost-effectiveness analysis is a quantitative technique for comparing the costs and effectiveness of a new therapy (Treatment) relative to some relevant standard (Willan and O'Brien, 1996). Cost-effectiveness analyses are common not only in the health care sector, but are becoming increasingly popular in many other areas (Commonwealth of Australia, 1992; Detsky, 1993; Leaf, 1989). This trend has been accompanied by an increase in the number of economic evaluations that are conducted as part of randomized clinical trials (Drummond and Davis, 1991; Eisenberg et al, 1994; Glick, 1995; Morris and Schulman, 1995; Polsky et al., 1997). The basis for economic evaluation relies on the data of effectiveness of the

therapy being evaluated relative to some standard therapy (Drummond et al., 1987; Luce, and Elixhauser, 1990; Willan and O'Brien, 1996).

2.1.2 Economic Analysis and Randomized Clinical Trials

Randomized controlled trials (RCT), used either as single studies or combined in meta-analyses, are valuable sources of evidence on effectiveness of therapies (Chaudhary and Stearns, 1996, L'Abbe et al., 1987). There are two general ways in which RCT data can be incorporated into an economic evaluation:

1. combining RCT effectiveness data retrospectively with cost data from secondary non-trial sources into a decision analysis model; or

2. collecting effectiveness and cost data on the same patients prospectively as part of an RCT.

The first method is referred to as a deterministic model, which for many years has been the standard for CEAs (Willan and O'Brien, 1996). This approach uses non-sampled secondary data (e.g. published literature, insurance claims databases, and expert opinion). Implicitly this data has an inherited uncertainty, which makes the interpretive task for cost-effectiveness more complex. The method that is widely recommended for assessing data uncertainty in economic appraisals of this type and allied evaluative techniques such as clinical decision analysis is sensitivity analysis (Commonwealth of Australia, 1992; Destky, 1993; Drummond et al., 1987; Leaf, 1989, Weinstein, 1980). The purpose is to examine the robustness of an estimated result over a range of plausible alternative values for unknown variables (Willan and O'Brien, 1996).

There are three major limitations with sensitivity analysis. First, the analyst has discretion as to which variables and what alternative values are included in the sensitivity analysis, creating the potential for selection bias (conscious or otherwise). Second, the interpretation of a sensitivity analysis is essentially arbitrary because there are no guidelines or standards as to what degree of variation in results is acceptable evidence that the analysis is 'robust'. Third, variation of uncertain parameters one at a time carries a risk that interactions between variables may not be captured (Willan and O'Brien, 1996).

Although more sophisticated simulation approaches to sensitivity analysis based on Monte Carlo methods exist to analyze deterministic models (Doubilet et. al., 1985), the second method, (i.e. the collection of prospective data as part of clinical trials) creates the opportunity to examine cost and effectiveness data with less uncertainty. This in turn allows analysts to perform cost-effectiveness analysis using conventional statistical inference. These methods are collectively called stochastic analyses (Willan and O'Brien, 1996). Due to the fact that they are based on a sample from the population, the costs, effects and cost-effectiveness ratio reported from an RCT are estimates of the true population values (Polsky et al., 1997). The degree of precision of these estimates is related to the size of the sample, sampling and measurement error (Polsky et al., 1997). This is one of the reasons many cost-effectiveness analyses are now conducted alongside clinical trials. The consequence of this union is the availability of patient-level cost data and health outcomes to analysts.

2.2 Model

In a two-arm randomized control trial let e_{ji} and c_{ji} be the respective measures of effectiveness and cost for patient i on therapy j = T (Treatment), S (Standard); $i = 1,2,...,n_j$; and n_j is the respective sample size.

$$\text{Let}\quad E\begin{pmatrix} e_{ji} \\ c_{ji} \end{pmatrix} = \begin{pmatrix} \mu_j \\ \upsilon_j \end{pmatrix}; V\begin{pmatrix} e_{ji} \\ c_{ji} \end{pmatrix} = \begin{pmatrix} \sigma_j^2 & \rho_j\sigma_j\omega_j \\ \rho_j\sigma_j\omega_j & \omega_j^2 \end{pmatrix}; \quad \text{and}\quad \Delta_e = \mu_T - \mu_S ; \Delta_c = \upsilon_T - \upsilon_S ,$$

where E is the expected value and V is the variance-covariance function. Typically, e_{ji} is the patient's survival time (perhaps quality-adjusted) from randomization to death (Willan and Lin, 2000).

2.3 Methods for Health Economic Evaluation

There are three approaches to health economic evaluation for comparing two therapies. These are:

2.3.1 Cost Minimization Analysis (CMA)

In cost minimization analysis (CMA) one assumes or observes no difference in effectiveness or no value is placed on difference in effectiveness (Willan and Lin, 2000). This means that $\Delta_e = 0$ or the analyst does not care about Δ_e (i.e. $\lambda = 0$). Therefore, in cost minimization Δ_c is the parameter of interest and measures the additional cost per patient from using T rather than S. A CMA can only be carried out without ambiguity if it

is based on exiting (medical) evidence of effectiveness (Drummond et al., 2000). Furthermore, cost minimization is a special case of incremental net benefit analysis (topic discussed later), because it assumes that there is zero willingness-to-pay for a unit of effectiveness, i.e. $\lambda = 0$.

2.3.1.1 Problems with Cost Minimization Analysis

Cost minimization analysis has considerable appeal to analysts and decision-makers because it keeps studies simple: if two treatments have the same effectiveness, then the lowest cost treatment is the treatment of choice (Briggs and O'Brien, 2001). However, the assumptions made for this simplification are not realistic because there are rare circumstances in which one can assume zero difference in effectiveness or that there is no willingness-to-pay. Its use is also inappropriate when sampled data on costs and effects are available (Briggs and O'Brien, 2001). For instance, Donaldson et al. (1996) noted that when designing a prospective economic evaluation, it is impossible to specify the technique of analysis (i.e. CEA versus CMA) because the data are unknown. Furthermore, Briggs and O'Brien (2001) contend that even when the data are known, the use of CMA is rarely appropriate as a method of economic analysis. They indicate that one possible circumstance where it might be viewed as legitimate to conduct CMA is where a randomized trial has been designed to test the explicit hypothesis of equivalence in outcome between two therapies. However, this form of CMA (conducted alongside an equivalence trial) is the exception (Briggs and O'Brien 2001). Equivalence trials are rare

because they typically require a much larger sample size than those designed to test for differences (Senn, 1997).

It is argued that unless a study has been specifically designed to show the equivalence of treatments (in terms of costs or effects), it would be inappropriate to conduct CMA on the basis of an observed lack of significance in either the effect or cost differences between treatments (Briggs and O'Brien, 2001).

2.3.2 Incremental Cost-Effectiveness Ratio (ICER)

In an incremental cost-effectiveness analysis, the parameter of interest is the incremental cost-effectiveness ratio (ICER) (Chaudhary and Stearns, 1996; Willan and Lin, 2000; Stinnett and Mullahy, 1998; Willan et al., 2001), defined as:

$$R = \frac{\Delta_c}{\Delta_e} = \frac{\upsilon_T - \upsilon_S}{\mu_T - \mu_S} \qquad (2.1)$$

If $R > 0$ and $\Delta_e > 0$ (i.e. treatment cost and effect exceed standard cost and effect), then the ratio represents the cost per additional unit of outcome achieved by using treatment rather than standard (Briggs and Fenn, 1998; Chaudhary and Stearns, 1996; Stinnett and Mullahy, 1998; Willan and Lin, 2000). In similar fashion, if $R > 0$ and $\Delta_e < 0$, then the resulting positive ratio reflects the cost per additional unit of outcome achieved by using standard rather than treatment.

If $R < 0$ and $\Delta_e > 0$, then treatment is both more effective and less expensive. This is otherwise known as the "win-win" situation (Willan and Lin, 2000). If

$R < 0$ and $\Delta_e < 0$, treatment is less effective and more costly, and is otherwise known as the "lose-lose" situation (Willan and Lin, 2000).

The parameter R is estimated using the "analogy" estimator:

$$\hat{R} = \frac{\hat{\Delta}_c}{\hat{\Delta}_e} \qquad (2.2)$$

where $\hat{\Delta}_c$ and $\hat{\Delta}_e$ represent sample estimates of the difference in means for the cost and effect from the RCT, respectively (Briggs and Fenn, 1998; Chaudhary and Stearns, 1996; Manski, 1988; Stinnett and Mullahy, 1998; Willan and Lin, 2000).

The ratio estimator is biased because it is a non-linear statistic but because the estimator is consistent it is possible to ignore the bias for large samples (Chaudhary and Stearns, 1996; Cochran, 1977). However, for moderate sample sizes, the distribution of $\hat{R}$ has been found to be positively skewed (Chaudhary and Stearns, 1996). Nonetheless, the limiting distribution of the ratio estimate is normal as the sample size becomes very large, which is subject to some mild restrictions (Cochran, 1977).

2.3.2.1 Statistical Properties of the ICER Statistic

Often in practice, new treatments are both more effective and more costly than standard therapy. Therefore, the majority of cost-effectiveness analyses find it necessary to report a point estimate of the ICER (Chaudhary and Stearns, 1996). Where patient-level data are available, it is natural to also present a confidence interval to represent uncertainty due to sampling variation (Briggs and Fenn, 1998). The purpose of collecting such data and estimating the ICER statistic is to make inferences about the population

ICER. Essentially, the estimated ICER statistic is constructed from four sample means (the mean costs and effects from each patient group). Although the underlying data may not follow a well-behaved distribution in general, the *central limit theorem* states that the means will approach a normal distribution with increasing sample size (Briggs and Fenn, 1998). Therefore, the larger the sample size, the closer the relevant estimators are to having normal distributions (Briggs and Fenn, 1998; Cochran, 1977).

In addition, statistical theory states that the difference of two normal variables is itself normally distributed. Therefore, it is possible to assume that the estimators of the incremental costs and effects (i.e. the numerator and denominator of the ICER) (equation 2) are also normally distributed:

$$\begin{pmatrix} \hat{\Delta}_e \\ \hat{\Delta}_c \end{pmatrix} \sim N\left[\begin{pmatrix} \Delta_e \\ \Delta_c \end{pmatrix}, \begin{pmatrix} V(\hat{\Delta}_e) & C(\hat{\Delta}_e, \hat{\Delta}_c) \\ C(\hat{\Delta}_e, \hat{\Delta}_c) & V(\hat{\Delta}_c) \end{pmatrix} \right] \qquad (2.3)$$

where,

$$V(\hat{\Delta}_e) = \frac{\sigma_T^2}{n_T} + \frac{\sigma_S^2}{n_C}$$

$$V(\hat{\Delta}_c) = \frac{\omega_T^2}{n_T} + \frac{\omega_S^2}{n_S}$$

$$C(\hat{\Delta}_e, \hat{\Delta}_c) = \frac{\rho_T \sigma_T \omega_T}{n_T} + \frac{\rho_S \sigma_S \omega_S}{n_S}$$

It is clear from the above description that the estimated ICER statistic (equation 2) is a ratio of two asymptotically normal variables, which Wakker and Klaassen (1995) point out will be Cauchy distributed (a t-distribution with one degree of freedom), where the mean and the variance are undefined. This would indicate that the above sample

estimates will be unstable (Stinnett and Mullahy, 1998; Wakker and Klaassen, 1995). However, it is the ratio of two independent normal distributions that has a Cauchy distribution (Briggs and Fenn, 1998; Chaudhary and Stearns, 1996; Wakker and Klaassen, 1995).

2.3.2.2 Confidence Intervals for ICERs

A $(1 - \alpha)100\%$ confidence interval (CI) is a statistical measure of precision for estimates with sample variation (Polsky et al., 1997). This interval defines a range within which one can be $(1 - \alpha)100\%$ confident the true value lies in this CI; that is, in repeated sampling the true value will be contained in the CI $(1 - \alpha)$ 100% of the time (Polsky et al., 1997).

Computing a $(1 - \alpha)100\%$ confidence interval for the estimate of either costs or effects depends on the distribution of the variable, its mean, variance, and sample size. Formulae for computing these intervals are readily available and are reliable because unbiased and efficient estimates are available since the distributions of the sample mean costs and effects are approximately normal when the sample size is sufficiently large (Polsky et al., 1997; Van Hout et al., 1994). However, the sampling distribution of the ICER statistic may not be known or it may not be well behaved, therefore, the estimation of confidence intervals cannot be done using routine statistical methods (Briggs and Fenn, 1998; O'Brien et al., 1994; Polsky et al., 1997; Van Hout et al, 1994).

In an attempt to overcome these difficulties, a number of authors have proposed alternative methods for estimating confidence limits for the ICER given sampled data on

the cost and effects. Three methods commonly used in the computation of CI for cost-effectiveness ratios are: the Taylor series method, the Fieller's theorem method and the non-parametric bootstrap method.

2.3.2.2.1 The Taylor Series Method

This method is also called the "delta method". It was one of the first proposed by O'Brien et al. (O'Brien et al., 1994), and estimates the variance of $\hat{R}$ using a second order Taylor series approximation. A two-tailed $(1 - \alpha)$ CI can be constructed as:

$$\hat{R} \pm z_{1-\alpha/2} \sqrt{Var(\hat{R})} \tag{2.4}$$

where $z_{1-\alpha/2}$ is the upper percentile of the standard normal distribution and $Var(\hat{R})$ is the estimated variance of the ratio. This method assumes that the cost-effectiveness ratio estimate is normally distributed, which among other properties requires that the confidence interval be symmetric about the ratio. O'Brien et al. (1994) argue that, although the assumption of a normal distribution may be justified in the case of large samples, it is unlikely that the distribution of every estimated ICER will follow a well-behaved distribution in general and thus one should remain cautious when using Taylor's approximation to calculate the CI for ICERs.

In addition to the above limitations, a Monte Carlo experiment performed by Polsky et al. (1997) that compared this method with other methods concluded that the Taylor series technique provided confidence intervals that asymmetrically underestimated the upper limit of the interval.

2.3.2.2.2 Fieller's Theorem Method

This approach has been advocated for use in calculating confidence intervals for ICERs by Chaudhary and Sterns (1996) and by Willan and O'Brien (1996).

