MEASURING UNCERTAINTY IN ECONOMIC EVALUATIONS: A CASE STUDY IN LIVER TRANSPLANTATION

A thesis submitted for the degree of
Doctor of Philosophy

by
Tracey Anne Young

Health Economics Research Group (HERG)
Brunel University

May 2007

Abstract

It is important to account for all sources of uncertainty when evaluating the clinical or cost-effectiveness of health care technologies. Therefore, this thesis takes as its basis a cost-effectiveness study in liver transplantation and identifies two previously unexplored issues that can arise in clinical and cost-effectiveness studies. A literature review of studies evaluating the effectiveness, costs or cost-effectiveness of solid organ transplantation confirmed that these issues were important and relevant to other transplantation studies.

The first issue concerns the selection of an appropriate method for estimating mean study costs in the presence of incomplete (censored) data. Twelve techniques were identified and their accuracy was compared across artificially created mechanisms and levels of censoring. Lin's method with known cost histories and short interval lengths is recommended for accurately estimating mean costs and their uncertainty. It is assumed that these findings are generalisable to any solid organ transplant study where censoring is an issue.

The second issue explored in this thesis relates to methods for measuring uncertainty around survival, HRQL and cost estimates derived from prognostic models in the absence of observed data. Probabilistic sensitivity analysis is recommended for measuring prognostic model parameter uncertainty and estimating individual patient outcomes and their uncertainties, as it is able to incorporate the additional uncertainty from using prognostic models to estimate control group outcomes.

This thesis shows the quantitative importance of these issues and the methodological guidance offered should enable decision makers to have more confidence in clinical and cost-effectiveness estimates. Providing decision makers with a fuller estimate of the uncertainty around clinical and cost effectiveness estimates will aid them in decisions about the necessity of conducting further research in to the clinical or cost-effectiveness of health care technologies.

Publications and Authorship

The research presented in Chapter 2 denote the starting point for the work presented in this thesis and the results in Chapter 2 are the result of a collaborative effort on the part of a multi-disciplinary research team [Longworth *et al,* 2002; Longworth *et al*, 2003]. I played a substantive role in this research team, performing the statistical analysis and contributing to the cost-effectiveness analysis, writing and dissemination of the study results. Whilst acknowledging the contribution of my colleagues, it should be emphasised that for all of the research presented in the thesis, I have been the lead researcher, designing the main component studies and undertaking all analyses.

The research presented in Chapter 4 of this thesis has been presented in *Pharmacoeconomics* [Young, 2005]. The results in Chapter 4 differ from those in the paper as they contain additional methods for estimating costs in the presence of censoring, they compare methods for different levels of censoring (e.g.10%, 30% and 50%) and compare methods for differing interval lengths (dividing the study in to alternative interval lengths). This resulted in slightly different conclusions to those published in the paper.

The research presented in Chapter 5 of this thesis has been published in the *International Journal of Technology Assessment in Health Care* [Young & Thompson, 2004].

Table of Contents

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

List of Abbreviations

ALD	Alcoholic liver disease
CABG	Coronary Artery Bypass Graft
CADTH	Canadian Agency for Drugs and Technologies in Health
CELT	Cost-effectiveness in liver transplantation
CI	Confidence intervals
CINHAL	Cumulative Index to Nursing and Allied Health Limited
DEAL	Declining exponential function
DF	Degrees of Freedom
DOH	Department of Health
ELTR	European Liver Transplant Registry
HRQL	Health related quality of life
IBSS	International Bibliography of the Social Sciences
ICER	Incremental cost-effectiveness ratio
IQR	Inter-quartile range
KCH	Known cost histories
MAR	Missing at random
MCAR	Missing completely at random
MESH	Medical Subject Headings
MSE	Mean square error
NHS	National Health Service
NHS EED	NHS Economic Evaluation Databases
NICE	National Institute for Health and Clinical Excellence
OHE HEED	Office of Health Economics Health Economics Evaluation Database
PBC	Primary biliary cirrhosis
PH	Proportional hazards
PN	Parenteral nutrition
PSC	Primary sclerosing cholangitis
PSA	Probabilistic sensitivity analysis
QALY	Quality adjusted life years
RCT	Randomised controlled trials
SBP	Spontaneous bacterial peritonitis
SE	Standard error
SEE	Sampling standard error
SF	Short form
SPK	Simultaneous pancreas kidney transplant
UCH	Unknown cost histories
UK	United Kingdom
US	United States

Acknowledgments

My thanks go to my supervisor Martin Buxton who has guided me through this thesis. I would also like to thank Simon Thompson who gave me advice and worked with me on a paper closely related to Chapter 5, and to Alex Sutton, Keith Lambert, the HESG audience and various anonymous referees who have given helpful comments on various aspects of this thesis.

