APPROVAL SHEET

Title of Dissertation: Some Contributions to Aggregate and Individualized Cost-Effectiveness Analysis

Name of Candidate: Aryana Arsham
Doctor of Philosophy, Statistics, 2019

Dissertation and Abstract Approved: ______________________
Thomas Mathew
Professor of Statistics
Department of Mathematics and Statistics

Date Approved: ______________

ABSTRACT

Title of Dissertation: Some Contributions to Aggregate and Individualized Cost-Effectiveness Analysis

Aryana Arsham
Doctor of Philosophy, 2019

Dissertation directed by: Thomas Mathew
Professor of Statistics
Department of Mathematics & Statistics
University of Maryland, Baltimore County

Ionut Bebu
Associate Research Professor
Epidemiology & Biostatistics
The George Washington University

The field of cost-effectiveness analysis (CEA) addresses the question of whether a new treatment provides value for the money compared to a standard treatment. The relevant statistical methodologies are based on outcomes of cost and effectiveness, typically obtained from a clinical trial. These methodologies assist health policy-makers in deciding if a new treatment should be assigned to patients, and what amount of investment can provide cost-effectiveness for the new treatment. There are several commonly used metrics to quantify a new treatment's cost-effectiveness. The most widely used metric is the incremental cost-effectiveness ratio (ICER), which is the ratio of the average incremental costs to the average incremental effectiveness between two competing treatments. Another popular metric for CEA is the incremental net benefit (INB), which is the difference between the incremental effectiveness and the incremental cost, after multiplying the former with a *willingness-to-pay* parameter, which is the maximum amount a policy-maker is willing to pay for a unit of effectiveness gained under the new treatment. Yet another CEA metric is the cost-effectiveness proportion (CEP), which is the proportion of patients for whom the new treatment is less costly and more effective, up to specified margins of cost and effectiveness. Sometimes net monetary benefits (NMBs) are also compared for CEA; the NMB for a treatment is the difference between the effectiveness and the cost, after multiplying the former with the willingness-to-pay parameter.

This thesis develops statistical methodologies for CEA that enable decision making at the population and patient levels. Aggregate analysis carried out at the population level aims to provide information of a new treatment's value for an entire population of patients. Such an analyses often omits the heterogeneity amongst patients. The realization that between-patient variability can affect cost-effectiveness has underscored the need for individualized CEA, new metrics and methods are necessary for such an investigation. In

addition, there has also been interest in developing cost-effectiveness metrics when there are multiple effectiveness measures. The research reported herein deals with individualized criteria for CEA, as well as aggregate criteria when there are multiple effectiveness measures.

For individualized cost-effectiveness analysis, the CEA literature currently recommends subgroup analysis based on a stratification approach for constructing the subgroups. However, these stratification methods are somewhat arbitrary, and there is no clear way of constructing the subgroups in a well-defined fashion. In our work, we have considered a multivariate regression model for incorporating the patient-level covariates, and covariate specific CEA metrics are then defined and investigated. This appears to be a natural approach for individualized CEA, and avoids the need to construct subgroups. The individualized criteria that we have investigated include the INB, CEP and NMB. In terms of comparing the NMBs at the population-level, we have explored the stochastic comparison of the NMB distributions of the new treatment and the standard treatment.

In the presence of multiple effectiveness measures, the traditional approach, labelled *multi-criteria decision analysis*, has been to combine the different effectiveness measures into a single quantity by taking a weighted linear combination; however, the weighting is clearly subjective. Our research on this topic has focused on the CEP metric, modified to take into account the availability of multiple effectiveness measures.

This work deals with the case of only continuous cost and effectiveness random variables. For the CEP metrics investigated, the major focus is interval estimation. Both parametric and non-parametric approaches are investigated. The parametric set up assumes a joint normal distribution for the cost and effectiveness, where the normality may hold only after a monotone-transformation, notably a log-transformation for costs. In the parametric set up, a fiducial approach, the delta method, and the bootstrap are all investigated and compared for the interval estimation of the relevant CEA metrics. The non-parametric method that has been investigated is based on U-statistics. Extensive numerical results are reported in order to assess the accuracy of the confidence intervals, and the results are all illustrated with examples.

SOME CONTRIBUTIONS TO AGGREGATE AND INDIVIDUALIZED COST-EFFECTIVENESS ANALYSIS

By

Aryana Arsham

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, Baltimore County, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2019

Advisory Committee:
Dr. Thomas Mathew, Chair/Advisor
Dr. Ionut Bebu, Co-advisor
Dr. Yi Huang
Dr. DoHwan Park
Dr. Anindya Roy

Dedication

To my father, Dr. Professor Hossein Arsham, for his love, support, and wisdom.

Acknowledgements

I am most grateful to the University of Maryland Baltimore County (UMBC) for an education that has afforded me a solid foundation in modelling; first natural phenomena through a Bachelors degree in Physics, and then decision making under uncertainty by obtaining a Masters and PhD in Statistics.

Dr. Thomas Mathew, my thesis advisor, has guided me through the process of research in a most gracious manner. In times of difficulty, he has encouraged me to overcome obstacles, with ultimate patience on his part. I feel most fortunate to have had the opportunity to learn from such a great scientist, teacher, and caring individual. Through his valuable time and guidance he has shaped my mind to better understand my own research interests, and statistics as a whole. His advice has far surpassed my thesis research and has also enhanced my knowledge of how to do research and how to choose an academic career path.

Dr. Ionut Bebu, my thesis Co-advisor, has provided valuable insights and guidance on my thesis research. This work is based on joint research with Drs. Thomas Mathew and Ionut Bebu. They have introduced me to a research topic that is of continuing interest to me. I am most grateful to them for inspiring me to push myself, and for allowing me to learn from their expertise. A series of manuscripts based on this research is being completed and I am most honoured to collaborate with Drs. Thomas Mathew and Ionut Bebu on this. I also wish to express my gratitude to Dr. Andrew Willan for sharing the Canadian implantable defibrillator study data utilized in this research.

The valuable suggestions from Drs. Ionut Bebu, Yi Huang, DoHwan Park, and Anindya Roy, who are the thesis committee members and have been most helpful. I will always be grateful to the Mathematics & Statistics department of UMBC for the knowledge they have transferred to me. The dedication to research and teaching in the department is inspiring. In particular, I am grateful to the Statistics faculty: Drs. Yi Huang, Yaakov Malinovsky, Thomas Mathew, Nagaraj Neerchal, DoHwan Park, Junyong Park, Anindya Roy, Bimal Sinha, and Elizabeth Stanwyck. The department staff have been immensely helpful during my time in the department and I am thankful to: Boris Alemi, Janet Burgee, Maggie Kennedy, and Marshall Turner.

The support and care of two pivotal individuals have helped me through my PhD journey. It is my great honour to thank my father, Dr. Hossein Arsham, and my life long partner, Jason Charles Parkins, Esq.

Contents

List of Tables

List of Figures

List of Algorithms

Chapter 1

Introduction

1.1 Background

Cost-effectiveness analysis (CEA) consists of statistical methodologies that compare competing treatments for a given disease, relative to their cost and effectiveness. The field of CEA has become increasingly relevant, as healthcare policy-makers, faced with increasing health care expenditures, require evidence that treatments are cost-effective. Indeed, cost-effectiveness analysis has been developed to assist in the decision making process for resource allocation under budgetary constraints. Various criteria are utilized to quantify cost-effectiveness; these include the incremental cost-effectiveness ratio (ICER), incremental net benefit (INB), individual net monetary benefit (NMB), and cost-effectiveness proportion (CEP). These will be defined in the next section. Cost is typically measured in monetary terms, and the effectiveness measure is disease specific. Some examples of effectiveness measures are survival time, number of lives saved, response or no response (in the case of binary effectiveness), and biomarkers of disease. A generic and popular measure of effectiveness is Quality-Adjusted Life Years (QALYs), which is a function of various indicators of the quality of life and length of life, combined into a single index. Other similar measures of effectiveness include Disability-Adjusted Life Years (DALYs), Quality-Adjusted Life Weeks

(QALWs), Health-Adjusted Life Years (HALYs), etc. Here we shall not go into the details on the computation of these. In addition, this thesis deals with only continuous effectiveness measures.

In the next section, we shall define the various criteria mentioned above for the assessment of cost-effectiveness, as they are defined in current literature. Throughout the thesis, we will be considering the comparison of only two treatments. It should be noted that the criteria ICER, INB, NMB, and CEP are population level criteria. Recent literature on CEA has brought up the issue of individualizing the criteria, i.e., defining the criteria from the perspective of the individual patient. Such an individualized analysis is a major topic taken up in this work. We address statistical inference problems associated with various population level and individual level criteria for CEA. We want to point out that the models and problems taken up here are for data obtained from clinical trials only, and not from observational studies. We shall now provide a summary of the criteria and models commonly used for CEA.

1.2 Criteria and models

In this section, we shall first define the various cost-effectiveness criteria mentioned in the previous section, keeping in mind that we are comparing just two treatments: Treatment 1 (a standard treatment), and Treatment 2 (an existing treatment). We shall then mention some of the statistical inference for CEA, available in the literature.

1.2.1 Criteria for CEA

Let (C_i, E_i) be a bivariate random variable denoting the cost and effectiveness for patients on the i^{th} treatment, $i = 1, 2$. Let (μ_{Ci}, μ_{Ei}) denote the mean of (C_i, E_i). The incremental cost-effectiveness ratio (ICER) is defined as the ratio of the average incremental cost (denoted by Δ_C) to the average incremental effectiveness (denoted by Δ_E). That is,

$$ICER = \frac{\Delta_C}{\Delta_E} = \frac{\mu_{C1} - \mu_{C2}}{\mu_{E1} - \mu_{E2}}. \tag{1.2.1}$$

We shall assume that the denominator of the ICER is positive; that is, the new treatment (Treatment 1) is more effective than the standard treatment (Treatment 2), on average. Thus, Treatment 1 is cost-effective compared to Treatment 2 if the ICER is not too large, i.e., it is below a threshold. In view of this, cost-effectiveness can be assessed by computing an upper confidence limit for the ICER, and verifying if the upper confidence limit is below the specified threshold. The threshold is the *willingness-to-pay* (WTP) parameter, denoted by the constant λ. The WTP, λ, is the maximum acceptable ICER, i.e. the maximum acceptable amount to pay per unit of health gained. The health care policy-maker typically decides the value of WTP. It should be noted that being a ratio metric, ICER presents challenges when it comes to statistical inference.

Noting the difficulties associated with the ICER, another metric has been introduced in the literature, referred to as the incremental net benefit (INB). The INB is defined as the difference of the incremental effectiveness (multiplied by the WTP, λ) and the incremental cost. That is,

$$INB = \lambda\Delta_E - \Delta_C = \lambda(\mu_{E1} - \mu_{E2}) - (\mu_{C1} - \mu_{C2}). \tag{1.2.2}$$

We note that the INB is linear in $(\mu_{E1} - \mu_{E2})$ and $(\mu_{C1} - \mu_{C2})$. It is clear that Treatment 1 is cost-effective compared to Treatment 2 if $INB > 0$. Thus, the cost-effectiveness can be assessed by computing a lower confidence limit for the INB and verifying if the lower confidence limit is positive. We refer to Willan and Briggs (2006) for a book-length treatment of the statistical methodologies used in CEA. In particular, statistical inference procedures concerning the ICER and INB are discussed in detail in the book.

A probabilistic metric for the assessment of cost-effectiveness has recently been introduced by Bebu, Mathew and Lachin (2016). The motivation for the criterion is as follows.

For two subjects, one from each treatment group, the first treatment is cost-effective compared to the second treatment if

$$C_1 \leq C_2 \ \& \ E_1 \geq E_2.$$

In view of this, the authors defined a cost-effectiveness proportion (CEP) as the probability that Treatment 1 is cost-effective compared to Treatment 2. That is,

$$CEP = P\left(C_1 \leq C_2, \quad E_1 \geq E_2\right). \tag{1.2.3}$$

Clearly, large values of the CEP are desirable. It turns out that CEP and its variants offer considerable flexibility in the assessment of cost-effectiveness. For example, if the first treatment is anticipated to be more effective, but also more costly, CEP can be modified as

$$CEP(\delta_C) = P\left(C_1 \leq C_2 + \delta_C, \quad E_1 \geq E_2\right), \tag{1.2.4}$$

for a specified threshold δ_C for the cost. Cost-effectiveness can be concluded if $CEP(\delta_C)$ is large. The quantity δ_C can be thought of as a willingness-to-pay parameter, i.e. the total additional cost one is willing to pay for the new treatment over the standard treatment. Conversely, one can also estimate the threshold δ_C after specifying a value for $CEP(\delta_C)$. For example, suppose a health care policy-maker wants a more effective treatment to be cost-effective for 80% of the patients. What willingness-to-pay quantity δ_C is necessary to meet this requirement? Note that δ_C is now an unknown parameter. Thus, we want to do statistical inference on δ_C satisfying $CEP(\delta_C) = 0.80$. If there is reason to believe that the value of δ_C satisfying such a requirement is too high, the cost-effectiveness of Treatment 1 becomes doubtful under the above 80% requirement.

A further modification of the CEP is

$$CEP(\delta_C, \delta_E) = P\left(C_1 \leq C_2 + \delta_C, \quad E_1 \geq E_2 + \delta_E\right), \tag{1.2.5}$$

which includes another threshold δ_E for the effectiveness. The motivation for including such a threshold is that it may not be enough to have improved effectiveness; a minimum *clinically meaningful* improved effectiveness is required, specified in terms of δ_E. Another possibility is to consider conditional probabilities such as $P(C_1 \leq C_2 + \delta_C \,|\, E_1 \geq E_2)$, which can be used for investigating how likely is it that the first treatment is less costly (up to a margin δ_C) for a subject for whom the new treatment is more effective. An advantage of the CEP metric is that it is invariant under monotone transformations of cost and effectiveness outcomes.

The criteria mentioned above do not take into account any covariates that could affect the cost and effectiveness. A natural way to incorporate covariates is to model the bivariate (cost, effectiveness) data using a linear regression model; see Willan, Briggs and Hoch (2004). The criteria defined above can be extended to a regression context; however, they have to be defined at a specified covariate value. An alternative approach that has been tried in the literature consists of defining an individual level net monetary benefit, referred to as the net monetary benefit, (NMB), and model it as a function of the covariates. For the j^{th} patient in the i^{th} group, let (C_{ij}, E_{ij}) denote the cost and effectiveness, the net monetary benefit, say NMB_{ij}, is defined as

$$NMB_{ij} = \lambda E_{ij} - C_{ij}, \tag{1.2.6}$$

$i = 1,\ 2$, where λ is the willingness-to-pay parameter defined earlier. Note that the INB defined in (1.2.2) is simply the mean of $NMB_{1j} - NMB_{2k}$. In addition to inference concerning such a mean, it appears reasonable to appropriately compare the distributions of NMB_{1j} and NMB_{2k} for assessing cost-effectiveness. As already noted, an approach taken by several authors in order to account for covariates consists of modelling NMB_{ij} as a function of the covariates, and then consider the mean of $NMB_{1j} - NMB_{2k}$ at a specified covariate value; see Hoch, Briggs and Willan (2002), Hoch and Dewa (2014), and Hounton and Newlands (2012). However, we are of the opinion that when covariates influence the cost and effectiveness of individual patients, they should be incorporated into the analysis via a bivariate regression model for the (cost, effectiveness) random variable, and then the various criteria should be extended to such a regression scenario.

1.2.2 The case of multiple effectiveness criteria

Cost-effectiveness studies sometimes have multiple measures of effectiveness when assessing the cost-effectiveness of competing treatments. Thus far, cost-effectiveness analysis literature has focused on combining the multiple measures of effectiveness into a single scalar quantity. However, when combining multiple measures of effectiveness into a single quantity, important information may be lost. Moreover the method of combining such measures is subjective. The emerging field of *multi-criteria decision analysis* (MCDA) focuses on evaluating treatments based on multiple criteria (multiple effectiveness measures, for example). As already noted, the recommendation in the literature is to combine the different effectiveness measures into a single quantity by forming a weighted combination; see Thokala and Duenas (2012) and Thokala et al. (2016). The weights can also be constructed so as to reflect patient preferences; see Broekhuizen et al. (2017) . Alternatively, one can think of performing cost-effectiveness analysis by appropriately modelling and analysing the multivariate data resulting from multiple effectiveness measures. This is especially desirable if there is no clearly defined approach for choosing the weights for combining the different effectiveness measures. In our work, we shall take up such a multivariate modelling for CEA.

1.2.3 A parametric model for CEA

We shall now specify a commonly used parametric model for CEA, when the cost and effectiveness are both continuous random variables. We shall also give the expressions for the ICER and INB under these models. Let C_i and E_i denote the cost and effectiveness random variables for the patients in the i^{th} group ($i = 1, 2$). It has been noted that cost data often exhibit skewness, and the log-normal distribution is very often appropriate. In other words, $\ln(C_i)$ follows a normal distribution. If the effectiveness measure also follows a normal distribution, then we have the following bivariate model for (cost, effectiveness), in the absence of covariates:

$$\begin{pmatrix} ln(C_i) \\ E_i \end{pmatrix} \sim N_2 \left(\mu_i = \begin{pmatrix} \mu_{Ci} \\ \mu_{Ei} \end{pmatrix}, \Sigma_i = \begin{pmatrix} \sigma_{iCC} & \sigma_{iCE} \\ \sigma_{iCE} & \sigma_{iEE} \end{pmatrix} \right) \tag{1.2.7}$$

$i = 1,\ 2$, where the mean vectors and covariance matrices are unknown parameters. This model is referred to as the lognormal-normal model. We note that μ_{Ci} now refers to the mean of the log-transformed cost; earlier (in (1.2.1) and (1.2.2), for example) we used the same notation for the mean cost before making any transformation. We hope this does not cause any confusion.

In view of the lognormality of the C_i, the mean of C_i is given by

$$E(C_i) = \exp\left(\mu_{Ci} + \frac{1}{2}\sigma_{iCC} \right),$$

where σ_{iCC} is the first diagonal element of the covariance matrix $\mathbf{\Sigma}_i$, $i = 1,\ 2$. Under the lognormal-normal model, we then have

$$ICER = \frac{\exp\left\{ \mu_{C1} + \frac{\sigma_{1CC}}{2} \right\} - \exp\left\{ \mu_{C2} + \frac{\sigma_{2CC}}{2} \right\}}{\mu_{E1} - \mu_{E2}} \tag{1.2.8}$$

and

$$INB = \lambda \left[\mu_{E1} - \mu_{E2}\right] - \left[\exp\left\{\mu_{C1} + \frac{\sigma_{1CC}}{2}\right\} - \exp\left\{\mu_{C2} + \frac{\sigma_{2CC}}{2}\right\}\right], \tag{1.2.9}$$

where λ is the willingness-to-pay parameter.

Now suppose the effectiveness also follows a lognormal distribution, so that we have the lognormal-lognormal model:

$$\begin{pmatrix} ln(C_i) \\ ln(E_i) \end{pmatrix} \sim N_2 \left(\mu_i = \begin{pmatrix} \mu_{Ci} \\ \mu_{Ei} \end{pmatrix}, \Sigma_i = \begin{pmatrix} \sigma_{iCC} & \sigma_{iCE} \\ \sigma_{iCE} & \sigma_{iEE} \end{pmatrix} \right) \tag{1.2.10}$$

Under the model (1.2.10), the ICER and INB are given by

$$ICER = \frac{\exp\left\{\mu_{C1} + \frac{\sigma_{1CC}}{2}\right\} - \exp\left\{\mu_{C2} + \frac{\sigma_{2CC}}{2}\right\}}{\exp\left\{\mu_{E1} + \frac{\sigma_{1EE}}{2}\right\} - \exp\left\{\mu_{E2} + \frac{\sigma_{2EE}}{2}\right\}} \tag{1.2.11}$$

and

$$INB = \lambda \left[\exp\left\{\mu_{E1} + \frac{\sigma_{1EE}}{2}\right\} - \exp\left\{\mu_{E2} + \frac{\sigma_{2EE}}{2}\right\}\right] - \left[\exp\left\{\mu_{C1} + \frac{\sigma_{1CC}}{2}\right\} - \exp\left\{\mu_{C2} + \frac{\sigma_{2CC}}{2}\right\}\right], \tag{1.2.12}$$

When covariates are present, a natural approach consists of modelling the μ_{Ci} and μ_{Ei} as linear functions of the covariates, i.e., we have a bivariate linear regression model. The ICER, INB, CEP, and NMB can now be defined at a specified covariate value. We shall take this up later in the thesis.

1.3 Statistical inference

The criteria ICER, INB, and CEP, have intuitive appeal and the calculation of point estimates is straightforward. However, the computation of interval estimation can be challenging. Being a ratio parameter, this is especially the case with the ICER. Various approaches are available in the literature in order to construct confidence intervals for ICER and INB, and these include parametric and non-parametric methods. The parametric methods include delta-method based asymptotic procedures, parametric bootstrap methods, and an application of Fieller's Theorem (for the ICER); the assumed parametric model is the lognormal-normal model.

The non-parametric method that has been utilized is the non-parametric bootstrap. The implementation of the above parametric and non-parametric approaches is quite straightforward, and we refer to the book by Willan and Briggs (2006) for more details. We would also like to point out the recent work of Bebu et al. (2016a) where confidence intervals are constructed using the fiducial approach (or the generalized pivotal quantity approach) for both the ICER and the INB (the fiducial approach is briefly explained shortly). The authors note that the fiducial approach is very satisfactory in terms of maintaining the coverage probability. The examples discussed in Bebu et al. (2016a) also show that different approaches for the interval estimation of ICER and INB can yield drastically different solutions when applied to real data; this is especially the case for the ICER. We would also like to point out that for the CEP criterion, introduced in Bebu, Mathew and Lachin (2016), the authors have once again derived satisfactory confidence limits following the fiducial approach. A non-parametric solution, based on U-statistics, is also proposed in Bebu, Mathew and Lachin (2016) for the CEP criterion presented in (1.2.3).

It appears that the NMB criterion has not received as much attention in the literature. However, as already noted, a regression model for the NMB criterion is formulated in Hoch, Briggs and Willan (2002) and Hoch and Dewa (2014). We want to once again emphasize that the authors have not investigated the NMB criterion under a regression model for the data on cost and effectiveness. Rather, the regression models used in these works model the NMB random variable directly as a function of covariates.

1.3.1 Fiducial inference

Fiducial inference for a parameter is based on the percentiles of a *fiducial quantity* (also referred to as a *generalized pivotal quantity*; see Weerahandi (1993)). The description that follows is not in the most general framework; rather, we shall introduce the fiducial methodology as it applies to our current research. We refer to Hannig (2009) for a very general discussion of fiducial inference, including theoretical developments and numerous applications. We also refer to Weerahandi (1993), where the concept was introduced, and applications were explored, under the terminologies *generalized pivotal quantity* and *generalized inference*.

To define a fiducial quantity, let $F_X(x, \theta, \delta)$ denote the CDF of a random variable X, depending on a parameter of interest θ, and a nuisance parameter δ (where δ could be a vector). Let $\boldsymbol{X}$ be a random sample from the distribution $F_X(x, \theta, \delta)$, and let $\boldsymbol{x}$ denote the observed value of $\boldsymbol{X}$. A fiducial quantity for θ is a function of $\boldsymbol{X}$, $\boldsymbol{x}$, θ and δ, and will be denoted as $\tilde{\theta} = G(\boldsymbol{X}, \boldsymbol{x}, \theta)$, where we have suppressed the possible dependence on δ. The quantity $G(\boldsymbol{X}, \boldsymbol{x}, \theta)$ is required to satisfy two conditions: (i) given the observed data $\boldsymbol{x}$, the distribution of $G(\boldsymbol{X}, \boldsymbol{x}, \theta)$ is free of any unknown parameters, and (ii) when $\boldsymbol{X}$ is replaced with $\boldsymbol{x}$, $G(\boldsymbol{X}, \boldsymbol{x}, \theta)$ simplifies to θ, the parameter of interest; i.e., $G(\boldsymbol{x}, \boldsymbol{x}, \theta) = \theta$. Under these two conditions, the percentiles of the fiducial quantity $\tilde{\theta} = G(\boldsymbol{X}, \boldsymbol{x}, \theta)$ can be computed and used as confidence limits for θ, when θ is a scalar. Such confidence intervals are referred to as fiducial intervals.

In a series of papers, Hannig and co-authors have rigorously investigated the asymptotic performance of fiducial intervals. They have shown that fiducial intervals maintain the coverage probability asymptotically. However, in practical applications where fiducial inference has been applied, the small sample performance of fiducial intervals must be investigated numerically. In many situations fiducial intervals do provide satisfactory coverage probabilities in small sample size scenarios. We refer to the article by Hannig et al. (2016) for a recent review along with several applications.

1.3.2 Percentile bootstrap

The percentile bootstrap is a simple numerical approach that can be used for computing confidence limits for a parameter θ. Based on a random sample, let $\hat{\theta}$ denote a point estimate of θ obtained in a parametric or non-parametric set up. Based on B bootstrap samples (generated parametrically or non-parametrically), let $\hat{\theta}_b^*$ denote the point estimate of θ obtained from the b^{th} bootstrap sample, $b = 1,\ 2,\ ....,\ B$. The percentile bootstrap method consists of using appropriate percentiles of $\hat{\theta}_b^*$ as confidence limits for θ.

1.3.3 Delta method

In some of the interval estimation problems that we have addressed, we shall also consider large sample confidence limits derived using the asymptotic normality of appropriate statistics, where the latter is obtained by applying the delta method.

1.3.4 Tolerance intervals and tolerance limits

A tolerance interval is an interval computed using a random sample, intended to capture a specified proportion or more of a population, with a given confidence level. The specified proportion is referred to as the *content* of the tolerance interval. We shall denote the content by p, and the confidence level by $1-\alpha$. To provide a formal definition, let X denote a scalar random variable, and suppose tolerance limits are required for the distribution of X. Let $\boldsymbol{X}$ denote a random sample from the distribution of X. A lower tolerance limit, say $L(\boldsymbol{X})$, satisfies the following condition, stated in terms of the content p and the confidence level by $1-\alpha$:

$$P_{\boldsymbol{X}}[P_X(X \geq L(\boldsymbol{X})|X) \geq p] = P_{\boldsymbol{X}}[L(\boldsymbol{X}) \leq q_{1-p}] = 1-\alpha, \tag{1.3.1}$$

where q_{1-p} is the $(1-p)^{th}$ quantile of X. Thus $(L(\boldsymbol{X}), \infty)$ is a one-sided tolerance interval having lower limit $L(\boldsymbol{X})$, and $L(\boldsymbol{X})$ is also a $100(1-\alpha)\%$ lower confidence limit for q_{1-p}. An upper tolerance limit, and a two-sided tolerance interval can be similarly defined, and an upper tolerance limit with content p and confidence level $1-\alpha$ is also a $100(1-\alpha)\%$ upper confidence limit for the p^{th} quantile of X. However, a two-sided tolerance interval does not reduce to a confidence interval for any parameter. A book-length discussion of tolerance limits and regions is found in Krishnamoorthy and Mathew (2009).

1.3.5 Stochastic dominance

Let C_i and E_i denote the cost and effectiveness random variables for the patients in the i^{th} group $(i = 1, 2)$. Then $NMB_i = \lambda E_i - C_i$ is the random variable representing the net monetary benefit for the i^{th} treatment. By definition, NMB_1 is stochastically larger than NMB_2 if $P(NMB_1 \geq t) \geq P(NMB_2 \geq t)$ for all t. If so, the first treatment can be deemed cost-effective. This is clearly a strong requirement compared to a comparison of the means, which results in the INB criterion. However, in practice such a stochastic dominance condition may not hold for all values of t. Furthermore, not all values of t will have practical relevance. Consequently, it may be of interest to test if the stochastic dominance condition holds for all values of t belonging to an interval that has practical relevance. This will be addressed later in the thesis.

1.3.6 U-statistics

The theory of U-statistics provides a very versatile non-parametric approach for inference in many practical problems, including CEA. In fact the work of Bebu, Mathew and Lachin (2016) develops non-parametric inference for the CEP based on U-statistics. The discussion given below is very brief, providing the definition and some basic results that will be used in our application. We refer to the book by Kowalsk and Tu (2008) for a detailed treatment of the topic; see also Chapter 6 of Lehmann (1999).

We start with the definition of a U-statistic in the context of a single sample consisting of the iid observations X_1, X_2, ..., X_n. A U-statistic with kernel h and order k is defined as

$$U_1 = \frac{1}{\binom{n}{k}} \sum \cdots \sum_{i_1,\dots,i_k} h(X_{i_1}, X_{i_2}, ..., X_{i_k}),$$

where the function h is symmetric in its arguments, $(i_1, i_2, ...i_k)$ is a subset of $(1, 2, ..., n)$, and the summation is over all such subsets. If $\theta = E\left[h(X_1, X_2, ..., X_k)\right]$, then $E(U_1) = \theta$. In order to write the variance of U_1, let

$$h_i(X_1, X_2, ..., X_i) = E\left[h(X_1, X_2, ..., X_k)|X_1, X_2, ..., X_i\right]$$
$$\text{and} \;\; \sigma_i^2 = V\left[h_i(X_1, X_2, ..., X_i)\right], i = 1, 2, ..., k$$

It can be shown that

$$\sigma_i^2 = \text{Cov}\left[h(X_1, X_2, ..., X_i, X_{i+1}, ..., X_k), h(X_1, X_2, ..., X_i, X'_{i+1}, ..., X'_k)\right],$$

where the X'_{i+1}, ..., X'_k are independent, and independent of the X_i's, having the same distribution as that of the X_i's. Then it is known that

$$V(U_1) = \frac{1}{\binom{n}{k}} \sum_{i=1}^{k} \binom{k}{i}\binom{n-k}{k-i}\sigma_i^2.$$

Also, $V(U_1) = k^2\sigma_1^2/n$, asymptotically (provided all the σ_i^2 are finite). Furthermore, the asymptotic distribution of $\sqrt{n}\,(U_1 - \theta)$ is $N(0, k^2\sigma_1^2)$.

The definition of U_1 has been extended to a variety of practical scenarios; a two-sample formulation is as follows. Consider two samples consisting of the iid observations X_{11}, X_{12}, ..., X_{1n_1} of size n_1, and X_{21}, X_{22}, ..., X_{2n_2} of size n_2. Let $(i_1, i_2, ...i_{k_1})$ be a subset of $(1, 2, ..., n_1)$, and $(j_1, j_2, ...j_{k_2})$ be a subset of $(1, 2, ..., n_2)$. A two-sample U-statistics with kernel h can now be defined as

$$U_{12} = \frac{1}{\binom{n_1}{k_1}} \frac{1}{\binom{n_2}{k_2}} \sum \sum h(X_{1i_1}, X_{1i_2}, ..., X_{1i_{k_1}}; X_{2j_1}, X_{2j_2}, ..., X_{2j_{k_2}}),$$

where the kernel h is assumed to be symmetric with respect to the arguments from each sample, and the summations in the definition are over all possible subsets of sizes k_1 and k_2, respectively, from $(1, 2, ..., n_1)$ and $(1, 2, ..., n_2)$. If $\theta = E\left[h(X_{11}, X_{12}, ..., X_{1k_1}; X_{21}, X_{22}, ..., X_{2k_2})\right]$, then $E(U_{12}) = \theta$, and

$$Var(U_{12}) = \sum_{i=0}^{k_1} \sum_{j=0}^{k_2} \frac{\binom{k_1}{i}\binom{n_1-k_1}{k_1-i}}{\binom{n_1}{k_1}} \frac{\binom{k_2}{j}\binom{n_2-k_2}{k_2-j}}{\binom{n_2}{k_2}} \sigma_{ij}^2,$$

with

$$\begin{aligned} \sigma_{ij}^2 = Cov\big[&h(X_{11}, \ldots, X_{1i}, X_{1i+1}, \ldots, X_{1k_1}; X_{21}, \ldots, X_{2j}, X_{2j+1}, \ldots, X_{2k_2}), \\ &h(X_{11}, \ldots, X_{1i}, X_{1i+1}^{'}, \ldots, X_{1k_1}^{'}; X_{21}, \ldots, X_{2j}, X_{2j+1}^{'}, \ldots, X_{2k_2}^{'})\big], \end{aligned}$$

where the $X_{1i+1}^{'}, \ldots, X_{1k_1}^{'}$ are independent and identically distributed as the X_{1i}'s; and $X_{2j+1}^{'}, \ldots, X_{2k_2}^{'}$ are independent and identically distributed as the X_{2j}'s.

If $\frac{n_1}{n_1+n_2} \to \rho$ as $n_1, n_2 \to \infty$, then $\sqrt{N}(U_{12} - \theta)$ is asymptotically distributed as $N(0, \sigma^2)$, where $\sigma^2 = \frac{k_1^2}{\rho}\sigma_{10}^2 + \frac{k_2^2}{1-\rho}\sigma_{01}^2$ and $N = n_1 + n_2$.

1.4 Summary of the thesis

The focus of the thesis is an investigation of the INB, CEP and NMB criteria for aggregate level (i.e., population level) and patient level (i.e., individualized) CEA. The ICER metric will not be considered in our work, given the acknowledged difficulties associated with this criterion.

Chapter 2 is on individualized CEA to facilitate patient level comparison of treatments with respect to cost-effectiveness. A bivariate regression model is proposed to incorporate covariates, and the INB and CEP criteria are then defined at a specified covariate value. In the parametric set up, i.e., lognormal-normal or lognormal-lognormal models, confidence intervals are constructed for these individualized cost-effectiveness metrics using the fiducial method. The proposed fiducial inference is straightforward to implement and provides accurate confidence limits. Another problem addressed in this chapter is the estimation of the thresholds δ_C and δ_E in (1.2.5) so that $CEP(\delta_C, \delta_E)$ defined in (1.2.5) assumes a specified value. A Newton-Raphson algorithm is used to solve for the thresholds required to obtain a desired value of $CEP(\delta_C, \delta_E)$ for a particular patient. The methodologies are illustrated using an application where the problem of interest is to compare two treatments for schizophrenia. The results indicate that cost-effectiveness for the new treatment varies considerably among patients. Furthermore, estimated values of the thresholds δ_C and δ_E show that differing amounts of investment are required for the new treatment to be cost-effective for various patients, a result that is useful for policy-makers with limited budgets.

The topic of multi-criteria decision analysis (MCDA) for CEA has focused on methods to analyze and make inferences about the cost-effectiveness of treatments when there are multiple effectiveness measures. The main focus of the MCDA literature thus far has been on various weighting schemes used to consolidate multiple effectiveness measures into a single scalar quantity. When there are multiple stake holders in the resource allocation decision process, the differing preferences amongst the decision makers could result in different weighting combinations leading to contradictory conclusions regarding treatment recommendations. This subjectivity has been actively debated in the associated MCDA literature. In order to circumvent this subjectivity, we develop two probabilistic criteria in Chapter 3, which are adaptations of the usual CEP, and are free of weights. One criterion prioritizes value for the money and the other prioritizes effectiveness of the treatment. In the parametric set up, we apply the fiducial approach and percentile bootstrap methods to construct confidence limits for the criteria. In addition, a non-parametric approach is implemented using U-statistics. The methods are applied to data from a clinical trial on irritable bowel syndrome (IBS). The results indicate that the parametric and non-parametric methods are accurate for moderate and large sample sizes.

Chapter 4 and Chapter 5 explore the comparison of the distributions of the net monetary benefit (NMB) random variables for the two treatment groups. We recall that the NMB for a treatment group is the difference between the random variables representing the effectiveness and cost for that group, after multiplying the effectiveness with a willingness-to-pay parameter. The problem addressed in Chapter 4 is under a parametric set up assuming a bivariate regression model, and we discuss the comparison of specific percentiles of the NMB random variables for the two treatment groups at specified covariate values. We have essentially compared lower tolerance limits of the two NMB distributions. A larger lower tolerance limit for a specific patient (i.e., at a specified set of covariates) under the new treatment indicates that the treatment is cost-effective. The results are applied to the schizophrenia example mentioned earlier. The specific problem addressed in Chapter 5 is the stochastic comparison of the NMB distribution of the new treatment with that of the standard treatment. Under a parametric set up, we have developed procedures to assess the stochastic dominance of the NMB for the new treatment compared to the standard treatment within an interval of values deemed relevant. We have accomplished this using the fiducial and percentile bootstrap methods for the interval estimation of an appropriate parameter. The methods are then applied to a data set from a study on malnutrition.

The results obtained in Chapters 2−5 assume that we have data obtained from a single randomized controlled trial, which is a typical scenario in CEA. However, there are situations where cost-effectiveness data comes from multi-center trials, and this calls for the development of cost-effectiveness methods based on such data. Under multi-center trials between-center heterogeneity is to be expected, and must be accounted for in the data analysis. Modelling the cost and effectiveness outcomes from such studies using a multilevel model is one possible approach to account for such heterogeneity. We develop a multivariate multilevel model for cost and effectiveness outcomes of patients in multi-center trials. Data analysis is then carried out under the model in order to assess the cost-effectiveness of a new treatment for a specified patient using the patient-specific INB. Model parameters are estimated using the restricted iterative generalized least squares (RIGLS). Confidence intervals for the patient-specific INB are computed using the delta method. The analysis is applied to the Canadian implantable defibrillator study. Our investigation in this chapter is under a parametric set up.

We conclude this introductory chapter with the observation that even though we have investigated different criteria for cost-effectiveness analysis, we are not in a position to recommend one criterion over the other. This will require input from health economists and other policy-makers. Consequently, we have refrained from advocating any particular criterion. Our goal has been the development of accurate inference only, for the various CEA metrics: aggregate as well as individualized.

Chapter 2

Individualized cost-effectiveness analysis

The need for an individualized analysis of cost-effectiveness data has recently been emphasized in the health economics literature. A major motivation for this emphasis is the observation that a treatment that is cost-effective for the whole population may not be so for a subgroup. Conversely, a treatment that is not cost-effective at the population-level may be cost-effective for a sub-population. In other words, the choice of treatment that maximizes the population's cost-effectiveness is not necessarily the best choice for an individual. Cost-effectiveness analysis metrics traditionally used in practice have been at the population/aggregate level. Clearly, population-level metrics cannot account for important inter-individual differences that affect the values of such metrics for a particular treatment intervention (Espinoza et al. (2018)). Thus, the literature on cost-effectiveness analysis has been emphasizing the need to have metrics relevant for individualized decision making (Mihaylova et al. (2011)). Several strong reasons for individualized cost-effectiveness analysis are noted in Ioannidis and Garber (2011). In spite of these observations, a well accepted framework and criteria for individualized cost-effectiveness analysis are still lacking. This is the main motivation for the work in this chapter.

In the literature, one suggestion for individualizing CEA is to form subgroups of patients, and then apply the traditional criteria to the subgroups. However, an obvious drawback of this approach is that the construction of the subgroups could be somewhat arbitrary. Secondly, such methods ignore information shared among patients belonging to different stratum. Clearly, individualization of CEA amounts to the incorporation of covariates into the analysis. An obvious way to do this is to use a regression model for the cost and effectiveness data. The present work proposes such a model for performing individualized CEA. In the case of skewed data where log-normality is often appropriate, we shall propose the regression model for the log-transformed data. Thus we will have a bivariate normal linear regression model for the cost and effectiveness data from each patient. For any specified covariate value, CEA metrics such as the incremental cost-effectiveness ratio (ICER), incremental net benefit (INB), and cost-effectiveness proportion (CEP) can now be defined. Thus what we propose to do is a covariate specific CEA, as opposed to a subgroup specific CEA. We have focused on the INB and CEP criteria, and interval estimation is addressed using the fiducial approach. As noted in the introductory chapter, the ICER criterion is not taken up here due to the acknowledged difficulties and drawbacks associated with the ICER.

In Bebu et al. (2016b) the authors successfully applied the fiducial methodology for the construction of confidence limits for the INB and ICER, and later extended it to the CEP parameter in Bebu, Mathew and Lachin (2016). Both of these works employed fiducial methods under bivariate models that did not include any covariates. Two problems are addressed in this chapter: (i) an extension of the methods proposed in the above two papers for individualized cost-effectiveness analysis, by incorporating covariates into the model, and (ii) the determination of meaningful cost and effectiveness thresholds required to obtain a desired CEP; these are the quantities δ_C and δ_E appearing in the definition of $CEP(\delta_C, \delta_E)$ given in (1.2.5). We recall that the threshold δ_C for the cost can be viewed as the additional cost one is willing to pay for the new treatment over the standard treatment. The threshold δ_E for the effectiveness refers to the minimum increase in effectiveness required for the new treatment to be deemed effective. Clearly, large values of $CEP(\delta_C, \delta_E)$ indicate that the proportion of patients for which the treatment is cost-effective is large. We aim to answer the question of what thresholds are required to obtain a desired CEP. The solution to this inverse problem aids health care policy-makers in determining which combination of thresholds are necessary for cost-effectiveness to hold, as specified by a value of $CEP(\delta_C, \delta_E)$.

2.1 Model for cost and effectiveness

In this section, we shall explore CEA under regression models for the (cost, effectiveness) data, to accommodate individual level covariates. Let us start with the model.

Let C_{ij} and E_{ij}, respectively, denote the cost and effectiveness for the j^{th} patient assigned to the i^{th} treatment intervention ($i = 1, 2; j = 1, \ldots, n_i$). Also let $\mathbf{w}_{ij}$ denote a $p \times 1$ vector of covariates associated with the j^{th} individual belonging to the i^{th} treatment intervention. We shall assume the following lognormal-lognormal model:

$$X_{ij} = \begin{pmatrix} \ln[C_{ij}] \\ \ln[E_{ij}] \end{pmatrix} \sim N_2 \left(B_i \mathbf{w}_{ij} = \begin{pmatrix} B'_{iC} \\ B'_{iE} \end{pmatrix} \mathbf{w}_{ij}, \ \Sigma_i = \begin{pmatrix} \sigma_{iCC} & \sigma_{iCE} \\ \sigma_{iCE} & \sigma_{iEE} \end{pmatrix} \right), \tag{2.1.1}$$

where B_i is a $2 \times p$ parameter matrix, and the two rows of B_i, namely, B'_{iC} and B'_{iE} are $p-$dimensional vectors. Write

$$X_i = (X_{i1}, X_{i2}, ..., X_{in_i}) \quad \text{and} \quad W_i = (\mathbf{w}_{i1}, \mathbf{w}_{i2}, ..., \mathbf{w}_{in_i}), \tag{2.1.2}$$

so that X_i and W_i are $2 \times n_i$ and $p \times n_i$ matrices, respectively. We shall assume that $\text{rank}(W_i) = p$. We then have

$$E(X_i) = W_i B_i \ \text{ and } \text{Cov}[\text{vec}(X_i)] = I_{n_i} \otimes \Sigma_i. \tag{2.1.3}$$

Under the above model, unbiased estimators of B_i and Σ_i are given by

$$\hat{B}_i = X_i W_i'[W_i W_i']^{-1}, \text{ and } \hat{\Sigma}_i = \frac{(X_i - \hat{B}_i W_i)(X_i - \hat{B}_i W_i)'}{n_i - p}, \tag{2.1.4}$$

$i = 1, 2$.

The various cost-effectiveness metrics can now be defined at a specified covariate vector $\mathbf{w}_0$. For example, under the lognormal-lognormal model the INB for an individual with covariate $\mathbf{w}_0$ is given by

$$\begin{aligned} INB(\mathbf{w}_0) &= \lambda\left[\exp\{B_{1E}'\mathbf{w}_0 + \sigma_{1EE}/2\} - \exp\{B_{2E}'\mathbf{w}_0 + \sigma_{2EE}/2\}\right] \\ &\quad - \left[\exp\{B_{1C}'\mathbf{w}_0 + \sigma_{1CC}/2\} - \exp\{B_{2C}'\mathbf{w}_0 + \sigma_{2CC}/2\}\right]. \end{aligned} \tag{2.1.5}$$

The CEP parameter can also be similarly defined.

2.2 Fiducial inference

We shall now develop the fiducial approach for inference concerning the INB and CEP parameters in the bivariate regression model. Fiducial quantities for a normal mean vector and covariance matrix are derived on Bebu and Mathew (2008), and we shall simply adopt these after making obvious modifications for the regression context. We first note that with $\hat{B}_i$ and $\hat{\Sigma}_i$ defined in (2.1.4),

$$\begin{aligned} &\text{vec}(\hat{B}_i) \sim N\left[\text{vec}(B_i), (W_i W_i')^{-1} \otimes \Sigma_i\right] \\ &(n_i - p_i)\hat{\Sigma}_i \sim Wishart\left[\Sigma_i, n_i - p_i\right], i = 1, 2. \end{aligned} \tag{2.2.1}$$

Let $\hat{B}_{io}$ and $\hat{\Sigma}_{io}$ denote the observed values of $\hat{B}_i$ and $\hat{\Sigma}_i$, respectively, $i = 1, 2$. Then a set of fiducial quantities for Σ_i and B_i, say $\tilde{\Sigma}_i$ and $\tilde{B}_i$ are given by

$$\begin{aligned}
\tilde{\Sigma}_i &= H_i^{-1}, \text{ where } H_i \sim W\left(\left\{(n_i - p)\hat{\Sigma}_{io}\right\}^{-1}, n_i - p\right) \\
\tilde{B}_i &= \hat{B}_{io} - \tilde{\Sigma}_i^{1/2}\Sigma_i^{-1/2}(\hat{B}_i - B_i)(W_i W_i')^{1/2}(W_i W_i')^{-1/2} \\
&= \hat{B}_{io} - \tilde{\Sigma}_i^{1/2} Z_i (W_i W_i')^{-1/2}, \qquad (2.2.2)
\end{aligned}$$

where Z_i is a $2 \times p$ matrix whose elements are independent standard normal random variables. A derivation of $\tilde{\Sigma}_i$ is given in Bebu and Mathew (2008).

The fiducial method can now be applied for the individualized INB and CEP metrics. Since $INB(\mathbf{w}_0)$ in (2.1.5) is a function of the B_is and Σ_is, a fiducial quantity for $INB(\mathbf{w}_0)$ can be obtained by replacing the B_is and Σ_is with the corresponding fiducial quantities exhibited in (2.2.2). The 5^{th} percentile of the fiducial quantity for $INB(\mathbf{w}_0)$ provides a 95% lower confidence limit for $INB(\mathbf{w}_0)$. The following algorithm provides the necessary computational steps.