This method is an application of Fieller's theorem (Fieller, 1954) and is a parametric method that computes confidence intervals of a ratio based on the assumption that the numerator and denominator of the ICER follow a bivariate normal distribution, so that $(\hat{\Delta}_c) - R(\hat{\Delta}_e)$ is normally distributed with mean 0 (Chaudhary and Stearns, 1996; Polsky et al., 1997). Consequently,

$$\frac{\left[\hat{\Delta}_c - R\hat{\Delta}_e\right]^2}{V(\hat{\Delta}_c) + R^2 V(\hat{\Delta}_e) - 2RC(\hat{\Delta}_e, \hat{\Delta}_c)} \leq \chi^2_{1,1-\alpha} = z^2_{1-\alpha/2} \tag{2.5}$$

By equating this expression to $z^2_{1-\alpha/2}$, and solving for R, Chaudhary and Sterns (1996) determined the following formula for calculating the $(1-\alpha)$ 100% confidence interval for R (Briggs and Fenn, 1998; Chaudhary and Stearns, 1996):

$$\text{Confidence Limits} = \frac{\hat{\Delta}_c}{\hat{\Delta}_e} \left(\frac{1 - z^2_{1-\alpha/2}c \pm z_{1-\alpha/2}\sqrt{a + b - 2c - z^2_{1-\alpha/2}(ab - c^2)}}{1 - z^2_{1-\alpha/2}a} \right) \tag{2.6}$$

where,

$$a = \frac{\hat{V}(\hat{\Delta}_e)}{\hat{\Delta}_e^2}, \quad b = \frac{\hat{V}(\hat{\Delta}_c)}{\hat{\Delta}_c^2} \quad \text{and} \quad c = \frac{\hat{C}(\hat{\Delta}_e, \hat{\Delta}_c)}{\hat{\Delta}_e \hat{\Delta}_c}$$

The advantage of Fieller's method over the Taylor series expansion is that it takes into account the potential for skewness in the sampling distribution of the ratio estimator and may, therefore, be symmetrically positioned around the point estimate (Willan and Lin, 2000).

In contrast to other methods, Fieller's theorem provides an exact solution subject to the joint normality assumption (i.e. bivariate normal distribution) (Briggs and Fenn, 1998). Despite these positive characteristics, some analysts have argued that the assumption of bivariate normality distribution may be hard to justify, particularly when the sample sizes are small (Chaudhary and Stearns, 1996).

2.3.2.2.3 Non-Parametric Bootstrap Method

This method involves re-sampling with replacement from the study sample and computing cost-effectiveness ratios in each of the samples (Chaudhary and Stearns, 1996; O'Brien et al., 1994; Polsky et al., 1997).

The underlying principle of the non-parametric bootstrap technique is that a random sample of size n (for instance the patients in a RCT) provides an empirical distribution function that estimates the probability distribution of the estimator (Efron, 1993).

The validity of the bootstrap approach rests on two asymptotics (Briggs et al., 1997): (i) as the original sample size approaches the population size, the sample distribution approaches the population distribution, and, given this, (ii) the number of bootstrap replications tends to infinity so the bootstrap estimate of the sampling distribution of a

statistic approaches the true sampling distribution (Briggs et al., 1997; Mooney and Duval, 1993).

This technique works in the following manner: let the vectors x_1, x_2, ...x_B be the B re-sampled estimates of $\left(\Delta_e, \Delta_c\right)$. Also, let $x = \left(\hat{\Delta}_e, \hat{\Delta}_c\right)$ and,

$$\theta_k = \begin{cases} +1 : x_k \text{ above the line through x and the origin} \\ -1 : x_k \text{ below the line through x and the origin} \end{cases} * \{\text{angle between x and } x_k\}$$

Therefore $\theta_k = \text{sign}(x'x_k) \left| \cos^{-1}\left(\frac{x'x_k}{\sqrt{x'x \, x'_k x_k}} \right) \right|$,

Following these results, the non-parametric bootstrap confidence interval are $\tan(\theta_{\{0.025\}})$ and $\tan(\theta_{\{0.975\}})$, where $\theta_{\{\alpha\}}$ is the $100\alpha^{\text{th}}$ percentile of the θ_i's. Figure 2.1 is a schematic representation of these results.

2.3.2.3 Problems with the ICER Statistic

There are statistical and conceptual difficulties that complicate the estimation of the incremental cost-effectiveness ratio. First of all, the estimates of this technique reflect sampling uncertainty and, although, there are different methods to obtain CIs for ICERs, unfortunately all these methods are hampered by the statistical difficulties inherent in estimating ratios and finding CIs for two random variables (Heitjan, 2000).

With respect to the conceptual problems of ICERs, the cost-effectiveness trade-off represented by the ratio of two positive differences is not necessarily equivalent to the trade-off represented by an equal ratio of negative differences (Heitjan, 2000). Furthermore, in terms of the CE plane, two totally opposite results can have the same

ICER (Willan and Lin, 2000). This problem is depicted in Figure 2.2. One of the ratios is in quadrant II of the CE plane indicating that the new treatment is more effective and less costly and therefore should be integrated. However, the same ICER can be obtained in quadrant IV of the CE plane. In this case the new treatment is more costly and less effective and should not be adopted. These quadrants, II and IV, generate negative ICERs; however, as presented above the implications for decision-making are exactly the opposite in each quadrant (Briggs and Fenn, 1998). Without examining the sign of the numerator and denominator of the ICER, it is impossible to distinguish negative ICERs in quadrant II from negative ICERs in quadrant IV.

In terms of the CE plane introduced in Chapter 1 (Figure 1.1), confidence intervals for the ICER may only be defined in quadrants I and III (Briggs and Fenn, 1998) because the ICERs are positive. Also CIs for ICER can include undefined values or may even be completely undefined (Willan and Lin, 2000). This problem is depicted in Figure 2.2, where the CI of an ICER can be in two different quadrants, making it difficult to interpret the results.

A more complex problem is the interpretation and ranking of negative ICERs. First, the interpretation of negative ICERs is difficult and their magnitude is meaningless (Briggs and Fenn, 1998). Second, negative ICERs are not properly ordered. Therefore, the summarisation to a single number in order to find the superiority of a treatment over a standard may be misleading (Heitjan, 2000; Willan and Lin, 2000). In order to explain and visualize this problem, an example is presented in Table 2.1. In this example three treatments are compared to the standard treatment, their differences in effectiveness and

costs are calculated and these results are used to calculate their respective ICERs. Three negative ICERs are obtained and all of them are in quadrant II (see Figure 2.3). The real order of treatment superiority is T2 is more cost effective than T3, which is more cost effective than T1. However, in terms of the ratios, ICER1 is equal to ICER2 and both are better than ICER3. These results, according to the theory of cost-effectiveness ratios, indicate that T1 and T2 are the same in terms of cost and effects, whereas T3 is the least effective among them (see Figure 2.3). The misinterpretation of these results is clear in this case: two treatments with the same ICER, theoretically, indicate that both treatments are appropriate and should be opted. However, by looking at the numerator and denominator of the two ratios as well as the graph of the CE plane in Figure 2.3, T2 is more effective and less costly than T1.

The extent of these problems has lead to the search for alternative methods that avoid the problems associated with estimating confidence limits for the ICER statistic.

2.3.3 Incremental Net Benefit (INB)

In response to the above problems and others associated with inference based on the distribution of ICERs, new approaches have been proposed for the analysis of uncertainty in the economic evaluation of interventions.

Much of the literature reviewed above has focused on the problems caused for statistical inference when hypotheses are constructed in terms of ratios of two random variables: in this case observations on effectiveness and costs of competing therapies (Briggs and Fenn, 1998). An alternative for economic evaluation is the incremental net

benefit (INB) approach, defined as the difference in mean effectiveness of a new Treatment (T) compared with the Standard (S), adjusted for cost difference (Willan and Lin, 2000). In this approach the effectiveness units are multiplied by the willingness-to-pay for a unit of effectiveness (WTP) and thus the effectiveness units are converted to the same units as costs (i.e. monetary units) (Briggs and Fenn, 1998; Tambour et al., 1998; Willan and Lin, 2000). Therefore, the new random variable has been created as the monetary equivalent of the incremental health benefits. The major difference between the ICER statistic and the INB statistic is that the latter is expressed in a linear form, while the former remains as a ratio. The INB is given by:

$$b(\lambda) = (\mu_T - \mu_S)\lambda - (v_T - v_S) = \Delta_e\lambda - \Delta_c \tag{2.7}$$

where λ is the WTP for a unit of effectiveness (Willan and Lin, 2000). The quantity $b(\lambda)$ is the net benefit, expressed in money, of giving a patient T rather than S. Typically λ is varied in a sensitivity analysis, and INB is expressed as $b(\lambda)$ (Willan and Lin, 2000). The question of interest is whether or not $b(\lambda) > 0$.

The hypothesis testing is simpler when dealing with well-established sampling distributions, which is the case of the hypotheses that are presented in linear terms. In recent literature Tambour et al. (1998) chose to use this approach to generate a net benefit measure on a cost scale where the decision rule is that the new treatment should be implemented if,

$$b(\lambda) = \Delta_e \lambda - \Delta_c > 0. \qquad\qquad (2.8)$$

The net-benefit approach compares the incremental treatment benefits with the incremental costs and the difference between these two gives the net benefit of treatment (Briggs and Fenn, 1998). In simple terms, the goal of this economic analysis technique is to identify whether a treatment's benefits exceed its costs, a positive incremental net benefit (as in equation 2.8) indicates that the treatment is cost-effective.

Similar to the CEA, the net-benefit approach can be represented in the CE plane (see Figure 2.4) where the horizontal axis measures the incremental effectiveness (Δ_e) and the vertical axis measures the incremental cost (Δ_c) (Stinnett and Mullahy, 1998). In Figure 2.4, the slope of the line is λ, interpreted as the WTP of a unit of effectiveness. In quadrant II the incremental net benefit is more than zero regardless of the value of λ. This is the win-win situation in which the treatment (T) is cost-effective and thus should be implemented. On the other hand, in quadrant IV of the CE plane, $b(\lambda) < 0$ regardless the value of λ, and treatment is said to be cost-ineffective and should not be implemented. However, these situations are the simplest. The other two situations occur in quadrants I and III and are the most common cases. In these quadrants, $b(\lambda) > 0$ and $b(\lambda) < 0$, depending on the values of λ, but the problem in these cases is that there is no general agreement regarding the value of λ. This value is what determines whether the T should be implemented or not. When $b(\lambda) > 0$, T is deemed cost-effective and thus should be selected for implementation. Similarly, when $b(\lambda) < 0$, then T is said to be cost-

ineffective. In the CE plane, Figure 2.4, for all the values above the line, $b(\lambda) < 0$, whereas for all the values below the line, $b(\lambda) > 0$.

Earlier in this chapter the concept of cost minimization was introduced as well as the fact that this method is a special case of the INB approach (Willan and Lin, 2000). When $\lambda = 0$, the INB formula reduces to $b(0) = -\Delta_c$, which is the formula for cost-minimization (see Figure 2.5). In addition, another important feature of this formula is that $b(R) = 0$ (i.e. INB is zero when the WTP = ICER) (Willan and Lin, 2000).

2.3.3.1 Estimation of Confidence Intervals for INB

Estimating confidence intervals for cost-effectiveness (CE) ratios is complicated because of the statistical problem of estimating a confidence interval for a ratio between two random variables (Stinnett and Mullahy, 1998). However, by explicitly incorporating the WTP into the analysis, the ratio estimation problem can be avoided. Multiplying the effectiveness units by λ converts the effectiveness units to the same units as costs (Willan and Lin, 2000). Therefore, the advantage of the net benefits approach is that the (1-α)100% confidence intervals can be easily determined using standard statistical techniques (Briggs and Fenn, 1998).

If all patients are followed until death or for the entire duration of interest, then the observed measures of effectiveness and cost are not censored (Willan and Lin, 2000). In this case the sample means, variances and covariances can be used to estimate the model parameters in statistical standard fashion.

The estimator of net benefit is a linear combination of two asymptotically normal random variables. In the parametric approach, the estimator of $b(\lambda)$ and its estimated variance are given by the following formulas,

$$\hat{b}(\lambda) = (\hat{\mu}_T - \hat{\mu}_S)\,\lambda - (\hat{v}_T - \hat{v}_S) = \hat{\Delta}_e \lambda - \hat{\Delta}_c \qquad (2.9)$$

$$\hat{V}\big[\hat{b}(\lambda)\big] = \sum_{j=S,T} \frac{1}{n_j}\big(\hat{\sigma}_j^2 \lambda^2 + \hat{\omega}_j^2 - 2\hat{\sigma}_j \hat{\omega}_j \hat{\rho}_j \lambda\big) \qquad (2.10)$$

The $(1-\alpha)100\%$ confidence limits are given by,

$$\hat{b}(\lambda) \pm z_{(1-\alpha/2)} \big\{\hat{V}\big[\hat{b}(\lambda)\big]\big\}^{\frac{1}{2}} \qquad (2.11)$$

where $z_{1-\alpha/2}$ is the $(1-\alpha/2)100\text{th}$ percentile for the standard normal distribution (Briggs and Fenn, 1998; Willan and Lin, 2000).

2.3.3.2 Net Benefit- Analysis

In INB one is interested to find whether treatment is cost effective, and thus the hypotheses for net benefit analysis to be tested are,

$$H_0 : \lambda \Delta_e - \Delta_c \le 0 \qquad \text{versus} \qquad H_1 : \lambda \Delta_e - \Delta_c > 0$$

The estimators of $b(\lambda)$ and $V(b(\lambda))$ can be calculated using equations (2.9) and (2.10), and H_0 and H_1 can be tested using a one sided z-test at level α, where z is given by,

$$Z = \frac{\hat{b}(\lambda)}{\sqrt{\hat{V}(\hat{b}(\lambda))}} \qquad (2.12)$$

The confidence intervals are calculated using equation (2.11).

It is important to note that these approaches calculate the confidence intervals for the net benefit analysis assuming no censoring. An example with an uncensored data is provided in Chapter 4.

2.3.3.3 Advantages and Disadvantages of INB

An advantage of the INB in comparison with the ICER is that costs and benefits for a treatment can be compared directly because they are both in monetary units (Diener et al., 1998; Willan and Lin, 2000).

The other important advantage of INB over the ICER approach is that the negative net benefit values are properly ordered in the quadrants where $R < 0$ (Willan and Lin, 2000). The same example used to explain the problems with negative ICERs is used in Figure 2.6. By projecting parallel lines to the standard line with slope $= \lambda$, it is possible to obtain the order of the different INB values against the vertical axis. The result in this case is that INB2 is better than INB3, which is better than INB1. This implies that T2 is more cost effective than T3, and this is more cost effective than T1, which as indicated in section 2.3.2.3 of this chapter is the correct order (see Figure 2.6).