I am especially grateful to all the staff of the Health Economics Research Group for supporting me through this and for letting me hijack their computers every evening and holiday period for two years to do simulation work. If not for their help Chapter 4 would still be awaiting results! Thanks also to those of you in HEDS, ScHARR for your support and encouragement and for kindly allowing me the time to write up this thesis.

Much of the research reported in this thesis was undertaken as part of a programme grant to HERG from the Department of Health. The research was also supported by a project grant awarded by the Department of Health.

I would also like to thank Terry Therneau, Joanne Benson, and Rolland Dickson from the Mayo clinic, Rochester, Minnesota USA for providing me with raw historical data on patients with primary biliary cirrhosis (PBC) which I use in Chapters 5 to 7.

Finally, I would like to thank all my family and friends for their help and support throughout this thesis. Last but definitely not least my thanks go to Mike who has given helpful comments on earlier drafts of this thesis and supported me through the last six years, I could not have done this without him.

CHAPTER 1

INTRODUCTION

1.1 MEASURING UNCERTAINTIES IN ECONOMIC EVALUATIONS OF HEALTH CARE TECHNOLOGIES

Guidelines on the analysis of economic evaluations of health care technologies encourage analysts to include information on the "uncertainty associated with clinical and cost-effectiveness information" [National Institute for Health and Clinical Excellence (NICE), 2004; See: Drummond & Jefferson, 1996; NICE, 2004; Canadian Agency for Drugs and Technologies in Health (CADTH), 2006].

These guidelines state that uncertainty should be dealt with "both systematically and thoroughly" so that informed decisions can be made for new and existing health care technologies [CADTH, 2006]. Results based on patient level data should include details of the statistical tests performed and confidence intervals (CI) should be presented around the main outcomes of interest [Drummond & Jefferson, 1996; CADTH, 2006]. Further, guidelines recommend that the uncertainty around economic evaluation results

be presented in the form of "confidence ellipses, scatter plots on the cost-effectiveness plane and cost effectiveness acceptability curves" [NICE, 2006].

In economic evaluations, uncertainties in sample estimates are usually expressed by standard errors or confidence intervals around point estimates. Briggs *et al,* classify this type of uncertainty as sampling variation [Briggs *et al,* 1994]. This is also referred to in the literature as first order uncertainty or stochastic uncertainty [Briggs, 2000; Coyle *et al,* 2003].

In addition to sampling variation, Briggs and colleagues classify three further types of uncertainty in stochastic analysis – studies where observed data are available [Briggs *et al,* 1994]. The first type of uncertainty that arises is methodological uncertainty, which occurs due to the assumptions made in the process of estimating the outcome of interest, for example choosing a discount rate for future costs and effects of 3.5% when rates of 0% or 6% may be equally applicable. The second source of uncertainty occurs when generalising the study results to alternative populations or settings.

Typically, economic evaluation studies are concerned with observing the lifetime costs and effects of health care technologies, however, the majority of studies only collect information over a fixed (short-term) time period. Therefore, methods are required for extrapolating results beyond the study time period and Briggs and colleagues note that the third source of uncertainty arises from the need to credibly extrapolate results over a longer time frame.