Algorithm 1: Fiducial lower confidence limit for $INB(\mathbf{w}_0)$

1 From the sample of each treatment group compute the estimates $\hat{B}_{io}$ and $\hat{\Sigma}_{io}$, $i = 1, 2$, using (2.1.4).

2 For $i = 1, 2$, independently generate: $H_i \sim W\left(\left\{(n_i - p)\hat{\Sigma}_{io}\right\}^{-1}, n_i - p\right)$ and $Z_i \sim 2 \times p$ matrix of iid $N(0, 1)$ random variates.

3 Compute $\tilde{\Sigma}_i = H_i^{-1}$ and $\tilde{B}_i = \hat{B}_{io} - \tilde{\Sigma}_i^{1/2} Z_i (W_i W_i')^{-1/2}$.

4 Compute the fiducial quantity $\widetilde{INB}(\mathbf{w}_0)$ for $INB(\mathbf{w}_0)$ by using its expression in (2.1.5), and replacing the elements from B_i and Σ_i with the corresponding elements from $\tilde{B}_i$ and $\tilde{\Sigma}_i$.

5 Repeat steps 2-4 M times, obtaining M values of $\widetilde{INB}(\mathbf{w}_0)$.

6 A lower $100(1-\alpha)\%$ confidence limit for $INB(\mathbf{w}_0)$ corresponds to the α^{th} percentile of the M values of $\widetilde{INB}(\mathbf{w}_0)$.

The development of the fiducial inference for the CEP is very similar to that of the INB, and we shall describe it briefly. Thus consider cost-effectiveness data following the regression model (2.1.1), and suppose we are interested in assessing the cost-effectiveness through the CEP at a fixed covariate vector $\mathbf{w}_0$. Let $(C_{i0}, E_{i0})'$ be a random variable denoting the corresponding cost and effectiveness for a patient having covariate vector $\mathbf{w}_0$, assigned to the i^{th} treatment group ($i = 1, 2$). The CEP under the model (2.1.1), denoted by $CEP(\mathbf{w}_0)$, is given by

$$\begin{aligned} CEP(\mathbf{w}_0) &= P[ln(C_{10}) \leq ln(C_{20}), ln(E_{10}) \geq ln(E_{20})] \\ &= \Phi[0; (B_{1C} - B_{2C})'\mathbf{w}_0, \sigma_{1CC} + \sigma_{2CC}] - \Phi_2\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}; \begin{pmatrix} B_{1C}' - B_{2C}' \\ B_{1E}' - B_{2E}' \end{pmatrix}\mathbf{w}_0, \Sigma_1 + \Sigma_2\right] \end{aligned} \tag{2.2.3}$$

In (2.2.3), $\Phi(0; \mu, \sigma^2)$ denotes the cumulative distribution function (CDF) of $N(\mu, \sigma^2)$ evaluated at zero, and $\Phi_2((0\ 0)'; \boldsymbol{\mu}, \Sigma)$ denotes the CDF of $N_2(\boldsymbol{\mu}, \Sigma)$ evaluated at $(0,\ 0)'$. A fiducial quantity for $CEP(\mathbf{w}_0)$ can now be developed similar to that for the INB by simply replacing the unknown parameters with the corresponding fiducial quantities.

The accuracy of the fiducial approach for the individualized metrics will be assessed by computation of coverage probabilities. The algorithm for the coverage probability computation is provided in the next algorithm for $INB(\mathbf{w}_0)$; the algorithm is similar for $CEP(\mathbf{w}_0)$.

Algorithm 2: Coverage probability of the fiducial lower confidence limit for $INB(\mathbf{w}_0)$

---

1 Specify covariate vector $\mathbf{w}_0$ and the design matrices W_i, for $i = 1, 2$.

2 Specify values for the parameters B_i and Σ_i, and compute $INB(\mathbf{w}_0)$.

3 Generate $\hat{B}_{io}$ and $\hat{\Sigma}_{io}$ using 2.2.1.

4 Implement steps 2-6 in algorithm 1.

5 Repeat the steps 2-4 given above M times, resulting in M lower limits for $INB(\mathbf{w}_0)$: $\widetilde{INB}(\mathbf{w}_0)_{1L}, \ldots, \widetilde{INB}(\mathbf{w}_0)_{ML}$.

6 Compute the coverage probability as: $\frac{1}{M} \sum_{m=1}^{M} 1\left[\widetilde{INB}(\mathbf{w}_0)_{mL} < INB(\mathbf{w}_0)\right]$.

An appeal of the CEP is it's flexibility to incorporate the thresholds δ_c and δ_e. Thus it maybe of interest to solve the converse problem: what values of δ_c and δ_e are required to obtain a specified value of CEP? The solution to this question is clearly not unique; after specifying one threshold, the other can be determined using a Newton-Raphson algorithm. Since we shall work under the lognormal-lognormal model (2.1.1), multiplicative thresholds are easier to handle. Thus we shall consider the Newton-Raphson root finding method to find δ_c after fixing a value for δ_e.

Let ω be the pre-specified value of $CEP(\mathbf{w}_0)$. Then

$$\begin{aligned} CEP(\mathbf{w}_0) &= \omega = P[ln(C_{10}) - ln(C_{20}) - ln(\delta_c) \leq 0, ln(E_{10}) - ln(E_{20}) - ln(\delta_e) \geq 0] \\ &= \Phi[ln(\delta_c); (B_{1C} - B_{2C})'\mathbf{w}_0, \sigma_{1CC} + \sigma_{2CC}] - \Phi_2\left[\begin{pmatrix} ln(\delta_c) \\ ln(\delta_e) \end{pmatrix}; \begin{pmatrix} B'_{1C} - B'_{2C} \\ B'_{1E} - B'_{2E} \end{pmatrix}\mathbf{w}_0, \Sigma_1 + \Sigma_2\right] \end{aligned} \tag{2.2.4}$$

Now let

$\mu_{\delta_e} = (B'_{1C} - B'_{2C})\mathbf{w}_0, \;\; \mu_{\delta_c} = (B'_{1E} - B'_{2E})\mathbf{w}_0$

$\sigma_{\delta_c} = \sqrt{\sigma_{1CC} + \sigma_{2CC}}, \;\; \sigma_{\delta_e} = \sqrt{\sigma_{1EE} + \sigma_{2EE}}$

$\rho = \frac{\sigma_{1CE}+\sigma_{2CE}}{\sigma_{\delta_c}\sigma_{\delta_e}}$

Using the above notation the derivative of $CEP(\mathbf{w}_0)$ with respect to δ_c is

$$\begin{aligned}&\frac{dCEP(\mathbf{w}_0)}{d\delta_c} = \\ &\frac{\phi[ln(\delta_c), \mu_{\delta_c}, \sigma_{\delta_c}]}{d\delta_c}\left\{1 - \text{Erfc}\left[\frac{-ln(\delta_e)\sigma_{\delta_c} + \mu_{\delta_e}\sigma_{\delta_c} - \mu_{\delta_c}\rho\sigma_{\delta_e} + \rho\sigma_{\delta_e}ln(\delta_c)}{\sqrt{2-2\rho^2}\sigma_{\delta_c}\sigma_{\delta_e}}\right]\right\}\end{aligned} \tag{2.2.5}$$

such that Erfc[.] is the complementary error function defined as

$$\text{Erfc}[z] = 2\Phi[-\sqrt{2}z] \tag{2.2.6}$$

and Φ[.] is the CDF of the standard normal distribution. Now we can apply the algorithm to solve for δ_c after fixing ω and δ_e. For illustration we set $\ln(\delta_e) = 0$. Let f denote the function for which we want to find the root, and let f' denote the derivative of this function. In this context f and f' are, respectively,

$$\begin{aligned} f(\delta_c) &= CEP(\mathbf{w}_0) - \omega \\ f'(\delta_c) &= \frac{dCEP(\mathbf{w}_0)}{d\delta_c}\end{aligned} \tag{2.2.7}$$

Let $\delta_{c,n}$ denote the root at the n^{th} iteration. The root $\delta_{c,n}$ is updated at iteration n as follows:

$$\delta_{c,n} = \delta_{c,n-1} - \frac{f(\delta_{c,n-1})}{f'(\delta_{c,n-1})} \tag{2.2.8}$$

The steps in the algorithm are as follows.

Algorithm 3: Newton-Raphson method for computing the margin δ_c

For each treatment group compute: $\hat{B}_{i0}$ and $\hat{\Sigma}_{i0}; i = 1, 2.$

Select value of δ_e, ω, and a tolerance a

Provide an initial guess for the root δ_c.

while $D \geq a$ **do**

  $\delta_{c,n} = \delta_{c,n-1} - \frac{f(\delta_{c,n-1})}{f'(\delta_{c,n-1})}$

  $D_n =\mid \delta_{c,n} - \delta_{c,n-1} \mid$

end

The final solution of $\delta_{c,n}$ is the desired δ_c, and is achieved when $\mid \delta_{c,n} - \delta_{c,n-1} \mid < a$

Alternatively, one can simply take the derivative with respect to $ln(\delta_c)$ and repeat the steps in algorithm 3, in which case the final result will need to be exponentiated. Notably, as ω increases the denominator $f'(x_{n-1})$ becomes closer to zero. This indicates that for large values of ω no root exist. In terms of cost-effectiveness this means that the $CEP(\mathbf{w}_0)$ cannot attain certain values of ω.

In light of this, it is also of interest to determine the maximum CEP that a person can attain under the new treatment; this can be done analytically since CEP is monotone in δ_e and δ_c. We note that

$$\begin{aligned} CEP(\mathbf{w}_0) &= P[ln(C_{10}) - ln(C_{20}) - ln(\delta_c) \leq 0, ln(E_{10}) - ln(E_{20}) - ln(\delta_e) \geq 0] \\ &= \Phi[ln(\delta_c); \mu_{\delta_c}, \sigma_{\delta_c}] - \Phi_2 \left[\begin{pmatrix} ln(\delta_c) \\ ln(\delta_e) \end{pmatrix}; \begin{pmatrix} \mu_{\delta_c} \\ \mu_{\delta_e} \end{pmatrix}, \Sigma_1 + \Sigma_2 \right] \\ &\leq P[ln(E_{10}) - ln(E_{20}) - ln(\delta_e) \geq 0] \leq 1 - \Phi[0; \mu_{\delta_e}, \sigma_{\delta_e}] \end{aligned} \tag{2.2.9}$$

Hence, the maximum $CEP(\mathbf{w}_0)$ is equal to (2.2.10) given below, which is attained when $\delta_c \to \infty$ and $\ln(\delta_e) = 0$.

$$max[CEP(\mathbf{w}_0)] = 1 - \Phi[0; (B_{1E} - B_{2E})'\mathbf{w}_0, \sigma_{1EE} + \sigma_{2EE}] \tag{2.2.10}$$

In regards to decision making based on these methods, a range of values for δ_e should be selected that are considered reasonable. Corresponding solutions of δ_c should be computed so that policy-makers and health care experts can make decisions on which combination of δ_e and δ_c is most appropriate and feasible.

2.3 An example

This example is taken from Tunis et al. (2006), and is on the cost-effectiveness of treatments for schizophrenia. The same application was later taken up by Faries et al. (2010). Here the cost-effectiveness is to be evaluated for a treatment that uses the anti-psychotic drug Olanzapine versus a "fail-first approach" where conventional anti-psychotics are first administered, followed by Olanzapine if necessary. We shall refer to the latter as the standard treatment, and treatment using Olanzapine as the new treatment. For more details, we refer to the original study reported in Tunis et al. (2006). In the the clinical trial for this investigation, patients with schizophrenia or schizo-affective disorder were randomized to the two groups. The new treatment group had 202 patients and the standard treatment group had 174 patients. The subject specific covariates collected from the patients were age, duration of psychiatric problems in years, duration (in months) of hospitalization before study, baseline BPRS (brief psychiatric rating scale) level that indicates if a patient has been diagnosed with schizophrenia. Other covariates were also included in the study, such as gender and indicator of substance abuse; however, these were not found to be statistically significant in modelling the bivariate outcomes. The outcomes are total one-year cost and the effectiveness defined in terms of the number of "responder days" for a patient during a one-year period. The latter was calculated using a clinical response (based on the BPRS score) and a social response (in terms of being highly satisfied with social relationships). Details on the computation of such an effectiveness measure is given in Tunis et al. (2006, pg. 80). The authors have also noted that both the cost and effectiveness outcomes exhibited positive skewness. The CEA reported in Faries et al. (2010) for the same application is based on cost and effectiveness data simulated from the original data to obtain lognormal costs and lognormal outcomes, respectively. We will first note that a traditional population-level CEA leads to the conclusion that the new treatment is not cost-effective. However, our individualized analysis does show that the new treatment is indeed cost-effective at the individual level for certain covariate values.

We shall now consider an aggregate analysis which utilizes an aggregate version of the INB, assuming a lognormal-lognormal model without covariates, i.e. (1.2.12). The estimated mean vectors and covariance matrices for the two groups, based on the log-transformed data, are

$$\hat{\mu}_1 = \begin{pmatrix} 9.3289 \\ 4.1725 \end{pmatrix} \; ; \; \hat{\Sigma}_1 = \begin{pmatrix} 1.1618 & -0.1562 \\ -0.1562 & 2.7015 \end{pmatrix}$$
$$\hat{\mu}_2 = \begin{pmatrix} 9.2070 \\ 4.1644 \end{pmatrix} \; ; \; \hat{\Sigma}_2 = \begin{pmatrix} 1.8630 & -0.0166 \\ -0.0166 & 2.7184 \end{pmatrix}$$

Under log-normality, the estimated mean costs $\widehat{E(C_i)}$, and the estimated mean effectiveness $\widehat{E(E_i)}$, $i = 1, 2$, are,

$$\widehat{E(C_1)} = 20127.4000 \; , \; \widehat{E(E_1)} = 250.4655$$
$$\widehat{E(C_2)} = 25296.7400 \; , \; \widehat{E(E_2)} = 250.5288$$

These numerical results indicate that the new treatment, consisting of treatment with Olanzapine, is less costly and somewhat less effective, compared to the standard treatment. This is also reflected in the values of the estimated INB: $\widehat{INB} = \lambda\left[\widehat{E(E_1)} - \widehat{E(E_2)}\right] - \left[\widehat{E(C_1)} - \widehat{E(C_2)}\right]$. Table 2.1 gives the estimated INBs corresponding to a few values of the willingness-to-pay quantity λ (the $\lambda-$ values in Table 2.1 are arbitrarily chosen). As expected, the estimated INB lower limits are all negative (lower limits are obtained using 5000 bootstrap samples). Clearly, from an aggregate perspective, the new treatment is not cost-effective.

Table 2.1 Aggregate level analysis: Estimated INBs and 95% lower confidence limits

λ	$I\hat{N}B$	Lower limit
25	5167.745	−1921.690
50	5166.161	−3392.060
100	5162.993	−8510.540
200	5156.657	−20 364.240

2.3.1 INB for individualized cost-effectiveness analysis

We shall now perform an individualized analysis of the same data, taking into account the following covariates: age, psychological duration, baseline BPRS, and months spent in hospital during the year prior to the study. Thus we have four covariates; however, the regression model that we shall consider will also include an intercept term; thus $p = 5$ in the notations of the model (1.2.10). We shall consider sixteen patients having the covariate values provided in Table 2.2. Now the goal of the analysis is to assess the effect of the covariates on the cost-effectiveness of the new treatment.

Table 2.2 Covariate values for sixteen patients

Patient ID	Age	Psyc duration	Baseline BPRS	Months in Hospital
1	40	0	0	0
2	40	0	1	0
3	40	0	0	0.25
4	40	0	1	0.25
5	40	5	1	0
6	40	5	1	0.25
7	40	5	1	0.50
8	40	5	1	0.75
9	60	0	0	0
10	60	0	1	0
11	60	0	0	0.25
12	60	0	1	0.25
13	60	15	1	0
14	60	15	1	0.25
15	60	15	1	0.50
16	60	15	1	0.75

Using the simulated lognormal-lognormal data, we fitted the multivariate regression model (1.2.10) after log-transforming the data from the two groups. We note that the regression matrix B_i is now a 2×5 matrix, and its columns will correspond to the intercept, and the four covariates age, psychological duration, baseline BPRS, and months spent in hospital during the year prior to the study, respectively. The least squares estimates of B_i and Σ_i are given by

$$\hat{B}_1 = \begin{pmatrix} 9.5580 & -0.0089 & 0.0256 & -0.4731 & 0.1345 \\ 5.6988 & 0.0235 & -0.0357 & -1.7997 & -0.2062 \end{pmatrix},$$

$$\hat{B}_2 = \begin{pmatrix} 7.8333 & 4.0842e-05 & 0.0332 & 0.7911 & -0.0830 \\ 5.4350 & -0.0064 & 0.0163 & -1.3958 & -0.1325 \end{pmatrix},$$

$$\hat{\Sigma}_1 = \begin{pmatrix} 1.0721 & -0.0721 \\ -0.0721 & 2.3638 \end{pmatrix}, \quad \hat{\Sigma}_2 = \begin{pmatrix} 1.6502 & -0.0186 \\ -0.0186 & 2.5081 \end{pmatrix}.$$

The estimated average incremental responder days Δ_e and average incremental cost Δ_c for the sixteen patients are provided in Table 2.3.

Table 2.3 Estimated average incremental responder days and incremental costs for the sixteen patients

Patient ID	$\widehat{\Delta}_e$	$\widehat{\Delta}_c$
1	1874.06	11 201.33
2	258.65	−2149.44
3	1768.92	11 899.91
4	242.93	−1526.75
5	178.03	−3005.02
6	166.13	−2285.78
7	154.93	−1558.82
8	144.38	−823.53
9	3450.38	8437.54
10	525.42	−3878.97
11	3267.37	9041.88
12	496.63	−3314.84
13	213.60	−7960.42
14	199.80	−7085.73
15	186.81	−6204.69
16	174.56	−5316.58

The results in Table 2.3 indicate that all patients have effectiveness that is, on average, larger under the new treatment than the standard treatment. However, the amount of the incremental benefit in effectiveness varies drastically among the patients. Similarly, the average cost of the new treatment compared to the standard treatment also varies significantly amongst the patients, with some patients having large positive average incremental costs and others having large negative incremental costs. All patients have less average costs when using the new treatment as compared to the standard treatment, except for patients having baseline BPRS equal to zero. Older patients tend to experience more benefit and savings from the new treatment compared to younger patients, on average. In addition, patients who have spent more time in hospital in the preceding year have higher average costs under the new treatment compared to those who have spent less time. These patients also tend to benefit less on average from the new treatment.

The INB estimates for the sixteen patients are obtained using (2.1.5), and are provided in Table 2.4 for a few values of the willingness-to-pay parameter λ. Corresponding 95% lower confidence limits are given in Table 2.5.

Table 2.4 INB estimates for the sixteen patients for different values of λ

Patient ID	$\lambda = 25$	$\lambda = 50$	$\lambda = 100$	$\lambda = 200$
1	35 650.2	82 501.8	176 204.9	363 611.2
2	8615.6	15 081.7	28 013.9	53 878.4
3	32 323.2	76 546.3	164 992.4	341 884.7
4	7600.0	13 673.2	25 819.7	50 112.6
5	7455.7	11 906.4	20 807.9	38 610.7
6	6439.1	10 592.3	18 898.9	35 512.0
7	5432.0	9305.2	17 051.6	32 544.4
8	4433.0	8042.6	15 261.6	29 699.6
9	77 822.0	164 081.6	336 600.8	681 639.0
10	17 014.5	30 150.1	56 421.1	108 963.3
11	72 642.4	154 326.8	317 695.4	644 432.7
12	15 730.6	28 146.4	52 977.9	102 641.0
13	13 300.3	18 640.2	29 319.9	50 679.5
14	12 080.9	17 076.0	27 066.2	47 046.7
15	10 874.9	15 545.0	24 885.4	43 566.0
16	9680.5	14 044.5	22 772.4	40 228.2

Table 2.5 95% lower confidence limits for the INB, using 5000 fiducial quantities for different values of λ

Patient ID	$\lambda = 25$	$\lambda = 50$	$\lambda = 100$	$\lambda = 200$
1	−7936.7	−798.3	11 605.9	35 586.3
2	794.0	3548.3	7975.0	15 697.2
3	−9564.0	−3274.8	7558.2	28 793.8
4	−215.1	2159.6	5906.5	12 074.0
5	371.0	1988.9	3960.4	6952.8
6	−504.0	1090.9	2764.2	5409.2
7	−1423.6	18.6	1739.1	4067.6
8	−2340.1	−1018.1	608.6	2634.0
9	5063.1	21 273.0	52 348.7	114 904.0
10	2747.0	7187.0	14 252.3	27 330.2
11	3080.6	18 108.9	47 087.1	103 590.8
12	1679.1	6044.3	12 619.8	24 548.8
13	2248.2	3882.4	5168.2	6874.3
14	1159.0	2768.2	3913.9	4863.4
15	100.4	1674.6	2871.8	3121.3
16	−568.5	870.1	1982.1	2506.7

We now note that unlike in the case of the aggregate analysis, all the INB-estimates are positive, indicating that for fixed subgroups of patients having certain covariate values, the new treatment is cost-effective. A careful examination of the INB-estimates in Table 2.4 also suggests a pattern regarding which patients benefit the most (least). In particular when comparing older patients to younger patients for a fixed willingness-to-pay value λ, and having the other covariate values (i.e., other than age) being common, the older patients have higher estimated values of INB, indicating that the new treatment is more cost-effective for older patients than for younger patients.

Further analysis of affect of patient-level covariates on cost-effectiveness can be assessed by analysing Table 2.5. After the covariate age, the covariate which appears to be a dominating factor in the cost-effectiveness is the psychiatric duration. Patients having psychiatric duration equal to zero have higher INB estimates than those with the same covariate values but having such history. Therefore, the new treatment is most likely to be value for the money for patients who have less history of psychiatric issues compared to those with a longer history. We also notice that for younger patients the new treatment is more cost-effective for patients who have baseline BPRS equal to one when λ is low, however as λ increases those with BPRS equal to zero benefit more. In addition, we find that when all other covariate values are kept fixed, those that have longer hospital stay in year prior to study tend to have smaller INB values. Overall, we find that older patients benefit more, and patients who have had less affliction from schizophrenia tend to have the largest cost-benefit when using the new treatment, especially as λ increases. In order to visualize the differences and trends in the INBs as certain covariate values are changed, we have plotted the estimated INB values against a particular covariate, when the others are held fixed. These plots are given in Appendix B.

The results in Table 2.5 show that even though the INB estimates are all positive, such is not the case with the lower confidence limits. We can clearly see the influence of λ in achieving cost-effectiveness for certain covariate values. The trends evident in the INB estimates are also seen here, with older patients who are less affected by psychiatric issues benefiting the most at any value of λ. For $\lambda = 25$ patients with age equal to 60 achieve cost-effectiveness under the new treatment, except for the patient who has spent 0.75 months of the past year in hospital. For $\lambda = 50$ all older patients have positive lower limits indicating that the new treatment is cost-effective. In contrast, younger patients have negative lower limits at $\lambda = 25$, with the exception of patients having baseline BPRS equal to one and months in hospital equal to zero. However by $\lambda = 50$, most younger patients have positive lower limits except patients with baseline BPRS equal to zero, as well as the patient having months in hospital equal to 0.75. For $\lambda = 100$ and 200 all patients have positive lower limits. It is clear that the different patients achieve cost-effectiveness under the new treatment for different values of λ, thus underscoring the importance of individualizing cost-effectiveness criteria.

2.3.2 The delta method for the interval estimation of the INB

For assessing cost-effectiveness based on the INB at an individual level, the analysis reported so far uses the fiducial approach. One can clearly think of other standard approaches, such as large sample solutions based on the delta method. The delta method is briefly outlined in Appendix A, as it applies to our problem. The resulting lower confidence limits are given in Table 2.6.

Table 2.6 95% lower confidence limits for the INB using the delta method for different values of λ

Patient ID	$\lambda = 25$	$\lambda = 50$	$\lambda = 100$	$\lambda = 200$
1	−20 503.7	−27 974.3	−43 614.5	−75 251.7
2	303.5	2645.7	5785.2	10 908.8
3	−21 233.7	−28 568.7	−44 012.3	−75 295.4
4	−454.5	1774.3	4707.3	9406.9
5	166.8	1922.4	3951.7	6698.3
6	−630.4	1022.5	2888.8	5313.7
7	−1469.2	73.8	1755.8	3809.7
8	−2350.8	−922.1	559.5	2202.2
9	−28 310.3	−46 909.1	−84 523.1	−159 960.9
10	824.3	2691.8	4353.2	6400.6
11	−28 376.9	−46 335.4	−82 712.0	−155 697.1
12	119.7	1923.7	3446.5	5191.9
13	1410.3	3323.1	4754.3	5174.8
14	475.3	2275.9	3561.6	3708.4
15	−499.8	1184.8	2309.9	2146.1
16	−1515.6	51.2	1005.1	502.2

Comparing with the results in Table 2.5, we see that the delta method based lower confidence limits are smaller than those resulting from the fiducial approach; in fact quite a few of the lower confidence limits in Table 2.6 are negative, suggesting that the new treatment is not cost-effective at the respective covariate values, when in fact we reach the opposite conclusion from Table 2.5. In the next section, we shall see that the delta method actually provides a very conservative lower confidence limit, leading to a conclusion against cost-effectiveness, when in fact cost-effectiveness holds. It is the fiducial method that we recommend for the assessment of cost-effectiveness, and this conclusion is consistent with what is noted in Bebu et al. (2016b) and Bebu, Mathew and Lachin (2016).

2.3.3 CEP for individualized cost-effectiveness analysis

If we assume a lognormal-lognormal model without covariates, the CEP can be estimated using the expression

$$
\begin{aligned}
CEP &= P[C_1 \leq C_2, E_1 \geq E_2] \\
&= \Phi[0; (\mu_{1C} - \mu_{2C}), \sigma_{1CC} + \sigma_{2CC}] - \Phi_2 \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}; \begin{pmatrix} \mu_{1C} - \mu_{2C} \\ \mu_{1E} - \mu_{2E} \end{pmatrix}, \Sigma_1 + \Sigma_2 \right],
\end{aligned}
$$

where we recall that Φ and Φ_2 denote the CDF of the univariate normal and bivariate normal distributions for specified values of the mean and variance (respectively, mean vector and covariance matrix). For our simulated data, the estimated CEP has the value $\widehat{CEP}$ = 0.2435, using the above expression. Now let's see how the CEP depends on covariates. Table 2.7 gives the estimated CEPs for the sixteen patients considered earlier, along with the corresponding lower confidence limits, where the latter are obtained using the fiducial approach, with 5000 fiducial quantities.

Table 2.7 CEP estimates and fiducial 95% lower confidence limits for the sixteen patients

Patient ID	CEP estimate	95% lower confidence limit
1	0.154	0.067
2	0.328	0.236
3	0.147	0.063
4	0.318	0.228
5	0.314	0.240
6	0.304	0.233
7	0.294	0.225
8	0.285	0.216
9	0.196	0.071
10	0.404	0.237
11	0.187	0.067
12	0.392	0.228
13	0.361	0.257
14	0.350	0.249
15	0.340	0.240
16	0.330	0.232

Based on the results in Table 2.7, we can draw conclusions regarding the dependence of the CEP on covariates (in particular by analysing the lower limits). As was the case with the INB, age appears to be a dominating factor, with older patients having higher CEPs. The second dominating factor is BPRS. Patients with BPRS equal to 1 (diagnosed with schizophrenia disorder) tend to have higher cost-effectiveness. Patients who spent less time in the hospital during the previous year have higher CEPs when fixing other covariates. Those with longer history of psychological issues appear to benefit more from the new treatment, compared to those with no such prior history.

2.3.4 Coverage probabilities

We shall now report estimated coverage probabilities of the lower confidence limits for the INB, obtained using the fiducial approach and the delta method, and those for the CEP, obtained using the fiducial approach. All the coverage probabilities have been estimated using 10,000 simulated samples, and the fiducial lower confidence limits have been estimated using 1000 fiducial quantities. For the purpose of simulation, the parameter estimates reported in the previous section, i.e., $\hat{B}_1$, $\hat{B}_2$, $\hat{\Sigma}_1$ and $\hat{\Sigma}_2$, will be taken as the true values of B_1, B_2, Σ_1 and Σ_2, respectively. Furthermore, we carried out the simulations under two sample sizes and covariate scenarios: (i) sample sizes and covariates corresponding to the patients in the data that we analysed in the previous section, so that we have $n_1 = 202$ and $n_2 = 174$ patients in the two groups, and (ii) reduce the sample sizes to $n_1 = n_2 = 50$ in each group by randomly selecting 50 patients from each group, and use the corresponding covariate values (these will be held fixed throughout the simulation).

Table 2.8 and Table 2.9 give the estimated coverage probabilities of the lower confidence limits for the INB in the two scenarios mentioned above; Table 2.8 gives the coverage probabilities under the sample sizes $n_1 = 202$ and $n_2 = 174$, and Table 2.9 provides the same under the reduced sample sizes $n_1 = n_2 = 50$ mentioned earlier. The coverage probabilities are reported corresponding to the covariate values for the same sixteen patients considered in the previous section. It should be clear that the fiducial approach satisfactorily maintains the coverage probabilities even under $n_1 = n_2 = 50$.

Table 2.8 Coverage probabilities of the 95% fiducial lower confidence limits for the INB when $n_1 = 202$; $n_2 = 174$ for different values of λ

Patient ID	$\lambda = 25$	$\lambda = 50$	$\lambda = 100$	$\lambda = 200$
1	0.957	0.957	0.956	0.956
2	0.948	0.949	0.949	0.951
3	0.959	0.957	0.956	0.956
4	0.950	0.952	0.954	0.955
5	0.952	0.951	0.951	0.953
6	0.951	0.950	0.951	0.953
7	0.950	0.951	0.951	0.952
8	0.950	0.950	0.952	0.953
9	0.958	0.957	0.956	0.956
10	0.946	0.952	0.957	0.960
11	0.957	0.956	0.957	0.957
12	0.946	0.951	0.956	0.959
13	0.944	0.947	0.950	0.953
14	0.945	0.947	0.950	0.954
15	0.945	0.947	0.950	0.953
16	0.946	0.949	0.949	0.953

Table 2.9 Coverage probabilities of the 95% fiducial lower confidence limits for the INB when $n_1 = n_2 = 50$ for different values of λ

Patient ID	$\lambda = 25$	$\lambda = 50$	$\lambda = 100$	$\lambda = 200$
1	0.967	0.967	0.966	0.966
2	0.944	0.950	0.956	0.957
3	0.968	0.967	0.967	0.965
4	0.946	0.954	0.958	0.960
5	0.948	0.951	0.956	0.959
6	0.948	0.952	0.955	0.958
7	0.949	0.953	0.956	0.959
8	0.948	0.952	0.956	0.959
9	0.964	0.963	0.963	0.962
10	0.944	0.952	0.954	0.959
11	0.964	0.964	0.964	0.963
12	0.945	0.951	0.954	0.958
13	0.946	0.947	0.952	0.956
14	0.946	0.948	0.954	0.958
15	0.947	0.950	0.954	0.958
16	0.948	0.953	0.956	0.958

Table 2.10 gives the coverage probabilities for the delta method based lower confidence limits for the INB when $n_1 = 202, n_2 = 174$. The results show that the delta method is providing very conservative lower confidence limits, consistent with the numerical results noted in the previous section. In other words, the delta method is likely to provide evidence against cost-effectiveness in situations where cost-effectiveness holds. In view of the conservatism of the delta method noted for the case $n_1 = 202, n_2 = 174$, we did not estimate the coverage probabilities for the case $n_1 = n_2 = 50$.

Table 2.10 Coverage probabilities of the 95% delta method based lower confidence limits for the INB when $n_1 = 202$; $n_2 = 174$ for different values of λ

Patient ID	$\lambda = 25$	$\lambda = 50$	$\lambda = 100$	$\lambda = 200$
1	1.000	1.000	1.000	1.000
2	0.973	0.977	0.982	0.984
3	1.000	1.000	1.000	1.000
4	0.973	0.977	0.984	0.985
5	0.973	0.971	0.977	0.977
6	0.971	0.970	0.975	0.979
7	0.970	0.971	0.975	0.978
8	0.967	0.969	0.974	0.978
9	1.000	1.000	1.000	1.000
10	0.992	0.997	0.998	0.999
11	1.000	1.000	1.000	1.000
12	0.992	0.997	0.998	0.999
13	0.981	0.978	0.982	0.985
14	0.980	0.977	0.981	0.985
15	0.978	0.977	0.982	0.985
16	0.978	0.976	0.981	0.985

Table 2.11 gives the estimated coverage probabilities of the fiducial lower confidence limits for the CEP for the sample sizes $n_1 = 202$, $n_2 = 174$ and $n_1 = n_2 = 50$. The coverage probabilities are satisfactory for the case $n_1 = 202$, $n_2 = 174$, but some conservatism is noted for the case $n_1 = n_2 = 50$.

Table 2.11 Coverage probabilities of the 95% fiducial lower confidence limits for the CEP when $n_1 = 202$; $n_2 = 174$ and $n_1 = n_2 = 50$

Patient ID	Coverage probability	
	$n_1 = 202, n_2 = 174$	$n_1 = n_2 = 50$
1	0.959	0.970
2	0.958	0.973
3	0.958	0.970
4	0.957	0.971
5	0.956	0.968
6	0.955	0.968
7	0.954	0.967
8	0.955	0.967
9	0.961	0.975
10	0.965	0.982
11	0.962	0.976
12	0.965	0.983
13	0.962	0.974
14	0.962	0.973
15	0.963	0.972
16	0.962	0.972

The overall picture that emerges from the simulations is that the fiducial approach is quite satisfactory for the interval estimation of the individualized cost-effectiveness metrics considered herein. In the next section, the results of the analysis for cost and effectiveness thresholds of the CEP are reported.

2.3.5 Estimation of the thresholds δ_c and δ_e

We shall now explain the determination of the cost threshold δ_c that provides a desired value of $CEP(\mathbf{w}_0)$, denoted by ω, for a fixed effectiveness threshold, denoted by δ_e. We apply algorithm 3 using $\delta_e = 1$ (i.e. $ln(\delta_e) = 0$), and for a few values of ω. In addition, we complement this analysis by determining the maximum $CEP(\mathbf{w}_0)$ that each of the sixteen patients can obtain. The maximum $CEP(\mathbf{w}_0)$ is determined analytically using (2.2.10). The results of the analysis are provided in Table 2.12.

Table 2.12 Maximum $CEP(\mathbf{w}_0)$ and δ_c values for different values of ω when $\ln(\delta_e) = 0$

Patient ID	Max[CEP($\mathbf{w}_0$)]	$\omega = 0.1$	$\omega = 0.2$	$\omega = 0.3$	$\omega = 0.4$	$\omega = 0.5$	$\omega = 0.6$	$\omega = 0.7$	ω=0.8
1	0.746	0.621	1.391	2.564	4.483	7.982	15.865	49.006	
2	0.684		0.441	0.842	1.546	2.999	7.371		
3	0.743	0.658	1.476	2.724	4.771	8.520	17.043	54.219	
4	0.681		0.468	0.896	1.650	3.217	8.030		
5	0.641		0.465	0.913	1.756	3.728	12.847		
6	0.638		0.494	0.974	1.879	4.021	14.432		
7	0.635		0.525	1.038	2.011	4.339	16.290		
8	0.632		0.558	1.106	2.152	4.684	18.490		
9	0.825	0.472	1.025	1.826	3.050	5.063	8.809	17.787	72.345
10	0.773			0.571	0.981	1.698	3.190	7.940	
11	0.823	0.499	1.086	1.936	3.237	5.382	9.389	19.077	81.447
12	0.771			0.606	1.043	1.810	3.415	8.631	
13	0.654			0.682	1.291	2.655	7.965		
14	0.651		0.372	0.726	1.380	2.858	8.821		
15	0.648		0.395	0.773	1.475	3.077	9.796		
16	0.645		0.420	0.824	1.577	3.315	10.911		

The first column of Table 2.12 shows the maximum attainable $CEP(\mathbf{w}_0)$ for each patient. The remaining columns are the values of δ_c determined by the Newton-Raphson algorithm. Values of ω that a given patient cannot obtain are empty. The empty cells indicate that the specified value of ω of $CEP(\mathbf{w}_0)$ is unattainable in the corresponding set up, and they correspond to relatively small or relatively large values of ω. For some patients the relatively low values of ω cannot be obtained due to the constraint on the effectiveness threshold, $ln(\delta_e) \geq 0$. For large values of ω the denominator of $\frac{dCEP(\mathbf{w}_0)}{d\delta_c}$ in (2.2.5) becomes close to zero, indicating that a patient's $CEP(\mathbf{w}_0)$ increases as δ_c increases and reaches a maximum, after which point increasing δ_c results in no added benefit. We recall that we are in a lognormal-lognormal set up. Thus the threshold δ_c is for the ratio of the costs in the original scale.

From the results of Table 2.12 it is clear that different patients require differing values of δ_c to attain a desired value of ω. The Max[CEP($\mathbf{w}_0$)] values show that older patients attain higher cost-effectiveness compared to younger patients. Those with baseline BPRS equal to zero tend to have larger Max[CEP($\mathbf{w}_0$)]. Also, larger time spent in hospital prior to study result in lower possible cost-effectiveness.

With respect to the trends of ω, patients with lower psychological duration, less time spent in hospital in year prior to study, and those with BPRS at baseline tend to have lower values for the threshold δ_c. Age also affects the value of the threshold; in particular, older patients require small values of δ_c for a fixed ω. In addition to the numerical analysis presented in this section, plots of δ_c as a function of δ_e are provided in Appendix C.

2.4 Discussion

Cost-effectiveness analysis (CEA) is a topic that has considerable contemporary relevance and significance, and continues to be an active area of research due to its implications for health care policy-making. Within the area of cost-effectiveness analysis, an emerging topic of practical interest is that of individualized cost-effectiveness analysis. Traditional CEA based on various aggregate metrics is a well-developed topic having widely accepted practical guidelines and criteria. However, this is not the case with individualized CEA. In the present work, we have developed a framework for individualized CEA, suggesting appropriate criteria and developing the relevant inference. The application discussed clearly demonstrates that in many situations, traditional CEA may not demonstrate cost-effectiveness; however, individualizing it based on covariates does identify the covariate values at which cost-effectiveness holds. This should be of considerable practical relevance for targeted application of new treatments.

When cost and effectiveness responses are affected by covariates, a natural approach is to consider a regression model that will incorporate the covariates. This is the approach we have pursued in the present work; we have proposed a bivariate regression model. We believe that this approach avoids the subjective formulation of subgroups within the population. Furthermore, once the bivariate model has been estimated, it is possible to assess the influence of the covariate values on the CEA criteria, such as the incremental net benefit and the cost-effectiveness probability. Appropriate plots can give quick visual guidance on this, and this can be followed up with a more formal analysis, as we have carried out in this chapter. It maybe possible to effectively use the regression model and the covariate values to construct subgroups. We have not taken this up in the present work, but should be a topic of interest for future research.

Some of the literature on individualized CEA have proposed regression models for certain criteria such as the net monetary benefit. We believe that the original cost and effectiveness data should be modelled using a regression model, and the dependence of various criteria on the covariates should be investigated in the framework of such a model. In addition to the investigation of the regression model for individualized CEA, we also want to emphasize that our proposed inference under the regression model is based on the fiducial idea. As already noted in this chapter, the fiducial approach has already been fruitfully employed by Bebu et al. (2016b) and Bebu, Mathew and Lachin (2016) for aggregate level CEA in the absence of covariates. Our analysis shows that the fiducial approach is remarkably accurate for addressing interval estimation problems relevant to individualized CEA.

In addition to individualizing this cost-effectiveness criteria, we have also developed a method to estimate the thresholds for the cost and effectiveness (δ_c and δ_e, respectively) to obtain a desired value of the CEP for a specified patient. We believe that the determination of δ_c and δ_e could be of importance for individualized CEA as it can aid policy-makers decide on the values of the thresholds that can provide desired values of CEP, and if such desired values are feasible.

Thus our contribution is three-fold: the formulation of the regression framework for individualized CEA, the development of accurate inference for both the INB and CEP criteria, employing the fiducial idea, and the determination of thresholds for the cost and effectiveness required to achieve a desired cost-effectiveness level in terms of the CEP. It is hoped that this work will stimulate further research on individualized CEA.

2.5 Appendix

2.5.1 Appendix A: The delta method

If $\hat{\beta}$ is an estimator of a vector parameter β based on a sample of size n, and if $\sqrt{n}\Big[\hat{\beta}-\beta\Big] \to N\Big[0, \Sigma^*\Big]$ then, by the delta method, $\sqrt{n}\Big[h(\hat{\beta})-h(\beta)\Big] \to N\Big[0, \nabla h^T(\beta)\Sigma^*\nabla h(\beta)\Big]$ where ∇h denotes the gradient of h.

In our application in this chapter, we have $\beta = [vec(B_1)', vec(\Sigma_1)', vec(B_2)', vec(\Sigma_2)']'$. With $\hat{B}_i$ and $\hat{\Sigma}_i$, $i = 1, 2$, given in (2.1.4), we have $\hat{\beta} = \Big[vec(\hat{B}_1)', vec((\hat{\Sigma}_1)', vec((\hat{B}_2)', vec((\hat{\Sigma}_2)'\Big]'$. Then $\hat{\beta}$ has an asymptotic normal distribution with mean β and covariance matrix Σ^* given by

$$\Sigma^* = diag\Big[(W_1W_1')^{-1} \otimes \Sigma_1, \frac{\Omega_1}{n_1 - p}, (W_2W_2')^{-1} \otimes \Sigma_2, \frac{\Omega_2}{n_2 - p}\Big]. \tag{2.5.1}$$

We note that the multivariate normal distribution associated with $vec(\hat{B}_1)$ and $vec(\hat{B}_2)$ are exact. We shall now give the expressions for Ω_1 and Ω_2, For this, let K_{22} be the commutation matrix defined as

$$K_{22} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{2.5.2}$$

Then $\Omega_i = \Big[\Sigma_i \otimes \Sigma_i\Big][K_{22} + I_4]$, $i = 1, 2$.

We note that the INB given in (2.1.5) is an explicit function of β. Thus let's write INB $= h(\beta)$, so that the estimated INB is given by $h(\hat{\beta})$. By the delta method, we can thus conclude that $h(\hat{\beta})$ has an asymptotic normal distribution with mean $h(\beta)$ and variance $\nabla h'(\beta)\Sigma^*\nabla h(\beta)$ where ∇h denotes the gradient of h, which has been determined analytically.

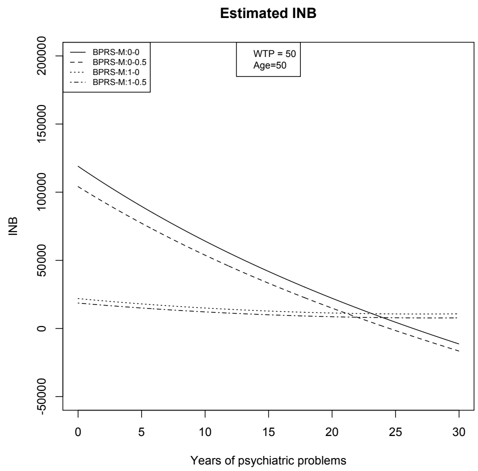
2.5.2 Appendix B: INB plots

In this Appendix, we shall give the plots of the INB estimates against various covariate values, keeping the other covariate values fixed. This will allow us to visualize the trends and patterns in the INB, as a function of the covariates.

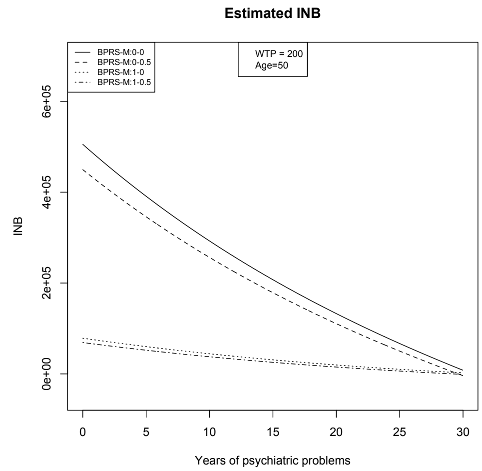
In Figure 2.1, estimated INBs are plotted for four patients having the following combinations of the covariates: baseline BPRS equal to 0 and 1, and months in hospitalization during the year preceding the study having the values 0 months and 0.5 months. All patients have age 50 years and the plots correspond to $\lambda = 50$, and 200. The estimated INBs are plotted against the single covariate "duration of psychiatric problems prior to the start of the study".

From the plot corresponding to $\lambda = 50$, we may conclude the following. The INBs show a decreasing trend as a function of the duration of psychiatric problems. Those with BPRS equal to 0 have much higher values of the INB when psychiatric duration is low. Increased psychiatric duration results in drastically lower INBs for patients with BPRS equal to 0, compared to those with BPRS equal to 1. Starting around 22 years of psychiatric duration, patients with BPRS equal to 1 have higher INBs than those with BPRS equal to 0. In addition, those with BPRS equal to 1 have positive INBs for all years of psychiatric duration. Furthermore, those with no months of hospitalization tend to have higher INBs than those who have been hospitalized.

The plot corresponding to $\lambda = 200$ follows a similar pattern as the one for $\lambda = 50$. However, the INB values are considerably higher for those with BPRS equal to 0 compared to those with BPRS equal to 1, indicating that spending more for such patients results in greater cost-effectiveness than those diagnosed with schizophrenia.