The main drawback of the net-benefits approach is the difficulty inherent in obtaining valid values for the WTP for decision-making purposes. Hence, it is only possible to interpret the estimated net-benefit statistic in terms of the value of WTP used in its definition (Briggs and Fenn, 1998). Different approaches to overcome this obstacle are presented in Chapter 3.
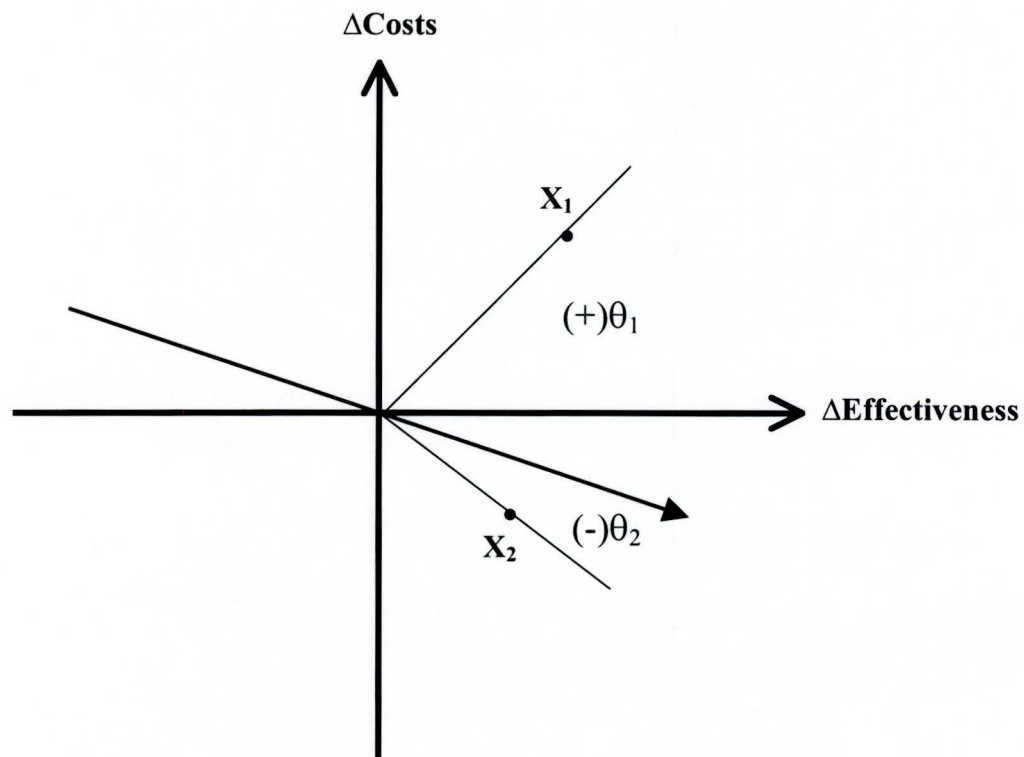
Figure 2.1: Cost-Effectiveness plane showing the definition of θ_k for the bootstrap method.

ΔCosts

ΔEffectiveness

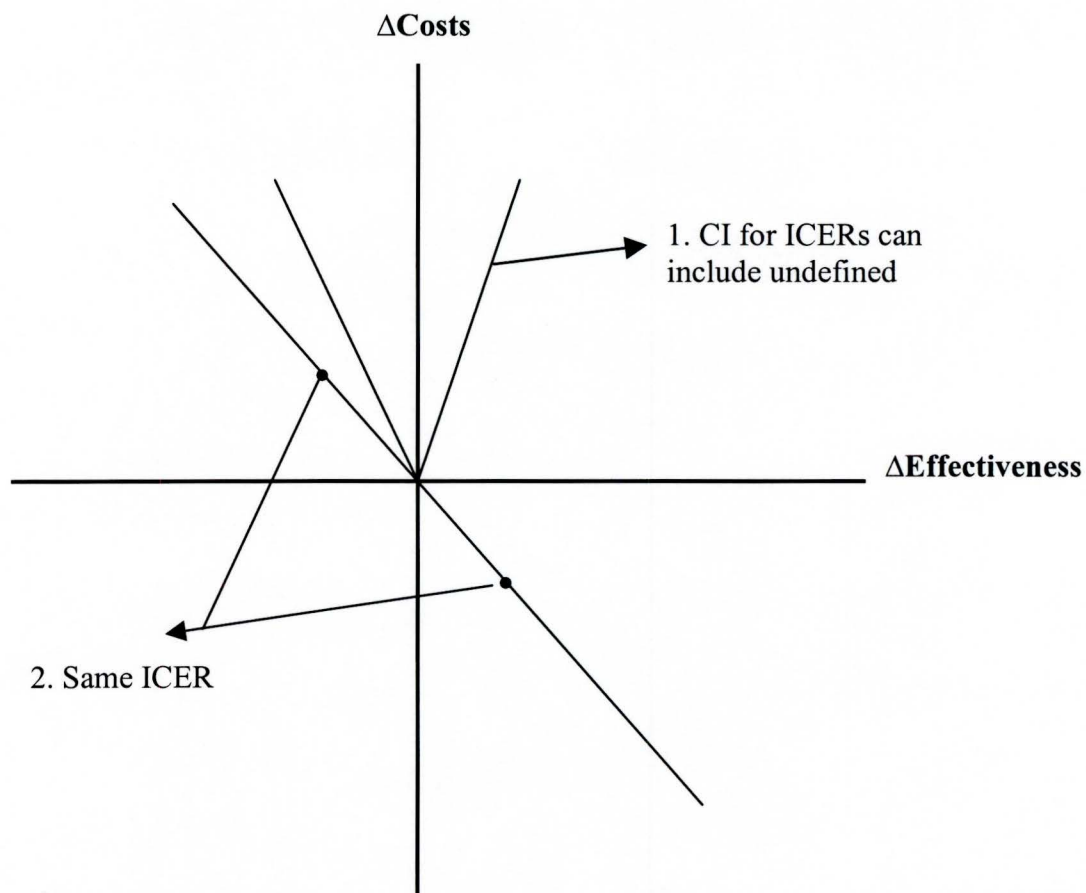1. CI for ICERs can include undefined

2. Same ICER

Figure 2.2: Cost-effectiveness plane and some of the problems with ICERS. It can be seen in the graph that sometimes the confidence intervals of ICERs can include undefined values, or the same value for an ICER can have two totally different interpretations as it does in the graph; it can be in the win-win quadrant or in the lose-lose quadrant.

Table 2.1: Example of ordering three treatments using ICERs. This is an example showing the effectiveness (yr) and costs ($1000) for three treatments where the change for each is also presented. Using these values the ICERs are calculated (ICER $= \Delta_c / \Delta_e$). All of these values are in the same quadrant.

	Effectiveness (yr)	Cost ($1000s)	
T1	5	8	
T2	7	6	
T3	7	8	
S	3	10	**ICER**
Δ1	2	-2	-1
Δ2	4	-4	-1
Δ3	4	-2	-0.5

ΔCosts

ΔEffectiveness

3

1

slope = -0.5

2

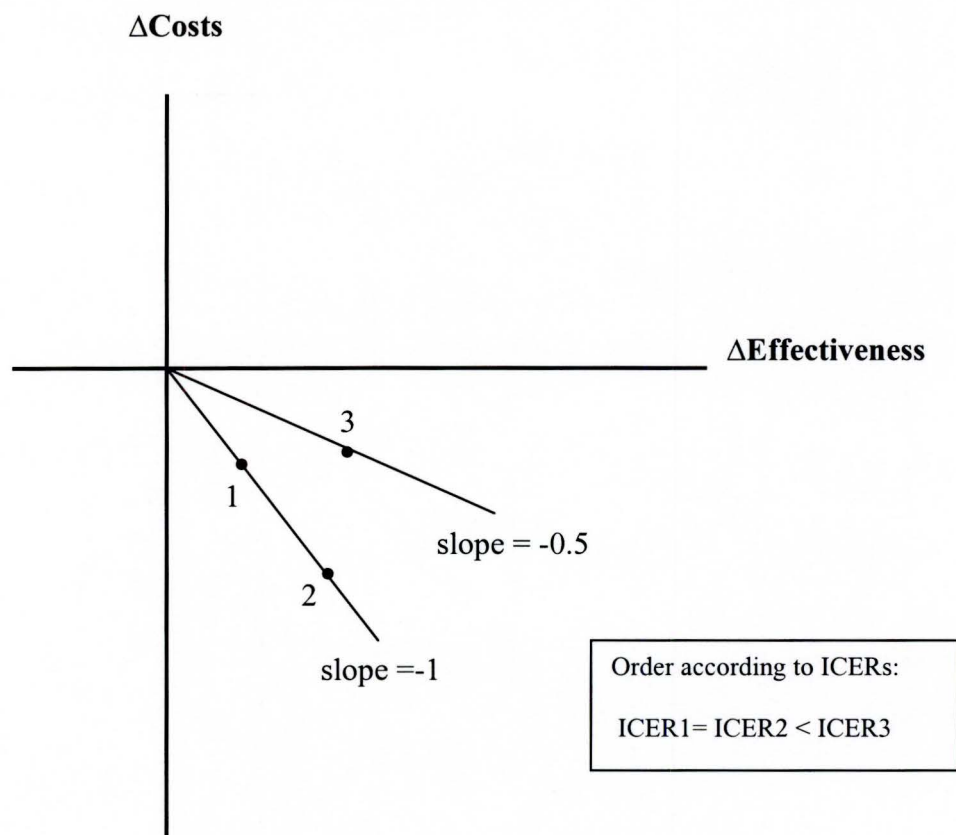slope =-1

Order according to ICERs:

ICER1= ICER2 < ICER3

Figure 2.3: Schematic representation of ordering of treatments using ICERs. Another problem with the ICERs is that they are not properly ordered outside the trade-off quadrants. The ICERs for the three treatments compared in Table 2.1 were placed in the CE plane. The order according to the ICERs is that ICER1 is equal to ICER2 and that both are less than ICER3. The real order is that T2 is more cost effective than T3, which is more cost effective than T1.
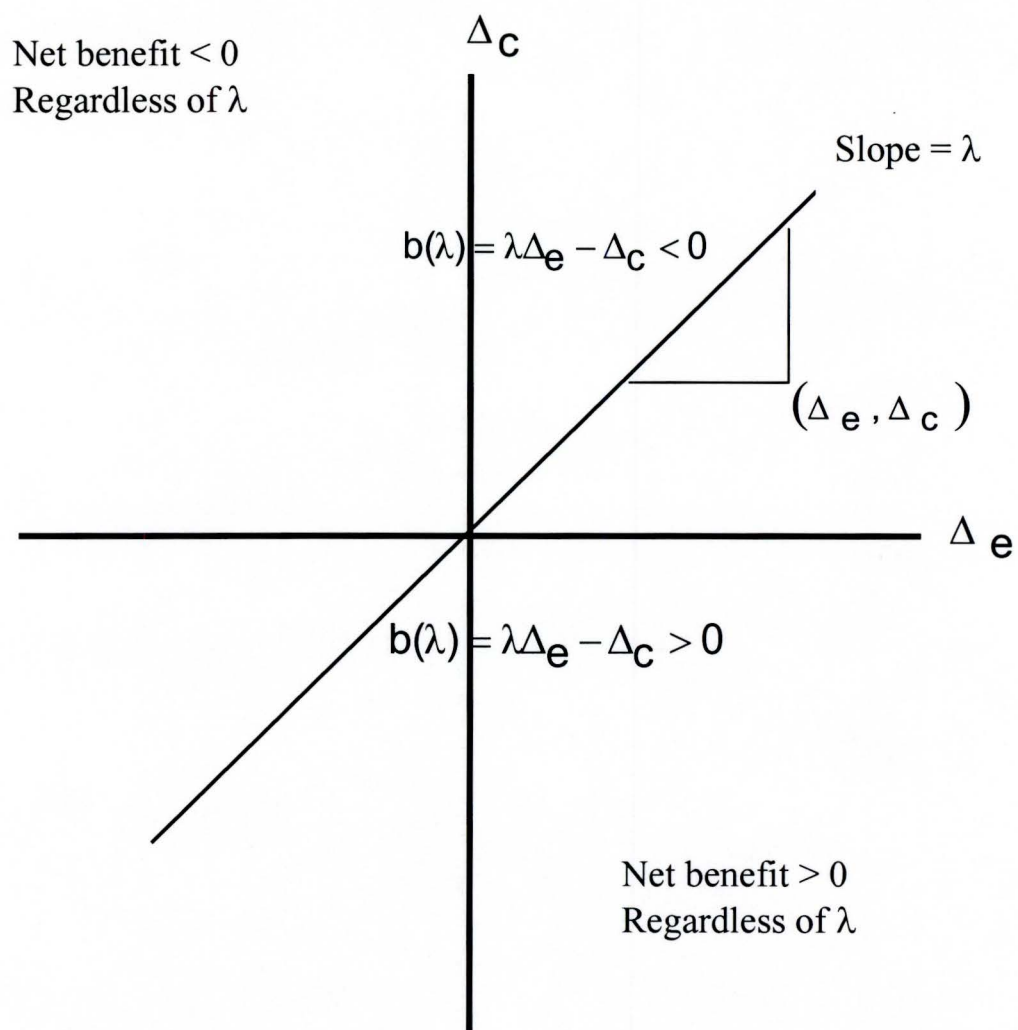
Figure 2.4: Graphical representation of INB in the CE plane.
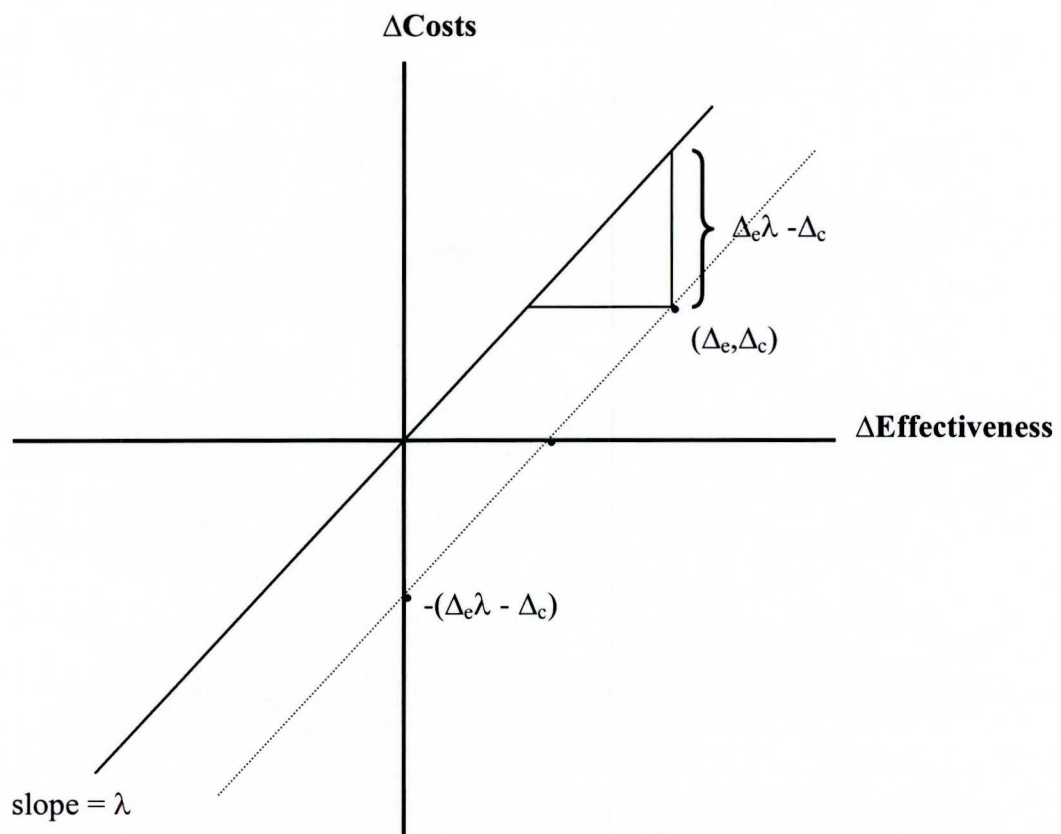
Figure 2.5: INB is the vertical distance between (Δ_e, Δ_c) and the line through the origin with slope λ.
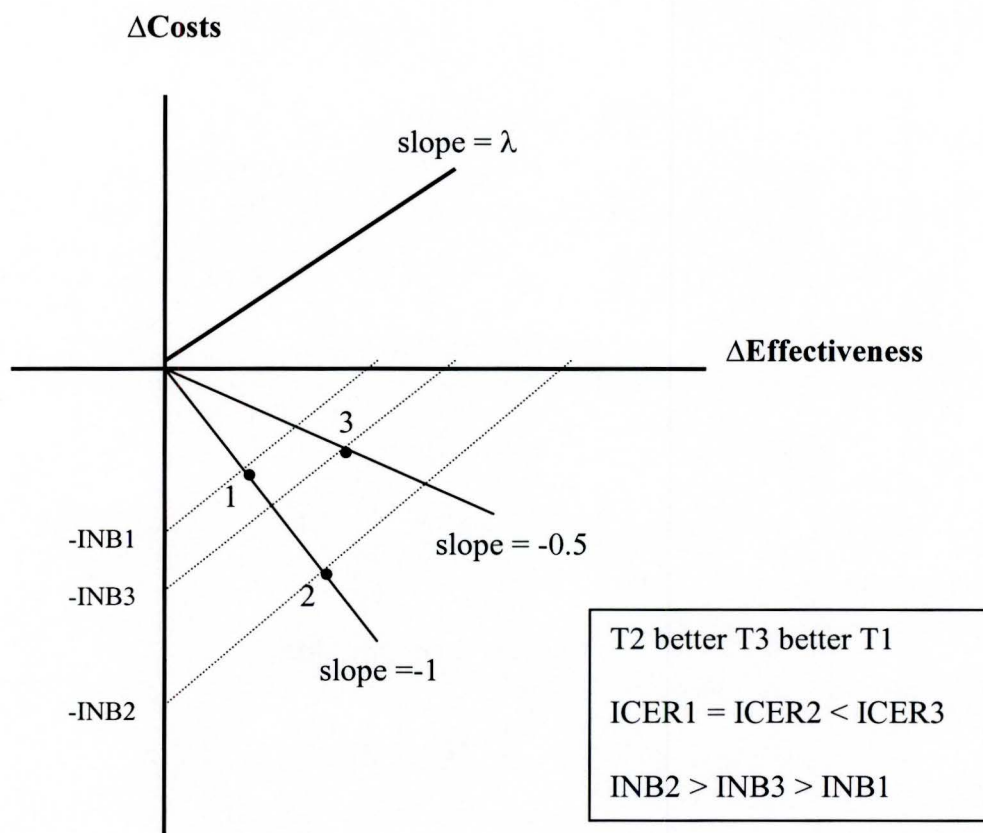
Figure 2.6: Schematic representation of ordering of treatments using INB. In this cost-effectiveness plane the same example used to state the problems with ICERs is used to present one of the many advantages of INB values. The net benefit values, unlike ICERs, are properly ordered outside trade-off quadrants. The results with this approach are, INB2 is better than INB3, which is better than INB1, indicating that T2 is more cost effective than T3 which is more cost effective than T1. The order obtained with the INB method is the same as the real order.