"The process of extrapolating the results of economic evaluations is invariably undertaken using modelling exercises" [Briggs & Gray, 1999]. Further, economic evaluations of health care technologies usually apply some form of mathematical modelling "to progress from the within-trial results to the economic results of interest" [Briggs, 2000]. This introduces three further sources of uncertainty, namely: model parameter uncertainty, model structure uncertainty and uncertainties in the model process [Manning *et al,* 1996]. Model input values, also known as model parameters, might include the probability of an event occurring, unit cost estimates or regression coefficients for a prognostic model for estimating patient survival and these values are subject to (model parameter) uncertainty. Uncertainties in the model structure arise when selecting the mathematical structure of the model. For example, a researcher aiming to model the hazard or risk of death needs to decide between analysing survival

at a fixed time point (using logistic regression or a similar model) and analysing the actual survival time (using Cox regression or a similar model). Model process uncertainties arise because there is more than one solution to a problem. For example, if a problem is assigned to two (or more) analysts to solve independently, the analysts may use different approaches (processes) to derive a solution to the problem.

Guidelines on measuring uncertainties in economic evaluations of health care technologies recommend the use of deterministic and probabilistic sensitivity analysis to explore methodological, extrapolation and modelling uncertainties [Drummond & Jefferson, 1996; NICE, 2004; CADTH, 2006]. When sensitivity analysis is conducted "details should be given of the approach used ... and justification given for the choice of variables and the ranges over which they are varied" [Drummond & Jefferson, 1996].

It is important to identify and measure all sources of uncertainty in economic evaluations and various methodological techniques for measuring sources of uncertainty in economic evaluations have been presented in the literature. Probably the most well debated problem pertains to measuring the sampling variation around incremental cost-effectiveness ratios (ICER), the ratio of the difference in costs to difference in effects between two or more treatment and control groups in a study. Methods for estimating CI include: confidence boxes [O'Brien *et al*, 1994; Polsky *et al*, 1997], confidence ellipses [van Hout *et al*, 1994], Taylor series [O'Brien *et al*, 1994], Fieller's theorem [Fieller, 1932; Fieller, 1954; Willan & O'Brien, 1996], sampling [Mullahy & Manning, 1995] and bootstrapping [Efron & Tibshirani, 1993; Manning *et al*, 1996; Briggs *et al*, 1997]. These techniques have been discussed thoroughly in the literature [See: Briggs & Fenn, 1997; Briggs & Gray, 1999] and will not be dealt with in more detail here.

It is not the focus of this thesis to identify and review all the methodology that has been proposed in the literature for identifying and measuring uncertainty in economic evaluations. However, from the analyst's perspective there are still unresolved issues pertaining to the methodology for measuring various forms of uncertainty in economic evaluations. A recent economic evaluation study in liver transplantation [Longworth *et al*, 2003] identified two such issues, neither of which are unique to this one study. From the analysts perspective this created an ongoing problem that hindered the ideal of allowing for all uncertainty associated with a study, and consequently formed the motivation for this thesis.

1.2 FOCUS OF THIS THESIS

This thesis takes as its basis a cost-effectiveness study in liver transplantation and explores some specific methodological, model parameter estimate and model uncertainty issues that arose during the course of this study, namely:

- the selection of an appropriate method for estimating mean total costs in the presence of censoring[1] (methodological uncertainty)
- methods for estimating uncertainty around non-transplant survival, quality adjusted life years (QALY) and cost estimates derived from a prognostic model[2] (methodological, model parameter estimate and model uncertainty)

This thesis first reviews a series of methods proposed in the literature for estimating mean total costs in the presence of censoring. The comparison of methods aids researchers by providing a guide to the most appropriate technique to use under particular circumstances. Although the methods of estimation set under censoring are not themselves original, this is the first time any study has statistically compared the existing methods in an empirical application and attempted to derive recommendations that can be applied in future economic evaluations where censoring is an issue.

The second contribution of this thesis is to provide original methodology for incorporating prognostic model uncertainty in to cost and effectiveness estimates in studies where no observed control group is available. These uncertainties can be incorporated in to economic evaluations in organ transplantation and the results presented to decision makers to aid their decisions. The application of these prognostic model techniques is generalisable beyond the area of organ transplantation to other studies where there is an absence of an obvious control group.

1.3 STRUCTURE OF THIS THESIS

This study has been divided in to seven further chapters, as set out below.

