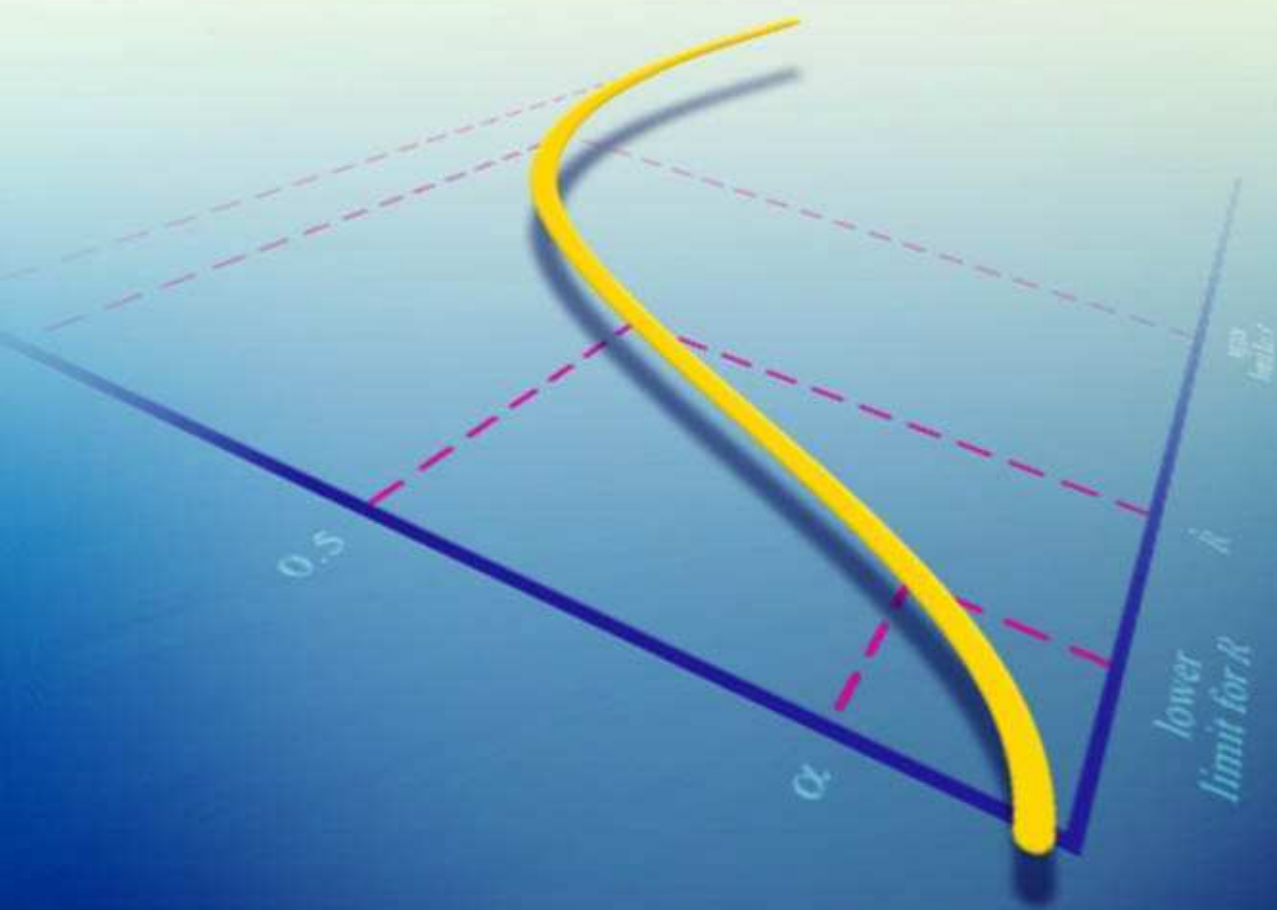


ANDREW R. WILLAN AND ANDREW H. BRIGGS

Statistical Analysis

of Cost-effectiveness Data



 WILEY

STATISTICS IN PRACTICE

***Statistical Analysis of
Cost-effectiveness Data***

STATISTICS IN PRACTICE

Advisory Editors

Stephen Senn

University of Glasgow, UK

Marian Scott

University of Glasgow, UK

Peter Bloomfield

North Carolina State University, USA

Founding Editor

Vic Barnett

Nottingham Trent University, UK

Statistics in Practice is an important international series of texts which provide detailed coverage of statistical concepts, methods and worked case studies in specific fields of investigation and study.

With sound motivation and many worked practical examples, the books show in down-to-earth terms how to select and use an appropriate range of statistical techniques in a particular practical field within each title's special topic area.

The books provide statistical support for professionals and research workers across a range of employment fields and research environments. Subject areas covered include medicine and pharmaceuticals; industry, finance and commerce; public services; the earth and environmental sciences, and so on.

The books also provide support to students studying statistical courses applied to the above areas. The demand for graduates to be equipped for the work environment has led to such courses becoming increasingly prevalent at universities and colleges.

It is our aim to present judiciously chosen and well-written workbooks to meet everyday practical needs. The feedback of views from readers will be most valuable to monitor the success of this aim.

A complete list of titles in this series appears at the end of the volume.

Statistical Analysis of Cost-effectiveness Data

Andrew R. Willan

University of Toronto, Canada

Andrew H. Briggs

University of Glasgow, UK



John Wiley & Sons, Ltd

Copyright © 2006 John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester,
West Sussex PO19 8SQ, England

Telephone (+44) 1243 779777

Email (for orders and customer service enquiries): cs-books@wiley.co.uk

Visit our Home Page on www.wiley.com

All Rights Reserved. No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning or otherwise, except under the terms of the Copyright, Designs and Patents Act 1988 or under the terms of a licence issued by the Copyright Licensing Agency Ltd, 90 Tottenham Court Road, London W1T 4LP, UK, without the permission in writing of the Publisher.

Requests to the Publisher should be addressed to the Permissions Department, John Wiley & Sons, Ltd, The Atrium, Southern Gate, Chichester, West Sussex PO19 8SQ, England, or emailed to permreq@wiley.co.uk, or faxed to (+44) 1243 770620.

Designations used by companies to distinguish their products are often claimed as trademarks. All brand names and product names used in this book are trade names, service marks, trademarks or registered trademarks of their respective owners. The Publisher is not associated with any product or vendor mentioned in this book.

This publication is designed to provide accurate and authoritative information in regard to the subject matter covered. It is sold on the understanding that the Publisher is not engaged in rendering professional services. If professional advice or other expert assistance is required, the services of a competent professional should be sought.

Other Wiley Editorial Offices

John Wiley & Sons Inc., 111 River Street, Hoboken, NJ 07030, USA

Jossey-Bass, 989 Market Street, San Francisco, CA 94103-1741, USA

Wiley-VCH Verlag GmbH, Boschstr. 12, D-69469 Weinheim, Germany

John Wiley & Sons Australia Ltd, 42 McDougall Street, Milton, Queensland 4064, Australia

John Wiley & Sons (Asia) Pte Ltd, 2 Clementi Loop #02-01, Jin Xing Distripark,
Singapore 129809

John Wiley & Sons Canada Ltd, 22 Worcester Road, Etobicoke, Ontario, Canada M9W 1L1

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic books.

Library of Congress Cataloging-in-Publication Data

(to follow)

British Library Cataloguing in Publication Data

A catalogue record for this book is available from the British Library

ISBN-13 978-0-470-85626-0 (HB)

ISBN-10 0-470-85626-2 (HB)

Typeset in 11/13pt Photina by TechBooks, New Delhi, India

Printed and bound in Great Britain by TJ International Ltd, Padstow, Cornwall

This book is printed on acid-free paper responsibly manufactured from sustainable forestry in which at least two trees are planted for each one used for paper production.

For Bernie

***All author proceeds donated
to the Emma and Lucy O'Brien Education Fund.***

Contents

Preface	xi
1 Concepts	1
1.1 Introduction	1
1.2 Cost-effectiveness data and the parameters of interest	2
1.3 The cost-effectiveness plane, the ICER and INB	5
1.4 Outline	8
2 Parameter Estimation for Non-censored Data	11
2.1 Introduction	11
2.2 Cost	12
2.2.1 Sample means for estimating incremental cost	12
2.2.2 Using multiple regression models	14
2.2.3 Transformation (and the retransformation problem)	15
2.2.4 Generalized linear models	17
2.2.5 Two-part models for excess zeros	18
2.2.6 Cost prediction models	19
2.3 Effectiveness	20
2.3.1 Probability of surviving	21
2.3.2 Mean survival time	21
2.3.3 Mean quality-adjusted survival time	22
2.3.4 Mean quality-adjusted survival: controlling for baseline utility	24
2.4 Summary	25
3 Parameter Estimation for Censored Data	27
3.1 Introduction	27
3.2 Mean Cost	28
3.2.1 Direct (Lin) method	29
3.2.2 Inverse-probability weighting	31
3.3 Effectiveness	34
3.3.1 Probability of surviving	34

viii Contents

3.3.2	Mean survival time	36
3.3.3	Mean quality-adjusted survival time	39
3.4	Summary	42
4	Cost-effectiveness Analysis	43
4.1	Introduction	43
4.2	Incremental cost-effectiveness ratio	44
4.3	Incremental net benefit	49
4.4	The cost-effectiveness acceptability curve	51
4.5	Using bootstrap methods	54
4.6	A Bayesian incremental net benefit approach	57
4.7	Kinked thresholds	60
4.8	Summary	64
5	Cost-effectiveness Analysis: Examples	67
5.1	Introduction	67
5.2	The CADET-Hp trial	67
5.3	Symptomatic hormone-resistant prostate cancer	72
5.4	The Canadian implantable defibrillator study (CIDS)	77
5.5	The EVALUATE trial	82
5.6	Bayesian approach applied to the UK PDS study	86
5.7	Summary	90
6	Power and Sample Size Determination	93
6.1	Introduction	93
6.2	Approaches based on the cost-effectiveness plane	94
6.2.1	Briggs and Gray	95
6.2.2	Willan and O'Brien	98
6.2.3	Gardiner <i>et al.</i>	101
6.3	The classical approach based on net benefit	103
6.3.1	The method	103
6.3.2	Example	105
6.4	Bayesian take on the classical approach	106
6.4.1	The Method	106
6.4.2	Example	107
6.5	The value of information approach	108
6.5.1	The method	108
6.5.2	Example	114
6.6	Summary	116
7	Covariate Adjustment and Sub-group Analysis	117
7.1	Introduction	117
7.2	Non-censored data	118
7.2.1	Example, non-censored data	121

7.3	Censored data	129
7.3.1	Cost	131
7.3.2	Quality-adjusted survival time	132
7.3.3	Survival time	134
7.3.4	The Canadian implantable defibrillator study (CIDS)	135
7.3.5	The evaluate trial	138
7.4	Summary	142
8	Multicenter and Multinational Trials	145
8.1	Introduction	145
8.2	Background to multinational cost-effectiveness	147
8.3	Fixed effect approaches	151
8.3.1	Willke <i>et al.</i>	151
8.3.2	Cook <i>et al.</i>	152
8.4	Random effects approaches	154
8.4.1	Aggregate level analysis: multicenter trials	154
8.4.2	Aggregate level analysis: multinational trials	156
8.4.3	Hierarchical modeling	162
8.5	Summary	164
9	Modeling Cost-effectiveness	165
9.1	Introduction	165
9.2	A general framework for modeling cost-effectiveness results	166
9.3	Case study: an economic appraisal of the goal study	167
9.3.1	The GOAL study	168
9.3.2	Standard approach to estimating cost-effectiveness	170
9.3.3	An alternative approach to estimating cost-effectiveness	171
9.3.4	Comparing the two analyses of GOAL	179
9.4	Summary	180
	References	183
	Author Index	193
	Subject Index	195
	Series List	197

Preface

This book describes statistical methods applied to cost-effectiveness analysis. It represents the experience over many years of the author's involvement in the application and methodology of health economic evaluation. The focus on randomised clinical trials reflects the fact that the trend towards collecting not only clinical, but also economic, data alongside clinical trials was the driving force behind many of the methodological developments described in the text. Health economics is a relatively young discipline and the use of clinical trials as a vehicle for economic evaluations began in earnest only twenty years ago. As a consequence, there has been a high degree of methodological development since then, with most of the reporting confined to journal articles. The aim of this book is to draw together those developments in a single source which we hope will be of interest to students of statistics, keen to understand more about health economics, and students of health economics, keen to understand the statistical methods required for undertaking economic evaluation of health care interventions. The exposition is at a technical level roughly equivalent to that found in final year undergraduate mathematics and statistics courses or postgraduate social sciences courses.

The book itself naturally divides into two parts. The first part (up to Chapter 5) deals with the established approach for the presentation of cost-effectiveness analyses, with a focus on estimating health outcomes and resource use costs. The second part of the book (Chapters 6 through 9) handles specific issues in more depth to give a fuller understanding of the nuances of a modern cost-effectiveness analysis where patient-level data are available.

xii ***Preface***

In the preparation of any book there are numerous colleagues and students who have provided the inspiration and insight, as well as friends and family who have provided the encouragement and support, necessary to bring such a project to fruition. We are extremely grateful to all those people who have helped us over the years and aided us to a greater or lesser extent in supporting our endeavours and correcting our mistakes. However, one person stands out as the true inspiration for this book. A friend and colleague who had a major influence on both of our careers in the area of health economic evaluation, albeit in different ways, Bernie O'Brien was a rare person – someone with a keen intellect, an infectious enthusiasm, and a generosity of ideas that could not fail to rub off on those around him. His untimely death on the 13th of February, 2004 was a terrible shock and leaves a vacuum in the health economics community, as well as for his wife Karen and daughters, Emma and Lucy. We dedicate this book to Bernie's memory.

1

Concepts

1.1 INTRODUCTION

There is a growing expectation from health care policymakers that evidence supporting the cost-effectiveness of new health care interventions, particularly pharmaceuticals, be provided along with the customary data on efficacy and safety. In Australia (Commonwealth of Australia, 1990) and Canada (Detsky, 1993) there are formal requirements that pharmaceutical companies present evidence of cost-effectiveness before a drug is granted reimbursement status on a formulary. In the United States there is demand for such economic data from third-party insurers, see Leaf (1989).

There are two general approaches to performing an economic evaluation of a health care intervention, see O'Brien (1996). One approach combines the efficacy and safety data from randomized clinical trials (RCTs) with cost data from secondary, non-trial sources in a decision analysis model. In such models the problem of inferential uncertainty is addressed using sensitivity analyses to determine what effect varying the model assumptions has on the results, see Briggs *et al.* (1994). The other approach uses health care utilization data collected on individual patients prospectively as part of an RCT. The health care utilization data combined with the appropriate price weights yield a measure of cost for each patient. Measuring effectiveness and cost at the patient level permits the use of more conventional methods of statistical inference to quantify the uncertainty due to sampling and measurement error. Since the early 1990s, when such data became

2 Concepts

more common, numerous articles have been published in the area of the statistical analysis of cost-effectiveness data. Initially, efforts were concentrated on providing confidence intervals for incremental cost-effectiveness ratios, but more recently, due to concerns regarding ratio statistics, the concept of incremental net benefit has been proposed as an alternative.

The purpose of this book is to provide an illustrated summary of some of the key developments published in the last 10 years that deal with statistical issues related to the cost-effectiveness comparison of two groups when measures of effectiveness and cost are observed at the subject level. The context used throughout the book is that of patients in a two-arm RCT where patients are randomized to Treatment (T) or Standard (S), but the methods apply to the comparison of any two groups, subject to the concerns one might have regarding bias due to the lack of random group allocation.

1.2 COST-EFFECTIVENESS DATA AND THE PARAMETERS OF INTEREST

In a cost-effectiveness analysis (CEA), whether an incremental cost-effectiveness ratio (ICER) or an incremental net benefit approach is taken, five parameters need to be estimated. Two of the parameters are the differences between treatment arms of mean effectiveness and costs, denoted by Δ_e and Δ_c , respectively. The other three parameters are the variances and covariance of those estimators. With the estimators of these five parameters, a CEA, based on either the incremental cost-effectiveness ratio or incremental net benefit, can be performed. For non-censored data the estimators are simple functions of the sample means, variances and covariance. For censored data estimation procedures are decidedly more complex.

Typically, the measure of effectiveness in a CEA is associated with a clinical event experienced by the patient. Quite commonly the event is death, but it could be relapse or reaching a pre-specified level of symptom relief. For simplicity, unless otherwise noted, we assume that the event is death. The simplest measure of effectiveness based on event data is the probability of the event not occurring within a

specified period of time from randomization. The specified period of time is often referred to as the duration of interest and denoted by τ . Let the random variable D_{Ti} be the time from randomization to the event for the i th patient on arm T and let $S_T(t) = \Pr(D_{Ti} \geq t)$, then the measure of effectiveness is given by $S_T(\tau)$ and denoted as π_T . $S_T(t)$ is the survival function for patients on arm T . Defining D_{Si} , $S_S(t)$ and π_S similarly for patients on arm S , the parameter of interest for effectiveness, denoted Δ_e , is given by

$$\Delta_e = S_T(\tau) - S_S(\tau) = \pi_T - \pi_S \tag{1.1}$$

The quantity Δ_e is the absolute risk reduction, and $1/\Delta_e$ is the mean number of patients that need to be treated with T rather than S to prevent a death. This quantity is usually referred to informally as the ‘number-needed-to-treat’ or more simply as the NNT. If the probability of surviving 5 years is 0.6 for patients on arm T and only 0.5 for patients on arm S , then we say that 10 (i.e. $1/0.1$) patients need to be treated with T rather than S to prevent one death, or more simply that the NNT is 10.

Another measure of effectiveness based on event data is the mean survival time over the duration of interest, otherwise referred to as the restricted mean survival time. The restricted mean survival time is sensitive to the entire survival curve from 0 to τ , and not just its value at τ . The restricted mean survival time for a particular arm, denoted as μ_j , $j = T, S$, is the area under the respective survival curve from 0 to τ , i.e.

$$\mu_j = \int_0^\tau S_j(t)dt$$

and the parameter of interest to be estimated for effectiveness is given by

$$\Delta_e = \int_0^\tau S_T(t)dt - \int_0^\tau S_S(t)dt = \mu_T - \mu_S \tag{1.2}$$

For mean survival time quantity $1/\Delta_e$ is the NNT to gain one year of life over the duration of interest. If the restricted mean survival time over 5 years for a patient on arm T is 4 and only 3.75 for a patient on

4 Concepts

arm S , then the NNT to gain one year of life is 4 (i.e. $1/0.25$). The mean total restricted survival time of 4 patients on arm T is $4 \times 4 = 16$, while on arm S the mean total restricted survival time of 4 patients is $4 \times 3.75 = 15$.

The third measure of effectiveness based on survival data is the mean quality-adjusted survival time over the duration of interest, otherwise referred to as the restricted mean quality-adjusted survival time. Quality-adjusted survival time is based on the concept that patients experience, at any given time, a certain quality of life based on a utility scale for which 1 is perfect health and 0 is death, see Torrance (1986). Negative values are used to allow for states of health considered worse than death. If the quality of life at time t for a patient on a particular treatment arm is given by $Q_j(t)$, $j = T, S$, then the restricted mean quality-adjusted survival time is given by

$$\varphi_j \equiv \int_0^{\tau} Q_j(t) dt$$

and the parameter of interest to be estimated for effectiveness is given by

$$\Delta_e = \int_0^{\tau} Q_T(t) dt - \int_0^{\tau} Q_S(t) dt = \varphi_T - \varphi_S \quad (1.3)$$

For quality-adjusted survival time the quantity $1/\Delta_e$ is the NNT to gain one quality-adjusted life-year over the duration of interest.

The quantity Δ_e is the difference between treatment arms with respect to effectiveness, and is a different function of the survival curves, depending on which measure of effectiveness is of interest.

If we let ν_j be the mean cost over the duration of interest for a patient in arm j , $j = T, S$, then the parameter of interest to be estimated for cost is given by

$$\Delta_c = \nu_T - \nu_S \quad (1.4)$$

The observed cost for a given patient is simply the sum of the amounts of each resource consumed by the patient multiplied by the respective price weight. Which resources are included depends

on the perspective taken by the analysis. If the analysis takes the perspective of the health care system, only resources covered under the system would be included. However, if a broader societal perspective were taken, then costs not covered under the system and items such as time lost from work and care by a family member could also be included; for a fuller discussion the reader is referred to Drummond *et al.* (1997). Estimation methods for Δ_e , Δ_c and the corresponding variances and covariance are given in Chapter 2 for non-censored data and in Chapter 3 for censored data.

1.3 THE COST-EFFECTIVENESS PLANE, THE ICER AND INB

Researchers have long used the cost-effectiveness plane to explore the policy interpretation of cost-effectiveness analyses. The cost-effectiveness (CE) plane is a graph with Δ_c and Δ_e plotted on the vertical axis and horizontal axis, respectively, as illustrated in Figure 1.1. For more discussion on the cost-effectiveness plane the reader is referred to Black (1990). Let $\Delta = (\Delta_e, \Delta_c)^T$. If, for a particular Treatment/Standard comparison, the point Δ is located in the Southeast (SE) quadrant (i.e. $\Delta_e > 0$, $\Delta_c < 0$), Treatment is said to dominate Standard because it is more effective and less costly, and the argument to adopt it to replace Standard is self-evident. By contrast, if Δ lies in the Northwest (NW) quadrant (i.e. $\Delta_e < 0$, $\Delta_c > 0$) Treatment is dominated by Standard, and its rejection as a replacement for Standard is the rational policy choice. It is in the Northeast (NE) and Southwest (SW) quadrants, referred to as the trade-off quadrants, that the magnitudes of Δ_e and Δ_c need to be considered to determine if Treatment is cost-effective.

To assist in this determination researchers have traditionally used the incremental cost-effectiveness ratio. The ICER is defined as $R \equiv \Delta_c / \Delta_e$, but can be written as

$$\frac{1}{\Delta_e} \Delta_c = \text{NNT} \times \Delta_c$$

It is easy to see then that the ICER is the product of the number of patients that need to be given Treatment to achieve an extra unit

6 Concepts

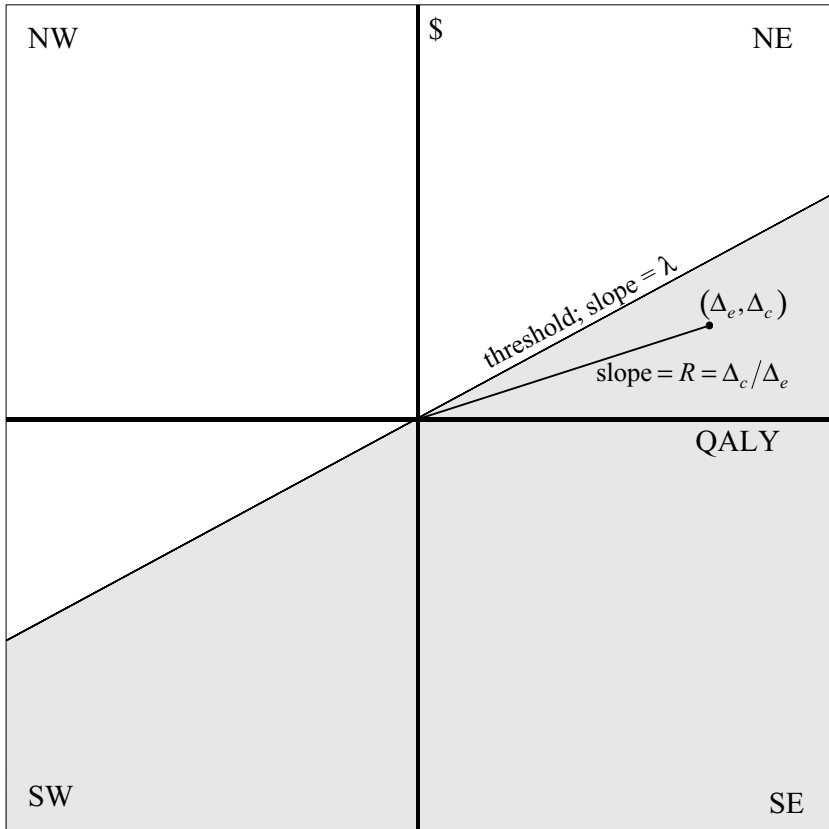


Figure 1.1 The cost-effectiveness plane

of effectiveness and the incremental cost of treating each of those patients, and is therefore the incremental cost of achieving a unit of effectiveness from using Treatment rather than Standard. On the CE plane the ICER is the slope of the line between the origin and the point Δ , see Figure 1.1. If the measure of effectiveness is the probability of surviving then the ICER is cost of saving a life (or preventing a death). If the measure of effectiveness is mean survival or mean quality-adjusted survival, then the ICER is the cost of achieving an extra year or quality-adjusted year of life (QALY), respectively. Essentially the ICER is the cost of an additional unit of effectiveness if Treatment is adopted over

Standard. This, as in any transaction, needs to be compared with what a policymaker is willing to pay.

The amount a policymaker is willing to pay is referred to as the willingness-to-pay (WTP), and is denoted by λ . The concept of WTP is discussed by Pauly (1995), and methods for quantifying it can be found in O'Brien and Gafni (1996), Johnson *et al.* (1998) and Hanley *et al.* (2003). By drawing a line through the origin with slope λ , the CE plane can be divided into two regions. For convenience this line will be referred to as the threshold. For points on the plane below and to the right of the threshold (the shaded area in Figure 1.1), Treatment is considered cost-effective, but for those above and to the left it is not. Since λ is positive, points in the SE quadrant are always below the threshold and therefore correspond to comparisons for which Treatment is cost-effective. On the other hand, points in the NW are always above the threshold and correspond to comparisons for which Treatment is not cost-effective. It is in the NE and SW quadrants that the concept of WTP allows for trade-off between effectiveness and costs. In the NE quadrant the slope of any point below the line is less than λ , i.e. $\Delta_c/\Delta_e < \lambda$ which implies that $\Delta_c < \Delta_e\lambda$. Therefore, the increase in value ($\Delta_e\lambda$) is greater than the increase in cost, making Treatment cost-effective. In the SW quadrant the slope of any point below the line is greater than λ , and since Δ_e and Δ_c are both negative (i.e. treatment is less effective and less costly), we have $\Delta_c/\Delta_e = |\Delta_c|/|\Delta_e| > \lambda$ which implies that $|\Delta_c| > |\Delta_e\lambda|$. Therefore, the value lost ($|\Delta_e\lambda|$) is less than the amount saved ($|\Delta_c|$), making Treatment cost-effective. In summary, Treatment is cost-effective if

$$A : \frac{\Delta_c}{\Delta_e} < \lambda \quad \text{if } \Delta_e > 0; \quad \text{or} \quad \frac{\Delta_c}{\Delta_e} > \lambda \quad \text{if } \Delta_e < 0 \quad (1.5)$$

Expression (1.5) (Hypothesis A) defines the region below the threshold and can be thought of as the alternative hypothesis for the null Hypothesis H, given by:

$$H : \frac{\Delta_c}{\Delta_e} \geq \lambda \quad \text{if } \Delta_e > 0; \quad \text{or} \quad \frac{\Delta_c}{\Delta_e} \leq \lambda \quad \text{if } \Delta_e < 0 \quad (1.6)$$

Rejecting H in favour of A would provide evidence to adopt Treatment. These expressions are somewhat awkward and can be simplified considerably by the introduction of incremental net benefit.

8 Concepts

The incremental net benefit (INB) is a function of λ , and is defined as

$$b_\lambda \equiv \Delta_e \lambda - \Delta_c \quad (1.7)$$

b_λ is the incremental net benefit because it is the difference between incremental value ($\Delta_e \lambda$) and incremental cost (Δ_c). Treatment is cost-effective if, and only if, $b_\lambda > 0$, regardless of the sign of Δ_e . To see this, both inequalities involving the ICER in Expression (1.5) can be rearranged to the inequality $\Delta_e \lambda - \Delta_c > 0$. Similarly, both inequalities involving the ICER in Expression 1.6 can be rearranged to the inequality $\Delta_e \lambda - \Delta_c \leq 0$. Therefore, in terms of INB the null and alternative hypotheses become

$$H : \Delta_e \lambda - \Delta_c \leq 0 \quad \text{versus} \quad A : \Delta_e \lambda - \Delta_c > 0 \quad (1.8)$$

On the CE plane b_λ is the vertical distance from the point Δ to the threshold, being positive if it is below the line and negative otherwise. Because it has slope λ , the point on the threshold with abscissa equal to Δ_e is $(\Delta_e, \Delta_e \lambda)$ and so the vertical distance between it and Δ is $\Delta_e \lambda - \Delta_c$, see Figure 1.2.

The incremental net health benefit (INHB) is defined as $\Delta_e - \Delta_c / \lambda = b_\lambda / \lambda$ and measures net benefit in units of effectiveness. Since INHB is simply a positive constant times INB, statistical inference made on one will be identical to statistical inference made on the other. INB has the advantage of being linear in λ . Therefore in a sensitivity analysis of WTP, the plots of INB by λ , are straight lines. Another advantage of INB is that it generalizes to more than one outcome. In a trial of patients at risk of thrombosis, if Δ_{e1} , Δ_{e2} and Δ_{e3} are the differences of the probability of avoiding death, thrombosis and stroke, respectively, and if λ_1 , λ_2 and λ_3 are the corresponding WTP values, then INB is defined as $\Delta_{e1} \lambda_1 + \Delta_{e2} \lambda_2 + \Delta_{e3} \lambda_3 - \Delta_c$. A corresponding formulation in INHB is not possible.