CHAPTER 3

WILLINGNESS TO PAY AND WILLINGNESS TO ACCEPT

3.1 Health's Unique Nature

It is a legitimate and important task to study consumers' willingness-to-pay for health care products or programmes. The purpose of this research is to determine the relationship between the price and the quantity of the product demanded (Gafni, 1991). However, measuring the benefits of health and lifesaving programs is a difficult task due to the special nature of health and the fact that there is not an actual market where such studies can be done.

The most obvious distinguishing characteristic of a person's demand for health care is that, unlike most goods and services, it is unpredictable (Arrow, 1971). Another distinguishing characteristic of health is that the provision of medical services does not always result in a cure (Gafni, 1991). In addition, from the perspective of the individual patient, the outcome of any health care intervention is based on probability (Arrow, 1971; Weinstein et al., 1980). Therefore, the insurance procedure or the social decision based on expected values ignoring the risk of the individual cannot be applied to the case of health care programs (Gafni, 1991) and thus other methods are necessary to account for such factors.

3.2 Methods used in the Economic Evaluation of Health Programs

Generally, most economic evaluations are performed primarily to inform decision-makers about the relationship between the costs and benefits of new and existing technologies (Ament and Baltussen, 1997). However, one major obstacle to the optimal allocation of health care resources is that the value that society attaches to health (or a unit of effectiveness) is not directly observable. In the literature, this value is referred to as the "willingness-to-pay" (WTP), 'cut off' level of permissible cost or "willingness-to-accept" (WTA) per unit of effects (Ament and Baltussen, 1997). It is important to note the value that society attaches to a unit of effectiveness has a high impact on the priority ranking of medical interventions (Ament and Baltussen, 1997).

At the most basic level, cost-benefit analysis requires the monetary valuation of all the effects of a programme on the welfare of all individual members of society. If this could be done for all health care programmes, a net benefit value (benefit minus resource cost) could be calculated for each programme, such calculations are of considerable use in health care policy making (Donaldson, 1990). However, no consensus has been reached on the methods that should be used to calculate such a value.

One way to address the uncertainty problem of monetary valuation in health care programs and studies is by undertaking an economic evaluation of the procedure (Lindholm et al., 1994; Neumann and Johannesson, 1994). Traditionally this was done using CE analysis where the cost per unit of health effects such as life-years or quality-adjusted life-years (QALYs) gained is estimated (Neumann and Johannesson, 1994). An alternative method is to carry out an INB analysis where the value of health interventions

are measured based on individuals' WTP (Lindholm et al., 1994; Neumann and Johannesson, 1994; Willan and Lin, 2000). In theory such an approach is consistent with individual preferences and welfare economics (Lindholm et al., 1994). The recent growing interest in the application of INB analysis as a technique for the economic evaluation results from the fact that INB expresses costs and benefits in the same units (Diener et al., 1998; O'Brien, 1998).

Recent literature on economic evaluation in health care has shown that a popular method for estimating values for health care programs is the use of WTP or WTA survey techniques known as Contingent Valuation (CV) (Gafni, 1991; O'Brien and Gafni, 1996; O'Brien, 1998; Olsen and Smith; 2001). Willingness-to-pay questions are used to evaluate the benefits of a given program, whereas WTA should be used to evaluate the costs (Gafni, 1991). The challenge for CV is to elicit a monetary value for the benefits of a programme as if a market for such programme benefits did exist, in essence to "replace" the missing market (Diener et al., 1998; Gafni, 1991; Olsen and Smith, 2001).

Economists have long measured individual WTP by examining the price of goods and services bought and sold in the marketplace (Neumann and Johannesson, 1994). It is more difficult, however, to measure the value of commodities, which are not typically traded in private markets. Furthermore, attempts to place values on "priceless" items, such as health, have sometimes been controversial. Observers have argued, for example, that there are some items on which attaching a value or putting a price on a non-market item may itself reduce its value (Kelman, 1981). Nonetheless, some researchers have

turned to survey methods to investigate the WTP factor for such items (Diener et al., 1998; Neumann and Johannesson, 1994).

Contingent valuation surveys involve posing questions such that responses are contingent upon hypothetical markets described to respondents (Diener et al., 1998; Jones-Lee et al., 1985; Mitchell and Carson, 1989). Although developed and primarily used to value environmental changes, researchers have also used this methodology to examine how patients value health programs ranging from mobile coronary care units to ultrasound, hypertension treatment and *in vitro* fertilization (IVF) (Appel et al., 1990; Berwick and Weinstein, 1985; Donaldson, 1990; Eastaugh, 1991; Johannesson, 1991; Johannesson et al., 1993; Thompson et al., 1984).

Contingent valuation involves asking individuals directly in a hypothetical survey the maximum amount they are WTP to have the commodity in question, or the minimum amount they would be WTA in compensation to be deprived of it. Among the various theoretical measures that exist WTP is the most widely applied (Bailey, 1980; Diener et al., 1998; O'Brien and Gafni, 1996; Olsen and Smith, 2001).

Economic analysts have long argued that the benefits of a program are best measured by the amount that potential beneficiaries would be willing to pay for the program in question (Appel et al., 1990; Landefield and Seskin, 1982). So far, there are two methods available for measuring WTP, direct survey and revealed preferences (O'Brien and Gafni, 1996; O'Brien, 1998). The first method involves direct questioning of persons about the amount they are willing to pay to reduce the risk of death or improve the quality of life (O'Brien, 1998). This amount, usually expressed in dollar terms,

represents how much of other goods and services a person is willing to give up. The second method is an indirect measurement and involves inferring from a person's behavior what amount he or she is willing to pay for such gains (Gafni, 1991).

3.3 WTP Estimation Techniques

Of the two methods available for measuring WTP, the revealed preferences approach, in which one infers WTP from the actual market behavior of individuals is preferred by most economists (Appel et al., 1990; Jones-Lee, 1976; Landefield and Seskin, 1982; O'Brien and Gafni, 1996). However, there is considerable conceptual and statistical variation in the assessment of WTP, which is why there are more methods of elicitation for WTP, and each approach has strengths and weaknesses but a common goal, to reveal the underlying monetary valuation. Each estimation technique varies in terms of measurement properties of precision and bias. Some of the most common ones are:

3.3.1 Open-Ended Questions

The measurement task is to find out the maximum that an individual (or group) would be willing to pay for the new program. The open-ended question format is the most direct format for determining this value from an individual. However, the open-ended format tends to produce large numbers of non-responses or protest zero responses to WTP questions and for these reasons choices are presented in the surveys to make the market scenario more realistic (O'Brien, 1998).

3.3.2 Bidding Games

Bidding games are the oldest and until recently the most widely used elicitation method in CV survey techniques (Mitchell and Carson, 1989). This procedure is similar to an auction; an initial starting money value is bid up or down by the respondent. This technique has not only market realism as the respondent requires a Yes/No response at each bid level, but also enables the respondent to consider more fully the value of the program (Hoehn and Randall, 1987). A major disadvantage is the potential for bias because the starting bid, which is chosen by the researcher, tends to imply a value for the good (Mueser and Dow, 1997; O'Brien, 1998).

3.3.3 Payment Cards

This technique was developed as an alternative to the bidding games procedure (Mitchell and Carson, 1998). The payment card is a visual aid, which contains a large array of potential WTP amounts ranging from $0 to some large amount (O'Brien, 1998). A related technique is the checklist method, where respondents indicate which of a list of payment ranges includes their WTP amount (O'Brien, 1998).

3.3.4 Dichotomous Choice (take it or leave it)

This method uses a large number of predetermined prices and each respondent is asked if he/she is willing to pay a single one of these prices (Yes or No) for the program with no further iteration (Bishop and Heberlin, 1979). The prices are randomly assigned to respondents so that one can use statistical techniques such as probit analysis to model

bid acceptance as a function of respondent characteristics and determine median WTP (Cameron and James, 1987). A disadvantage in this approach is that it requires a much larger sample size for the same level of statistical precision of the other methods (O'Brien, 1998).

3.3.5 Double-Bounded Dichotomous Choice

This technique, also known as take it or leave it with follow-up, is an extension to the previous method to improve its statistical efficiency. Here the follow-up bid question is asked from the respondent, higher or lower, conditional upon the response to the first bid; the higher or lower bid is randomly selected from a range (Hanemann et al., 1991). In this case probit analysis can also be used for the analysis (Cameron and James, 1987).

3.3.6 Conjoint Analysis

Conjoint Analysis is a method used widely in consumer economics to establish consumer preferences over attributes of commodities (Green and Krieger, 1996). The method works by first defining a number of attributes of the product or program and then asking the respondent to choose between hypothetical pairs (e.g. program A vs. program B) that vary in their attribute composition and where one of the attributes is how much the individual would have to pay. Conjoint analysis has been used successfully in the evaluation of consumer products where health outcomes are attributes (Magat et al., 1996).

Contingent valuation survey researchers stress that WTP questions should be constructed so that respondents view their answers as accurate representations of how they would behave if confronted with an actual market for the good (Mitchell and Carson, 1989; Neumann and Johannesson, 1994; Thompson et al., 1984; Thompson, 1986). Thus, the scenario presented should be plausible and meaningful to respondents. The survey should present details about the hypothetical market, including conditions for provision of the good, the frequency of payments, and who will have access (O'Brien, 1998). Others stress that the survey should assure subjects of anonymity and confidentiality, and ensure that respondents understand that "willingness-to-pay" means their own personal willingness to pay (Neumann and Johannesson, 1994; Thompson, 1986).

As mentioned earlier, these methods are vulnerable to a number of biases. For instance, starting point bias or "anchoring" may occur if respondents are influence by the initial amount presented to them. Responses may also be influenced by the order or manner in which questions are asked. Other biases may occur if respondents incorrectly perceive the market or good being valued or lack other cognitive capabilities to understand the questions (Neumann and Johannesson; 1994).

The CV methodology is still experimental and has a number of limitations such as possibility of response bias with the information given to respondents (O'Brien, 1998). However, this technique remains as the most potentially useful tool in understanding how people value items not typically traded in private markets.