(a) $\lambda = 50$

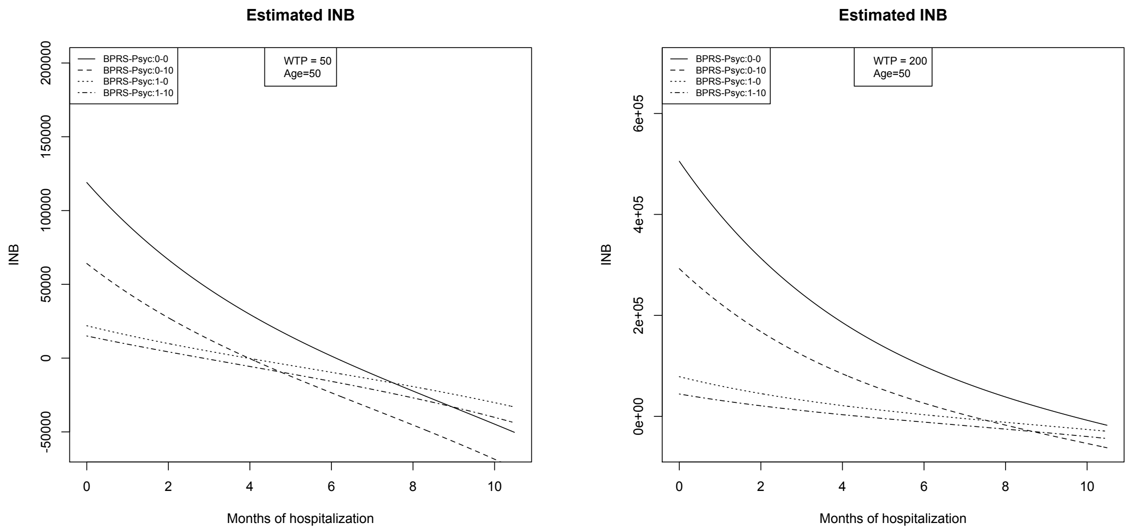


(b) $\lambda = 200$

Figure 2.1: Estimated INB as function of duration of psychiatric problems for patients aged 50 for a few combinations of Baseline BPRS and months in hospital during year preceding the study

Next, Figure 2.2 gives some plots of the estimated INB as a function of months spent in hospitalization in the year preceding the study. The four patients used in Figure 2.2 are also 50 years old and correspond to combinations of BPRS equal to 0, 1, and duration of psychiatric problems equal to 0 and 10 years .

The plot corresponding to $\lambda = 50$ shows that patients with BPRS equal to 0 have higher INB values when months spent in hospitalization are lower. However, for individuals who have spent almost a year in hospitalization, those with BPRS equal to 1 have slightly higher INB values compared with those having BPRS equal to zero. In addition, those with psychiatric duration equal to 0 tend to benefit more than those with psychiatric duration equal to 10. The plot corresponding to $\lambda = 200$ also shows a similar pattern. However, the INB values are greatly increased for patients with BPRS equal to 0 when going from $\lambda = 50$ to $\lambda = 200$.



(a) $\lambda = 50$ **(b)** $\lambda = 200$

Figure 2.2: Estimated INB as function of months in hospital in year preceding the study for patients aged 50 for various combinations of Baseline BPRS and psychological duration

2.5.3 Appendix C: Plots of cost and effectiveness thresholds

In this Appendix we continue the analysis of patient-specific cost and effectiveness thresholds required to obtain desired levels of CEP. Plots of the cost threshold, δ_c, against the effectiveness threshold, δ_e, for eight patients are provided. The plots corresponding to patients 1, 3, 5, and 7 are shown in Figure 2.3. For patients 9, 11, 13, and 15, the plots are shown in Figure 2.4. Similar analysis can be conducted for the remaining eight patients in Table 2.2.

The plot for each patient includes various values of desired $\omega = CEP$, corresponding to the possible values attainable for that patient (see Table 2.12). As noted earlier, different combinations of the cost and effectiveness thresholds can be utilized to obtain a desired CEP value. Further, graphical representations of possible choices for the thresholds δ_e and δ_c could be of practical use to decision makers.

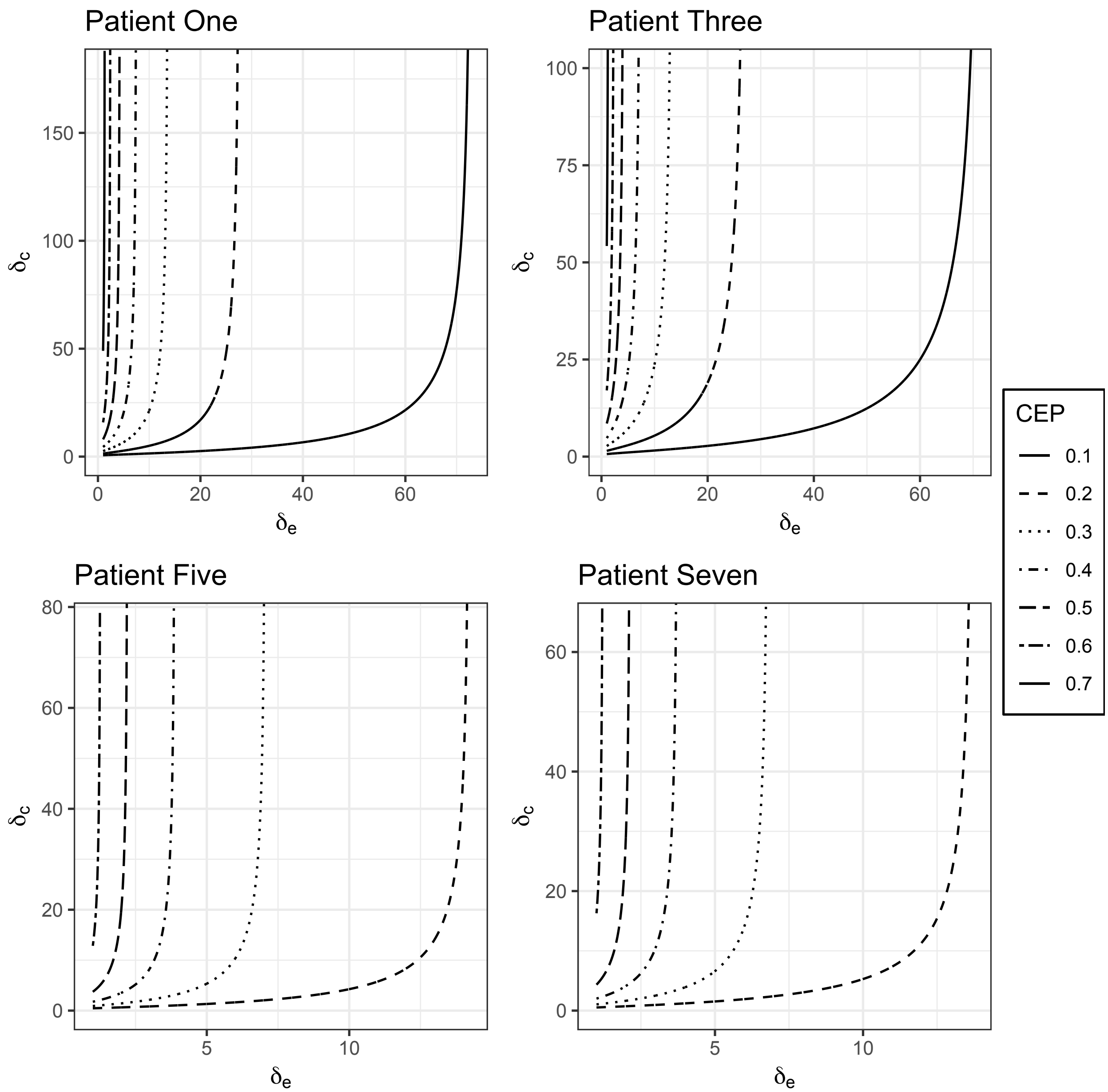


Figure 2.3: Plots of CEP as a function of δ_c and δ_e for patients aged 40

As noticed in Table 2.12, patients who have not been diagnosed with schizophrenia at the start of the study are able to attain higher values of CEP; however, for fixed δ_e the required δ_c tends to be higher for such patients. In addition, upon close examination we notice that for fixed δ_e and ω, months spent in hospital during the year preceding the study increase δ_c. Comparing the plots for patients 1 and 3 with patients 5 and 7, we observe that the latter two patients have lower δ_c for a fixed δ_e in order to achieve the same value of CEP. Thus, patients who are more severely affected by the disorder can obtain the same CEP value for the same minimum effectiveness and the lower investment, compared to healthier patients. Results for patients 9, 11, 13, and 15 are provided in the next figure.

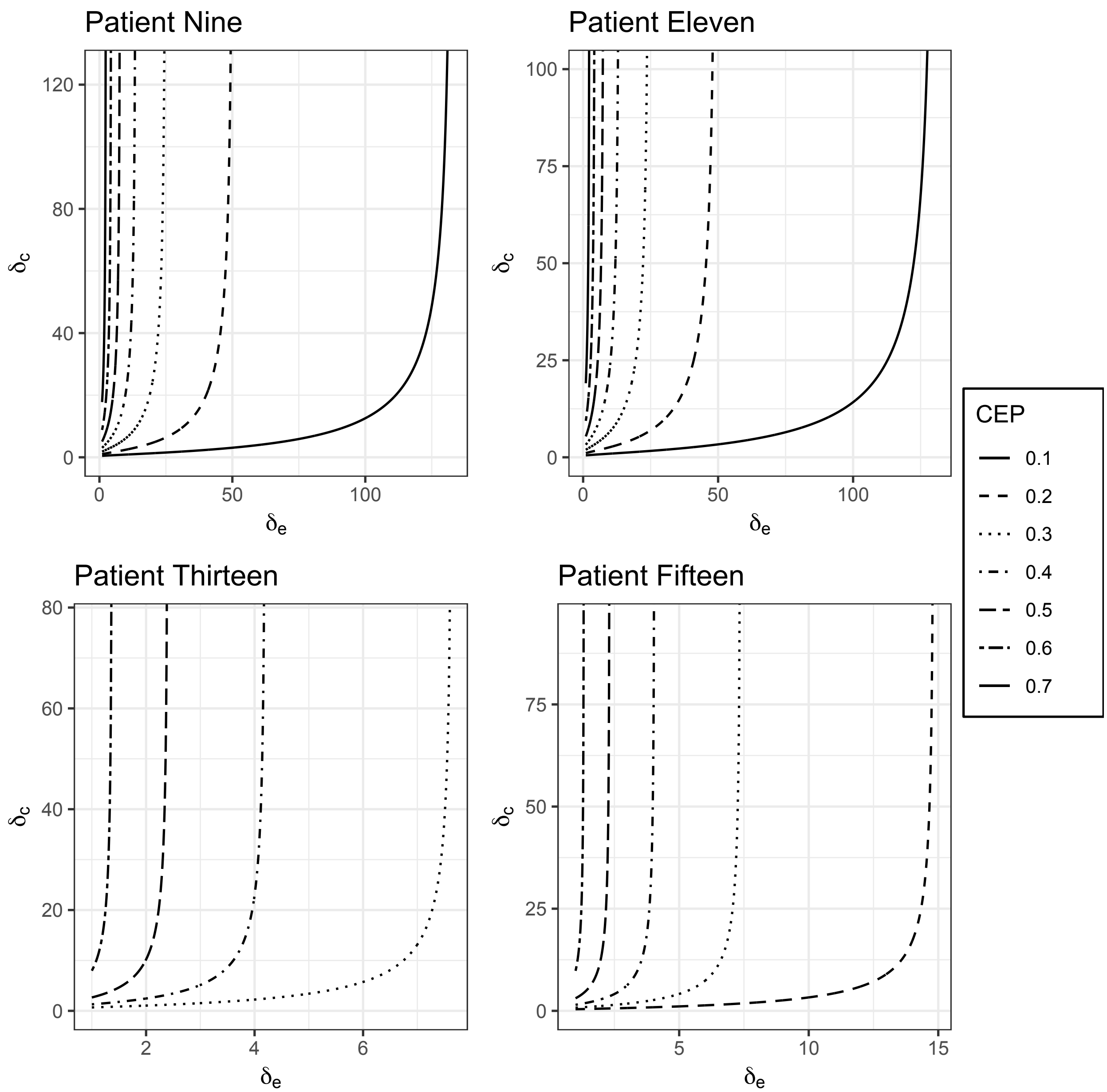


Figure 2.4: Plots of CEP as a function of δ_c and δ_e for patients aged 60

The same trends noted from Figure 2.3 are present in the above figure for older patients. The covariate age has a decreasing relationship with δ_c for a fixed δ_e. That is, in order to attain the same cost-effectiveness under a fixed minimum effectiveness, the total investment decreases as patients get older. In addition, the effect of months spent in hospital during the year preceding the study appears to affect older patients more than younger patients.

The above graphical analyses show the trend of δ_c as a function of δ_e. Increasing the effectiveness threshold increases the cost threshold as a convex function until a saturation level is reached, indicating that beyond a certain minimum effectiveness threshold, there is practically no impact on δ_c. In addition, fixing the δ_e and increasing the desired CEP results in a higher δ_c. In other words, the required amount of increased investment under the new treatment increases for a fixed minimum effectiveness in order to obtain an increase in cost-effectiveness.

Chapter 3

Multi-criteria decision analysis: A probabilistic approach

Applications of multi-criteria decision analysis (MCDA) have become increasingly important in aiding decision making in health economics. Health care economics studies often have multiple measures of effectiveness when assessing the cost-effectiveness of competing treatments. Thus far relevant literature on the topic has focused on combining these multiple measures of effectiveness into scalar quantities such as QALYs. However, when combining multiple measures of effectiveness into a single quantity, important information may be lost. Moreover, this method of combining measures is subjective. The emerging field of MCDA focuses on evaluating treatments based on multiple criteria (e.g. multiple effectiveness measures). The number of MCDA-related publications has increased at an annual rate of 83.23% from 2006 to 2017, see Wu et al. (2018), demonstrating the need for accurate and precise inference methods.

Generally, the recommendation is to combine the different criteria into a single criterion by forming a weighted combination; see Thokala and Duenas (2012) and Thokala et al. (2016). In the latter article, the authors note:

> Weighting involves eliciting stakeholders' preferences between criteria. Weights rep-

resent trade-offs between criteria and are used to combine the scores on individual criterion into a measure of total value. Thokala et al. (2016, pg. 9).

Additionally, weights can be constructed as to reflect patient preferences; see Broekhuizen et al. (2017). Different weighting schemes have been discussed extensively in Schey et al. (2017). The authors noted the prevalence of thirteen commonly used weighting procedures. They concluded that disease severity is one criteria that health experts recommend prioritizing, while healthcare policy-makers prioritize value for the money.

In their work, Thokala et al. (2016) highlight the importance of MCDA in the process of choosing between competing treatments. The authors note that selecting a procedure for weighting is not obvious, especially when considering differing preferences of decision makers. However, Thokala and Duenas (2012) have made recommendations regarding the assessment of which criteria and weights should be utilized. The authors recommend choosing weights that are robust and perform well using a sensitivity analysis. Based on their analysis, Wen, Zhang and Yang (2014) recommend MCDA methods that account for correlation between criteria. A Bayesian approach is considered in O Meachair and Walsh (2014). Clearly, there is an abundance of MCDA methods; however, no clear consensus has been reached on selection of criteria and relative weights to be used.

Alternatively, one can think of performing cost-effectiveness analysis by appropriately modelling and analysing the multivariate data resulting from multiple effectiveness measures. This is especially desirable if there is no clearly defined methodology for choosing the weights utilized in combining the different effectiveness measures. In this chapter, we shall take up such a multivariate modelling. We propose two new metrics which are adaptations of the cost-effectiveness probability (CEP) metric, and they avoid the use of weights. To account for the different preferences of decision makers, we formulate two probabilistic measures, one which prioritizes effectiveness, the other prioritizing value for the money. Our proposed approach is then applied to a motivating example involving two effectiveness measures where ($\ln(\text{Cost})$, Effect_1, Effect_2) follows a multivariate normal distribution. A fiducial approach will be pursued for the interval estimation, and it will be noted that the fiducial methodology provides accurate inference. In addition, a non-parametric solution is provided based on U-statistics, and its accuracy will also be assessed.

3.1 The MCDA set-up

In this section, we shall first review various weighting schemes proposed in the MCDA literature. Then, we define our cost-effectiveness proportion (CEP) cost-effectiveness criteria keeping in mind that we are comparing just two treatments: Treatment 1 (a new treatment), and Treatment 2 (an existing treatment).

3.1.1 MCDA weighting schemes

The MCDA literature has thus far relied on weighting schemes to assess the performance of competing treatments. Criteria are numerical measures of performance by which treatments are evaluated, and are often scores from experts. These criteria are selected as relevant properties from the decision maker's point-of-view. This means that when comparing the same treatments two decision makers may select different criteria. Suppose we have g criteria, and e_{ik} is the k^{th} criterion for the i^{th} treatment with an associated weight w_k, where these have been decided by a policy-maker, a "value" for the i^{th} treatment, say V_i, is defined as

$$V_i = \sum_{k=1}^{g} w_k e_{ik} \tag{3.1.1}$$

, $i = 1, 2$. Treatment T_1 is preferred to treatment T_2 if $V_1 > V_2$. This synthesizing approach has numerous variations in the MCDA literature. The brief outline provided above is summarized in Baltussen and Niessen (2006). In Thokala and Duenas (2012) the authors use performance scores that are between 0 to 1. The authors then assign weights independently of the performance scores. They also note that the weighting combinations need not sum to one. Huang, Keisler and Linkov (2011) state:

> These approaches share common mathematical elements, i.e., values for alternatives [treatments] are assigned for a number of dimensions [preference scores], and then multiplied by weights and finally combined to produce a total score. The approaches differ significantly in the details of how values are assigned and combined, meaning that the processes have different information - and knowledge -requirements and the calculated scores have different mathematical properties and thus slightly different meanings. (p. 3579)

From the brief overview of this widely used method, the issues associated with the method should be clear: subjectivity of the weights and the criteria. This could affect the conclusion on which treatment is to be preferred; later we shall give an example that illustrates this. We develop two metrics that do not use any weighting of criteria; these metrics are probabilistic. In addition, we only use actual effectiveness outcomes as criteria. This is an important distinction from the current MCDA framework.

3.2 A model for cost and effectiveness

We shall now specify a parametric model for CEA; the model is a lognormal-normal model, similar to the parametric model used in the previous chapter, except that we are now considering two effectiveness measures. The case of more than two effectiveness measures can be handled similarly.

Let C_{ij}, E_{1ij} and E_{2ij}, respectively, denote the cost, and the two effectiveness measures for the j^{th} patient in the i^{th} treatment group; $j = 1, 2, ..., n_i$, $i = 1, 2$. In the absence of covariates, we shall consider the following "lognormal-normal-normal" model:

$$X_{ij} = \begin{pmatrix} \ln(C_{ij}) \\ E_{1ij} \\ E_{2ij} \end{pmatrix} \sim N \left(\mu_i = \begin{pmatrix} \mu_{iC} \\ \mu_{E_{1i}} \\ \mu_{E_{2i}} \end{pmatrix}, \boldsymbol{\Sigma}_i \right), \ i = 1, 2 \tag{3.2.1}$$

For later use, we shall denote the elements of the covariance matrix in (3.2.1) as

$$\Sigma_i = \begin{bmatrix} \sigma_{C_i} & \sigma_{C_i,E_{1i}} & \sigma_{C_i,E_{2i}} \\ \sigma_{C_i,E_{1i}} & \sigma_{E_{1i}} & \sigma_{E_{1i},E_{2i}} \\ \sigma_{C_i,E_{2i}} & \sigma_{E_{1i},E_{2i}} & \sigma_{E_{2i}} \end{bmatrix} \tag{3.2.2}$$

Next, we introduce two probabilistic criteria. As noted by other authors, there tends to be differing preferences amongst stake holders. Therefore, to account for such differing preferences, we construct two probabilistic metrics, one that prioritizes effectiveness and the other that prioritizes cost. The two metrics are conditional probabilities, modifying the CEP parameter defined in (1.2.3). For simplicity of notation, let

$$\mu_\Delta = \begin{pmatrix} \mu_{1C} - \mu_{2C} \\ \mu_{E_{12}} - \mu_{E_{11}} \\ \mu_{E_{22}} - \mu_{E_{21}} \end{pmatrix} = \begin{pmatrix} \mu_{\Delta_C} \\ \mu_{\Delta_{E_1}} \\ \mu_{\Delta_{E_2}} \end{pmatrix} \tag{3.2.3}$$

$$\begin{aligned}\Sigma_\Delta = \Sigma_1 + \Sigma_2 &= \left[\begin{array}{c|c} \sigma_{\Delta_C} & \Sigma_{\Delta_C,(\Delta_{E_1},\Delta_{E_2})} \\ \hline \Sigma_{\Delta_C,(\Delta_{E_1},\Delta_{E_2})} & \Sigma_{\Delta_{E_1},\Delta_{E_2}} \end{array}\right] = \\ &\left[\begin{array}{c|c} \sigma_{1C} + \sigma_{2C} & \begin{pmatrix} \sigma_{C_1,E_{11}} + \sigma_{C_2,E_{12}} & \sigma_{C_1,E_{21}} + \sigma_{C_2,E_{22}} \end{pmatrix} \\ \hline \begin{pmatrix} \sigma_{C_1,E_{11}} + \sigma_{C_2,E_{12}} \\ \sigma_{C_1,E_{21}} + \sigma_{C_2,E_{22}} \end{pmatrix} & \begin{pmatrix} \sigma_{E_{11}} + \sigma_{E_{12}} & \sigma_{E_{12},E_{12}} + \sigma_{E_{11},E_{21}} \\ \sigma_{E_{12},E_{12}} + \sigma_{E_{11},E_{21}} & \sigma_{E_{22}} + \sigma_{E_{21}} \end{pmatrix} \end{array}\right]\end{aligned} \tag{3.2.4}$$

The first conditional probability metric we shall define is conditional on those individuals for whom the new treatment is less costly than the standard treatment (3.2.5). The second metric is conditional on those individuals for whom the new treatment is more effective than the standard treatment (3.2.6). In terms of the lognormal-normal-normal model, the two metrics can be written as

$$
\begin{aligned}
CEP_1 &= P[E_{11} - E_{12} \geq 0, E_{21} - E_{22} \geq 0 | \ln(C_1) - \ln(C_2) \leq 0] \\
&= \frac{\Phi_3(0; \mu_\Delta, \Sigma_\Delta)}{\Phi(0, \mu_{\Delta_C}, \sigma_{\Delta_C})}
\end{aligned} \tag{3.2.5}
$$

$$
\begin{aligned}
&CEP_2 = P[\ln(C_1) - \ln(C_2) \leq 0 | E_{11} - E_{12} \geq 0, E_{21} - E_{22} \geq 0] = \\
&\frac{\Phi_3(0; \mu_\Delta, \Sigma_\Delta)}{\Phi_2\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}; \begin{pmatrix} \mu_{\Delta_{E_1}} \\ \mu_{\Delta_{E_2}} \end{pmatrix}, \Sigma_{\Delta_{E_1}, \Delta_{E_2}}\right]},
\end{aligned} \tag{3.2.6}
$$

where the various quantities are defined in (3.2.3) and (3.2.4), $\Phi_3(0; \mu, \Sigma)$ denotes the CDF of the trivariate normal distribution with mean μ and covariance matrix Σ, where the CDF is evaluated at $(0, 0, 0)'$, and Φ_2 denotes the CDF for the bivariate normal distribution. The conditional CEP in (3.2.5), CEP_1, is of interest to decision makers whom prioritize treatments based on total cost. The second CEP, CEP_2 in (3.2.6), is of primary interest to decision makers prioritizing the effectiveness of treatment. In the next section, we shall implement the fiducial approach and the percentile bootstrap for the interval estimation of the proposed metrics.

3.3 Parametric inference

We shall now explore the fiducial approach and the parametric bootstrap for inference concerning the CEP parameters under the model (3.2.1). The aim is to construct lower confidence limits for each CEP metric. Fiducial quantities for a normal mean vector and covariance matrix are derived in Bebu and Mathew (2008), and we modify their results for the multiple effectiveness model.

With X_{ij}s as defined in (3.2.1), unbiased estimates of μ_i and Σ_i, say $\hat{\mu}_i$ and $\hat{\Sigma}_i$, respectively, are given by

$$
\begin{aligned}
\hat{\mu}_i &= \frac{1}{n_i}\sum_{j=1}^{n_i} X_{ij} \sim N\left[\mu_i, \frac{\Sigma_i}{n_i}\right] \\
\hat{\Sigma}_i &= \frac{1}{n_i-1}\sum_{j=1}^{n_i}(X_{ij}-\hat{\mu}_i)(X_{ij}-\hat{\mu}_i)' \sim Wishart\left[\Sigma_i, n_i-1\right],
\end{aligned} \tag{3.3.1}
$$

$i = 1,\ 2$. Let $\hat{\mu}_{io}$ and $\hat{\Sigma}_{io}$ denote the observed values of $\hat{\mu}_i$ and $\hat{\Sigma}_i$, respectively, for $i = 1,\ 2$. Then a set of fiducial quantities for Σ_i and μ_i, say $\tilde{\Sigma}_i$ and $\tilde{\mu}_i$ are given by

$$
\begin{aligned}
\tilde{\Sigma}_i &= H_i^{-1}, \quad \text{where} \quad H_i \sim W\left(\left\{(n_i-1)\hat{\Sigma}_{io}\right\}^{-1}, n_i-1\right) \\
\tilde{\mu}_i &= \hat{\mu}_{io} - \tilde{\Sigma}_i^{1/2} Z_i \frac{1}{\sqrt{n_i}},
\end{aligned} \tag{3.3.2}
$$

where Z_i is a 3×1 vector whose elements are independent standard normal random variables. A derivation of the above fiducial quantities is given in Bebu and Mathew (2008).

Since CEP_1 and CEP_2 are functions of the elements of μ_is and Σ_is, a fiducial quantity for each can be obtained by replacing μ_is and Σ_is with the corresponding fiducial quantities exhibited in (3.3.2). The 5^{th} percentile of the fiducial quantity of each probabilistic metric provides a 95% lower confidence limit for that metric. Next, we present the algorithm that provides the necessary computational steps. The algorithm 4 shows the required steps to obtain fiducial limits for CEP_1; the algorithm is similar for CEP_2.

Algorithm 4: Fiducial lower confidence limit for CEP_1

1 From the sample of each treatment group compute the estimates $\hat{\mu}_{i0}$ and $\hat{\Sigma}_{i0}$, $i = 1, 2$, using the trivariate model (3.2.1).

2 For $i = 1, 2$, independently generate: $H_i \sim W\left(\left\{(n_i - 1)\hat{\Sigma}_{i0}\right\}^{-1}, n_i - 1\right)$ and $Z_i \sim 3 \times 1$ matrix of $N(0, 1)$ random variates.

3 Compute $\tilde{\Sigma}_i = H_i^{-1}$ and $\tilde{\mu}_i = \hat{\mu}_{io} - \tilde{\Sigma}_i^{1/2} Z_i \frac{1}{\sqrt{n_i}}$.

4 Compute the fiducial quantity $\widetilde{CEP_1}$ for CEP_1 by replacing the elements from μ_i and Σ_i with the corresponding elements from $\tilde{\mu}_i$ and $\tilde{\Sigma}_i$ using expression (3.2.5).

5 Repeat steps 2-4 M times, obtaining M values of $\widetilde{CEP_1}$.

6 A lower $100(1-\alpha)\%$ confidence limit for CEP_1 corresponds to the α^{th} percentile of the M values of $\widetilde{CEP_1}$.

In addition to the fiducial approach, we will also use the percentile bootstrap method for confidence interval construction, the algorithm for this is provided next.

Algorithm 5: Percentile bootstrap lower confidence limit for CEP_1

1 From the sample of each treatment group compute the estimates $\hat{\mu}_{i0}$ and $\hat{\Sigma}_{i0}$, $i = 1, 2$, using the trivariate model (3.2.1).

2 For $i = 1, 2$, independently generate: $\hat{\mu}_i \sim N\left[\hat{\mu}_{i0}, \frac{\hat{\Sigma}_{i0}}{n_i}\right]$ and $\hat{\Sigma}_i \sim \frac{1}{n_i - 1} W[\hat{\Sigma}_{io}, n_i - 1]$ random variates.

3 Compute the bootstrap quantity $\widehat{CEP_1}$ for CEP_1 by replacing the elements from μ_i and Σ_i with the corresponding elements from $\hat{\mu}_i$ and $\hat{\Sigma}_i$ using expression (3.2.5).

4 Repeat steps 2-3 M times, obtaining M values of $\widehat{CEP_1}$.

5 A lower $100(1-\alpha)\%$ confidence limit for CEP_1 corresponds to the α^{th} percentile of the M values of $\widehat{CEP_1}$.

We will now present the U-statistics approach next.

3.4 Non-parametric inference

In this section we formulate the U-statistics approach that will be used to obtain non-parametric lower confidence limits for CEP_1 and CEP_2. A general discussion of multivariate U-statistics is presented in Yu et al. (2018). We shall first exhibit bivariate U-statistics such that estimates of CEP_1 and CEP_2 are ratios of the two components of the bivariate U-statistics. Then we utilize the asymptotic normality of the bivariate U-statistics along with the delta method to obtain approximate lower confidence limits for CEP_1 and CEP_2.

Let the U-statistic corresponding to the numerator of CEP_1 defined in (3.2.5) (and also CEP_2 defined in (3.2.6)) be denoted by U_n, and the U-statistics in the denominator of CEP_1 and CEP_2 be denoted by U_{1d} and U_{2d}, respectively. We shall shortly define U_n, U_{1d} and U_{2d}. Before doing so, let

$$U_{CEP_1} = \frac{U_n}{U_{d1}} \tag{3.4.1}$$

and

$$U_{CEP_2} = \frac{U_n}{U_{d2}} \tag{3.4.2}$$

so that U_{CEP_1} and U_{CEP_2} are estimates of CEP_1 and CEP_2, respectively. Here are the expressions for U_n, U_{1d} and U_{2d}:

$$U_n = \frac{1}{n_1 n_2} \sum_{j=1}^{n_1} \sum_{k=1}^{n_2} h_1 \left[(C_{1j}, E_{11j}, E_{21j}); (C_{2k}, E_{12k}, E_{22k})\right] \tag{3.4.3}$$

$$U_{d1} = \frac{1}{n_1 n_2} \sum_{j=1}^{n_1} \sum_{k=1}^{n_2} h_2 \left[(C_{1j}); (C_{2k})\right] \tag{3.4.4}$$

$$U_{d2} = \frac{1}{n_1 n_2} \sum_{j=1}^{n_1} \sum_{k=1}^{n_2} h_3 \left[(E_{11j}, E_{21j}); (E_{12k}, E_{22k})\right] \tag{3.4.5}$$

, where the kernels h_1, h_2 and h_3 are given by

$$h_1\left[(C_{1j}, E_{11j}, E_{21j}); (C_{2k}, E_{12k}, E_{22k})\right] = \begin{cases} 1 & \text{if } C_{1j} - C_{2k} \leq 0; E_{11j} - E_{12k} \geq 0; E_{21j} - E_{22k} \geq 0 \\ 0 & \text{otherwise} \end{cases} \tag{3.4.6}$$

$$h_2\left[(C_{1j}); (C_{2k})\right] = \begin{cases} 1 & \text{if } C_{1j} - C_{2k} \leq 0 \\ 0 & \text{otherwise} \end{cases} \tag{3.4.7}$$

$$h_3\left[(E_{11j}, E_{21j}); (E_{12k}, E_{22k})\right] = \begin{cases} 1 & \text{if } E_{11j} - E_{12k} \geq 0; E_{21j} - E_{22k} \geq 0 \\ 0 & \text{otherwise} \end{cases} \tag{3.4.8}$$

Let

$$U_1 = \begin{pmatrix} U_n \\ U_{d1} \end{pmatrix}, \text{ and } E(U_1) = \theta_1. \tag{3.4.9}$$

We note that the bivariate vector θ_1 consists of the numerator and denominator of CEP_1 defined in (3.2.5) If $N = n_1 + n_2$, then

$$\sqrt{N}\left[U_1 - \theta_1\right] \sim N_2\left[0, \Sigma_{U_1}\right], \text{ asymptotically.} \tag{3.4.10}$$

We shall write

$$\Sigma_{U_1} = \begin{pmatrix} \sigma^2_{U_n} & \sigma_{U_n,U_{d1}} \\ \sigma_{U_n,U_{d1}} & \sigma^2_{U_{d1}} \end{pmatrix} \tag{3.4.11}$$

Similarly, let

$$U_2 = \begin{pmatrix} U_n \\ U_{d2} \end{pmatrix}, \text{ and } E(U_2) = \theta_2, \tag{3.4.12}$$

where θ_2 consists of the numerator and denominator of CEP_2 defined in (3.2.6). We then have

$$\sqrt{N}\left[U_2 - \theta_2\right] \sim N_2\left[0, \Sigma_{U_2}\right], \text{ asymptotically.} \tag{3.4.13}$$

We shall write

$$\Sigma_{U_2} = \begin{pmatrix} \sigma^2_{U_n} & \sigma_{U_n,U_{d2}} \\ \sigma_{U_n,U_{d2}} & \sigma^2_{U_{d2}} \end{pmatrix} \tag{3.4.14}$$

We will invoke the above asymptotic results to obtain lower confidence limits for CEP_1 and CEP_2. The estimates of the covariance matrices, Σ_{U_1} and Σ_{U_2} will now be given. First we will consider the estimation of $\sigma^2_{U_n}$. Standard theory of U-statistics give

$$\sigma^2_{U_n} = \frac{N}{n_1}\epsilon_{10,n} + \frac{N}{n_2}\epsilon_{01,n} + \frac{N}{n_1 n_2}\epsilon_{11,n},$$

where

$$\begin{aligned} \epsilon_{10,n} &= Cov[h_1(X_1,Y_1); h_1(X_1,Y_1^{'})] \\ \epsilon_{01,n} &= Cov[h_1(X_1,Y_1); h_1(X_1^{'},Y_1)] \\ \epsilon_{11,n} &= Cov[h_1(X_1,Y_1); h_1(X_1,Y_1)] \end{aligned} \tag{3.4.15}$$

where $X_1 = (C_{1j}, E_{11j}, E_{21j})$ and $X_1^{'}$ refer to values of cost and effectiveness for two different subjects from the first group, and likewise $Y_1, Y_1^{'}$ from the second group, all independent. Let X $\succ$ Y denote the event that subject X has lower cost and higher effectiveness measures than subject Y. Then, we can write

$$\begin{aligned} \epsilon_{10,n} &= P[X_1 \succ Y_1 \ \& \ X_1 \succ Y_1^{'}] - P_1^2(X_1 \succ Y_1) \\ \epsilon_{01,n} &= P[X_1 \succ Y_1 \ \& \ X_1^{'} \succ Y_1] - P_1^2(X_1 \succ Y_1) \\ \epsilon_{11,n} &= P[X_1 \succ Y_1 \ \& \ X_1 \succ Y_1] - P_1^2(X_1 \succ Y_1) \end{aligned} \tag{3.4.16}$$

Let $S_{jk} = h_1(X_j, Y_k)$; $S_{j.} = \sum_{k=1}^{n_2} S_{jk}$; $S_{.k} = \sum_{j=1}^{n_1} S_{jk}$. The estimates of the quantities in (3.4.16) are as follows.

$$\hat{P}(X_1 \succ Y_1) = \frac{1}{n_1 n_2}\sum_{j=1}^{n_1}\sum_{k=1}^{n_2} S_{jk} = U_n \tag{3.4.17}$$

$$\hat{P}(X_1 \succ Y_1 \ \& \ X_1 \succ Y_1^{'}) = \frac{1}{n_1 n_2 (n_2 - 1)} \sum_{j=1}^{n_1} S_{j.}(S_{j.} - 1) \tag{3.4.18}$$

$$\hat{P}(X_1 \succ Y_1 \ \& \ X_1^{'} \succ Y_1) = \frac{1}{n_1 n_2 (n_1 - 1)} \sum_{k=1}^{n_2} S_{.k}(S_{.k} - 1) \tag{3.4.19}$$

$$\hat{P}(X_1 \succ Y_1 \ \& \ X_1 \succ Y_1) = U_n \tag{3.4.20}$$

The above estimates provide us with an estimate of $\sigma^2_{U_n} = \frac{N}{n_1}\epsilon_{10,n} + \frac{N}{n_2}\epsilon_{01,n} + \frac{N}{n_1 n_2}\epsilon_{11,n}$.

Now we will show how to estimate $\sigma^2_{U_{d1}}$. For this we use the observation $\sigma^2_{U_{d1}} = \frac{N}{n_1}\epsilon_{10,d1} + \frac{N}{n_2}\epsilon_{01,d1} + \frac{N}{n_1 n_2}\epsilon_{11,d1}$, where

$$\begin{aligned}
\epsilon_{10,d1} &= Cov[h_2(X_1, Y_1); h_2(X_1, Y_1^{'})] \\
\epsilon_{01,d1} &= Cov[h_2(X_1, Y_1); h_2(X_1^{'}, Y_1)] \\
\epsilon_{11,d1} &= Cov[h_2(X_1, Y_1); h_2(X_1, Y_1)]
\end{aligned} \tag{3.4.21}$$

which can be expressed as

$$\begin{aligned}
\epsilon_{10,d1} &= P[X_1 \succ Y_1 \ \& \ X_1 \succ Y_1^{'}] - P_2^2(X_1 \succ Y_1) \\
\epsilon_{01,d1} &= P[X_1 \succ Y_1 \ \& \ X_1^{'} \succ Y_1] - P_2^2(X_1 \succ Y_1) \\
\epsilon_{11,d1} &= P[X_1 \succ Y_1 \ \& \ X_1 \succ Y_1] - P_2^2(X_1 \succ Y_1)
\end{aligned} \tag{3.4.22}$$

Let $R_{jk} = h_2(X_j, Y_k)$; $R_{j.} = \sum_{k=1}^{n_2} R_{jk}$; $R_{.k} = \sum_{j=1}^{n_1} R_{jk}$. The estimates of the terms in (3.4.22) are as follows

$$\hat{P}(X_1 \succ Y_1) = \frac{1}{n_1 n_2} \sum_{j=1}^{n_1} \sum_{k=1}^{n_2} R_{jk} = U_{d1} \tag{3.4.23}$$

$$\hat{P}(X_1 \succ Y_1 \ \& \ X_1 \succ Y_1^{'}) = \frac{1}{n_1 n_2 (n_2 - 1)} \sum_{j=1}^{n_1} R_{j.}(R_{j.} - 1) \tag{3.4.24}$$

$$\hat{P}(X_1 \succ Y_1 \ \& \ X_1^{'} \succ Y_1) = \frac{1}{n_1 n_2 (n_1 - 1)} \sum_{k=1}^{n_2} R_{.k}(R_{.k} - 1) \tag{3.4.25}$$

$$\hat{P}(X_1 \succ Y_1 \ \&\ X_1 \succ Y_1) = U_{d1} \tag{3.4.26}$$

We will now determine the covariance term $\sigma_{U_{n1},U_{d1}}$. Note that $\sigma_{U_{n1},U_{d1}} = \frac{N}{n_1}\delta_{10} + \frac{N}{n_2}\delta_{01} + \frac{N}{n_1 n_2}\delta_{11}$, where

$$\begin{aligned}
\delta_{10} &= Cov[h_1(X_1, Y_1) \ \&\ h_2(X_1, Y_1^{'})] \\
\delta_{01} &= Cov[h_1(X_1, Y_1) \ \&\ h_2(X_1^{'}, Y_1)] \\
\delta_{11} &= Cov[h_1(X_1, Y_1) \ \&\ h_2(X_1, Y_1)].
\end{aligned} \tag{3.4.27}$$

The above can be expressed as

$$\begin{aligned}
\delta_{10} &= P[X_1 \succ Y_1 \ \&\ X_1 \succ Y_1^{'}] - P_1[X_1 \succ Y_1]P_2[X_1 \succ Y_1] \\
\delta_{01} &= P[X_1 \succ Y_1 \ \&\ X_1^{'} \succ Y_1] - P_1[X_1 \succ Y_1]P_2[X_1 \succ Y_1] \\
\delta_{11} &= P[X_1 \succ Y_1 \ \&\ X_1 \succ Y_1] - P_1[X_1 \succ Y_1]P_2[X_1 \succ Y_1].
\end{aligned} \tag{3.4.28}$$

Equivalently,

$$\begin{aligned}
\delta_{10} &= P[X_1 \succ Y_1 \ \&\ X_1 \succ Y_1^{'}] - U_n U_{d1} \\
\delta_{01} &= P[X_1 \succ Y_1 \ \&\ X_1^{'} \succ Y_1] - U_n U_{d1} \\
\delta_{11} &= P[X_1 \succ Y_1 \ \&\ X_1 \succ Y_1] - U_n U_{d1}
\end{aligned} \tag{3.4.29}$$

Estimates of the terms in (3.4.29) are as follows.

$$\hat{P}(X_1 \succ Y_1 \ \&\ X_1 \succ Y_1^{'}) = \frac{1}{n_1 n_2 (n_2 - 1)} \sum_{j=1}^{n_1} R_{j.}(S_{j.} - 1) \tag{3.4.30}$$

$$\hat{P}(X_1 \succ Y_1 \ \&\ X_1^{'} \succ Y_1) = \frac{1}{n_1 n_2 (n_1 - 1)} \sum_{k=1}^{n_2} R_{.k}(S_{.k} - 1) \tag{3.4.31}$$

$$\hat{P}(X_1 \succ Y_1 \ \&\ X_1 \succ Y_1) = U_n \tag{3.4.32}$$

The results for the elements of Σ_{U_2} follow similarly to those of Σ_{U_1}. Let's start with $\sigma^2_{U_{d2}}$. Note that $\sigma^2_{U_{d2}} = \frac{N}{n_1}\epsilon_{10,d2} + \frac{N}{n_2}\epsilon_{01,d2} + \frac{N}{n_1 n_2}\epsilon_{11,d2}$, where

$$\begin{aligned}
\epsilon_{10,d2} &= Cov[h_3(X_1, Y_1); h_3(X_1, Y_1^{'})] \\
\epsilon_{01,d2} &= Cov[h_3(X_1, Y_1); h_3(X_1^{'}, Y_1)] \\
\epsilon_{11,d2} &= Cov[h_3(X_1, Y_1); h_3(X_1, Y_1)]
\end{aligned} \tag{3.4.33}$$

These can be expressed as

$$\begin{aligned}
\epsilon_{10,d2} &= P[X_1 \succ Y_1 \;\&\; X_1 \succ Y_1^{'}] - P_3^2(X_1 \succ Y_1) \\
\epsilon_{01,d2} &= P[X_1 \succ Y_1 \;\&\; X_1^{'} \succ Y_1] - P_3^2(X_1 \succ Y_1) \\
\epsilon_{11,d2} &= P[X_1 \succ Y_1 \;\&\; X_1 \succ Y_1] - P_3^2(X_1 \succ Y_1)
\end{aligned} \tag{3.4.34}$$

Let $T_{jk} = h_3(X_j, Y_k)$; $T_{j.} = \sum_{k=1}^{n_2} T_{jk}$; $T_{.k} = \sum_{j=1}^{n_1} T_{jk}$. The estimates of the terms in (3.4.34) are as follows.

$$\hat{P}(X_1 \succ Y_1) = \frac{1}{n_1 n_2} \sum_{j=1}^{n_1} \sum_{k=1}^{n_2} T_{jk} = U_{d2} \tag{3.4.35}$$

$$\hat{P}(X_1 \succ Y_1 \;\&\; X_1 \succ Y_1^{'}) = \frac{1}{n_1 n_2 (n_2 - 1)} \sum_{j=1}^{n_1} T_{j.}(T_{j.} - 1) \tag{3.4.36}$$

$$\hat{P}(X_1 \succ Y_1 \;\&\; X_1^{'} \succ Y_1) = \frac{1}{n_1 n_2 (n_1 - 1)} \sum_{k=1}^{n_2} T_{.k}(T_{.k} - 1) \tag{3.4.37}$$

$$\hat{P}(X_1 \succ Y_1 \;\&\; X_1^{'} \succ Y_1) = U_{d2} \tag{3.4.38}$$

Next we note that $\sigma_{U_n,U_{d2}} = \frac{N}{n_1}\eta_{10} + \frac{N}{n_2}\eta_{01} + \frac{N}{n_1 n_2}\eta_{11}$, where

$$\begin{aligned} \eta_{10} &= Cov[h_1(X_1, Y_1) \ \& \ h_3(X_1, Y_1^{'})] \\ \eta_{01} &= Cov[h_1(X_1, Y_1) \ \& \ h_3(X_1^{'}, Y_1)] \\ \eta_{11} &= Cov[h_1(X_1, Y_1) \ \& \ h_3(X_1, Y_1)] \end{aligned} \tag{3.4.39}$$

The above can be expressed as

$$\begin{aligned} \eta_{10} &= P[X_1 \succ Y_1 \ \& \ X_1 \succ Y_1^{'}] - P_1[X_1 \succ Y_1]P_3[X_1 \succ Y_1] \\ \eta_{01} &= P[X_1 \succ Y_1 \ \& \ X_1^{'} \succ Y_1] - P_1[X_1 \succ Y_1]P_3[X_1 \succ Y_1] \\ \eta_{11} &= P[X_1 \succ Y_1 \ \& \ X_1 \succ Y_1] - P_1[X_1 \succ Y_1]P_3[X_1 \succ Y_1] \end{aligned} \tag{3.4.40}$$

Equivalently,

$$\begin{aligned} \eta_{10} &= P[X_1 \succ Y_1 \ \& \ X_1 \succ Y_1^{'}] - U_n U_{d2} \\ \eta_{01} &= P[X_1 \succ Y_1 \ \& \ X_1^{'} \succ Y_1] - U_n U_{d2} \\ \eta_{11} &= P[X_1 \succ Y_1 \ \& \ X_1 \succ Y_1] - U_n U_{d2} \end{aligned} \tag{3.4.41}$$

The estimates of terms in (3.4.41) are

$$\hat{P}(X_1 \succ Y_1 \ \& \ X_1 \succ Y_1^{'}) = \frac{1}{n_1 n_2 (n_2 - 1)} \sum_{j=1}^{n_1} T_{j.}(S_{j.} - 1) \tag{3.4.42}$$

$$\hat{P}(X_1 \succ Y_1 \ \& \ X_1^{'} \succ Y_1) = \frac{1}{n_1 n_2 (n_1 - 1)} \sum_{k=1}^{n_2} T_{.k}(S_{.k} - 1) \tag{3.4.43}$$

$$\hat{P}(X_1 \succ Y_1 \ \& \ X_1 \succ Y_1) = U_n \tag{3.4.44}$$

We point out that all estimates of the elements in the covariance matrices, Σ_{U_1} and Σ_{U_2} include three terms (for example, the terms $\epsilon_{10,n}$, $\epsilon_{01,n}$ and $\epsilon_{11,n}$ in the expression for $\sigma^2_{U_n}$). It is readily seen that the last term approaches zero asymptotically. However, the inclusion of the third term improves the accuracy for moderate samples.