Chapter 2 introduces the United Kingdom (UK) cost-effectiveness in liver transplantation (CELT) study and presents the original study methodology and results over a 2.25 year time frame [Longworth *et al*, 2003]. Consideration is then given to

[1] Censoring occurs when a number of patients in a study do not have the end-point of interest within the study period and are lost to follow-up prior to the end of the study.
[2] Prognostic models are mathematical models that can be used to estimate patients outcomes, e.g. survival, in the absence of observed data (See Chapter 5 for further details of prognostic models).

extending the CELT study over a five-year time frame and this chapter identifies two areas of concern in extending the study: 1) selecting an appropriate method that accurately estimates mean study costs in the presence of censoring and 2) measuring prognostic model uncertainty when estimating control group effectiveness and costs in the absence of an observed control group. These two issues will be taken forward in the remainder of this thesis.

Solid organ transplantation is currently considered as the treatment of choice for certain patients with end-stage organ failure and has never been the subject of a randomised controlled trial (RCT). **Chapter 3** discusses alternatives to the RCT study design for cost studies, effectiveness studies or cost-effectiveness studies in solid organ transplantation. The chapter highlights some of the problems in applying an observational study design within solid organ transplantation. This chapter also reviews the literature on solid organ transplant studies, focussing on studies where transplantation is compared with an alternative treatment for patients with end-stage organ failure. The aim of this review was to confirm that the two areas of uncertainty, identified in the CELT study and taken forward in this thesis, were general statistical issues arising in solid organ transplantation studies. The review identifies that the majority of studies are poorly designed and fail to either acknowledge or incorporate uncertainties associated with their choice of study design.

Chapter 4 compares a series of methods proposed in the literature for estimating the mean total costs in the presence of censoring. The methods are compared across different levels of censoring and different censoring mechanisms, since the accuracy of methods may vary according to the proportion of censored cases and the reason for censoring. Censoring is simulated for an observed cohort of patients for whom complete liver transplant costs over a 2.25 year study period have been observed. Censored estimates of mean total costs and standard errors are compared with the observed mean estimates and their standard errors over the 2.25 year study period prior to censoring. This chapter makes recommendations on the most appropriate censored cost techniques to use, based upon the nature of the study cost data.

Chapter 5 sets out a simulation technique that accounts for prognostic model parameter uncertainties. The impact of *parameter uncertainty* on non-transplant survival and the survival gain from liver transplantation are compared and the effect of allowing for model uncertainty is discussed. This chapter also considers a series of

steps for selecting an appropriate prognostic model that can be applied to a series of transplant patients in order to estimate, what would have been, their survival in the absence of transplantation. The implementation of *model uncertainty* in to the *model selection* process is also considered. Three models for predicting survival in patients with end-stage primary biliary cirrhosis (PBC), a type of liver disease, are used to demonstrate these methods.

Although the prognostic models applied in the original CELT study estimate individual patient survival lengths they did not estimate patient specific outcomes (survival/death) over the 2.25 year study period. **Chapter 6** proposes a series of techniques for estimating individual patient outcomes over a fixed time period. The need to estimate individual outcome predictions may arise due to: estimating patient level survival, QALYs or costs at the individual level, estimating survival, QALY or cost gains at the individual level, or adjusting for costs or health related quality of life (HRQL) data for a time period prior to death. Any approach for estimating outcomes should maximise the use of the control group information that is available and, given that the individual patient outcome is an estimate, should allow for uncertainties around the estimate. This chapter also investigates the impact of estimating patient level outcomes on QALY and cost estimates.

Chapter 5 and 6 considered three types of prognostic model uncertainty; prognostic model parameter uncertainty, prognostic model outcome uncertainty, and prognostic model selection uncertainty. Additionally, estimates of individual patient survival are conditional on the survival predictions obtained from prognostic models and **Chapter 7** investigates the impact of all three sources of prognostic model uncertainty on non-transplant survival, QALY and cost estimates using probabilistic sensitivity analysis (PSA) over a five-year study period for a possible extension of the CELT study.

In **Chapter 8** the emerging methodological themes are brought together. The contribution of this research to the evaluation of transplantation and its wider application within health services research are discussed.

CHAPTER 2

COST-EFFECTIVENESS OF LIVER TRANSPLANTATION: A CASE STUDY

2.1 INTRODUCTION

Liver transplantation is a medical intervention that has become the accepted standard treatment for end-stage liver failure across several liver disease groups. Despite this, there are still issues and controversies that remain in this area. Section 2.2 explains how liver transplantation has become a treatment of choice despite a lack of RCTs evaluating the effectiveness, costs or cost-effectiveness of liver transplantation. Section 2.2 also considers how the assessment of the clinical effectiveness and cost-effectiveness of liver transplantation has had to address this.