3.4 The WTP versus WTA disparity (WTA>WTP)

One of the most robust findings in experimental analysis in economics is that subjects often display a large discrepancy between the dollar value they are willing-to-accept in order to sell an item (WTA) and the dollar value they are willing-to-pay to purchase it (WTP) (Kahneman et al., 1990; Mueser and Dow, 1997).

A recent review by Diener et al. (1998) states that most monetary health valuation studies have measured compensating (rather than equivalent) variation using WTP rather than WTA. Another recent publication (O'Brien, Goeree, Gafni, et al., 1998), which is the first health study to measure both WTP and WTA for the same probabilistic increase/decrease in health, showed that the mean WTA is 2.7 times greater than WTP. This result is not surprising when set in the broader context of the non-health contingent valuation literature, where WTA is repeatedly found to exceed WTP (Morrison, 1997; Shogren et al., 1994).

The difference between WTP and WTA has also been studied experimentally, and these studies suggest that paying more may accurately reflect market values than accepting, but that the two measures tend to converge in a mature market setting (Coursey et al., 1987; Harless, 1989; Lindholm et al., 1994).

The context for the WTP-WTA disparity is that, in theory, and aside from small income effects, the Hicksian measures of consumer surplus measured by WTP and WTA should be the same (Cummings et al., 1995; Mueser and Dow, 1997). In practice they differ, with WTA exceeding WTP by quite a margin depending upon the benefit being

valued. Several explanations have been proposed to explain this observed gap, some of the most accepted are:

3.4.1 Measurement Artifact

Since the argument is that all contingent valuation studies suffer from measurement biases, this is particularly true for WTA where respondents are asked (often unrealistically in the case of public goods) to assume a property right over the commodity being valued (Mueser and Dow, 1997). This problem of hypothetical bias is compounded by potential strategic bias, such that a respondent may "bid up" WTA compensation if they have some concerns or suspicions about who is compensating for the loss (e.g. industry or government) (O'Brien et al., 2001).

3.4.2 Substitute Commodities

In a modification to economic theory, Hanemann (1991) proposed that the degree of disparity between WTA and WTP depends upon the availability of substitute commodities. Specifically, when there are few substitute commodities (for instance public goods such as clean air in environmental studies) the prediction is that the WTA-WTP difference will be larger.

3.4.3 Endowment Effect

Based upon the original work of Tversky and Kahneman (1991) on loss aversion and framing effects. This is a psychological theory to explain the disparity, which assumes that the utility of a loss is greater than the utility of an equivalent gain. According to this theory, receipt of ownership, an "endowment", changes the subject's reference point, not only shifting a subject's position on the indifference map but also altering the shape of the indifference curves (Fox and Tversky, 1995; Hanemann, 1991). In this framework the compensation required to compensate for a loss (WTA) exceeds the amount a person would pay for the equivalent gain that has lower utility value.

3.5 Methods for Cost-Effectiveness Analysis when the WTA is greater than the WTP

In order to use cost-effectiveness data for making decisions, it is necessary to make a valuation on society's willingness-to-pay for a unit of health such as a quality-adjusted life-year (QALY). Based on the repeated observations found in the literature that WTA>WTP, O'Brien and others (accepted for publication) argue that there is a good reason to question the current convention of assuming a fixed threshold for accept-reject decisions for both quadrants I and III of the CE-plane(see Figure 3.1). Based on their empirical review, they argue that the "selling price" of a QALY is at least twice as great as its "buying price" (see Figure 3.2). This value disparity is best explained by an endowment effect, although experimental work continues to attempt to test the validity of the other competing themes presented above (Morrison, 1997; Morrison, 1998).

The observation that WTA exceeds WTP has an impact on the decision rules of cost effectiveness. The main implication being that the "reject" region in quadrant III becomes larger if one recognizes that relinquishing QALYs requires greater compensation (see Figure 3.2) (O'Brien, accepted for publication). To our knowledge the economic evaluation literature has not previously presented methods that accommodate the implications of the WTA/WTP disparity. In this thesis we present a modified version of the INB, which accounts for this disparity and its implications in terms of cost effectiveness. Specifically, we took into account the "kink" in the accept-reject threshold for cost-effectiveness in the CE plane (see Figures 3.1 and 3.2) which is a result of the WTA/WTP disparity.

Currently, the threshold through the origin of the CE plane in Figure 3.1 is drawn as a straight line with constant slope λ in both quadrants I and III, meaning that the acceptable "buying price" and "selling price" of quality-adjusted life-years (QALYs) is assumed to be same. But if WTA>WTP, then the compensation required to relinquish QALYs should be greater than what we would be willing to pay to acquire them. This would imply a "kink" in the cost effectiveness threshold with two different values of λ, depending on whether the treatment increases effectiveness (λ) in quadrant I (λ_{QI}) or decreases effectiveness ($\gamma\lambda$), where $\gamma>1$, in quadrant III (λ_{QIII}) (see Figure 3.2).

In Chapter 2, the method of net benefit analysis, which explicitly incorporates the monetary value of health benefits, was introduced; in this Chapter we show that this method can be adapted to accommodate the WTA/WTP disparity.

The model introduced in Chapter 2 for a two–arm randomized control trial, where e_{ji} and c_{ji} are the respective measures of effectiveness and cost for patient i on therapy j, is also used to introduce the statistical methods for the cost-effectiveness analysis to allow for the WTA/WTP disparity. To conduct a cost-effectiveness analysis the quantities Δ_e and Δ_c must be estimated; their estimators are denoted by $\hat{\Delta}_c$ and $\hat{\Delta}_e$, respectively. In addition, the quantities $V(\hat{\Delta}_e)$, $V(\hat{\Delta}_c)$ and $C(\hat{\Delta}_e, \hat{\Delta}_c)$, where C is the covariance function, must be estimated. The methods to estimate these variables for uncensored data were presented in Chapter 2.

Firstly, a Bayessian approach is adopted assuming normality (Briggs, 1998; Willan et al., 2001), and using a non-informative prior, the posterior distribution, $\Pr(\text{INB}(\lambda) < b)$ is given by

$$G(b) = \int\limits_{-\infty}^{\infty} \int\limits_{\lambda e - b}^{\infty} f(e, c)\, dc\, de$$

where f is the density function of a bivariate normal random variable with mean $(\hat{\Delta}_e, \hat{\Delta}_c)$ and variance-covariance $\begin{pmatrix} \hat{V}(\hat{\Delta}_e) & \hat{C}(\hat{\Delta}_e, \hat{\Delta}_c) \\ \hat{C}(\hat{\Delta}_e, \hat{\Delta}_c) & \hat{V}(\hat{\Delta}_c) \end{pmatrix}$. The quantity $G(b)$ is the density contained in the shaded area in the left side of Figure 3.3. Defining $b_{G\theta}$ such that $G(b_{G\theta}) = \theta$, $\text{INB}(\lambda)$ is estimated by $b_{G, 0.5}$, with $100(1-\alpha)\%$ Bayesian interval given by $b_{G, \alpha/2}$ and $b_{G, 1-\alpha/2}$. This confidence interval is presented in the right side of Figure 3.3. Because of the assumption of normality and the use of a non-informative prior,

$b_{G,0.5} = \hat{\Delta}_e \lambda - \hat{\Delta}_c$ and the Bayesian interval is identical to the confidence interval given in Section 2.3.3.1.

In order to adjust the statistical analysis for the situation where the WTA is greater than the WTP, the Bayesian approach is extended for WTA $= \gamma$WTP (Willan et al., 2001) where γ is a proportionality constant. It represents the ratio by which WTA is greater than WTP. If the WTA $= \gamma$WTP, where $\gamma>1$, then the posterior distribution, $\Pr(INB(\lambda)<b)$ is given by

$$H(b) = \int_{-\infty}^{0} \int_{\gamma\lambda e - b}^{\infty} f(e,c)\,dc\,de + \int_{0}^{\infty} \int_{\lambda e - b}^{\infty} f(e,c)\,dc\,de$$

This area is shown as the density in the shaded area in the left side of Figure 3.4. It is important to note that the slope of the border of the shaded area to the left of the vertical axis is equal to $\gamma\lambda > \lambda$, and, therefore, more of the cost-effectiveness plane is included in the region in which $INB(\lambda, \gamma) < b$. Since $\gamma > 1, H(b) > G(b)$ for all b, consequently $b_{H\theta} < b_{G\theta}$ for all θ, where $b_{H\theta}$ is defined such that $H(b_{H\theta}) = \theta$. Thus as γ increases, the estimate of $INB(\lambda)$ and the Bayesian limits decrease, this effect can be appreciated in the graph located in the right side of Figure 3.4, here the two curves are presented one where $\gamma=1$ and the other with $\gamma >1$.

The way this adjustment works and why it works can be explained by comparing the net benefit analysis when $\gamma=1$ with $\gamma >1$. This can be accomplished by drawing parallel lines in the cost-effectiveness plane (see Figure 3.5). In the first graph all the lines are drawn with constant slope $= \lambda$, this assumes WTA=WTP. As the lines are moved parallel

to each other from right to left one can observe that there is a decrease in effect and an increase in cost. Applying the Bayesian approach, the density of the area to the left of these lines decrease and according to the above results, this increases the Bayesian limits and the $INB(\lambda)$ estimate. In the second graph, however, the parallel lines have different slopes assuming that WTA>WTP by some proportionality. As these kinked lines are moved parallel to each other from right to left there is still a decrease in effectiveness and an increase in cost, but because of the kink in the lines, the increase in cost is less when compared to the other model.