An approximate 100(1-α)% lower confidence limit for CEP_1 can be obtained by using the delta method, noting the asymptotic normality of U_1. This gives

$$\sqrt{N}[U_{CEP_1} - CEP_1] \sim N_2\left[0, \left(\frac{1}{U_{d_1}}, -\frac{U_n}{U_{d_1}}\right)^T \Sigma_{U_1}\left(\frac{1}{U_{d_1}}, -\frac{U_n}{U_{d_1}}\right)\right]. \tag{3.4.45}$$

Therefore a lower $100(1-\alpha)\%$ confidence limit for CEP_1 is given by

$$U_{CEP_1} - z_{1-\alpha}\sqrt{\left(\frac{1}{U_{d_1}}, -\frac{U_n}{U_{d_1}}\right)^T \Sigma_{U_1}\left(\frac{1}{U_{d_1}}, -\frac{U_n}{U_{d_1}}\right)}, \tag{3.4.46}$$

where $z_{1-\alpha}$ is the $(1-\alpha)^{th}$ percentile of the standard normal distribution. Similarly, an approximate $100(1-\alpha)\%$ lower confidence limit for CEP_2 is given by

$$U_{CEP_2} - z_{1-\alpha}\sqrt{\left(\frac{1}{U_{d_2}}, -\frac{U_n}{U_{d_2}}\right)^T \Sigma_{U_2}\left(\frac{1}{U_{d_2}}, -\frac{U_n}{U_{d_2}}\right)}, \tag{3.4.47}$$

The following algorithm gives the steps necessary for computing the 100(1-α)% lower confidence limit for CEP_1; the algorithm is similar for CEP_2.

Algorithm 6: Lower confidence limits for CEP_1 based on U-statistics

1 For $j = 1, \ldots, n_1$ and $k = 1, \ldots, n_2$ compute:

$$S_{jk} = \begin{cases} 1 \text{ if } C_j - C_k \leq 0 \text{ \& } E_{1j} - E_{1k} \geq 0 \text{ \& } E_{2j} - E_{2k} \geq 0 \\ 0 \text{ otherwise} \end{cases}$$

$$R_{jk} = \begin{cases} 1 \text{ if } C_j - C_k \leq 0 \\ 0 \text{ otherwise} \end{cases}$$

2 Now compute the quantities in Steps 3−10:

3 $S_{j.} = \sum_{k=1}^{n_2} S_{jk}$ and $S_{.k} = \sum_{j=1}^{n_1} S_{jk}$

4 $R_{j.} = \sum_{k=1}^{n_2} R_{jk}$ and $R_{.k} = \sum_{j=1}^{n_1} R_{jk}$

5 $U_N = \frac{\sum_{j=1}^{n_1} \sum_{k=1}^{n_2} S_{jk}}{n_1 n_2}$

6 $U_{d_1} = \frac{\sum_{j=1}^{n_1} \sum_{k=1}^{n_2} R_{jk}}{n_1 n_2}$

7 $U_{CEP_1} = \frac{U_N}{U_{d_1}}$

8 $\sigma^2_{U_n} = \frac{N}{n_1} \left(\frac{\sum_{j=1}^{n_1} S_{j.}(S_{j.}-1)}{n_1 n_2 (n_2-1)} - U_n^2 \right) + \frac{N}{n_2} \left(\frac{\sum_{k=1}^{n_2} S_{.k}(S_{.k}-1)}{n_1 n_2 (n_1-1)} - U_n^2 \right) + \frac{N}{n_1 n_2} \left(U_n - U_n^2 \right)$

9 $\sigma^2_{U_{d_1}} = \frac{N}{n_1} \left(\frac{\sum_{j=1}^{n_1} R_{j.}(R_{j.}-1)}{n_1 n_2 (n_2-1)} - U_{d_1}^2 \right) + \frac{N}{n_2} \left(\frac{\sum_{k=1}^{n_2} R_{.k}(R_{.k}-1)}{n_1 n_2 (n_1-1)} - U_{d_1}^2 \right) + \frac{N}{n_1 n_2} \left(U_{d_1} - U_{d_1}^2 \right)$

10 $\sigma_{U_{n1}, U_{d1}} =$

$$\left(\frac{N}{n_1} \frac{\sum_{j=1}^{n_1} R_{j.}(S_{j.}-1)}{n_1 n_2 (n_2-1)} - U_n U_{d_1} \right) + \frac{N}{n_2} \left(\frac{\sum_{k=1}^{n_2} R_{.k}(S_{.k}-1)}{n_1 n_2 (n_1-1)} - U_n U_{d_1} \right) + \frac{N}{n_1 n_2} \left(U_n - U_n U_{d_1} \right)$$

11 Plug-in results from steps 8-10 into Σ_{U_1}: $\Sigma_{U_1} = \begin{pmatrix} \sigma^2_{U_n} & \sigma_{U_n, U_{d1}} \\ \sigma_{U_n, U_{d1}} & \sigma^2_{U_{d1}} \end{pmatrix}$

12 The $100(1-\alpha)\%$ lower confidence limit for CEP_1 is:

$$U_{CEP_1} - z_{1-\alpha} \sqrt{ \left(\frac{1}{U_{d_1}}, -\frac{U_n}{U_{d_1}} \right)^T \Sigma_{U_1} \left(\frac{1}{U_{d_1}}, -\frac{U_n}{U_{d_1}} \right) }$$

In the next section, the parametric methods developed in the previous section and non-parametric methods developed above will be applied to data obtained from a trial on irritable bowel syndrome.

3.5 An example

This example uses data from Creed et al. (2003). The authors compare the cost-effectiveness of three competing treatments for patients suffering from severe irritable bowel syndrome (IBS). The three treatments are the drug Paroxetine, psychotherapy, and a usual standard care. In our analysis we assess cost-effectiveness of the new treatment Paroxetine against the usual standard care. Several effects on participants in the study have been recorded using a survey questionnaire. Two of the effect measures collected are 'SF-36 physical component score' and 'SF-36 mental component score'; the random variables E_1 and E_2, respectively, will denote these effectiveness measures. These are the change from baseline to 15 months. The authors provide summary data shown in Table 3.1. In our analysis, the new treatment is Paroxetine and the standard treatment is usual care, referred to as treatment one and treatment two, respectively.

Table 3.1 Summary data from the IBS trial

Treatment	Paroxetine (new)	Usual Treatment (standard)
Sample size	72	61
E(E_1)	5.8	-0.3
SE(E_1)	1.00	1.17
E(E_2)	2.2	3.8
SE(E_2)	1.55	1.57
E(Cost)	1252	1663
SD(Cost)	1616	3177

Based on the summary results, we note that the average effectiveness measure corresponding to E_1 (i.e., changes from baseline in the SF-36 physical component score) appears to be higher for patients taking Paroxetine compared to patients under the standard care. However, the average effectiveness measure for E_2 (i.e., changes from baseline in the SF-36 mental component score) appears to be higher for patients under the standard care compared to that for patients taking Paroxetine.

Earlier, we mentioned the weighting scheme often used in the MCDA framework, and noted that the method is subjective and may result in different conclusions on the cost-effectiveness of competing treatments. We shall demonstrate this for the present example. The weights are assigned to the mean cost and the two mean effectiveness measures based on scores assigned by experts. A set of scores and the resulting weights are given in Table 3.2.

Table 3.2 A set of performance scores and weights

Treatment \Criteria	$E(Cost)$	$E(E_1)$	$E(E_2)$
Paroxetine	1	1	0
Standard	0	0	2
Total	1	1	2
Weight	1/4	1/4	2/4

The scores in Table 3.2 are based on expert opinion as follows: $E(E_1) < 0$ receives a score of 0, $E(E_1) > 0$ receives 1; $E(E_2) < 2.5$ receives 0, $2.5 < E(E_2) < 3$ receives 1, $E(E_2) > 3$ receives 2; $E(Cost) > 1550$ receives 0 , and $E(Cost) < 1550$ receives 1. The weights in Table 3.2 are used to obtain estimates of the values V_1 and V_2 for treatment one and treatment two, respectively; see (3.1.1).

Consider another expert's opinion that uses the same scores as in Table 3.2 for E(Cost) and $E(E_1)$ and a different scoring scheme for $E(E_2)$. The scoring based on this expert is: $E(E_2) < 3$ receives 0, $E(E_2) > 3$ receives 1. This gives the scores and weights in Table 3.3.

Table 3.3 A second set of performance scores and weights

Treatment \Criteria	$E(Cost)$	$E(E_1)$	$E(E_2)$
Paroxetine	1	1	0
Standard	0	0	1
Total	1	1	1
Weight	1/3	1/3	1/3

The estimated values for each treatment based on the two different weighting combinations (Tables 3.2 and 3.3), are provided in Table 3.4.

Table 3.4 Overall performance scores based on two weighting combinations

Weighting Scheme \Treatment Score	$\widehat{V}_1$	$\widehat{V}_2$	Preferred Treatment
1	1/2	1	Standard
2	2/3	1/3	Paroxetine

Table 3.4 reports the overall values for the two treatments based on the two sets of expert scores. The scores defined by each only differ with regards to $E(E_2)$. Even this slight difference in scoring results in contradictory conclusions regarding treatment recommendation. The first weighting scheme indicates that the standard treatment should be preferred because $\widehat{V}_2 > \widehat{V}_1$. In contrast, the scoring by the second expert concludes that $\widehat{V}_1 > \widehat{V}_2$, indicating that Paroxetine treatment is preferred. Now we turn to our probabilistic criteria.

In Creed et al. (2003, pg. 305) the authors note that the cost data are "highly skewed". However, we don't have the original data to obtain estimates of the lognormal parameters, and we approximate these based on the summary statistics in Table 3.1. The mean and standard deviation of the log(cost) for treatment one is computed to be 6.642 and 0.990, respectively. The mean and standard deviation of the log(cost) for treatment two is similarly obtained as 6.648 and 1.239, respectively. In addition, we also need the estimated correlations among log(cost), E_1 and E_2. For this, we simulated data using the model (3.2.1) with correlation values supplemented by those provided in Table 3.5. We have considered eighteen correlation combinations for each treatment group, which are denoted by 'cases' in Table 3.5. The means and standard deviations used to simulate the data are the values mentioned earlier in this paragraph. Estimates for CEP_1 and CEP_2 are calculated for each case and also reported in the table.

Table 3.5 Correlations and the corresponding estimated CEPs based on simulated data under the model (3.2.1)

Case ID	ρ_{e_1,e_2}	$\rho_{e_1,c}$	$\rho_{e_2,c}$	$\widehat{CEP_1}$	$\widehat{CEP_2}$
1	0.100	0.100	0.100	0.299	0.450
2	−0.500	0.100	0.100	0.208	0.424
3	0.100	0.100	−0.100	0.343	0.515
4	−0.500	0.100	−0.100	0.252	0.513
5	0.500	0.100	−0.100	0.404	0.518
6	0.100	−0.100	0.100	0.324	0.488
7	−0.500	−0.100	0.100	0.240	0.490
8	0.100	−0.100	−0.100	0.368	0.553
9	−0.500	−0.100	−0.100	0.284	0.579
10	0.500	−0.100	−0.100	0.424	0.543
11	0.100	−0.500	−0.500	0.514	0.773
12	0.100	−0.500	0.500	0.278	0.418
13	−0.100	−0.500	−0.500	0.492	0.808
14	−0.100	−0.500	0.500	0.259	0.425
15	−0.100	0.500	−0.500	0.352	0.578
16	0.500	−0.500	−0.500	0.563	0.722
17	−0.500	0.500	−0.500	0.279	0.569
18	−0.500	−0.500	0.500	0.213	0.434

Based on the results of Table 3.5, we note that the different combinations of the correlations ρ_{e_1,e_2} , $\rho_{e_1,c}$, and $\rho_{e_2,c}$ result in different cost-effectiveness probabilities: CEP_1 ranges from 0.208 to 0.563, and CEP_2 ranges from 0.418 to 0.808. The new treatment, Paroxetine, is deemed cost-effective when the probabilities can be considered 'large'. We recall that the probability CEP_1 is of interest to decision makers who prioritize cost. Based on our results, we note that among patients who incur less cost when taking the new treatment Paroxetine, over the standard treatment, the probability of improved effectiveness with respect to both E_1 and E_2 is mostly less than 0.5, often by a significant margin. On the other hand, among patients form whom the new treatment Paroxetine is more effective, the probability of incurring a lower cost is mostly more than 0.5. The probabilities clearly depend on the correlations.

The effect of the correlations on the estimated CEP_1 and CEP_2 values is not clear cut. However, both metrics result in largest values under the cases 11, 13, and 16 in Table 3.5. These cases all have $\rho_{e_1,c}$ = -0.5 and $\rho_{e_2,c}$ = -0.5. This indicates that, when these correlations are negative and large in magnitude, cost-effectiveness is higher for such patients. Comparing the corresponding estimates (fixing the case ID) of CEP_1 and CEP_2 in Table 3.5, the estimates of CEP_2 are considerably larger.

3.5.1 Parametric confidence intervals

We now present the results of the fiducial approach for the interval estimation of CEP_1 and CEP_2. The 95% lower limits are obtained using algorithm 4 based on 5000 fiducial quantities. The confidence limits are reported in Table 3.6 for one set of simulated data using the parameter combinations (including the correlations) mentioned earlier. In addition, we also estimated the coverage probabilities of our confidence limits using 5000 simulated samples, and 1000 fiducial quantities. The coverage probabilities and the expected values of the confidence limits are also reported in Table 3.6.

Table 3.6 Fiducial 95% lower confidence limits with corresponding coverage probabilities and expected lower limits

Case ID	Lower conf. limit CEP_1	Lower conf. limit CEP_2	Coverage prob. CEP_1	Coverage prob. CEP_2	Expected limit CEP_1	Expected limit CEP_2
1	0.224	0.350	0.964	0.958	0.222	0.349
2	0.150	0.323	0.968	0.953	0.151	0.324
3	0.263	0.412	0.961	0.959	0.263	0.413
4	0.189	0.413	0.968	0.955	0.189	0.409
5	0.315	0.419	0.959	0.955	0.316	0.418
6	0.243	0.388	0.958	0.950	0.246	0.387
7	0.179	0.388	0.965	0.957	0.179	0.387
8	0.284	0.456	0.958	0.958	0.285	0.451
9	0.217	0.479	0.971	0.956	0.216	0.475
10	0.335	0.445	0.965	0.957	0.333	0.444
11	0.398	0.664	0.963	0.950	0.399	0.659
12	0.193	0.289	0.959	0.953	0.193	0.291
13	0.377	0.698	0.958	0.956	0.376	0.693
14	0.179	0.295	0.958	0.955	0.180	0.299
15	0.268	0.444	0.972	0.956	0.267	0.440
16	0.450	0.614	0.960	0.948	0.449	0.611
17	0.209	0.435	0.974	0.955	0.209	0.434
18	0.148	0.307	0.962	0.953	0.147	0.307

Comparing the coverage probabilities of Table 3.6, we note the coverage tends to be conservative for CEP_1, while those for CEP_2 tend to be closer to nominal. For each case ID (combination of correlation values) the coverage is larger for CEP_1 than for CEP_2, with exception of case ID 10. Overall, the coverage probabilities appear to be satisfactory as they are close to nominal or slightly larger than nominal.

An alternative to the fiducial methodology for obtaining lower parametric confidence limits is to implement the percentile bootstrap method. The corresponding results are shown in Table 3.7.

Table 3.7 Percentile bootstrap 95% lower confidence limits with corresponding coverage probabilities and expected limits

Case ID	Lower conf. limit CEP_1	Lower conf. limit CEP_2	Coverage prob. CEP_1	Coverage prob. CEP_2	Expected limit CEP_1	Expected limit CEP_2
1	0.227	0.351	0.958	0.949	0.224	0.351
2	0.151	0.325	0.960	0.944	0.152	0.325
3	0.267	0.417	0.953	0.951	0.264	0.412
4	0.192	0.411	0.949	0.950	0.192	0.412
5	0.316	0.418	0.953	0.947	0.317	0.420
6	0.246	0.383	0.959	0.947	0.247	0.387
7	0.182	0.390	0.960	0.953	0.180	0.387
8	0.283	0.451	0.952	0.948	0.287	0.452
9	0.218	0.477	0.955	0.956	0.219	0.475
10	0.337	0.445	0.953	0.955	0.335	0.444
11	0.397	0.663	0.957	0.946	0.399	0.657
12	0.194	0.289	0.950	0.945	0.194	0.292
13	0.375	0.691	0.956	0.956	0.376	0.689
14	0.180	0.297	0.957	0.948	0.180	0.298
15	0.272	0.442	0.962	0.948	0.269	0.442
16	0.443	0.611	0.955	0.954	0.450	0.608
17	0.214	0.437	0.961	0.952	0.212	0.436
18	0.148	0.306	0.956	0.946	0.148	0.307

The lower limits using the percentile bootstrap method for the eighteen cases in Table 3.5 are provided in Table 3.7 and appear to be similar to those obtained by the fiducial method for CEP_1. The coverage probabilities are obtained using 5000 simulated samples with 1000 parametric bootstrap estimates under each simulated sample. Both the fiducial and percentile bootstrap methods tend to have larger coverage for CEP_1 compared to CEP_2. The percentile bootstrap method for CEP_1 is also conservative. The coverage probabilities from the two methods are comparable for CEP_2, and are less conservative under both methods. Comparing the coverage probabilities of the fiducial and percentile bootstrap methodologies reveals that the fiducial approach tends to be more conservative. The expected lower limits for the fiducial and percentile bootstrap methods are comparable for cases in which the two methods have similar coverage probabilities.

For comparing the two approaches, we also calculated 95% two-sided confidence limits, even though this won't be of practical interest in the context of our probability metrics. The results for the two-sided interval are provided in Table 3.8 for the fiducial methodology, and in Table 3.9 for the percentile bootstrap method.

Table 3.8 Fiducial 95% two-sided confidence limits with corresponding coverage probabilities and expected lengths

Case ID	Two-sided conf. limit CEP_1	Two-sided conf. limit CEP_2	Cov. prob. CEP_1	Cov.prob. CEP_2	Expected Length CEP_1	Expected Length CEP_2
1	(0.210, 0.392)	(0.330, 0.568)	0.956	0.956	0.183	0.239
2	(0.139, 0.280)	(0.303, 0.543)	0.959	0.955	0.138	0.241
3	(0.250, 0.440)	(0.394, 0.635)	0.957	0.960	0.191	0.241
4	(0.179, 0.328)	(0.395, 0.636)	0.954	0.954	0.149	0.244
5	(0.299, 0.515)	(0.400, 0.637)	0.957	0.957	0.209	0.238
6	(0.230, 0.422)	(0.369, 0.608)	0.955	0.954	0.189	0.241
7	(0.168, 0.314)	(0.369, 0.610)	0.958	0.958	0.148	0.243
8	(0.267, 0.471)	(0.437, 0.669)	0.955	0.954	0.199	0.239
9	(0.206, 0.368)	(0.461, 0.697)	0.959	0.956	0.162	0.241
10	(0.319, 0.534)	(0.424, 0.660)	0.958	0.953	0.213	0.235
11	(0.379, 0.654)	(0.637, 0.885)	0.955	0.952	0.271	0.242
12	(0.177, 0.383)	(0.266, 0.589)	0.953	0.953	0.204	0.317
13	(0.355, 0.635)	(0.675, 0.914)	0.953	0.957	0.274	0.238
14	(0.164, 0.357)	(0.265, 0.587)	0.953	0.960	0.190	0.315
15	(0.254, 0.440)	(0.423, 0.724)	0.953	0.955	0.190	0.314
16	(0.430, 0.696)	(0.592, 0.839)	0.953	0.953	0.264	0.247
17	(0.198, 0.357)	(0.413, 0.718)	0.953	0.954	0.159	0.312
18	(0.136, 0.294)	(0.282, 0.593)	0.949	0.955	0.156	0.312

Table 3.9 Percentile bootstrap 95% two-sided confidence limits with corresponding coverage probabilities and expected lengths

Case ID	Two-sided conf. limit CEP_1	Two-sided conf. limit CEP_2	Cov. prob. CEP_1	Cov. prob. CEP_2	Expected Length CEP_1	Expected Length CEP_2
1	(0.210, 0.392)	(0.334, 0.565)	0.949	0.949	0.183	0.237
2	(0.141, 0.279)	(0.307, 0.549)	0.949	0.951	0.138	0.239
3	(0.252, 0.439)	(0.398, 0.634)	0.946	0.946	0.190	0.239
4	(0.182, 0.331)	(0.391, 0.637)	0.946	0.951	0.148	0.241
5	(0.302, 0.507)	(0.398, 0.633)	0.947	0.945	0.208	0.236
6	(0.231, 0.421)	(0.368, 0.611)	0.949	0.943	0.188	0.239
7	(0.172, 0.319)	(0.373, 0.608)	0.944	0.951	0.147	0.241
8	(0.269, 0.468)	(0.433, 0.673)	0.941	0.952	0.198	0.237
9	(0.207, 0.370)	(0.455, 0.694)	0.948	0.950	0.161	0.239
10	(0.316, 0.531)	(0.427, 0.661)	0.948	0.951	0.213	0.233
11	(0.376, 0.649)	(0.641, 0.882)	0.942	0.946	0.270	0.243
12	(0.178, 0.387)	(0.261, 0.588)	0.946	0.946	0.204	0.315
13	(0.355, 0.633)	(0.666, 0.911)	0.946	0.951	0.273	0.239
14	(0.165, 0.360)	(0.274, 0.594)	0.951	0.942	0.190	0.313
15	(0.258, 0.445)	(0.414, 0.737)	0.952	0.949	0.190	0.312
16	(0.422, 0.696)	(0.591, 0.838)	0.949	0.953	0.264	0.247
17	(0.202, 0.360)	(0.413, 0.726)	0.946	0.948	0.158	0.310
18	(0.136, 0.295)	(0.281, 0.590)	0.949	0.945	0.156	0.311

The two-sided coverage probabilities corresponding to both the methods appear satisfactory, as most are near nominal 0.95. Based on the numerical results, we note that there is no substantial difference between the two approaches.

We also utilized the delta method for the interval estimation of CEP_1 and CEP_2. We implemented the delta method with the original sample sizes as well as the larger sample sizes $n_1 = n_2 = 500$ and $n_1 = n_2 = 1000$. Even under these larger sample sizes the expected lower limits for CEP_1 and CEP_2 were negative. The corresponding coverage probabilities under these various sample sizes resulted in coverage probabilities equal to one. However, when the sample sizes were increased the lower limits did increase as one would expect. Our conclusion based on these results is that the delta method is too conservative to be useful in our context. The details of the delta method are outlined in the Appendix to this chapter.

3.5.2 Coverage probabilities of U-statistics based confidence intervals

Now, we present estimated coverage probabilities of the U-statistics based confidence intervals corresponding to the parameter combinations used in the previous subsection. However, two sample size scenarios will be considered: the sample sizes corresponding to the example, i.e., $n_1 = 72$, $n_2 = 61$ and the somewhat larger sample size combination $n_1 = n_2 = 250$. The coverage probabilities are estimated using 5000 simulations. For $n_1 = 72$, $n_2 = 61$, estimated coverage probabilities of the 95% lower confidence limits are provided in Table 3.10. In the same set up, estimated coverage probabilities of the 95% two-sided confidence intervals are provided in Table 3.11

Table 3.10 Estimated coverage probabilities and expected values of the U-statistics based 95% lower confidence limits when $n_1 = 72$, $n_2 = 61$

Case ID	Coverage prob. CEP_1	Coverage prob. CEP_2	Expected limit CEP_1	Expected limit CEP_2
1	0.970	0.961	0.215	0.339
2	0.974	0.963	0.142	0.310
3	0.964	0.959	0.253	0.402
4	0.972	0.957	0.180	0.397
5	0.962	0.952	0.307	0.410
6	0.966	0.954	0.236	0.376
7	0.968	0.958	0.168	0.372
8	0.965	0.953	0.276	0.440
9	0.970	0.955	0.207	0.463
10	0.957	0.954	0.327	0.436
11	0.957	0.937	0.423	0.689
12	0.970	0.958	0.196	0.316
13	0.958	0.940	0.404	0.730
14	0.972	0.963	0.178	0.318
15	0.968	0.953	0.264	0.471
16	0.960	0.940	0.467	0.633
17	0.968	0.954	0.203	0.456
18	0.974	0.958	0.144	0.322

Based on the results in the above table, we note that the confidence limits are conservative, especially for CEP_1. The results also indicate that the correlation has some affect on the accuracy of the results, however no clear trends are evident.

Table 3.11 Estimated coverage probabilities and expected lengths of the U-statistics based 95% two-sided confidence intervals when $n_1 = 72$, $n_2 = 61$

Case ID	Coverage prob. CEP_1	Coverage prob. CEP_2	Expected length CEP_1	Expected length CEP_2
1	0.953	0.954	0.204	0.266
2	0.955	0.960	0.159	0.276
3	0.955	0.953	0.214	0.267
4	0.959	0.959	0.174	0.281
5	0.961	0.952	0.230	0.259
6	0.957	0.952	0.211	0.268
7	0.954	0.957	0.174	0.282
8	0.958	0.955	0.219	0.266
9	0.957	0.959	0.184	0.276
10	0.956	0.954	0.232	0.258
11	0.959	0.955	0.216	0.197
12	0.959	0.958	0.201	0.247
13	0.957	0.956	0.208	0.180
14	0.954	0.957	0.192	0.256
15	0.956	0.956	0.211	0.255
16	0.961	0.955	0.225	0.214
17	0.957	0.957	0.183	0.271
18	0.952	0.958	0.167	0.271

The coverage probabilities tend to be closer to nominal for the two-sided confidence limits compared to the one-sided limits.

Table 3.12 Estimated coverage probabilities and expected values of the U-statistics based 95% lower confidence limits when $n_1 = n_2 = 250$

Case ID	Coverage prob. CEP_1	Coverage prob. CEP_2	Expected limit CEP_1	Expected limit CEP_2
1	0.961	0.953	0.258	0.396
2	0.964	0.955	0.176	0.369
3	0.956	0.952	0.299	0.462
4	0.960	0.955	0.217	0.457
5	0.955	0.949	0.357	0.466
6	0.961	0.956	0.281	0.433
7	0.961	0.952	0.206	0.434
8	0.957	0.950	0.323	0.499
9	0.959	0.947	0.247	0.524
10	0.953	0.953	0.376	0.491
11	0.953	0.941	0.470	0.734
12	0.965	0.957	0.237	0.369
13	0.953	0.943	0.450	0.773
14	0.959	0.959	0.220	0.375
15	0.960	0.952	0.308	0.527
16	0.950	0.949	0.517	0.678
17	0.962	0.948	0.243	0.516
18	0.964	0.957	0.179	0.380

The results of Table 3.12 indicate that an increase in sample size provides coverage probabilities that are closer to nominal compared to using smaller sample sizes. However, there is still some level of conservatism.

Table 3.13 Estimated coverage probabilities and expected lengths of the U-statistics based 95% two-sided confidence intervals when $n_1 = n_2 = 250$

Case ID	Coverage prob. CEP_1	Coverage prob. CEP_2	Expected length CEP_1	Expected length CEP_2
1	0.954	0.952	0.099	0.129
2	0.951	0.952	0.076	0.131
3	0.951	0.956	0.104	0.130
4	0.956	0.956	0.084	0.134
5	0.956	0.952	0.113	0.126
6	0.948	0.951	0.103	0.130
7	0.952	0.954	0.083	0.134
8	0.953	0.949	0.107	0.128
9	0.950	0.957	0.088	0.131
10	0.951	0.955	0.114	0.125
11	0.949	0.950	0.104	0.093
12	0.953	0.954	0.098	0.119
13	0.950	0.955	0.100	0.083
14	0.954	0.958	0.093	0.122
15	0.952	0.955	0.103	0.122
16	0.951	0.953	0.109	0.102
17	0.952	0.953	0.087	0.129
18	0.953	0.953	0.080	0.128

These results also indicate that the level of conservatism of the coverage probabilities has decreased and the results are closer to nominal coverage with an increase in sample sizes, as expected. We also note that the increase in the sample size also improved the expected value of the lower confidence limit, and the expected width of the two-sided interval, once again, as expected. We also estimated the coverage probabilities when $n_1 = n_2 = 500$, and further improvement was noted in the coverage probabilities; these results are not reported here.

In regards to inclusion of the third term for the covariance and variance elements of the multivariate U-statistics, simulation results obtained by not including this term indicate that the inclusion of the third term improves the coverage probability. These simulation results are not reported here.

3.6 Discussion and conclusions

The literature on multi-decision criteria analysis (MCDA) has highlighted the importance and the complexities of making inferences regarding the effectiveness of treatments under the presence of multiple effectiveness measures. Previous MCDA metrics have been based on weighting schemes that combine outcome measures into a scalar quantity. The varying preferences of various decision makers, including patients and health care policy-makers, have sparked a debate among experts regarding the appropriate effectiveness measures and the weights to be used. The subjectivity of the approach is clearly cause for concern, especially since conclusions can be reversed simply by a slight change in the weighting scheme.

We have proposed two metrics to meet the preferences of decision makers who prioritize effectiveness, and those who prioritize cost. Our metrics are probabilistic and are straightforward to interpret. We believe that our metrics are objective, in that they avoid the use of any weights. Even though we have investigated the case of only two effectiveness measures, our metrics can be extended to include more than two effectiveness measures. It is also possible to incorporate thresholds such as δ_C for the cost, and similar thresholds for the effectiveness measures. In short, our metrics appear to be versatile and objective.

A fiducial approach has been utilized to construct accurate lower confidence limits for each proposed metric, assuming the normal distribution (perhaps after an appropriate transformation). The percentile bootstrap method has also been implemented in the parametric framework. Both approaches produce coverage probabilities close to nominal even under relatively small sample sizes, such as those corresponding to the example in this chapter. The proposed probabilistic metrics can be easily extended for individualized inference under the regression framework of Chapter 2. A non-parametric U-statistics based approach is also developed in this chapter. Overall, such an approach provided satisfactory confidence limits, as evidenced by the estimated coverage probabilities.

3.7 Appendix: The delta method

The delta method used in this chapter is very similar to that explained in Appendix A of Chapter 2.

For the application in this chapter, we have $n = n_1 + n_2$, $\beta = [vec(\mu_1)', vec((\Sigma_1)', vec(\mu_2)', vec((\Sigma_2)']'$, and $\hat{\beta}$ is the unbiased estimator of β. Furthermore, the matrix Σ^* is defined as:

$$\Sigma^* = diag\left(\frac{\Sigma_1}{n_1}, \frac{\Omega_1}{n_1 - 1}, \frac{\Sigma_2}{n_2}, \frac{\Omega_2}{n_2 - 1}\right), \tag{3.7.1}$$

where the matrices Ω_1 and Ω_2 will now be defined. For this, let K_{33} be the commutation matrix defined as:

$$K_{33} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \tag{3.7.2}$$

Then $\Omega_i = \left[\Sigma_i \otimes \Sigma_i\right][K_{33} + I_9]$. Since CEP_1 and CEP_2 are functions of β, the delta method can be applied to conclude the asymptotic normality of the estimators $\widehat{CEP}_1$ and $\widehat{CEP}_2$, where these estimators are obtained by replacing β with $\hat{\beta}$. However, the gradients required to compute the asymptotic variance of these estimators have no analytic expressions. In view of this, we calculated the gradients numerically by using the `gradient` function in `R`.

Chapter 4

Individualized net monetary benefit: A tolerance limits approach

The net-benefit framework for cost-effectiveness analysis was first proposed by Stinnett and Mullahy (1998), and has gained some popularity since then. The 'net monetary benefit' (NMB) for a particular patient is defined as the difference between the patient's effectiveness and cost, after multiplying the former with the willingness-to-pay parameter λ. For the j^{th} patient in the i^{th} group the net monetary benefit is defined in (1.2.6). To account for covariates, the associated literature has incorporated patient characteristics through a regression analysis carried out directly on the NMB random variable defined in (1.2.6). This is the approach adopted in the works Hoch, Briggs and Willan (2002), Hoch and Dewa (2014), and Hounton and Newlands (2012). In Hoch, Briggs and Willan (2002) the authors propose three net-benefit regression approaches. The first includes a treatment indicator only, whereas, the second and third utilize a treatment indicator and incorporate patient-level covariates. The third approach incorporates an interaction of patient-level covariates and the treatment indicator. Later we shall comment on this approach of carrying out a regression analysis directly on the NMB random variable. In their work, the above authors utilize the cost-effectiveness acceptability curve (CEAC) to assess a new treatment's cost-effectiveness.

In fact the inference related to the NMB has been dominated by the use of the CEAC. The CEAC is a common method for illustrating the value of competing interventions. It shows the probability that a new treatment is cost-effective relative to a competing treatment, for a range of willingness-to-pay values. Briggs, O'Brien and Blackhouse (2002) note that the use of the CEAC in the net-benefit framework is useful when performing stratified analysis. Hounton and Newlands (2012) also highlight that use of the net-benefit framework is advantageous because it allows for subgroup analysis and enables utilizing the CEACs as a graphical approach towards this aim. Our approach differs from the typical NMB analysis in two ways: (i) we do not use a stratified approach; and (ii) we do not rely on graphical methods such as the CEAC for inference.

Difficulties associated with the stratification approach are already noted in Chapter 2. Secondly, reliance on the CEAC for determination of treatment allocation has been under scrutiny in recent literature. In an article by Barton, Briggs and Fenwick (2008), the authors note a common misinterpretation often associated with analysis of the CEAC:

> With regard to the claim that the CEAC shows the probability that one option dominates another option, it can be seen that this is untrue because the CEAC is determined by the net benefit of each option and it is possible for an option to have a higher net benefit without dominating another option. (p. 888)

Furthermore, in Koerkamp (2009, ch. 5), the author devotes an entire chapter to the limitations of the CEAC. One major criticism of the CEAC is that since it plots the probability of cost-effectiveness directly, it does not adequately account for precision of estimates. An alternative approach suggested by the author involves using confidence intervals to provide the decision maker with information about the precision and magnitude of an estimate.

In our research, we model the cost and effectiveness outcomes directly using a bivariate regression model, as done earlier in the thesis. The NMB for an individual patient under a specified treatment can then be computed at the specified covariate value. Because treatment allocation for a specified patient is of interest, we compare the NMBs for this patient under the new and standard treatment. The new treatment is determined to be cost-effective if the resulting NMB is 'larger'. Since the NMB is a random variable, an approach that can be adopted for the NMB comparison is to use specific scalar quantities associated with the NMB distributions. For example, a comparison of the means of the NMB distributions leads to the INB parameter. In this chapter we shall explore another comparison: the lower tail percentiles and their lower confidence limits. Such lower confidence limits are referred to as lower tolerance limits of the distribution. We shall thus address a comparison of the lower tolerance limits of the NMB distributions for patients having the same covariate values under the new treatment and under the standard treatment. The choice of the percentile is at the discretion of the policy-maker. If the lower tolerance limit associated with the new treatment is larger than that of the standard treatment then the new treatment is considered cost-effective.

Let $NMB_1(\mathbf{w}_0)$ and $NMB_2(\mathbf{w}_0)$ denote the net monetary benefit of a patient having covariate vector $\mathbf{w}_0$ under the new treatment and standard treatment, respectively. A model that has been predominately used to incorporate covariates into the NMB has been proposed by Hoch, Briggs and Willan (2002). The model is given by

$$NMB_i(w_0) = \alpha + \beta_i'\mathbf{w}_0 + \delta k_j + \epsilon_j, \tag{4.0.1}$$

In the model (4.0.1), α is an intercept term, β_i is a $p\times$ 1 vector of regression coefficients, $\mathbf{w}_0$ is a $p \times 1$ vector of covariates, and ϵ_j is an error term. Furthermore, k_j is an indicator function for the treatment ($k_j = 1$ corresponds to the j^{th} patient allocated to the new treatment, i.e., $i = 1$, and $k_j = 0$, otherwise). Note that in a lognormal-normal scenario, the NMB is a linear combination of a normally distributed quantity (namely, the effectiveness) and a lognormally distributed quantity (namely, the cost). Formulating a linear regression model for such a random variable, especially, assuming a normal distribution for the error term ϵ_j, appears to be difficult to justify. The same can be said in the lognormal-lognormal scenario also. We are of the opinion that when covariates influence the cost and effectiveness of individual patients, they should be incorporated into the analysis via a bivariate regression model for the (cost,effectiveness) random variable, and then the various criteria should be extended to such a regression scenario. This is what we propose to do this chapter, as was done earlier in the thesis. We propose a comparison of $NMB_1(\mathbf{w}_0)$ and $NMB_2(\mathbf{w}_0)$ based on lower tolerance limits, assuming a regression model for the cost and effectiveness data, after a log transformation if necessary.

4.1 The computation of lower tolerance limits

Let C_{ij} and E_{ij}, respectively, denote the cost and effectiveness for the j^{th} individual belonging to the i^{th} treatment intervention ($i = 1, 2; j = 1, \ldots, n_i$). We shall use the lognormal-lognormal model in (2.1.1). The unknown parameters are the regression coefficient matrices B_i and the covariance matrices Σ_i, $i = 1,\ 2$. Let C_i and E_i be the cost and effectiveness for a patient having covariate vector $\mathbf{w}_0$ and belonging to the i^{th} treatment group. Under the lognormal-lognormal model, we then have the distribution

$$\begin{pmatrix} ln[C_i] \\ ln[E_i] \end{pmatrix} \sim N_2 \left[B_i\mathbf{w}_0, \quad \Sigma_i \right],$$

$i = 1, 2$. Thus the NMB for a patient from the i^{th} treatment group and having covariate $\mathbf{w}_0$, denoted by $NMB_i(\mathbf{w}_0)$, is given by

$$NMB_i(\mathbf{w}_0) = \lambda E_i - C_i. \qquad (4.1.1)$$

We recall that our aim is to compute lower tolerance limits for the distributions of $NMB_i(\mathbf{w}_0)$, $i = 1,\ 2$. That is, we wish to compute lower confidence limits for the left-tail percentiles of these distributions. Thus suppose we want to compute a $100(1-\alpha)\%$ lower confidence limit for the $(1-p)^{th}$ percentile of $NMB_i(\mathbf{w}_0)$ (typically, $1-p$ is 0.10, 0.05, 0.01). Recall the lower tolerance limit is defined by content, p, and confidence level, $1-\alpha$. In particular, we will be computing the 95% lower confidence limit for the 10^{th} percentile of $NMB_i(\mathbf{w}_0)$ which is equivalent to a lower ($p = 0.9, 1-\alpha = 0.95$) tolerance limit. Note that the percentiles do not have an analytical expression. However, both the fiducial approach and a parametric bootstrap approach can be applied for a numerical result. We shall now explain this computation.

Under the assumed model, let $\hat{B}_i$ and $\hat{\Sigma}_i$ denote the unbiased estimators of B_i and Σ_i, respectively, based on samples of n_1 patients and n_2 patients assigned to the first and second treatments, respectively. Furthermore, let $\tilde{B}_i$ and $\tilde{\Sigma}_i$ denote the corresponding fiducial quantities. The expressions for these quantities are given in Chapter 2, and are not repeated here. A fiducial quantity for the $(1-p)^{th}$ percentile of $NMB_i(\mathbf{w}_0)$ can be implicitly computed as follows. Generate data under the assumed bivariate regression model corresponding to the covariate vector $\mathbf{w}_0$, with B_i and Σ_i replaced by the corresponding fiducial quantities $\tilde{B}_i$ and $\tilde{\Sigma}_i$, respectively. Based on the data generated, compute the value of $NMB_i(\mathbf{w}_0)$, say $\widetilde{NMB}_i(\mathbf{w}_0)$. Repeat these steps C_1 times, resulting in C_1 values of $\widetilde{NMB}_i(\mathbf{w}_0)$, say $\widetilde{NMB}_{ik}(\mathbf{w}_0)$, $k = 1, 2, ..., C_1$. The $(1-p)^{th}$ percentile of these C_1 values is a fiducial quantity for the $(1-p)^{th}$ percentile of $NMB_i(\mathbf{w}_0)$. Repeating this procedure, we generate the fiducial quantity for the $(1-p)^{th}$ percentile of $NMB_i(\mathbf{w}_0)$ C_2 times. The α^{th} percentile of these C_2 fiducial quantities give a $100(1-\alpha)\%$ lower confidence limit for the $(1-p)^{th}$ percentile of $NMB_i(\mathbf{w}_0)$, which is the required lower tolerance limit.

The above computational steps are summarized in algorithm 7.

Algorithm 7: Fiducial approach for the construction of a lower tolerance limit for $NMB_i(\mathbf{w}_0)$

1 From the sample of each treatment group compute the observed values $\hat{B}_{io}$ and $\hat{\Sigma}_{io}$ of the estimates $\hat{B}_i$ and $\hat{\Sigma}_i$, $i = 1, 2$.

2 For $i = 1, 2$, independently generate: $H_i \sim W\left(\left\{(n_i - p)\hat{\Sigma}_{io}\right\}^{-1}, n_i - p\right)$ and $Z_i \sim 2 \times p$ matrix of $N(0,1)$ random variates.

3 Compute $\tilde{\Sigma}_i = H_i^{-1}$ and $\tilde{B}_i = \hat{B}_{io} - \tilde{\Sigma}_i^{1/2} Z_i (W_i W_i')^{-1/2}$.

4 For each $i = 1, 2$ generate C_1 samples:

$$\begin{pmatrix} ln[\tilde{C}_{ik}] \\ ln[\tilde{E}_{ik}] \end{pmatrix} \sim N_2\left[\tilde{B}_i \mathbf{w}_0, \quad \tilde{\Sigma}_i\right], k = 1, \ldots C_1$$

5 For each combination of i and k compute $\widetilde{NMB}_{ik}(\mathbf{w}_0) = \lambda\tilde{E}_{ik} - \tilde{C}_{ik}$

6 Compute the $(1-p)^{th}$ percentile of the C_1 values of $\widetilde{NMB}_{ik}(\mathbf{w}_0)$ computed in step 5. Denote this as $\widetilde{NMB}_{ik}(\mathbf{w}_0; 1-p)$

7 Repeat steps 2-6 C_2 times obtaining C_2 values of $\widetilde{NMB}_{ik}(\mathbf{w}_0; 1-p)$.

8 Compute the α^{th} percentile of the C_2 values of $\widetilde{NMB}_{ik}(\mathbf{w}_0; 1-p)$. This is a $100(1-\alpha)\%$ lower tolerance limit for $NMB_i(\mathbf{w}_0)$ having content p.

In addition to the fiducial approach outlined in algorithm 7, we also implemented a percentile bootstrap method. The necessary steps for the percentile bootstrap approach are provided in algorithm 8.

Algorithm 8: Percentile Bootstrap approach for the construction of a lower tolerance limit for $NMB_i(\mathbf{w}_0)$

1 From the sample of each treatment group compute the observed values $\hat{B}_{io}$ and $\hat{\Sigma}_{io}$ of the estimates $\hat{B}_i$ and $\hat{\Sigma}_i$, $i = 1, 2$.

2 For $i = 1, 2$, independently generate: $vec(\hat{B}_i^{\,*}) \sim N\left[vec(\hat{B}_{io}), (W_iW_i')^{-1} \otimes \hat{\Sigma}_{io}\right]$ and $\hat{\Sigma}_i^* \sim \frac{1}{n_i - p} W[\hat{\Sigma}_{io}, n_i - p]$ random variates.

3 For each $i = 1, 2$ generate C_1 samples:

$$\begin{pmatrix} ln[C_{ik}^*] \\ \\ ln[E_{ik}^*] \end{pmatrix} \sim N_2 \left[\hat{B}_i^* \mathbf{w}_0, \quad \hat{\Sigma}_i^*\right], \, k = 1, \ldots C_1$$

4 For each combination of i and k compute $\widehat{NMB}_{ik}^{*}(\mathbf{w}_0) = \lambda E_{ik}^* - C_{ik}^*$

5 Compute the $(1-p)^{th}$ percentile of the C_1 values of $\widehat{NMB}_{ik}^{*}(\mathbf{w}_0)$, computed in step 4. Denote this as $\widehat{NMB}_{ik}^{*}(\mathbf{w}_0; 1-p)$

6 Repeat steps 2-5 C_2 times obtaining C_2 values of $\widehat{NMB}_{ik}^{*}(\mathbf{w}_0; 1-p)$.

7 Compute the α^{th} percentile of the C_2 values of $\widehat{NMB}_{ik}^{*}(\mathbf{w}_0; 1-p)$. This is a lower $100(1-\alpha)\%$ lower tolerance limit for $NMB_i(\mathbf{w}_0)$ with content p.

Later it will be necessary to estimate the percentiles of $NMB_i(\mathbf{w}_0)$. While this estimation is fairly straightforward, we shall give an algorithm to do so; the algorithm is given below.

Algorithm 9: Estimation of the $(1-p)^{th}$ percentile of $NMB_i(\mathbf{w}_0)$

1. From the sample of each treatment group compute the estimates $\hat{B}_i$ and $\hat{\Sigma}_i$, $i = 1, 2$.
2. For each $i = 1, 2$, independently generate: $\begin{pmatrix} ln[\hat{C}_i] \\ ln[\hat{E}_i] \end{pmatrix} \sim N_2 \left[\hat{B}_i \mathbf{w}_0, \quad \hat{\Sigma}_i \right]$
3. Compute $\widehat{NMB}_i(\mathbf{w}_0) = \lambda \hat{E}_i - \hat{C}_i$
4. Repeat steps 2-3 M times, obtaining M values of $\widehat{NMB}_i(\mathbf{w}_0)$.
5. For each $i = 1, 2$ compute the $(1-p)^{th}$ percentile of the M values of $\widehat{NMB}_i(\mathbf{w}_0)$s.

Later we shall assess the accuracy of the fiducial and percentile bootstrap approaches based on their respective coverage probabilities.

4.2 An example

We shall now apply our proposed individualized cost-effectiveness analysis to the schizophrenia effectiveness study mentioned in Chapter 2 (see Section 2.3). We will evaluate the cost-effectiveness of the new treatment consisting of the anti-psychotic drug Olanzapine as compared to that of the standard treatment where conventional anti-psychotics are first administered, followed by Olanzapine if necessary. Details of the study, including a description of the covariates, are given in Section 2.3. We shall apply the tolerance limit methods in the previous section to twelve specified patients for an individualized analysis. The covariates of the twelve patients are provided in Table 4.1.