1.4 OUTLINE

The remainder of the book is organized as follows. Methods for estimating Δ_e and Δ_c and their variances and covariances for non-censored data are given in Chapter 2. The methods make use of simple statistics,

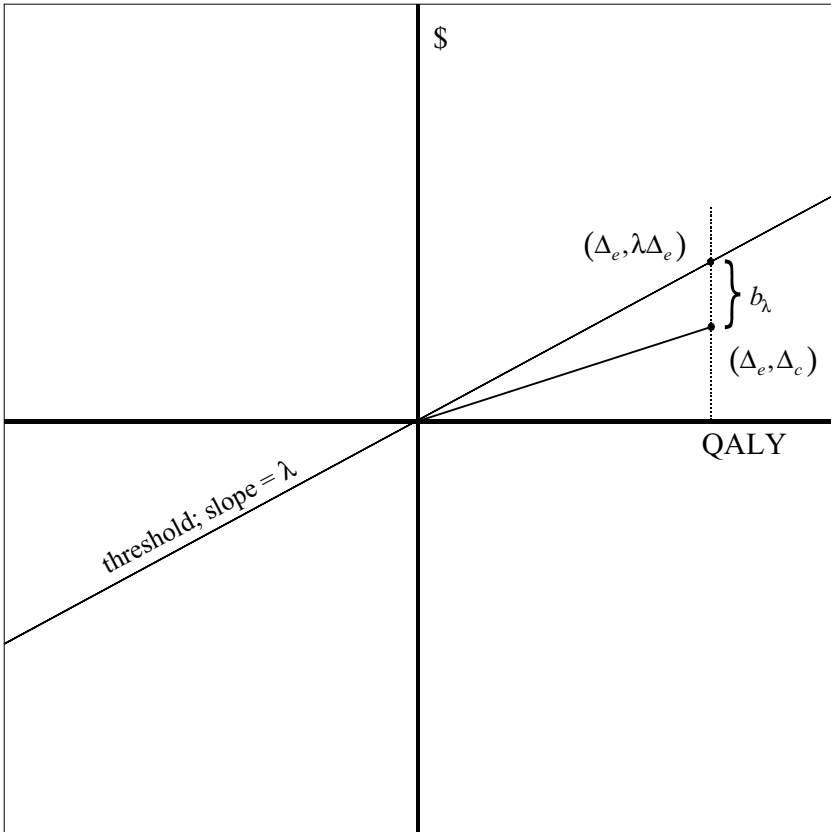


Figure 1.2 INB on the cost-effectiveness plane

such as proportions and sample means and variances. Estimation methods for censored data are given in Chapter 3. The methods include life-table procedures, the direct method of Lin *et al.* (1997) and inverse probability weighting. How the parameters are used in a cost-effectiveness analysis is described in Chapter 4. Emphasis is placed on estimating the ICER and INB, along with their confidence limits, and constructing cost-effectiveness acceptability curves. In Chapter 5 the methods of Chapters 2, 3 and 4 are illustrated with examples. Methods for determining sample sizes, using both classical and Bayesian

10 *Concepts*

approaches, are given in Chapter 6. In Chapter 7 regression methods for covariate adjustment and testing for treatment by prognostic factor interactions are described, along with several examples. The issues regarding multicenter and multinational trials are the subject of Chapter 8. In Chapter 9 a more general framework of statistical modeling is proposed, which is based on modeling the separate components cost-effectiveness to build indirect estimates of incremental cost and effectiveness.

Parameter Estimation for Non-censored Data

2.1 INTRODUCTION

Data from a clinical trial are said to be non-censored if all patients were followed either until death or for the total duration of interest. This implies that no patients were lost to follow-up and that the delay in time between when the last patient was randomized and when the analysis was conducted was sufficiently long that all patients were followed for the duration of interest. Consequently, the measure of effectiveness and total cost are known for all patients. Trials of treatments for acute health events, such as myocardial infarction and asthma exacerbations, are likely to have non-censored data since the duration of interest is usually relatively short. As the duration lengthens, however, the probability that some patients will be lost to follow-up increases. In addition, for trials with long durations of interest and staggered patient entry, analysis will likely occur prior to all patients having been followed for the entire duration of interest, resulting in administrative censoring.

The statistical methods for estimating the five parameters required to conduct a cost-effectiveness analysis when the data are non-censored are given in the remaining sections of this chapter. For brevity, in the remainder of the chapter and throughout the rest of the book the terms 'mean cost' or 'mean effectiveness' will be used in place of the term 'restricted (to the duration of interest) mean'.

12 *Parameter estimation for non-censored data*

Estimation of the between-treatment difference in mean cost is dealt with in the Section 2.2, beginning with the common approach in clinical trials of using the difference in sample means. However, in practice cost data can be skewed and will often have excess zeros, which can lead to inefficiency in the standard least-squares approach. The most common approaches advocated for overcoming these problems are outlined in the remainder of the section. All three functions of survival—probability of surviving, mean survival time and mean quality-adjusted survival time—are considered in Section 2.3. Particular attention is given to the importance of controlling for baseline utility in the estimation of QALY gains. Because the data are non-censored, only simple statistics, such as proportions and sample means, variances and covariances, are required. The methods for handling censoring are the subject of Chapter 3.

2.2 COST

Incremental cost is often estimated by the between-treatment difference in the sample means. However, the nature of health service resource use is that cost data are often highly skewed and can exhibit a large proportion of zero values. Nevertheless, the fundamental interest of the economist in means was outlined in the Chapter 1 and leads to a rejection of rank order statistics, such as the median, and associated rank sum tests for location, when analyzing such data. Possible solutions include the use of transformation and smearing, the direct transformation of the expectation of cost and the use of two-part models for separating out zero observations.

2.2.1 Sample means for estimating incremental cost

For non-censored data the total cost over the duration of interest is observed for all patients, and the mean cost in each arm of the trial is often estimated by the sample mean. If we let C_{ji} be the observed total cost over the duration of interest for the j th patient, where the j th patient is the i th patient on the j th arm, where $j = T, S$; $i =$

1, 2, . . . n_j; and n_j is the number of patients randomized to the jth arm, then v_j, the mean cost for the jth arm, is estimated by

$$\hat{v}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} C_{ji}$$

Therefore, the between-arm difference in mean cost is estimated by

$$\hat{\Delta}_c = \hat{v}_T - \hat{v}_S = \frac{1}{n_T} \sum_{i=1}^{n_T} C_{Ti} - \frac{1}{n_S} \sum_{i=1}^{n_S} C_{Si} \quad (2.1)$$

The variance of $\hat{\Delta}_c$ is estimated by

$$\begin{aligned} \hat{V}(\hat{\Delta}_c) &= \hat{V}(\hat{v}_T) + \hat{V}(\hat{v}_S) \\ &= \frac{1}{n_T(n_T - 1)} \sum_{i=1}^{n_T} (C_{Ti} - \hat{v}_T)^2 + \frac{1}{n_S(n_S - 1)} \sum_{i=1}^{n_S} (C_{Si} - \hat{v}_S)^2 \end{aligned} \quad (2.2)$$

The formulation in Equation (2.2) allows the between-patient variance in cost to differ between arms.

Because of the right-skewing which is usually present in cost data, criticism is often leveled at the use of least-squares methods such as sample means and variances, see O’Hagan and Stevens (2003), Briggs and Gray (1998a), Thompson and Barber (2000), Nixon and Thompson (2005) and Briggs *et al.* (2005), and transformations, such as the logarithm and square root, are sometimes proposed. However, such transformations provide estimates on a scale not relevant to decision makers, see Manning and Mullahy (2001) and Thompson and Barber (2000).

One approach to this issue is to consider ‘rules of thumb’ for when skewness may cause concern for standard least-squares methods of analysis. Although it is commonly considered that the *central limit theorem* applies for samples with greater than 30 observations, ensuring a normal distribution of the sample mean whatever the distribution in the underlying population, this rule of thumb may not apply to non-symmetric distributions typical of cost data. An alternative rule of thumb for situations where the ‘principal deviation from normality consists of marked positive skewness’ is provided by Cochran (1977) with the suggestion that $n > 25\eta^2$ where η is the skewness coefficient

14 *Parameter estimation for non-censored data*

in the sample and n is the sample size. The guideline was devised such that a 95% confidence interval will have an error probability no greater than 6%.

Additionally, a number of investigations into the issue of skewed data, using mostly simulated data, have drawn the conclusion that least-squares methods provide valid estimates of mean cost and the between-arm difference in mean cost. Lumley *et al.* (2002) provide a review of such investigations. Willan and O'Brien (1996) and Willan, Briggs and Hoch (2004) address specifically the issue of skewed cost data and find that, even when cost data is distributed as log-normal to the base 10, the distribution of $\hat{\Delta}_c$ exhibits little skewness and kurtosis for sample sizes as small as 25 per arm.

Nonetheless, the blind application of sample means and variances to cost data with extreme outliers is likely to lead to misleading conclusions. Faith in the robustness of least-squares methodology is no substitute for careful examination of the data using box-plots and histograms. Furthermore, although least-squares methods may provide valid estimators of mean cost, the estimators may be inefficient in the presence of right skewing.

2.2.2 Using multiple regression models

In the presence of high levels of skewness, it is natural to consider whether, by modeling cost on covariates, the skewness in the data can be explained. It is helpful to recognize that, at the simplest level, the simple comparison of means from Section 2.2.1 can be represented by the following linear model

$$C_i = \alpha + \Delta_c t_i + \varepsilon_i$$

where α is an intercept term, t_i a treatment dummy, taking the values zero for the standard treatment (S) and one for the new treatment (T), and ε a random error term. The coefficient Δ_c on the treatment dummy gives the estimated incremental cost of treatment exactly as in Equation (2.1). Similarly, the standard error of the coefficient will be the square root of the variance given in Equation (2.2).

The advantage of the regression framework is that it is straightforward to add covariates in addition to the treatment indicator. For example, baseline patient characteristics measured prior to randomization can be employed to make allowance for prognostic information in the treatment comparison using the multiple regression framework given by

$$C_i = \alpha + \sum_{k=1}^p \beta_k x_{ik} + \Delta_c t_i + \varepsilon_i,$$

where x_{ik} is value of the k th covariate for the i th patient. As before, the coefficient Δ_c for the treatment indicator gives the incremental cost controlling for the covariates. In the context of an experimental design such as a randomized controlled trial, the randomization process is expected to ensure a balance of both observed and unobserved potentially confounding factors across the treatment arms. Therefore, the use of prognostic covariates will not usually materially affect the magnitude of the estimated incremental cost, but may account for some of the skewness and improve the precision of the estimate, leading to narrower confidence intervals, see Altman (1985) and Pocock (1984).

2.2.3 Transformation (and the retransformation problem)

In practice, adding covariates to a cost regression rarely results in a model with a high explanatory power. The consequence is that residual skewness in the error term of the regression often remains and consideration can be given to transforming the cost with the aim of fitting a superior model. However, with any transformation of the data, it is important to recognize that health care policy decisions must be made concerning costs on the untransformed scale and so retransformation from the scale of estimation back to the original cost scale will be required. Consider fitting a linear model to some transformation of the cost data $Z_i = g(C_i)$. Although the back-transformation $h(\cdot) = g^{-1}(\cdot)$ can be employed it is well-known that $E[h(\cdot)] \neq h(E[\cdot])$

16 *Parameter estimation for non-censored data*

for non-linear transformations, therefore estimating coefficients from the regression model

$$Z_i = \alpha^Z + \Delta_c^Z t_i + \varepsilon_i$$

in order to predict the cost on the untransformed scale as

$$\hat{C}_i = h(\hat{\alpha}^Z + \Delta_c^Z t_i)$$

would give a biased estimate since

$$\begin{aligned} E[h(Z_i)] &= E[C_i] \\ &= E[h(\alpha^Z + \Delta_c t_i + \varepsilon_i)] \\ &\neq E[h(\alpha^Z + \Delta_c t_i)] \end{aligned}$$

2.2.3.1 A Taylor series approximation

Commonly, the Taylor series (or Delta) method is used to provide an approximation of the expectation of a random variable under a non-linear transformation. Letting $\mu_Z = E[Z]$, we have:

$$\begin{aligned} E[h(Z)] &= E[h(\mu_Z + Z - \mu_Z)] \\ &= E \left[h(\mu_Z) + (Z - \mu_Z) h'(\mu_Z) + \frac{(Z - \mu_Z)^2}{2!} h''(\mu_Z) \right. \\ &\quad \left. + \frac{(Z - \mu_Z)^3}{3!} h'''(\mu_Z) + \dots \right] \simeq h(\mu_Z) + \frac{1}{2} h''(\mu_Z) \text{var}[Z] \end{aligned}$$

Thus the Taylor series approximation to the expectation on the untransformed scale suggests using a bias correction term of one half of the second derivative of the back-transformation function $h(\cdot)$ multiplied by the variance of Z , which can be estimated by $\text{var}(\hat{\varepsilon}_i^Z)$.

2.2.3.2 A non-parametric ‘smearing’ estimator

While the Taylor series approximation may be adequate, there is an alternative approach known as non-parametric smearing that does not require a specific distribution for the error term, see Duan (1983). Note that the expectation of the untransformed cost can be written as a functional of the cumulative distribution function (cdf) of the errors

on the transformed scale

$$\begin{aligned} E [C_i] &= E [h (\alpha^Z + \Delta_c^Z t_i + \varepsilon_i^Z)] \\ &= \int h (\alpha^Z + \Delta_c^Z t_i + \varepsilon_i^Z) dF (\varepsilon_i^Z), \end{aligned}$$

where $F (\varepsilon_i^Z)$ represents the cumulative distribution function (cdf) of the error on the transformed scale. Rather than assume a parametric form for the cdf, it can be estimated by the empirical cdf of the estimated residuals

$$\hat{F}_t (e) = \frac{1}{n} \sum_{i=1}^n I \{ \hat{\varepsilon}_i^Z \leq e \},$$

where $I \{ \cdot \}$ is the indicator function, $\hat{\varepsilon}_i^Z = Z_i - (\hat{\alpha}^Z + \hat{\Delta}_c^Z t_i)$, $Z_i = g(C_i)$, $n = n_S + n_T$, and $\hat{\alpha}^Z$ and $\hat{\Delta}_c^Z$ are the OLS estimated regression coefficients of Z_i on t_i .

The expected cost $E [C_i]$ is estimated by substituting the empirical cdf for the unknown cdf and substituting the OLS estimates of the regression coefficients to yield

$$\hat{E} [C_i] = \int h (\hat{\alpha}^Z + \hat{\Delta}_c^Z t_i + \hat{\varepsilon}_i^Z) d\hat{F}_t (\hat{\varepsilon}_i^Z) = \frac{1}{n} \sum_{i=1}^n h (\hat{\alpha}^Z + \hat{\Delta}_c^Z t_i + \hat{\varepsilon}_i^Z),$$

which gives the ‘smearing’ estimate for the expectation on the untransformed scale and provides an estimate of the between-treatment difference given by

$$\hat{\Delta}_c = \frac{1}{n} \sum_{i=1}^n h (\hat{\alpha}^Z + \hat{\Delta}_c^Z + \hat{\varepsilon}_i^Z) - \frac{1}{n} \sum_{i=1}^n h (\hat{\alpha}^Z + \hat{\varepsilon}_i^Z).$$

2.2.4 Generalized linear models

An alternative approach to transforming the data is to work within the class of generalized linear models (GLMs) where a linear predictor, $x^T \beta$, is related to the expectation of the outcome of interest through a link function $g(\cdot)$, such that $g (E [y|x]) = x^T \beta$. Since, in the GLM framework, it is the expectation that is subject to the transformation (rather than the data), back-transformation to the

18 *Parameter estimation for non-censored data*

original scale is straightforward using the inverse of the link function: $E[y|x] = g^{-1}(x^T \beta)$. Given the fundamental interest of the economist in the mean cost/expenditure, this makes the GLM class of models particularly attractive. In addition, GLMs are extremely flexible, allowing a range of different distributions for the data to be coupled with different link functions.

For the purposes of the exposition here, the multiplicative log link GLM is assumed such that the model for cost is given by

$$\ln(E[C_i]) = \alpha + \sum_{k=1}^p \beta_k x_{ik} + \Delta_c t_i$$

or, equivalently

$$\begin{aligned} E[C_i] &= \exp \left\{ \alpha + \sum_{j=1}^p \beta_j x_{ij} + \Delta_c t_i \right\} \\ &= \exp \left\{ \alpha + \sum_{k=1}^p \beta_k x_{ik} \right\} \cdot \exp \{ \Delta_c t_i \} \end{aligned} \quad (\text{Model 4})$$

What should be clear for the expression given above, is that the coefficient on the treatment dummy from a multiplicative model of cost is a factor, such that the incremental cost is obtained by multiplying this factor by the average cost in the absence of treatment. This means that the estimated incremental cost will vary according to the baseline cost in the standard care arm. If covariates have an important impact on this baseline cost, then the incremental cost of treatment may also vary substantially by covariate pattern. Thus, it is the lack of treatment interaction with baseline covariates in a multiplicative model that indicates important subgroup effects.

2.2.5 Two-part models for excess zeros

One approach to the problem of excess zeros is to form a two-part model. In the first part of the model, a logistic GLM is employed using the indicator variable for positive cost to predict which costs will be

positive and which costs will be zero. This gives the first part of the model as

$$\ln\left(\frac{\pi_i}{1 - \pi_i}\right) = \alpha^\pi + \Delta^\pi t_i$$

with the predicted probability that the cost is positive given by

$$\pi_i = \frac{\exp\{\alpha^\pi + \Delta^\pi t_i\}}{1 + \exp\{\alpha^\pi + \Delta^\pi t_i\}}$$

The second part of the model then involves modeling the positive costs only using the transformed positive costs C_i^+ , the transformed positive costs $Z_i^+ = g(C_i^+)$ using OLS, or most flexibly with a GLM for the expectation of positive cost.

The overall predicted value for a two-part model is simply the product of the expectations of its two parts, given by

$$\hat{C}_i = \hat{\pi}_i \hat{C}_i^+ = \frac{\exp\{\hat{\alpha}^\pi + \hat{\Delta}^\pi t_i\}}{1 + \exp\{\hat{\alpha}^\pi + \hat{\Delta}^\pi t_i\}} g^{-1}\left(\hat{\alpha}^{C^+} + \hat{\Delta}^{C^+} t_i\right)$$

where the regression of positive cost is assumed undertaken from a GLM with $g^{-1}(\cdot)$ representing the inverse of the link function. The difference between treatment arms is estimated by

$$\frac{\exp\{\hat{\alpha}^\pi + \hat{\Delta}^\pi\}}{1 + \exp\{\hat{\alpha}^\pi + \hat{\Delta}^\pi\}} g^{-1}\left(\hat{\alpha}^{C^+} + \hat{\Delta}^{C^+}\right) - \frac{\exp\{\hat{\alpha}^\pi\}}{1 + \exp\{\hat{\alpha}^\pi\}} g^{-1}\left(\hat{\alpha}^{C^+}\right)$$

A similar result holds if the second part of the two-part model is based on transformed cost; however, as discussed above, consideration should be given to a Taylor series or smearing correction in order to give an unbiased estimators for untransformed cost.

2.2.6 Cost prediction models

Collecting all health care utilization data in a clinical trial can be very expensive. As an alternative, investigators may choose to collect a small number of high-impact utilization variables, and use them, along with some baseline and outcome variables, to predict total cost

20 *Parameter estimation for non-censored data*

for each patient. This approach relies on the existence of the results of a regression analysis using a previously existing dataset, in which total cost has been regressed on the predictor variables in question.

Let x_{jki} be the observed value of predictor k ($k = 1, \dots, p$) on patient i ($i = 1, \dots, n_j$) from treatment arm j ($j = T, S$). Then $x_{ji} = (x_{j1i}, x_{j2i}, \dots, x_{jpi})^T$ is the vector of predictor variables for patient i from treatment arm j . Let $\bar{x}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} x_{ji}$, $\bar{x} = \bar{x}_T - \bar{x}_S$, $\hat{\Sigma}_{xj}$ be the sample variance-covariance matrix based on the x_{ji} values, and $\hat{\Sigma}_x = \frac{1}{n_T} \hat{\Sigma}_{xT} + \frac{1}{n_S} \hat{\Sigma}_{xS}$. Let $\hat{\beta}$ be the vector of dimension p of parameter estimates corresponding to the p predictor variables resulting from the regression analysis performed on the existing data, with estimated variance-covariance matrix given by $\hat{\Sigma}_\beta$. The quantity $\hat{\beta}_0 + \hat{C}_{ji}$ is the predicted cost for patient i from treatment arm j , where $\hat{C}_{ji} = x_{ji}^T \hat{\beta}$ and $\hat{\beta}_0$ is the estimate of the intercept from the regression analysis performed on the existing data. The estimator of between-arm difference in mean cost is $\bar{\hat{C}}_T - \bar{\hat{C}}_S$, where $\bar{\hat{C}}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} \hat{C}_{ji}$. A naïve estimator for the variance of $\bar{\hat{C}}_T - \bar{\hat{C}}_S$ is given by

$$\frac{\sum_{i=1}^{n_S} (\hat{C}_{Si} - \bar{\hat{C}}_S)^2}{n_S (n_S - 1)} + \frac{\sum_{i=1}^{n_T} (\hat{C}_{Ti} - \bar{\hat{C}}_T)^2}{n_T (n_T - 1)}$$

but this ignores the fact that $\hat{\beta}$ is a random vector. The proper estimator for the variance of $\bar{\hat{C}}_T - \bar{\hat{C}}_S$ is given by $x^T \hat{\Sigma}_\beta x + \hat{\beta}^T \hat{\Sigma}_x \hat{\beta} + \text{trace}(\hat{\Sigma}_\beta \hat{\Sigma}_x)$, (see Willan and O'Brien, 2001). Setting $\hat{\Sigma}_\beta = 0$ yields the naïve estimator given above.

2.3 EFFECTIVENESS

In dealing with the effectiveness side of cost-effectiveness analysis, consideration is given to each of the three possible measures of effectiveness typically employed in clinical trials that include economic analyses: probability of surviving, mean survival time and mean quality adjusted survival time. Due to the role of baseline utility measures in the QALY calculus, the appropriate adjustment of QALY measures is also discussed.

2.3.1 Probability of surviving

Let the random variable D_{ji} be the time from randomization to death for the j th patient, and let $S_j(t) = \Pr(D_{ji} \geq t)$, then the probability of surviving the duration of interest for a patient on the j th arm is given by $S_j(\tau)$, which is denoted more simply by π_j . For non-censored data $S_j(\tau)$ is estimated by the proportion of patients who survived the duration of interest. Let $\bar{\delta}_{ji} = 1$ if the j th patient was observed to survive the duration of interest, and 0 if not. That is, $\bar{\delta}_{ji} = I\{D_{ji} \geq \tau\}$, where $I\{\text{expression}\}$ is the indicator function, equaling 1 if the expression is true, and 0 otherwise. The probability of surviving the duration of interest for a patient on the j th arm is estimated by $\hat{\pi}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} \bar{\delta}_{ji}$. Therefore, the estimated between-arm difference is given by

$$\hat{\Delta}_e = \hat{\pi}_T - \hat{\pi}_S = \frac{1}{n_T} \sum_{i=1}^{n_T} \bar{\delta}_{Ti} - \frac{1}{n_S} \sum_{i=1}^{n_S} \bar{\delta}_{Si} \quad (2.3)$$

The variance of $\hat{\Delta}_e$ is estimated by

$$\hat{V}(\hat{\Delta}_e) = \hat{V}(\hat{\pi}_T) + \hat{V}(\hat{\pi}_S) = \frac{\hat{\pi}_T(1 - \hat{\pi}_T)}{n_T} + \frac{\hat{\pi}_S(1 - \hat{\pi}_S)}{n_S} \quad (2.4)$$

and the covariance between $\hat{\Delta}_e$ and $\hat{\Delta}_c$ by

$$\begin{aligned} \text{Cov}(\hat{\Delta}_e, \hat{\Delta}_c) &= \text{Cov}(\hat{\pi}_T, \hat{v}_T) + \text{Cov}(\hat{\pi}_S, \hat{v}_S) \\ &= \frac{\left[\left(\sum_{i=1}^{n_T} \bar{\delta}_{Ti} C_{Ti} \right) - n_T \hat{\pi}_T \hat{v}_T \right]}{n_T(n_T - 1)} \\ &\quad + \frac{\left[\left(\sum_{i=1}^{n_S} \bar{\delta}_{Si} C_{Si} \right) - n_S \hat{\pi}_S \hat{v}_S \right]}{n_S(n_S - 1)} \end{aligned} \quad (2.5)$$

2.3.2 Mean survival time

Since there is no censoring, all patients have their survival time observed and sample means can be used to estimate the mean survival time. Let $X_{ji} = \min(D_{ji}, \tau)$; i.e. X_{ji} is the duration of interest or the time

22 Parameter estimation for non-censored data

from randomization to death, whichever is smaller. The mean survival for a patient on the j th treatment, denoted by μ_j , is estimated by: $\hat{\mu}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} X_{ji}$. Therefore, the estimated between-arm difference is given by

$$\hat{\Delta}_e = \hat{\mu}_T - \hat{\mu}_S = \frac{1}{n_T} \sum_{i=1}^{n_T} X_{Ti} - \frac{1}{n_S} \sum_{i=1}^{n_S} X_{Si} \quad (2.6)$$

The variance of $\hat{\Delta}_e$ is estimated by

$$\begin{aligned} \hat{V}(\hat{\Delta}_e) &= \hat{V}(\hat{\mu}_T) + \hat{V}(\hat{\mu}_S) \\ &= \frac{1}{n_T(n_T - 1)} \sum_{i=1}^{n_T} (X_{Ti} - \hat{\mu}_T)^2 + \frac{1}{n_S(n_S - 1)} \sum_{i=1}^{n_S} (X_{Si} - \hat{\mu}_S)^2 \end{aligned} \quad (2.7)$$

and the covariance between $\hat{\Delta}_e$ and $\hat{\Delta}_c$ by

$$\begin{aligned} \text{C}\hat{\text{ov}}(\hat{\Delta}_e, \hat{\Delta}_c) &= \text{C}\hat{\text{ov}}(\hat{\mu}_T, \hat{v}_T) + \text{C}\hat{\text{ov}}(\hat{\mu}_S, \hat{v}_S) \\ &= \frac{\sum_{i=1}^{n_T} (X_{Ti} - \hat{\mu}_T)(C_{Ti} - \hat{v}_T)}{n_T(n_T - 1)} \\ &\quad + \frac{\sum_{i=1}^{n_S} (X_{Si} - \hat{\mu}_S)(C_{Si} - \hat{v}_S)}{n_S(n_S - 1)} \end{aligned} \quad (2.8)$$

2.3.3 Mean quality-adjusted survival time

To measure a patient's quality-adjusted life-years (QALYs) over the duration of interest, we assume that the patients quality of life (QoL) is measured at various times during this period, most likely, though not necessarily at $t = 0$ (at randomization), at $t = \tau$, and perhaps at times in between. The measurements are on a scale in which 1 corresponds to perfect health, 0 to death, and negative values to states of health worse than death. For a more complete discuss of measuring quality of life, the reader is referred to Weinstein and Stason (1977) and Torrance (1976; 1986; 1987).

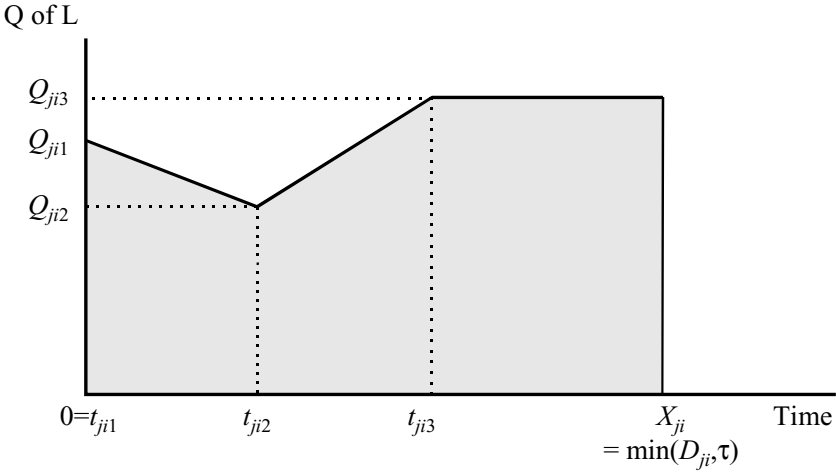


Figure 2.1 Quality of life by time

Suppose there are m_{ji} QoL measurements on the j th patient, taken at times: $0 \leq t_{j1} < t_{j2} < \dots < t_{jm_{ji}} \leq \tau$, with corresponding values $Q_{j1}, Q_{j2}, \dots, Q_{jm_{ji}}$. The observed QALY on the j th patient, denoted by q_{ij} , is simply the area under the curve, between 0 and X_{ji} , of the plot of the Q_{ij} by the t_{ij} values, as shown in Figure 2.1. (As would be expected with most clinical trials, in Figure 2.1 we have set $t_{j1} = 0$.) Therefore, $q_{ji} = \int_0^{X_{ji}} Q_{ji}(t)dt$, where $Q_{ji}(t)$ is defined as

$$Q_{ji}(t) = \begin{cases} Q_{j1} : & 0 \leq t < t_{j1} \\ Q_{jih} + \frac{(Q_{ji,h+1} - Q_{jih})(t - t_{jih})}{t_{ji,h+1} - t_{jih}} : & t_{jih} \leq t < t_{ji,h+1} \\ Q_{jm_{ji}} : & t_{jm_{ji}} \leq t < X_{ji} \\ 0 : & t \geq X_{ji} \end{cases} \quad (2.9)$$

More simply put, q_{ij} is just the sum of the areas of the shaded rectangles shown in Figure 2.1, i.e. $q_{ji} = (t_{j2} - t_{j1}) \times \left(\frac{Q_{j1} + Q_{j2}}{2}\right) + (t_{j3} - t_{j2}) \times \left(\frac{Q_{j2} + Q_{j3}}{2}\right) + (X_{ji} - t_{j3}) \times Q_{j3}$. If t_{j1} is greater than 0, the definition in Equation (2.9) sets the QoL between 0 and t_{j1} to a constant value of Q_{j1} . Similarly, if X_{ji} is greater than $t_{jm_{ji}}$ then Equation (2.9) sets the QoL between $t_{jm_{ji}}$ and X_{ji} to a constant value of $Q_{jm_{ji}}$.

24 Parameter estimation for non-censored data

The mean QALY for a patient on the j th arm, denoted by φ_j , is estimated by: $\hat{\varphi}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} q_{ji}$. Therefore, the estimated between-arm difference is given by:

$$\hat{\Delta}_e = \hat{\varphi}_T - \hat{\varphi}_S = \frac{1}{n_T} \sum_{i=1}^{n_T} q_{Ti} - \frac{1}{n_S} \sum_{i=1}^{n_S} q_{Si} \quad (2.10)$$

The variance of $\hat{\Delta}_e$ is estimated by

$$\begin{aligned} \hat{V}(\hat{\Delta}_e) &= \hat{V}(\hat{\varphi}_T) + \hat{V}(\hat{\varphi}_S) \\ &= \frac{1}{n_T(n_T - 1)} \sum_{i=1}^{n_T} (q_{Ti} - \hat{\varphi}_T)^2 + \frac{1}{n_S(n_S - 1)} \sum_{i=1}^{n_S} (q_{Si} - \hat{\varphi}_S)^2 \end{aligned} \quad (2.11)$$

and the covariance between $\hat{\Delta}_e$ and $\hat{\Delta}_c$ by

$$\begin{aligned} \text{Cov}(\hat{\Delta}_e, \hat{\Delta}_c) &= \text{Cov}(\hat{\varphi}_T, \hat{v}_T) + \text{Cov}(\hat{\varphi}_S, \hat{v}_S) \\ &= \frac{\sum_{i=1}^{n_T} (q_{Ti} - \hat{\varphi}_T)(C_{Ti} - \hat{v}_T)}{n_T(n_T - 1)} + \frac{\sum_{i=1}^{n_S} (q_{Si} - \hat{\varphi}_S)(C_{Si} - \hat{v}_S)}{n_S(n_S - 1)} \end{aligned} \quad (2.12)$$

2.3.4 Mean quality-adjusted survival time: controlling for baseline utility

In principle, the methods of Section 2.2.2 which described the use of regression models for cost apply equally to effectiveness outcomes, in that even where perfect balance is achieved in the randomization process, the inclusion of additional covariates into a regression framework can be expected to reduce the variance and improve the power of the analysis. However, in terms of QALY outcomes there is an important distinction to be made between baseline measures of covariates that, while prognostic for outcomes, are independent of treatment (through the randomization process) and baseline measures of utility, which forms part of the QALY calculus. Being part of the QALY calculation,

and therefore correlated with the QALY outcomes, any imbalance in baseline utility measures will potentially bias estimates of $\hat{\Delta}_e$ above.