This chapter provides the reader with a brief overview of the original CELT study. Sections 2.3 to 2.5 introduce the original study and describe the study methodology, results, sensitivity analysis and conclusions over the 2.25 year study period. In the absence of an observed control group it is inevitable that particular uncertainties exist when estimating the cost-effectiveness of liver transplantation and Section 2.6

discusses the areas of uncertainty that were addressed within the original CELT study. Section 2.6 also considers the implications of extending the study time frame and the potential impact of the statistical uncertainties on the cost-effectiveness of liver transplantation over an extended period.

2.2 LIVER TRANSPLANTATION: BACKGROUND

The first successful human liver transplant was performed in 1963 by Thomas E. Starzl in Denver, Colerado, United States (US) [Starzl *et al,* 1963]. Five years later Sir Roy Calne successfully performed the first UK liver transplant in Cambridge [Calne & Williams, 1968]. Unfortunately, early liver transplantations often resulted in a relatively poor outcome for reasons such as poor donor organ quality, technical complications and frequent graft failure [Foster & Burton, 1989; Neuberger & Lucey, 1994]. Postoperative complications were frequent with patients having a poor tolerance to the immunosuppressive regime required following transplantation, which commonly led to biliary complications and sepsis.

By the 1980s several factors had led to both an increase in the number of liver transplants being performed throughout the world and an increase in survival following this procedure [Neuberger & Lucey, 1994]. Improvements in surgical techniques and the introduction of the immunosuppressive drug cyclosporine A led to large improvements in survival after liver transplantation. Furthermore, the general adoption of brain stem death as acceptable grounds for instigating liver donation lead to an increase in both the number and the condition of donor organs becoming available [Neuberger & Lucey, 1994]. In the early years of transplantation, an organ could be retrieved from the donors only after the donor heart had stopped beating, by which point a liver may already be damaged due to the human deterioration or "shutting down" process. The acceptance of the definition of brain stem death improved the condition of donor organs, which are now retrieved from donors earlier [Hockerstedt, 1990]. Additionally, the 1983 National Institute of Health Consensus Development Conference recommended that liver transplantation should no longer be regarded as an experimental procedure, but be considered as a "therapeutic modality for end-stage liver disease" [National Institute of Health, 1984].

The number of liver transplants in Europe has therefore increased dramatically over the last 25 years. The number of transplants in Europe was 49 in 1980, increasing to 2,107 in 1990 and 4,624 in the year 2000 [European Liver Transplant Registry (ELTR)

website, 2006]. Prior to 1985, one-year and five-year survival rates were 34% and 21% respectively. These rates had more than doubled to 83% at one-year and 71% at five years for the period 1995 to 1999 [ELTR website, 2006].

Liver transplantation is currently considered the treatment of choice for patients with end-stage liver disease. It is also recognised as being a costly procedure due to the nature of the surgery involved, the need for intensive care post surgery, the cost of the immunosuppressive drug regimes and the necessity and frequency of clinical monitoring post transplant [Hockerstedt & Kankaanpaa, 1986; O'Grady & Williams, 1986; Foster & Burton, 1989; Neuberger & Lucey, 1994]. Several authors have recognised the necessity of exploring the cost-effectiveness of liver transplantation [See: Hockerstedt & Kankaanpaa, 1986; Burroughs *et al,* 1991]. Burton and Heyse noted the following:

> "Biomedical progress makes it possible to save more lives than ever before: however, costly medical innovations such as liver transplantation are sophisticated, but expensive, new diagnostic technologies are bringing us to the point where somebody, not just the physician, but economists, planners, the public, and their elected political leaders will have to make decision on how to distribute the limited material and fiscal resources for health care among the competing medical technologies and treatments needed for and desired by the patient population and health care providers alike." [Burton & Heyse, 1985]

2.2.1 Impracticalities of the RCT Study Design for an Economic Evaluation in Liver Transplantation

It is currently considered unethical to withhold liver transplantation from patients who are eligible for it [Neuberger & Lucey, 1994]. For this reason a RCT in liver transplantation has never been undertaken. In addition to this, there are practicalities in the design of a theoretical RCT that seeks to randomise patients to either a treatment (liver transplant) or control group given the current allocation system of donor livers in the UK [UK Transplant, 2005a].