Table 4.1 Covariate values of twelve patients

Patient ID	Age	Pysc. Duration	Months
1	40	0	0.25
2	40	0	0
3	40	5	0.25
4	40	5	0
5	40	10	0.25
6	40	10	0
7	60	0	0.25
8	60	0	0
9	60	10	0.25
10	60	10	0
11	60	15	0.25
12	60	15	0

Based on a set of simulated lognormal-lognormal data under the assumed model, the estimates of the unknown parameters are given by

$$\hat{B}_1 = \begin{pmatrix} 9.1197 & -0.0091 & 0.0259 & 0.1303 \\ 4.0311 & 0.0225 & -0.0344 & -0.222 \end{pmatrix}$$

$$\hat{B}_2 = \begin{pmatrix} 8.5866 & -0.0025 & 0.03619 & -0.06789 \\ 4.0106 & -0.0019 & 0.0111 & -0.1591 \end{pmatrix}$$

$$\hat{\Sigma}_1 = \begin{pmatrix} 1.0793 & -0.0233 \\ -0.0233 & 2.5358 \end{pmatrix}$$

$$\hat{\Sigma}_2 = \begin{pmatrix} 1.6956 & -0.1157 \\ -0.1157 & 2.6649 \end{pmatrix}$$

Next, we will explain the computation of the lower tolerance limit with content p = 0.90 and confidence level $1 - \alpha = 0.95$. That is, we will compute a 95% lower confidence limit for the $(1-p)^{th} = 10^{th}$ percentiles of $NMB_1(\mathbf{w}_0)$ and $NMB_2(\mathbf{w}_0)$ corresponding to the covariate vectors $\mathbf{w}_0$ specified in Table 4.1. Before doing so, we shall give estimates of the 10^{th} percentiles of $NMB_1(\mathbf{w}_0)$ and $NMB_2(\mathbf{w}_0)$, obtained using algorithm 9, and based on 10,00,000 simulations. These are provided in Table 4.2 and Table 4.3, respectively.

Table 4.2 Estimates of the 10^{th} percentiles of $NMB_1(\mathbf{w}_0)$ for the twelve patients for different values of λ

Patient ID	$\lambda = 250$	$\lambda = 500$	$\lambda = 1000$	$\lambda = 2000$
1	8145.1	16 712.3	33 852.5	68 100.0
2	8635.8	17 743.8	35 947.4	72 365.0
3	6892.0	14 121.2	28 621.1	57 612.5
4	7222.3	14 824.9	30 045.2	60 476.1
5	5762.1	11 831.2	23 975.7	48 249.8
6	6077.0	12 471.1	25 250.8	50 845.5
7	8144.5	16 704.2	33 840.9	68 120.8
8	13 458.6	27 605.2	55 911.0	112 563.7
9	9068.2	18 610.8	37 717.2	75 933.3
10	9594.8	19 668.6	39 830.4	80 176.7
11	7629.3	15 648.5	31 695.2	63 796.5
12	8109.4	16 654.1	33 721.3	67 885.4

Based on the results in Table 4.2, we can assess the effects of the covariates on the 10^{th} percentiles of the NMB_i distribution for the twelve patients in Table 4.1, when they are assigned to the new treatment. Older patients tend to have higher percentile values compared to younger patients, especially as the willingness-to-pay, λ, increases. The exception to this trend is patient one. Patients with shorter history of psychological duration attain higher percentile values for any given λ. In addition, patients who have spent less time in hospital in the year prior to the study have higher percentile values. Increasing the willingness-to-pay parameter results in larger 10^{th} percentile values, indicating that the increased investment results in an increased benefit for all twelve patients being analysed. Next, we present the estimates of the 10^{th} percentiles for the twelve patients in Table 4.1 when they are under the second treatment, i.e. the standard treatment.

Table 4.3 Estimates of the 10^{th} percentiles of $NMB_2(\mathbf{w}_0)$ for the twelve patients for different values of λ

Patient ID	$\lambda = 250$	$\lambda = 500$	$\lambda = 1000$	$\lambda = 2000$
1	3206.9	6637.8	13 486.1	27 180.0
2	3325.9	6887.2	13 985.5	28 186.1
3	3385.3	7013.0	14 245.2	28 718.1
4	3524.0	7282.8	14 795.3	29 812.7
5	3596.0	7433.2	15 102.6	30 445.2
6	3722.6	7695.8	15 634.1	31 523.7
7	3202.6	6624.0	13 456.5	27 113.4
8	3219.3	6660.9	13 540.6	27 290.2
9	3448.5	7130.4	14 481.7	29 172.5
10	3607.3	7447.3	15 141.9	30 509.6
11	3633.9	7525.8	15 294.2	30 832.7
12	3760.4	7775.4	15 804.1	31 852.4

From Table 4.3, we once again notice that older patients have higher percentile values than the younger patients, with exception of patients one and two. However, the effect of age is noticeably less compared to what we observed under the new treatment. In contrast to the effect of psychological duration on the percentile values for patients under the new treatment, patients under the standard treatment with longer psychological duration benefit more compared to those with shorter history. However, with regards to months spent in hospital in year prior to the study, both treatments provide greater benefit for patients that have spent less time in the hospital the year preceding the study. Comparing the 10^{th} percentiles of the twelve patients in Table 4.2 and Table 4.3, it is clear that the values are larger for patients under treatment one as compared to those under treatment two. Therefore, it is anticipated that the lower tolerance limits will follow a similar trend. The lower ($p = 0.9$, $1 - \alpha = 0.95$) tolerance limits are computed for each treatment using the fiducial methodology provided in algorithm 7, and the results for the first and second treatments are provided in Table 4.4 and Table 4.5, respectively. The results are obtained using algorithm 7 with $C_1 = 1000$ and $C_2 = 2000$.

Table 4.4 Fiducial lower ($p = 0.90$, $1 - \alpha = 0.95$) tolerance limits for the twelve patients under treatment one for different values of λ

Patient ID	$\lambda = 250$	$\lambda = 500$	$\lambda = 1000$	$\lambda = 2000$
1	4750.8	9827.8	19 936.3	40 168.2
2	4892.5	10 131.7	20 646.4	41 866.5
3	4359.1	8956.9	18 265.8	36 874.9
4	4646.9	9628.4	19 424.1	39 232.8
5	4031.4	8349.4	16 890.6	34 192.4
6	4263.5	8816.6	17 884.5	36 061.1
7	4773.1	9722.7	19 695.4	39 549.9
8	5851.6	11 954.4	24 238.6	48 786.7
9	4662.3	9645.2	19 617.3	39 298.3
10	5081.1	10 387.9	21 005.7	42 296.5
11	4316.9	8873.1	18 188.6	36 449.8
12	4574.7	9385.2	19 068.8	38 401.0

Table 4.5 Fiducial lower ($p = 0.90$, $1 - \alpha = 0.95$) tolerance limits for the twelve patients under treatment two for different values of λ

Patient ID	$\lambda = 250$	$\lambda = 500$	$\lambda = 1000$	$\lambda = 2000$
1	1904.1	3950.3	7960.9	16 079.7
2	1947.7	4043.8	8209.3	16 563.7
3	2215.7	4627.9	9379.6	18 922.3
4	2249.4	4729.2	9583.7	19 410.7
5	2434.8	5097.7	10 411.1	20 834.4
6	2542.3	5316.8	10 833.4	21 861.6
7	1872.1	3878.8	7880.1	15 899.4
8	1267.8	2646.0	5387.1	10 814.4
9	1746.9	3652.9	7441.7	15 009.9
10	1782.6	3676.0	7460.1	15 086.1
11	1961.5	4042.8	8213.6	16 567.9
12	2047.0	4300.7	8766.3	17 588.6

Regarding the effects of the covariates on the lower limits, the trends are similar to those described in reference to Table 4.2 and Table 4.3. As expected, an increasing λ results in larger lower limits. Using the percentile bootstrap method (i.e., algorithm 8), similar results on the lower ($p = 0.9$, $1 - \alpha = 0.95$) are provided in Table 4.6 and Table 4.7. These lower limits are also computed using algorithm 8 with $C_1 = 1000$ and $C_2 = 2000$.

Table 4.6 Percentile bootstrap lower ($p = 0.90$, $1 - \alpha = 0.95$) tolerance limits for the twelve patients under treatment one for different values of λ

Patient ID	$\lambda = 250$	$\lambda = 500$	$\lambda = 1000$	$\lambda = 2000$
1	5030.9	10 309.7	21 109.9	42 391.8
2	5319.8	10 902.2	22 165.8	44 750.1
3	4671.4	9693.7	19 579.1	39 292.1
4	4866.1	10 000.9	20 324.1	41 026.4
5	4152.0	8551.8	17 426.0	35 326.8
6	4405.5	9039.2	18 392.8	36 984.9
7	5081.9	10 512.7	21 233.6	42 772.6
8	5908.1	12 164.3	24 597.9	49 442.3
9	4775.8	9756.3	19 991.2	40 268.8
10	5017.9	10 538.5	21 119.2	42 575.9
11	4415.0	9060.7	18 546.6	37 206.7
12	4631.9	9509.4	19 183.1	38 576.6

Table 4.7 Percentile bootstrap lower ($p = 0.90$, $1 - \alpha = 0.95$) tolerance limits for the twelve patients under treatment two for different values of λ

Patient ID	$\lambda = 250$	$\lambda = 500$	$\lambda = 1000$	$\lambda = 2000$
1	1916.4	3977.8	8086.3	16 278.7
2	2021.0	4177.2	8482.5	17 044.0
3	2226.4	4637.2	9329.8	18 840.5
4	2278.3	4726.6	9595.7	19 363.7
5	2515.4	5278.2	10 755.8	21 600.3
6	2594.9	5409.0	10 995.3	22 297.9
7	1909.7	3915.6	7940.9	15 947.8
8	1361.9	2828.9	5736.7	11 506.9
9	1764.6	3619.1	7341.2	14 778.2
10	1858.7	3848.9	7913.0	15 885.6
11	2061.6	4292.8	8699.8	17 591.1
12	2132.9	4448.6	9087.5	18 311.7

Comparing the lower limits in Table 4.4 and Table 4.6, and those in Table 4.5 and Table 4.7, it is clear that the percentile bootstrap approach results in larger limits. There are two cases in which the lower limits in Table 4.4 and Table 4.5 are larger than those in Table

4.6 and Table 4.7, respectively; specifically, when $\lambda = 250$ for patients 3 and 10. Comparing the resulting limits from Table 4.5 with those in Table 4.7, the percentile bootstrap approach generally provides larger lower limits. However, there are four exceptions which occur when $\lambda = 2000$ for patients 3, 4, and 9; and also for patient 4 when $\lambda = 500$. Overall, these differences do not appear to be large, indicating that the fiducial and percentile bootstrap approaches provide similar results.

Additionally, the cost-effectiveness of the new treatment is also assessed using the cost-effectiveness acceptability curve (CEAC). Figure 4.1 shows the CEAC curve for patient one, and the plots for the other eleven patients are nearly identical. What is plotted is the probability $P[NMB_{11}(\mathbf{w}_0) - NMB_{21}(\mathbf{w}_0) > 0]$, estimated using 10,000 simulated samples. The plot is given for λ ranging from 0 to 50; the probabilities appear to plateau at $\lambda = 50$. The maximum probability that the new treatment is cost effective occurs at the largest λ $= 2000$, around 0.61 for all patients (not shown in the plot). However, based on the results in Table 4.4 , Table 4.5, Table 4.6 , and Table 4.7; the lower limits for the patients indicate that the new treatment is cost-effective; as the lower limits are always greater for the new treatment. In addition, the results we have obtained demonstrate that the cost-effectiveness of the new treatment increases as a function of λ.

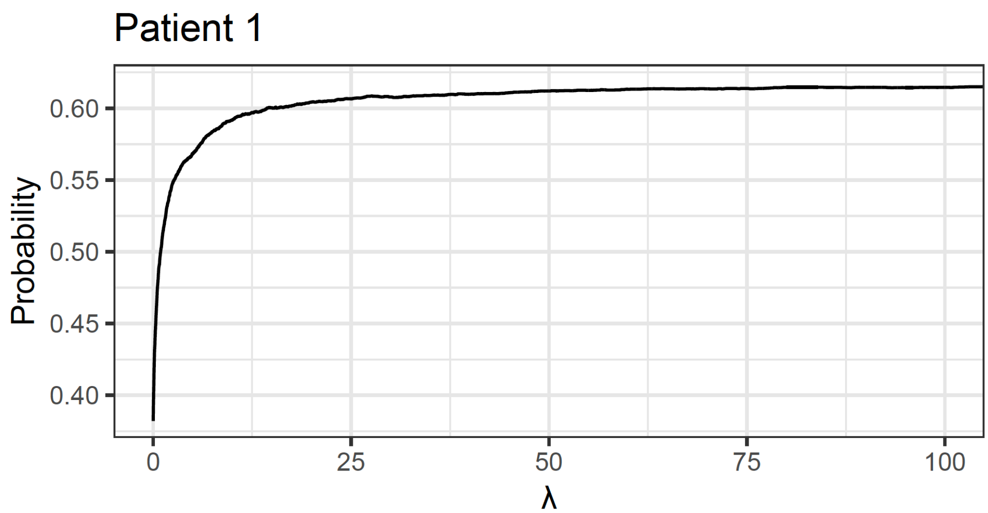


Figure 4.1: CEAC for patient one

An approach to complement the tolerance limit analysis is to consider a plot of the estimated CDF of the NMBs for fixed willingness-to-pay values. The results for the twelve patients were computed using 10,000 simulated samples. As the results for all patients are very similar, we provide only the results for patient one in Figure 4.2.

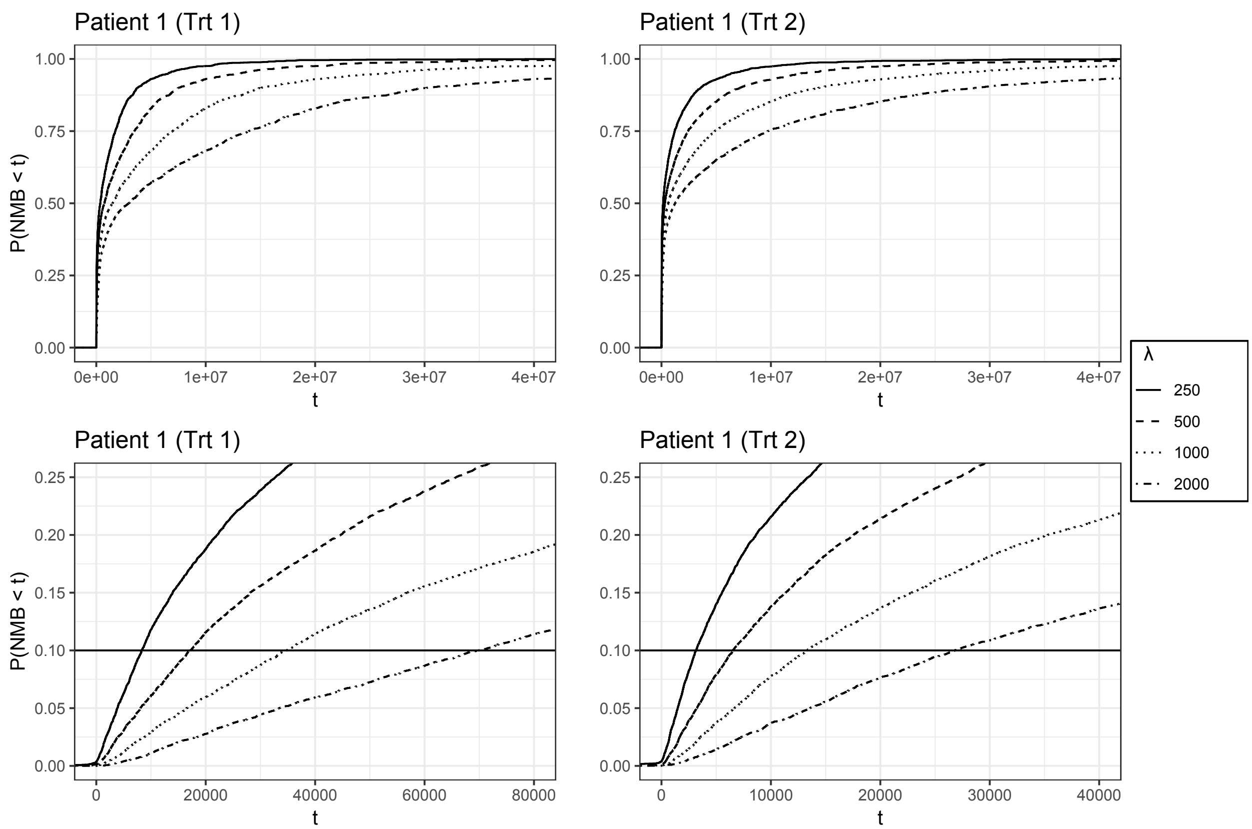


Figure 4.2: ECDF plots for patient one

[Figure 4.2] displays the CDFs for patients under the new treatment, treatment one, and standard treatment, treatment two. A close-up is provided for values around the 10^{th} percentile. The 10^{th} percentile is denoted by a horizontal line and the results of the plots are in agreement with the results in Table 4.2 and Table 4.3. A noted criticism of the CEAC is that it is often misinterpreted, i.e. probability of cost-effectiveness for a new treatment implies that it dominates the standard treatment. To investigate the stochastic dominance of the new treatment to the standard treatment four separate plots are shown in Figure 4.3, each for a given λ value, corresponding to patient one. The remaining eleven patients exhibit similar trends.

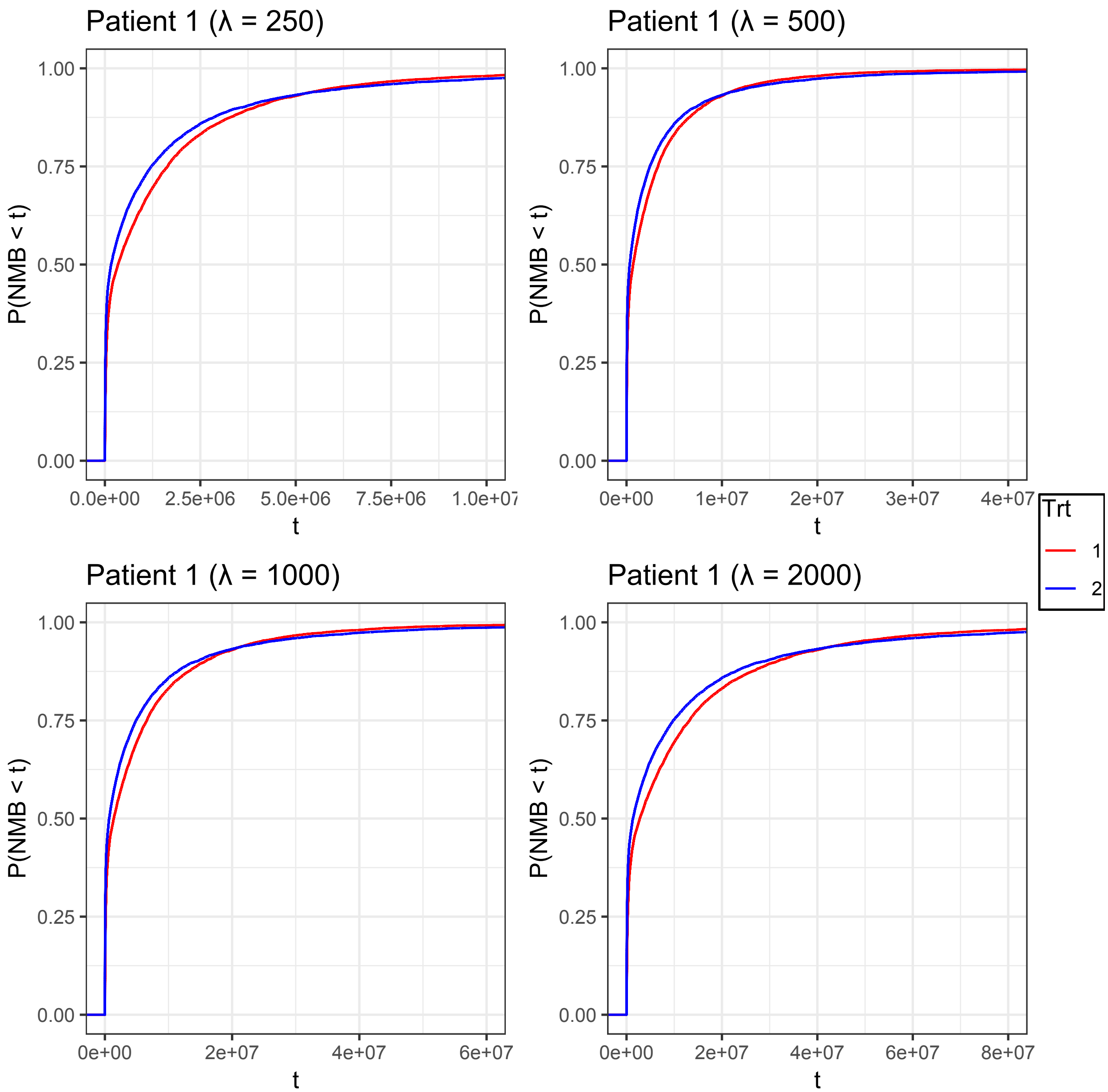


Figure 4.3: Stochastic dominance comparison for patient one

From Figure 4.3 it is clear that the new treatment is not stochastically dominant for all values of NMB. The results indicate that the new treatment is stochastically dominant for small and moderate NMB values, however as the NMB increases to large values the standard treatment becomes dominant. In addition, as λ increases, the point at which the standard treatment becomes stochastically dominant increases. That is, as the amount of investment in both treatments increases, the stochastic dominance of the new treatment holds for more NMB values. This means that increased investment in the new treatment results in cost-effectiveness for more patients under the new treatment than compared to the standard treatment. Finally, as noted in the related literature, we have demonstrated that the comparison of the CDFs is more fruitful in treatment allocation compared to the CEAC. Next, we will further compare the fiducial and percentile bootstrap approaches based on coverage probabilities and expected lower limits, reported in the next section.

4.2.1 Coverage probabilities and expected lower limits

We estimated the coverage probabilities associated with the lower tolerance limits using 5000 simulated samples, when the tolerance limits are computed using the fiducial approach and the percentile bootstrap approach, for $1-p=0.10$ and $1-\alpha=0.95$. The samples were simulated by choosing the estimates of the parameters obtained in the previous subsection as the true values. To begin with, we used the sample sizes corresponding to the example; that is, $n_1 = 202$ for the first treatment group, and $n_2 = 174$ for the second treatment group. Furthermore, the covariate values used (i.e., the matrices W_1 and W_2) were the same as those for the example. For each simulated sample, we used algorithm 7 to compute the lower tolerance limit with $C_1 = 1000$ and $C_2 = 2000$. In order to estimate the coverage probability, we also need the true value of the corresponding parameter, namely the 10^{th} percentile (since $1-p=0.10$). The true value of the 10^{th} percentile was computed using algorithm 9. The coverage probability results for the fiducial lower tolerance limit of the first treatment group are reported in Table 4.8 and those pertaining to the second treatment group are reported in Table 4.9.

Table 4.8 Coverage probabilities of the fiducial lower tolerance limits ($p = 0.90$, $1 - \alpha = 0.95$) for treatment one under $n_1 = 202$ and $n_2 = 174$ for different values of λ

Patient ID	$\lambda = 250$	$\lambda = 500$	$\lambda = 1000$	$\lambda = 2000$
1	0.951	0.950	0.950	0.950
2	0.953	0.953	0.952	0.951
3	0.957	0.955	0.955	0.954
4	0.954	0.955	0.954	0.952
5	0.955	0.956	0.955	0.954
6	0.956	0.955	0.954	0.954
7	0.951	0.950	0.949	0.949
8	0.953	0.952	0.952	0.951
9	0.949	0.949	0.948	0.949
10	0.954	0.953	0.952	0.952
11	0.953	0.954	0.953	0.953
12	0.956	0.956	0.956	0.957

Table 4.9 Coverage probabilities of the fiducial lower tolerance limits ($p = 0.90$, $1 - \alpha = 0.95$) for treatment two under $n_1 = 202$ and $n_2 = 174$ for different values of λ

Patient ID	$\lambda = 250$	$\lambda = 500$	$\lambda = 1000$	$\lambda = 2000$
1	0.949	0.949	0.948	0.948
2	0.949	0.949	0.948	0.947
3	0.951	0.950	0.949	0.949
4	0.950	0.949	0.948	0.947
5	0.958	0.957	0.957	0.956
6	0.956	0.955	0.955	0.955
7	0.949	0.948	0.947	0.948
8	0.953	0.951	0.950	0.951
9	0.953	0.952	0.951	0.951
10	0.953	0.952	0.952	0.952
11	0.955	0.954	0.953	0.953
12	0.954	0.954	0.953	0.952

The results in Table 4.8 and Table 4.9 indicate that the coverage probabilities under the fiducial method are satisfactory under the sample sizes $n_1 = 202$ and $n_2 = 174$.

We shall now report the estimated coverage probabilities for the percentile bootstrap approach. The set up is the same as that used for Table 4.8 and Table 4.9. We used 5000 simulated samples, and also used $C_1 = 1000$ and $C_2 = 2000$ while applying (algorithm 8) in order to compute the lower tolerance limit. The results for the new and the standard treatment are provided in Table 4.10 and Table 4.11.

Table 4.10 Coverage probabilities of the percentile bootstrap lower tolerance limits ($p = 0.90$, $1 - \alpha = 0.95$) for treatment one under $n_1 = 202$ and $n_2 = 174$ for different values of λ

Patient ID	$\lambda = 250$	$\lambda = 500$	$\lambda = 1000$	$\lambda = 2000$
1	0.948	0.948	0.947	0.947
2	0.947	0.947	0.947	0.946
3	0.952	0.951	0.951	0.951
4	0.947	0.947	0.948	0.948
5	0.954	0.955	0.954	0.955
6	0.952	0.951	0.951	0.951
7	0.948	0.948	0.948	0.947
8	0.945	0.944	0.943	0.944
9	0.944	0.944	0.944	0.944
10	0.946	0.946	0.946	0.946
11	0.946	0.947	0.947	0.947
12	0.949	0.948	0.947	0.947

Table 4.11 Coverage probabilities of the percentile bootstrap lower tolerance limits ($p = 0.90$, $1 - \alpha = 0.95$) for treatment two under $n_1 = 202$ and $n_2 = 174$ for different values of λ

Patient ID	$\lambda = 250$	$\lambda = 500$	$\lambda = 1000$	$\lambda = 2000$
1	0.948	0.948	0.947	0.948
2	0.947	0.947	0.947	0.946
3	0.950	0.950	0.950	0.949
4	0.951	0.949	0.948	0.947
5	0.953	0.952	0.951	0.952
6	0.950	0.949	0.949	0.949
7	0.948	0.947	0.947	0.947
8	0.945	0.946	0.946	0.945
9	0.947	0.946	0.946	0.945
10	0.947	0.947	0.947	0.946
11	0.947	0.946	0.946	0.945
12	0.946	0.944	0.944	0.944

The percentile bootstrap coverage probabilities are also close to the nominal 0.95. The overall pattern seems to be that the fiducial approach provides coverages that are slightly larger than the nominal level, and the percentile bootstrap approach provides coverages that are slightly smaller than the nominal level. However, we can conclude that both the fiducial and percentile bootstrap approaches provide satisfactory coverage probabilities under the sample sizes $n_1 = 202$ and $n_2 = 174$. Thus it of interest to compare the expected values of the lower limits obtained using the two approaches. The expected values of lower tolerance limits obtained using the fiducial approach are provided in Table 4.12 and Table 4.13, corresponding to the new treatment and standard treatment, respectively.

Table 4.12 Expected values of the fiducial lower tolerance limits ($p = 0.90$, $1 - \alpha = 0.95$) for treatment one under $n_1 = 202$ and $n_2 = 174$ for different values of λ

Patient ID	$\lambda = 250$	$\lambda = 500$	$\lambda = 1000$	$\lambda = 2000$
1	5256.6	10 826.2	21 977.1	44 293.1
2	5540.6	11 409.9	23 164.9	46 682.2
3	4731.9	9750.7	19 802.6	39 916.5
4	4984.6	10 270.1	20 856.8	42 040.5
5	4223.2	8709.0	17 693.4	35 673.4
6	4445.2	9165.8	18 622.4	37 546.5
7	5254.3	10 820.8	21 967.6	44 275.4
8	6944.1	14 287.0	28 987.1	58 402.7
9	5300.5	10 908.6	22 137.1	44 611.0
10	5591.6	11 510.4	23 359.6	47 077.5
11	4738.4	9755.1	19 798.9	39 900.0
12	5000.9	10 296.8	20 900.0	42 115.1

Table 4.13 Expected values of the fiducial lower tolerance limits ($p = 0.90$, $1 - \alpha = 0.95$) for treatment two under $n_1 = 202$ and $n_2 = 174$ for different values of λ

Patient ID	$\lambda = 250$	$\lambda = 500$	$\lambda = 1000$	$\lambda = 2000$
1	1968.7	4095.5	8348.3	16 850.6
2	2039.3	4241.8	8646.2	17 453.7
3	2241.6	4667.4	9516.3	19 214.4
4	2321.3	4832.6	9853.0	19 895.5
5	2528.7	5269.9	10 749.5	21 712.1
6	2615.9	5451.4	11 120.4	22 457.5
7	1968.9	4095.8	8348.4	16 852.7
8	1545.5	3211.4	6541.7	13 199.2
9	1913.5	3978.0	8105.3	16 359.1
10	1985.8	4127.8	8411.3	16 972.4
11	2164.4	4501.4	9173.9	18 517.6
12	2246.8	4672.9	9522.7	19 220.5

We note that since we are computing a lower tolerance limit, the larger the expected value, the better (provided the coverage probability is satisfactory). Comparing the expected lower tolerance limits in Table 4.12 and Table 4.13 with the lower limit results obtained in Table 4.4 and Table 4.5, respectively, the expected limits are larger. However, the difference is minimal considering the magnitude of the limits.

Since the lower limits and coverage probabilities from the percentile bootstrap and fiducial methods were similar, we anticipate that the expected lower limits under the percentile bootstrap will exhibit behaviour similar to that under the fiducial approach. The expected lower limits obtained using the percentile bootstrap method are provided in Table 4.14 and Table 4.15, corresponding to the new and standard treatment, respectively.

Table 4.14 Expected values of the percentile bootstrap lower tolerance limits ($p = 0.90$, $1 - \alpha = 0.95$) for treatment one under $n_1 = 202$ and $n_2 = 174$ for different values of λ

Patient ID	$\lambda = 250$	$\lambda = 500$	$\lambda = 1000$	$\lambda = 2000$
1	5310.2	10 925.2	22 167.6	44 666.0
2	5600.7	11 523.3	23 379.1	47 103.4
3	4786.4	9852.9	19 998.3	40 298.4
4	5044.2	10 384.0	21 075.6	42 469.8
5	4275.5	8807.8	17 884.1	36 045.5
6	4504.1	9278.6	18 838.5	37 971.5
7	5310.2	10 924.2	22 165.0	44 655.8
8	7047.5	14 486.9	29 378.1	59 183.7
9	5376.3	11 053.6	22 420.8	45 169.0
10	5682.7	11 684.0	23 700.5	47 742.3
11	4812.8	9899.2	20 082.7	40 462.6
12	5086.6	10 461.6	21 224.4	42 759.0

Table 4.15 Expected values of the percentile bootstrap lower tolerance limits ($p = 0.90$, $1 - \alpha = 0.95$) for treatment two under $n_1 = 202$ and $n_2 = 174$ for different values of λ

Patient ID	$\lambda = 250$	$\lambda = 500$	$\lambda = 1000$	$\lambda = 2000$
1	2008.7	4170.0	8491.8	17 132.0
2	2080.7	4321.0	8797.9	17 752.8
3	2285.9	4751.0	9678.3	19 529.1
4	2368.3	4920.5	10 022.5	20 224.6
5	2578.5	5363.8	10 931.4	22 064.5
6	2668.3	5549.9	11 311.1	22 831.7
7	2007.7	4169.0	8489.5	17 129.9
8	1582.7	3282.1	6677.8	13 467.4
9	1960.0	4066.3	8276.2	16 694.2
10	2033.7	4219.2	8589.2	17 327.1
11	2216.3	4600.2	9365.6	18 895.6
12	2299.9	4774.4	9720.8	19 611.2

The expected lower tolerance limits are slightly larger than those obtained by the fiducial methodology. This is expected, since the coverage probabilities under the percentile bootstrap approach were slightly less than those obtained when utilizing the fiducial approach. The differences in the expected lower limits do not appear to be significant between the two methods.

For further comparison between the two approaches for small sample sizes, we shall now consider $n_1 = n_2 = 50$. The estimated coverage probabilities are given in Table 4.16, Table 4.17, Table 4.18, and Table 4.19.

Table 4.16 Coverage probabilities of the fiducial lower tolerance limits ($p = 0.90$, $1 - \alpha = 0.95$) for treatment one under $n_1 = n_2 = 50$ for different values of λ

Patient ID	$\lambda = 250$	$\lambda = 500$	$\lambda = 1000$	$\lambda = 2000$
1	0.954	0.954	0.953	0.952
2	0.957	0.955	0.955	0.954
3	0.958	0.955	0.955	0.955
4	0.957	0.956	0.955	0.955
5	0.959	0.957	0.956	0.956
6	0.959	0.958	0.958	0.957
7	0.955	0.954	0.952	0.952
8	0.956	0.956	0.955	0.955
9	0.955	0.954	0.954	0.953
10	0.955	0.954	0.954	0.953
11	0.958	0.957	0.956	0.955
12	0.958	0.958	0.957	0.957

Table 4.17 Coverage probabilities of the fiducial lower tolerance limits ($p = 0.90$, $1 - \alpha = 0.95$) for treatment two under $n_1 = n_2 = 50$ for different values of λ

Patient ID	$\lambda = 250$	$\lambda = 500$	$\lambda = 1000$	$\lambda = 2000$
1	0.960	0.959	0.958	0.958
2	0.960	0.959	0.957	0.957
3	0.961	0.960	0.958	0.958
4	0.961	0.958	0.958	0.957
5	0.963	0.961	0.960	0.960
6	0.963	0.960	0.959	0.958
7	0.960	0.959	0.957	0.957
8	0.960	0.959	0.959	0.959
9	0.958	0.957	0.957	0.956
10	0.958	0.957	0.957	0.957
11	0.958	0.956	0.956	0.956
12	0.957	0.957	0.956	0.956

We note that when $n_1 = n_2 = 50$, the fiducial approach gives accurate coverages, though slightly more conservative compared to the large sample results.

Next, we present the coverage probability results under small sample sizes for the percentile bootstrap procedure.

Table 4.18 Coverage probabilities of the percentile bootstrap lower tolerance limits ($p = 0.90$, $1 - \alpha = 0.95$) for treatment one under $n_1 = n_2 = 50$ for different values of λ

Patient ID	$\lambda = 250$	$\lambda = 500$	$\lambda = 1000$	$\lambda = 2000$
1	0.933	0.933	0.933	0.933
2	0.933	0.933	0.933	0.933
3	0.932	0.932	0.932	0.932
4	0.931	0.930	0.931	0.931
5	0.929	0.929	0.929	0.929
6	0.931	0.930	0.930	0.929
7	0.933	0.933	0.932	0.933
8	0.934	0.933	0.933	0.933
9	0.930	0.931	0.931	0.931
10	0.930	0.931	0.930	0.931
11	0.932	0.931	0.931	0.931
12	0.933	0.934	0.934	0.934

Table 4.19 Coverage probabilities of the percentile bootstrap lower tolerance limits ($p = 0.90$, $1 - \alpha = 0.95$) for treatment two under $n_1 = n_2 = 50$ for different values of λ

Patient ID	$\lambda = 250$	$\lambda = 500$	$\lambda = 1000$	$\lambda = 2000$
1	0.937	0.937	0.937	0.937
2	0.936	0.936	0.936	0.936
3	0.935	0.935	0.934	0.934
4	0.937	0.936	0.935	0.935
5	0.938	0.936	0.936	0.936
6	0.937	0.936	0.937	0.936
7	0.936	0.936	0.936	0.936
8	0.944	0.944	0.944	0.943
9	0.941	0.941	0.941	0.941
10	0.942	0.941	0.942	0.941
11	0.937	0.938	0.937	0.938
12	0.939	0.937	0.937	0.937

The coverage probabilities in Table 4.18 and Table 4.19 are generally below the nominal level. From the above four tables it is clear that the fiducial method is to be preferred for small sample sizes. Thus our overall recommendation is the fiducial methodology. We also note that the computational effort is essentially the same for both approaches.

4.3 Discussion and conclusions

In this chapter we have developed methods for individualized inference using the distribution of the NMB random variable for a patient. Lower tolerance limits are constructed using the fiducial approach and a percentile bootstrap approach. As expected, the lower tolerance limits are affected by the patient level covariates. The choice of the percentile to be used (i.e., the value of p) is clearly subjective. The treatment with a larger lower tolerance limit is recommended over its competitor. Under the choice $1 - p = 0.10$, if the new treatment shows a larger lower tolerance limit for the new treatment for a particular patient, we conclude with 95% confidence that among such patients, 90% or more have a larger NMB, compared to the standard treatment. We want to emphasize again that as was the case in Chapter 2, we are not modelling the NMB random variable directly; rather, we model the cost and effectiveness data using a bivariate regression model, where log-transformation can be first applied when necessary, and then consider the distribution of the NMB under such a model. This appears more appropriate to us, especially when the cost and/or effectiveness is lognormally distributed. Lastly, we have also demonstrated through the example that the numerical analysis presented can be complemented by graphical analysis, our recommendation is to use CDFs over CEAC.

Chapter 5

Aggregate net monetary benefit: A stochastic dominance approach

In the previous chapter, we considered a comparison of the distributions of the NMB random variables based on a specified percentile, in terms of lower tolerance limits. In this chapter we explore a stronger comparison of the NMB distributions using stochastic ordering. Clearly, if the NMB distribution corresponding to the new treatment is stochastically larger compared to that for the standard treatment, we have a strong reason to conclude the cost-effectiveness of the new treatment. This approach has not been explored in the CEA literature. In Stinnett and Mullahy (1998) the authors note that stochastic dominance is an approach to consider in the net-benefit framework. The authors note that stochastic dominance assessed graphically can complement other analyses; for example, confidence limits for various criteria that are functions of the parameters of the NMB distribution (we note that both the INB and ICER are functions of the means of the NMB distributions for the new treatment and the standard treatment). In addition, the authors highlight the usefulness of first-order and second-order stochastic dominance in health economics:

> Stochastic dominance is a powerful analytic tool because it allows one to identify cases in which a decision maker should unambiguously prefer one alternative over another despite the presence of uncertainty, with only very general assumptions required re-

garding the decision maker's utility function. (p. S75)

In a paper by Leshno and Levy (2004), the authors assess competing health care options using stochastic dominance (of first and second order) utilizing empirical distribution functions. Multivariate first-order dominance analysis is presented in Hussain, Jørgensen and Østerdal (2016). In this work, the authors investigate dominance comparisons between population groups based on a comparison of empirical distribution functions that account for patient-level characteristics. Most approaches (in economics literature) utilizing stochastic dominance seek testing procedures to accompany their methods or rely solely on graphical analysis of the CDFs. In this chapter, we propose to develop a methodological framework to assess the stochastic dominance of the NMB for the new treatment compared to the standard treatment. However, our analysis will only be an aggregate level analysis, even though the methodology can be individualized.

5.1 The stochastic dominance criterion

We shall consider the lognormal-normal model for the cost C_i and effectiveness E_i for a patient in the i^{th} treatment group:

$$X_i = \begin{pmatrix} \ln(C_i) \\ E_i \end{pmatrix} \sim N\left(\mu_i = \begin{pmatrix} \mu_{iC} \\ \mu_{iE} \end{pmatrix}, \Sigma_i = \begin{pmatrix} \sigma_{iC}^2 & \rho\sigma_{iC}\sigma_{iE} \\ \rho\sigma_{iC}\sigma_{iE} & \sigma_{iE}^2 \end{pmatrix}\right), \tag{5.1.1}$$

$i = 1, 2$. Let $\hat{\mu}_i$ and $\hat{\Sigma}_i$ denote the unbiased estimators of the respective parameters; see Section 3.3. If C_i and E_i, respectively, denote the cost and effectiveness for a randomly chosen patient from the i^{th} treatment group, the corresponding NMB, say NMB_i, is the random variable given by

$$NMB_i = \lambda E_i - C_i. \tag{5.1.2}$$

We are interested in comparing the distributions of NMB_1 and NMB_2 in terms of stochastic dominance. We recall that NMB_1 is stochastically larger than NMB_2 if

$$P(NMB_1 \geq t) \geq P(NMB_2 \geq t),$$

for all t. Before we proceed with this definition, we note that requiring the above to hold for all t maybe unrealistic. Furthermore, from a practical point of view, all values of t may not be of interest; for example, very large values of the NMB may not be possible in practice, and negative values of t may not be of interest. Thus we shall require the above to hold for all t in an interval $(t_1,\ t_2)$, where t_1 and t_2 are known bounds. Clearly, a decision maker has to decide what is a reasonable interval. We shall proceed under the assumption that t_1 and t_2 are known bounds. Thus our criterion is

$$P(NMB_1 \geq t) - P(NMB_2 \geq t) \geq 0, \quad \text{for all } t \in (t_1,\ t_2). \tag{5.1.3}$$

We note that the condition (5.1.3) is equivalent to

$$\min_{t\in(t_1,t_2)} [P(NMB_1 \geq t) - P(NMB_2 \geq t)] \geq 0. \tag{5.1.4}$$

We shall be working with the condition (5.1.4). Note that the left hand side of (5.1.4) involves unknown parameters; thus we shall compute a lower confidence limit for the left hand side of (5.1.4), and conclude stochastic dominance if the lower confidence limit is positive. The fiducial approach as well as the percentile bootstrap approach will be pursued for computing the required lower confidence limit.

5.2 The assessment of stochastic dominance

We shall first develop a fiducial-based numerical approach for the assessment of stochastic dominance using the condition (5.1.4). We shall assume that $\tilde{\mu}_i$ and $\tilde{\Sigma}_i$ are the fiducial quantities for μ_i and Σ_i, respectively, $i = 1, 2$. These fiducial quantities are exhibited in Section 3.3 of Chapter 3.

Since we are in the lognormal-normal scenario, let $Y_{1i} = ln(C_i)$ and $Y_{2i} = E_i$. Then

$$P(NMB_i \geq t) = P(\lambda Y_{2i} - e^{Y_{1i}} \geq t) = P\Big(Y_{2i} \geq \frac{t + e^{Y_{1i}}}{\lambda}\Big) \tag{5.2.1}$$

Since

$$Y_{2i}|Y_{1i} = y_{1i} \sim N\Big[\mu_{iE} + \rho_i \frac{\sigma_{iE}}{\sigma_{iC}}(y_{1i} - \mu_{Ci}); \sqrt{(1-\rho_i^2)\sigma_{iE}^2}\Big], \tag{5.2.2}$$

we have

$$P(NMB_i \geq t) = 1 - E_{Y_{1i}}\Bigg\{\Phi\Bigg[\frac{(\frac{t+e^{y_{1i}}}{\lambda}) - \mu_{iE} - \rho_i \frac{\sigma_{iE}}{\sigma_{iC}}(y_{1i} - \mu_{iC})}{\sqrt{(1-\rho_i^2)\sigma_{iE}^2}}\Bigg]\Bigg\} \tag{5.2.3}$$

Our goal is to construct a 100(1-α)% lower confidence limit for $P(NMB_1 \geq t^*) - P(NMB_2 \geq t^*)$ where t^* is the value of t that minimizes $P(NMB_1 \geq t) - P(NMB_2 \geq t)$. Note that t is a function of unknown parameters, and there is no analytic form for t. Nevertheless, the fiducial approach can be numerically implemented to compute the required lower confidence limit. The required steps are given in algorithm 10. In the algorithm, $\tilde{\mu}_i$ and $\tilde{\Sigma}_i$ will denote the fiducial quantities for μ_i and Σ_i, respectively. Furthermore, the elements of $\tilde{\mu}_i$ and $\tilde{\Sigma}_i$ will be denoted accordingly. For example, with $\mu_i = (\mu_{iC}, \mu_{iE})'$, we will write $\tilde{\mu}_i = (\tilde{\mu}_{iC}, \tilde{\mu}_{iE})'$.

Algorithm 10: Fiducial lower confidence limits for $\min_{t \in (t_1, t_2)} \left[P(NMB_1 \geq t) - P(NMB_2 \geq t) \right]$

1 Compute the observed values $\hat{\mu}_{io}$ and $\hat{\Sigma}_{io}$ of the estimates $\hat{\mu}_i$ and $\hat{\Sigma}_i$, $i = 1,\, 2$.

2 For $i = 1,\, 2$, independently generate: $H_i \sim W\left(\left\{(n_i - 1)\hat{\Sigma}_{io}\right\}^{-1}, n_i - 1\right)$ and $Z_i \sim 2 \times 1$ matrix of $N(0, 1)$ random variates.

3 Compute $\tilde{\Sigma}_i = H_i^{-1}$ and $\tilde{\mu}_i = \hat{\mu}_{io} - \tilde{\Sigma}_i^{1/2} \frac{Z_i}{\sqrt{n_i}}$.

4 Generate K random variates $\tilde{Y}_{1ik} \sim N(\tilde{\mu}_{iC}, \tilde{\sigma}^2_{iC})$, $i = 1,\, 2$; $k = 1, 2, ..., K$.

5 Referring to (5.2.3), compute

$$R_{ik} = \Phi\left[\frac{\left(\frac{t + e^{\tilde{y}_{1ik}}}{\lambda}\right) - \tilde{\mu}_{iE} - \tilde{\rho}_i \frac{\tilde{\sigma}_{iE}}{\tilde{\sigma}_{iC}}(\tilde{y}_{1ik} - \tilde{\mu}_{iC})}{\sqrt{(1 - \tilde{\rho}_i^2)\tilde{\sigma}^2_{iE}}}\right],$$

$i = 1,\, 2$; $k = 1, 2, ..., K$.

6 Referring once again to (5.2.3), we see that for a fixed t, a fiducial quantity for $P(NMB_i \geq t)$ is now given by $1 - \frac{1}{K}\sum_{k=1}^{K} R_{ik}$.

7 For a grid of values of $t \in (t_1, t_2)$, repeat steps 2-6 and compute the fiducial quantities for $P(NMB_1 \geq t)$ and for $P(NMB_2 \geq t)$, and hence for the difference $P(NMB_1 \geq t) - P(NMB_2 \geq t)$.

8 Take the minimum of the fiducial quantities of $P(NMB_1 \geq t) - P(NMB_2 \geq t)$ over the chosen grid of values of $t \in (t_1, t_2)$. This gives a fiducial quantity for $\min_{t \in (t_1, t_2)} \left[P(NMB_1 \geq t) - P(NMB_2 \geq t) \right]$.