In order to adjust appropriately for potential imbalance, the baseline utility can be added to the multiple regression equation to give the QALY estimate as

$$q_i = \alpha + \sum_{k=1}^p \beta_k x_{ik} + \Delta_e t_i + \gamma Q_i(0) + \varepsilon_i$$

where t_i is a treatment indicator variable as before, $Q_i(0)$ is the baseline utility score, and the x_{ij} are other potential prognostic covariates (as in Section 2.2.2). Further discussion on the use of regression adjustment for handling imbalance at baseline can be found in Altman (1985) and Senn (1997). Manca *et al.* (2005a) discuss the baseline adjustment of QALYs in particular and present simulation results to show that regression adjustment outperforms simple measures of change from baseline, which is sometimes used as a method for adjusting for baseline imbalance in the outcome of interest.

2.4 SUMMARY

Methods for estimating the five parameters required for a cost-effectiveness comparison of two groups using non-censored data were discussed in this chapter. The main focus was on the simple comparison of two arms of a clinical trial, but emphasizing the potential importance of regression methods for increasing the precision of estimates of treatment effect, adjusting for imbalance, and avoiding bias associated with baseline utility imbalances affecting the QALY calculations. The potential treatment of skewness and excess zeros in the analysis of cost data were also discussed, with an emphasis on the use of GLMs to focus on the fundamental interest on mean values. The utilization of these five parameters in a cost-effectiveness analysis is the subject of Chapter 4, while the use of covariates in a cost-effectiveness analysis and the examination of sub-group effects is the subject of Chapter 7.

Parameter Estimation for Censored Data

3.1 INTRODUCTION

The analysis of a clinical trial involves censored data if some patients are not followed for the entire duration of interest. Some patients may be lost to follow-up, either because they refuse to attend follow-up clinic visits or because they move out of the jurisdiction covered by trial management resources. Because of staggered entry, patients also can be censored because the analysis is performed prior to them being followed for the duration of interest. This type of censoring, known as administrative, is present more frequently for trials with relatively long durations of interest, say 3–5 years.

Lost-to-follow-up (LTF) censoring is usually considered more problematic, since it is harder to assume that the distributions for the outcomes (cost and effectiveness) for LTF patients are the same as those not lost to follow-up. If the survival experience of those LTF does differ, then the censoring is informative, leading to biased parameter estimators. This potential source of bias is exacerbated if LTF rates or the causes of LTF differ between treatment arms. Administrative censoring is less problematic than FTF censoring if it can be assumed that patients randomized later in the trial, and therefore more likely to be censored administratively, have the same survival and cost distributions as those randomized earlier. This may not always be true, since co-interventions that improve survival or alter costs may become

28 *Parameter estimation for censored data*

available during the recruitment phase of the trial, improving the survival times or altering the costs for patients randomized later. Also, entry criteria are often changed during the trial, and consequently, survival and cost distributions may depend on when patients were randomized. Nonetheless, most trials with censored data tend to assume that censoring is uninformative, or more accurately, assume that the extent of the informative censored is unlikely to bias the estimates of treatment effect appreciably.

Parameter estimators for censored data are given in the remaining sections of this chapter. Standard life-table methods are used for probability of surviving and the mean survival time. Two methods are proposed for mean cost and mean quality-adjusted survival time. The direct (Lin) method makes use of the survival function combined with the cost histories, while the method of inverse-probability weighting makes use of the censoring function together with the cost histories. Inverse-probability weighting can also be used for estimating mean survival time.

Recall that D_{ji} is the time from randomization to death for the j th patient. Let U_{ji} be the time from randomization to censoring for the j th patient. The j th patient is censored if $U_{ji} \leq D_{ji}$ and is observed to die if $D_{ji} < U_{ji}$. Further, let $X_{ji} = \min(D_{ji}, U_{ji})$, $\delta_{ji} = I\{D_{ji} < U_{ji}\}$ and $\bar{\delta}_{ji} = 1 - \delta_{ji}$. Therefore X_{ji} is the time on study, δ_{ji} is the indicator for observed death, and $\bar{\delta}_{ji}$ is the indicator for censoring.

3.2 MEAN COST

Because patients accrue costs at different rates, the cost accrued to death and the cost accrued to censoring will be positively correlated, even though the times to death and censoring are not. A patient accruing cost at a fast rate will tend to have a large cost at death and a large cost at censoring; and conversely, a patient accruing cost at a slow rate will tend to have a small cost at death and a small cost at censoring. As a result informative censoring is induced on the cost scale, and the use of standard life-table methods for estimating mean cost, similar to those for estimating mean survival time (see Section 3.3.2.1), will lead to positively biased estimators. Because patients with smaller cost are

more likely to be censored, the estimator of mean cost will be positively biased, providing an overestimate of the mean cost. Two methods have been proposed to account for this induced informative censoring on the cost scale, the direct (Lin) method and inverse-probability weighting.

3.2.1 Direct (Lin) method

This method was first proposed by Lin *et al.* (1997), and requires that the duration of interest be divided into K intervals $[a_k, a_{k+1})$, where $0 = a_1 < a_2 < \dots < a_{K+1} = \tau$, in which the cost of each patient is known, at least until the patient is censored. The set of costs incurred during these intervals is often referred to as the cost histories. The interval $[a_k, a_{k+1})$ includes all those values of t such that $a_k \leq t < a_{k+1}$. These intervals, which need not be of equal length, usually coincide with the follow-up clinic visits when health care utilization data are collected.

Let C_{jki} be the cost of the j th patient during the k th interval. If the j th patient is censored during the k th interval, then C_{jki} is unknown for the k th and subsequent intervals. If the j th patient dies during the k th interval, then C_{jki} is zero for all subsequent intervals. The assumption is made here that the cost for the interval in which a patient dies is known. Let $\delta_{jki}^{***} = 1$ if the j th patient enters the k th interval (i.e. $X_{ji} > a_k$) and either dies ($\delta_{ji} = 1$) or is not censored during the interval ($X_{ji} > a_{k+1}$); otherwise $\delta_{jki}^{***} = 0$. The value of δ_{jki}^{***} is 1 for all patients on whom complete cost data for the k th interval is observed. The value of δ_{jki}^{***} is 0 for all patients who either died prior to a_k or were censored prior to a_{k+1} , i.e. those patients for whom cost data was not observed during the k th interval. Algebraically, $\delta_{jki}^{***} = I \{ X_{ji} > a_k \text{ and } (X_{ji} > a_{k+1} \text{ or } \delta_{ji} = 1) \}$. Let \tilde{C}_{jk} be the average of the cost of the patients who have complete cost data observed in the k th interval, i.e. those for whom $\delta_{jki}^{***} = 1$. Thus $\tilde{C}_{jk} = (\sum_{i=1}^{n_j} \delta_{jki}^{***})^{-1} \sum_{i=1}^{n_j} \delta_{jki}^{***} C_{jki}$. Assuming that the censored data in the k th interval are missing completely at random, \tilde{C}_{jk} provides an unbiased estimator of the mean cost during the k th interval of patients on the j th arm who survive beyond a_k . The estimator of v_j , the mean cost over the duration of interest, is simply the sum of the \tilde{C}_{jk} , weighted

30 *Parameter estimation for censored data*

by the probability of surviving to the beginning of the interval. That is,

$$\hat{v}_j = \sum_{k=1}^K S_j(a_k) \tilde{C}_{jk} \quad (3.1)$$

The estimator is unbiased if the censoring occurs only at the boundaries of the intervals. In practice the survival function $S_j(\cdot)$ must be estimated, and so Equation (3.1) becomes

$$\hat{v}_j = \sum_{k=1}^K \hat{S}_j(a_k) \tilde{C}_{jk}. \quad (3.2)$$

The estimator in Equation (3.2) is consistent if the censoring occurs only at the boundaries of the intervals.

The survival function can be estimated using the product-limit method. Let $t_{j1}, t_{j2}, t_{j3}, \dots$ be the unique times at which deaths occur on the j th arm, let $d_{j1}, d_{j2}, d_{j3}, \dots$ be the number of deaths occurring at those times, and let $n_{j1}, n_{j2}, n_{j3}, \dots$ be the number at risk of dying (i.e. not previously dead or censored) at those times. Then the estimator of survival function is given by

$$\hat{S}_j(t) = \prod_{h|t_{jh} < t} \left(1 - \frac{d_{jh}}{n_{jh}}\right) \quad \text{for } t_{j,h-1} \leq t < t_{jh} \quad (3.3)$$

$\hat{S}_j(\cdot)$ is a step function that decreases at those times at which deaths occur, see Figure 3.1. The height of $\hat{S}_j(\cdot)$ at the beginning of the interval provides the weight in the sum of the \tilde{C}_{jk} values. The estimator of the between-arm difference in mean cost is given by $\hat{\Delta}_c = \hat{v}_T - \hat{v}_S$.

The estimator of the variance of \hat{v}_j is given by

$$\hat{V}(\hat{v}_j) = \sum_{i=1}^{n_j} \sum_{k=1}^K \sum_{g=1}^K \hat{\xi}_{jki}^{(c)} \hat{\xi}_{jgi}^{(c)} \quad (3.5)$$

where

$$\begin{aligned} \hat{\xi}_{jki}^{(c)} = & \left(\sum_{i=1}^{n_j} \delta_{jki}^{**} \right)^{-1} (C_{ijk} - \tilde{C}_{jk}) \hat{S}_j(a_k) \delta_{jki}^{**} \\ & - \hat{S}_j(a_k) \tilde{C}_{jk} \left(\frac{I\{X_{ji} \leq a_k\} \delta_{ji}}{R_{ji}} - \sum_{g=1}^{n_j} \frac{I\{X_{jg} \leq \min(X_{ji}, a_k)\} \delta_{jg}}{R_{jg}^2} \right) \end{aligned} \quad (3.6)$$

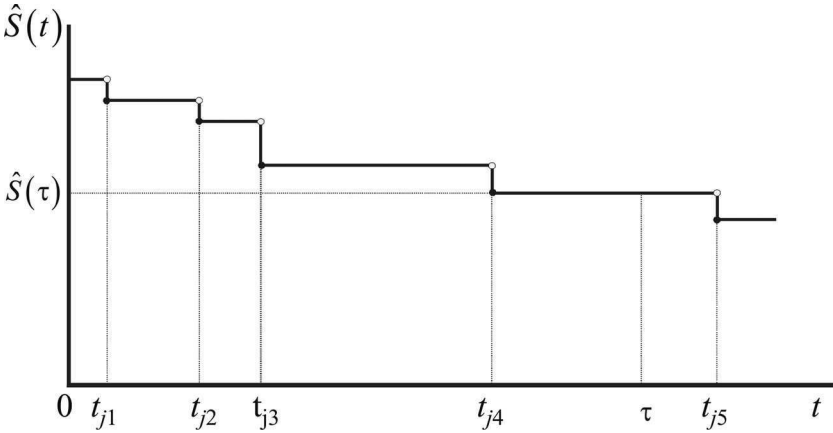


Figure 3.1 Estimated survival curve

and $R_{ji} = \sum_{g=1}^{n_j} I \{X_{jg} \geq X_{ji}\}$, see Lin *et al.* (1997) and Willan and Lin (2001). Therefore the estimator of the variance of $\hat{\Delta}_c$ is given by

$$\hat{V}(\hat{\Delta}_c) = \hat{V}(\hat{v}_T) + \hat{V}(\hat{v}_S) = \sum_{i=1}^{n_T} \sum_{k=1}^K \sum_{g=1}^K \hat{\xi}_{Tki}^{(c)} \hat{\xi}_{Tgi}^{(c)} + \sum_{i=1}^{n_S} \sum_{k=1}^K \sum_{g=1}^K \hat{\xi}_{Ski}^{(c)} \hat{\xi}_{Sgi}^{(c)} \tag{3.7}$$

If cost histories are not available and only total cost for the duration of interest is observed, then there is a single interval $[0, \tau)$, and \hat{v}_j is the sample mean for the non-censored patients and $\hat{V}(\hat{v}_j)$ is the sample variance for the non-censored patients, divided by the number of non-censored patients.

3.2.2 Inverse-probability weighting

The method of inverse-probability weighting (IPW) can be used as an alternative to the direct method for estimating mean cost in the presence of censoring, see Bang and Tsiatis (2000), Lin (2000), Zhao and Tian (2001) and Willan *et al.* (2002). The principle behind IPW can be illustrated by the following simple example.

32 Parameter estimation for censored data

Let the random variables Y_i , $i = 1, 2, \dots, N$ be normally distributed (μ, σ^2) , and let $\delta_i = I\{Y_i \text{ is observed}\}$, i.e. δ_i is 0 if Y_i is censored. Let n be the number of non-censored observations, i.e. $n = \sum_{i=1}^N \delta_i$. If we assume that the censored data is missing at random, then μ is estimated by the average of the non-censored observations, and can be written, rather clumsily, as

$$\hat{\mu} = \left(\sum_{i=1}^N \delta_i \right)^{-1} \sum_{i=1}^N \delta_i Y_i = \left(\sum_{i=1}^N \frac{\delta_i}{n/N} \right)^{-1} \sum_{i=1}^N \frac{\delta_i Y_i}{n/N} \quad (3.8)$$

The last equality holds because the numerator and denominator are both being divided by n/N , albeit unnecessarily. Thus, $\hat{\mu}$ is a weighted average of the Y_i values, where the weight for each observation is δ_i divided by the n/N , i.e. δ_i divided by the probability of not being censored. The factor n/N is not really required in this example because the probability of not being censored is the same for all observations. However, in a follow-up study the probability of an observation not being censored will depend on when that observation is taken.

We will apply IPW to provide an estimator of mean cost in the presence of censoring. Let v_{jk} be the mean cost for a patient on the j th arm during the k th interval. Therefore, $v_j = \sum_{k=1}^K v_{jk}$. We use IPW to provide consistent estimators of the v_{jk} , which can be summed to provide a consistent estimator of v_j . Let $G_j(t) = \Pr(U_{ji} \geq t)$, i.e. $G_j(t)$ is the probability of a patient on the j th treatment not being censored by time t . Let δ_{jki}^* equal to 1 if the j th patient dies (i.e. $\delta_{ji} = 1$) or is not censored before a_{k+1} (i.e. $X_{ji} \geq a_{k+1}$). For a patient who dies, δ_{jki}^* is equal to 1 for all intervals, including those following his or her death. For a patient who is censored, δ_{jki}^* is equal to 1 for intervals prior to the censoring, and equal to 0 for the interval in which the censoring occurs and all subsequent intervals. Algebraically,

$$\delta_{jki}^* = \delta_{ji} + \bar{\delta}_{ji} I \{ X_{ji} \geq a_{k+1} \} \quad (3.9)$$

Further, let $X_{jki}^* = \min(X_{ji}, a_{k+1})$. The value of X_{jki}^* is equal to the upper boundary of the interval for those intervals prior to the patient's death or censoring, and equal to the time on study, i.e. X_{ji} , for the interval in which the death or censoring occurs and for all subsequent

intervals. An unbiased estimator of v_{jk} is given by

$$\hat{v}_{jk} = \left(\sum_{i=1}^{n_j} \frac{\delta_{jki}^*}{G(X_{jki}^*)} \right)^{-1} \sum_{i=1}^{n_j} \frac{\delta_{jki}^* C_{jki}}{G(X_{jki}^*)} \tag{3.10}$$

Equation (3.10) is analogous to Equation (3.9). The value of δ_{jki}^* is 1 if C_{jki} is observed (recall that the cost during an interval following death is set to 0, since it is assumed that patients stop accumulating costs after death). The divisor $G(X_{jki}^*)$ is the probability of not being censored, evaluated at the end of the interval if the patient is on study until the end, or at the time of death if the patient dies. If a patient is censored during the interval, δ_{jki}^* is equal to 0, and the patient will not contribute to the estimated mean for that interval. The role of the divisor is to inflate the observed C_{jki} to account for the censored data. If there is no censoring $G(X_{jki}^*)$ equals 1 and no inflation occurs. If $G(X_{jki}^*)$ is 0.8, for sake of argument, then the value of C_{jki} is inflated by 1.25 (i.e. $1/0.8$) to make up for the censored data when determining the total cost incurred by all patients, both censored and non-censored. The total is then divided by the sum of the weights to derive the mean.

It is usually necessary to use the product-limit to estimate $G_j(\cdot)$. Similarly to estimating $S_j(\cdot)$,

$$\hat{G}_j(t) = \prod_{h|t_{jh}^* < t} \left(1 - \frac{d_{jh}^*}{n_{jh}^*} \right) \quad \text{for } t_{jh-1}^* \leq t < t_{jh}^* \tag{3.11}$$

where the t_{jh}^* 's are the unique censoring times for the j th arm, and n_{jh}^* and d_{jh}^* are, respectively, the number of patients at risk for censoring and the number of patients censored at time t_{jh}^* . Equation (3.10) becomes

$$\hat{v}_{jk} = \left(\sum_{i=1}^{n_j} \frac{\delta_{jki}^*}{\hat{G}(X_{jki}^*)} \right)^{-1} \sum_{i=1}^{n_j} \frac{\delta_{jki}^* C_{jki}}{\hat{G}(X_{jki}^*)} \tag{3.12}$$

34 Parameter estimation for censored data

and is an estimator for the mean cost during interval k . The estimator for ν_j is given by

$$\hat{\nu}_j = \sum_{k=1}^K \hat{\nu}_{jk} \quad (3.13)$$

The estimator given in Equation (3.13) is consistent, regardless of the pattern of censoring. This is in contrast to the estimator provided by the direct method, which is consistent only if the censoring occurs at the boundaries of the intervals.

The estimator of the between-arm difference in mean cost is given by $\hat{\Delta}_c = \hat{\nu}_T - \hat{\nu}_S$, and the variance of $\hat{\Delta}_c$ is given by

$$\hat{V}(\hat{\Delta}_c) = \hat{V}(\hat{\nu}_T) + \hat{V}(\hat{\nu}_S) = \sum_{i=1}^{n_T} \sum_{k=1}^K \sum_{g=1}^K \hat{\xi}_{Tki}^{(c)} \hat{\xi}_{Tgi}^{(c)} + \sum_{i=1}^{n_S} \sum_{k=1}^K \sum_{g=1}^K \hat{\xi}_{Ski}^{(c)} \hat{\xi}_{Sgi}^{(c)} \quad (3.14)$$

where

$$\hat{\xi}_{jki}^{(c)} = \frac{1}{n_j} \left(\frac{\delta_{jki}^* (C_{jki} - \hat{\nu}_{jk})}{\hat{G}(X_{jki}^*)} + \bar{\delta}_{ji} B_{jki} - \sum_{g=1}^{n_j} \frac{\bar{\delta}_{jg} I\{X_{jg} \leq X_{ji}\} B_{jkg}}{R_{jg}} \right) \quad (3.15)$$

and $B_{jki} = \frac{1}{R_{ji}} \sum_{g=1}^{n_j} \frac{I\{X_{jkg} > X_{ji}\} \delta_{jkg}^* (C_{jkg} - \hat{\nu}_{jk})}{\hat{G}(X_{jkg}^*)}$, see Bang and Tsiatis (2000),

Lin (2000), Zhao and Tian (2001) and Willan *et al.* (2002). To ease the burden of notation $\hat{\xi}_{jki}^{(c)}$ has two definitions, one in Equation (3.6) for the direct method and one in Equation (3.15), for IPW. This should not cause confusion since for any particular application only one method would be used.

3.3 EFFECTIVENESS

3.3.1 Probability of surviving

The probability of surviving the duration of interest can be estimated by $\hat{S}_j(\tau)$ and can be read from the estimated survival curve as shown

in Figure 3.1. Formally, it is defined as

$$\hat{\pi}_j = \hat{S}_j(\tau) = \prod_{h|t_{jh} < \tau} \left(1 - \frac{d_{jh}}{n_{jh}}\right) \quad (3.16)$$

The estimator of the variance of $\hat{\pi}_j$ is given by $\hat{V}(\hat{\pi}_j) = \sum_{i=1}^{n_j} (\hat{\xi}_{ji}^{(p)})^2$, where

$$\hat{\xi}_{ji}^{(p)} = -\hat{\pi}_j \left(\frac{I\{X_{ji} \leq \tau\} \delta_{ji}}{R_{ji}} - \sum_{g=1}^{n_j} \frac{I\{X_{jg} \leq \min(\tau, X_{ji})\} \delta_{jg}}{R_{jg}^2} \right) \quad (3.17)$$

see Willan *et al.* (2003). Therefore, when the measure of effectiveness is the probability of surviving the duration of interest, Δ_e is estimated by

$$\hat{\Delta}_e = \hat{\pi}_T - \hat{\pi}_S \quad (3.18)$$

and the variance of $\hat{\Delta}_e$ is given by

$$\hat{V}(\hat{\Delta}_e) = \hat{V}(\hat{\pi}_T) + \hat{V}(\hat{\pi}_S) = \sum_{i=1}^{n_T} (\hat{\xi}_{Ti}^{(p)})^2 + \sum_{i=1}^{n_S} (\hat{\xi}_{Si}^{(p)})^2 \quad (3.19)$$

The estimator of the covariance between $\hat{\Delta}_e$ and $\hat{\Delta}_c$ is given by

$$\begin{aligned} \hat{C}(\hat{\Delta}_e, \hat{\Delta}_c) &= \hat{C}(\hat{\pi}_T, \hat{v}_T) + \hat{C}(\hat{\pi}_S, \hat{v}_S) = \sum_{i=1}^{n_T} \hat{\xi}_{Ti}^{(p)} \sum_{k=1}^K \hat{\xi}_{Tki}^{(c)} \\ &+ \sum_{i=1}^{n_S} \hat{\xi}_{Si}^{(p)} \sum_{k=1}^K \hat{\xi}_{Ski}^{(c)} \end{aligned} \quad (3.20)$$

see Willan *et al.* (2003). The definition of $\hat{\xi}_{jki}^{(c)}$ used in Equation (3.20) will depend on which method was used to estimate the mean cost. The appropriate definition is given in Equation (3.6) for the direct method and in Equation (3.15) for IPW.

3.3.2 Mean survival time

3.3.2.1 Area under the survival curve

The mean survival time over the duration of interest is the area under the survival curve from 0 to τ , i.e. $\int_0^\tau S_j(t) dt$, and can be estimated by the area under the estimated survival curve as shown in Figure 3.2. The estimator is simply the sum of the area of all the shaded rectangles, and, defining $t_{j0} = 0$, can be written algebraically as

$$\hat{\mu}_j = \int_0^\tau \hat{S}_j(u) du = \sum_{h|t_{jh} < \tau} [\hat{S}_j(t_{jh}) (\min(t_{j,h+1}, \tau) - t_{j,h})] \quad (3.21)$$

The estimator of the variance of $\hat{\mu}_j$ is given by

$$\hat{V}(\hat{\mu}_j) = \sum_{i=1}^{n_j} (\hat{\xi}_{ji}^{(m)})^2 \quad (3.22)$$

where

$$\hat{\xi}_{ji}^{(m)} = - \left(\frac{I\{X_{ji} \leq \tau\} \delta_{ji} A_j(X_{ji})}{R_{ji}} - \sum_{g=1}^{n_j} \frac{I\{x_{jg} \leq \min(X_{ji}, \tau)\} \delta_{jg} A_j(X_{jg})}{R_{jg}^2} \right) \quad (3.23)$$

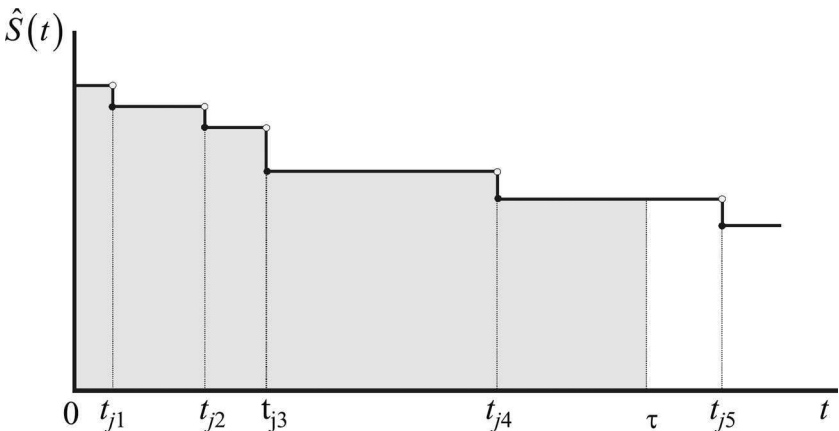


Figure 3.2 Shaded area equal to estimated mean survival time from 0 to τ

and

$$\begin{aligned}
 A_j(t) &= \int_t^\tau \hat{S}_j(u) du \\
 &= \int_0^\tau \hat{S}_j(u) du - \int_0^t \hat{S}_j(u) du \\
 &= \hat{\mu}_j - \sum_{h|t_{jh} < t} [\hat{S}_j(t_{jh}) (\min(t_{j,h+1}, t) - t_{j,h})]
 \end{aligned}$$

see Willan *et al.* (2002). Therefore, when the measure of effectiveness is mean survival time, Δ_e is estimated by

$$\hat{\Delta}_e = \hat{\mu}_T - \hat{\mu}_S \tag{3.24}$$

and the variance of $\hat{\Delta}_e$ is given by

$$\hat{V}(\hat{\Delta}_e) = \hat{V}(\hat{\mu}_T) + \hat{V}(\hat{\mu}_S) = \sum_{i=1}^{n_T} \left(\hat{\xi}_{Ti}^{(m)} \right)^2 + \sum_{i=1}^{n_S} \left(\hat{\xi}_{Si}^{(m)} \right)^2 \tag{3.25}$$

The estimator of the covariance between $\hat{\Delta}_e$ and $\hat{\Delta}_c$ is given by

$$\begin{aligned}
 \hat{C}(\hat{\Delta}_e, \hat{\Delta}_c) &= \hat{C}(\hat{\mu}_T, \hat{\nu}_T) + \hat{C}(\hat{\mu}_S, \hat{\nu}_S) + \sum_{i=1}^{n_T} \left(\hat{\xi}_{Ti}^{(m)} \sum_{k=1}^K \hat{\xi}_{Tki}^{(c)} \right) \\
 &\quad + \sum_{i=1}^{n_S} \left(\hat{\xi}_{Si}^{(m)} \sum_{k=1}^K \hat{\xi}_{Ski}^{(c)} \right)
 \end{aligned} \tag{3.26}$$

see Willan *et al.* (2002). The definition of $\hat{\xi}_{jki}^{(c)}$ used in Equation (3.26) will depend on which method was used to estimate the mean cost. The appropriate definition is given in Equation (3.6) for the direct method and in Equation (3.15) for IPW.

3.3.2.2 Inverse-probability weighting

The method of IPW can be used to estimate mean survival time. Let $X_{ji}^* = \min(D_{ji}, U_{ji}, \tau)$, i.e. X_{ji}^* is the time to death, time to censoring or the duration of interest, whichever is the shortest. Let $\delta_{ji}^* = \delta_{ji} + \bar{\delta}_{ji} I \{U_{ji} \geq \tau\}$, i.e. δ_{ji}^* equals 1 if, and only if, the patient dies

38 *Parameter estimation for censored data*

or if the patient is censored after being followed for the duration of interest. The IPW estimator for mean survival time over the duration of interest is given by

$$\hat{\mu}_j = \left(\sum_{i=1}^{n_j} \frac{\delta_{ji}^*}{\hat{G}(X_{ji}^*)} \right)^{-1} \sum_{i=1}^{n_j} \frac{\delta_{ji}^* X_{ji}^*}{\hat{G}(X_{ji}^*)}. \quad (3.27)$$

This estimator is simply a weighted average of the times on study i.e. X_{ji}^*) for the patients who are not censored prior to τ . The weight is 1 dividend by the probability of not being censored at the time of death for those who died prior to τ and 1 over the probability of not being censored at τ for those who survived, uncensored, to τ .

The variance of $\hat{\mu}_j$ is given by

$$\hat{V}(\hat{\mu}_j) = \sum_{i=1}^{n_j} \left(\hat{\xi}_{ji}^{(m)} \right)^2 \quad (3.28)$$

where

$$\hat{\xi}_{ji}^{(m)} = \frac{1}{n_j} \left[\frac{\delta_{ji}^* (X_{ji}^* - \hat{\mu}_j)}{\hat{G}(X_{ji}^*)} + \bar{\delta}_{ji} F_{ji} - \sum_{g=1}^{n_j} \frac{\bar{\delta}_{jg} I \{ X_{jg} \leq X_{ji} \}}{R_{jg}} F_{jg} \right] \quad (3.29)$$

and

$$F_{ji} = \frac{1}{R_{ji}} \sum_{g=1}^{n_j} \frac{I \{ X_{jg} > X_{ji} \} \delta_{jg}^* (X_{jg}^* - \hat{\mu}_j)}{\hat{G}(X_{jg}^*)}.$$

The estimator for the covariance between $\hat{\Delta}_e$ and $\hat{\Delta}_c$ is given by

$$\begin{aligned} \hat{C}(\hat{\Delta}_e, \hat{\Delta}_c) &= \hat{C}(\hat{\mu}_T, \hat{\nu}_T) + \hat{C}(\hat{\mu}_S, \hat{\nu}_S) = \sum_{i=1}^{n_T} \left(\hat{\xi}_{Ti}^{(m)} \sum_{k=1}^K \hat{\xi}_{Tki}^{(c)} \right) \\ &\quad + \sum_{i=1}^{n_S} \left(\hat{\xi}_{Si}^{(m)} \sum_{k=1}^K \hat{\xi}_{Ski}^{(c)} \right) \end{aligned} \quad (3.30)$$

where $\hat{\xi}_{jki}^{(c)}$ is defined in Equation (3.15).