These differences could have important implications in health policy making. For instance in the right panel of Figure 3.4, the curve where the WTA>WTP, the 95% Bayesian CI and the estimate of the $INB(\lambda)$ are shifted to left of the standard curve, making these values smaller. This indicates that a more precise estimate of the $INB(\lambda)$ value with smaller CIs can be obtained with this theory, if there is indeed a difference in the WTA and WTP. In addition, this adjustment theory can help policy makers in decisions of which treatments or programs to replace when they are very close in terms of cost. Differences of one dollar or less can make this process less complex and thus be more beneficial to the public in general. To elucidate more on this topic, an INB analysis of an RCT including this theory is presented in Chapter 4.
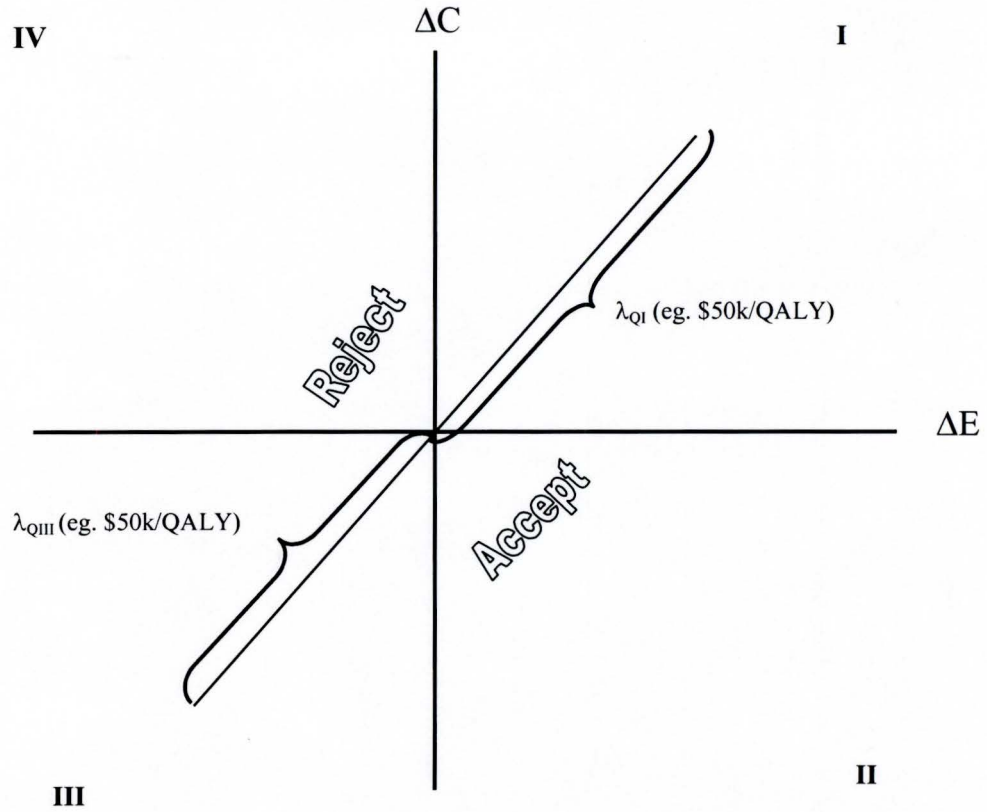
Figure 3.1: The CE-plane and QALY. This CE plane shows the space of incremental cost (ΔC) and incremental effect (ΔE), here represented as quality-adjusted life-years (QALYs). In this figure the arbitrary threshold of $50k per QALY is a straight line through the origin from quadrant I to quadrant II.
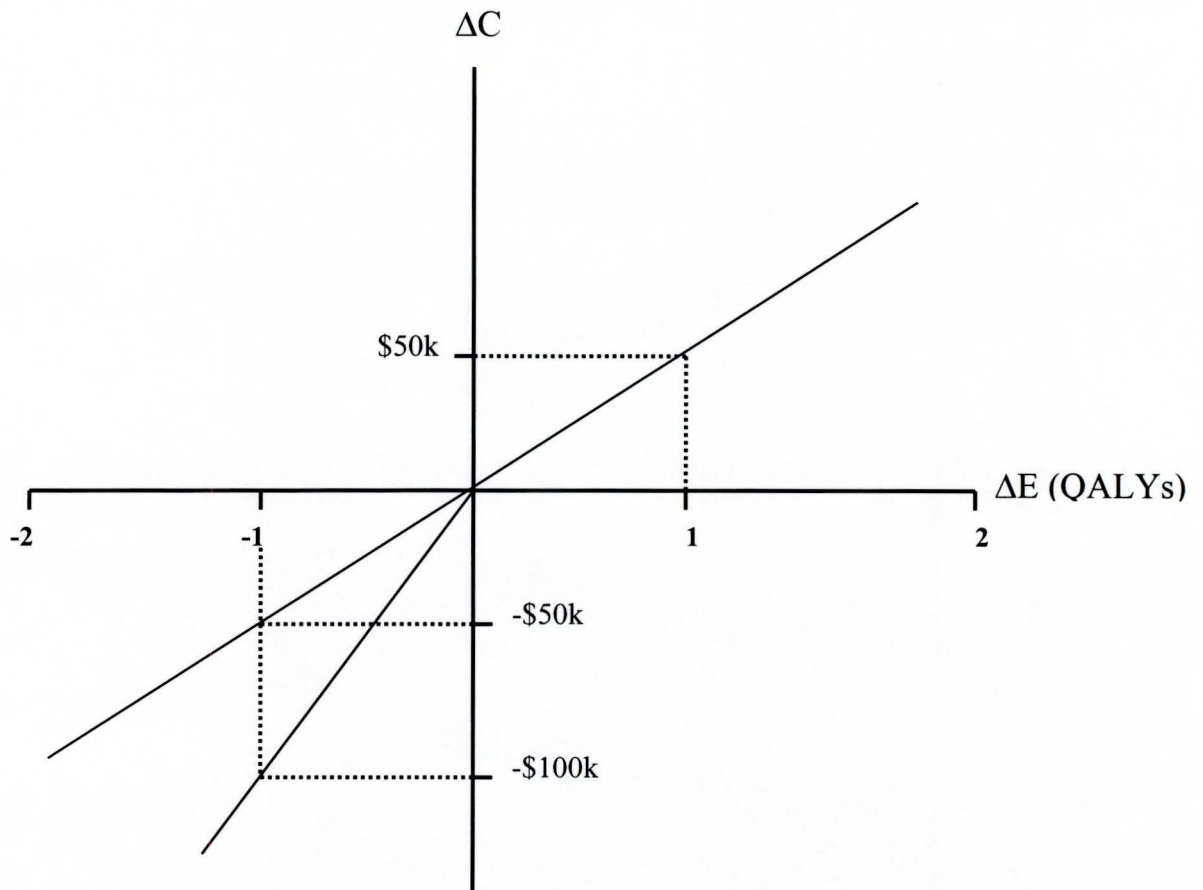
Figure 3.2: Selling QALYs at twice the price a person would be willing to buy them.
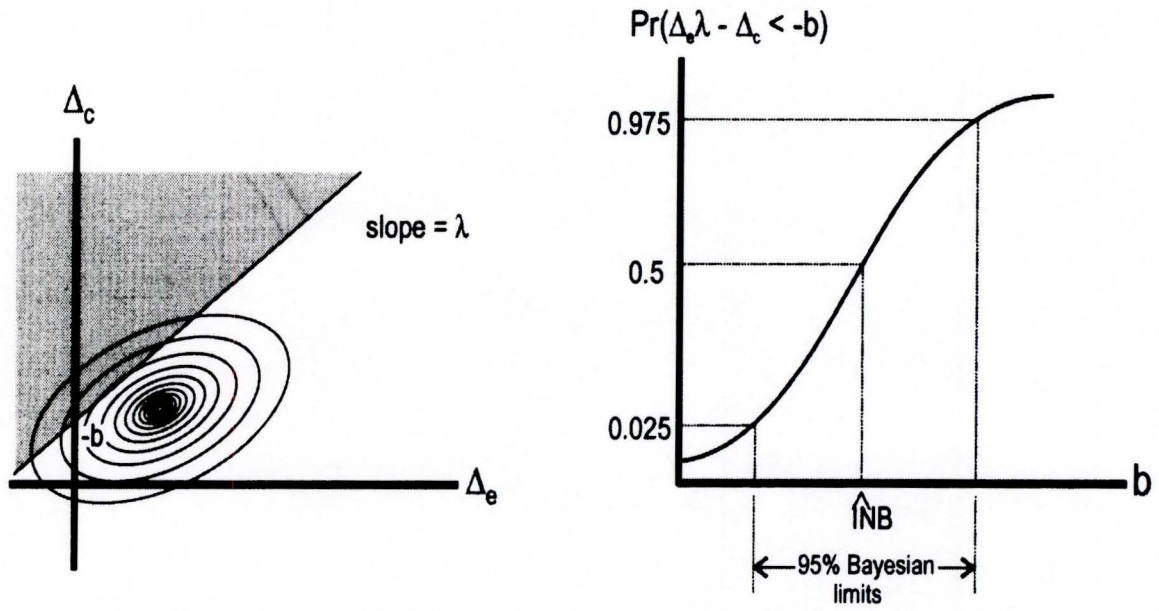
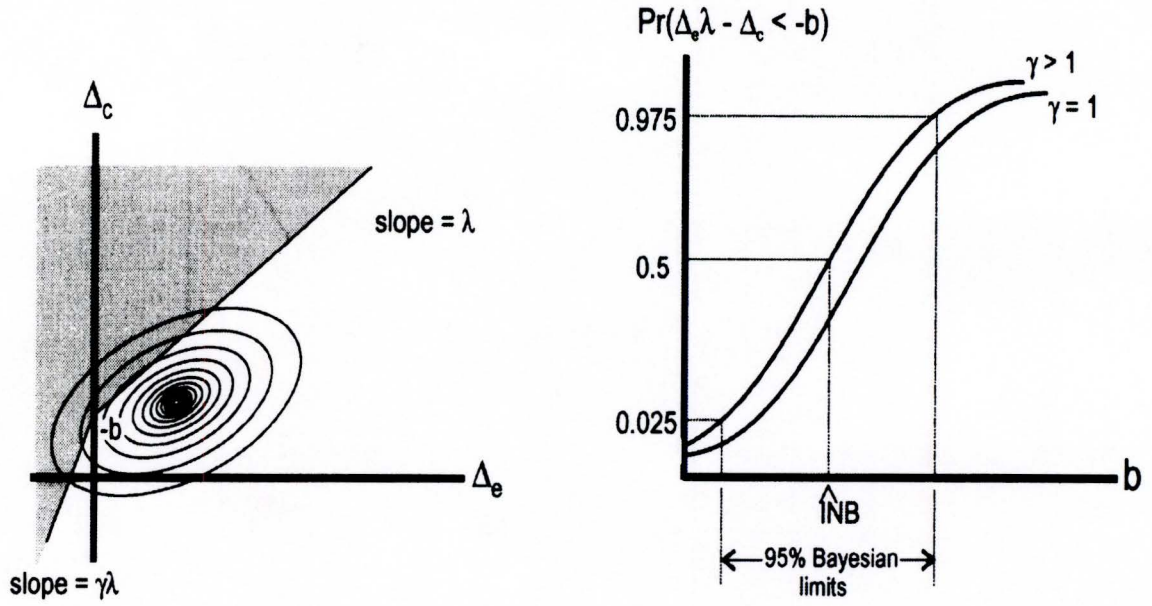**Figure 3.3:** Posterior distribution for net benefit, WTA=WTP

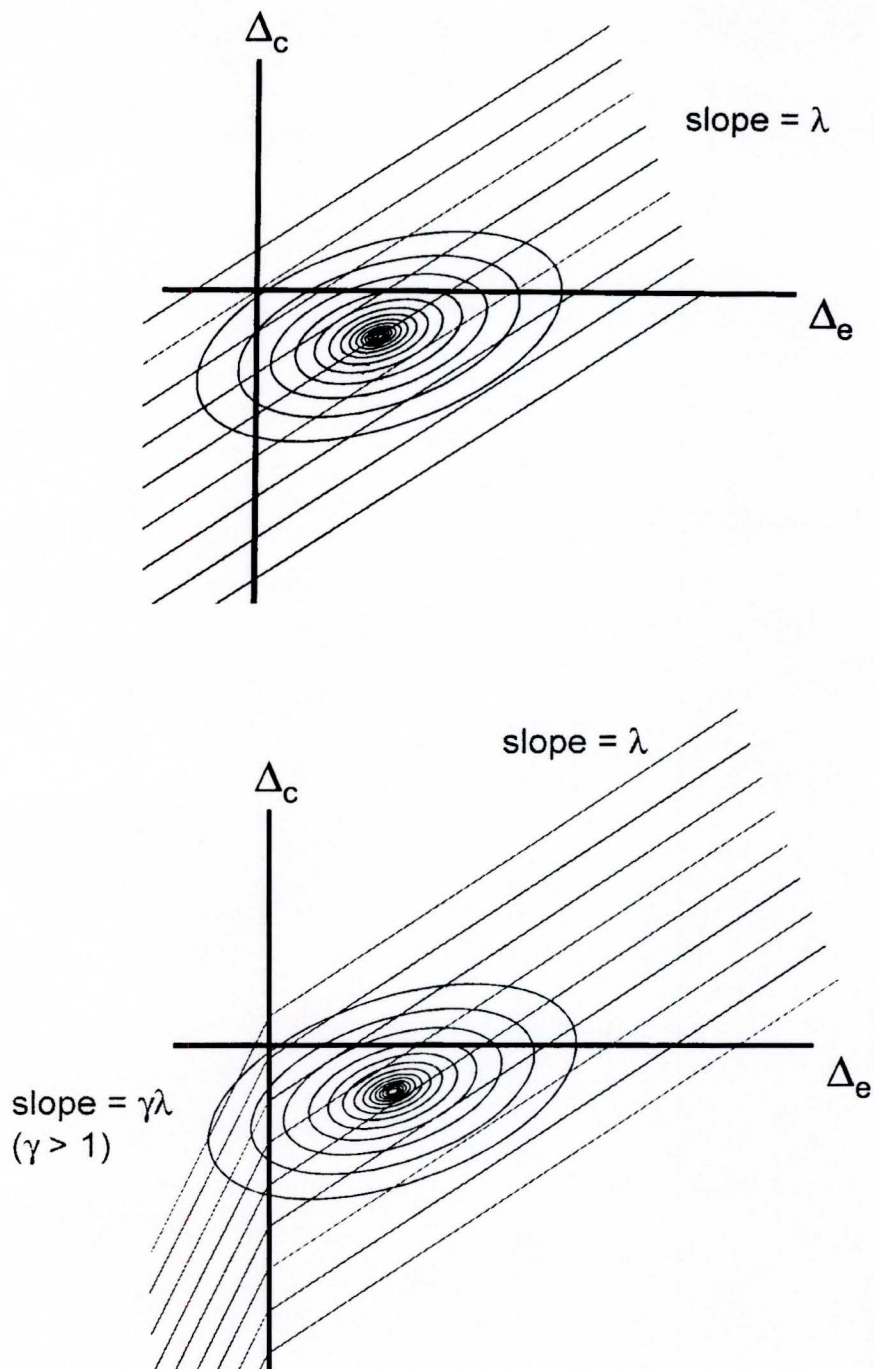**Figure 3.4:** Posterior distribution for net benefit, WTA>WTP

Figure 3.5: Comparison of net benefit analysis when the WTA=WTP with WTA=γ WTP, where γ >1.

CHAPTER 4

EXAMPLE

4.1 Uncensored Data

If all patients are followed until death or for the entire duration of interest, then the observed measures of effectiveness and cost are not censored (Willan and Lin, 2000). In this case the sample means, variance and covariances can be used to estimate the model parameters and the estimator of $b(\lambda)$ and its estimated variance are given by:

$$\hat{b}(\lambda) = (\hat{\mu}_T - \hat{\mu}_S)\,\lambda - (\hat{v}_T - \hat{v}_S) = \hat{\Delta}_e\lambda - \hat{\Delta}_c$$

$$\hat{V}\!\left[\hat{b}(\lambda)\right] = \sum_{j=S,T}\frac{1}{n_j}\left(\hat{\sigma}_j^2\lambda^2 + \hat{\omega}_j^2 - 2\hat{\sigma}_j\,\hat{\omega}_j\,\hat{\rho}_j\,\lambda\right)$$

where,

$$\hat{\sigma}_j = \frac{1}{n_j(n_j - 1)}\sum_{i=1}^{n_j}\left(e_{ji} - \hat{\mu}_j\right)^2$$

$$\hat{\omega}_j = \frac{1}{n_j(n_j - 1)}\sum_{i=1}^{n_j}\left(c_{ji} - \hat{v}_j\right)^2$$

$$\hat{\sigma}_j\hat{\omega}_j\hat{\rho}_j = \frac{1}{n_j(n_j - 1)}\sum_{i=1}^{n_j}\left(e_{ji} - \hat{\mu}_j\right)\left(c_{ji} - \hat{v}_j\right)$$

The 100(1- α)% confidence limits are given by:

$$\hat{b}(\lambda) \pm z_{(1-\alpha/2)}\left\{\hat{V}\!\left[\hat{b}(\lambda)\right]\right\}^{\frac{1}{2}}$$

where $z_{(1-\alpha/2)}$ is the $(1-\alpha/2)100^{th}$ percentile of the standard normal distribution.

Alternatively, identical results can be achieved by defining net benefit for each patient as $b_{ji}(\lambda) = e_{ji}\lambda + c_{ji}$, and defining $\hat{b}_j(\lambda)$ and $\hat{V}_j(\lambda)$ as the sample mean and variance, respectively, the estimator of $b(\lambda)$ and its estimated variance are given by:

$$\hat{b}(\lambda) = \hat{b}_T(\lambda) - \hat{b}_S(\lambda)$$

$$\hat{V}[\hat{b}(\lambda)] = \frac{1}{n_j}\frac{1}{n_j - 1}\sum_{i=1}^{n_j}[b_{ji}(\lambda) - \hat{b}_j(\lambda)]$$

$$= \sum_{j=S,T}\frac{1}{n_j(n_j - 1)}\left[\lambda^2\sum_{i=1}^{n_j}(e_{ji} - \hat{\mu}_j)^2 + \sum_{i=1}^{n_j}(c_{ji} - \hat{v}_j)^2 - 2\lambda\sum_{i=1}^{n_j}(e_{ji} - \hat{\mu}_j)(c_{ji} - \hat{v}_j)\right]$$

Thus, net benefit analysis for uncensored data can be accomplished using common two-sample procedures.

4.2 Incremental Net Benefit for Uncensored Data-Prostate

When analyzing data from a clinical trial, expressing incremental net benefit (INB) in terms of cost allows the investigator to examine all three approaches: cost minimization, ICER and INB in a single graph, complete with the corresponding statistical inferences (Willan and Lin, 2000). This is achieved by calculating $\hat{b}(\lambda)$ and the confidence limits for a wide range of λ, and graphing them as a function of λ. In addition, inference regarding incremental net benefit can be made as a function of λ. The estimate and confidence limits for the difference in costs are given by the vertical intercepts and for the ICER by the horizontal intercepts (see Figure 4.1). To show the above advantage of using INB in terms of cost, we analyzed an uncensored clinical trial with palliative outcomes.

In a trial of symptomatic hormone resistant prostate cancer (HRPC) (Bloomfield et al., 1998; Tannock et al., 1996), 161 patients were randomized between prednisone alone (S) and prednisone plus mitoxantrone (T). Although there was no statistically significant difference in survival, there was better palliation with T (Tannock et al., 1996). Cost data, including hospital admissions, outpatient visits, investigations, therapies and palliative care, were collected retrospectively on the 114 patients from the three largest centres. Survival was quality-adjusted and determined using the European Organization for Research and Treatment of Cancer (EORTC) quality-of-life questionnaire QLQ-C30 (Bloomfield et al., 1998; Tannock et al., 1989; Aaroson et al., 1993). All patients were followed until death.