9 Repeat steps 2-8 M times

10 A lower $100(1 - \alpha)\%$ lower confidence limit for $\min_{t \in (t_1, t_2)} \left[P(NMB_1 \geq t) - P(NMB_2 \geq t) \right]$ corresponds to the α^{th} percentile of the M fiducial quantities of $\min_{t \in (t_1, t_2)} \left[P(NMB_1 \geq t) - P(NMB_2 \geq t) \right]$.

The steps required to implement the percentile bootstrap approach are identical, except that instead of using the fiducial quantities for μ_i and Σ_i, we shall use parametric bootstrap samples based on the distributions of $\hat{\mu}_i$ and $\hat{\Sigma}_i$. For completeness, the steps are given in algorithm 11.

Algorithm 11: Percentile bootstrap lower confidence limits for $\min_{t\in(t_1,t_2)}\left[P(NMB_1 \geq t) - P(NMB_2 \geq t)\right]$

---

1 Compute the the estimates $\hat{\mu}_i$ and $\hat{\Sigma}_i$, $i = 1, 2$.

2 For $i = 1, 2$, independently generate: $\hat{\mu}_{i*} \sim N\left[\hat{\mu}_i, \frac{1}{n_i}\hat{\Sigma}\right]$ and $\hat{\Sigma}_{i*} \sim \frac{1}{n_i-1}W\left[\hat{\Sigma}_i, n_i - 1\right]$.

3 Generate K random variates $Y^*_{1ik} \sim N(\hat{\mu}_{iC*}, \hat{\sigma}^2_{iC*})$, $i = 1, 2$; $k = 1, 2, ..., K$.

4 Referring to (5.2.3), compute

$$R^*_{ik} = \Phi\left[\frac{(\frac{t+e^{y^*_{1ik}}}{\lambda}) - \mu_{iE*} - \rho_{i*}\frac{\sigma_{iE*}}{\sigma_{iC*}}(y^*_{1ik} - \mu_{iC*})}{\sqrt{(1-\rho^2_{i*})\sigma^2_{iE*}}}\right]$$

$i = 1, 2$; $k = 1, 2, ..., K$.

5 A parametric bootstrap version of $P(NMB_i \geq t)$ is given by $1 - \frac{1}{K}\sum_{k=1}^{K} R^*_{ik}$.

6 For a grid of values of $t \in (t_1, t_2)$, repeat the above steps and compute the parametric bootstrap versions of $P(NMB_1 \geq t)$ and for $P(NMB_2 \geq t)$, and hence for the difference $P(NMB_1 \geq t) - P(NMB_2 \geq t)$.

7 Take the minimum of the parametric bootstrap versions of $P(NMB_1 \geq t) - P(NMB_2 \geq t)$ over the chosen grid of values of $t \in (t_1, t_2)$. This gives a parametric bootstrap version of $\min_{t\in(t_1,t_2)}\left[P(NMB_1 \geq t) - P(NMB_2 \geq t)\right]$.

8 Repeat steps 2-7 M times

9 A lower $100(1-\alpha)\%$ lower confidence limit for $\min_{t\in(t_1,t_2)}\left[P(NMB_1 \geq t) - P(NMB_2 \geq t)\right]$ corresponds to the α^{th} percentile of the M parametric bootstrap versions of $\min_{t\in(t_1,t_2)}\left[P(NMB_1 \geq t) - P(NMB_2 \geq t)\right]$.

We will be assessing the accuracy of the proposed methods by estimating coverage probabilities attained from each method. The true values will be computed using a parametric bootstrap approach, the steps are provided in the next algorithm.

Algorithm 12: Computing $\min_{t\in(t_1,t_2)}\left[P(NMB_1 \geq t) - P(NMB_2 \geq t)\right]$ via parametric bootstrap

---

1. From the sample of each treatment group compute the estimates $\hat{\mu}_{io}$ and $\hat{\Sigma}_{io}$, $i = 1, 2$.
2. For each $i = 1, 2$, generate B bootstrap samples:
$$\begin{pmatrix} ln[C^*_{ib}] \\ \\ E^*_{ib} \end{pmatrix} \sim N_2\left[\hat{\mu}_{io} \quad , \hat{\Sigma}_{io}\right], b = 1, \ldots, B$$
3. Compute $NMB^*_{ib} = \lambda E^*_{ib} - C^*_{ib}$
4. For each $t \in (t_1, t_2)$ compute: $\gamma(t) = \frac{1}{B}\sum_{b=1}^{B}\left(\left[NMB^*_{1b} \geq t\right] - \left[NMB^*_{2b} \geq t\right]\right)$
5. Compute $\min_{t\in(t_1,t_2)}\gamma(t) = \delta_{NMB}$

We note that even though our methodology is developed for the lognormal-normal model, it can be adopted to the lognormal-lognormal model, or to any model under which the cost and the effectiveness can be individually transformed so that we have bivariate normality for the transformed random variables.

Remark It is clear that for a fixed t, the quantity $P(NMB_1 \geq t) - P(NMB_2 \geq t)$ is a function of parameters in the model (5.1.1). Consequently, fiducial inference concerning $P(NMB_1 \geq t) - P(NMB_2 \geq t)$ is straightforward to develop. Here we would like to note that even without the bivariate normality assumption in (5.1.1), it is possible to develop confidence limits for $P(NMB_1 \geq t) - P(NMB_2 \geq t)$ using the binomial distribution, provided the sample NMBs for the first treatment, namely $NMB_{1j} = \lambda E_{1j} - C_{1j}, j = 1, \ldots, n_1$, are iid, and the same assumption can be made for the samples NMBs for the second treatment, namely $NMB_{2k} = \lambda E_{2k} - C_{2k}, k = 1, \ldots, n_2$. Let $p_1 = P(NMB_1 \geq t)$ and $p_2 = P(NMB_2 \geq t)$, so that the parameter of interest is $p_1 - p_2$. If $B_1 = \#\,(NMB_1 \geq t)$ and $B_2 = \#\,(NMB_2 \geq t)$, then under the iid condition, we have $B_1 \sim Binomial(n_1, p_1)$ and $B_2 \sim Binomial(n_2, p_2)$.

Furthermore, we also have the estimates $\hat{p}_1 = \frac{B_1}{n_1}$ and $\hat{p}_2 = \frac{B_2}{n_2}$. Thus confidence limits for $p_1 - p_2$ can be constructed using a normal approximation for $p_1 - p_2$. It is also possible to develop fiducial inference for $p_1 - p_2$ using fiducial quantities for p_1 and p_2; see Krishnamoorthy and Lee (2010) and Bebu, Mathew, Lachin and Agan (2016). Here we shall not explore this further.

5.3 An example

The following example is obtained from Kruizenga et al. (2005), in which well-nourished and malnourished patients were placed under a new treatment and standard intervention. The aim of the study was the early recognition of malnourishment in hospitalized patients. The study took place from February to June 2003. The new treatment was a 'Short Nutritional Assessment Questionnaire' ('SNAQ') malnutrition screening tool commonly utilized during hospital admissions. The standard treatment was defined as standardized nutritional care protocol. The two effectiveness outcomes are weight gained and number of days in hospital. Positive percentage of weight gain and lower hospital days are associated with the effectiveness of a treatment. Costs are measured in euros (€).

The authors provided summary results for the malnourished patients in the control and intervention group. The summary statistics obtained prior to transformation are reported in Table 5.1. The summary statistics pertaining to the outcomes of percent weight change and cost of dietitian for the new and standard intervention are also provided.

Table 5.1 Raw summary statistics from the malnutrition trial

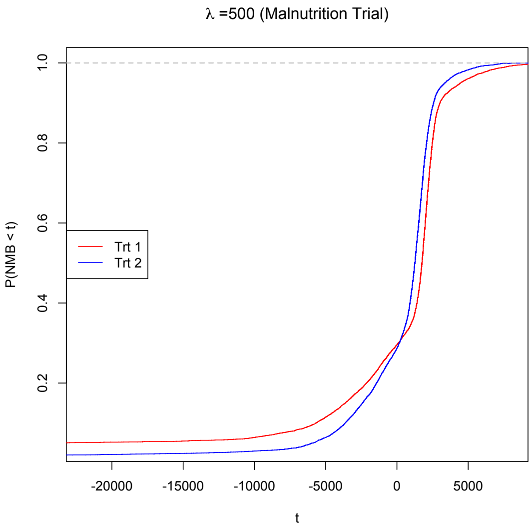
Outcome	New treatment	Standard treatment
E[Cost of Dietitian (€)]	118.2	104.7
SD[Cost of Dietitian (€)]	136.3	174.7
E[Weight change %]	-0.1	-0.3
SD[Weight change %]	7.9	5.9
Sample size	$n_1 = 297$	$n_2 = 291$

The authors note that costs were 'skewed to the right'; see Kruizenga et al. (2005, pg. 1084). Therefore, to illustrate the developed methods, we model cost as lognormal and percent weight change as normal. The mean and standard deviation for the lognormal cost are computed and reported in the following table. The correlation of the cost and effectiveness outcomes were not provided in the paper. However, we used a correlation of $\rho = 0.1$ within our analysis.

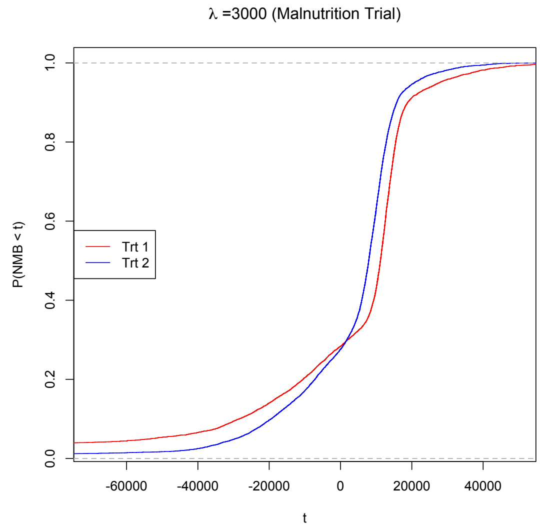
Table 5.2 Summary statistics for the malnutrition trial based on the log-transformed cost

Outcome	Intervention	Control
E[ln(Cost of Dietitian)]	4.349	3.320
SD[ln(Cost of Dietitian)]	0.920	1.154
E[Weight change %]	-0.1	-0.3
SD[Weight change %]	7.9	5.9
Sample size	$n_1 = 297$	$n_2 = 291$
Correlation	$\rho_1 = 0.1$	$\rho_2 = 0.1$

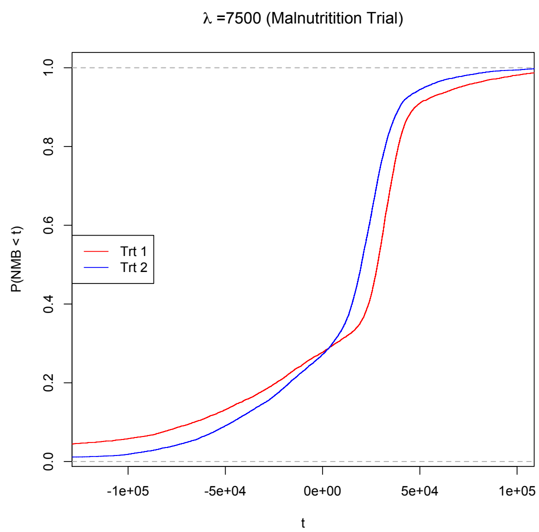
We begin with a visual analysis of the ECDFs under three values of λ. The resulting ECDFs are obtained using using 10000 simulated samples and are provided in Figure 5.1.



(a) $\lambda = 500$



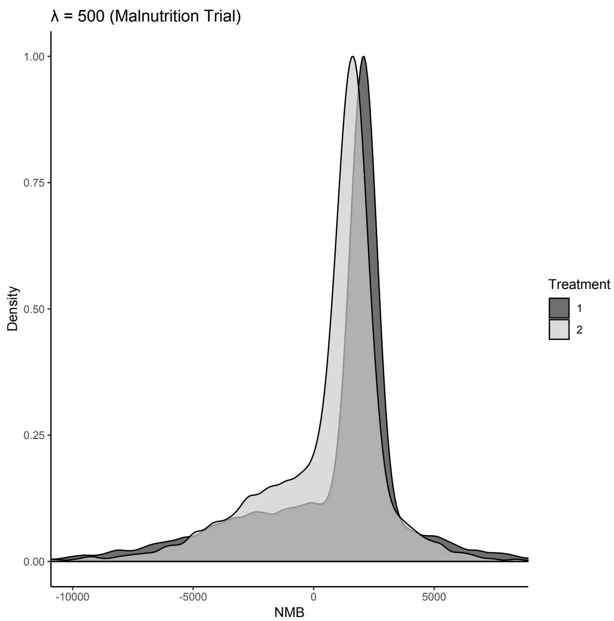
(b) $\lambda = 3000$



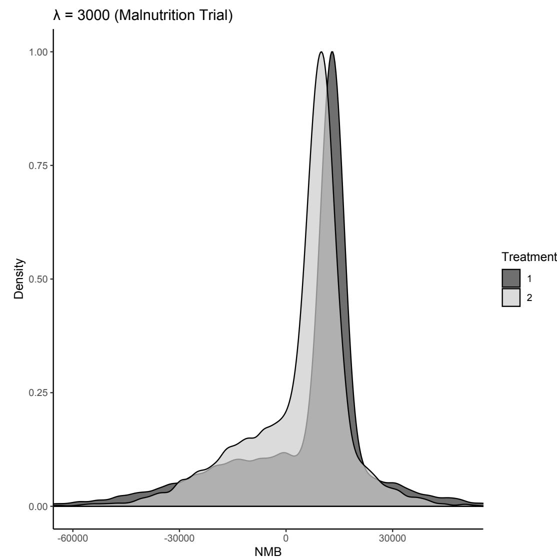
(c) $\lambda = 7500$

Figure 5.1: ECDFs for the malnutrition trial under different values of λ

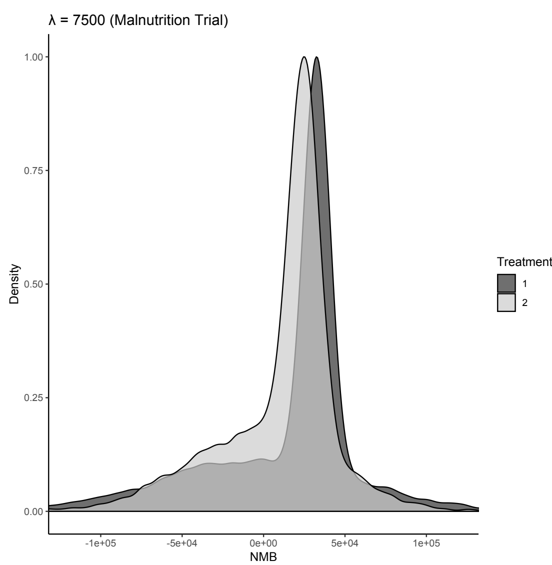
Based on these ECDFs, there are clear points at which the stochastic dominance changes from standard treatment to the new intervention. As the willingness-to-pay increases, the distance between the two distribution functions increases. The density functions are plotted in Figure 5.2 and are also computed using 10000 simulated samples.



(a) $\lambda = 500$



(b) $\lambda = 3000$



(c) $\lambda = 7500$

Figure 5.2: Empirical densities for the malnutrition trial under different values of λ

We will now implement the method developed earlier in this chapter to assess if the new treatment, SNAQ screening, is stochastically dominant over the usual care. The plots indicate that stochastic dominance is likely to hold for a certain range of values, denoted by t.

5.3.1 Lower confidence limits

For each willingness-to-pay value, we used three intervals for t that correspond to patients having low, midrange, and high NMB values. Along with the three values of λ, this results in nine cases for the analysis. Using algorithm 10 and algorithm 11, we estimated the lower confidence limits for the confidence levels 0.95 and 0.90. In order to compute the required minimum over the range for t, we chose 50 values in the range t, and evaluated the difference of the NMB tail probabilities for the two treatment groups. Estimates of the minimum difference between the two probabilities, (5.1.3), obtained using algorithm 12, are provided in Table 5.3.

Table 5.3 Estimated minimum differences for the malnutrition trial for different intervals (t_1, t_2) and for λ = 500, 3000 and 7500

Case ID	(t_1, t_2)	λ	ST	Estimated Min. Diff.
1	(-10000, -2500)	500	2	0.006
2	(1250, 2500)	500	1	0.049
3	(2500, 5000)	500	1	0.056
4	(-45000, -12000)	3000	2	0.024
5	(7500, 15000)	3000	1	0.053
6	(15000, 32500)	3000	1	0.053
7	(-75000, -37500)	7500	2	0.055
8	(18750, 37500)	7500	1	0.053
9	(37500, 87500)	7500	1	0.047

In Table 5.3 the column 'ST' denotes the treatment which appears dominant for a specified interval based on the graphical results of Figure 5.1, such that 1 is the new treatment and 2 is the standard treatment. The estimated differences are positive for all nine cases. The smallest difference occurs for patients with a low NMB range when λ = 500, and the largest difference occurs for patients with the high NMB range under the same willingness-to-pay value. Overall, these differences appear to be close to zero. To determine if the stochastic ordering holds, the lower confidence limits were computed and are given in Table 5.4. We have used the values $M = 2500$ and $K = 5000$ in algorithm 10 and algorithm 11.

In the table, the superscripts F and B denote fiducial method and parametric bootstrap, respectively. Furthermore, $Q^F_{0.05}$ denotes the required 95% lower confidence limit computed by the fiducial approach, and so on.

Table 5.4 Fiducial & percentile bootstrap lower confidence limits for stochastic dominance for the malnutrition trial

Case ID	(t_1, t_2)	λ	ST	$Q^F_{0.05}$	$Q^F_{0.10}$	$Q^B_{0.05}$	$Q^B_{0.10}$
1	(-10000, -2500)	500	2	0.0029	0.0035	0.0028	0.0035
2	(1250, 2500)	500	1	-0.0038	0.0089	-0.0026	0.0089
3	(2500, 5000)	500	1	0.0233	0.0319	0.0239	0.0324
4	(-45000, -12000)	3000	2	-0.0043	0.0078	-0.0036	0.0073
5	(7500, 15000)	3000	1	0.0023	0.0137	0.0023	0.0140
6	(15000, 32500)	3000	1	0.0286	0.0344	0.0263	0.0333
7	(-75000, -37500)	7500	2	0.0076	0.0188	0.0093	0.0183
8	(18750, 37500)	7500	1	0.0001	0.0115	0.0021	0.0126
9	(37500, 87500)	7500	1	0.0245	0.0304	0.0246	0.0305

Based on the results in Table 5.4, it appears that the fiducial and percentile bootstrap lower limits are close. The standard treatment, denoted by 'ST= 2', is stochastically dominant for cases 1, 4, and 7, when using confidence level 0.90. For the smallest and largest λ the stochastic dominance also holds when the confidence level is 0.95. For case 2, the new treatment is not cost-effective at 95% confidence but is cost-effective at 90% confidence. Except for cases 1, 2, 4, and 7, the new treatment is cost-effective, as the lower limits for the minimum difference of stochastic dominance are positive. In particular, these results permit the conclusion that all patients with large NMB values benefit using treatment one instead of treatment two. The results also indicate that patients with lower NMB values tend to benefit more when taking the standard treatment as compared to the new treatment.

In addition, the new treatment tends to be cost-effective for patients in the middle range as well. From Figure 5.1 and Figure 5.2, we notice that as λ increases, the distance between the ECDFs and empirical density functions appear to be increasing as well. The results in Table 5.4 support this observation, as comparison of the 90% lower confidence limits from the percentile bootstrap approach for each type of patient (low, mid-range, and large NMBs) has increasing limits as a function of λ (with exception of patients with larger NMB values). This result cannot be stated as strongly for the other three lower limits, as some decrease for certain values of λ, depending on the range of the NMB.

5.3.2 Coverage probabilities and expected lower limits

In order to assess the accuracy of the lower confidence limits obtained using the fiducial and percentile bootstrap approaches, we estimated the coverage probabilities using 5000 simulated samples. We also chose $M = 1000$, $K = 5000$ in the algorithms. The lognormal-normal model was used to generate the data, and the parameter estimates given in Table 5.2 were used as the true values.

Table 5.5 Coverage probabilities of the lower confidence limits obtained by the fiducial & percentile bootstrap methods; $n_1 = 297, n_2 = 291$

Case ID	$Q^F_{0.05}$	$Q^F_{0.10}$	$Q^B_{0.05}$	$Q^B_{0.10}$
1	0.960	0.914	0.962	0.910
2	0.947	0.900	0.945	0.893
3	0.959	0.915	0.960	0.916
4	0.969	0.931	0.971	0.933
5	0.946	0.900	0.945	0.893
6	0.959	0.916	0.960	0.915
7	0.968	0.934	0.969	0.935
8	0.946	0.901	0.945	0.893
9	0.959	0.914	0.961	0.917

The coverage probability results are mostly similar for both approaches; however, the fiducial method appears to provide coverage probabilities closer to the nominal level, compared to the percentile bootstrap method. The cases under analysis also have an effect on the resulting coverage probabilities, with some cases having conservative coverage probabilities. For example, coverage probabilities for cases 4 and 7 tend to be the most conservative compared to other cases. These two cases correspond to patients with lower

NMB values when standard treatment is stochastically dominant.

The two methods can be further compared based on the expected lower limits. The results of are provided in Table 5.6. Larger lower limits indicate higher precision.

Table 5.6 Expected values of the lower confidence limits using fiducial & percentile bootstrap methods; $n_1 = 297, n_2 = 291$

Case ID	$Q^F_{0.05}$	$Q^F_{0.10}$	$Q^B_{0.05}$	$Q^B_{0.10}$
1	-0.0045	-0.0008	-0.0045	-0.0008
2	-0.0031	0.0082	-0.0025	0.0088
3	0.0208	0.0289	0.0213	0.0294
4	-0.0080	0.0000	-0.0082	-0.0001
5	0.0006	0.0120	0.0012	0.0126
6	0.0219	0.0293	0.0223	0.0297
7	0.0071	0.0168	0.0068	0.0166
8	0.0011	0.0124	0.0017	0.0130
9	0.0198	0.0264	0.0202	0.0268

The results in Table 5.6 indicate that the expected lower limits tend to be larger under the percentile bootstrap approach. However, the differences between the expected lower limits appear slight. To further investigate the accuracy between the two methods we provide results for small sample sizes, $n_1 = n_2 = 50$.

Table 5.7 Coverage probabilities of the lower confidence limits obtained by the fiducial & percentile bootstrap methods; $n_1 = n_2 = 50$

Case ID	$Q^F_{0.05}$	$Q^F_{0.10}$	$Q^B_{0.05}$	$Q^B_{0.10}$
1	0.972	0.940	0.987	0.959
2	0.953	0.909	0.954	0.907
3	0.967	0.933	0.977	0.941
4	0.980	0.952	0.986	0.963
5	0.952	0.908	0.955	0.908
6	0.969	0.934	0.978	0.946
7	0.972	0.937	0.972	0.940
8	0.952	0.908	0.955	0.908
9	0.971	0.936	0.979	0.949

Table 5.8 Expected values of the lower confidence limits using fiducial & percentile bootstrap methods; $n_1 = n_2 = 50$

Case ID	$Q^F_{0.05}$	$Q^F_{0.10}$	$Q^B_{0.05}$	$Q^B_{0.10}$
1	−0.0660	−0.0466	−0.0623	−0.0439
2	−0.0753	−0.0482	−0.0780	−0.0509
3	−0.0431	−0.0212	−0.0440	−0.0228
4	−0.0804	−0.0574	−0.0771	−0.0546
5	−0.0720	−0.0448	−0.0748	−0.0475
6	−0.0414	−0.0201	−0.0425	−0.0218
7	−0.0632	−0.0395	−0.0590	−0.0359
8	−0.0716	−0.0444	−0.0744	−0.0471
9	−0.0421	−0.0214	−0.0432	−0.0232

Based on the small sample size results presented in Table 5.7 and Table 5.8 it is clear that the coverage probabilities are generally conservative when using the fiducial and percentile bootstrap methodologies. However, under these sample sizes the fiducial approach does appear to be slightly less conservative than the percentile bootstrap approach. An important point to note is that all of the expected lower confidence limits are negative under these small sample sizes, when they were previously largely positive when utilizing the larger sample sizes (original sample sizes $n_1 = 297, n_2 = 291$). Therefore, it is clear that our methodology performs satisfactorily for larger sample sizes, but not for small sample size scenarios.

5.4 Discussion

Decision making based on stochastic dominance has been under-utilized in the CEA literature. In this chapter, we have develop a framework for doing so, by considering the stochastic dominance of the distributions of NMBs from competing treatments. The methodology is developed and illustrated under a lognormal-normal model. As already noted, the methodology can be modified for other distributions that can be transformed to normality, at least approximately. For example, if the cost data are distributed as gamma, a cube-root transformation can be used to achieve approximate normality (referred to as the Wilson-Hilferty transformation). Application of our methodology does require the specification of a range over which stochastic domination of the NMB is required to hold.

The analysis we have carried out is under a model without covariates; that is, it is an aggregate level analysis. The approach can be individualized by considering the bivariate regression model that was investigated elsewhere in the thesis.

5.5 Appendix- Further simulation results

In this Appendix, we shall demonstrate our approach by considering six cases of cost and effectiveness outcomes where transformations have to be applied in order to achieve normality or approximate normality. The parameters used for the simulation analysis are provided in Table 5.9. The parameter values reported are for the parameters of the bivariate normal distribution after applying transformations (if necessary) to the cost and/or effectiveness.

Table 5.9 Treatment parameters for simulation analysis

Treatment	μ_{iC}	μ_{iE}	σ_{iC}	σ_{iE}	ρ
1	11.439	4.996	0.097	0.022	0.100
2	10.618	4.598	0.083	0.018	0.100

The six cases considered for this analysis are provided in Table 5.10, sample sizes used are $n_1 = n_2 = 200$.

Table 5.10 The six cases for simulation analysis

Case ID	Cost Dist.	Effect. Dist.
1	Gamma	Lognormal
2	Gamma	Gamma
3	Gamma	Normal
4	Lognormal	Lognormal
5	Lognormal	Gamma
6	Lognormal	Normal

A set of data corresponding to each of the six cases in Table 5.10 were simulated using a bivariate normal model similar to (5.1.1), with modification to account for the specific transformations used. In particular, we point out that the transformation of the outcomes with marginal gamma distributions is the cube-root transformation. For the six cases of the cost and effectiveness outcomes, the following are the expressions for the NMB tail probabilities computed in accordance with (5.2.3).

Case 1 :

$$\begin{aligned} &Y_1 = C^{1/3}; Y_2 = ln(E) \\ &P(NMB \geq t) = 1 - E_{Y_1}\left\{\Phi\left[\frac{ln\left[\frac{t+y_1^3}{\lambda}\right] - \mu_E - \rho\frac{\sigma_E}{\sigma_C}(y_1 - \mu_C)}{\sqrt{(1-\rho^2)\sigma_E^2}}\right]\right\} \end{aligned} \tag{5.5.1}$$

Case 2:

$$\begin{aligned} &Y_1 = C^{1/3}; Y_2 = E^{1/3} \\ &P(NMB \geq t) = 1 - E_{Y_1}\left\{\Phi\left[\frac{\left[\frac{t+y_1^3}{\lambda}\right]^{1/3} - \mu_E - \rho\frac{\sigma_E}{\sigma_C}(y_1 - \mu_C)}{\sqrt{(1-\rho^2)\sigma_E^2}}\right]\right\} \end{aligned} \tag{5.5.2}$$

Case 3 :

$$Y_1 = C^{1/3}; Y_2 = E$$
$$P(NMB \geq t) = 1 - E_{Y_1}\left\{\Phi\left[\frac{[\frac{t+y_1^3}{\lambda}] - \mu_E - \rho\frac{\sigma_E}{\sigma_C}(y_1 - \mu_C)}{\sqrt{(1-\rho^2)\sigma_E^2}}\right]\right\} \tag{5.5.3}$$

Case 4:

$$Y_1 = ln(C); Y_2 = ln(E)$$
$$P(NMB \geq t) = 1 - E_{Y_1}\left\{\Phi\left[\frac{ln\left[\frac{t+e^{y_1}}{\lambda}\right] - \mu_E - \rho\frac{\sigma_E}{\sigma_C}(y_1 - \mu_C)}{\sqrt{(1-\rho^2)\sigma_E^2}}\right]\right\} \tag{5.5.4}$$

Case 5 :

$$Y_1 = ln(C); Y_2 = E^{1/3}$$
$$P(NMB \geq t) = 1 - E_{Y_1}\left\{\Phi\left[\frac{\left[\frac{t+e^{y_1}}{\lambda}\right]^{1/3} - \mu_E - \rho\frac{\sigma_E}{\sigma_C}(y_1 - \mu_C)}{\sqrt{(1-\rho^2)\sigma_E^2}}\right]\right\} \tag{5.5.5}$$

Case 6:

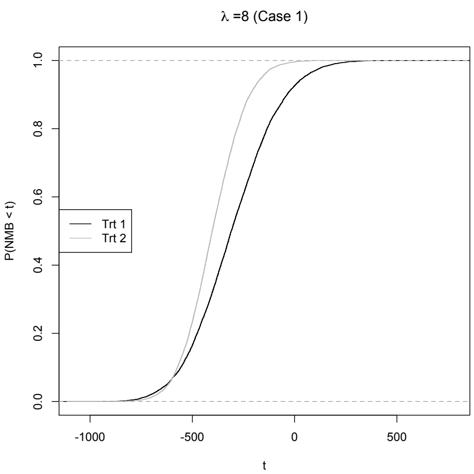
$$Y_1 = ln(C); Y_2 = E$$
$$P(NMB \geq t) = 1 - E_{Y_1}\left\{\Phi\left[\frac{[\frac{t+e^{y_1}}{\lambda}] - \mu_E - \rho\frac{\sigma_E}{\sigma_C}(y_1 - \mu_C)}{\sqrt{(1-\rho^2)\sigma_E^2}}\right]\right\} \tag{5.5.6}$$

We apply algorithm 10 and algorithm 11 to the six cases in Table 5.10. For each of the six cases, there will be nine estimates of the lower confidence limits corresponding to three values of λ and three intervals specified for t. The three intervals for t represent low, mid-range, and high NMB values, and were selected based on ECDFs. The ECDFs plotted in this section are based on 10000 simulations.

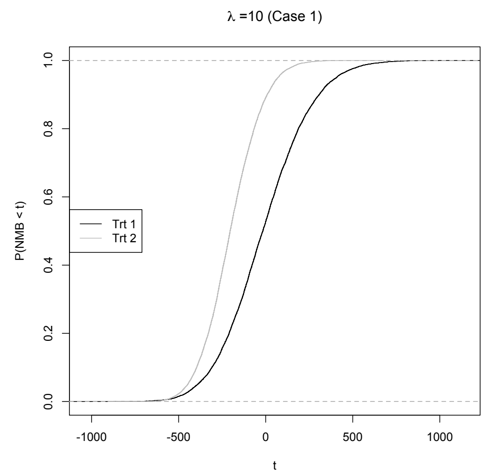
All the cases demonstrate that as the willingness-to-pay parameter increases the stochastic dominance shifts from the second treatment to the first treatment. Each case in

Table 5.10 is evaluated for cost-effectiveness using three values of λ. The values of λ were chosen on a case by case basis to determine when the stochastic dominance changes from the second treatment to the first treatment.

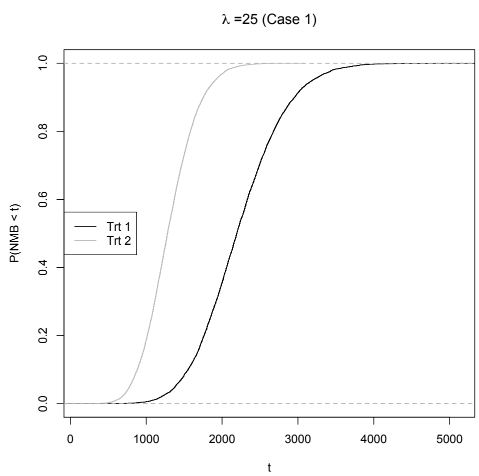
Each table in this section corresponds to one of the six cases provided in Table 5.10. The analysis of these cases indicates that overall the percentile bootstrap and fiducial methods provide lower confidence limits for stochastic dominance that are similar. Each table contains the following columns: Interval, λ, Increment, ST, $Q^F_{0.05}$, $Q^F_{0.10}$, $Q^B_{0.05}$, and $Q^B_{0.10}$. The Interval column indicates the interval of NMB values of interest. The distance between consecutive t values is denoted by 'Increment'. The stochastically dominant treatment is denoted by 'ST' (based on ECDFs). The 95% and 90% lower confidence limits obtained by the fiducial method are denoted as $Q^F_{0.05}$ and $Q^F_{0.10}$, respectively. The 95% and 90% lower confidence limits obtained by the percentile bootstrap method are denoted as $Q^B_{0.05}$ and $Q^B_{0.10}$, respectively.



(a) $\lambda = 8$



(b) $\lambda = 10$



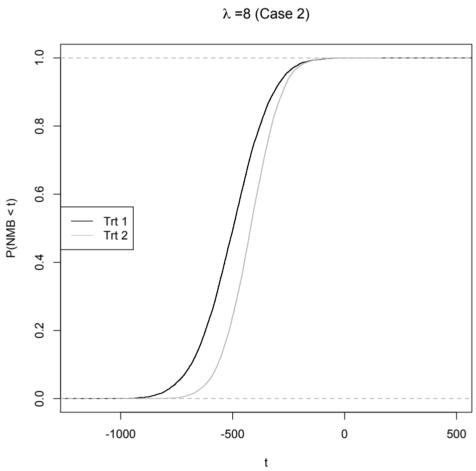
(c) $\lambda = 25$

Figure 5.3: ECDFs for case 1 under different values of λ

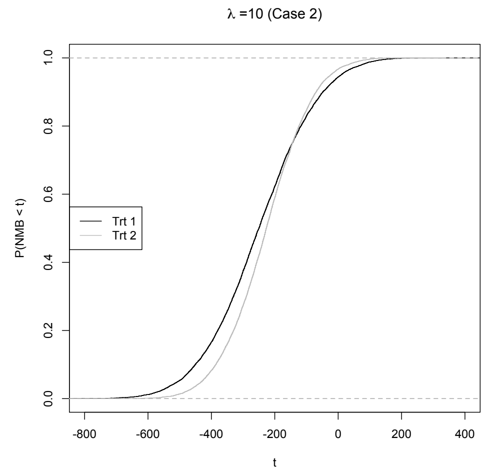
Table 5.11 Fiducial & percentile bootstrap lower limits for case 1 for different intervals (t_1, t_2) and for $\lambda = 8$, 10 and 25

Case ID	(t_1, t_2)	λ	Increment	ST	$Q^F_{0.05}$	$Q^F_{0.10}$	$Q^B_{0.05}$	$Q^B_{0.10}$
1	(-500, -333)	8	25	1	0.008	0.019	0.009	0.020
1	(-333, -167)	8	25	1	0.156	0.165	0.154	0.165
1	(-167, 0.00)	8	25	1	0.057	0.063	0.056	0.062
1	(-300, -67)	10	25	1	0.105	0.115	0.103	0.115
1	(-67, 167)	10	25	1	0.184	0.192	0.182	0.190
1	(167, 400)	10	25	1	0.041	0.045	0.039	0.043
1	(10007, 1667)	25	50	1	0.147	0.154	0.144	0.153
1	(1667, 2333)	25	50	1	0.365	0.374	0.363	0.372
1	(2333, 3000)	25	50	1	0.072	0.077	0.071	0.071

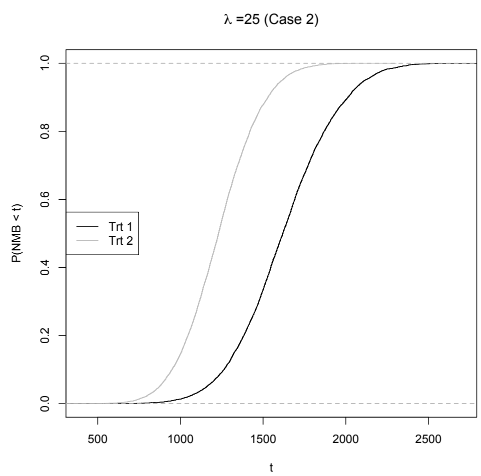
The results for Table 5.11 pertain to the case in which the cost has a gamma distribution and the effectiveness is lognormally distributed. Based on the ECDF results in Figure 5.3, one expects that lower limits should be positive for all interval and λ combinations. The results in the above table are in agreement with this visual assessment. Hence, for this analysis the first treatment is stochastically dominant for all patients under the given values of λ.



(a) $\lambda = 8$



(b) $\lambda = 10$



(c) $\lambda = 25$

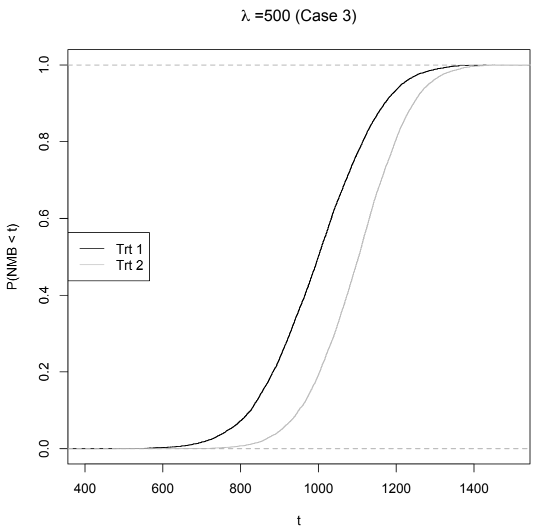
Figure 5.4: ECDFs for case 2 under different values of λ

Table 5.12 Fiducial & percentile bootstrap lower limits for case 2 for different intervals (t_1, t_2) and for $\lambda = 8$, 10 and 25

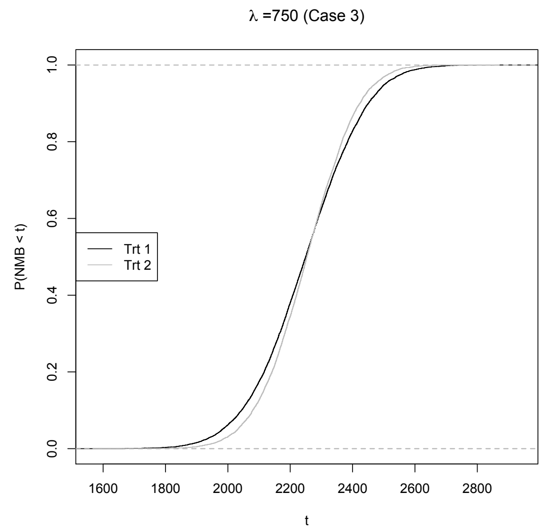
Case ID	(t_1, t_2)	λ	Increment	ST	$Q^F_{0.05}$	$Q^F_{0.10}$	$Q^B_{0.05}$	$Q^B_{0.10}$
2	(-600, -533)	8	10	2	0.138	0.147	0.138	0.148
2	(-533, -467)	8	10	2	0.185	0.197	0.182	0.196
2	(-467, -400)	8	10	2	0.131	0.146	0.135	0.151
2	(-500, -375)	10	25	2	0.022	0.027	0.022	0.027
2	(-375, -250)	10	25	2	0.013	0.027	0.015	0.030
2	(-50, 50)	10	25	1	-0.011	-0.004	-0.013	-0.005
2	(1000, 1333)	25	50	1	0.101	0.108	0.099	0.105
2	(1333, 1667)	25	50	1	0.386	0.397	0.386	0.397
2	(16677, 2000)	25	50	1	0.095	0.101	0.093	0.098

Case 2 is the scenario where the cost and effectiveness outcomes each follow a gamma distribution. The ECDFs in Figure 5.4 corresponding to this case indicate that the second treatment group is stochastically dominant for most NMB values when $\lambda = 8, 10$. The results in the proceeding table indicate that this dominance holds. For the interval (-50, 50) and $\lambda = 10$ the lower limits are negative, indicating that treatment one is not stochastically dominant. Inspecting the ECDFs it is clear that the distance between the distribution functions for this interval is small. The stochastic dominance of treatment one for the largest λ value holds, as one would anticipate from a visual inspection of the ECDFs. Hence, the second treatment is recommended for all patients when using the lower values of λ. The first treatment is recommended for all patients when the value of λ is increased to 25.

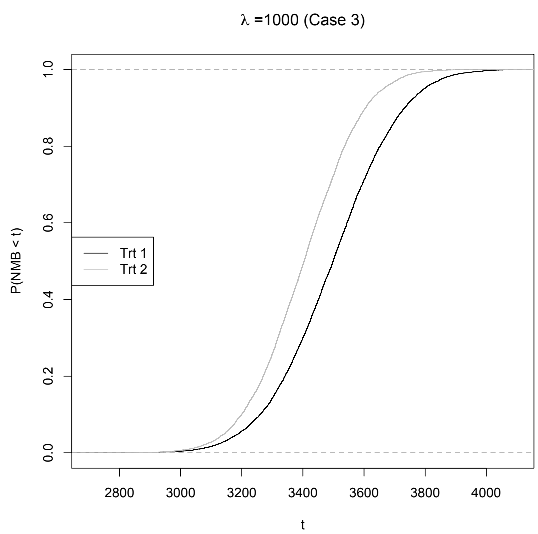
The ECDF for case 3, under three values of λ are provided in the next figure.



(a) $\lambda = 500$



(b) $\lambda = 750$



(c) $\lambda = 1000$

Figure 5.5: ECDFs for case 3 under different values of λ

The results obtained for case 3 are reported in Table 5.13.

Table 5.13 Fiducial & percentile bootstrap lower limits for case 3 for different intervals (t_1, t_2) and for $\lambda = 500$, 750 and 1000

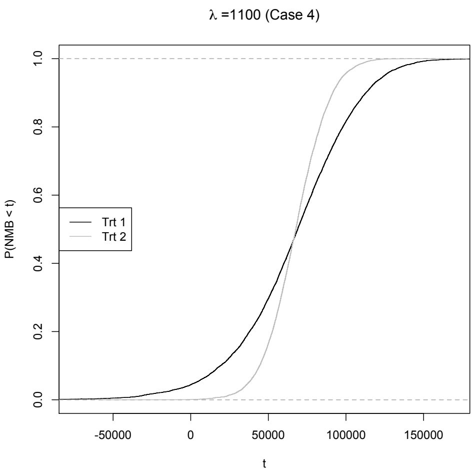
Case ID	(t_1, t_2)	λ	Increment	ST	$Q^F_{0.05}$	$Q^F_{0.10}$	$Q^B_{0.05}$	$Q^B_{0.10}$
3	(800, 942)	500	25	2	0.050	0.054	0.048	0.053
3	(942, 1083)	500	25	2	0.200	0.209	0.199	0.209
3	(1083, 1225)	500	25	2	0.067	0.076	0.068	0.076
3	(2000, 2075)	750	10	2	0.001	0.008	−0.001	0.007
3	(2075, 2150)	750	10	2	−0.017	−0.004	−0.016	−0.003
3	(2400, 2550)	750	10	1	−0.010	−0.000	−0.011	0.000
3	(3200, 3400)	1000	50	1	0.017	0.024	0.016	0.024
3	(3400, 3600)	1000	50	1	0.127	0.138	0.127	0.139
3	(3600, 3800)	1000	50	1	0.026	0.028	0.024	0.028

In all cases in Table 5.13, the fiducial and percentile bootstrap lower limits are relatively close. For case 3 and $\lambda = 500$, the standard treatment is stochastically dominant for low, mid-range, and high NMB values. This indicates that this treatment is cost-effective for all patients.

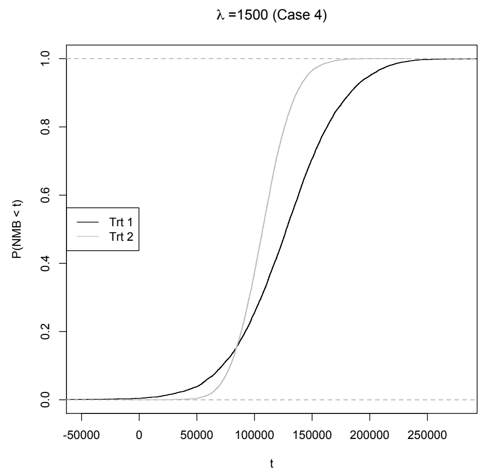
For $\lambda = 750$, treatment two is stochastically dominant for patients with lower NMB values, when using the fiducial approach, but not so for the percentile bootstrap results. In particular, the lower 95% confidence limit is slightly greater than 0, 0.001, while the percentile bootstrap approach is slightly negative, -0.001. However, the 90% lower confidence limits are positive for both approaches. The lower limits for patients with mid-range NMB values are negative for both methods and confidence levels, indicating the second treatment is not cost-effective for these patients. The lower limits for patients with high values of NMB at $\lambda = 750$ are negative when using a confidence level of 0.95 and almost equal zero under a confidence level of 0.90.

When $\lambda = 1000$, the new treatment dominates as the lower limits are positive. Analysing the results for this case leads to the following conclusions: (i) what visually appears to be stochastically dominant based on ECDFs may not be statistically so; and (ii) choice of confidence level may influence the results.

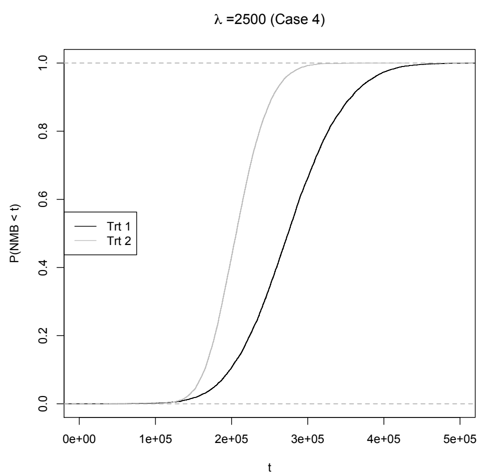
Next we consider case 4.