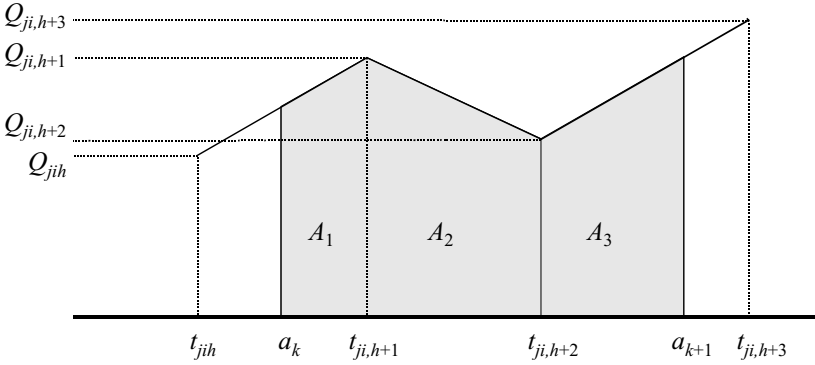


Figure 3.3 QALYs for K th interval

3.3.3 Mean quality-adjusted survival time

Since patients accrue quality-adjusted survival (QAS) at different rates, much the same as they accrue costs at different rates, the application of life-table methods to estimate mean QAS time produces biased results. So, as with cost, we can utilize either the direct method or inverse-probability weighting to provide estimators for mean QAS time. Recall from Section 2.3.3 that there are m_{ji} QoL measurements on the j th patient, taken at times: $0 \leq t_{ji1} < t_{ji2} < \dots < t_{jim_{ji}} \leq \tau$, with corresponding values $Q_{ji1}, Q_{ji2}, \dots, Q_{jim_{ji}}$. The observed QALY on the j th patient during the k th interval, denoted by q_{jki} , is simply the area under the curve of the plot of the Q_{ij} values by the t_{ij} values between a_k and a_{k+1} , as shown in Figure 3.3. Therefore $q_{jki} = \int_{a_k}^{a_{k+1}} Q_{ij}(t) dt$, where $Q(t)$ is defined as in Equation (2.9.) Put more simply, q_{jki} is the sum of the shaded rectangles shown in Figure 3.3, i.e.

$$\begin{aligned}
 q_{jki} &= A_1 + A_2 + A_3 \\
 &= \left(Q_{ji,h+1} + \frac{(Q_{ji,h+1} - Q_{jih})(a_k - t_{ji,h+1})}{2(t_{ji,h+1} - t_{jih})} \right) \times (t_{ji,h+1} - a_k) \\
 &\quad + (Q_{ji,h+1} + Q_{ji,h+2})(t_{ji,h+2} - t_{ji,h+1})/2 \\
 &\quad + \left(Q_{ji,h+2} + \frac{(Q_{ji,h+3} - Q_{jih+2})(a_{k+1} - t_{ji,h+2})}{2(t_{ji,h+3} - t_{ji,h+2})} \right) \\
 &\quad \times (a_{k+1} - t_{ji,h+2})
 \end{aligned}$$

40 Parameter estimation for censored data

To apply the direct method or IPW, one need only replace C_{jki} by q_{jki} in the appropriate equations in Sections 3.2.1 and 3.2.2, as illustrated in the following sections.

3.3.3.1 The direct method

The estimator of the between-treatment difference of mean quality-adjusted life using the direct method is given by

$$\hat{\Delta}_e = \hat{\varphi}_T - \hat{\varphi}_S = \sum_{k=1}^K \hat{S}_T(a_k) \tilde{q}_{Tk} - \sum_{k=1}^K \hat{S}_S(a_k) \tilde{q}_{Sk} \quad (3.31)$$

where $\tilde{q}_{jk} = \left(\sum_{i=1}^{n_j} \delta_{jki}^{**} \right)^{-1} \sum_{i=1}^{n_j} \delta_{jki}^{**} q_{jki}$.

The estimator of the variance of $\hat{\Delta}_e$ is given by, see Willan *et al.* (2003).

$$\hat{V}(\hat{\Delta}_e) = \hat{V}(\hat{\varphi}_T) + \hat{V}(\hat{\varphi}_S) = \sum_{i=1}^{n_T} \sum_{k=1}^K \sum_{g=1}^K \hat{\xi}_{Tki}^{(q)} \hat{\xi}_{Tgi}^{(q)} + \sum_{i=1}^{n_S} \sum_{k=1}^K \sum_{g=1}^K \hat{\xi}_{Ski}^{(q)} \hat{\xi}_{Sgi}^{(q)} \quad (3.32)$$

where

$$\begin{aligned} \hat{\xi}_{jki}^{(q)} &= \left(\sum_{i=1}^{n_j} \delta_{jki}^{**} \right)^{-1} (q_{ijk} - \tilde{q}_{jk}) \hat{S}_j(a_k) \delta_{jki}^{**} \\ &\quad - \hat{S}_j(a_k) \tilde{q}_{jk} \left(\frac{I\{X_{ji} \leq a_k\} \delta_{ji}}{R_{ji}} - \sum_{g=1}^{n_j} \frac{I\{X_{jg} \leq \min(X_{ji}, a_k)\} \delta_{jg}}{R_{jg}^2} \right) \end{aligned} \quad (3.33)$$

The estimator for the covariance between $\hat{\Delta}_e$ and $\hat{\Delta}_c$ is given by

$$\begin{aligned} \hat{C}(\hat{\Delta}_e, \hat{\Delta}_c) &= \hat{C}(\hat{\varphi}_T, \hat{v}_T) + \hat{C}(\hat{\varphi}_S, \hat{v}_S) = \sum_{i=1}^{n_T} \sum_{k=1}^K \sum_{g=1}^K \hat{\xi}_{Tki}^{(q)} \hat{\xi}_{Tgi}^{(c)} \\ &\quad + \sum_{i=1}^{n_S} \sum_{k=1}^K \sum_{g=1}^K \hat{\xi}_{Ski}^{(q)} \hat{\xi}_{Sgi}^{(c)} \end{aligned} \quad (3.34)$$

where $\hat{\xi}_{jki}^{(c)}$ is defined in Equation (3.6.) The reader is referred to Willan *et al.* (2003) for more details.

3.3.3.2 Inverse-probability weighting

The estimator of the between-treatment difference of mean quality-adjusted life using IPW is given by

$$\hat{\Delta}_e = \hat{\varphi}_T - \hat{\varphi}_S = \sum_{k=1}^K \hat{\varphi}_{Tk} - \sum_{k=1}^K \hat{\varphi}_{Sk} \tag{3.35}$$

where

$$\hat{\varphi}_{jk} = \left(\sum_{i=1}^{n_j} \frac{\delta_{jki}^*}{G(X_{jki}^*)} \right)^{-1} \sum_{i=1}^{n_j} \frac{\delta_{jki}^* q_{jki}}{G(X_{jki}^*)} \tag{3.36}$$

The variance of $\hat{\Delta}_e$ is given by

$$\hat{V}(\hat{\Delta}_e) = \hat{V}(\hat{\varphi}_T) + \hat{V}(\hat{\varphi}_S) = \sum_{i=1}^{n_T} \sum_{k=1}^K \sum_{g=1}^K \hat{\xi}_{Tki}^{(q)} \hat{\xi}_{Tgi}^{(q)} + \sum_{i=1}^{n_S} \sum_{k=1}^K \sum_{g=1}^K \hat{\xi}_{Ski}^{(q)} \hat{\xi}_{Sgi}^{(q)} \tag{3.37}$$

where

$$\hat{\xi}_{jki}^{(q)} = \frac{1}{n_j} \left(\frac{\delta_{jki}^* (q_{jki} - \hat{\varphi}_{jk})}{\hat{G}(X_{jki}^*)} + \bar{\delta}_{ji} D_{jki} - \sum_{g=1}^{n_j} \frac{\bar{\delta}_{jg} I\{X_{jg} \leq X_{ji}\} D_{jkg}}{R_{jg}} \right) \tag{3.38}$$

and $D_{jki} = \frac{1}{R_{ji}} \sum_{g=1}^{n_j} \frac{I\{X_{jkg}^* > X_{ji}\} \delta_{jkg}^* (q_{jkg} - \hat{\varphi}_{jk})}{\hat{G}(X_{jkg}^*)}$. Again, to ease

the burden of notation, $\hat{\xi}_{jki}^{(q)}$ has two definitions, one in Equation (3.32) for the direct method and one in Equation (3.37), for the IPW method. The estimator for the covariance between $\hat{\Delta}_e$ and $\hat{\Delta}_c$ is given by

$$\begin{aligned} \hat{C}(\hat{\Delta}_e, \hat{\Delta}_c) &= \hat{C}(\hat{\varphi}_T, \hat{v}_T) + \hat{C}(\hat{\varphi}_S, \hat{v}_S) = \sum_{i=1}^{n_T} \sum_{k=1}^K \sum_{g=1}^K \hat{\xi}_{Tki}^{(q)} \hat{\xi}_{Tgi}^{(c)} \\ &+ \sum_{i=1}^{n_S} \sum_{k=1}^K \sum_{g=1}^K \hat{\xi}_{Ski}^{(q)} \hat{\xi}_{Sgi}^{(c)} \end{aligned} \tag{3.39}$$

where $\hat{\xi}_{jki}^{(c)}$ is defined in Equation (3.15), see Willan *et al.* (2002).

3.4 SUMMARY

Methods for estimating the five parameters required for a cost-effectiveness comparison of two groups using censored data are given in Sections 3.2 and 3.3. Life-table methods are used for the probability of surviving and the mean survival time. Inverse-probability weighting is used for mean cost, mean survival time and mean quality-adjusted survival time. The direct (Lin) method is used for mean cost and mean quality-adjusted survival time. In general, inverse-probability weighting is preferred to the direct method because it provides consistent estimators regardless of the censoring pattern. Furthermore, inverse-probability weighting can be generalized to include covariate adjustment, as discussed in Chapter 7. A discussion on the use of these parameters in cost-effectiveness analysis is provided in the following chapter. The use of incremental cost-effectiveness ratios, incremental net benefit and Bayesian methods are discussed.

4

Cost-effectiveness Analysis

4.1 INTRODUCTION

Statistical inference in cost-effectiveness analysis, using the five parameter estimates calculated from the methods given in Chapters 2 and 3, are presented in this chapter. Traditionally, cost-effectiveness analysis has summarized the value for money of the treatment under evaluation using the incremental cost-effectiveness ratio (ICER). Statistical inference using the ICER is covered in Section 4.2, focusing on the many difficulties and complexities that arise when using ratio statistics. The concerns regarding inference on the ICER have led investigators to the use incremental net benefit (INB) as an alternative. Statistical inference based on INB is discussed in Section 4.3. The equivalence of the appropriate inference using either statistic is emphasized through the equivalence of the regions on the cost-effectiveness plane. The presentation of results using the cost-effectiveness acceptability curve (CEAC) is presented in Section 4.4 and the use of bootstrap methods is discussed in Section 4.5. The Bayesian approach, under which the CEAC has its most natural interpretation, is presented in Section 4.6. Lastly, in Section 4.7 consideration is given to the fact that the decision threshold may differ, depending on whether a policy decision involves implementing a new treatment or withdrawing an existing treatment.

4.2 INCREMENTAL COST-EFFECTIVENESS RATIO

The ICER is estimated by $\hat{R} = \hat{\Delta}_c / \hat{\Delta}_e = \text{slope of } \hat{\Delta}$, where $\hat{\Delta} = (\hat{\Delta}_e, \hat{\Delta}_c)^T$. The estimator \hat{R} is consistent, although not unbiased. In a standard test of hypothesis approach one looks for evidence to reject the following null hypothesis:

$$H : \frac{\Delta_c}{\Delta_e} \geq \lambda \quad \text{if } \Delta_e > 0; \quad \text{or} \quad \frac{\Delta_c}{\Delta_e} \leq \lambda \quad \text{if } \Delta_e < 0 \quad (4.1)$$

in favor of

$$A : \frac{\Delta_c}{\Delta_e} < \lambda \quad \text{if } \Delta_e > 0; \quad \text{or} \quad \frac{\Delta_c}{\Delta_e} > \lambda \quad \text{if } \Delta_e < 0. \quad (4.2)$$

These rather complex hypotheses result from the ICER being a ratio and this is one of the arguments in favor of using an analysis based on INB (see Section 4.3). Nevertheless, the reason for this formulation of the cost-effectiveness hypothesis is clear from the cost effectiveness plane (introduced in Figure 1.1). Figure 4.1 shows the cost-effectiveness plane including the threshold line with slope λ and passing through the origin. The shaded area below and to the right of this line is the acceptance region (corresponding to the alternative hypothesis of Equation 4.2) and the unshaded area above and to the left of the threshold line is the rejection region (corresponding to the null hypothesis of Equation 4.1). The corresponding ICERs for points a_1 and a_2 are both less than the threshold, but a_1 falls in the rejection region and a_2 in the acceptance region. Similarly, the corresponding ICERs for b_1 and b_2 are both greater than the threshold, but with opposing decision implications. In order to understand this apparent anomaly it is necessary to consider the denominator of the ratio. Where Δ_e is positive (the NE and SE quadrants in Figure 4.1) Treatment is cost-effective if, and only if, the ICER $< \lambda$. However, where Δ_e is negative (the NW and SW quadrants) Treatment is cost-effective if, and only if the ICER $> \lambda$.

Inference for the ICER is limited to constructing its confidence interval $(R_{2\alpha}^{LL}, R_{2\alpha}^{UL})$ where the superscript refers to the lower and upper

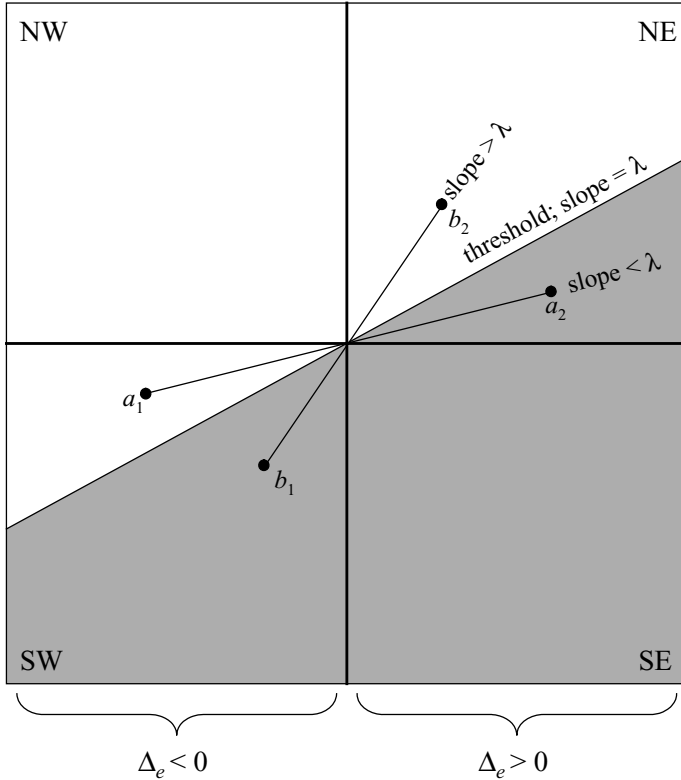


Figure 4.1 The cost-effectiveness plane

limit respectively and 2α is the level that provides a $(1 - 2\alpha)100\%$ confidence interval. The conclusion that there is evidence that Treatment is cost-effective is made if and only if either: (i) $\hat{\Delta}_e > 0$ and $R_{2\alpha}^{UL} < \lambda$; or (ii) $\hat{\Delta}_e < 0$ and $R_{2\alpha}^{LL} > \lambda$.

However, confidence interval construction for the ICER is not straightforward. In particular, Δ_e can be 0, which means that R is undefined and therefore not properly estimable. Some authors have proposed using a Taylor series expansion to estimate the variance of \hat{R} , see O'Brien *et al.* (1994), Chaudhary and Stearns (1996), Briggs

46 Cost-effectiveness analysis

and Fenn (1998) and Briggs *et al.* (1999). Using the Taylor series expansion the estimator for the variance of \hat{R} is given by

$$\hat{V}(\hat{R}) \simeq \hat{R}^2 [\hat{V}(\hat{\Delta}_e)/\hat{\Delta}_e^2 + \hat{V}(\hat{\Delta}_c)/\hat{\Delta}_c^2 - 2\hat{C}(\hat{\Delta}_e, \hat{\Delta}_c)/\hat{\Delta}_e\hat{\Delta}_c]$$

and $100(1 - 2\alpha)\%$ confidence interval can be constructed by $\hat{R} \pm z_{1-\alpha}\sqrt{\hat{V}(\hat{R})}$, where $z_{1-\alpha}$ is the $100(1 - \alpha)$ th percentile of a standard normal random variable. However, the accuracy of this approximation cannot be relied on when the sample sizes are small, or when $\hat{\Delta}_e/\sqrt{\hat{V}(\hat{\Delta}_e)}$ or $\hat{\Delta}_c/\sqrt{\hat{V}(\hat{\Delta}_c)}$ is less than 0.1.

An application of Fieller's theorem provides an alternative to the use of the Taylor series expansion for calculating confidence intervals for R , see Willan and O'Brien (1996), Chaudhary and Sterns (1996), Briggs and Fenn (1998), Briggs *et al.* (1999), Laska EM *et al.* (1997). The best way to illustrate the derivation of the Fieller confidence interval is to recognize that if $\hat{\Delta}_e$ and $\hat{\Delta}_c$ are unbiased, then

$$\begin{aligned} E(\hat{\Delta}_c - R\hat{\Delta}_e) &= E(\hat{\Delta}_c) - \frac{\Delta_c}{\Delta_e} E(\hat{\Delta}_e) \\ &= \Delta_c - \frac{\Delta_c}{\Delta_e} \Delta_e \\ &= 0 \end{aligned}$$

and, if $\hat{\Delta}_e$ and $\hat{\Delta}_c$ are normally distributed, then so is $\hat{\Delta}_c - R\hat{\Delta}_e$, and therefore

$$\begin{aligned} &(\hat{\Delta}_c - R\hat{\Delta}_e) / \sqrt{V(\hat{\Delta}_c - R\hat{\Delta}_e)} \\ &= (\hat{\Delta}_c - R\hat{\Delta}_e) / \sqrt{V(\hat{\Delta}_c) + R^2V(\hat{\Delta}_e) - 2R \times C(\hat{\Delta}_c, \hat{\Delta}_e)} \end{aligned}$$

is distributed as a standard normal random variable. Consequently,

$$|\hat{\Delta}_c - R\hat{\Delta}_e| / \sqrt{V(\hat{\Delta}_c) + R^2V(\hat{\Delta}_e) - 2R \times C(\hat{\Delta}_c, \hat{\Delta}_e)} \leq z_{1-\alpha} \quad (4.3)$$

with probability $(1 - 2\alpha)$. By substituting in the estimates for the variances and the covariance and squaring both sides, Equation (4.3)

becomes a quadratic equation in R . Solving for R provides the $100(1 - 2\alpha)\%$ confidence interval for ICER, the limits of which are given by

$$\hat{R} \left\{ \left(1 - z_{1-\alpha}^2 c \pm z_{1-\alpha} \sqrt{a + b - 2c - z_{1-\alpha}^2 (ab - c^2)} \right) / (1 - z_{1-\alpha}^2 a) \right\} \quad (4.4)$$

where $a = \hat{V}(\hat{\Delta}_e) / \hat{\Delta}_e^2$, $b = \hat{V}(\hat{\Delta}_c) / \hat{\Delta}_c^2$ and $c = \hat{C}(\hat{\Delta}_e, \hat{\Delta}_c) / (\hat{\Delta}_e \hat{\Delta}_c)$.

The Fieller confidence interval is a 'bow tie' defined by the two lines through the origin and includes the wedge in which the observed point lies and the wedge opposite, see Figure 4.2 for illustration. This is

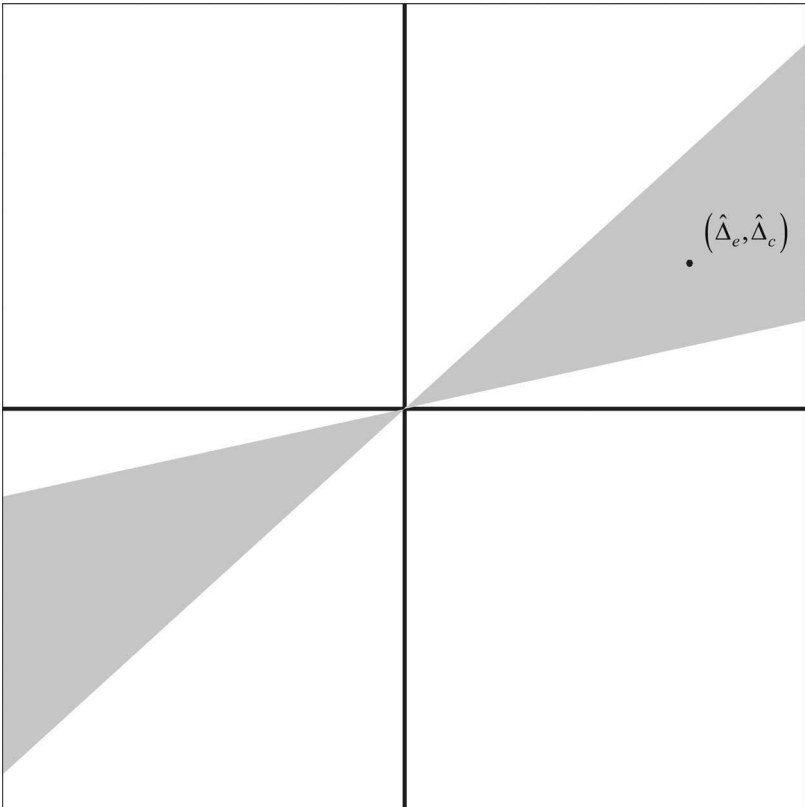


Figure 4.2 Bow tie ICER confidence region

because the slope of any point in either wedge satisfies Equation (4.3). This is particularly vexing when the observed point is in one of the non-trade-off quadrants (i.e. S.E. or N.W.) because then the confidence interval will include parts from both the win–win and lose–lose regions of the cost-effectiveness plane. In practice, when the observed point is sufficiently far away from the origin, and the confidence interval is narrow, very little probability will lie in the opposite wedge.

The confidence interval can include part of the vertical axis, in which case the corresponding limit is considered to be arbitrarily large if it is the upper limit and arbitrarily small if it is the lower limit. If $a + b - 2c - z_{1-\alpha}^2(ab - c^2) = 0$, the upper and lower limits are the same and are equal to $\hat{R}(1 - z_{1-\alpha}^2c)/(1 - z_{1-\alpha}^2a)$, in which case, the confidence interval defines the entire CE plane. If $a + b - 2c - z_{1-\alpha}^2(ab - c^2) < 0$, no solution exists, and again the confidence interval defines the entire CE plane. Such are the problems when dealing with ratio statistics.

The extent to which the Taylor series expansion is an approximation is seen by rewriting it as $\hat{R}(1 \pm z_{1-\alpha}\sqrt{a + b - 2c})$ and contrasting it with Expression (4.4). The divisor $(1 - z_{1-\alpha}^2a)$ tends to dominate the contrast, and although it is not obvious from routine inspection of the expressions, the Fieller solution will always provide a wider confidence interval, see Willan and O'Brien (1996). In general the Fieller solution is preferred since it is not an approximation and relies only on the estimators $\hat{\Delta}_e$ and $\hat{\Delta}_c$ being unbiased and normally distributed. As was discussed in Chapter 2, cost data in particular may often exhibit substantial skewness such that there may be some concern over the legitimacy of the normal assumption for $\hat{\Delta}_c$. However, Willan and O'Brien (1996) provide simulation evidence that, even with sample sizes as small as 100 and cost data that are log-normal to the base 10, the distribution of $\hat{\Delta}_c$ is normally distributed. Other studies have come to the same conclusion, see Thompson and Barber (2000), Lumley *et al.* (2002), Willan *et al.* (2004), and Briggs *et al.* 2005. In general, these simulation studies support Cochran's rule of thumb that n must be $25\eta^2$ for situations where the 'principal deviation from normality consists of marked positive skewness' (Cochran 1977) and where η is the skewness coefficient in the sample. The guideline was devised such that a 95% confidence interval will have an error probability no greater than 6%.

Table 4.1 Three treatment comparison falling in the win–win quadrant

	Δ_e	Δ_c	ICER
<i>T1</i>	0.1	–2000	–20 000
<i>T2</i>	0.2	–4000	–20 000
<i>T3</i>	0.2	–2000	–10 000

The ICER is not properly ordered in the SE and NW quadrants, i.e. in the win–win and lose–lose quadrants, respectively. To see this, consider comparing *S* to three different treatments (*T1*, *T2* and *T3*) for which the values of Δ_e and Δ_c are shown in Table 4.1. The measure of effectiveness is the probability of survival. *T1* increases the probability of survival by 0.1 and reduces the cost by 2000, for an ICER of –20 000. *T2* increases the probability of survival by 0.2 and reduces the cost by 4000, and is clearly superior to *T1*, yet it has the same ICER. Therefore, two treatments compared to the same standard can have the same ICER, yet one may be clearly superior to the other. It gets worse. *T3* increases the probability of survival by 0.2 and reduces the cost by 2000, and is superior to *T1* because, although it reduces cost by the same amount, its increase in the probability of survival is double. However, the ICER for *T3* is larger than for *T1*. It is not possible to interpret the confidence interval for a parameter over a range in which it is not properly ordered, and this concern is one more argument in favour of an INB approach.

4.3 INCREMENTAL NET BENEFIT

The concerns encountered when making statistical inference on the ICER can be addressed by using an INB approach. The estimator of INB is given by $\hat{b}_\lambda = \hat{\Delta}_e \lambda - \hat{\Delta}_c$, and is unbiased if $\hat{\Delta}_e$ and $\hat{\Delta}_c$ are. The variance of \hat{b}_λ is estimated by $v_\lambda \equiv \hat{V}(\hat{b}_\lambda) = \lambda^2 \hat{V}(\hat{\Delta}_e) + \hat{V}(\hat{\Delta}_c) - 2\lambda \hat{C}(\hat{\Delta}_e, \hat{\Delta}_c)$. The null hypothesis $H: \Delta_e \lambda - \Delta_c \leq 0$ can be rejected in favour of the alternative hypothesis $A: \Delta_e \lambda - \Delta_c > 0$ at the level α if the test statistic $\hat{b}_\lambda / \sqrt{v_\lambda}$ exceeds $z_{1-\alpha}$. Additionally, the $100(1 - 2\alpha)\%$

50 *Cost-effectiveness analysis*

confidence limits for INB are given by

$$\hat{b}_\lambda \pm z_{1-\alpha} \sqrt{v_\lambda} \quad (4.6)$$

and H can be rejected in favor of A, equivalently, if the lower limit exceeds 0. Note that the hypotheses H and A are, respectively, exactly the same as the hypotheses defined in Equations (4.1) and (4.2). In terms of Figure 4.1, note that the acceptance and rejection regions relating to the INB formulation exactly correspond to those under the ICER interpretation with points above and to the left of the threshold line associated with $\text{INB} < 0$ (i.e. not cost-effective) and those below and to the right having $\text{INB} > 0$ (i.e. cost-effective). It should be straightforward to appreciate the ease with which the INB approach can test H in comparison with the ICER approach.

Another advantage to the INB approach is that a sensitivity analysis, varying the value of λ , can be performed. Therefore, by estimating \hat{b}_λ and calculating the corresponding confidence limits for a large number of values of λ , and plotting them as a function of λ , as seen in Figure 4.3, readers can apply the value of λ they feel most appropriate, reading from the vertical axis the estimate of INB and the confidence limits. Furthermore, \hat{b}_λ crosses the horizontal axis at \hat{R} (i.e. $\hat{b}_{\hat{R}} = 0$), and the confidence limits for b_λ crosses the horizontal axis at the Fieller limit for R , allowing readers to make inference on the ICER. This can be seen by setting Expression (4.6) to 0, which becomes the same equation in λ as the Inequality (4.3) is in R . The correspondence between the Fieller and net-benefit solution has been noted in the literature (Heitjan, 2000; Zethraeus and Lothgren, 2000), but should be unsurprising given the similar configurations of $\hat{\Delta}_e$ and $\hat{\Delta}_c$ and the assumption of joint-normality employed in both approaches. In Figure 4.3 the lower limit for b_λ fails to cross the horizontal axis then the upper limit for the ICER does not exist and is considered to be arbitrarily large. If the upper limit for b_λ fails to cross the horizontal axis then the lower limit for the ICER does not exist and is considered to be arbitrarily small. If neither limit for b_λ crosses the horizontal axis then neither limit for the ICER exists because, as discussed in Section 4.2, $a + b - 2c - z_{1-\alpha}^2(ab - c^2) = 0$. Finally, \hat{b}_λ crosses the vertical axis at $-\hat{\Delta}_c$, and the lower and upper limits for b_λ cross the vertical axis at the lower and upper limit for $-\hat{\Delta}_c$ allowing inference on Δ_c .

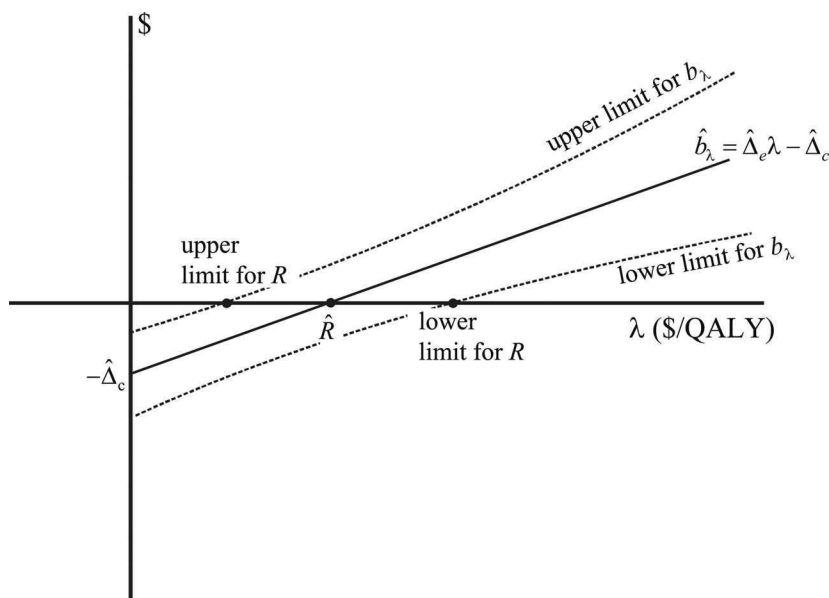


Figure 4.3 Incremental net benefit as a function of WTP

It should also be straightforward to see that the INB could be bootstrapped as an alternative to assuming a normal distribution. While the necessity for bootstrapping the INB, given that it does not suffer from the problems associated with ratio statistics is less clear, it is worth noting that a bootstrap analysis of net benefit and an appropriate bootstrap analysis of the ICER should exactly correspond in terms of inference regarding cost-effectiveness, just as the parametric INB and parametric Fieller method correspond.