The sample means and the sample variance-covariance information can be found in Table 4.1. Cost is given in Canadian dollars (CAD$) and effectiveness in quality-adjusted life-weeks (QALW). Using the values given in Table 4.1 $\hat{b}(\lambda)$ and the corresponding 90% confidence limits were calculated for various values of λ and plotted in Figure 4.1. For $\lambda = 1,000$ CAD$/QALW (approximately $50,000/QALY) the estimated of net benefit is $14,517 with confidence interval $3,662 to $25,372. Ninety percent confidence intervals were used to be consistent with a required 5% test of the hypothesis $b(\lambda) = 0$ versus $b(\lambda) > 0$. The slope of $\hat{b}(\lambda)$ is positive (i.e. $\hat{\Delta}_e = 12.8 > 0$), illustrating that Treatment was observed to increase effectiveness. The vertical intercept is also positive (i.e. $\hat{\Delta}_c = -1,717 < 0$), illustrating that Treatment was observed to decrease cost. This "win-win" observation is illustrated by a negative value for the estimated ICER,

given by the horizontal intercept of -134 CAD\$/QALW (see Figure 4.1). The graphs provide confidence intervals for Δ_c of -7,946 to 4,512 (the negative of the vertical intercepts in Figure 4.1) and for the ICER of -1764 to 378 (the horizontal intercepts in Figure 4.1). The confidence intervals of the ICER, provided by the horizontal intercepts, are identical to those provided by Fieller's Theorem (Chaudhary and Stearns, 1996; Heitjan, 2000; Willan and O'Brien, 1996; Willan and Lin, 2000). The lower 90% confidence limit for the incremental net benefit is positive for values of λ greater than 378 CAD\$/QALW, that is, greater than approximately 20,000 CAD\$/QALY.

From the above results, it is observed that the new treatment reduces symptoms and improves quality of life; therefore, it has the potential to reduce cost in other areas.

If there are concerns regarding the sample estimates in the presence of right skewing, as is often the case with cost data, the non-parametric bootstrap estimates provide an alternative (Briggs et al., 1997). The bootstrap method was applied to this example separately to this thesis and the estimates, using 1500 re-samples, provided a 90% confidence interval of the ICER of -1802 to 362, which is very close to the Fieller's interval (Willan and Lin, 2000).

4.3 INB Calculation when WTA>WTP for Uncensored Data – Prostate

The incremental net benefit procedure, as stated in earlier chapters, has more consistent interpretation and is amenable to routine statistical procedures. However, these procedures assume that the WTA compensation for a loss of a unit of health benefit is the same as the WTP for this same unit of health. But theoretical and empirical evidence

suggest that in health care the willingness-to-accept is about twice as the willingness-to-pay (O'Brien et al., 2001). In order to illustrate this disparity, we used the statistical procedure introduced in Chapter 3 to compare the two arms of the prostate randomized clinical trial allowing the WTP and the WTA to have different values.

For this analysis the sample means and the sample variance-covariance information as well as the difference in costs and effects provided in Table 4.1 were also used to calculate the INB of this section.

The estimate of INB for $\lambda = 400$ CAD\$/QALW and various values of γ are given in the top panel of Table 4.2. Since the estimate $\hat{\Delta}_e$ is statistically significant, little of the probability density of $\left(\hat{\Delta}_e, \hat{\Delta}_c\right)$ lies to the left of the vertical axis (see Figure 3.4). Consequently, as γ increase, only small changes are seen in the estimate of INB and the corresponding 90% Bayesian interval. However, the hypothesis INB (400) < 0 (vs INB(400) > 0) can be rejected for $\gamma = 1$ or 2 (lower limit greater than zero), but not for $\gamma = 5$ or 10 (lower limit less than zero). For illustration purposes, we also worked the sample when $\hat{\Delta}_e = 0$, and re-ran the analysis. The results are shown in the bottom panel of Table 4.1. In this case, since half the density is to the left of the vertical axis, sizable differences are seen in the estimate of the incremental net benefit and its lower limit as γ increases.

Table 4.1: Sample sizes and parameter estimates for prostate example

	n_j	$\hat{\mu}_j$	$\hat{v}_j$	$\hat{V}(\hat{\mu}_j)$	$\hat{V}(\hat{v}_j)$	$\hat{C}(\hat{\mu}_j, \hat{v}_j)$
Standard	53	28.1	29 039	16.4	7 872 681	2 876
Treatment	61	40.9	27 322	24.1	6 466 351	2 771
	114	$\hat{\Delta}_e = 12.8$	$\hat{\Delta}_c = -1717$	$\hat{V}(\hat{\Delta}_e) = 40.5$	$\hat{V}(\hat{\Delta}_c) = 14\,339\,032$	$\hat{C}(\hat{\Delta}_e, \hat{\Delta}_c) = 5\,647$
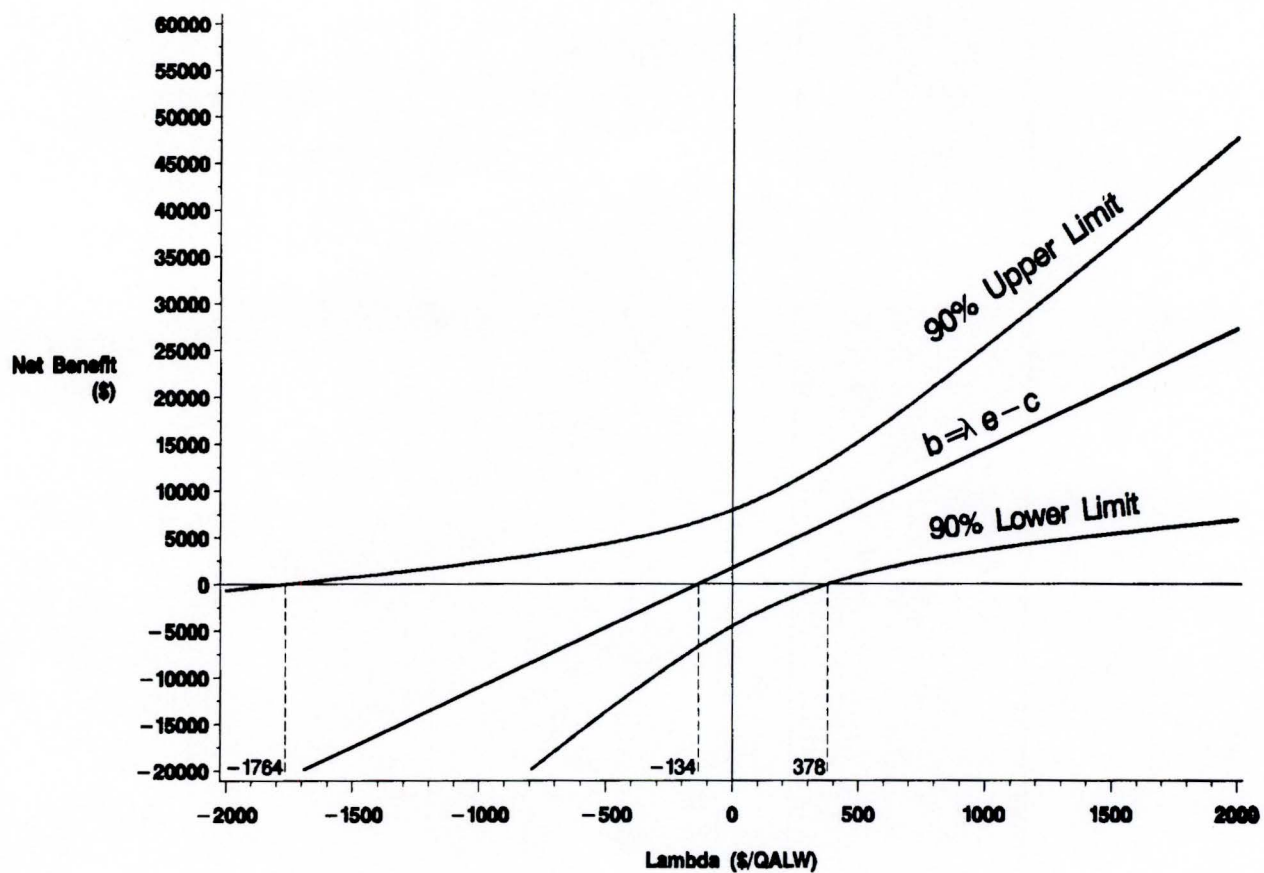
Figure 4.1: Net benefit as a function of lambda for the uncensored Prostate data

Table 4.2: Estimate of INB and Bayesian Interval for $\lambda = 400/\text{QALW}$

	γ	$\hat{\text{INB}}$	90% Bayesian Interval
$\hat{\Delta}_e = 12.8$	1	6 836	195 to 13 477
	2	6 828	116 to 13 477
	5	6 817	-92 to 13 476
	10	6 811	-281 to 13476
$\hat{\Delta}_e = 0$	1	1 715	-4 925 to 6 891
	2	915	-7 215 to 6 629
	5	-448	-18 679 to 6 409
	10	-1 515	-39 146 to 6 328

CHAPTER 5

DISCUSSION

Interest in the cost-effectiveness of health care interventions is growing rapidly. There are different methods available to conduct economic analysis for clinical trials with binary outcomes. In this work we discuss the advantages and disadvantages of the three most appealing approaches to health economic evaluators. These are, cost minimization, incremental cost-effectiveness ratio and incremental net benefit (Briggs and Fenn, 1998; Briggs and O'Brien, 2001; Polsky et al., 1997; Willan and Lin, 2000). The convention in economic evaluation studies comparing two treatments is to calculate the ICER, however, there are inherent difficulties associated with modeling the distribution of this ratio as illustrated in Chapter 2. Recent work argues in favor of using the INB methodology for analyzing cost-effectiveness data where both cost and effectiveness are measured at the patient level (Willan and Lin, 2000) so that the patient-specific units of health gain are weighted by their monetary value. The main advantage of using this technique is that INB can be expressed in units of costs thus reducing the problem to a single metric. In addition, expressing the INB in units of costs allows the investigator to examine all three approaches in a single graph (Willan and Lin, 2000), complete with their corresponding statistical inferences (see Figure 4.1). Therefore, the ICER analysis derived from the INB methodology provides a better insight where the inferential and presentational goal is to show incremental net benefit as a function of λ.

Although the INB approach has many advantages over the other methods, to use it one needs to confront first a valuation issue that is explicit to cost benefit analysis - what is society willing to pay for a unit of health such as a quality-adjusted life-year (QALY)? Secondly, one needs to consider the puzzling but consistent result that in health care WTP for a gain in health is markedly less than WTA compensation for the same reduction in health (Bhattacharya, 1982; Coursey, 1987; Kolstad and Guzman, 1998; Mueser and Dow, 1997; O'Brien et al., 2001). This creates a problem when calculating incremental net benefit of a new treatment where the joint density of (Δ_e, Δ_c) is not confined to positive values of Δ_e, because in doing this calculation we need to allow a downward "kink" in the line on the CE plane for which the incremental net benefit is constant, as shown in Figures 3.4 and 3.5.

The dominant context of this thesis was focused around the implication of the disparity between WTA and WTP in the economic analysis evaluation of comparing therapies and to provide an approach that considered this difference. By adopting a Bayesian framework, the above problem becomes one of estimating the posterior probability distribution for the incremental net benefit for a given λ. Assuming an uninformative (flat) prior, this can be achieved by splitting the integration of the bivariate joint density (Δ_e, Δ_c) into two parts (see Chapters 3 and 4), thus reflecting the observation that WTA>WTP.

Analytically, it was shown that the incremental net benefit is a declining function of γ, which is the ratio of WTA to WTP. Recent evidence provided by a meta-analysis of different economic studies using the INB method (O'Brien et al., 2001) showed that γ is

approximately two for health care studies and as high as seven for environmental studies. In the randomized controlled trial used in this thesis, the difference in effects was calculated to be 12.8 and the values for γ were set at 1, 2, 5, and 10. Looking at Table 4.2, the incremental net benefit value and its intervals do not change as much with the values 1 and 2, but the difference is seen when γ takes values of 5 and 10, especially in the lower bounds of the confidence intervals. Given the uncertainty around the values of γ and based on the results of Chapter 4 (see table 4.2), it is recommended that analysts present results over a range of values for this parameter in a sensitivity analysis. The work should include the "unkinked" value of $\gamma = 1$. But as noted above and in Table 4.2, values for $\gamma > 1$ will only have impact on the results if negative values for Δ_e have positive density.

The full advantages of the modified INB method are notwithstanding further work is required. The analysis of the data presented in this thesis was done retrospectively. Therefore, it would be interesting and useful to use this approach in a prospective economic study. Additionally, these statistical methods need to be extended to the case of censored data.

REFERENCES

Aaronson N., Ahmedzai S., Bergman B., et al. The European Organization for Research and Treatment of Cancer QLQ-C30: A quality-of-life instrument for use in international trials of oncology. *J. Natl. Cancer Inst* 1993; **85**: 365-376.

Anderson J.P., Bush J.W., Chen M., Dolenc D. Policy space areas and properties of benefit-cost/utility analysis. *Journal of the American Medical Association* 1986; **255**(6): 794-795.

Ament A., Baltussen R. The interpretation of results of economic evaluation: Explicating the value of health. *Health Economics* 1997; **6**: 625-635.

Appel L.J., Steinberg E.P., Powe N.R., et al. Risk reduction from low osmolality contrast media: What do patients think it is worth? Medical Care 1990; **28**: 324-330.

Arrow K.J. *Uncertainty and the welfare economics of medical care.* In: Arrow K.J., ed. Essays in the theory of risk-bearing. Chicago: Markham Publishing Company, 1971; 177.

Bailey M.J. Reducing risks to life. Washington DC: *American Enterprise Institute for Public Policy Research*, 1980.

Berwick D.M., Weinstein M.C. What do patients value? Willingness to pay for ultrasound in normal pregnancies. Medical Care 1985; **23**: 881-890.

Bhattacharya G. The importance of learning in planning with exhaustible resources, in *Essays in the Economics of Renewable Resources.* Mirman L. and Spulber D. (eds), Elservier, Amsterdam 1982.

Bishop R.C., Heberlin T.A. Measuring values of extra-market goods: are indirect measures biased? *American Journal of Agricultural Economics* 1979; **61**: 926-930.