(a) $\lambda = 1100$



(b) $\lambda = 1500$



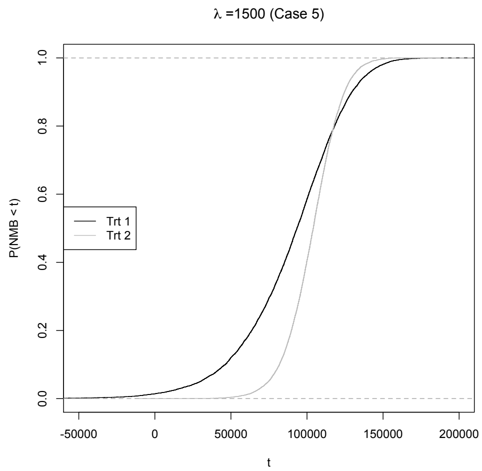
(c) $\lambda = 2500$

Figure 5.6: ECDFs for case 4 under different values of λ

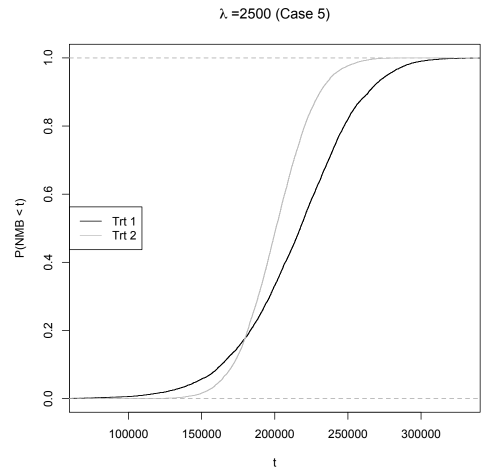
Table 5.14 Fiducial & percentile bootstrap lower limits for case 4 for different intervals (t_1, t_2) and for $\lambda = 1100$, 1500 and 2500

Case ID	(t_1, t_2)	λ	Increment	ST	$Q^F_{0.05}$	$Q^F_{0.10}$	$Q^B_{0.05}$	$Q^B_{0.10}$
4	(0, 25000)	1100	500	2	0.030	0.033	0.029	0.032
4	(25000, 50000)	1100	500	2	0.077	0.086	0.075	0.084
4	(75000, 112000)	1100	500	1	0.026	0.041	0.027	0.040
4	(40000, 65000)	1500	200	2	0.015	0.017	0.014	0.016
4	(100000, 145000)	1500	200	1	0.053	0.066	0.051	0.064
4	(145000, 190000)	1500	200	1	0.049	0.053	0.046	0.051
4	(175000, 241667)	2500	200	1	0.090	0.099	0.089	0.097
4	(241667, 308333)	2500	200	1	0.241	0.249	0.236	0.246
4	(308333, 375000)	2500	200	1	0.042	0.044	0.039	0.044

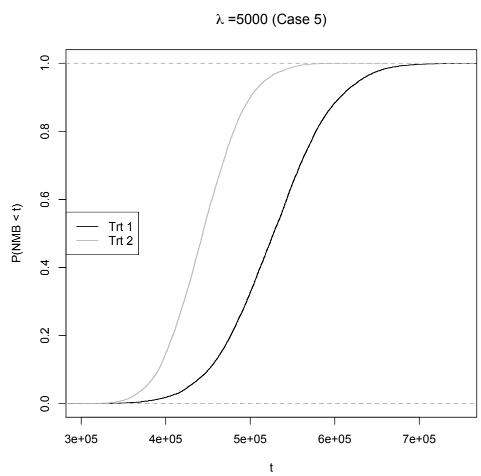
The results of Table 5.14 are based on cost and effectiveness measures that are each lognormally distributed. Based on the ECDFs in Figure 5.6 it appears that the second treatment is stochastically dominant for patients with lower NMB values when $\lambda = 1100, 1500$. Patients with mid-range NMB values attain cost-effectiveness under the second treatment when $\lambda = 1100$, and achieve cost-effectiveness under the first treatment when $\lambda = 1500, 2500$. For $\lambda = 2500$ the first treatment appears to be stochastically dominant for all patients. All of the lower limits are positive in Table 5.14 confirming the visual inspection.



(a) $\lambda = 1500$



(b) $\lambda = 2500$



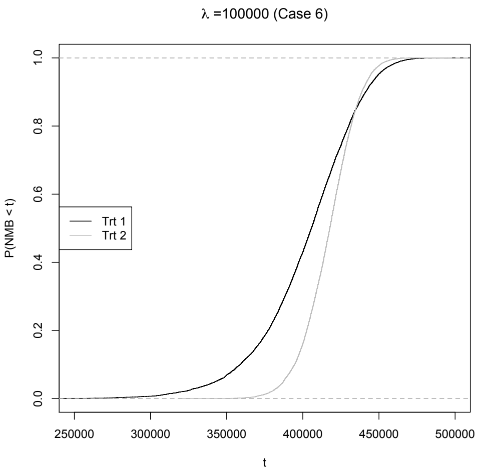
(c) $\lambda = 5000$

Figure 5.7: ECDFs for case 5 under different values of λ

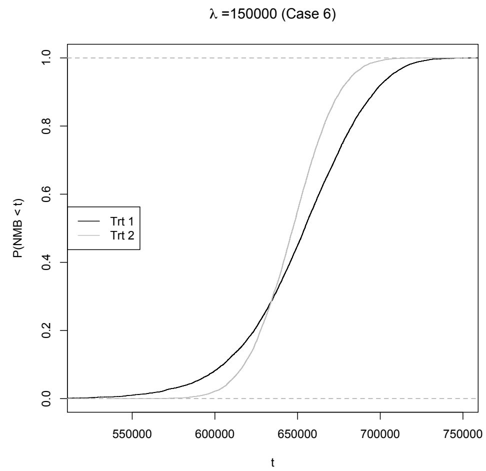
Table 5.15 Fiducial & percentile bootstrap lower limits for case 5 for different intervals (t_1, t_2) and for $\lambda = 1500$, 2500 and 5000

Case ID	(t_1, t_2)	λ	Increment	ST	$Q^F_{0.05}$	$Q^F_{0.10}$	$Q^B_{0.05}$	$Q^B_{0.10}$
5	(37500, 68750)	1500	200	2	0.049	0.053	0.048	0.052
5	(68750, 100000)	1500	200	2	0.116	0.134	0.120	0.132
5	(131250, 140000)	1500	200	1	0.014	0.020	0.013	0.019
5	(137500, 167000)	2500	200	2	0.001	0.010	0.001	0.010
5	(200000, 231000)	2500	200	1	0.085	0.101	0.087	0.103
5	(231000, 262000)	2500	200	1	0.075	0.081	0.075	0.081
5	(400000, 466667)	5000	400	1	0.100	0.106	0.096	0.103
5	(466667, 533333)	5000	400	1	0.387	0.398	0.387	0.398
5	(533333, 600000)	5000	400	1	0.091	0.097	0.089	0.096

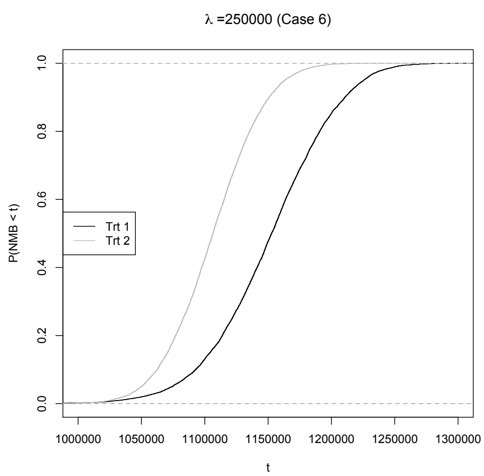
Case 5 corresponds to the scenario where the cost is lognormally distributed and the effectiveness is gamma distributed. The resulting ECDFs in Figure 5.7 indicate that the second treatment is stochastically dominant when $\lambda = 1500$ and the NMB values are not high. For $\lambda = 2500$ the first treatment is stochastically dominant for all patients, except those with low NMB values. When $\lambda = 5000$ the first treatment is stochastically dominant for all patients. Because the resulting lower limits in Table 5.15 are all positive the preceding visual assessment holds.



(a) $\lambda = 100000$



(b) $\lambda = 150000$



(c) $\lambda = 250000$

Figure 5.8: ECDFs for case 6 under different values of λ

Table 5.16 Fiducial & percentile bootstrap lower limits for case 6 for different intervals (t_1, t_2) and for λ = 100000, 150000 and 250000

Case ID	(t_1, t_2)	λ	Increment	ST	$Q^F_{0.05}$	$Q^F_{0.10}$	$Q^B_{0.05}$	$Q^B_{0.10}$
6	(350000, 381250)	100 000	200	2	0.049	0.053	0.047	0.051
6	(381250, 412500)	100 000	200	2	0.143	0.158	0.145	0.158
6	(440000, 460000)	100 000	200	1	−0.020	−0.012	−0.021	−0.013
6	(575000, 612500)	150 000	200	2	0.016	0.019	0.015	0.018
6	(662000, 681000)	150 000	100	1	0.093	0.104	0.093	0.104
6	(681000, 700000)	150 000	100	1	0.047	0.051	0.045	0.050
6	(1080000, 1120000)	250 000	200	1	0.109	0.118	0.108	0.118
6	(1120000, 1160000)	250 000	200	1	0.320	0.332	0.324	0.335
6	(1160000, 1200000)	250 000	200	1	0.109	0.115	0.106	0.112

The last case for the simulation analysis occurs when the cost and effectiveness outcomes follow a lognormal and normal distribution, respectively. The visual analysis of the ECDFs in Figure 5.8 indicates that when $\lambda = 100000$ the second treatment is stochastically dominant for all patients, with the exception of those with very high NMB values. For $\lambda = 150000$ the first treatment dominates when the NMB values are midrange and large, in contrast, lower NMB values result in the second treatment dominating. When λ is increased to 250000 the first treatment is stochastically dominant or all patients. The fiducial and percentile bootstrap results in Table 5.16 are in agreement regarding dominance for all intervals and values of λ. The visual analysis discussed for this case appears to hold as all lower confidence limits are positive, with one exception. When $\lambda = 100000$ the patients with the highest NMB values, corresponding to the interval, (440000, 460000), do not achieve cost-effectiveness under treatment one. For this interval Figure 5.8 shows that the distance between the ECDFs is very slight, hence the resulting lack of dominance is not surprising.

Chapter 6

Individualized cost-effectiveness analysis for multi-center trials

The literature on cost-effectiveness analysis has been dominated by methods that focus on randomized controlled trials (RCTs). However, outcomes from multi-center trials and cluster-randomized trials have also been considered. This chapter focuses on cost-effectiveness methods for data obtained from multi-center trials. Such trials are clinical trials conducted over multiple centers, where patients in a particular center are randomized to one of two treatments. Therefore, multi-center trials have a hierarchical structure, and patients within the same center could have correlated outcomes. Thus, multi-center trials have two types of variabilities: (i) between-patient variability within a center, and (ii) between-center variability. The first variability is a measure of heterogeneity amongst patients. Similarly, the latter is a measure of heterogeneity amongst centers. Accordingly, analysis of multi-center trial data must account for both types of variations.

The variation amongst centers may stem from multiple sources. Petri, M., et al. (2005) point out that in the context of health economics analysis such trials often exhibit heterogeneity of cost outcomes across centers. Particularly, the authors highlight that even though protocol surrounding treatment administration can be standardized across centers, costs often vary depending on location of centers and organizational aspects. In addition, Manju, Candel and Berger (2015) note that costs and medical resources associated with treatments may vary across centers. In essence, the distribution among outcomes may differ across centers and such differences should be accounted for.

We note that related literature on multi-center trials include analysis of correlated and clustered data. In particular, some literature refer to centers as clusters. However, cluster-randomized trials and multi-center trials differ in the randomization unit, an important distinction. Cluster-randomized trials (CRTs) most often refer to clinical trials in which the randomization of treatment occurs at the cluster-level. In contrast, in multi-center trials the randomization of treatment occurs at the patient-level. Therefore, in CRTs patients in the same cluster all receive the same treatment; whereas, in multi-center trials, patients within the same center may receive different treatments. To highlight this distinction, throughout this work we refer to this effect as center-effect instead of cluster-effect. The differences and similarities between these two trial designs are noted in Moerbeek, van Breukelen and Berger (2003) and Manca et al. (2005). In particular, Manca et al. (2005) note:

> In trials that randomise by location rather than by patient (i.e. cluster-randomised trials), the hierarchical nature of the data available for economic analysis is an inevitable implication of the design of the study. However, at least for economic analysis, some degree of clustering is also likely to exist in trials where the patient is the 'unit of randomisation' due to variation between locations in clinical and economic parameters. (p.474)

Numerous models and methods have been proposed for inference based on multi-center data, highlighting the importance of accounting for between-center heterogeneity. Ignoring this source of variation may have several consequences, one being that results may indicate more substantial differences between treatments than what truly exist (Petri, M., et al. (2005)). The approach used in the literature consists regression models of various varieties to model the outcomes. One approach involves synthesizing summary statistics from the centers to obtain information about each treatment. This analysis can be classified as an aggregate meta-analysis approach. Our work focuses on assessment of treatments using patient-level and center-level information, which may be classified as an individual patient data meta-analysis. Two general approaches are utilized to model outcomes associated with multi-center trials: (i) fixed-effects model, and (ii) mixed-linear models that include random-effect(s) at the center-level. Associated literature has compared such methods, their advantages and limitations. Such comparisons are included in Kahan and Harhay (2015), Basagana et al. (2018), and Moerbeek, van Breukelen and Berger (2003). Bayesian approaches have also been adopted; see Thompson, Turner and Warn (2001).

The simplest models are fixed effect regression models that incorporate both patient-level and center-level characteristics through covariates. Typically, a single center-effect is incorporated in the model using a dummy variable. These models are straightforward to analyze as they only utilize sampling error at the patient-level. However, analysis under the fixed-effects regression framework omits center-specific random effects. In Manca et al. (2005) the authors note how such models ignore important correlation amongst trials:

> The key implication of clustering in economic data in multi-centre and multinational trials is that the cost-effectiveness of the interventions of interest may vary between locations. Most trial- based economic studies of this type, however, ignore this potential source of variability. (p.472)

In the context of multi-center trials, models that include center-level random effects are referred to as multi-level models (MLMs) or hierarchical linear models. One advantage of MLMs is that these models utilize within and between center information. In particular, the analysis accounts for the dependence of outcomes from patients nested within the same center while treating the patient as the unit of analysis. The inclusion of a center-level random effect allows for generalizability of results to similar centers that are not included in the analysis. Under this general framework of MLMs, authors have proposed univariate and multi-variate models. The net monetary benefit (NMB), considered earlier in the thesis, is modelled directly using an MLM approach in Manca et al. (2005). Their regression model utilizes random-slopes and random-intercepts to account for center-level heterogeneity, and models the NMB directly. The approach presented in this chapter utilizes the MLM framework in which we model patient-level outcomes directly, as we have done earlier in the thesis.

In addition to proposing models for outcomes from multi-center trials, authors have also compared various estimation procedures for such models. These methods include ordinary least squares (OLS), full maximum likelihood (MLE), and restricted maximum likelihood (REML). Of these three estimation methods, the REML is utilized in our work as it provides estimates that have reduced bias compared to that of the MLE approach. In Goldstein (1986) the author shows that the iterative generalised least squares procedure (IGLS) results in estimates that are equivalent to the maximum likelihood estimates for normal outcomes. In Goldstein (1989) and Goldstein (2011) the author demonstrates that using the restricted iterative generalized least squares procedure (RIGLS) results in estimates equivalent to the restricted maximum likelihood estimates for normally distributed outcomes. The RIGLS procedure only requires a simple modification to the residuals. A book-length discussion on MLMs and the associated estimation procedures is available in Goldstein (2011) .

6.1 Multi-level model and estimation

Suppose bivariate data on the cost and effectiveness are available on two treatments from g centers, and let n_{jk} denote the number of patients assigned to the k^{th} treatment at the g^{th} center; $k = 1,\ 2,\ j = 1, \ldots, g$; and $i = 1, \ldots n_{jk}$. Let C^i_{jk} and E^i_{jk}, respectively, denote the cost and effectiveness measures for the i^{th} patient assigned to the k^{th} treatment at the j^{th} center. We shall use the lognormal-lognormal model. The bivariate outcome,

$X_{jk}^i = (ln(C_{jk}^i), ln(E_{jk}^i))'$, is modelled as

$$X_{jk}^i = \begin{pmatrix} ln(C_{jk}^i) \\ ln(E_{jk}^i) \end{pmatrix} = \gamma_j + B_k^T w_{jk}^i + \epsilon_{jk}^i + u_j; k = 1, 2, j = 1, \ldots, g; i = 1, \ldots n_{jk} \tag{6.1.1}$$

where $\gamma_j = (\gamma_j^C, \gamma_j^E)^T$ is a center-specific intercept, w_{jk}^i is a subject-level $p \times 1$ fixed effects vector (covariates), B_k is a $p \times 2$ matrix of unknown parameters, u_j is a center-level random effect, and ϵ_{jk}^i is the subject-specific random error. We shall denote the number of patients in center j as $n_{j.}$, and $n_{.k}$ will denote the number of subjects assigned to treatment k among all the centers. In addition, we shall write $N = \sum_{k=1}^2 \sum_{j=1}^g n_{jk} = n_{.1} + n_{.2}$.

Model (6.1.1) differs from other cited MLMs in two key aspects: (i) inclusion of treatment-specific regression parameters, and (ii) specification of the covariance structure. Typically, MLMs in the literature have simply included treatment as a fixed effect dummy variable and patient-level covariates are incorporated to estimate average outcomes at the center-level. However, our model avoids the assumption that patient characteristics affect outcomes of differing treatments in the same manner. Further, as seen in (6.1.2) we also do not assume that subject-level error covariance matrix is the same across centers or treatments. Our distributional assumptions are as follows:

$$\begin{aligned} \epsilon_{jk}^i &\sim N_2 \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad \Sigma_{jk} \right], \Sigma_{jk} = \begin{pmatrix} \sigma_{jkCC} & \sigma_{jkCE} \\ \sigma_{jkCE} & \sigma_{jkEE} \end{pmatrix} \\ u_j &\sim N_2 \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad \Psi \right], \Psi = \begin{pmatrix} \tau_{CC} & \tau_{CE} \\ \tau_{CE} & \tau_{EE} \end{pmatrix} \end{aligned} \tag{6.1.2}$$

The heterogeneity amongst the centers is represented by Ψ. The cost and effectiveness outcomes within a center share a common random effect u_j, which accounts for correlation among these outcomes.

Let X_{jk} denote the $2 \times n_{jk}$ matrix of bivariate outcomes of cost and effectiveness from all n_{jk} subjects belonging to center j and treatment group k. Let w_{jk} and ϵ_{jk} be defined similarly. The regression model for X_{jk} is

$$X_{jk} = \gamma_j 1_{n_{jk}}^T + B_k^T w_{jk} + \epsilon_{jk} + u_j 1_{n_{jk}}^T;\; k = 1, 2; j = 1, \ldots, g, \tag{6.1.3}$$

where 1_m denotes an $m \times 1$ vector of ones.

If we consider the case in which there are only two centers, $j = 1,\ 2$, then we can express the outcomes of all patients in each center using the models

$$\begin{aligned} X_1 &= \begin{pmatrix} X_{11} & X_{12} \end{pmatrix} = \begin{pmatrix} \gamma_1 & \gamma_2 & B_1^T & B_2^T \end{pmatrix} D_1 + u_1 1_{1\times n_1} + \begin{pmatrix} \epsilon_{11} & \epsilon_{12} \end{pmatrix} \\ X_2 &= \begin{pmatrix} X_{21} & X_{22} \end{pmatrix} = \begin{pmatrix} \gamma_1 & \gamma_2 & B_1^T & B_2^T \end{pmatrix} D_2 + u_2 1_{1\times n_2} + \begin{pmatrix} \epsilon_{21} & \epsilon_{22} \end{pmatrix}, \end{aligned} \tag{6.1.4}$$

where D_1 and D_2 are defined as:

$$D_1 = \begin{pmatrix} 1_{1\times n_{11}} & 1_{1\times n_{12}} \\ 0_{1\times n_{11}} & 0_{1\times n_{12}} \\ w_{11} & 0 \\ 0 & w_{12} \end{pmatrix},\; D_2 = \begin{pmatrix} 0_{1\times n_{21}} & 0_{1\times n_{22}} \\ 1_{1\times n_{21}} & 1_{1\times n_{22}} \\ w_{21} & 0 \\ 0 & w_{22} \end{pmatrix} \tag{6.1.5}$$

In (6.1.4) X_1 is the $(2 \times n_{1.})$ matrix of the cost and effectiveness outcomes from all the patients in center one, and X_2 is the $(2 \times n_{2.})$ matrix of the cost and effectiveness outcomes from all the patients in center two.

Now let $r_j = u_j 1^T_{(n_{j.}\times 1)} + (\epsilon_{j1}, \epsilon_{j2})$. Then r_j has the following distribution:

$$
\begin{aligned}
r_j &\sim N_2\left[\begin{pmatrix}0\\0\end{pmatrix}, V_j\right]\\
V_j &= J_{(n_j\times n_j)} \otimes \Psi + \begin{pmatrix} I_{(n_{j1}\times n_{j1})} & 0\\ 0 & 0\end{pmatrix} \otimes \Sigma_{j1} + \begin{pmatrix} 0 & 0\\ 0 & I_{(n_{j2}\times n_{j2})}\end{pmatrix} \otimes \Sigma_{j2}\\
V_{j_{(2n_j\times 2n_j)}} &= \begin{bmatrix} \Sigma_{j1}+\Psi & \dots & \Psi\\ \vdots & \ddots & \vdots\\ \Psi & \dots & \Sigma_{j2}+\Psi \end{bmatrix}
\end{aligned}
\tag{6.1.6}
$$

An iterative generalized least squares (IGLS) estimation approach will now be discussed for estimating parameters in the model (6.1.1). This approach has been shown to produce consistent estimators (Goldstein (1986)). The method estimates the covariance parameters and regression parameters together in an iterative fashion. In presenting the IGLS methodology of estimation for the multivariate multilevel model (6.1.1), we utilize the notation presented in Veiga, Smith and Brown (2014). In doing so, we will express V_j in (6.1.6) as a linear function of the covariance parameters. The elements of the vector θ are the distinct parameters in the covariance matrices V_j. Since the matrices Ψ, Σ_{j1}, and Σ_{j2}, are each 2×2, there is a total of $3 + 3g + 3g = 3 + 6g$ parameters to be estimated for the covariance matrices when there are g centers. Write

$$
\Psi = \begin{pmatrix} \tau_{CC} & \tau_{CE}\\ \tau_{CE} & \tau_{EE}\end{pmatrix},\ \Sigma_{j1} = \begin{pmatrix} \sigma_{j1CC} & \sigma_{j1CE}\\ \sigma_{j1CE} & \sigma_{j1EE}\end{pmatrix} \text{ and } \Sigma_{j2} = \begin{pmatrix} \sigma_{j2CC} & \sigma_{j2CE}\\ \sigma_{j2CE} & \sigma_{j2EE}\end{pmatrix}
$$

Let θ be the vector of unknown covariance parameters. In order to have a convenient notation for θ, let

$$\begin{aligned}
&\tilde{\theta}_0 = (\theta_1, \theta_2, \theta_3)^T = (\tau_{CC}, \tau_{CE}, \tau_{EE})^T \\
&\tilde{\theta}_j = (\theta_{j4}, \theta_{j5}, \theta_{j6}, \theta_{j7}, \theta_{j8}, \theta_{j9})^T = \\
&(\sigma_{j1CC}, \sigma_{j1CE}, \sigma_{j1EE}, \sigma_{j2CC}, \sigma_{j2CE}, \sigma_{j2EE})^T, \\
&\theta_j = (\tilde{\theta}_0^T, \tilde{\theta}_j^T)^T
\end{aligned} \tag{6.1.7}$$

$j = 1, 2, ..., g$, so that

$$\theta = (\tilde{\theta}_0^T, \tilde{\theta}_1^T, \tilde{\theta}_2^T, ..., \tilde{\theta}_g^T)^T. \tag{6.1.8}$$

The covariance matrix V_j in (6.1.6) can be written as

$$\begin{aligned}
&V_j = \sum_{s=1}^{3} \theta_s G_{sj} + \sum_{s=4}^{9} \theta_{js} G_{js} \\
&G_{js} = J_{(n_j \times n_j)} \otimes H_{js} + \begin{pmatrix} I_{(n_{j1} \times n_{j1})} & 0 \\ 0 & 0 \end{pmatrix}_{n_j \times n_j} \otimes \Delta_{js} + \begin{pmatrix} 0 & 0 \\ 0 & I_{(n_{j2} \times n_{j2})} \end{pmatrix}_{n_j \times n_j} \otimes \Gamma_{js} \\
&H_{j1} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, H_{j2} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, H_{j3} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, H_{js} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \text{ for s= 4, 5, 6, 7, 8, 9} \\
&\Delta_{j4} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \Delta_{j5} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \Delta_{j6} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \Delta_{js} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \text{ for s = 1, 2, 3, 7, 8, 9} \\
&\Gamma_{j7} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \Gamma_{j8} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \Gamma_{j9} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \Gamma_{js} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \text{ for s = 1, 2, 3, 4, 5, 6}
\end{aligned} \tag{6.1.9}$$

The IGLS methodology iterates between estimation of $\hat{B}$ (stage 1) and $\hat{\theta}$ (stage 2) until convergence is reached. Here B is the matrix consisting of all the fixed effect parameter vectors in the model (6.1.3). Let superscript r denote the r^{th} iteration, where $r =$

0 corresponds to the initial values. The IGLS algorithm for multivariate multilevel models will now be outlined. Vectorized notation will be used for computational purposes. The vectorized model for center j is given by:

$$\tilde{X}_j = \tilde{D}_j\beta + (1_{n_j\times 1}\otimes I_2)u_j + \tilde{\epsilon}_j, \tag{6.1.10}$$

where

$$\beta = \begin{pmatrix} \gamma_1^C \\ \gamma_1^E \\ \vdots \\ \gamma_g^C \\ \gamma_g^E \\ vec(B_1^T)_{2p\times 1} \\ vec(B_2^T)_{2p\times 1} \end{pmatrix}_{(4p+2g)\times 1} ;\tilde{D}_j = (D_j^T\otimes I_2); \qquad \tilde{\epsilon}_j = vec\Big[\,(\epsilon_{j1}, \quad \epsilon_{j2})\,\Big], \tag{6.1.11}$$

where D_j is defined similar to D_1 and D_2 given in (6.1.5).

Next, we provide the estimates and the steps used in the IGLS algorithm. To begin, let the vectorized residuals for patients in center j be denoted by $\tilde{r}_j$:

$$\tilde{r}_j = \tilde{X}_j - \tilde{D}_j\beta, r_j^* = \tilde{r}_j\tilde{r}_j^T, E(r_j^*) = V_j = \sum_{s=1}^{3}\theta_s G_{sj} + \sum_{s=4}^{9}\theta_{js}G_{js} \tag{6.1.12}$$

Now define

$$V_j^* = V_j\otimes V_j, r_j^{**} = vec(r_j^*), E(r_j^{**}) = E[vec(r_j^*)] = H_j\theta_j. \tag{6.1.13}$$

In (6.1.13) H_j is a $4n_{j.}^2\times 9$ matrix and θ_j is the 9×1 vector defined in (6.1.7). These will be used to express the expected value of the vectorized residuals. H_j can be expressed as

$$
\begin{aligned}
&H_j = \left(H_{j0} \vdots H_{j1}\right) \\
&H_{j0} = [vec(G_{j1}), vec(G_{j2}), vec(G_{j3})] \\
&H_{j1} = [vec(G_{j4}), vec(G_{j5}), vec(G_{j6}), vec(G_{j7}), vec(G_{j8}), vec(G_{j9})]
\end{aligned}
\tag{6.1.14}
$$

Using (6.1.14) and (6.1.7), the expectation of r_j^{**} can be expressed as

$$
E(r_j^{**}) = E[vec(r_j^*)] = H_j\theta_j = H_{j0}\tilde{\theta}_0 + H_{j1}\tilde{\theta}_j. \tag{6.1.15}
$$

We note that $r_j^{**} = vec(r_j^*) = vec(\tilde{r}_j\tilde{r}_j^T) = \tilde{r}_j \otimes \tilde{r}_j$. We now introduce $r^{**} = (r_1^{**}, \ldots, r_g^{**})^T$, which is the vectorized form of all residuals. It follows that the expectation of r^{**} is given by

$$
\begin{aligned}
&E(r^{**}) = H\theta \\
&H = \begin{pmatrix} H_{10} & H_{11} & 0 & 0\ldots & 0 \\ H_{20} & 0 & H_{21} & \ldots & 0 \\ \vdots & 0 & 0 & 0 & 0 \\ H_{g0} & 0 & 0 & 0 & \ldots H_{g1} \end{pmatrix}_{4\sum_{j=1}^{g} n_j^2 \times 3(1+2g)} \\
&V^{**} = \begin{pmatrix} V_1^* & 0 & \ldots & 0 \\ 0 & V_2^* & \ldots & 0 \\ \vdots & 0 & \ddots & \vdots \\ 0 & 0 & \ldots & V_g^* \end{pmatrix}_{4\sum_{j=1}^{g} n_j^2 \times 4\sum_{j=1}^{g} n_j^2},
\end{aligned}
\tag{6.1.16}
$$

where θ is defined in (6.1.7), and contains all of the variance-covariance parameters estimated from all g centers in the multi-center trial, namely the 3+6g unknown parameters. The IGLS estimate of θ is given by

$$\hat{\theta} = R^{-1}S, R = H^T {V^{**}}^{-1} H, S = H^T {V^{**}}^{-1} r^{**} \tag{6.1.17}$$

Under the multivariate normal model, the maximum likelihood estimates are equivalent to the estimates obtained using the iterative generalized least squares method. The resulting estimates are biased, because the sampling variation of $\hat{\beta}$ is unaccounted for in the estimation process. Specifically, we have the following result concerning the bias:

$$\begin{aligned} &\tilde{\alpha}_j = \tilde{X}_j - \tilde{D}_j\hat{\beta} \\ &E(\tilde{\alpha}_j\tilde{\alpha}_j^T) = E[(\tilde{X}_j - \tilde{D}_j\hat{\beta})(\tilde{X}_j - \tilde{D}_j\hat{\beta})^T] = V_j - \tilde{D}_j Cov(\hat{\beta})\tilde{D}_j^T \\ &= V_j - \tilde{D}_j(\sum_{j=1}^{g} \tilde{D}_j^T V_j^{-1} \tilde{D}_j)^{-1} \tilde{D}_j^T \end{aligned} \tag{6.1.18}$$

The result (6.1.18) shows that the estimation of V_j based on estimated residuals $\tilde{\alpha}_j$, leads to an underestimation. To correct for such a bias one can implement a restricted iterative generalized least squares (RIGLS) approach. The RIGLS estimates will be equivalent to the restricted maximum likelihood estimates under normality (Goldstein (2011)). Rather than using the residuals in (6.1.18) the residuals defined in (6.1.19) will be used. These residuals correct for the bias and therefore, yield a bias corrected estimate of V_j.

$$\begin{aligned} &R_j^* = (\tilde{X}_j - \tilde{D}_j\hat{\beta})(\tilde{X}_j - \tilde{D}_j\hat{\beta})^T + \tilde{D}_j(\sum_{j=1}^{g} \tilde{D}_j^T V_j^{-1} \tilde{D}_j)^{-1} \tilde{D}_j^T \\ &E(R_j^*) = V_j \end{aligned} \tag{6.1.19}$$

Let $R_j^{**} = vec(R_j^*)$ and define $R^{**} = (R_1^{**}, \ldots, R_g^{**})^T$.

At each iteration of the algorithm, we must check for positive-definiteness of V_j. We do so by checking that the eigenvalues of matrices Σ_{jk} are positive, for $j = 1, \ldots, g$ and

$k = 1, 2$, and the eigenvalues of A_j are also positive, where

$$A_j = \begin{pmatrix} \Sigma_{j1} + n_{j1}\Psi & \sqrt{n_{j1}n_{j2}}\Psi \\ \sqrt{n_{j1}n_{j2}}\Psi & \Sigma_{j2} + n_{j2}\Psi \end{pmatrix}_{4\times 4}, \tag{6.1.20}$$

$j = 1, \ldots, g$. The matrix A_j is determined by applying an orthogonal and permutation transformation to V_j, details of which are developed in Appendix B. The multivariate RIGLS algorithm is provided in algorithm 13. The iteration is initialized using the ordinary least squares (OLS) estimates of β and Σ_{jk}s, as noted in Goldstein (1986).

Algorithm 13: RIGLS algorithm for estimating the parameters of the MLM

1 **Initialization stage (r = 0):** Initialize estimates of β and θ.

Use OLS for the initial estimate of β: $\hat{\beta}^{T(0)} = \left[\sum_{j=1}^{g} \tilde{D}_j^T \tilde{D}_j\right]^{-1} \left[\sum_{j=1}^{g} \tilde{D}_j \tilde{X}_j\right]$

2 Initialize $\hat{\theta}$ and set $\hat{\Psi}^{(0)}$ to the zero matrix.

The residuals are calculated as follows:

$\hat{\epsilon}^{i(0)} = X_{jk}^i - B_k^T w_{jk}^i - \gamma_j,\ \hat{\epsilon}^{(0)} = \sum_{i=1}^{n_{jk}} \frac{1}{n_{jk}} \hat{\epsilon}^{i(0)}$

for each $j = 1, \ldots, g$ and $k = 1, 2$.

3 The initial estimate of $\hat{\Sigma}_{j1}$ and $\hat{\Sigma}_{j2}$ are: $\hat{\Sigma}_{jk} = \sum_{i=1}^{n_{jk}} \frac{(\hat{\epsilon}^{i(0)} - \hat{\epsilon}^{(0)})(\hat{\epsilon}^{i(0)} - \hat{\epsilon}^{(0)})^T}{n_{jk} - p - 1}$

for each combination of $j = 1, \ldots, g$, and $k = 1, 2$.

4 **Iteration r^{th} stage 1:** At this stage $\hat{\beta}^{(r)}$ is calculated as follows:

$\hat{\beta}^{(r)} = {P^{(r)}}^{-1} Q^{(r)}$, where $P^{(r)} = \sum_{j=1}^{g} \tilde{D}_j^T [V_j^{-1}(\hat{\theta}^{(r-1)})] \tilde{D}_j$

$Q^{(r)} = \sum_{j=1}^{g} \tilde{D}_j^T [V_j^{-1}(\hat{\theta}^{(r-1)})] \tilde{X}_j$

5 **Iteration r^{th} stage 2:** At this stage $\hat{\theta}^{(r)}$ is calculated as follows:

$\hat{\theta}^{(r)} = {R^{(r)}}^{-1} S^{(r)}\ R^{(r)} = H^T {V^{**}}^{-1} H, S^{(r)} = H^T {V^{**}}^{-1} R^{**}$

6 Check that eigenvalues of A_j and Σ_{jk}s are positive for $j = 1, \ldots, g$ and $k = 1, 2$.

7 Repeat stages one and two until convergence is reached. Convergence is defined as: $\left\| \hat{\beta}^{(r)} - \hat{\beta}^{(r-1)} \right\| \leq \epsilon$ and $\left\| \hat{\theta}^{(r)} - \hat{\theta}^{(r-1)} \right\| \leq \epsilon$.

When we implemented the above algorithm, we noticed that the estimates of Σ_{jk} and A_j are positive definite, $j = 1 \dots g$ and $k = 1, 2$, where A_j is defined in (6.1.20). However, the resulting estimate of Ψ may not necessarily be non-negative definite. This result is typical for multilevel models, as Ψ is not a variance-covariance matrix in the typical sense. In our approach, we decided not to add extra computational steps to force Ψ to be non-negative definite, since the algorithm produced positive definite estimates of V_j, which is satisfactory. Further discussion on this is provided in Appendix B.

6.2 Interval estimation of the INB

Under the model (6.1.1), we have

$$\begin{aligned} E[C^i_{jk}] = \mu^C_{jk} = exp\Big[\frac{\sigma_{jkCC} + \tau_{CC}}{2} + \gamma^C_j + (B^C_k)^T w^i_{Cjk}\Big] \\ E[E^i_{jk}] = \mu^E_{jk} = exp\Big[\frac{\sigma_{jkEE} + \tau_{EE}}{2} + \gamma^E_j + (B^E_k)^T w^i_{Ejk}\Big] \end{aligned} \tag{6.2.1}$$

We shall now define the incremental net benefit (INB) for a patient with covariate vector $\mathbf{w}_0$. Let C^0_{jk} and E^0_{jk}, respectively, denote the patient's cost and effectiveness if the patient is assigned to the treatment k at the j^{th} center. The mean cost and effectiveness are given by

$$\begin{aligned} E[C^i_{jk}] = \mu^C_{jk0} = exp\Big[\frac{\sigma_{jkCC} + \tau_{CC}}{2} + \gamma^C_j + (B^C_k)^T \mathbf{w}_0\Big] \\ E[E^i_{jk}] = \mu^E_{jk0} = exp\Big[\frac{\sigma_{jkEE} + \tau_{EE}}{2} + \gamma^E_j + (B^E_k)^T \mathbf{w}_0\Big] \end{aligned} \tag{6.2.2}$$

Note that the means given in (6.2.2) are center-dependent. In order to formulate a patient-specific INB, we shall average the means in (6.2.2) across centers. Thus let

$$\mu^C_{k0} = \frac{1}{g}\sum_{j=1}^{g} \mu^C_{jk0}, \;\; \mu^E_{k0} = \frac{1}{g}\sum_{j=1}^{g} \mu^E_{jk0} \tag{6.2.3}$$

The average incremental cost and effectiveness are defined as:

$$\Delta_E = \mu_{10}^E - \mu_{20}^E,\ \Delta_C = \mu_{10}^C - \mu_{20}^C. \tag{6.2.4}$$

Using the above average incremental differences, the INB for patients with covariate vector $\mathbf{w}_0$ is defined by:

$$INB(\mathbf{w}_0) = \lambda\Delta_E - \Delta_C \tag{6.2.5}$$

For the assessment of cost-effectiveness, we will construct a lower confidence limit for $INB(\mathbf{w}_0)$ using the delta method. The asymptotic distribution of the parameters in the model (6.1.1) is

$$\left[\begin{pmatrix}\hat{\beta}\\ \hat{\theta}\end{pmatrix} - \begin{pmatrix}\beta\\ \theta\end{pmatrix}\right] \xrightarrow{D} N\left[0, \begin{pmatrix}Cov(\hat{\beta}) & 0\\ 0 & Cov(\hat{\theta})\end{pmatrix}\right], \tag{6.2.6}$$

where

$$Cov[\hat{\beta}] = \left(\sum_{j=1}^{g} \tilde{D}_j^T V_j^{-1} \tilde{D}_j\right)^{-1} \tag{6.2.7}$$

$$Cov[\hat{\theta}] = 2\left(H^T (V^{**})^{-1} H\right)^{-1} \tag{6.2.8}$$

The covariance matrices in (6.2.7) and (6.2.8) are derived in the Appendix A. We note that this derivation is based on the full maximum likelihood rather than the restricted maximum likelihood approach. However, the estimates of β and θ remain bias corrected, as noted in Goldstein (2011). Applying the delta method the asymptotic distribution of $INB(\mathbf{w}_0) =$

$INB(\beta, \theta)$ is given by

$$\left[INB(\hat{\beta}, \hat{\theta}) - INB(\beta, \theta)\right] \xrightarrow{D} N\left[0, \nabla^T INB(\beta, \theta)\begin{pmatrix} Cov(\hat{\beta}) & 0 \\ 0 & Cov(\hat{\theta}) \end{pmatrix}\nabla INB(\beta, \theta)\right]$$
$$= N\left[0, \sigma^2_{INB}\right] \tag{6.2.9}$$

The resulting $100(1-\alpha)\%$ asymptotic lower confidence limit is:

$$INB(\hat{\beta}, \hat{\theta}) - z_{1-\alpha}\hat{\sigma}_{INB}, \tag{6.2.10}$$

where $z_{1-\alpha}$ denotes the $(1-\alpha)$ quantile of the standard normal. If the lower confidence limit is positive then the new treatment is considered cost-effective for the individual with covariate vector $\mathbf{w}_0$.

6.3 An example

We apply our proposed model and methods to the Canadian implantable defibrillator study (CIDS). The study randomized patients at risk of cardiac arrest to two treatment groups: implantable cardioverter-defibrillator and amiodarone between October 1990 and January 1997. The primary effectiveness outcome is all-cause mortality. Number of response days is another effectiveness outcome that we will utilize in our analysis. Response days are defined as the number of days a patient was followed, either until death, loss-of-follow-up, or trial completion. The latter two end-points result in censored outcomes. Costs were collected from a subset of patients. The multi-center trial took place over 24 trials. However, sample sizes varied between 1 to 123 patients per center (not accounting for missing data). Cost-effectiveness analysis on the CIDS data has been carried out by Willan, Lin and Manca (2005). The authors account for censoring through inverse probability weighting.

We focus our analysis to just two centers. Based on these two centers, we fitted a lognormal-lognormal model for the cost and effectiveness outcomes having the form of (6.1.1). In our analysis, we have excluded patients with missing data values. Several patient-level characteristics were recorded at the start of the trial, including age, ejection fraction, gender, and indicator of congestive heart failure. The ejection fraction measures the ratio between blood pumped out of and into the heart. It demonstrates how well the heart is able to pump blood through the body. Low ejection fraction suggests that the heart is deficient in its pumping function and has been associated with congestive heart failure. The model fit indicated that the first two factors, i.e. age and ejection fraction, were statistically significant. Using the lognormal-lognormal model fitted with the covariates age and ejection fraction, outcomes were simulated for 42 patients within each center and each treatment group. The new treatment is the implantable cardioverter-defibrillator (treatment one), and the standard treatment is taken to be amiodarone (treatment two).

We fit model (6.1.1) with age and ejection fraction as covariates to the simulated data having 42 patients within each center and each treatment group, i.e. $n_{11} = n_{12} = n_{21} = n_{22} = 42$. We obtain estimates of the parameters using the restricted iterative generalized least squares (RIGLS) method, provided in algorithm 13, with stopping criterion $\epsilon = 0.0001$. The estimates are:

$$\hat{\gamma}_1 = \begin{pmatrix} \hat{\gamma}_1^C \\ \hat{\gamma}_1^E \end{pmatrix} = \begin{pmatrix} 10.9242 \\ 7.7250 \end{pmatrix}; \hat{\gamma}_2 = \begin{pmatrix} \hat{\gamma}_2^C \\ \hat{\gamma}_2^E \end{pmatrix} = \begin{pmatrix} 9.7450 \\ 7.6364 \end{pmatrix}$$

$$\hat{B_1}^T = \begin{pmatrix} \hat{B}^C_{1,eject} & \hat{B}^C_{1,age} \\ \hat{B}^E_{1,eject} & \hat{B}^E_{1,age} \end{pmatrix} = \begin{pmatrix} 0.0097 & -0.0004 \\ -0.0055 & -0.0026 \end{pmatrix}$$

$$\hat{B_2}^T = \begin{pmatrix} \hat{B}^C_{2,eject} & \hat{B}^C_{2,age} \\ \hat{B}^E_{2,eject} & \hat{B}^E_{2,age} \end{pmatrix} = \begin{pmatrix} 0.0092 & -0.0074 \\ 0.0077 & -0.0089 \end{pmatrix}$$

Based on the parameter estimates, we notice differences between the two centers as well as between the two treatments. Average cost at the baseline is slightly higher for patients in the first center compared to those in the second center. From the estimates $\hat{B}_1$ and $\hat{B}_2$, we note that both treatments have increasing average costs for younger patients and those who have a larger ejection fraction. Average number of response days are lower for older patients under both treatment groups. However, the effect of ejection fraction on the average number of response days differs between the two treatment groups. In particular, as the ejection fraction increases, the number of response days in the amiodarone treatment group increases. In contrast, a similar increase in ejection fraction has a decreasing affect for the average number of response days for implantable cardioverter-defibrillator. Willan, Briggs and Hoch (2004) also found that $\hat{B}^C_{1,eject} > 0$ and $\hat{B}^E_{1,eject} < 0$ under their analysis (see Willan, Lin and Manca (2005, p. 138-139)).

The estimated covariance matrices are given by

$$\hat{\Psi} = \begin{pmatrix} 0.000078 & 0.000028 \\ 0.000028 & 0.000042 \end{pmatrix}$$

$$\widehat{\Sigma_{11}} = \begin{pmatrix} 0.1178 & 0.0051 \\ 0.0051 & 0.1555 \end{pmatrix} \qquad \widehat{\Sigma_{12}} = \begin{pmatrix} 0.8401 & 0.1799 \\ 0.1799 & 0.1877 \end{pmatrix}$$

$$\widehat{\Sigma_{21}} = \begin{pmatrix} 0.9897 & 0.0640 \\ 0.0640 & 0.1456 \end{pmatrix} \qquad \widehat{\Sigma_{22}} = \begin{pmatrix} 1.0015 & 0.0843 \\ 0.0843 & 0.1407 \end{pmatrix}$$

The estimated covariance matrices are all positive definite. However, $\hat{\Psi}$ indicates that variability between centers is near zero. While these elements are small in magnitude, they still affect the resulting estimates, as others have noted in the literature. The estimated matrices A_1 and A_2 are positive definite as required and correspond to centers one and two, each is provided next

$$\widehat{A_1} = \begin{pmatrix} 0.1211 & 0.0063 & 0.0033 & 0.0012 \\ 0.0063 & 0.1573 & 0.0012 & 0.0018 \\ 0.0033 & 0.0012 & 0.8434 & 0.1810 \\ 0.0012 & 0.0018 & 0.1810 & 0.1894 \end{pmatrix}$$

$$\widehat{A_2} = \begin{pmatrix} 0.9930 & 0.0652 & 0.0033 & 0.0012 \\ 0.0652 & 0.1473 & 0.0012 & 0.0018 \\ 0.0033 & 0.0012 & 1.0047 & 0.0855 \\ 0.0012 & 0.0018 & 0.0855 & 0.1424 \end{pmatrix}$$

Next, we analyze the cost-effectiveness of the new treatment, implantable cardioverter-defibrillator against the standard treatment amiodarone. This will be carried out for twenty-three patients, i.e. corresponding to twenty-three covariate vectors $\mathbf{w}_0$ (i.e., 23 pairs of values for the age and ejection fraction). These are given in Table 6.1.