4.4 THE COST-EFFECTIVENESS ACCEPTABILITY CURVE

For the methods of inference discussed in Sections 4.2 and 4.3 the level of significance is fixed at some level α . This may not always be appropriate, and a Bayesian approach leading to the cost-effectiveness

52 **Cost-effectiveness analysis**

acceptability curve (CEAC) can provide a more flexible and a natural means for expressing the uncertainty of parameter estimators, see van Hout *et al.* (1994), Briggs and Fenn (1998) and Briggs (1999). The CEAC is a plot of the probability that Treatment is cost-effective as a function of willingness-to-pay. In Bayesian terms, it is the probability that the INB, for a particular value of the willingness-to-pay, is greater than zero, (i.e. $\Pr(\Delta_e \lambda - \Delta_c > 0)$), and is the probability that the point Δ falls below the line on the CE plane that passes through the origin with slope λ (the threshold), see Figure 4.4 for illustration. The concentric ellipses represent contours of equal probability for the

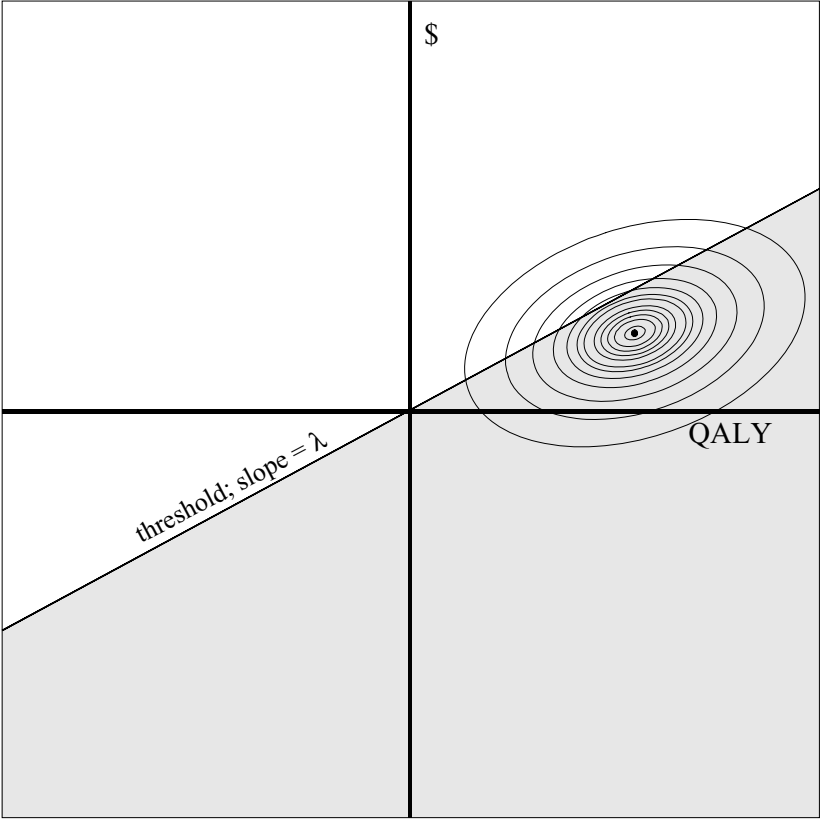


Figure 4.4 The CE plane; shaded area = $\Pr(b_\lambda > 0)$

joint posterior density function of Δ . The CEAC is the probability that the point Δ lies below the threshold as a function of λ . The CEAC = $\Pr(\Delta_c < 0)$ for, $\lambda = 0$ approaches $\Pr(\Delta_e < 0)$ as λ approaches $-\infty$ and approaches $\Pr(\Delta_e > 0)$ as λ approaches $+\infty$. The probabilities are determined from the posterior distribution of Δ , which is, if an uninformative prior is specified, the distribution of $\hat{\Delta}$. If normality is assumed, then the posterior distribution of Δ will be bivariate normal with mean $\hat{\Delta}$ variance-covariance matrix

$$\hat{V}(\hat{\Delta}) = \begin{pmatrix} \hat{V}(\hat{\Delta}_e) & \hat{C}(\hat{\Delta}_e, \hat{\Delta}_c) \\ \hat{C}(\hat{\Delta}_e, \hat{\Delta}_c) & \hat{V}(\hat{\Delta}_c) \end{pmatrix}$$

If we let $f_{\hat{\Delta}}(\cdot, \cdot)$ be the corresponding density function, then the CEAC for a particular value of λ is given by

$$\mathcal{A}(\lambda) \equiv \int_{-\infty}^{\infty} \int_{-\infty}^{\lambda e} f_{\hat{\Delta}}(e, c) dc de = \Phi(\hat{b}(\lambda)/\sqrt{v_{\lambda}}) \tag{4.7}$$

where $\Phi(\cdot)$ is the cumulative distribution function for a standard normal random variable. $\mathcal{A}(\lambda)$ has a frequentist interpretation as 1 minus the p -value for the test of hypothesis H: $\Delta_e \lambda - \Delta_c \leq 0$ versus A: $\Delta_e \lambda - \Delta_c > 0$. Clearly, $\mathcal{A}(\hat{R}) = 0.5$. In addition, the CEAC passes through Fieller limits for the ICER at α and $1 - \alpha$. If the CEAC does not pass through α or $1 - \alpha$, then the corresponding limit for the ICER does not exist and is considered to be arbitrarily large or small because, as described in Section 4.2, the interval includes the vertical axis. If the CEAC passes through neither α nor $1 - \alpha$, then neither limit for the ICER exists because, as discussed in Section 4.2, $a + b - 2c - z_{1-\alpha}^2(ab - c^2) < 0$. As before, the confidence limits for the ICER derived from the CEAC define a ‘bow tie’ region as illustrated in Figure 4.2.

The CEAC has three important strengths. It is a measure of both magnitude and uncertainty of cost-effectiveness. It expresses the uncertainty in terms of a probability statement about the cost-effectiveness of Treatment, which is often considered more natural to policymakers. Thirdly, it allows policymakers the ability to use different strengths of evidence depending on what Treatment is. If adopting Treatment is expected to have a substantial effect on the health care

54 **Cost-effectiveness analysis**

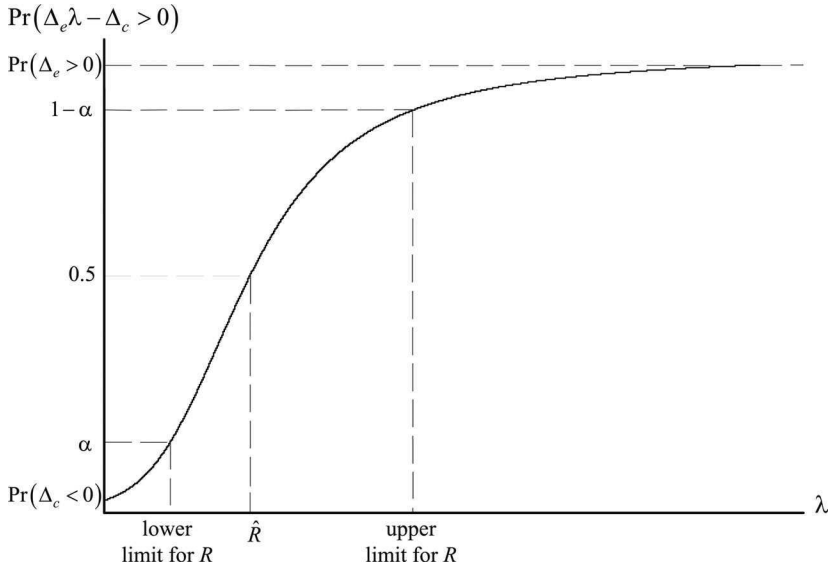


Figure 4.5 The Cost-effectiveness acceptability curve

system, because either Δ_c or the population to which it applies is large, then policymakers might want less uncertainty, i.e. a high value for $\mathcal{A}(\lambda)$, perhaps 0.99. Conversely, if Treatment is expected to have very little impact, lower values of $\mathcal{A}(\lambda)$ might be tolerated, perhaps 0.8.

Although the CEAC in Figure 4.5 looks similar to a probability distribution function, in practice a CEAC can take on a variety of shapes, including those with negative slope. For a more complete discussion on the CEAC and the various shapes that it can take, see Fenwick *et al.* (2004).

4.5 USING BOOTSTRAP METHODS

Bootstrapping has been proposed as another approach for calculating the CEAC and confidence limits for the ICER, see Chaudhary and Sterns (1996), Briggs *et al.* (1997), Briggs and Fenn (1998), Briggs *et al.* (1999). For the bootstrap approach the analyst re-samples the

data at random with replacement, sampling the same number of observations as are in the original data set, i.e. n_T from arm T and n_S from arm S . Consequently, some patients are sampled more than once and some patients not at all. The procedure is repeated so that the data are re-sampled B times. The validity of the bootstrap approach rests on two asymptotics: (i) as the original sample size approaches the population size so the sample distribution tends to the population distribution; and, given this, (ii) as B , the number of bootstrap replications approaches infinity so the bootstrap estimate of the sampling distribution of a statistic approaches the true sampling distribution (Mooney and Duval, 1993). The analyst can then calculate the estimates for Δ_e , Δ_c and the ICER for the re-sampled data; these are denoted as $\hat{\Delta}_{ei}^*$, $\hat{\Delta}_{ci}^*$ and R_i^* , $i = 1, 2, \dots, B$, respectively. The set of values of R_i^* provide an estimator for the distribution of \hat{R} , the estimator of the ICER from the original set of data.

Efron and Tibshirani (1993) suggest that the 'ideal' bootstrap estimate corresponds to an infinite number of bootstrap re-samples. In practice, although there are no formal rules regarding the number of bootstrap replications required for reliable estimation, they suggest that 50 replications are usually adequate to provide an informative estimate of variance and that very seldom are more than 200 replications required. For the estimation of percentile confidence intervals, more replications are required in order to better estimate the tails of the sampling distribution. It is generally agreed that between 2000 and 5000 re-samples are sufficient to achieve stability for percentile interval estimates. A number of different methods of confidence interval estimation based on bootstrap replication are possible including: normal approximation, percentile, bias-corrected and accelerated percentile and percentile- t (Efron and Tibshirani 1993; Mooney and Duval 1993). However, the normal approximation and percentile- t methods make use of the bootstrap estimate of variance of the ICER which is not recommended because, as ratios, the values of R_i^* are not normally distributed, and extreme values occur when $\hat{\Delta}_{ei}^*$ approaches zero. The bias corrected and accelerated methods adjust for the inherent bias and skewness of ratio statistics. Nevertheless, we concentrate here on the straightforward percentile method, since, as we will go on to argue, this method is consistent with decision making on the

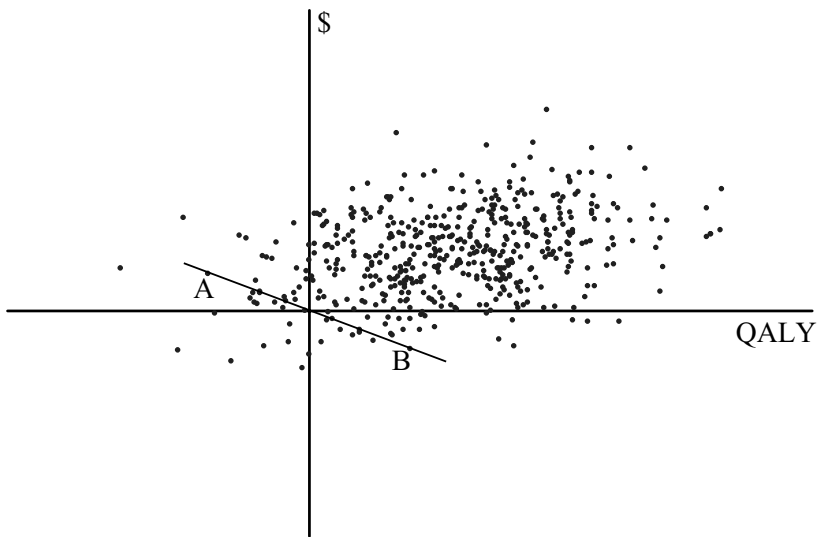


Figure 4.6 Plot of re-sampled $(\hat{\Delta}_{ei}^*, \hat{\Delta}_{ci}^*)$

cost-effectiveness plane and provides equivalent results using INB and acceptability curve approaches to handling uncertainty. For more details regarding general bootstrap methods the reader is referred to Efron and Tibshirani (1993). For specific details on the bootstrap estimation of confidence intervals for ICERs, including comparison of the four approaches to interval estimation outlined above, the reader is referred to Briggs *et al.* (1997) and Briggs *et al.* (1999).

An illustration of how the statistics from the re-samples are used to define a percentile confidence interval for R is given in Figure 4.6. Each point on the CE plane in Figure 4.6 represents a $(\hat{\Delta}_{ei}^*, \hat{\Delta}_{ci}^*)$ pair from one of the re-samples. To construct the $100(1 - 2\alpha)\%$ confidence interval the analyst must find two rays from the origin that enclose $\hat{\Delta}$ from the original data and $100(1 - 2\alpha)\%$ of the $(\hat{\Delta}_{ei}^*, \hat{\Delta}_{ci}^*)^T$ points from the re-samples. The percentile method is based on an ordering of the $(\hat{\Delta}_{ei}^*, \hat{\Delta}_{ci}^*)^T$ points. The ordering cannot be based on the R_i^* values because two different re-samplings of the data can have the same R_i^* , but be from different quadrants, and in the case of negative values can have totally different interpretations, one being in the win-win SE

quadrant and the other in the lose–lose NW quadrant. For example, consider points A and B in Figure 4.6. They have the same slope, but one (A) is a candidate to be in the upper tail and the other (B) in the lower. Whenever the re-samples lie on both sides of the vertical axis of the CE plane, confidence limits based on orderings of R_i^* will have inappropriate coverage. This point has often been overlooked in theoretical derivations and applications of bootstrap in cost-effectiveness analysis.

As an alternative to deriving bootstrap confidence limits for the ICER we first derive the CEAC from the bootstrap re-samples; so that

$$\mathcal{A}^B(\lambda) \equiv \sum_{i=1}^B I \left\{ \hat{\Delta}_{ci}^* \lambda - \hat{\Delta}_{ci}^* > 0 \right\} / B$$

and the confidence limits are given by $\mathcal{A}^B(\alpha)$ and $\mathcal{A}^B(1 - \alpha)$. As before, these limits define a ‘bow tie’ region as illustrated in Figure 4.2.

4.6 A BAYESIAN INCREMENTAL NET BENEFIT APPROACH

A Bayesian approach for INB can be employed by determining its posterior distribution, which is defined as $\mathcal{P}_\lambda(b) \equiv \Pr(b_\lambda \leq b)$. For an uninformative prior, and assuming normality, the posterior distribution of INB is normal with mean \hat{b}_λ and variance v_λ , and if we let $f_{b_\lambda}(\cdot)$ be the corresponding density function, then

$$\mathcal{P}_\lambda(b) = \int_{-\infty}^b f_{b_\lambda}(x) dx = \Phi \left((b - \hat{b}_\lambda) / \sqrt{v_\lambda} \right) \quad (4.8)$$

where $\Phi(\cdot)$ is the cumulative distribution function for a standard normal random variable. $\mathcal{P}_\lambda(b)$ is also the probability that the point $\hat{\Delta}$ lies above the line on the CE plane with slope λ and vertical intercept $-b$, since for all points above this line $\text{INB} < b$, see Figure 4.7. Again, the concentric ellipses represent contours of equal probability for the density function $f_{\hat{\Delta}}(\cdot, \cdot)$, and an alternative expression for $\mathcal{P}_\lambda(b)$ is

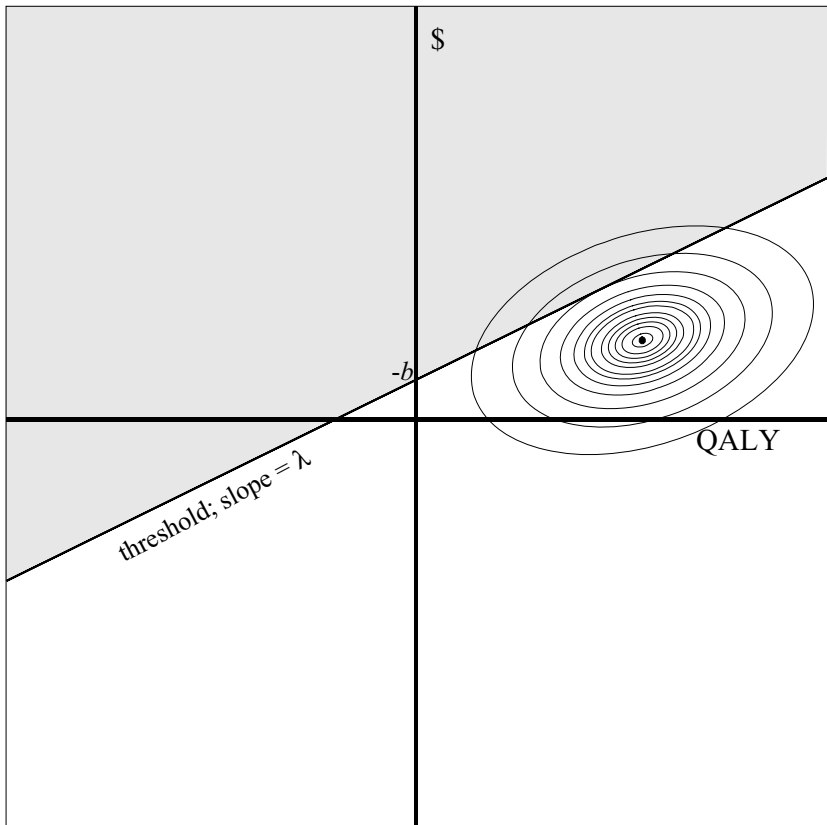


Figure 4.7 The CE plane; shaded area = $\Pr(b_\lambda < b)$

given by

$$\mathcal{P}_\lambda(b) = \int_{-\infty}^{\infty} \int_{e\lambda-b}^{\infty} f_{\hat{\Delta}}(e, c) dc de \quad (4.9)$$

Alternatively, bootstrapping can be employed; so that

$$\mathcal{P}_\lambda(b) = \frac{\sum_{i=1}^B I \left\{ \hat{\Delta}_{ei}^* \lambda - \hat{\Delta}_{ci}^* \leq b \right\}}{B}$$

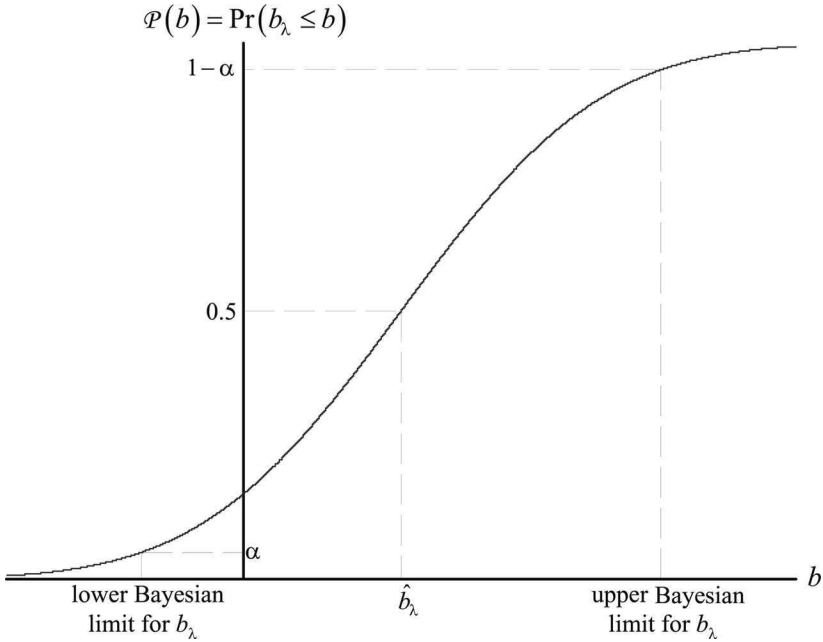


Figure 4.8 The posterior distribution for b_λ

The Bayesian estimate of INB is that value of b for which $\mathcal{P}_\lambda(b) = 0.5$, and the Bayesian lower and upper limits are the values of b for which $\mathcal{P}_\lambda(b) = \alpha$ and $\mathcal{P}_\lambda(b) = 1 - \alpha$, respectively. With an uninformative prior, $\mathcal{P}_\lambda(\hat{b}_\lambda) = 0.5$, $\mathcal{P}_\lambda(\hat{b}_\lambda - z_{1-\alpha}\sqrt{v_\lambda}) = \alpha$ and $\mathcal{P}_\lambda(\hat{b}_\lambda + z_{1-\alpha}\sqrt{v_\lambda}) = 1 - \alpha$, and the Bayesian and frequentist solution are the same, see Figure 4.8.

If the analyst has information on which to base a prior distribution for b_λ , a full Bayesian analysis is possible. If the prior distribution is normal with mean $b_{0\lambda}$ and variance $v_{0\lambda}$, then the posterior distribution for b_λ , given the data, is normal with mean $b_{1\lambda}$ and variance $v_{1\lambda}$, where

$$b_{1\lambda} = v_{1\lambda} \left(\frac{b_{0\lambda}}{v_{0\lambda}} + \frac{\hat{b}_\lambda}{\hat{v}_\lambda} \right) \quad \text{and} \quad v_{1\lambda} = \left(\frac{1}{v_{0\lambda}} + \frac{1}{\hat{v}_\lambda} \right)^{-1}$$

and if $f_{1b_\lambda}(\cdot)$ is the corresponding density function, then $\mathcal{P}_\lambda(b) = \int_{-\infty}^b f_{1b_\lambda}(x) dx = \Phi((b - b_{1\lambda})/\sqrt{v_{1\lambda}})$.

60 *Cost-effectiveness analysis*

The estimator of b_λ and the lower and upper Bayesian limits are the values of b that satisfy $\mathcal{P}_\lambda(b) = 0.5$, $\mathcal{P}_\lambda(b) = \alpha$ and $\mathcal{P}_\lambda(b) = 1 - \alpha$, respectively, and are given by $b_{1\lambda}$, $b_{1\lambda} - z_{1-\alpha}\sqrt{v_{1\lambda}}$ and $b_{1\lambda} + z_{1-\alpha}\sqrt{v_{1\lambda}}$, respectively. Performing these calculations for numerous values of λ in a sensitivity analysis will yield graphs as shown in Figure 4.3. The horizontal intercepts of these graphs provide posterior estimates and Bayesian limits for R , while the negative intercepts of the vertical axis provide posterior estimates and Bayesian limits for Δ_c .

4.7 KINKED THRESHOLDS

Under standard principles and assumptions of microeconomic theory, there should be little difference between a person's WTP for a commodity or program and the compensation they would demand—willingness to accept (WTA)—to relinquish the same commodity, see Johansson (1995). In practice there has been a wide and reproducible disparity between measured WTP and WTA values. In a recent meta-analysis of published studies by O'Brien *et al.* (2002), the ratio of WTA to WTP is approximately 7 for environmental studies and approximately 2 in the one health study. For a discussion of the competing theories as to why $WTA > WTP$ readers are referred to Hanemann (1991), Kahneman and Tversky (1979) and Morrison (1998).

Regardless of the reason, the implication is that the threshold dividing the cost-effectiveness plane into cost-effective and not-cost-effective regions is not a straight line through the origin with slope equal to the WTP, as seen in Figure 1.1. The threshold is now made up of two straight lines meeting at the origin. One, for $\Delta_e > 0$, has slope equal to the WTP, and the other, for $\Delta_e < 0$, has slope equal to the WTA. The new 'kinked' threshold is illustrated in Figure 4.9. The darkest shaded region is cost-effective if $WTA = WTP$, but not if $WTA > WTP$.

The cost-effectiveness analysis of a clinical trial needs to be adjusted if $WTA > WTP$. Let $WTP = \lambda$ and let $WTA = \gamma\lambda$, where $\gamma \geq 1$. The parameters Δ_e and Δ_c are estimated in the appropriate way, as discussed in previous chapters and earlier in this chapter, to yield $\hat{\Delta} = (\hat{\Delta}_e, \hat{\Delta}_c)^T$ with estimated covariance matrix $\hat{V}(\hat{\Delta})$. Recalling from Section 4.4

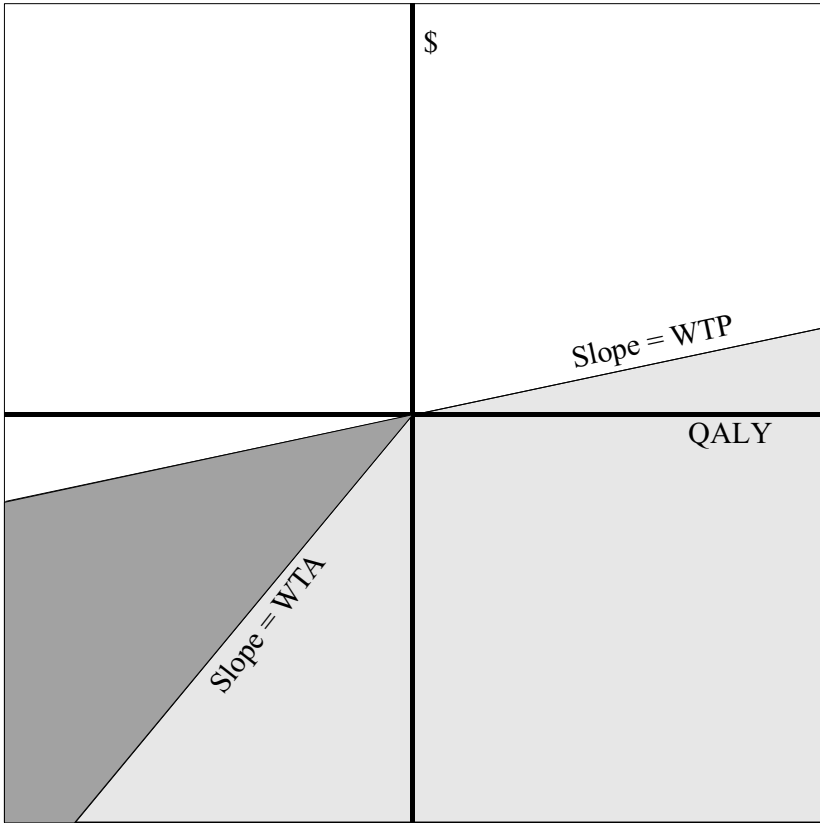


Figure 4.9 threshold for CE plane with $WTA > WTP$

that $f_{\hat{\Delta}}(\cdot, \cdot)$ is the density function for a bivariate normal with mean $\hat{\Delta}$ and covariance matrix $\hat{V}(\hat{\Delta})$ and that the CEAC for $\gamma = 1$ is given by

$$\mathcal{A}(\lambda) = \int_{-\infty}^{\infty} \int_{-\infty}^{\lambda e} f_{\hat{\Delta}}(e, c) dc de = \Phi(\hat{b}_{\lambda} / \sqrt{v_{\lambda}})$$

where \hat{b}_{λ} and v_{λ} are defined as in Section 4.3.

When $\gamma > 1$ the upper limit of the inner integral depends on whether e (i.e. Δ_e) is positive or negative, being λe when $e \geq 0$ and

62 Cost-effectiveness analysis

$\gamma\lambda e$ when $e < 0$. Therefore, to account for $\gamma > 1$ the CEAC becomes

$$\begin{aligned} \mathcal{A}_\gamma(\lambda) &= \int_{-\infty}^0 \int_{-\infty}^{\gamma\lambda e} f_{\hat{\Delta}}(e, c) \, dc \, de + \int_0^{\infty} \int_{-\infty}^{\lambda e} f_{\hat{\Delta}}(e, c) \, dc \, de \\ &= \mathcal{A}(\lambda) - \int_{-\infty}^0 \int_{\gamma\lambda e}^{\lambda e} f_{\hat{\Delta}}(e, c) \, dc \, de \\ &= \Phi(\hat{b}_\lambda / \sqrt{v_\lambda}) - \int_{-\infty}^0 \int_{\gamma\lambda e}^{\lambda e} f_{\hat{\Delta}}(e, c) \, dc \, de \end{aligned}$$

The last step is seen by recognizing that the difference between $\mathcal{A}_\gamma(\lambda)$ and $\mathcal{A}(\lambda)$ is the reduction of the probability density that lies to the left of the vertical axis in the ‘wedge’ between the lines through the origin with slopes $\gamma\lambda$ and λ , i.e. the darkest shaded area in Figure 4.9. The ICER, if it exists, equals $r_{\mathcal{A}_\gamma, 0.5}$ where $\mathcal{A}_\gamma(r_{\mathcal{A}_\gamma, \beta}) = \beta$ and the corresponding limits, if they exist, will equal $r_{\mathcal{A}_\gamma, \alpha}$ and $r_{\mathcal{A}_\gamma, 1-\alpha}$. Clearly $\mathcal{A}_\gamma(\lambda) < \mathcal{A}(\lambda)$, if $\gamma > 1$, and if both functions increase with λ (i.e. there is not too much probability density in the SW quadrant), then the estimate of the ICER and corresponding confidence limits will increase with γ , as illustrated in Figure 4.10.

To adjust the analysis of INB we must first recall that, for the case where $WTA = WTP = \lambda$, the estimate of INB and corresponding limits are given, respectively, by $b_{\mathcal{P}, 0.5}$, $b_{\mathcal{P}, \alpha}$ and $b_{\mathcal{P}, 1-\alpha}$, where $\mathcal{P}(b_{\mathcal{P}, \beta}) = \beta$ and

$$\mathcal{P}(b) = \int_{-\infty}^{\infty} \int_{\lambda e - b}^{\infty} f_{\hat{\Delta}}(e, c) \, dc \, de = \Phi((b - \hat{b}_\lambda) / \sqrt{v_\lambda}) \quad (4.10)$$

as defined in Equation (4.9). The function $\mathcal{P}(\cdot)$ is being treated as the posterior distribution for INB, i.e. $\mathcal{P}(b) = \Pr(\Delta_e \lambda - \Delta_e \leq b)$, and assuming normality and a flat prior yields Equation (4.10), see Willan, O’Brien and Leyva (2001). When $\gamma > 1$ the lower limit of the inner integral depends on whether e (i.e. Δ_e) is positive or negative, being $\lambda e - b$ when $e \geq 0$ and $\gamma\lambda e - b$ when $e < 0$. Therefore, to account

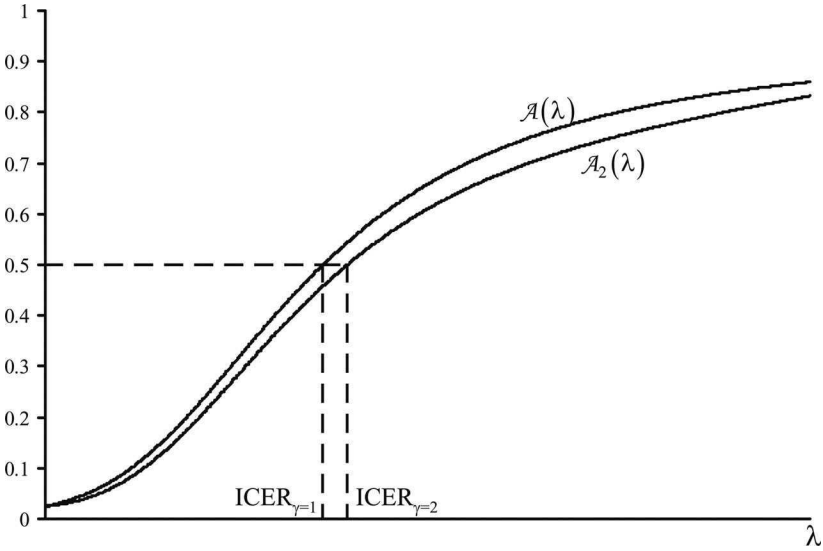


Figure 4.10 CEACs for $\gamma = 1$ and $\gamma = 2$

for $\gamma > 1$ the posterior distribution for INB becomes

$$\begin{aligned}
 \mathcal{P}_\gamma(b) &= \int_{-\infty}^0 \int_{\gamma\lambda e-b}^{\infty} f_{\hat{\Delta}}(e, c) dc de + \int_0^{\infty} \int_{\lambda e-b}^{\infty} f_{\hat{\Delta}}(e, c) dc de \\
 &= \mathcal{P}(b) + \int_{-\infty}^0 \int_{\gamma\lambda e-b}^{\lambda e-b} f_{\hat{\Delta}}(e, c) dc de \\
 &= \Phi((b - \hat{b}_\lambda)/\sqrt{v_\lambda}) + \int_{-\infty}^0 \int_{\gamma\lambda e-b}^{\lambda e-b} f_{\hat{\Delta}}(e, c) dc de
 \end{aligned}$$

The last step is seen by recognizing that the difference between $\mathcal{P}_\gamma(\lambda)$ and $\mathcal{P}(\lambda)$ is the addition of the probability density that lies to the left of the vertical axis in the ‘wedge’ between the lines through the point $(0, -b)$ with slopes $\gamma\lambda$ and λ , i.e. the darkest shaded area in

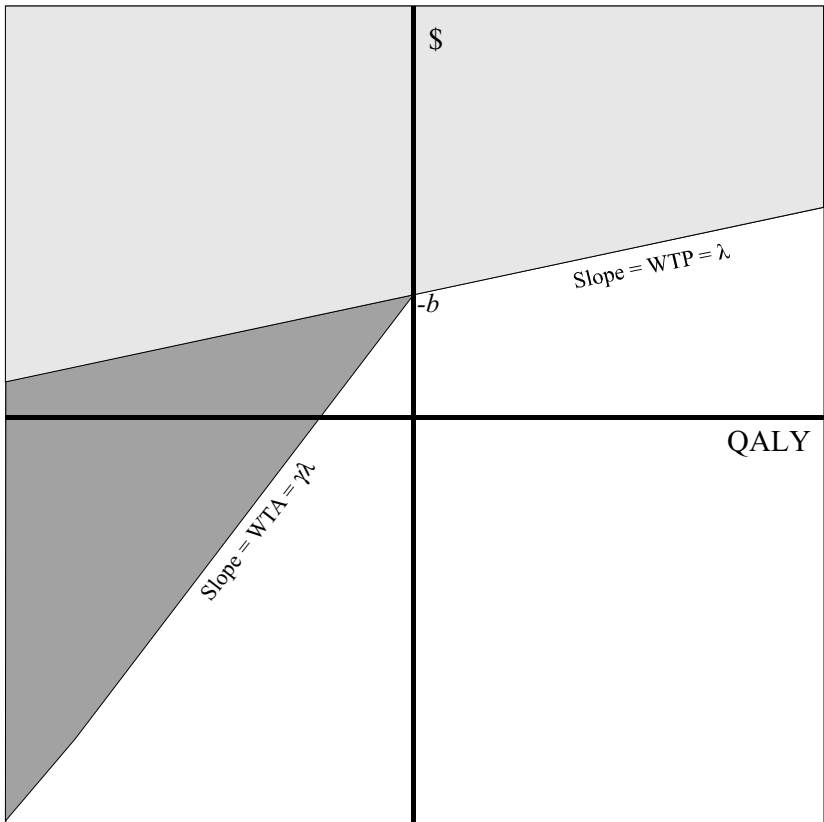


Figure 4.11 Region in which INB is less than or equal to b

Figure 4.11. The estimate of INB and corresponding limits are given by $b_{\mathcal{P}_\gamma, 0.5}$, $b_{\mathcal{P}_\gamma, \alpha}$ and $b_{\mathcal{P}_\gamma, 1-\alpha}$, respectively, where $\mathcal{P}_\gamma(b_{\mathcal{P}_\gamma, \beta}) = \beta$. Clearly $\mathcal{P}_\gamma(\lambda) > \mathcal{P}(\lambda)$, if $\gamma > 1$, and therefore estimate of the INB and corresponding confidence limits will decrease with γ , as illustrated in Figure 4.12.