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

Bloomfield D.J., Krahn M.D., Neogi T., Panzarella T., Warde P., Willan A.R., Ernst S., Moore M.J., Neville A., Tannock I.F. Economic evaluation of choemotherapy with mitoxantrone plus prednisone for symptomatic hormone-resistant prostate cancer: Based on a Canadian trial with palliative endpoints. *Journal of Clinical Oncology* 1998; **16**: 2272-2279.

Briggs A.H., Wonderling D.E., Mooney C.Z. Pulling cost-effectiveness analysis up by its bootstraps; a non-parametric approach to confidence interval estimation. *Health Economics* 1997; **6**: 327-340.

Briggs A.H., Fenn P. Confidence intervals or surfaces? Uncertainty on the cost-effectiveness plane. *Health Economics* 1998; **7**: 723-740.

Briggs A.H. A Bayesian approach to stochastic cost-effectiveness analysis. *Health Economics* 1999; **8**: 257-261.

Briggs A.H., Mooney C.Z., Wonderling D.E. Constructing confidence intervals around cost-effectiveness ratios: an evaluation of parametric and non-parametric methods using Monte Carlo simulation. *Statistics in Medicine* 1999; **18** (23): 3245-3244.

Briggs A.H., O'Brien B.J. The death of cost-minimization analysis? *Health Economics* 2001; **10**: 179-184.

Cameron T.A., James M.D. Efficient estimation methods for "closed-ended" contingent valuation surveys. *Review of Economics and Statistics* 1987; **69**: 269-276.

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

Cochran W.G. *Sampling Techniques*, Wiley, New York, 1977.

Commonwealth of Australia. *Guidelines for the Pharmaceutical Industry on Preparation of Submissions to the Pharmaceutical Benefits Advisory Committee*, p.13. Canberra: Woden (ACT) Department of Health, Housing and Community Services, 1992.

Coursey D.L., Hovis J.L., Schultze W.D. The disparity between willingness to accept and willingness to pay measures of value. *Quarterly Journal of Economics* 1987; **97**: 679-690.

Cummings R.G., Harrison G.W., Rutstrom E.E. Homegrown values, hypothetical surveys: is the dichotomous choice approach incentive-compatible? *American Economic Review* 1995; **85**(1): 260-266.

Detsky A.S. A guideline for preparation of economic analysis of pharmaceutical products: A draft document for Ontario and Canada. *Pharmaco-economics* 1993; **3**: 354-361.

Diener A., O'Brien B.J. , Gafni A. Health care contingent valuation studies: A review and classification of the literature. *Health Economics* 1998; **7**: 313-326.

Donaldson C. Willingness to pay for publicly-provided goods: A possible measure of benefit?. *Journal of Health Economics* 1990; **9**: 103-112.

Donaldson C., Hundley V., McIntosh E. Using economics alongside clinical trials: why we cannot choose the evaluation technique in advance [letter]. *Health Economics* 1996; **5**(3): 267-269.

Doubilet P., Begg C.B., Weinstein M.C., Braun P., McNeil B. Probabilistic sensitivity analysis using Monte Carlo simulation: a practical approach. *Medical Decision Making* 1985; **5**: 157-177.

Drummond M.F., Stoddart G.L., Torrance G.W. *Methods for the economic evaluation of health care programmes*. Oxford: Oxford University Press, 1987.

Drummond M.F., Davies L.M. Economic analysis alongside clinical trials: revisiting the methodological issues. *International Journal of Technology Assessing Health Care* 1991; **7**: 561-573.

Drummond M.F., O'Brien B., Stoddart G.L., et al. *Methods for the economic evaluation of health care programmes*, 2nd edition. Oxford: Oxford University Press, 1997.

Drummond M.F., O'Brien B.J., Stoddart G.L., Torrance G.W. Methods for the economic evaluation of health care programmes, Second Edition. Oxford University Press, Oxford: Oxford Medical Publications, 2000.

Eastaugh S.R. Valuations of the benefits of risk-free blood. *International Journal of Technology Assessment of Health Care* 1991; **7**: 51-65.

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

Eisenberg J.M., Schulman K.A., Glick H., Koffer H. Pharmacoeconomics: economic evaluation of pharmaceuticals . In: Strom, B. (ed.) *Pharmacoepidemiology*, 2nd edn, ch. 32. New York: Wiley, 1994.

Fieller E.C. Some problems in interval estimation with discussion. *J. Royal Stat. Soc.*, Series B 1954; **16**: 175-188.

Fox C.R., Tversky A. Ambiguity aversion and comparative ignorance. *Quarterly Journal of Economics* 1995; **110**(3): 585-603.

Gafni A. Willingness-to-pay as a measure of benefits: Relevant questions in the context of public decision making about health care programs. *Medical Care* 1991; **29** (12): 1246-1252.

Green P.E., Krieger A.M. Individualized hybrid models for conjoint analysis. *Management Science* 1996; **42**: 850-867.

Glick H. Strategies for economic assessment during the development of new drugs. *Drug Inf. J.* 1995; **29**: 1391-1403.

Hanemann W.M. Willingness to pay and willingness to accept: How much can they differ? *The American Economic Review* 1991; **18**(3): 635-647.

Hanemann M., Loomis J., Kanninen B. Statistical efficiency of double-bounded dichotomous choice contingent valuation. *American Agricultural Economics Association* 1991; **73**: 1255-1263.

Harless D.W. More laboratory evidence on the disparity between willingness to pay and compensation demanded. *Journal of Economic Behaviour and Organization* 1989; **11**: 359-379.

Heitjan D.F. Fieller's Method and net health benefits. *Health Economics* 2000; **9**: 327-335.

Hoehn J.P., Randall A. a satisfactory benefit cost indicator from contingent valuation. *Journal of Environmental Economics and Management* 1987; **14**: 226-247.

Johannesson M., Jönsson B., Borgquist L. Willingness to pay for antihypertensive therapy: Results of a Swedish pilot study: *Journal of Health Economics* 1991; **10**: 461-470.

Johannesson M., Johansson P-O, Kriström B., et al. Willingness to pay for lipid lowering: a health production function approach. *Applied Economics* 1993; **25**: 1023-1033.

Jones-Lee M.W. The value of life: *An economic analysis*. Chicago, IL: University of Chicago Press, 1976.

Jones-Lee M.W., Hammerton M., Phillips P.R. The value of safety: Results of a national sample survey. *Economic Journal* 1985; **95**: 49-72.

Kahneman D., Tversky A. Prospect theory: An analysis of decision under risk. *Econometrica* 1979; **47**(2): 263-291.

Kahneman D., Knetsch J.L., Thaler R.H. Experimental tests of the endowment effect and the coase theorem. *Journal of Political Economy* 1990; **93**(6): 1325-1348.

Kelman S. Cost-benefit analysis: An ethical critique. *Regulation* 1981; **(Jan/Feb)**: 33

Kolstad C.D., Guzman R.M. Information and the divergence between willingness-to-accept and willingness-to-pay. *Department of Economics and Bren School of Environmental Science and Management, University of California, Santa Barbara, and National Planning Office, Dominican Republic.* Revised: December 1998.

L'Abbe K.A., Detsky A.S., O'Rourke K. Meta-analysis in clinical research. Annals of Internal Medicine 1987; **107**: 224-233.

Landefield J.S., Seskin E.P. The economic value of life: linking theory to practice. *American Journal of Public Health* 1982; **72**: 555-567.

Leaf, A. Cost effectiveness as a criterion for Medicare coverage. *New England Journal of Med.* 1989; **321**: 898-900.

Lindholm L., Rosén M., Hellsten G. Are people willing to pay for a community-based preventive program. *International Journal of Technology Assessment in Health Care* 1994; **10** (2): 317-324.

Luce B.R., Elixhauser A. *Standards for socio-economic evaluations of health care products and services*, Berlin: Springer, 1990.

Magat W.A., Viscusi W.K., Huber J. Paired comparison and contingent valuation approaches to morbidity risk valuation. *Journal of Environmental Economics and Management* 1996; **15**: 395-411.

Manning W.G., Fryback D.G., Weinstein M.C. Reflecting uncertainty in cost effectiveness analysis. In Gold M.R., Siegel J.E., Russell L.B., Weinstein M.C. (eds). *Cost Effectiveness in Health and Medicine.* Oxford University Press: New York, 1996.

Manski C.F. *Analog estimation models in econometrics*. London, U.K.: Chapman and Hall, 1988.

Mitchell R.C., Carson R.T. Using surveys to value public goods: the contingent valuation method. Washington, D.C.: *Resources for the Future*, 1989.

Mooney C.Z., Duval R.D. Bootstrapping: A nonparametric approach to statistical inference. *Sage University Paper Series on Quantitative Applications in the Social Sciences*, Series No. 07-095. Newbury Park, CA: Sage, 1993.

Morrison G.C. Understanding the disparity between WTP and WTA: endowment effect, substitutability, or imprecise preferences? *Economic Letters* 1998; **59**: 189-194.

Morris C.B., Schulman K.A. Emerging role of the Food and Drug Administration in pharmacoeconomic evaluation during the drug development process. *Drug Inf. J.* 1995; **29**: 1105-1111.

Morrison G.C. Willingness to pay and willingness to accept: some evidence of an endowment effect. *Applied Economics* 1997; **29**: 411-417.

Morrison G.C. Understanding the disparity between WTP and WTA: Endowment effect, substitutability, or imprecise preferences? *Economics Letters* 1998; **59**: 189-194.

Mueser P.R., Dow J.K. Experimental evidence on the divergence between measures of willingness to pay and willingness to accept: The role of value of uncertainty. *University of Missouri, Department of Economics. MU working paper*; 1997.

Mullahy J., Manning W. Statistical issues of cost-effectiveness analysis. In Sloan F. (ed). *Valuing Health Care*. Cambridge University Press: Cambridge, 1994; 149-184.

Mullahy J. What you don't know can't hurt you? Statistical issues and standards for medical technology evaluation. *Medical Care* 1996; **34**(12 Suppl): DS124-DS135.

Neumman P.J., Johannesson M. The willingness to pay for in vitro fertilization: A pilot study using contingent valuation. Medical Care 1994; **32** (7): 686-699.

O'Brien B.J., Drummond M.F., Labelle R.J.,Willan A.R. In search of power and significance: Issues in the design and analysis of stochastic cost-effectiveness studies in health care. *Medical Care* 1994; **32**: 150-163.

O'Brien B.J., Gafni A. When do the "dollars" make sense? Toward a conceptual framework for contingent valuation studies in health care. *Medical Decision Making* 1996; **16**: 288-299.

O'Brien B.J. Willingness-to-pay in cost-benefit analysis of health care programs. In Colton T., Armitage P. (eds). *Encyclopedia of Biostatistics, Volume 6*. John Wiley & Sons: Chichester, 1998; pp 4750-4755.

O'Brien B., Goeree R., Gafni A., Torrance G.W., et al. Assessing the value of a new pharmaceutical: A feasibility study of contingent valuation in managed care. *Medical Care* 1998; **36** (370); 384- 397.

O'Brien B.J., Willan A.R., Gertsen K. Is there a kink in consumers' threshold value for cost-effectiveness in health care? *Health Economics* (submitted for publication)

Ontario Ministry of Health. *Guidelines for the Preparation of Economic Analysis to be Included in Submission to Drug Programs Branch for Listing in the Ontario Drug Benefit Formulary/Comparative Drug Index.* Toronto: Ministry of Health, 1991.

Olsen J.A., Smith R.D. Theory versus practice: A review of 'willingness-to-pay' in health and health care. *Health Economics* 2001; **10**: 39-52.

Poirier D.J. *Intermediate Statistics and Econometrics.* Cambridge, MA: MIT Press, 1995.

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

Senn S. *Statistical Issues in Drug Development.* Wiley: Chichester, 1997.

Shogren J.F., Shin S.Y., Hayes D.J., Kliebenstein J.B. Resolving differences in willingness to pay and willingness to accept. *The American Economic Review* 1994; **84**(1): 255-270.

Stinnett A.A., Mullahy J. Net health benefits: A new framework for the analysis of uncertainty in cost-effectiveness analysis. *Medical Decision Making* 1998; **18** (Suppl): S68-S80.

Tambour M., Zethraeus N., Johannesson M. A note on confidence intervals in cost-effectiveness analysis. *International Journal of Technology Assessment in Health Care* 1998; **14**(3): 467-471.

Tannock I.F., Gospodarowicz M., Meakin W., et al. Treatment of metastatic prostate cancer with low-dose prednisone: Evaluation of pain and quality of life as pragmatic indices of response. *Journal of Clinical Oncology* 1989; **7**: 590-597.

Tannock I.F., Osoba D., Stockler M.R., Ernst D.S., Neville A.J., Moore M.J., Armitage G.R., Wilson J.J., Venner P.M., Coppin C.M., Murphy K.C. Chemotherapy with mitoxantrone plus prednisone alone for symptomatic hormone-resistant prostate cancer: A Canadian trial with palliative endpoints. *Journal of Clinical Oncology* 1996; **14**: 1756-1764.

Thompson M.S., Read J.L., Liang M. Feasibility of willingness-to-pay measures in chronic arthritis. *Medical Decision Making* 1984; **4**: 195-205.

Thompson M.S. Willingness to pay and accept risks to cure chronic disease. *American Journal of Public Health* 1986; **76**: 392-403.

Tversky A., Kahneman D. Loss Aversion in riskless choice: A reference-dependent model. *Quarterly Journal of Economics* 1991; **104**(4): 1039-1062.

Van Hout B.A., Al, M.J., Gordon G.S., Rutten F.F.H. Cost, effects and C/E ratios alongside a clinical trial. *Health Economics* 1994; **3**: 309-319.

Wakker P., Klaassen M.P. Confidence intervals for cost-effectiveness ratios. *Health Economics* 1995; **4**(5): 373-382.

Weinstein M.C., Fineburg H.V. *Clinical decision analysis*. Philadelphia: W.B. Saunders, 1980.

Weinstein M.C., Sheppard D.S., Pliskin J.S. The economic value of changing mortality probabilities: A decision-theoretic approach. *The Quarterly Journal of Economics* 1980; **94**: 373-380.

Willan A.R., O'Brien B.J. Confidence intervals for cost-effectiveness ratios: An application of Fieller's theorem. *Health Economics* 1996; **5**: 297-305.

Willan A.R., Lin D.Y. Incremental net benefit in randomized clinical trials. Statistics in Medicine 2000; **20**: 1463-1574.

Willan A.R., O'Brien B.J., Leyva R.A. Cost-effectiveness analysis when the WTA is greater than the WTP. Statistics in Medicine 2001; **20**: 3251-3259.