Table 6.1 Covariate values of twenty-three patients

Patient ID	Lveject	Age
1	12.5	55
2	15	55
3	12.5	65
4	15	65
5	17.5	65
6	20	65
7	12.5	75
8	15	75
9	17.5	75
10	20	75
11	22.5	75
12	30	55
13	35	55
14	40	55
15	45	55
16	50	55
17	35	65
18	40	65
19	45	65
20	50	65
21	40	75
22	45	75
23	50	75

The twenty-three cases include patients having ejection fraction levels ranging from 12.5 to 50, and ages of 55, 65, and 75. In Willan, Lin and Manca (2005), the authors included ejection fraction as an indicator function, which indicates whether an individual has an ejection fraction less than or equal to 35. In our analysis, we include both covariates as continuous effects. The next table gives each patient's average incremental cost and average incremental effectiveness.

Table 6.2 Estimated average incremental number of response days and incremental costs for the twenty-three patients

Patient ID	$\widehat{\Delta}_E$	$\widehat{\Delta}_C$
1	308.6	5919.8
2	251.7	6117.1
3	396.2	8723.5
4	342.6	8985.9
5	288.8	9255.9
6	234.8	9533.6
7	473.5	11 311.9
8	423.0	11 634.4
9	372.3	11 965.8
10	321.4	12 306.5
11	270.3	12 656.6
12	−95.6	7434.4
13	−214.0	7928.9
14	−333.9	8453.9
15	−455.5	9011.2
16	−579.0	9602.7
17	−94.9	11 376.0
18	−207.3	12 063.0
19	−321.2	12 789.8
20	−436.7	13 558.8
21	−94.1	15 394.6
22	−200.8	16 277.9
23	−309.0	17 210.7

From the results in Table 6.2, the new treatment, implantable cardioverter-defibrillator, is more costly on average for all patients. The average cost is higher for older patients compared to that of younger patients. In addition, the costs are higher on average for patients having larger ejection fraction levels. The most noticeable differences occur for the average incremental effectiveness. The first eleven patients having lower ejection fraction levels have higher number of response days under the new treatment on average compared to the standard treatment. In contrast, the remaining patients all have negative average incremental effectiveness which indicates that the standard treatment is more effective for such patients.

For a fixed age, the average incremental effectiveness decreases under the new treatment as the ejection fraction increases.

Based on these results, we conclude that patients 12 to 23 clearly will not achieve cost-effectiveness under the new treatment, as the new treatment is more costly and less effective on average. For the first eleven patients, these preliminary results indicate that older patients with lower ejection fraction levels (more at risk of cardiac arrest) are most likely to benefit from the new treatment. Next, the estimated INB values for the twenty-three patients upon which our analysis focuses are provided in the following table.

Table 6.3 Estimated INB for the twenty-three patients for different values of λ

Patient ID	$\lambda = 250$	$\lambda = 500$	$\lambda = 1000$	$\lambda = 5000$
1	71 221.0	148 361.7	302 643.2	1 536 895.3
2	56 797.9	119 712.9	245 542.9	1 252 182.7
3	90 333.4	189 390.2	387 503.9	1 972 413.7
4	76 671.8	162 329.4	333 644.8	1 704 167.3
5	62 950.9	135 157.7	279 571.2	1 434 879.6
6	49 166.7	107 867.0	225 267.6	1 164 472.6
7	107 063.6	225 439.1	462 190.2	2 356 198.4
8	94 107.7	199 849.8	411 333.9	2 103 207.0
9	81 100.8	174 167.3	360 300.4	1 849 365.5
10	68 038.9	148 384.3	309 075.0	1 594 601.0
11	54 918.3	122 493.1	257 642.8	1 338 840.4
12	−31 330.9	−55 227.4	−103 020.5	−485 364.9
13	−61 420.4	−114 911.9	−221 894.9	−1 077 758.8
14	−91 928.1	−175 402.2	−342 350.5	−1 677 936.7
15	−122 892.2	−236 773.1	−464 534.9	−2 286 629.7
16	−154 351.9	−299 101.1	−588 599.4	−2 904 586.1
17	−35 103.6	−58 831.2	−106 286.4	−485 927.7
18	−63 893.4	−115 723.7	−219 384.4	−1 048 670.1
19	−93 092.2	−173 394.6	−333 999.3	−1 618 837.3
20	−122 737.1	−231 915.3	−450 271.9	−2 197 124.1
21	−38 909.0	−62 423.5	−109 452.4	−485 683.4
22	−66 478.2	−116 678.5	−217 079.1	−1 020 283.9
23	−94 448.5	−171 686.3	−326 162.0	−1 561 967.4

The INB estimates in Table 6.3 indicate that cost-effectiveness differs immensely amongst patients. The analysis is conducted under four willingness-to-pay values ($\lambda =$

250, 500, 1000, and 5000). The first eleven patients having lower ejection fraction levels, have positive and increasing INB values as a function of λ. The remaining patients all have negative INB values that decrease as willingness-to-pay increases. To assess the cost-effectiveness of the new treatment, we compute the lower 95% confidence limits for the patients in Table 6.1 using the delta method approach.

Table 6.4 95% lower confidence limits of the INB using the delta method for different values of λ

Patient ID	$\lambda = 250$	$\lambda = 500$	$\lambda = 1000$	$\lambda = 5000$
1	9504.3	26 587.9	59 479.0	319 514.6
2	1319.3	10 829.8	28 399.2	165 415.1
3	14 979.2	39 357.6	87 298.7	468 869.1
4	8302.5	26 556.2	62 179.1	345 031.4
5	1012.8	12 551.3	34 637.5	208 931.8
6	−7110.8	−3114.1	3734.3	55 742.6
7	17 485.5	46 352.1	103 471.6	558 960.6
8	11 878.9	35 677.1	82 637.0	456 792.5
9	5888.8	24 264.1	60 334.0	347 260.2
10	−602.5	11 876.4	36 080.0	227 894.9
11	−7760.2	−1821.4	9189.3	95 182.8
12	−82 606.7	−155 879.8	−305 456.9	−1 509 277.5
13	−126 056.8	−243 495.1	−481 662.5	−2 394 678.2
14	−174 008.2	−340 028.8	−675 641.3	−3 368 792.6
15	−225 074.6	−442 672.8	−881 808.0	−4 403 945.5
16	−278 659.6	−550 270.3	−1 097 893.9	−5 488 975.8
17	−89 942.4	−167 226.7	−324 505.1	−1 589 137.3
18	−131 916.2	−251 392.2	−493 436.0	−2 436 993.8
19	−178 210.2	−344 144.7	−679 491.2	−3 370 310.5
20	−227 584.6	−442 968.0	−877 658.5	−4 364 227.7
21	−97 386.6	−178 063.6	−342 109.7	−1 660 851.8
22	−138 015.0	−259 076.3	−504 353.7	−2 473 960.9
23	−182 809.4	−348 393.3	−683 170.9	−3 369 794.3

The lower confidence limits in Table 6.4 demonstrate that the new treatment, implantable cardioverter-defibrillator, is not cost-effective for patients with higher ejection fraction levels (greater than or equal to 30), i.e. patients 12 to 23. For these patients, increased investment in the new treatment results in lower cost-effectiveness, as expected based on the results in Table 6.2. Focusing on the first 11 patients with lower ejection fraction levels, we note that the lower limits increase when the willingness-to-pay increases. This indicates that increased investment in the new treatment results in higher cost-effectiveness for these patients.

Amongst the first 11 patients having ejection fraction levels equal or less than 22.5, the older patients with lower ejection fraction levels have the highest cost-effectiveness under the new treatment. Similar results regarding the effect of ejection fraction were reported in Willan, Lin and Manca (2005, p. 139). In our analysis, patient 7 (having ejection fraction of 12.5 and age 75) has the highest lower limit, followed by patient 3 (having ejection fraction of 12.5 and age 65), followed by patient 8 (having ejection fraction equal to 15 and age 75) regardless of willingness-to-pay value. Patients 6 and 11 have the largest ejection fraction levels and the smallest lower limits among the first eleven patients.

Certain trends regarding the willingness-to-pay are also evident. For lower values of λ, the covariate ejection fraction appears to be the dominating factor determining the cost-effectiveness of the new treatment. However, as λ increases, age becomes increasingly important. In particular, as λ increases, older patients achieve more benefit compared to younger patients (when ejection fraction is still low). Further, positive lower limits of the INB indicate that the new treatment is cost-effective. Out of the first eleven patients, all achieve cost-effectiveness by $\lambda = 250$, apart from patients 6, 10, and 11. These patients all have ejection fraction levels equal to 20 and 22.5, the highest in this subset of patients. Increasing the willingness-to-pay to $\lambda = 500$, patients 6 and 11 still do not achieve cost-effectiveness under the new treatment. However, when λ is increased to 1000, all of the first eleven patients have positive lower confidence limits for the INB, indicating that all of these patients have achieved cost-effectiveness under the new treatment.

In addition, we can also plot the estimated INB and the corresponding 95% confidence limits to visualize the effect of willingness-to-pay on cost-effectiveness. In Figure 6.1 the estimated INB and lower 95% confidence limits are plotted as a function of willingness-to-pay, λ.

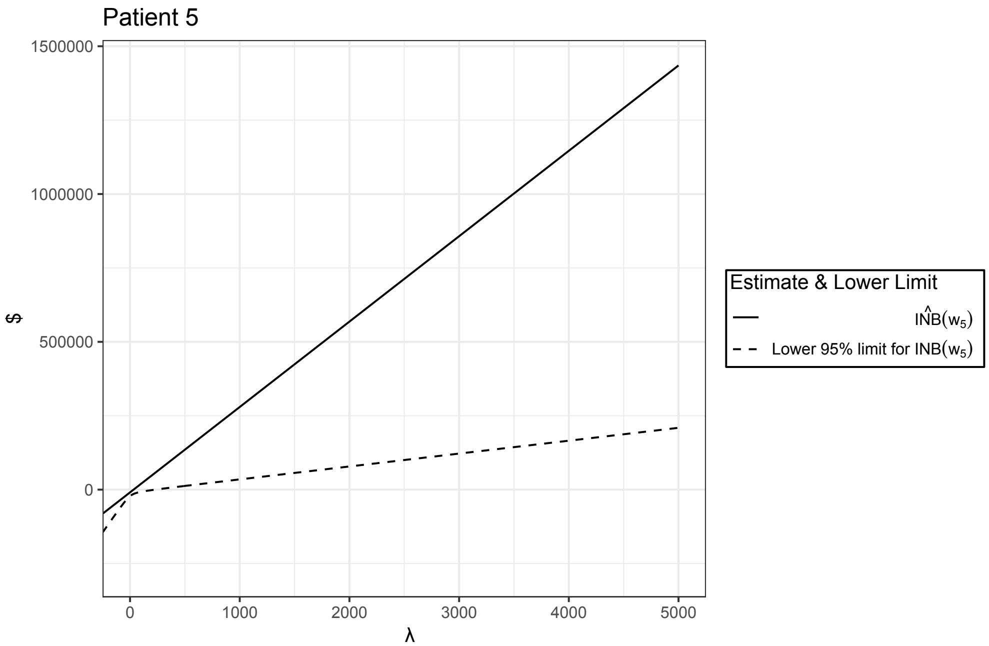


Figure 6.1: Estimated INB and 95% lower confidence limit for patient five

Figures similar to Figure 6.1 can be particularly useful to decision makers who desire to analyze cost-effectiveness for a particular patient in regards to the investment, the willingness-to-pay.

6.4 Discussion and conclusions

In this chapter, we developed a multivariate multi-level model for the outcomes from multi-center trials. Our model includes relevant patient-level covariates to enable a patient-level assessment of cost-effectiveness of a new treatment. The proposed model includes a center-specific intercept and treatment-specific regression parameters. The model is a linear mixed-effects model; the random effect is center-specific. The proposed model captures heterogeneity at both the center-level and subject-level. In addition, we model the outcomes of the multi-center trial directly, rather than modelling cost-effectiveness metrics such as the NMB. Further, the model includes treatment-specific regression parameters which are often excluded in most multi-level modelling. This is an important feature since patient characteristics can have a different effect on the outcomes of treatments.

We have developed a restricted iterative generalized least squares approach, which provides parameter estimates equivalent to the restricted maximum likelihood solution. The cost-effectiveness of the new treatment for a particular patient is assessed by calculating a lower confidence limit for that patient's INB. We constructed an asymptotic lower limit using the delta method. The results of our analysis indicate that patient-level covariates strongly affect the cost-effectiveness of a new treatment. In particular, it is possible to draw conclusions regarding for which patients the new treatment is cost-effective. Furthermore, our methods provide information on the amount of resources a decision maker should invest for a specified treatment and particular patient.

We believe that our methods and models can be further developed to suit other frameworks. For example, extension of our model to accommodate longitudinal data would be of particular interest for multi-center trials that have long-term follow-up periods.

6.5 Appendix

6.5.1 Appendix A- Derivation of the covariance matrices

Here we shall give a derivation of Cov($\hat{\beta}$) and Cov($\hat{\theta}$) under the multi-level model. The log-likelihood is given by

$$l(x|\theta,\beta) = log[L(x|\theta,\beta)] \propto \frac{-1}{2}\sum_{j=1}^{g}(\tilde{X}_j - \tilde{W}_j\beta)^T V_j^{-1}(\tilde{X}_j - \tilde{W}_j\beta)] - \frac{1}{2}\sum_{j=1}^{g} log|V_j| \qquad (6.5.1)$$

It is well known that the Fisher information matrix is block-diagonal, having two blocks corresponding to β and θ. From the above expression for the log-likelihood function, it follows that the block corresponding to β is simply

$$\sum_{j=1}^{g} \tilde{W}_j^T V_j^{-1} \tilde{W}_j$$

Thus we get

$$Cov[\hat{\beta}] = \left(\sum_{j=1}^{g} \tilde{W}_j^T \hat{V}_j^{-1} \tilde{W}_j \right)^{-1}$$

Derivation of Cov($\hat{\theta}$)

We will present the derivation for some of the elements in θ for brevity. We restrict the results here for the case when the number of centers is $g = 2$, however the results follow similarly when there are more than two centers.

The first order partial derivatives of the log-likelihood function with respect to the elements in θ are:

$$\begin{aligned}
\frac{\partial l(x|\beta,\theta)}{\partial \tau_{CC}} \propto & \frac{1}{2}(\tilde{X}_j - \tilde{W}_j\beta)^T V_1^{-1} \left[J_{(n_1 \times n_1)} \otimes \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \right] V_1^{-1} (\tilde{X}_j - \tilde{W}_j\beta) \\
& - \frac{1}{2} \sum_{j=1}^{g} Tr \left[V_1^{-1} J_{(n_1 \times n_1)} \otimes \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \right] \\
& + \frac{1}{2}(\tilde{X}_j - \tilde{W}_j\beta)^T V_2^{-1} \left[J_{(n_2 \times n_2)} \otimes \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \right] V_2^{-1} (\tilde{X}_j - \tilde{W}_j\beta) \\
& - \frac{1}{2} Tr \left[V_2^{-1} J_{(n_2 \times n_2)} \otimes \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \right]
\end{aligned} \tag{6.5.2}$$

$$
\begin{aligned}
\frac{\partial l(x|\beta,\theta)}{\partial \tau_{CE}} \propto & \frac{1}{2}(\tilde{X}_j - \tilde{W}_j\beta)^T V_1^{-1}\left[J_{(n_1\times n_1)} \otimes \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\right] V_1^{-1}(\tilde{X}_j - \tilde{W}_j\beta) \\
& - \frac{1}{2}\sum_{j=1}^{g} Tr\left[V_1^{-1} J_{(n_1\times n_1)} \otimes \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\right] \\
& + \frac{1}{2}(\tilde{X}_j - \tilde{W}_j\beta)^T V_2^{-1}\left[J_{(n_2\times n_2)} \otimes \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\right] V_2^{-1}(\tilde{X}_j - \tilde{W}_j\beta) \\
& - \frac{1}{2} Tr\left[V_2^{-1} J_{(n_2\times n_2)} \otimes \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\right]
\end{aligned} \tag{6.5.3}
$$

$$
\begin{aligned}
\frac{\partial l(x|\beta,\theta)}{\partial \tau_{EE}} \propto & \frac{1}{2}(\tilde{X}_j - \tilde{W}_j\beta)^T V_1^{-1}\left[J_{(n_1\times n_1)} \otimes \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}\right] V_1^{-1}(\tilde{X}_j - \tilde{W}_j\beta) \\
& - \frac{1}{2}\sum_{j=1}^{g} Tr\left[V_1^{-1} J_{(n_1\times n_1)} \otimes \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}\right] \\
& + \frac{1}{2}(\tilde{X}_j - \tilde{W}_j\beta)^T V_2^{-1}\left[J_{(n_2\times n_2)} \otimes \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}\right] V_2^{-1}(\tilde{X}_j - \tilde{W}_j\beta) \\
& - \frac{1}{2} Tr\left[V_2^{-1} J_{(n_2\times n_2)} \otimes \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}\right]
\end{aligned} \tag{6.5.4}
$$

$$
\begin{aligned}
\frac{\partial l(x|\beta,\theta)}{\partial \sigma_{j1CC}} \propto & -\frac{1}{2} Tr\left[V_j^{-1} \begin{pmatrix} I_{(n_{j1}\times n_{j1})} & 0 \\ 0 & 0 \end{pmatrix} \otimes \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}\right] \\
& + \frac{1}{2}(\tilde{X}_j - \tilde{W}_j\beta)^T V_j^{-1}\left[\begin{pmatrix} I_{(n_{j1}\times n_{j1})} & 0 \\ 0 & 0 \end{pmatrix} \otimes \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}\right] V_j^{-1}(\tilde{X}_j - \tilde{W}_j\beta)
\end{aligned} \tag{6.5.5}
$$

$$
\begin{aligned}
\frac{\partial l(x|\beta,\theta)}{\partial \sigma_{j1CE}} \propto & -\frac{1}{2}Tr\left[V_j^{-1}\begin{pmatrix} I_{(n_{j1}\times n_{j1})} & 0 \\ 0 & 0 \end{pmatrix} \otimes \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\right] \\
& + \frac{1}{2}(\tilde{X}_j - \tilde{W}_j\beta)^T V_j^{-1}\left[\begin{pmatrix} I_{(n_{j1}\times n_{j1})} & 0 \\ 0 & 0 \end{pmatrix} \otimes \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\right] V_j^{-1}(\tilde{X}_j - \tilde{W}_j\beta)
\end{aligned} \tag{6.5.6}
$$

$$
\begin{aligned}
\frac{\partial l(x|\beta,\theta)}{\partial \sigma_{j1EE}} \propto & -\frac{1}{2}Tr\left[V_j^{-1}\begin{pmatrix} I_{(n_{j1}\times n_{j1})} & 0 \\ 0 & 0 \end{pmatrix} \otimes \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}\right] \\
& + \frac{1}{2}(\tilde{X}_j - \tilde{W}_j\beta)^T V_j^{-1}\left[\begin{pmatrix} I_{(n_{j1}\times n_{j1})} & 0 \\ 0 & 0 \end{pmatrix} \otimes \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}\right] V_j^{-1}(\tilde{X}_j - \tilde{W}_j\beta)
\end{aligned} \tag{6.5.7}
$$

$$
\begin{aligned}
\frac{\partial l(x|\beta,\theta)}{\partial \sigma_{j2CC}} \propto & -\frac{1}{2}Tr\left[V_j^{-1}\begin{pmatrix} 0 & 0 \\ 0 & I_{(n_{j2}\times n_{j2})} \end{pmatrix} \otimes \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}\right] \\
& + \frac{1}{2}(\tilde{X}_j - \tilde{W}_j\beta)^T V_j^{-1}\left[\begin{pmatrix} 0 & 0 \\ 0 & I_{(n_{j2}\times n_{j2})} \end{pmatrix} \otimes \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}\right] V_j^{-1}(\tilde{X}_j - \tilde{W}_j\beta)
\end{aligned} \tag{6.5.8}
$$

$$
\begin{aligned}
\frac{\partial l(x|\beta,\theta)}{\partial \sigma_{j2CE}} \propto & -\frac{1}{2}Tr\left[V_j^{-1}\begin{pmatrix} 0 & 0 \\ 0 & I_{(n_{j2}\times n_{j2})} \end{pmatrix} \otimes \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\right] \\
& + \frac{1}{2}(\tilde{X}_j - \tilde{W}_j\beta)^T V_j^{-1}\left[\begin{pmatrix} 0 & 0 \\ 0 & I_{(n_{j2}\times n_{j2})} \end{pmatrix} \otimes \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\right] V_j^{-1}(\tilde{X}_j - \tilde{W}_j\beta)
\end{aligned} \tag{6.5.9}
$$

$$\frac{\partial l(x|\beta,\theta)}{\partial \sigma_{j2EE}} \propto -\frac{1}{2}Tr\left[V_j^{-1}\begin{pmatrix}0 & 0\\ 0 & I_{(n_{j2}\times n_{j2})}\end{pmatrix}\otimes\begin{pmatrix}0 & 0\\ 0 & 1\end{pmatrix}\right] + \frac{1}{2}(\tilde{X}_j-\tilde{W}_j\beta)^T V_j^{-1}\left[\begin{pmatrix}0 & 0\\ 0 & I_{(n_{j2}\times n_{j2})}\end{pmatrix}\otimes\begin{pmatrix}0 & 0\\ 0 & 1\end{pmatrix}\right]V_j^{-1}(\tilde{X}_j-\tilde{W}_j\beta) \tag{6.5.10}$$

Next, we derive the second order partial derivative of the log-likelihood with respect to τ_{CC}.

$$\begin{aligned}\frac{\partial l(x|\beta,\theta)}{\partial^2 \tau_{CC}} \propto & -Tr\left[V_1^{-1}J_{(n_1\times n_1)}\otimes\begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix}V_1^{-1}J_{(n_1\times n_1)}\otimes\begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix}V_1^{-1}(\tilde{X}_j-\tilde{W}_j\beta)(\tilde{X}_j-\tilde{W}_j\beta)^T\right]\\ & +\frac{1}{2}Tr\left[V_1^{-1}J_{(n_1\times n_1)}\otimes\begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix}V_1^{-1}J_{(n_1\times n_1)}\otimes\begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix}\right]\\ & -Tr\left[V_2^{-1}J_{(n_2\times n_2)}\otimes\begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix}V_2^{-1}J_{(n_2\times n_2)}\otimes\begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix}V_2^{-1}(\tilde{X}_j-\tilde{W}_j\beta)(\tilde{X}_j-\tilde{W}_j\beta)^T\right]\\ & +\frac{1}{2}Tr\left[V_2^{-1}J_{(n_2\times n_2)}\otimes\begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix}V_2^{-1}J_{(n_2\times n_2)}\otimes\begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix}\right]\end{aligned} \tag{6.5.11}$$

The second order partial derivatives follow similarly for the remaining parameters in θ.

Taking the expected value and after doing some simplifications, we obtain the following.

$$E\Big[\frac{\partial l(x|\beta,\theta)}{\partial^2\tau_{CC}}\Big] \propto -\frac{1}{2}vec^T\left[J_{(n_1\times n_1)}\otimes\begin{pmatrix}1&0\\0&0\end{pmatrix}\right]\left[V_1^{-1}\otimes V_1^{-1}\right]vec\left[J_{(n_1\times n_1)}\otimes\begin{pmatrix}1&0\\0&0\end{pmatrix}\right]$$
$$-\frac{1}{2}vec^T\left[J_{(n_2\times n_2)}\otimes\begin{pmatrix}1&0\\0&0\end{pmatrix}\right]\left[V_2^{-1}\otimes V_2^{-1}\right]vec\left[J_{(n_2\times n_2)}\otimes\begin{pmatrix}1&0\\0&0\end{pmatrix}\right] \quad (6.5.12)$$

$$E\Big[\frac{\partial l(x|\beta,\theta)}{\partial^2\tau_{CE}}\Big] \propto -\frac{1}{2}vec^T\left[J_{(n_1\times n_1)}\otimes\begin{pmatrix}0&1\\1&0\end{pmatrix}\right]\left[V_1^{-1}\otimes V_1^{-1}\right]vec\left[J_{(n_1\times n_1)}\otimes\begin{pmatrix}0&1\\1&0\end{pmatrix}\right]$$
$$-\frac{1}{2}vec^T\left[J_{(n_2\times n_2)}\otimes\begin{pmatrix}0&1\\1&0\end{pmatrix}\right]\left[V_2^{-1}\otimes V_2^{-1}\right]vec\left[J_{(n_2\times n_2)}\otimes\begin{pmatrix}0&1\\1&0\end{pmatrix}\right] \quad (6.5.13)$$

$$E\Big[\frac{\partial l(x|\beta,\theta)}{\partial^2\tau_{EE}}\Big] \propto -\frac{1}{2}vec^T\left[J_{(n_1\times n_1)}\otimes\begin{pmatrix}0&0\\0&1\end{pmatrix}\right]\left[V_1^{-1}\otimes V_1^{-1}\right]vec\left[J_{(n_1\times n_1)}\otimes\begin{pmatrix}0&0\\0&1\end{pmatrix}\right]$$
$$-\frac{1}{2}vec^T\left[J_{(n_2\times n_2)}\otimes\begin{pmatrix}0&0\\0&1\end{pmatrix}\right]\left[V_2^{-1}\otimes V_2^{-1}\right]vec\left[J_{(n_2\times n_2)}\otimes\begin{pmatrix}0&0\\0&1\end{pmatrix}\right] \quad (6.5.14)$$

$$E\Big[\frac{\partial l(x|\beta,\theta)}{\partial^2\sigma_{j1CC}}\Big] \propto -\frac{1}{2}vec^T\left[\begin{pmatrix}I_{(n_{j1}\times n_{j1})}&0\\0&0\end{pmatrix}\otimes\begin{pmatrix}1&0\\0&0\end{pmatrix}\right]\left[V_j^{-1}\otimes V_j^{-1}\right]vec\left[\begin{pmatrix}I_{(n_{j1}\times n_{j1})}&0\\0&0\end{pmatrix}\otimes\begin{pmatrix}1&0\\0&0\end{pmatrix}\right] \quad (6.5.15)$$

$$E\Big[\frac{\partial l(x|\beta,\theta)}{\partial^2\sigma_{j1CE}}\Big] \propto -\frac{1}{2}vec^T\Bigg[\begin{pmatrix} I_{(n_{j1}\times n_{j1})} & 0 \\ 0 & 0\end{pmatrix}\otimes\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}\Bigg]\Big[V_j^{-1}\otimes V_j^{-1}\Big]vec\Bigg[\begin{pmatrix} I_{(n_{j1}\times n_{j1})} & 0 \\ 0 & 0\end{pmatrix}\otimes\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}\Bigg] \tag{6.5.16}$$

$$E\Big[\frac{\partial l(x|\beta,\theta)}{\partial^2\sigma_{j1EE}}\Big] \propto -\frac{1}{2}vec^T\Bigg[\begin{pmatrix} I_{(n_{j1}\times n_{j1})} & 0 \\ 0 & 0\end{pmatrix}\otimes\begin{pmatrix}0 & 0\\ 0 & 1\end{pmatrix}\Bigg]\Big[V_j^{-1}\otimes V_j^{-1}\Big]vec\Bigg[\begin{pmatrix} I_{(n_{j1}\times n_{j1})} & 0 \\ 0 & 0\end{pmatrix}\otimes\begin{pmatrix}0 & 0\\ 0 & 1\end{pmatrix}\Bigg] \tag{6.5.17}$$

$$E\Big[\frac{\partial l(x|\beta,\theta)}{\partial^2\sigma_{j2CC}}\Big] \propto -\frac{1}{2}vec^T\Bigg[\begin{pmatrix} 0 & 0 \\ 0 & I_{(n_{j2}\times n_{j2})}\end{pmatrix}\otimes\begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix}\Bigg]\Big[V_j^{-1}\otimes V_j^{-1}\Big]vec\Bigg[\begin{pmatrix} 0 & 0 \\ 0 & I_{(n_{j2}\times n_{j2})}\end{pmatrix}\otimes\begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix}\Bigg] \tag{6.5.18}$$

$$E\Big[\frac{\partial l(x|\beta,\theta)}{\partial^2\sigma_{j2CE}}\Big] \propto -\frac{1}{2}vec^T\Bigg[\begin{pmatrix} 0 & 0 \\ 0 & I_{(n_{j2}\times n_{j2})}\end{pmatrix}\otimes\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}\Bigg]\Big[V_j^{-1}\otimes V_j^{-1}\Big]vec\Bigg[\begin{pmatrix} 0 & 0 \\ 0 & I_{(n_{j2}\times n_{j2})}\end{pmatrix}\otimes\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}\Bigg] \tag{6.5.19}$$

$$E\Big[\frac{\partial l(x|\beta,\theta)}{\partial^2\sigma_{j2CE}}\Big] \propto -\frac{1}{2}vec^T\Bigg[\begin{pmatrix} 0 & 0 \\ 0 & I_{(n_{j2}\times n_{j2})}\end{pmatrix}\otimes\begin{pmatrix}0 & 0\\ 0 & 1\end{pmatrix}\Bigg]\Big[V_j^{-1}\otimes V_j^{-1}\Big]vec\Bigg[\begin{pmatrix} 0 & 0 \\ 0 & I_{(n_{j2}\times n_{j2})}\end{pmatrix}\otimes\begin{pmatrix}0 & 0\\ 0 & 1\end{pmatrix}\Bigg] \tag{6.5.20}$$

$$E\Big[\frac{\partial l(x|\beta,\theta)}{\partial\tau_{CC}\partial\sigma_{j2CC}}\Big] \propto -\frac{1}{2}vec^T\Bigg[J_{(n_j\times n_j)}\otimes\begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix}\Bigg]\Big[V_j^{-1}\otimes V_j^{-1}\Big]vec\Bigg[\begin{pmatrix} 0 & 0 \\ 0 & I_{(n_{j2}\times n_{j2})}\end{pmatrix}\otimes\begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix}\Bigg] \tag{6.5.21}$$

$$E\Big[\frac{\partial l(x|\beta,\theta)}{\partial\sigma_{j2CC}\partial\sigma_{j1CC}}\Big] \propto -\frac{1}{2}vec^T\left[\begin{pmatrix} I_{(n_{j1}\times n_{j1})} & 0 \\ 0 & 0\end{pmatrix}\otimes\begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix}\right]\left[V_j^{-1}\otimes V_j^{-1}\right]vec\left[\begin{pmatrix}0 & 0\\ 0 & I_{(n_{j2}\times n_{j2})}\end{pmatrix}\otimes\begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix}\right] \tag{6.5.22}$$

$$\frac{\partial l(x|\beta,\theta)}{\partial\sigma_{12CC}\partial\sigma_{21CC}} = 0 \tag{6.5.23}$$

Completing the above derivation for all pairs of unknown elements in θ we obtain the Fisher information matrix for θ, which can be expressed in terms of H and V^{**}. In particular, H_{j0} and H_{j1} can be expressed as follows.

$$\begin{aligned} H_{j0} &= [vec(G_{j1}), vec(G_{j2}), vec(G_{j3})] \\ H_{j1} &= [vec(G_{j4}), vec(G_{j5}), vec(G_{j6}), vec(G_{j7}), vec(G_{j8}), vec(G_{j9})], \end{aligned} \tag{6.5.24}$$

where

$$G_{j1} = J_{(n_1\times n_1)}\otimes\begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix},\ G_{j2} = J_{(n_1\times n_1)}\otimes\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix} \tag{6.5.25}$$

$$G_{j3} = J_{(n_1\times n_1)}\otimes\begin{pmatrix}0 & 0\\ 0 & 1\end{pmatrix},\ G_{j4} = \begin{pmatrix} I_{(n_{j1}\times n_{j1})} & 0 \\ 0 & 0\end{pmatrix}\otimes\begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix} \tag{6.5.26}$$

$$G_{j5} = \begin{pmatrix} I_{(n_{j1}\times n_{j1})} & 0 \\ 0 & 0\end{pmatrix}\otimes\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix},\ G_{j6} = \begin{pmatrix} I_{(n_{j1}\times n_{j1})} & 0 \\ 0 & 0\end{pmatrix}\otimes\begin{pmatrix}0 & 0\\ 0 & 1\end{pmatrix} \tag{6.5.27}$$

$$G_{j7} = \begin{pmatrix} 0 & 0 \\ 0 & I_{(n_{j2}\times n_{j2})} \end{pmatrix} \otimes \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \; G_{j8} = \begin{pmatrix} 0 & 0 \\ 0 & I_{(n_{j2}\times n_{j2})} \end{pmatrix} \otimes \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \tag{6.5.28}$$

$$G_{j9} = \begin{pmatrix} 0 & 0 \\ 0 & I_{(n_{j2}\times n_{j2})} \end{pmatrix} \otimes \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \tag{6.5.29}$$

Using the above representation for H, we find that the Fisher information matrix, I_θ can be expressed as:

$$I_\theta = \frac{1}{2}\begin{pmatrix} H_{10}^T(V_1^*)^{-1}H_{10} + H_{20}^T(V_2^*)^{-1}H_{20} & H_{10}^T(V_1^*)^{-1}H_{11} & H_{20}^T(V_2^*)^{-1}H_{21} \\ H_{11}^T(V_1^*)^{-1}H_{10} & H_{11}^T(V_1^*)^{-1}H_{11} & 0 \\ H_{21}^T(V_2^*)^{-1}H_{20} & 0 & H_{21}^T(V_2^*)^{-1}H_{21} \end{pmatrix} = \frac{1}{2}\Big[H^T(V^{**})^{-1}H\Big] \tag{6.5.30}$$

Hence,

$$Cov(\hat{\theta}) = I_\theta^{-1} = 2\Big[H^T(V^{**})^{-1}H\Big]^{-1} \tag{6.5.31}$$

We note, that the results derived for Cov($\hat{\beta}$) and Cov($\hat{\theta}$) are analogous to the results provided by Goldstein (2011). We refer the reader to Appendix 2.1 of Goldstein's book.

6.5.2 Appendix B- Assessing positive definiteness of V_j

It was noted that in order to verify the positive definiteness of V_j, it is enough to verify if all the $\Sigma_{jk}s$ are positive definite and all A_js are positive definite, for $j = 1, \ldots, g$ and $k = 1, 2$, where A_j is defined in (6.1.20). We shall now prove this assertion. The proof will be presented for the case of two centers. The proof for the general case is similar.

For $k = 1,\, 2$, let Q_{jk} be a $2n_{jk} \times 2n_{jk}$ orthogonal matrix whose first two columns are given by $\frac{1}{\sqrt{n_{jk}}}1_{n_{jk}} \otimes I_2$, where $1_{n_{jk}}$ is an $n_{jk} \times 1$ matrix of ones. Let $Q_j = \text{diag}\,(Q_{j1}, Q_{j2})$

so that Q_j is a $2n_{j.} \times 2n_{j.}$ orthogonal matrix. By direct multiplication it can be verified that

$$Q_j^T V_j Q_j = \begin{pmatrix} Q_{j1}^T V_{j1} Q_{j1} & Q_{j1}^T \left[J_{(n_{j1} \times n_{j2})} \otimes \Psi \right] Q_{j2} \\ Q_{j2}^T \left[J_{(n_{j2} \times n_{j1})} \otimes \Psi \right] Q_{j1} & Q_{j2}^T V_{j2} Q_{j2} \end{pmatrix}, \tag{6.5.32}$$

where $J_{m_1 \times m_2}$ is an $m_1 \times m_2$ matrix of ones. The above expression for $Q_j^T V_j Q_j$ can be explicitly written as

$$Q_j^T V_j Q_j = \left[\begin{array}{c|c} \begin{pmatrix} \Sigma_{j1} + n_{j1}\Psi & 0 & 0 & 0 \\ 0 & \Sigma_{j1} & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & \Sigma_{j1} \end{pmatrix}_{2n_{j1} \times 2n_{j1}} & \begin{pmatrix} \sqrt{n_{j1}n_{j2}}\Psi & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix}_{2n_{j1} \times 2n_{j2}} \\ \hline \begin{pmatrix} \sqrt{n_{j1}n_{j2}}\Psi & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix}_{2n_{j12} \times 2n_{j1}} & \begin{pmatrix} \Sigma_{j2} + n_{j2}\Psi & 0 & 0 & 0 \\ 0 & \Sigma_{j2} & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & \Sigma_{j2} \end{pmatrix}_{2n_{j12} \times 2n_{j2}} \end{array} \right]$$

Now let F_j be an $n_{j.} \times n_{j.}$ permutation matrix obtained by permuting the 2^{nd} column and $(n_{jk}+1)^{th}$ column of $I_{n_{j.}}$. We can then verify the following:

$$\left(F_j \otimes I_2 \right)^T \left[Q_j^T V_j Q_j \right] \left(F_j \otimes I_2 \right) = \text{diag}\left(A_j, I_{n_{j1}-1} \otimes \Sigma_{j1}, I_{n_{j2}-1} \otimes \Sigma_{j2} \right). \tag{6.5.33}$$

It should now be clear that the positive definiteness of V_j is equivalent to that of the matrices A_j, Σ_{j1} and Σ_{j2}.

Bibliography

Baltussen, R., Niessen, L. (2006). Priority setting of health interventions: the need for multi-criteria decision analysis. *Cost Effectiveness and Resource Allocation*, 4:14.

Barton, G.R., Briggs, A.H., Fenwick, E.A. (2008). Optimal cost-effectiveness decisions: The role of the cost-effectiveness acceptability curve (CEAC), the cost-effectiveness acceptability frontier (CEAF), and the expected value of perfection information (EVPI). *Value in Health*, 5, 886−897.

Basagana, X. et al. (2018). Analysis of multicentre epidemiological studies: contrasting fixed or random effects modelling and meta-analysis, *International Journal of Epidemiology*, 47, 1343−1354.

Bebu, I., Mathew, T. (2008). Comparing the means and variances of a bivariate log-normal distribution. *Statistics in Medicine*, 27, 2684−2696.

Bebu, I., Luta, G., Mathew, T., Agan, B.K. (2016). Generalized confidence intervals and fiducial intervals for some epidemiological measures. *International Journal of Environmental Research and Public Health*, 13, 605.

Bebu I., Luta G., Mathew, T., Kennedy P.A., Agan B.K. (2016). Parametric cost-effectiveness inference with skewed data. *Computational Statistics & Data Analysis*, 94, 210−220.

Bebu, I., Mathew, T., Lachin, J.M. (2016). Probabilistic measures of cost-effectiveness. *Statistics in Medicine*, 35, 3976−3986.

Bebu, I., Mathew, T., Lachin, J.M., Agan, B.K. (2016). Generalized confidence intervals and fiducial intervals for some epidemiological measures. *International Journal of Environmental Research and Public Health*, 13, 605−617.

Briggs, A.H., O'Brien, B.J., Blackhouse, G. (2002). Thinking Outside the Box: Recent Advances in the Analysis and Presentation of Uncertainty in Cost-Effectiveness Studies. *Annual Review of Public Health*, 23, 377−401.

Broekhuizen, H. et al. (2017). Public Preferences for Lung Cancer Screening Policies. *Value in Health*, 20, 961−968.

Creed, F. et al. (2003). The cost-effectiveness of psychotherapy and paroxetine for severe irritable bowel syndrome. *Gastroenterology*, 124, 303−317.

Espinoza, M.A., Manca, A., Claxton, K., Sculpher, M. (2018). Social value and individual choice: The value of a choice-based decision-making process in a collectively funded health system. *Health Economics*, 27, 28−40.

Faries, D., Leon, A., Haro, J., Obenchain, R. (2010). *Analysis of Observational Health Care Data Using SAS*. SAS Institute.

Goldstein, H. (1986). Multilevel mixed linear model analysis using iterative generalized least squares. *Biometrika*, 73, 43−56.

Goldstein, H. (1989). Restricted unbiased iterative generalized least-squares estimation. *Biometrika*, 76, 622−623.

Goldstein, H. (2011). *Multilevel Statistical Models*, Chichester: John Wiley & Sons.

Hannig, J. (2009). On generalized fiducial inference. *Statistica Sinica*, 19, 491−544.

Hannig, J., Iyer, H., Lai, R.C.S., Lee, T.C.M. (2016). Generalized fiducial inference: A review and new results. *Journal of the American Statistical Association*, 111, 1346−1361.

Huang, I.B., Keisler, J., Linkov, I. (2011). Multi-criteria decision analysis in environmental sciences: Ten years of applications and trends. *Science of The Total Environment*, 409, 3578−3594.

Hussain, A.M., vJørgensen, M.M, Østerdal, L.P. (2016). Refining Population Health Comparisons: A Multidimensional First Order Dominance Approach. *Social Indicators Research: An International and Interdisciplinary Journal for Quality-of-Life Measurement*, 129, 739−759.

Hoch, J.S., Dewa, C.S. (2014). Advantages of the net benefit regression framework for economic evaluations of interventions in the workplace: A case study of the cost-effectiveness of a collaborative mental health care program for people receiving short-term disability benefits for psychiatric disorders. *Journal of Occupational and Environmental Medicine*, 56, 441−445.

Hoch, J.S., Briggs A.H., Willan AR. (2002). Something old, something new, something borrowed, something blue: a framework for the marriage of health econometrics and cost-effectiveness analysis. *Health Economics*, 11, 415−430.

Hounton, S., Newlands, D. (2012). Applying the net-benefit framework for analyzing and presenting cost-effectiveness analysis of a maternal and newborn health intervention. *PloS one*, 7(7), e40995.

Ioannidis, J.P.A., Garber, A.M. (2011). Individualized cost-effectiveness analysis. *PLOS Medicine*, 8(7):e1001058.

Kahan, B.C., Harhay, M.O. (2015). Many multicentre trials had few events per centre, requiring analysis via random-effects models or GEEs. *Journal of Clinical Epidemiology*, 68, 1504−1511.

Koerkamp, B.G. (2009). Uncertainty in medical decision making: Knowing how little you know.

Kowalski, J., Tu, X.M. (2008). *Modern applied U-Statistics*, Hoboken, NJ : John Wiley.

Krishnamoorthy, K., Lee, M. (2010). Inference for functions of parameters in discrete distributions based on fiducial approach: Binomial and Poisson cases. *Journal of Statistical Planning and Inference*, 140, 1182-1192.

Krishnamoorthy, K., Mathew, T. (2009). *Statistical tolerance regions: theory, applications, and computation*, Hoboken, NJ : John Wiley.

Kruizenga, H.M. et al. (2005). Effectiveness and cost-effectiveness of early screening and treatment of malnourished patients. *The American Journal of Clinical Nutrition*, 82, 1082−1089.

Lehmann, E.L. (1999). *Elements of large-sample theory*, New York, NY: Springer-Verlag.

Leshno, M., Levy, H. (2004). Stochastic Dominance and Medical Decision Making. *Health Care Management Science*, 7, 207−215.

Manca, A., Rice, N., Sculpher, M.J., Briggs, A.H. (2005). Assessing generalisability by location in trial-based cost-effectiveness analysis: the use of multilevel models. *Health Economics*, 14, 471−485.

Manju, M. A., Candel, M.J., Berger, M.P. (2015). Optimal and maximin sample sizes for multicentre cost-effectiveness trials. *Statistical Methods in Medical Research*, 24, 513−539.

Mihaylova, B., Briggs, A., O'Hagan, A., Thompson, S.G. (2011). Review of statistical methods for analysing healthcare resources and costs. *Health Economics*, 20, 897−916.

Moerbeek, M., van Breukelen, G.J.P., Berger, M.P.F. (2003). A comparison between traditional methods and multilevel regression for the analysis of multicenter intervention studies. *Journal of Clinical Epidemiology*, 56, 341−350.

O Meachair S., Walsh, C. (2014). Uncertainty and probabilistic methods in multi-criteria decision analysis. *Value Health*, 17(7), A579.

Petri, M., et al. (2005). Combined Oral Contraceptives in Women with Systemic Lupus Erythematosus. *New England Journal of Medicine*, 353, 2550−2558.

Schey, C., Krabbe, P.F.M., Postma, M.J., Connolly, M.P. (2017). Multi-criteria decision analysis (MCDA): testing a proposed MCDA framework for orphan drugs. *Orphanet Journal of Rare Diseases*,12.

Stinnett, A.A., Mullahy, J. (1998). Net Health Benefits: A New Framework for the Analysis of Uncertainty in Cost-Effectiveness Analysis. *Medical Decision Making*, 18, S68−S80.

Thokala, P. et al. (2016). Multiple Criteria Decision Analysis for Health Care Decision Making-An Introduction: Report 1 of the ISPOR MCDA Emerging Good Practices Task Force. *Value in Health*, 19, 1−13.

Thokala, P., Duenas, A. (2012). Multiple Criteria Decision Analysis for Health Technology Assessment. *Value in Health*,15, 1172−1181.

Thompson, S. G., Turner, R.M., Warn, D.E. (2001). Multilevel models for meta-analysis, and their application to absolute risk differences. *Statistical Methods in Medical Research*, 10, 375−392.

Tunis, S.L. et al. (2006). Cost- effectiveness of olanzapine as first line treatment for schizophrenia: results from a randomized, open- label, 1 year trial. *Value in Health*, 9, 77−89.

Veiga, A., Smith, P.W.F., Brown, J.J. (2014). The use of sample weights in multivariate multilevel models with an application to income data collected by using a rotating panel survey. *Royal Statistical Society, Series C*, 63, 65−84.

Weerahandi, S. (1993). Generalized confidence intervals. *Journal of the American Statistical Association*, 88, 899−905.

Wen, S., Zhang, L., Yang, B. (2014). Two Approaches to Incorporate Clinical Data Uncertainty into Multiple Criteria Decision Analysis for Benefit-Risk Assessment of Medicinal Products. *Value in Health*, 17, 619−628.

Willan, A., Briggs A. (2006). *Statistical analysis of cost-effectiveness data*, Hoboken, NJ : John Wiley.

Willan, A.R., Briggs, A.H., Hoch, J.S. (2004). Regression methods for covariate adjustment and subgroup analysis for non-censored cost-effectiveness data. *Health Economics*, 13, 461−475.

Willan, A.R., Lin, D.Y., Manca, A. (2005). Regression methods for cost-effectiveness analysis with censored data. *Statistics in Medicine*, 24, 131−145.

Wu, M., Hsu, J.C., Tseng, Y. (2018). The Past, Present and Future of Multi-Criteria Decision Analysis (MCDA) Applied in Medical Field. *Value in Health*, 21, S220−S221.

Yu, Q., Tang, W., Kowalski, J., Tu, X.M. (2011). Multivariate U-statistics: a tutorial with applications. *WIREs Computational Statistics*, 3, 457−471.