4.8 SUMMARY

Contained in this chapter are methods of statistical inference for cost-effectiveness trials using estimators of the five key parameters. Note

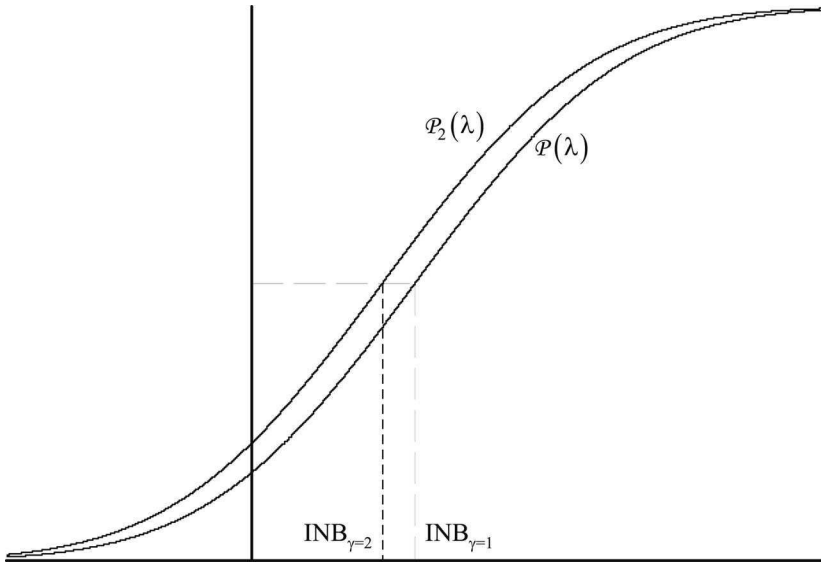


Figure 4.12 Posterior distribution for INB for $\gamma = 1$ and $\gamma = 2$

that the presence of censoring affects only how these five basic parameters are estimated (see Chapter 2 for non-censored data and Chapter 3 for censored data) not how they are subsequently used in analyses. Although the chapter has reviewed in detail both the ICER and INB approach to analyzing cost-effectiveness information and emphasized their formal equivalence, it should be clear that the INB approach offers a number of statistical and practical advantages, avoiding as it does the problems associated with ratio statistics. Nevertheless, in making use of the threshold decision rule, λ , which is generally unknown to the analyst, it is recommended that any analysis be presented as a function of λ . This could be either in terms of the INB and its associated confidence interval directly or as a CEAC which has its most natural interpretation within a Bayesian framework. In the following chapter the methods of estimation and inference reviewed in this chapter are illustrated using several examples.

Cost-effectiveness Analysis: Examples

5.1 INTRODUCTION

Five examples of cost-effectiveness analyses will be given in detail in this chapter, illustrating different aspects of the material covered in preceding chapters. For non-censored data, an example with a binary measure of effectiveness is given in Section 5.2 and an example with quality-adjusted survival time as the measure of effectiveness is given in Section 5.3. For censored data, an example with survival time as the measure of effectiveness is given in Section 5.4 and an example with quality-adjusted survival time as the measure of effectiveness is given in Section 5.6. A final example presents a Bayesian analysis contrasting the use of prior information and uninformative priors. Emphasis will be placed on using the appropriate methodology for estimating the five parameters of interest, and on using the plot of the incremental net benefit by willingness-to-pay and the cost-effectiveness acceptability curve for inference. The use of different decision thresholds in the SW and NE quadrants of the cost-effectiveness plane is also demonstrated.

5.2 THE CADET-Hp TRIAL

The CADET-Hp Trial is a double-blind, placebo-controlled, parallel-group, multicenter, randomized controlled trial performed in 36

68 Cost-effectiveness analysis: examples

family practitioner centres across Canada. The results are published in Chiba *et al.* (2002) and Willan (2004). Patients 18 years and over with uninvestigated dyspepsia of at least moderate severity presenting to their family physicians were eligible for randomization, provided they did not have any alarm symptoms and were eligible for empiric drug therapy. Patients were randomized between

- T: Omeprazole 20 mg, metronidazole 500 mg and clarithromycin 250 mg, and
- S: Omeprazole 20 mg, placebo metronidazole and placebo clarithromycin.

Both regimens were given twice daily for seven days. Treatment success was defined as the presence of no or minimal dyspepsia symptoms at one year. Total costs were determined from the societal perspective and are given in Canadian dollars.

A summary of the parameter estimates is given in Table 5.1. The elements in Table 5.1 are calculated by applying Equations (2.1–2.5). Costs are given in Canadian dollars (CAD\$). T is observed to increase success and decrease costs and, consequently, the point $(\hat{\Delta}_e, \hat{\Delta}_c)$ lies in the SE (win–win) quadrant and \hat{b}_λ is positive for all positive values of λ . For sake of argument suppose the WTP for a success is CAD\$ 1000. That is, if the true value of INB evaluated at 1000 is less than or equal to zero, T should not be adopted. This, then, becomes the null hypothesis, expressed below as H_0 . On the other hand, if INB evaluated at 1000 is greater than zero, T should be adopted. This becomes the

Table 5.1 Parameter estimates for the CADET-Hp trial

	$T(n_T = 142)$	$S(n_S = 146)$		Equation
$\hat{\pi}_j$	0.5070	0.3699	difference = $\hat{\Delta}_e = 0.1371$	(2.3)
\hat{v}_j	476.97	529.98	difference = $\hat{\Delta}_c = -53.01$	(2.1)
$\hat{V}(\hat{\pi}_j)$	0.00176	0.001596	sum = $\hat{V}(\hat{\Delta}_e)$ = 0.003356	(2.4)
$\hat{V}(\hat{v}_j)$	2167	2625	sum = $\hat{V}(\hat{\Delta}_c) = 4792$	(2.2)
$\hat{C}(\hat{\pi}_j, \hat{v}_j)$	-0.2963	-0.4166	sum = $\hat{C}(\hat{\Delta}_e, \hat{\Delta}_c)$ = -0.7129	(2.5)

alternative hypothesis, expressed below as H_1 . Thus the investigators wish to test the hypothesis

$$H_0: b_{1000} \leq 0 \quad \text{versus} \quad H_1: b_{1000} > 0$$

at the 5% level. Rejection of H_0 in favor of H_1 provides evidence that T is cost-effective compared with S . First, the investigators must determine:

$$\hat{b}_{1000} = 1000\hat{\Delta}_e - \hat{\Delta}_c = 1000 \times 0.1371 - (-53.01) = 190.1$$

and

$$\begin{aligned} v_{1000} &= \hat{V}(\hat{b}_{1000}) = 1000^2 \times \hat{V}(\hat{\Delta}_e) + \hat{V}(\hat{\Delta}_c) - 2 \times 1000 \\ &\quad \times C(\hat{\Delta}_e, \hat{\Delta}_c) = 1000^2 \times 0.003356 + 4792 - 2 \times 1000 \\ &\quad \times (-0.7129) = 9574 \end{aligned}$$

The z -statistic for testing H_0 is $\hat{b}_{1000}/\sqrt{v_{1000}} = 190.1/\sqrt{9574} = 1.943$. Since $1.943 > z_{0.95} (= 1.645)$, H_0 can be rejected in favour of H_1 , and the investigators can conclude that there is evidence that T is cost-effective compared with S if the WTP is at least CAD\$ 1000. The 90% confidence limits for b_{1000} are given by:

$$\hat{b}_{1000} \pm z_{0.95}\sqrt{v_{1000}} = 190.1 \pm 1.645 \times 97.85 = 29.15, 351.1.$$

Since the lower limit is greater than 0, we can, as above, reject H_0 in favour of H_1 , at the 5% level. A two-sided 90% confidence interval is constructed because we wish to test a one-sided null hypothesis at the 5% level.

For any value of λ , the quantities \hat{b}_λ and v_λ are given by $0.1371\lambda + 53.01$ and $0.003356\lambda^2 + 4792 + 2 \times 0.7129\lambda$, respectively, and the 90% confidence limits are given by $0.1371\lambda + 53.01 \pm 1.645\sqrt{0.003356\lambda^2 + 4792 + 2 \times 0.7129\lambda}$. The quantity \hat{b}_λ and corresponding 90% confidence limits can be calculated for a large range of λ and plotted as shown in Figure 5.1. The plot of \hat{b}_λ has slope 0.1371 ($= \Delta_e$), crosses the vertical axis at 53.01 ($= -\hat{\Delta}_c$) and the horizontal axis at $-53.01/0.1371 = -386.7$ ($= ICER$). By observing where the confidence limits cross the vertical axis, one can make inference regarding the difference between arms with respect to cost, in what is essentially a cost-minimization analysis. Since the

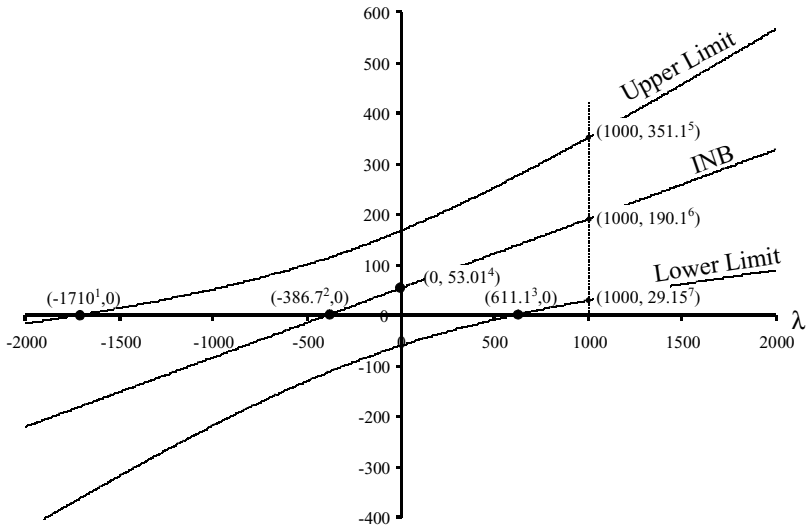


Figure 5.1 INB versus λ for the CADET-Hp trial. ¹Lower confidence limit for ICER; ²ICER; ³upper confidence limit for ICER; ⁴estimate of $-\Delta_c$; ⁵upper confidence limit for INB(1000); ⁶estimate of INB(1000); ⁷lower confidence limit for INB(1000)

confidence interval includes 0, the null hypothesis of no difference in mean cost cannot be rejected at the 10% level (note that this is a two-sided test). The horizontal intercepts of the confidence limits for b_λ define the Fieller confidence interval for the ICER. Therefore, by focusing attention on the horizontal axis one can perform a cost-effectiveness analysis using an ICER approach. The 90% confidence limits for the ICER are -1710 and 611.1 . The hypothesis that the ICER is greater than any value above 611.1 can be rejected at the 5% level since 611.1 is the upper limit of the 90% confidence interval. The Fieller 90% confidence limits for the ICER can be calculated directly from applying Equation (4.4). That is, the 90% confidence limits are given by:

$$\hat{R} \left\{ \left(1 - z_{1-0.05}^2 c \pm z_{1-0.05} \sqrt{a + b - 2c - z_{1-0.05}^2 (ab - c^2)} \right) / \left(1 - z_{1-0.05}^2 a \right) \right\}$$

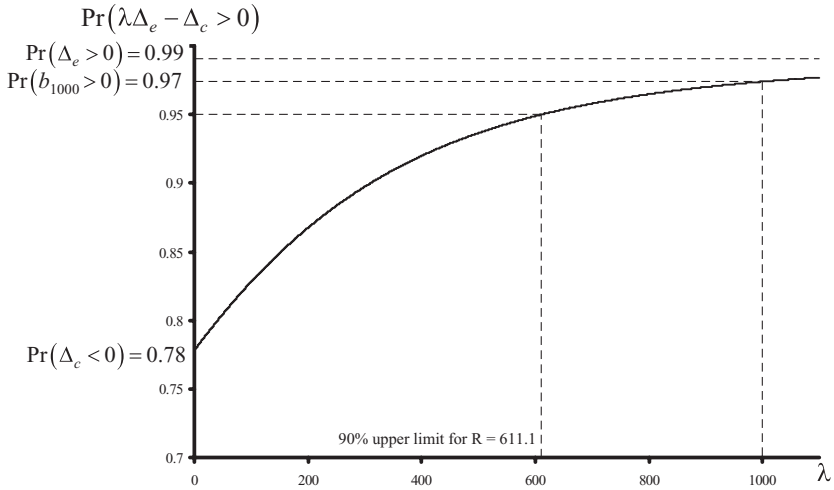


Figure 5.2 Cost-effectiveness acceptability curve for CADET-Hp trial

The CEAC is shown in Figure 5.2. Since the ICER and its lower 90% confidence limit are negative, they are not shown in the figure. The curve meets the vertical axis ($\lambda = 0$) at $\mathcal{A}(0) = \Phi(\hat{b}_0/\sqrt{\hat{v}_0}) = \Phi(-\hat{\Delta}_c/\sqrt{\hat{V}(\hat{\Delta}_c)}) = \Pr(\Delta_e < 0) = \Phi(53.01/\sqrt{4792}) = 0.7781$ and asymptotically approaches $\Pr(\Delta_e > 0) = \Phi(\hat{\Delta}_e/\sqrt{\hat{V}(\hat{\Delta}_e)}) = \Phi(0.1371/\sqrt{0.003356}) = 0.9910$. Since $\mathcal{A}(1000) = 0.9741$, if the WTP for a success is CAD\$ 1000, we could say that the probability that T is cost-effective is 97.41%.

5.3 SYMPTOMATIC HORMONE-RESISTANT PROSTATE CANCER

In a trial of symptomatic, hormone-resistant prostate cancer, 161 patients were randomized between prednisone alone (S) and prednisone plus mitoxantrone (T). The clinical results are published in Tannock *et al.* (1996) and the economics analysis in Bloomfield *et al.* (1998). Although there was no statistically significant difference in survival, patients experienced a better quality of life with T . Cost data,

Table 5.3 Parameter estimates for the prostate trial

	$T(n_T = 61)$	$S(n_S = 53)$		Equation
$\hat{\phi}_j$	40.89	28.11	difference = $\hat{\Delta}_e = 12.78$	(2.10)
\hat{v}_j	27 322	29 039	difference = $\hat{\Delta}_c = -1717$	(2.1)
$\hat{V}(\hat{\phi}_j)$	24.10	16.42	sum = $\hat{V}(\hat{\Delta}_e) = 40.52$	(2.11)
$\hat{V}(\hat{v}_j)$	6466 351	7872 681	sum = $\hat{V}(\hat{\Delta}_c) = 14 339 032$	(2.2)
$\hat{C}(\hat{\phi}_j, \hat{v}_j)$	2771	2876	sum = $\hat{C}(\hat{\Delta}_e, \hat{\Delta}_c) = 5647$	(2.12)

including hospital admissions, outpatient visits, investigations, therapies and palliative care, were collected retrospectively on the 114 patients from the three largest centres. Survival was quality-adjusted using the EORTC quality of life questionnaire QLQ-C30. All patients were followed until death.

A summary of the parameter estimates is given in Table 5.3. Costs are given in Canadian dollars and effectiveness in quality-adjusted life-weeks (QALW). T is observed to increase quality-adjusted survival and decrease costs, and consequently, the point $(\hat{\Delta}_e, \hat{\Delta}_c)$ lies in the SE (win-win) quadrant and \hat{b}_λ is positive for all positive values of λ . If the WTP for a QALW is CAD\$ 500, the investigators wish to test the hypothesis

$$H_0: b_{500} \leq 0 \quad \text{versus} \quad H_1: b_{500} > 0$$

at the 5% level, and conclude that there is evidence to adopt T if H_0 is rejected. First, the investigators must determine:

$$\hat{b}_{500} = 500 \times \hat{\Delta}_e - \hat{\Delta}_c = 500 \times 12.78 - (-1717) = 8107$$

and

$$\begin{aligned} v_{500} &= 500^2 \times \hat{V}(\hat{\Delta}_e) + \hat{V}(\hat{\Delta}_c) - 2 \times 500 \times C(\hat{\Delta}_e, \hat{\Delta}_c) \\ &= 500^2 \times 40.52 + 14\,339\,032 - 2 \times 500 \times 5647 \\ &= 18\,822\,032 \end{aligned}$$

The z-statistic for testing H_0 is $\hat{b}_{500} / \sqrt{v_{500}} = 8107 / \sqrt{18\,822\,032} = 1.869$. Since $1.869 > z_{0.95}(= 1.645)$, H_0 can be rejected in favour of

74 Cost-effectiveness analysis: examples

H_1 , and the investigators can conclude that there is evidence that T is cost-effective compared with S if the WTP is at least CAD\$ 500. The 90% confidence limits for b_{500} are given by:

$$\hat{b}_{500} \pm z_{0.95}\sqrt{v_{500}} = 8107 \pm 1.645 \times 4338 = 971.0, 15243$$

Since the lower limit is greater than 0, we can, as above, reject H_0 in favour of H_1 , at the 5% level. A two-sided 90% confidence interval is constructed because we wish to test a one-sided null hypothesis at the 5% level.

For any value of λ , the quantities \hat{b}_λ and v_λ are given by $12.78\lambda + 1717$ and $40.52\lambda^2 + 14\,339\,032 - 2 \times 5647\lambda$, respectively, and the 90% confidence limits are given by $12.78\lambda + 1717 \pm 1.645 \times \sqrt{40.52\lambda^2 + 14\,339\,032 - 2 \times 5647\lambda}$. The quantity \hat{b}_λ and corresponding 90% confidence limits can be calculated for a large range of λ and plotted as shown in Figure 5.3. The plot of \hat{b}_λ has slope 12.78

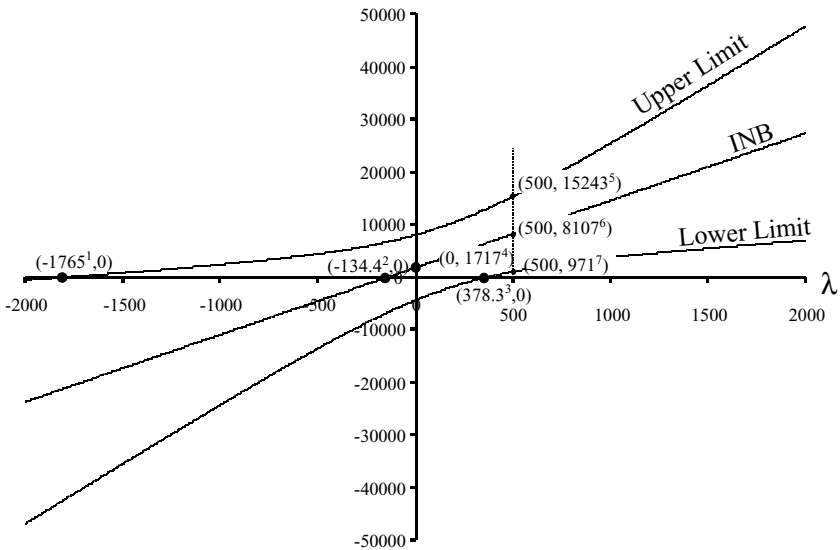


Figure 5.3 INB versus λ for the prostate trial. ¹Lower confidence limit for ICER; ²ICER; ³upper confidence limit for ICER; ⁴estimate of $-\Delta_c$; ⁵upper confidence limit for INB(500); ⁶estimate of INB(500); ⁷lower confidence limit for INB(500)

(= Δ_e), crosses the vertical axis at 1717 (= $-\hat{\Delta}_c$) and the horizontal axis at $-1717/12.78 = -134.4$ (= ICER). By observing where the confidence limits cross the vertical axis, one can make inference regarding the difference between arms with respect to cost, in what is essentially a cost-minimization analysis. Since the confidence interval includes 0, the null hypothesis of no difference in mean cost cannot be rejected at the 10% level (note that this is a two-sided test). The horizontal intercepts of the confidence limits define the Fieller confidence interval for the ICER. Therefore, by focusing attention on the horizontal axis one can perform a cost-effectiveness analysis using an ICER approach. The 90% confidence limits for the ICER are -1765 and 378.3 . The hypothesis that the ICER is greater than any value above 378.3 can be rejected at the 5% level since 378.3 is the upper limit of the 90% confidence interval. The 90% confidence limits for the ICER can be calculated directly from applying Equation (4.4). That is, 90% confidence limits are given by:

$$\hat{R} \left\{ \left(1 - z_{1-0.05}^2 c \pm z_{1-0.05} \sqrt{a + b - 2c - z_{1-0.05}^2(ab - c^2)} \right) / \right. \\ \left. (1 - z_{1-0.05}^2 a) \right\}$$

The associated elements for Equation (4.4) are given in Table 5.4.

Using Figure 5.3 one can perform a cost-effectiveness analysis using an INB approach for any value of λ . In particular, for $\lambda = 500$, the confidence interval contains only positive values, which means, as above, the null hypothesis that $b_{500} \leq 0$ can be rejected and that there is evidence to support the adoption of T . This is true for any value of

Table 5.4 Elements for calculation confidence limits for ICER for the prostate trial

$\hat{R} = \hat{\Delta}_c / \hat{\Delta}_e = -1717 / 12.78 = -134.4$
$a = \hat{V}(\hat{\Delta}_e) / \hat{\Delta}_e^2 = 40.52 / 12.78^2 = 0.2481$
$b = \hat{V}(\hat{\Delta}_c) / \hat{\Delta}_c^2 = 14\,339\,032 / (-1717)^2 = 4.864$
$c = \hat{C}(\hat{\Delta}_e, \hat{\Delta}_c) / (\hat{\Delta}_e \hat{\Delta}_c) = 5647 / (12.78 \times (-1717)) = -0.2573$
$z_{1-0.05} = 1.645$

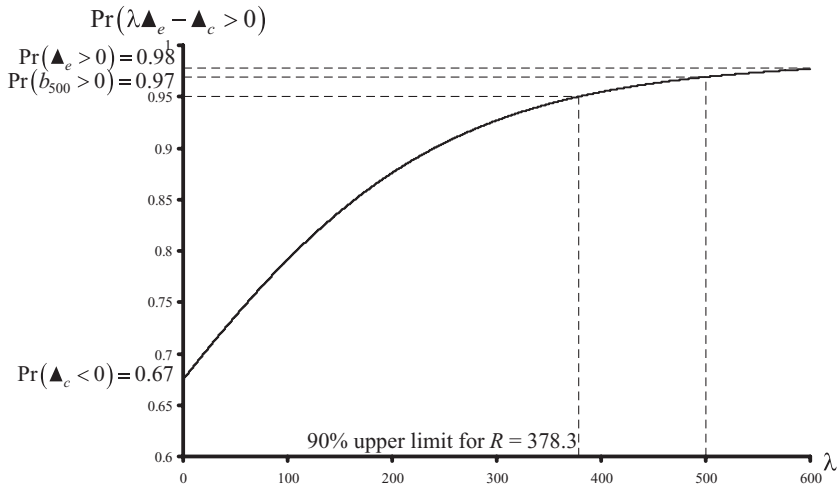


Figure 5.4 Cost-effectiveness acceptability curve for prostate trial

λ that is greater than 378.3 (upper limit of the ICER). For values of λ less than 378.3 the confidence interval includes negative values and the null hypothesis that $b_\lambda \leq 0$ cannot be rejected.

The cost-effectiveness acceptability curve can be calculated by:

$$A(\lambda) = \Phi\left(\frac{\hat{b}_\lambda \sqrt{v_\lambda}}{\sqrt{40.52\lambda^2 + 14\,339\,032 - 2 \times 5647\lambda}}\right) = \Phi\left(\frac{12.78\lambda + 1717}{\sqrt{40.52\lambda^2 + 14\,339\,032 - 2 \times 5647\lambda}}\right)$$

The curve is illustrated in Figure 5.4 and meets the vertical axis at $\Pr(\Delta_c < 0) = \Phi(1717/\sqrt{14\,339\,032}) = 0.6749$ and asymptotically approaches $\Pr(\Delta_e > 0) = \Phi(12.78/\sqrt{40.52}) = 0.9777$. Since $A(500) = 0.9692$, if the WTP for a QALW is CAD\$500, we could say that the probability that T is cost-effective is 96.92%.

The potential use of different threshold decision rules in the SW and NE quadrants of the cost-effectiveness plane is demonstrated for the prostate example in Table 5.5. In the upper panel of the table, the probability that Treatment is cost-effect and the estimate of INB, with 90% Bayesian credible intervals are presented under the assumption of a WTP of CAD\$400 per quality-adjusted life-week, for various values of γ , the ratio of WTP to WTA in the two quadrants (see Section 4.6).

Table 5.5 Estimate of INB and Bayesian interval for $\lambda = 400/\text{QALW}$

	γ	$A_\gamma(400)$	INB	90% credible interval
$\hat{\Delta}_e = 12.8$	1	0.9546	6836	195–13 477
	2	0.9526	6828	116–13 477
	5	0.9475	6817	–92–13 476
	10	0.9434	6811	–281–13 476
$\hat{\Delta}_e = 0$	1	0.6647	1715	–4925–691
	2	0.5759	917	–7215–6629
	5	0.4727	–448	–18 679–6409
	10	0.4324	–1515	–39 146–6328

The first row with $\gamma = 1$ corresponds to the conventional approach of using the same value of lambda throughout. Since the estimate $\hat{\Delta}_e$ is statistically significant, and little of the probability density of $(\hat{\Delta}_e, \hat{\Delta}_c)^T$ lies to the left of the vertical axis, as γ increases, only small changes are seen. However, the hypothesis $b_{400} \leq 0$ (versus $b_{400} > 0$) can be rejected for $\gamma = 1$ or 2, but not for $\gamma = 5$ or 10. For illustration, the analysis was rerun setting $\hat{\Delta}_e = 0$, so that 50% of the probability lies to the left of the vertical axis. The results, shown in the bottom panel in Table 5.5, illustrate that in this circumstance the results are quite sensitive to γ .

5.4 THE CANADIAN IMPLANTABLE DEFIBRILLATOR STUDY (CIDS)

In a trial of patients at risk of cardiac arrest, a total of 659 patients with resuscitated ventricular fibrillation or sustained ventricular tachycardia or with unmonitored syncope were randomized between amiodarone and implantable cardioverter defibrillator. Due to budgetary constraints, the costs were collected on the first 430 patients only. The clinical results are reported in Connolly *et al.* (2000), and the economic evaluation in O'Brien *et al.* (2000). For this example the measure of effectiveness is survival time, with the duration of interest set to 6.5 years, i.e. $\tau = 6.5$. Administrative censoring occurred and,

78 Cost-effectiveness analysis: examples

Table 5.6 Parameter estimates for the CIDS trial using inverse probability weighting

	$T(n_T = 212)$	$S(n_S = 218)$		Equation
$\hat{\mu}_j$	4.832	4.682	difference = $\hat{\Delta}_e = 0.1500$	(3.27)
\hat{v}_j	87 044	38 819	difference = $\hat{\Delta}_c = 48 225$	(3.12, 3.13)
$\hat{V}(\hat{\mu}_j)$	0.02418	0.02440	sum = $\hat{V}(\hat{\Delta}_e) = 0.04858$	(3.28)
$\hat{V}(\hat{v}_j)$	8462 152	6497 962	sum = $\hat{V}(\hat{\Delta}_c)$ = 14 960 114	(3.14)
$\hat{C}(\hat{\mu}_j, \hat{v}_j)$	125.8	20.42	sum = $\hat{C}(\hat{\Delta}_e, \hat{\Delta}_c)$ = 146.2	(3.30)

consequently, not all patients were followed for the entire duration of interest. Total costs are given in Canadian dollars, and were collected every three months for a total of 26 three-monthly intervals.

The estimators of the parameters of interest, using inverse probability weighting (IPW), are given in Table 5.6. By contrast the parameters of interest using life-table methods for effectiveness and the direct method for cost are given in Table 5.7. The estimates are almost identical and the following analyses are performed using the IPW estimates. T is observed to increase mean survival and increase costs, and consequently, the point $(\hat{\Delta}_e, \hat{\Delta}_c)$ lies in the NE (trade-off) quadrant. For any value of λ , the quantities \hat{b}_λ and v_λ are given

Table 5.7 Parameter estimates for the CIDS trial using life-table methods for effectiveness and the direct method for cost

	$T(n_T = 212)$	$S(n_S = 218)$		Equation
$\hat{\mu}_j$	4.832	4.682	difference = $\hat{\Delta}_e = 0.1500$	(3.21)
\hat{v}_j	87 103	38 864	difference = $\hat{\Delta}_c = 48 239$	(3.2)
$\hat{V}(\hat{\mu}_j)$	0.02443	0.02437	sum = $\hat{V}(\hat{\Delta}_e) = 0.04858$	(3.22)
$\hat{V}(\hat{v}_j)$	8461 538	6519 142	sum = $\hat{V}(\hat{\Delta}_c)$ = 14 980 680	(3.5)
$\hat{C}(\hat{\mu}_j, \hat{v}_j)$	124.9	14.20	sum = $\hat{C}(\hat{\Delta}_e, \hat{\Delta}_c)$ = 139.1	(3.26)

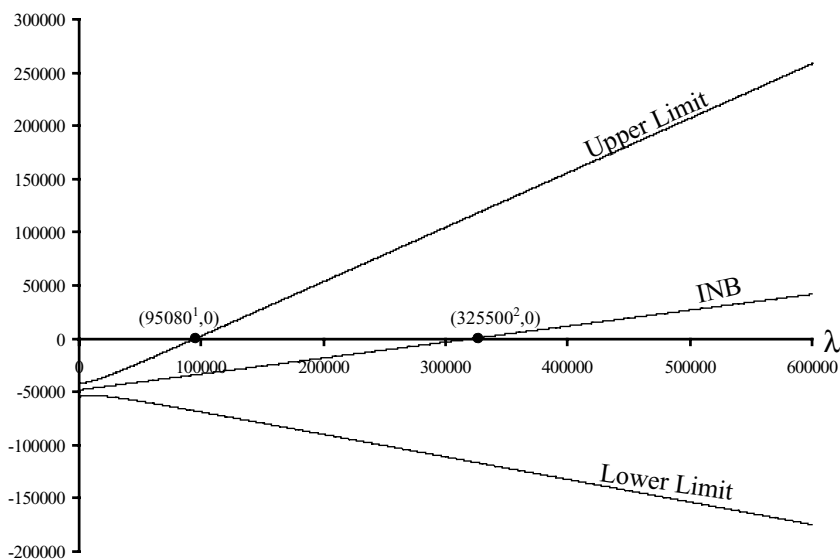


Figure 5.5 INB versus λ for the CIDS trial. ¹Lower confidence limit for ICER; ²ICER

by $0.15\lambda - 48225$ and $0.04858\lambda^2 + 14960114 - 2 \times 146.2\lambda$, respectively, and the 90% confidence limits are given by $0.15\lambda - 48225 \pm 1.645 \times \sqrt{0.04858\lambda^2 + 14960114 - 2 \times 146.2\lambda}$. The quantity \hat{b}_λ and corresponding 90% confidence limits are plotted in Figure 5.5. The plot of \hat{b}_λ has slope 0.15 ($= \Delta_e$), crosses the vertical axis at -48225 ($= -\hat{\Delta}_c$) and the horizontal axis at $48225/0.15 = 321500$ ($= \text{ICER}$). By observing where the confidence limits cross the vertical axis, one can make inference regarding the difference between arms with respect to cost, in what is essentially a cost-minimization analysis. Since the confidence interval includes only negative values, the null hypothesis of no difference in mean cost can be rejected at the 10% level, and the investigators can conclude that there is evidence that T increases costs. (Recall that the confidence interval defined on the vertical axis is for $-\Delta_c$, and since the confidence interval for $-\Delta_c$ includes only negative values, the confidence interval for Δ_c includes only positive values.)

Table 5.8 Elements for calculation confidence limits for ICER for the CIDS trial

$\hat{R} = \hat{\Delta}_c / \hat{\Delta}_e = 48\,225 / 0.1500 = 321\,500$
$a = \hat{V}(\hat{\Delta}_e) / \hat{\Delta}_e^2 = 0.04858 / 0.1500^2 = 2.159$
$b = \hat{V}(\hat{\Delta}_c) / \hat{\Delta}_c^2 = 14\,960\,114 / (48\,225)^2 = 0.006433$
$c = \hat{C}(\hat{\Delta}_e, \hat{\Delta}_c) / (\hat{\Delta}_e \hat{\Delta}_c) = 146.2 / (0.1500 \times 48\,225) = 0.02021$
$z_{1-0.05} = 1.645$

The horizontal intercepts of the confidence limits define the Fieller confidence interval for the ICER and, by focusing attention on the horizontal axis, one can perform a cost-effectiveness analysis using an ICER approach. The 90% lower limit for the ICER is 95 080. Since the lower limit of the INB fails to cross the horizontal axis, the upper limit of the ICER approaches $+\infty$, i.e. the confidence interval for the ICER includes the positive vertical axis of the cost-effectiveness plane. This is interpreted to mean that there is no value for the WTP that would make T cost-effective with 95% confidence. Likewise, since the lower limit of the INB fails to cross the horizontal axis, the confidence interval for INB always includes negative values, which leads to the same conclusion. The 90% confidence limits for the ICER can be calculated directly from applying Equation (4.4), whose associated elements are given in Table 5.8. The solution to Equation (4.4) yields the interval [95 080, $-220\,626$]. Since the lower limit is positive and the upper limit is negative, the confidence region on the CE plane includes part of the NW (lose–lose) quadrant, see Figure 5.6. Consequently, a proper representation of the interval is given by [95 080, $+\infty$), since the largest value in the interval approaches $+\infty$.

The cost-effectiveness acceptability curve can be calculated by:

$$\begin{aligned}
 A(\lambda) &= \Phi(\hat{b}_\lambda / \sqrt{v_\lambda}) \\
 &= \Phi\left((0.15\lambda - 48225) / \sqrt{0.04858\lambda^2 + 14960114 - 2 \times 146.2\lambda}\right)
 \end{aligned}$$

The curve, see Figure 5.7, meets the vertical axis ($\lambda = 0$) at $\Pr(\Delta_c < 0) = \Phi(-48\,225 / \sqrt{14\,960\,114}) = 5.564 \times 10^{-36}$ and asymptotically approaches $\Pr(\Delta_e > 0) = \Phi(0.15 / \sqrt{0.04858}) = 0.7519$.

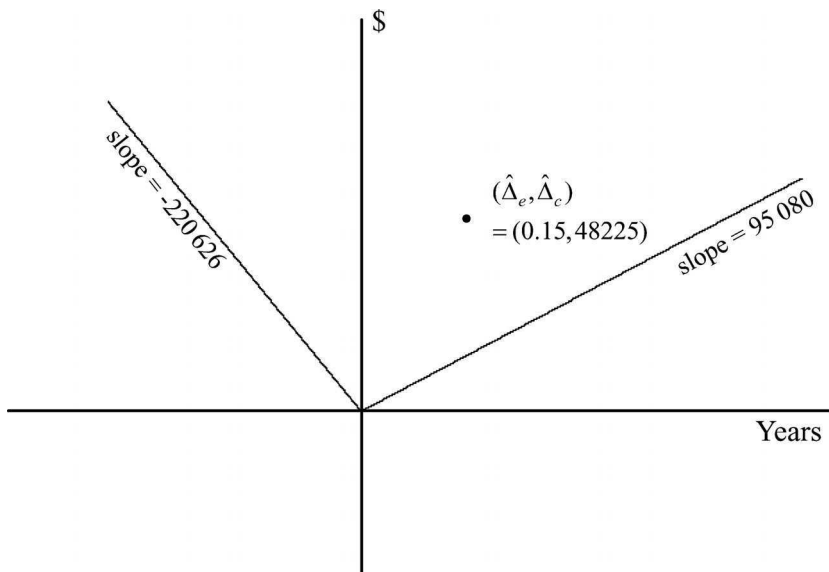


Figure 5.6 ICER confidence region on the CE plane for the CIDS trial

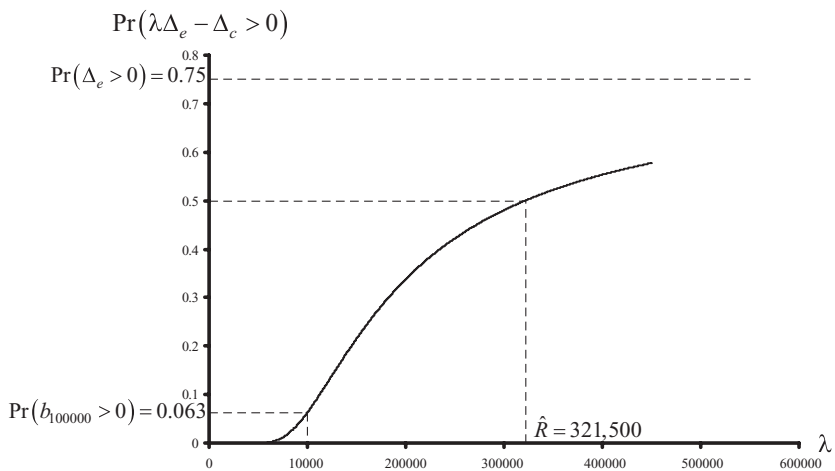


Figure 5.7 Cost-effectiveness acceptability curve for CIDS trial

82 Cost-effectiveness analysis: examples

Since $\mathcal{A}(100\,000) = 0.06314$, if the WTP for a year of life is CAD\$100,000, we could say that the probability that T is cost-effective is 6.314%.

5.5 THE EVALUATE TRIAL

The EVALUATE trial is a randomized comparison of laparoscopic-assisted hysterectomy (T) versus standard hysterectomy (S). Detailed reports of the trial can be found elsewhere, see Garry *et al.* (2004) and Sculpher *et al.* (2004). Patients were stratified into two groups, those undergoing vaginal hysterectomy and those undergoing abdominal hysterectomy, and the strata were analyzed and reported separately. Randomization was 2-to-1 in favor of laparoscopy. Allocation by treatment group and type of hysterectomy is given in Table 5.9. Effectiveness was expressed in quality-adjusted life-years, and measured using the EQ-5D questionnaire, see Kind (1996). EQ-5D measurements were taken at randomization, and at three follow-up visits: six weeks, four months and one year. These follow-up visits defined the three time intervals (i.e. $K = 3$) with $a_1 = 0$, $a_2 = 6$, $a_3 = 17$ and $a_4 = 52$, expressed in weeks. Since the boundaries of the time interval coincided with the EQ-5D measurements (as they often would), q_k , the quality-adjusted life-year score for interval k , equals

$$\frac{(Q_{k+1} + Q_k)}{2} \times \frac{(a_{k+1} - a_k)}{52}$$

where Q_k is the EQ-5D measurement at a_k . Total costs, expressed in UK pounds, were collected for these time intervals also. Although there were no deaths, several patients failed to complete all follow-up visits and were considered censored.

Table 5.9 Treatment by type of hysterectomy, EVALUATE example

	Laparoscopic	Standard
Vaginal ($N = 487$)	324	163
Abdominal ($N = 859$)	573	286

Table 5.10 Parameter estimates for the abdominal patients in the EVALUATE trial using inverse probability weighting

	$T(n_T = 573)$	$S(n_S = 286)$		Equation
$\hat{\varphi}_j$	0.8703	0.8617	difference = $\hat{\Delta}_e$ = 0.009148	(3.35)
\hat{v}_j	1705.4	1519.6	difference = $\hat{\Delta}_c = 185.8$	(3.12, 3.13)
$\hat{V}(\hat{\varphi}_j)$	0.00003242	0.00007121	sum = $\hat{V}(\hat{\Delta}_e)$ = 0.0001036	(3.37)
$\hat{V}(\hat{v}_j)$	3245	7099	sum = $\hat{V}(\hat{\Delta}_c) = 10\,344$	(3.14)
$\hat{C}(\hat{\varphi}_j, \hat{v}_j)$	-0.05912	-0.1748	sum = $\hat{C}(\hat{\Delta}_e, \hat{\Delta}_c)$ = -0.2339	(3.39)

The estimators of the parameters of interest for the abdominal patients, using inverse probability weighting, are given in Table 5.10. T is observed to increase mean quality-adjusted survival and increase costs and, consequently, the point $(\hat{\Delta}_e, \hat{\Delta}_c)$ lies in the NE (trade-off) quadrant. For any value of λ , the quantities \hat{b}_λ and v_λ are given by $0.009148\lambda - 185.8$ and $0.0001036\lambda^2 + 10\,344 - 2 \times (-0.2339)\lambda$, respectively, and the 90% confidence limits are given by

$$0.009148\lambda - 185.8 \pm 1.645\sqrt{0.0001036\lambda^2 + 10\,344 + 2 \times 0.2339\lambda}$$

The quantity \hat{b}_λ and corresponding 90% confidence limits are plotted in Figure 5.8. The plot of \hat{b}_λ has slope 0.009148 ($= \Delta_e$), crosses the vertical axis at $-185.8 (= -\hat{\Delta}_c)$ and the horizontal axis at $185.8/0.009148 = 20\,310 (= ICER)$. By observing where the confidence limits cross the vertical axis, one can make inference regarding the difference between arms with respect to cost, in what is essentially a cost-minimization analysis. Since the confidence interval includes only negative values, the null hypothesis of no difference in mean cost can be rejected at the 10% level, and the investigators can conclude that there is evidence that T increases costs. (Recall that the confidence interval defined on the vertical axis is for $-\hat{\Delta}_c$, and since the

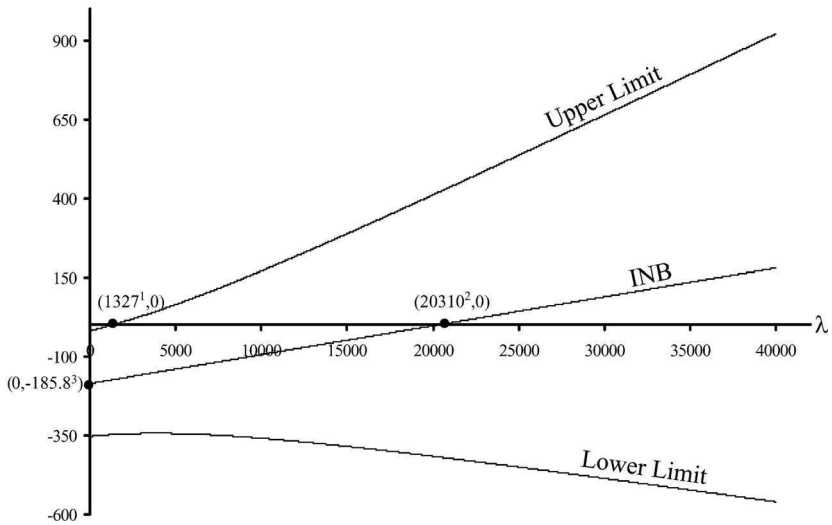


Figure 5.8 INB versus λ for the EVALUATE trial. ¹Lower confidence limit for ICER; ²ICER; ³ $-\hat{\Delta}_c$

confidence interval for $-\hat{\Delta}_c$ includes only negative values, the confidence interval for $\hat{\Delta}_c$ includes only positive values.)

The horizontal intercepts of the confidence limits define the Fieller confidence interval for the ICER and, by focusing attention on the horizontal axis, one can perform a cost-effectiveness analysis using an ICER approach. The 90% lower limit for the ICER is 1327. Since the lower limit of the INB fails to cross the horizontal axis, the upper limit of the ICER approaches $+\infty$, i.e. the confidence interval for the ICER includes the positive vertical axis of the cost-effectiveness plane. This is interpreted to mean that there is no value for the WTP that would make T cost-effective with 95% confidence. Similarly, since the lower limit of the INB fails to cross the horizontal axis, the confidence interval for INB always includes negative values, which leads to the same conclusion. The 90% confidence limits for the ICER can be calculated directly from applying Equation (4.4), whose associated elements are given in Table 5.11. The solution to Equation (4.4) yields the interval [1327, -25054]. Since the lower limit is positive and the upper limit is negative, the confidence region on the CE plane includes

Table 5.11 Element for calculation confidence limits for ICER for the abdominal patients in the EVALUATE trial

$\hat{R} = \hat{\Delta}_c / \hat{\Delta}_e = 185.8 / 0.009148 = 20\,310$
$a = \hat{V}(\hat{\Delta}_e) / \hat{\Delta}_e^2 = 0.0001036 / 0.009148^2 = 1.238$
$b = \hat{V}(\hat{\Delta}_c) / \hat{\Delta}_c^2 = 10\,344 / (185.8)^2 = 0.2996$
$c = \hat{C}(\hat{\Delta}_e, \hat{\Delta}_c) / (\hat{\Delta}_e \hat{\Delta}_c) = -0.2329 / (0.009148 \times 185.8) = -0.1370$
$z_{1-0.05} = 1.645$

part of the NW (lose–lose) quadrant, see Figure 5.9. Consequently, a proper representation of the interval is given by $[1327, +\infty)$, since the largest value in the interval approaches $+\infty$.

The cost-effectiveness acceptability curve can be calculated by

$$\begin{aligned}
 \mathcal{A}(\lambda) &= \Phi(\hat{b}_\lambda / \sqrt{v_\lambda}) \\
 &= \Phi\left((0.009148\lambda - 185.8) / \sqrt{0.0001036\lambda^2 + 10\,344 + 2 \times 0.2339\lambda}\right)
 \end{aligned}$$

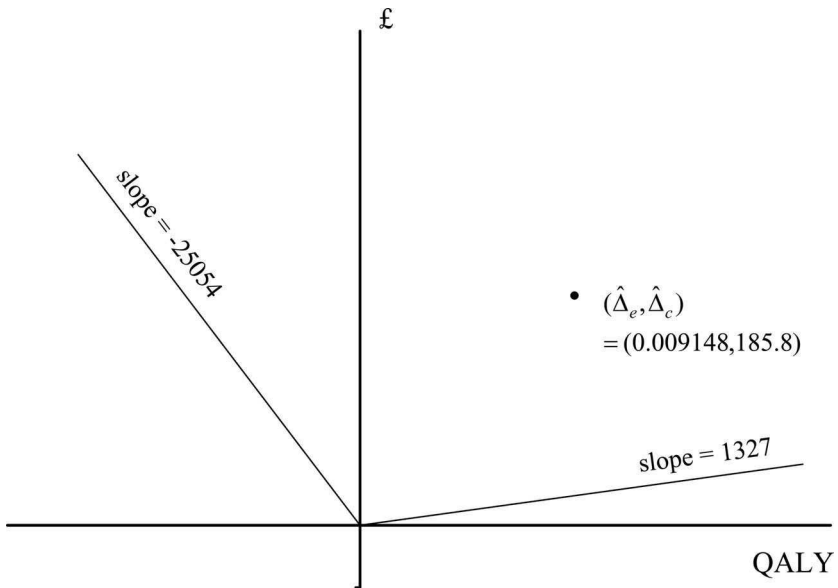


Figure 5.9 ICER confidence region on the CE plane for the EVALUATE trial

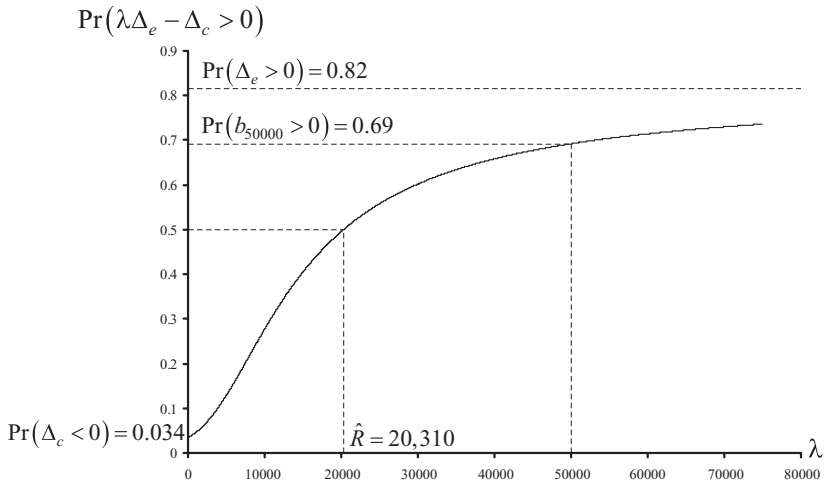


Figure 5.10 Cost-effectiveness acceptability curve for EVALUATE trial

The curve, see Figure 5.10, meets the vertical axis at $\Pr(\Delta_c < 0) = \Phi(-185.8/\sqrt{103\,444}) = 0.03386$ and asymptotically approaches $\Pr(\Delta_e > 0) = \Phi(0.009148/\sqrt{0.0001036}) = 0.8156$. Since $\mathcal{A}(50\,000) = 0.6922$, if the WTP for a year of life is CAD\$ 50 000, we could say that the probability that T is cost-effective is 69.22%.

5.6 BAYESIAN APPROACH APPLIED TO THE UK PDS STUDY

To illustrate the Bayesian approach to cost-effectiveness analysis, data from the UK prospective diabetes study (PDS) are employed. These data relate to a cost-effectiveness analysis designed to assess the efficiency of tight blood pressure control compared with less tight control in hypertensive patients with type 2 diabetes (Raikou *et al.*, 1998). Hypertension in subjects with type II diabetes is a risk factor for macrovascular complications. Although improved blood pressure control has been shown to reduce myocardial infarction and stroke in a diabetic sub-group of elderly patients with type II diabetes (SHEP Cooperative Research Group 1991) no information on younger patients or

the effect on complications of diabetes was available. The authors of the original cost-effectiveness analysis noted that, although cost-effectiveness of antihypertensive programs based on education and drugs have been reported for a number of populations: ‘... these analyses have mainly been based on models and lack information on effectiveness and use of resources from long term trials, and none has considered hypertensive patients with type II diabetes.’

The authors then go on to report the results of their study in the form of the incremental costs, incremental effects and employing the cost-effectiveness acceptability curve approach. These results, assuming the then standard discount rates of 6% for both costs and life-years, are reproduced in Table 5.12. Also included in Table 5.12 is an assessment of the net benefit of tight blood pressure control, assuming a value of λ of £20 000 per year of life gained (LYG). Although tight blood pressure control looks both more effective (in terms of LYG) and more costly, neither of these differences are statistically significant at conventional levels. Furthermore, the net benefit statistic shows a positive net benefit at $\lambda = £20\,000$, but is also

Table 5.12 Summary statistics for the cost-effectiveness of tight blood pressure control compared with less tight control in the UK prospective diabetes study

	Mean	SE	Lower 95% limit	Upper 95% limit
<i>Control Group</i>				
Effect (yr)	10.30	0.17	9.97	10.64
Cost (£)	6145	434	5294	6996
<i>Treatment Group</i>				
Effect (yr)	10.63	0.12	10.41	10.96
Cost (£)	6381	309	5775	6987
<i>Difference</i>				
Effect (yr)	0.33	0.21	-0.08	0.73
Cost (£)	236	533	-808	1280
ICER	720	N/A	N/A	N/A
INB*	6319	4169	-1853	14 490

*For $\lambda = £20\,000$ per year of life gained.

non-significant. A naïve economic analysis might conclude that there was nothing to choose between these two treatments, however, this would ignore the wealth of information built up in the UK PDS study over an 11-year median follow-up time that the effectiveness of treatment is approaching the standard level of statistical significance. Indeed, the clinical endpoint of time to a diabetes-related event was shown to have a statistically significant difference between the two treatments. Instead the authors presented the results of their analysis in terms of a cost-effectiveness acceptability curve showing, for example, that for $\lambda = \text{£}20\,000/\text{LYG}$, there is a 92% probability that the intervention is cost-effective.

Of course, such an interpretation is only possible using a Bayesian view of probability. The quote from the original paper above indicates that the authors of the original study had considered that there was little information on the cost-effectiveness of tight blood pressure control for diabetic patients prior to their reported results. However, a search of a database of cost-effectiveness analyses reporting cost per QALY and cost per life-year results (Briggs and Gray, 1999) for economic analyses of hypertension control conducted alongside a clinical trial identified an economic analysis of the Swedish Trial of Old patients with Hypertension (the STOP hypertension study; Dahlof *et al.*, 1991; Johannesson *et al.*, 1993). Clearly, the population studied (elderly Swedish patients without diabetes) differed from that of hypertensive patients in the UK PDS, but in the absence of other information, this trial might be seen as providing at least some information of the likely costs and effectiveness in the UK PDS population. Taking the average estimates of the life-years gained and costs reported in the economic analysis of the STOP study suggests that an estimated 0.16 life-years might be gained from treatment at an additional cost of approximately £1400 (converted from Swedish kronor and inflated to 1996 prices). Unfortunately, there was little information on sampling variation given in the economic analysis of the STOP study and standard errors on cost and effect differences were not given. Therefore, the standard errors were set arbitrarily to give a coefficient of variation on cost and effect equal to 2, in order to be conservative and to reflect the fact that the different methods employed in the economic analysis, the difference between the patient populations and the difference

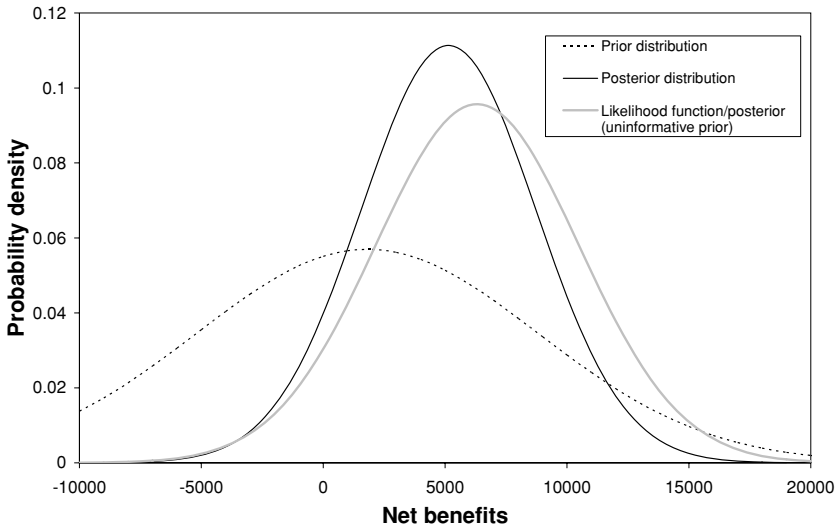


Figure 5.11 Bayesian approach to cost-effectiveness analysis of tight blood pressure control for hypertensive diabetic patients. Two posterior distributions for net benefits based on an informative prior distribution (shown) and an uninformative prior

between the two countries' health care systems will all increase the level of uncertainty associated with these prior estimates.

Using the net benefit approach, Figure 5.11 presents this prior information and the posterior distribution arising from employing this information together with the data from Table 5.12. This is the Bayesian approach with an informative prior. The alternative would be to employ an uninformative prior such that the posterior distribution produced is dominated by the observed data—either because no prior information is available, or because the analyst wishes to discard that information. The posterior distribution based on the uninformative prior is also shown in Figure 5.11, and exactly corresponds with the data likelihood. It is clear that incorporating the prior information reduces the variance of the posterior distribution, but that the point estimate of net benefit is weighted most heavily toward the UK PDS data.

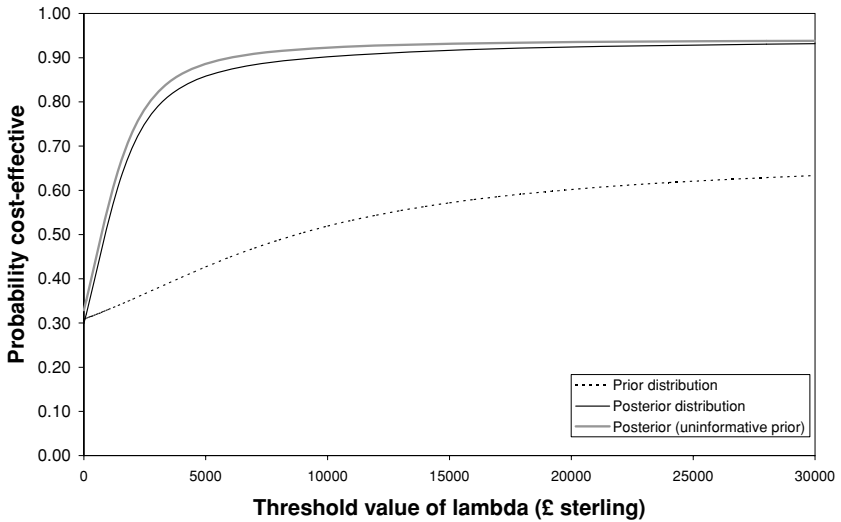


Figure 5.12 Bayesian approach to cost-effectiveness analysis of tight blood pressure control for hypertensive diabetic patients. Two posterior cost-effectiveness acceptability curves based on an informative prior distribution (shown) and an uninformative prior

Having estimated the prior and posterior distributions for net benefits using Bayesian methods, the probability of the intervention being cost-effective can then be plotted as a function of λ to generate cost-effectiveness acceptability curves and these are shown in Figure 5.12. Just as the Bayesian posterior distribution of net benefit with an uninformative prior corresponds to the data likelihood, so the cost-effectiveness acceptability curve plotted under the assumption of an uninformative prior will correspond exactly to a cost-effectiveness acceptability curve calculated by frequentist methods (involving the less natural interpretation of the curve based on P -values).

5.7 SUMMARY

This chapter has demonstrated different aspects of the material covered so far in the book using a range of different examples taken from

real-life analyses. The aim was to illustrate and emphasize the importance of the methods discussed to this point. Having covered the basic methods for estimating the five key parameters required to summarize cost-effectiveness data, the remaining chapters move on to more specialized topics: power and sample size, covariate adjustment and subgroup analyses, multicenter studies, and statistical modeling of trial data.

References

- Altman DG. (1985) Comparability of randomised groups. *Statistician*, **34**, 125–136.
- The ASSENT-3 Investigators. (2001) Efficacy and safety of tenecteplase in combination with enoxaparin, abciximab, or unfractionated heparin: the ASSENT-3 randomised trial in acute myocardial infarction. *Lancet*, **358**, 605–613.
- Bang H and Tsiatis AA. (2000) Estimating medical costs with censored data, *Biometrika*, **87**, 329–343.
- Bateman ED, Boushey HA, Bousquet J, Busse WW, Clark TJH, Pauwels RA, Pedersen SE and the GOAL Investigators Group (2004) Can guideline-defined asthma control be achieved?: The gaining optimal asthma control study". *American Journal of Respiratory and Critical Care Medicine*, **170**, 836–844.
- Black WC. (1990) The CE plane: a graphic representation of cost-effectiveness, *Medical Decision Making*, **10**, 212–214.
- Bloomfield DJ, Krahn MD, Neogi T, Panzarella T, Warde P, Willan AR, Ernst S, Moore MJ, Neville A, Tannock IF. (1998) Economic evaluation of chemotherapy with mitoxantrone plus prednisone for symptomatic hormone-resistant prostate cancer: Based on a Canadian trial with palliative endpoints, *Journal of Clinical Oncology*, **16**, 2272–2279.
- Briggs AH. (1999) A Bayesian approach to stochastic cost-effectiveness analysis, *Health Economics*, **8**, 257–261.
- Briggs A, Bousquet J, Wallace M, Busse WW, Clark TJH, Pederson SE, Bateman E. (2006) Cost effectiveness of asthma control: an economic appraisal of the GOAL study. *Allergy*, **61**, 531–536.
- Briggs AH and Fenn P. (1998) Confidence intervals or surfaces? Uncertainty on the cost-effectiveness plane, *Health Economics*, **7**, 723–740.

184 *References*

- Briggs AH and Gray AM. (1998a) The Distribution of Health Care Costs and Their Statistical Analysis for Economic Evaluation, *Journal of Health Services Research and Policy*, **3**, 233–245.
- Briggs AH and Gray AM. (1998b) Power and sample size calculations for stochastic cost-effectiveness analysis. *Medical Decision Making*, **18**, S81–S92.
- Briggs AH and Gray AM. (1999) Handling uncertainty when performing economic evaluation of health care interventions, *Health Technology Assessment*, vol. 3, no. 2.
- Briggs AH, Mooney CZ and Wonderling DE. (1999) Constructing confidence intervals for cost-effectiveness ratios: an evaluation of parametric and non-parametric techniques using monte carlo simulation, *Statistics in Medicine*, **18**, 3245–3262.
- Briggs A, Nixon R, Dixon S and Thompson S. (2005) Parametric modelling of cost data: some simulation evidence, *Health Economics*, **14**, 421–428.
- Briggs AH, Sculpher MJ and Buxton MJ. (1994) Uncertainty in the economic evaluation of health care technologies: the role of sensitivity analysis, *Health Economics*, **3**, 95–104.
- Briggs AH, Wonderling DE and Mooney CZ. (1997) Pulling cost-effectiveness analysis up by its bootstraps; a non-parametric approach to confidence interval estimation, *Health Economics*, **6**, 327–340.
- Brown HK and Kempton RA. (1994) The application of REML in clinical trials. *Statistics in Medicine*, **13**, 1601–1617.
- Chaudhary MA and Stearns SC. (1996) Confidence intervals for cost-effectiveness ratios: An example from a randomized trial, *Statistics in Medicine*, **15**, 1447–1458.
- Chiba N, van Zanten SJ, Sinclair P, Ferguson RA, Escobedo S, Grace E. (2002) Treating *Helicobacter pylori* infection in primary care patients with uninvestigated dyspepsia: the Canadian adult dyspepsia empiric treatment-*Helicobacter pylori* positive (CADET-Hp) randomised controlled trial. *BMJ*, **324**, 1012–1016.
- Clarke PM, Gray AM, Briggs, A, Farmer AJ, Fenn, P, Stevens RJ, Matthews DR, Stratton IM, and Holman, R. (2004) A model to estimate the lifetime health outcomes of patients with type 2 diabetes: the United Kingdom Prospective Diabetes Study (UKPDS) Outcomes Model (UKPDS no. 68), *Diabetologia*, **47**, 1747–1759.
- Claxton K. The irrelevance of inference: a decision-making approach to the stochastic evaluation of health care technologies. (1999) *Journal of Health Economics*, **18**, 341–364.

- Claxton K and Posnett J. (1996) An economic approach to clinical trial design and research priority setting. *Health Economics*, **5**, 513–524.
- Cochran WG. (1977) *Sampling techniques*. John Wiley & Sons, New York.
- Collins R, Gray R, Godwin J, Peto R. (1987) Avoidance of large biases and large random errors in the assessment of moderate treatment effects: the need for systematic overviews. *Statistics in Medicine*, **6**, 245–254.
- Commonwealth of Australia. (1990) *Guidelines for the pharmaceutical industry on preparation of submissions to the Pharmaceutical Benefits Advisory Committee: including submissions involving economic analyses*. Woden (ACT) Department of Health, Housing and Community Services.
- Connolly SJ, Gent M, Roberts RS, Dorian P, Roy D, Sheldon RS, Mitchell LB, Green MS, Klein GJ and O'Brien B. (2000) Canadian implantable defibrillator study (CIDS): a randomized trial of the implantable cardioverter defibrillator against amiodarone. *Circulation*, **101**, 1297–1302.
- Cook JR, Drummond M, Glick H and Heyse JF. (2003) Assessing the appropriateness of combining economic data from multinational clinical trials. *Statistics in Medicine*, **22**, 1955–1976.
- Dahlof B, Lindholm LH, Hansso L, Scherste, B, Ekblom T, and Wester PO. (1991), Morbidity and mortality in the Swedish Trial in Old Patients with Hypertension (STOP-Hypertension), *Lancet*, **338**, 1281–1285.
- Dersimonian R and Laird N. (1986) Meta-analysis in clinical trials. *Controlled Clinical Trials*, **7**, 177–188.
- Detsky AS. (1993) A guideline for preparation of economic analysis of pharmaceutical products: a draft document for Ontario and Canada. *Pharmacoeconomics* **3**, 354–361.
- Drummond ME, O'Brien BJ, Stoddart GL and Torrance GW. (1997) *Methods for the Economic Evaluation of Health Care Programmes*, 2nd ed. Oxford: Oxford University Press.
- Duan N. (1983) Smearing Estimate: A Nonparametric Retransformation Method. *Journal of the American Statistical Association*, **78**, 605–610.
- Efron B. (1987) Better bootstrap confidence intervals, *Journal of the American Statistical Association*, **82**, 171–200.
- Efron B and Morris C. (1972) Empirical Bayes on vector observations: An extension of Stein's method. *Biometrika*, **59**, 335–347.
- Efron B and Tibshirani R. (1993) *An Introduction to the Bootstrap*, Chapman and Hall, New York.
- Fenwick E, O'Brien BJ, Briggs AH. (2004) Cost-effectiveness acceptability curves – facts, fallacies and frequently asked questions, *Health Economics*, **13**, 405–415.

186 *References*

- Gail M and Simon R. (1985) Testing for Qualitative Interactions between Treatment Effects and Patient Subsets, *Biometrics*, **41**, 361–372
- Gardiner JC, Huebner M, Jetton J and Bradley CJ. (2000) Power and sample size assessments for tests of hypotheses on cost-effectiveness ratios. *Health Economics*, **9**, 227–234.
- Garry R, Fountain J, Mason S, et al. (2004) The EVALUATE study: two parallel randomised trials, one comparing laparoscopic with abdominal hysterectomy, the other comparing laparoscopic with vaginal hysterectomy. *BMJ* 2004, **328**, 129–133.
- Glasziou PP and Irwig LM. (1995) An evidence based approach to individualising treatment, *British Medical Journal*, **311**, 7016, 1356–1359.
- Global Initiative for Asthma (GINA) (1998) *Pocket guide for asthma management and prevention*, National Institutes for Health, National Heart, Lung and Blood Institute, Bethesda, Publication No. 95-3659-B.
- Greene WH. (1993) *Econometric Analysis* (2nd Ed). MacMillan: New York; 486–489.
- Grieve R, Nixon RM, Thompson SG and Normand C. (2005) Using multilevel models for assessing the variability of multinational resource use and cost data. *Health Economics*, **14**, 185–196.
- Grundy PM, Healy MJR and Rees DH. (1956) Economic choice of the amount of experimentation. *Journal of the Royal Statistical Society, Series A*, **18**, 32–48.
- Hanley N, Ryan M, and Wright R. (2003) Estimating the monetary value of health care: lessons from environmental economics, *Health Economics*, **12**, 3–16.
- Hannah ME, Hannah WJ, Hewson SH, Hodnett ED, Saigal S and Willan AR. (2000) Term Breech Trial: a multicentre international randomised controlled trial of planned caesarean section and planned vaginal birth for breech presentation at term. *The Lancet*, **356**, 1375–1383.
- Hanemann WM. (1991) Willingness to pay and willingness to accept: How much can they differ? *The American Economic Review*, **18**, 635–647.
- Hardy RJ and Thompson SG. (1998) Detecting and describing heterogeneity in meta-analysis. *Statistics in Medicine*, **17**, 841–885.
- Heitjan DF (2000), Fieller's method and net health benefits. *Health Economics*, **9**, 327–335.
- Hoch JS, Briggs AH, 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.

- Hutton EK, Kaufman K, Hodnett E, Amankwah K, Hewson SA, McKay D, Szalai JP and Hannah ME. (2003) External cephalic version beginning at 34 weeks' gestation versus 37 weeks' gestation: a randomized multicenter trial. *American Journal of Obstetrics and Gynecology*, **189**, 245–254.
- Johannesson M, Dahlof B, Lindholm LH, Ekblom T, Hansson L, Oden A, Schersten B, Wester PO, Jonsson B. The cost-effectiveness of treating hypertension in elderly people—an analysis of the Swedish Trial in Old Patients with Hypertension (STOP Hypertension), *Journal of Internal Medicine*, **234**, 317–323.
- Johansson PO. (1995) *Evaluating Health Risks: An Economic Perspective*. Cambridge University Press: Cambridge; 38.
- Johnson FR, Desvousges WJ, Ruby MC, Stieb D and De Civita P. (1998) Eliciting stated health preferences: an application to willingness to pay for longevity, *Medical Decision Making*, **18 Suppl**, S57–S67.
- Jonsson B, Johannesson M, Kjekshus J, Olsson AG, Pedersen TR, and Wedel H. (1996) Cost-effectiveness of cholesterol lowering. Results from the Scandinavian Simvastatin Survival Study (4S), *European Heart Journal*, **17**, 1001–1007.
- Kahneman D and Tversky A. (1979) Prospect theory: An analysis of decision under risk. *Econometrica*, **47**, 263–291.
- Kind P. (1996) The EuroQol instrument: an index of health-related quality of life. In *Quality of Life and Pharmacoeconomics in Clinical Trials, 2nd edition*, Spilker B (ed). Lippincott-Raven: Philadelphia, 191–201.
- Koopmanschap MA, Touw KCR and Rutten FFH. (2001) Analysis of cost and cost-effectiveness in multinational trials. *Health Policy*, **58**, 175–186.
- Kuntz KM and Weinstein MC. (2001) Modeling in economic evaluation. In *Economic Evaluation in Health Care: Merging Theory with Practice*, MF Drummond & A McGuire, (eds), Oxford University Press: Oxford, 141–171.
- Lachin JM. (1981) Introduction to sample size determination and power analysis for clinical trials, *Controlled Clinical Trials*, **2**, 93–113.
- Laska EM, Meisner M and Siegel C. (1997) Statistical inference for cost-effectiveness ratios, *Health Economics*, **6**, 229–242.
- Laupacis A, Feeny D, Detsky A and Tugwell P. (1992) How attractive does a new technology have to be to warrant adoption and utilization? Tentative guidelines for using clinical and economic evaluations, *Canadian Medical Association Journal*, **146**, 472–481.
- Leaf A. (1989) Cost-effectiveness as a criterion for Medicare coverage, *New England Journal of Medicine*, **321**, 898–900.

188 *References*

- Lehman AF, Dixon LB, Kernan E *et al.* (1997) A randomized trial of assertive community treatment for homeless persons with severe mental illness, *Archives of General Psychiatry*, **54**, 1038–1043.
- Lehman AF, Dixon LB, Hoch JS *et al.* (1999) Cost-effectiveness of assertive community treatment for homeless persons with severe mental illness, *British Journal of Psychiatry*, **174**, 346–352.
- Lemeshow S, Hosmer Jr. DW, Klar J and Lwanga SK. (1990) The adequacy of sample size in health studies. John Wiley and Sons, New York.
- Lin DY, Feuer EJ, Etzioni R and Wax Y. (1997) Estimating medical costs from incomplete follow-up data, *Biometrics*, **53**, 419–434.
- Lin DY. (2000) Linear regression analysis of censored medical costs, *Biostatistics*, **1**, 35–47.
- Lumley T, Diehr P, Emerson S and Chen L. (2002) The importance of the normality assumption in large public health data sets, *Annual Review of Public Health*, **23**, 151–169.
- Manca A, Hawkins N, Sculpher MJ. (2005a) Estimating mean QALYs in trial-based cost-effectiveness analysis: the importance of controlling for baseline utility. *Health Economics*, **14**, 487–496.
- Manca A, Rice N, Sculpher MJ and Briggs AH. (2005b). Assessing generalisability by location in trial-based cost-effectiveness analysis: the use of multilevel models. *Health Economics*, **14**, 471–485.
- Manning WG and Mullahy J. (2001) Estimating log models: to transform or not to transform? *Journal of Health Economics*, **20**, 461–494.
- McAlister FA. (2002) Commentary: relative treatment effects are consistent across the spectrum of underlying risks . . . usually, *International Journal of Epidemiology*, **31**, 76–77.
- Mooney CZ and Duval RD. (1993) *Bootstrapping: A Nonparametric Approach to Statistical Inference*. Sage, Newbury Park, CA.
- Morrison GC. (1998) Understanding the disparity between WTP and WTA: endowment effect, substitutability, or imprecise preferences? *Economics Letters* **59**, 189.
- National Institute for Clinical Excellence (NICE) 2004, *Guide to the methods of technology assessment* NICE, London.
- Nixon RM and Thompson SG. (2005) Incorporating covariate adjustment, subgroup analysis and between-centre differences into cost-effectiveness evaluations. *Health Economics*, **14**, 1217–1229.
- O'Brien BJ. (1996) Economic evaluation of pharmaceuticals: Frankenstein's monster of vampire of trials? *Medical Care*, **34**, DS99–DS108.
- O'Brien BJ. (1997) A tale of two (or more) cities: geographic transferability of pharmaco-economic data. *American Journal of Managed Care*, **3**, S33–S40.

- O'Brien BJ, Connolly SJ, Goeree R, Blackhouse G, Willan AR, Yee R, Roberts RS and Gent M. (2000) Cost-effectiveness of the implantable cardioverter defibrillator results from the Canadian implantable defibrillator study (CIDS). *Circulation*, **103**, 1416–1421.
- O'Brien BJ, Drummond MF, Labelle RJ and Willan AR. (1994) In search of power and significance: Issues in the design and analysis of stochastic cost-effectiveness studies in health care, *Medical Care*, **32**, 150–163.
- O'Brien BJ and Gafni A. (1996) When do the dollars make sense? Toward a conceptual framework for contingent valuation studies in health care, *Medical Decision Making*, **16**, 288–299.
- O'Brien BJ, Gertsen K, Willan AR and Faulkner L. (2002) Is there a kink in consumers' threshold value for cost-effectiveness in health care? *Health Economics*, **11**, 175–180.
- O'Hagan A and Stevens JW. (2001) Bayesian assessment of sample size for clinical trials of cost-effectiveness. *Medical Decision Making*, **21**, 219–230.
- O'Hagan A and Stevens JW. (2003) Assessing and comparing costs: how robust are the bootstrap and methods based on asymptotic normality? *Health Economics*, **12**, 33–49.
- Oxman AD, Guyatt GH, (1992) A consumer's guide to subgroup analyses. *Annals of Internal Medicine*, **116**, 78–84.
- Pauly MV. (1995) Valuing health care benefits in money terms. In Sloan FA (ed). *Valuing Health Care*, Cambridge University Press, Cambridge.
- Pezeshk H. (2003) Bayesian techniques for sample size determination in clinical trials: a short review. *Statistical Methodology in Medical Research* 2003, **12**, 489–504.
- Pinto EM, Willan AR and O'Brien BJ. (2005) Cost-effectiveness analysis for multinational clinical trials. *Statistics in Medicine*, **24**, 1965–1982.
- Pocock SJ. (1984) *Clinical Trials: A Practical Approach*. John Wiley & Sons Ltd., Chichester.
- Raiffa H and Schlaifer R. (2000) *Applied Statistical Decision Theory*. John Wiley and Sons, New York.
- Raikou M, Gray A, Briggs A, Steven, R, Cull C, McGuire A, Fenn P, Stratton I, Holman, R, and Turner R. (1998) Cost effectiveness analysis of improved blood pressure control in hypertensive patients with type 2 diabetic patients (HDS7): UKPDS 40, *British Medical Journal*, **317**, 720–726.
- Reed SD, Anstrom KJ, Bakhai A, Briggs AH, Califf RM, Cohen DJ, Drummond MF, Glick HA, Gnanasakthy A, Hlatky MA, O'Brien BJ, Torti FM, Tsiatas AA, Willan AR, Mark DB and Schulman KA. (2005) Conducting economic

190 *References*

- evaluations alongside multinational clinical trials: Toward a research consensus. *American Heart Journal*, **149**, 434–443.
- Scandinavian Simvastatin Survival Study Group (1994) Randomised trial of cholesterol lowering in 4444 patients with coronary heart disease: the Scandinavian Simvastatin Survival Study (4S). *Lancet*, **344**, 1383–1389.
- Sculpher M, Manca A, Abbott J, Fountain J, Mason S and Garry R. (2003) The cost-effectiveness of laparoscopic-assisted hysterectomy in comparison of standard hysterectomy: the EVALUATE trial. *BMJ*, 2004, **328**, 134–137.
- Senn S. (1997) *Statistical Issues in Drug Development*. John Wiley & Sons, Chichester.
- Skene AM and Wakefield JC. (1990) Hierarchical models for multicentre binary response studies. *Statistics in Medicine*, **9**, 919–929.
- Spiegelhalter, DJ, Thomas, A and Best, NG (1999). *WinBUGS Version 1.2 User Manual*. Cambridge: MRC Biostatistics Unit.
- Sutton AJ, Cooper NJ, Abrams KR, Lambert PC and Jones DR. (2005) A Bayesian approach to evaluating net clinical benefit allowed for parameter uncertainty, *Journal of Clinical Epidemiology*, **58**, 26–40.
- Tannock IF, Osoba D, Stockler MR, Ernst DS, Neville AJ, Moore MJ, Armitage GR, Wilson JJ, Venner PM, Coppin CM, Murphy KC. (1996) Chemotherapy with mitoxantrone plus prednisone or prednisone alone for symptomatic hormone-resistant prostate cancer: A Canadian trial with palliative endpoints. *Journal of Clinical Oncology*, **14**, 1756–1764.
- Thompson SG and Barber JA. (2000) How should cost data in pragmatic randomised trials be analysed? *BMJ*, **320**, 1197–1200.
- Torrance GW. (1976) Health status index models: a unified mathematical view, *Management Science*, **22**, 990–1001.
- Torrance GW. (1986) Measurement of health state utilities for economic appraisals: a review, *Journal of Health Economics*, **5**, 1–30.
- Torrance GW. (1987) Utility approach to measuring health related quality of life, *Journal of Chronic Disease*, **40**, 593–600.
- van Hout BA, Al MJ, Gordon GS, and Rutten FF. (1994) Costs, effects and C/E ratios alongside a clinical trial, *Health Economics*, **3**, 309–319.
- Weinstein MC and Stason WB. (1977) Foundation of cost-effective analysis for health and medical practices, *New England Journal of Medicine*, **296**, 716–721.
- Willke RJ, Glick HA, Polsky D and Schulman K. (1998) Estimating country-specific cost-effectiveness from multinational clinical trials. *Health Economics*, **7**, 481–493.

- Willan AR. (1994) Power function arguments in support of an alternative approach for analyzing management trials. *Controlled Clinical Trials*, **15**, 211–219.
- Willan AR. (2004) Incremental net benefit in the analysis of economic data from clinical trials with application to the CADET-Hp Trial. *European Journal of Gastroenterology and Hepatology*, **16**, 543–549.
- Willan AR, Briggs AH and Hoch JS. (2004) Regression methods for covariate adjustment and subgroup analysis for non-censored cost-effectiveness data. *Health Economics*, **13**, 461–475.
- Willan AR, Chen EB, Cook RJ and Lin DY. (2003) Incremental net benefit in randomized clinical trials with qualify-adjusted survival. *Statistics in Medicine*, **22**, 353–362.
- Willan AR, Cruess AF and Ballantyne M. (1996) Argon green vs krypton red laser photocoagulation of extrafoveal choroidal neovascular lesions: Three-year results in age-related macular generation. *Canadian Journal Ophthalmology*, **31**, 11–17.
- Willan AR and Lin DY. (2001) Incremental net benefit in randomized clinical trials. *Statistics in Medicine*, **20**, 1563–1574.
- Willan AR, Lin DY, Cook RJ and Chen EB. (2002) Using inverse-weighting in cost-effectiveness analysis with censored data. *Statistical Methods in Medical Research*, **11**, 539–551.
- Willan AR, Lin DY and Manca A. (2005) Regression methods for cost-effectiveness analysis with censored data. *Statistics in Medicine*, **24**, 131–145.
- Willan AR and O'Brien BJ. (1996) Confidence intervals for cost-effectiveness ratios: An application of Fieller's theorem. *Health Economics*, **5**, 297–305.
- Willan AR and O'Brien BJ. (1999) Sample size and power issues in estimating incremental cost-effectiveness ratios from clinical trials data. *Health Economics*, **8**, 203–211.
- Willan AR and O'Brien BJ. (2001) Cost prediction models for the comparison of two groups. *Health Economics*, **10**, 363–366.
- Willan AR, O'Brien BJ and Leyva RA. (2001) Cost-effectiveness analysis when the WTA is greater than the WTP. *Statistics in Medicine*, **20**, 3251–3259.
- Willan AR, Pinto EM. (2005) The value of information and optimal clinical trial design. *Statistics in Medicine*, **24**, 1791–1806.
- Willan AR, Pinto EM, O'Brien BJ, Kaul P, Goeree R, Lynd L and Armstrong PW. (2005) Country-specific cost comparisons from multinational clinical trials using empirical Bayesian shrinkage estimation: The Canadian ASSENT-3 Economic Analysis. *Health Economics*, **14**, 327–338.

192 *References*

- Zethraeus N and Lothgren M. (2000) *On the equivalence of the net benefit and the Fieller's methods for statistical inference in cost-effectiveness analysis*. Working paper no. 379, SSE/EFI Working Series Papers in Economics and Finance, Stockholm School of Economics.
- Zhao H and Tian L. (2001) On estimating medical cost and incremental cost-effectiveness ratios with censored data, *Biometrics*, **57**, 1002–1008.

Author Index

- Altman DG 15, 25, 117, 143
- Bang H 31, 34
- Bateman ED 167, 168
- Best NG 164
- Black WC 5
- Bloomfield 72
- Briggs AH 1, 11, 13, 14, 27, 43, 45, 46, 48, 52, 54, 56, 67, 88, 93, 94, 95, 96, 117, 119, 142, 145, 165, 167
- Brown HK 146, 162
- Chaudhary MA 45, 46, 54
- Chiba PM 68
- Clarke PM 181
- Claxton K 104, 109, 110
- Cochran WG 13, 48
- Collins R 143
- Connolly SJ 77
- Cook JR 146, 152, 153
- Dahlof B 88
- Dersimonian R 155
- Detsky AS 1
- Drummond MF 5
- Duan N 16
- Efron B 55, 56, 156
- Fenn P 46, 52, 54, 96
- Fenwick E 54
- Gafni A 7
- Gail M 152
- Gardiner JC 94, 101
- Garry R 82
- Glasziou PP 166
- Greene WH 177, 119, 172
- Grieve R 146, 162
- Gray AM 13, 88, 94, 95, 96
- Grundy PM 109
- Guyatt GH 143
- Hanemann WM 60
- Hanley N 7
- Hannah ME 108
- Hardy RJ 157
- Heitjan DF 51
- Hoch JS 14, 117, 119, 142
- Hutton EK 114
- Johannesson M 88
- Johansson PO 60
- Johnson FR 7
- Jonsson B 152
- Kahneman D 60
- Kind P 80, 172
- Koopmanschap MA 146
- Kuntz KM 165
- Lachin JM 104
- Laska M 46

194 *Author Index*

- Laupacis A 100
Leaf A 1
Lehman AF 121
Lemeshow S 104
Lin DY 9, 28, 29, 31, 34, 42, 118, 130,
132, 134
Lothgren M 51
Lumley T 14, 48

Manca A 25, 118, 130, 134, 143, 146,
162
Manning WG 13
McAlister FA 148
Mooney CZ 55, 56
Morrison GC 60
Mullahy J 13

Nixon RM 13, 146, 162, 163

O'Brien BJ 1, 7, 14, 20, 45, 46, 48, 60,
62, 77, 94, 95, 96, 98, 103, 146,
147, 156
O'Hagan A 13, 106, 107
Oxman AD 143

Pauly MV 7
Pezeshk H 109
Pinto EM 109, 110, 114, 146, 156,
157, 160
Pocock SJ 115, 117, 143

Raiffa H 109

Raikou M 86
Reed D 149, 150, 164

Schlaifer R 109
Sculpher M 82
Simon R 152
Skene AM 146, 162
Spiegelhalter DJ 164
Stevens JW 13, 106, 107
Sutton AJ 166

Tannock IF 72
Thomas A 164
Thompson SG 13, 48, 146, 157, 162,
163
Tian L 31, 34
Tibshirani R 55, 56
Torrance GW 4, 22
Tsiatis AA 31, 34

van Hout BA 52

Weinstein MC 22, 165, 166
Willke RJ 146, 151
Willan AR 1, 11, 14, 20, 27, 31, 34,
35, 37, 38, 40, 41, 43, 46, 48, 62,
67, 68, 93, 94, 98, 103, 104, 108,
109, 110, 114, 117, 118, 119, 130,
134, 142, 145, 146, 156, 165

Zethraeus N 51
Zhao H 31, 34

Subject Index

- absolute risk reduction 3
- aggregate level analysis 146, 154, 156, 157
- area under the curve 23, 36, 39

- Bayesian methods 9, 43, 52, 57, 67, 76, 86, 93, 106, 110
- Bayesian sample size 106
- bias 2, 16, 25, 27, 39, 56
- bias corrected 56
- bootstrap 43, 52, 54, 58, 177

- censored data 27, 67, 129
- completely at random 29
- components of cost-effectiveness 165
- confidence box 94
- confidence intervals 2, 44, 46
- consistent estimation 30, 32, 34, 42, 44, 119
- cost histories 28, 29, 31, 131
- cost minimization 69, 102
- cost prediction models 19
- cost-effectiveness acceptability curve 9, 43, 52, 67, 72, 76, 81, 86, 90, 162, 163
- cost-effectiveness plane 5, 6, 9, 43, 45, 60, 94, 95
- covariate adjustment 10, 117, 147
- covariates 14, 18, 24, 117, 163

- decision-analytic modelling 165
- direct (Lin) method 9, 28, 29
- duration of interest 3, 11

- EM algorithm 156
- empirical Bayes 156
- EQ5D 170
- excess zeros 12, 18
- expected value of perfect information 110–112
- expected value of sample information 109, 112, 113

- Fieller's theorem 46, 51, 54

- generalized least squares 120, 125
- generalized linear models 17

- health related quality of life 176
- heterogeneity 143, 152
- hierarchical modeling 154, 162

- incremental cost-effectiveness ratio 5, 44
- incremental net benefit 2, 7, 8, 43, 49, 57, 67
- informative censoring 27, 28
- informative priors 89, 108
- inverse probability weighting 9, 28, 31, 37, 39, 41, 78, 83, 129

196 *Subject Index*

- kinked thresholds 69
- life-table methods 9, 28, 39, 78
- lost to follow-up 11, 27

- mean quality-adjusted survival time 4, 12, 22, 39, 133
- mean survival time 3, 12, 21, 36, 134
- multicenter trials 145
- multinational trials 145, 180
- multinomial regression 172

- Newton-Raphson algorithm 156
- non-censored data 11, 67, 117
- non-linear 16
- number-needed-to-treat 3

- opportunity loss 109–115
- optimal sample size 112–115
- ordinary least squares 17, 119, 125

- parameters of interest 2, 67
- posterior distributions 52, 53, 57, 59, 62, 65, 89, 90, 106, 107, 111, 112, 156
- power 93
- price weights 1, 4, 158
- probability of surviving 21, 30, 93
- product limit method 30, 33, 132

- QALY 6, 12, 20, 22, 23, 39, 88
- quality of life 4, 22, 72, 132, 148, 166, 176

- random effects models 146, 154, 160

- regression models 10, 14, 17, 19, 20, 24, 117, 151, 170
- restricted maximum likelihood 156
- restricted mean survival time 3

- sample size 93
- seemingly unrelated regression 117, 119, 125
- sensitivity analysis 1, 8, 50, 60
- shrinkage estimator 156, 157, 160
- skewness 12, 13, 14, 15, 48, 147, 163
- smearing 16
- statistical modeling 165
- structural assumptions 179
- sub-group analysis 86, 117, 167, 181
- survival function 3, 4, 30, 31, 34, 36, 181

- Taylor series 16, 45, 48
- threshold WTP 6, 7, 43, 44, 52, 60, 76, 94, 100
- trade-off quadrants 5, 48
- transformations 12, 15, 152
- two-part models 18
- type I error 93, 104, 108, 114, 146
- type II error 86, 104, 109

- uninformative priors 52, 57, 59, 89
- utility 4, 12, 20, 110, 169, 176

- value of information 93, 116

- willingness-to-accept 60
- willingness-to-pay 7, 52, 67, 124

Statistics in Practice

Human and Biological Sciences

Berger – Selection Bias and Covariate Imbalances in Randomized Clinical Trials
Brown and Prescott – Applied Mixed Models in Medicine
Chevret(Ed) – Statistical Methods for Dose-Finding Experiments
Ellenberg, Fleming and DeMets – Data Monitoring Committees in Clinical Trials: A Practical Perspective
Lawson, Browne and Vidal Rodeiro – Disease Mapping with WinBUGS and MLwiN
Lui – Statistical Estimation of Epidemiological Risk
*Marubini and Valsecchi – Analysing Survival Data from Clinical Trials and Observation Studies
Parmigiani – Modeling in Medical Decision Making: A Bayesian Approach
Senn – Cross-over Trials in Clinical Research, Second Edition
Senn – Statistical Issues in Drug Development
Spiegelhalter, Abrams and Myles – Bayesian Approaches to Clinical Trials and Health-Care Evaluation
Whitehead – Design and Analysis of Sequential Clinical Trials, Revised Second Edition
Whitehead – Meta-Analysis of Controlled Clinical Trials
Willan and Briggs – Statistical Analysis of Cost-effectiveness Data

Earth and Environmental Sciences

Buck, Cavanagh and Litton – Bayesian Approach to Interpreting Archaeological Data
Glasbey and Horgan – Image Analysis in the Biological Sciences
Helsel – Nondetects and Data Analysis: Statistics for Censored Environmental Data
McBride – Using Statistical Methods for Water Quality Management
Webster and Oliver – Geostatistics for Environmental Scientists

Industry, Commerce and Finance

Aitken and Taroni – Statistics and the Evaluation of Evidence for Forensic Scientists, Second Edition

Balding – Weight-of-evidence for Forensic DNA Profiles

Lehtonen and Pahkinen – Practical Methods for Design and Analysis of Complex Surveys, Second Edition

Ohser and Mücklich – Statistical Analysis of Microstructures in Materials Science

Taroni, Aitken, Garbolino and Biedermann – Bayesian Networks and Probabilistic Inference in Forensic Science

*Now available in paperback