In the UK, the listing of transplant patients and allocation of donor organs is co-ordinated by UK Transplant [UK Transplant, 2005a]. At point of listing, patients are

classified as either emergency cases[1] or elective (routine) cases, and in the donor allocation system priority is given to emergency patients. When a suitable donor liver becomes available it is matched to recipients as closely as possible by blood group (A, B, AB or O), age and size of liver, in order to increase the likelihood of a successful transplant (minimising the possibility of rejection and complications post transplantation). This matching is performed by a computer programme that has been set up by UK Transplant [UK Transplant, 2006].

Once a liver has been matched it is firstly offered to emergency patients at any of the transplant centres in the UK. If no suitable recipient is available the organ is then offered to elective patients in the transplant centre in the zone where the donor organ originated[2]. If there is still no suitable match it is offered to the remaining transplant centres, where the centre with the highest points total gets priority[3]. Finally if no suitable match is found in the UK the liver is offered to transplant centres in the European Community and worldwide.

Clearly, the selection of patients for liver transplantation is anything but random and a randomised study design would therefore need to fit within the current allocation of donor livers in the UK. Current figures suggest that at any one time point there are approximately 270 patients on the waiting list for a liver transplant [UK Transplant, 2005]. It is reasonable to argue that a sample size of 270 patients is sufficient to use in a randomised controlled trial. However, patients need to be stratified by blood group (four groups), age and size or organ. For illustrative purposes let us assume that patients are stratified in to five age groups and five size groups, let us also assume that the probability of appearing in any group is equally likely across blood group, age and size. Thus, if an RCT in liver transplantation were conducted, patients would need to be stratified in to 100 groups (4x5x5) with two to three patients in each stratum. The randomisation of patients to receive transplantation or not, would then take place within each stratum. In reality, the probability of being in a particular stratum will not be evenly distributed across the 100 strata, with some strata having more patients in than others. Further, if more than one potential recipient existed the patients would additionally need to be in retrieval zones with equal point allocation for randomisation to take place.

[1] Emergency cases are patients who clinicians expect to die if they do not receive a transplant within three days of listing.
[2] For liver transplantation the UK is divided in to seven organ retrieval zones (one in Scotland and the remaining six in England and Wales), each zone contains one liver transplant centre.
[3] "A centre is awarded a point if it passes on a liver it cannot use to one of the other centres. There is a corresponding loss of a point when any centre receives a liver from outside the area" [UK Transplant, 2007].

Given the current system for matching and allocating donor organs, and the related lack of donor livers, it is easy to appreciate that the number of scenarios where there are more than one suitable recipient for randomisation are rare.

Two additional and foreseeable problems with a potential RCT in liver transplantation will be recruitment of patients and it is likely that patients would be reluctant to participate in a trial where they might not receive life saving treatment. Furthermore any patients who do agree to participate are likely to withdraw if they were randomised to a non-transplant group. Additionally, it would clearly be impossible to design a study in such a way that patients and their assessors were unaware of their allocated treatment arm (awareness of treatment is acknowledged as a source of study bias [Pocock, 1983]).

Therefore, an RCT in liver transplantation is considered to be both unethical and impractical to conduct and any evaluation needs to be based on a non-randomised design, comparing liver transplant patients with non-transplant patients. This raises the question of who to use as an appropriate non-transplant control group for comparing liver transplantation within an economic evaluation.

2.2.2 Commissioning of the CELT Study

In 1995, the UK Department of Health (DoH) commissioned the CELT study to attempt to estimate the clinical effectiveness and the cost-effectiveness of the National Health Service (NHS) liver transplant programme in England and Wales. Given the suggestive evidence that liver transplantation is a life saving procedure and the impracticalities of conducting an RCT, an RCT design where transplantation would be withheld from a randomly selected control group, was rejected and a non-randomised study design was used. This approach is described in Section 2.3 and 2.4.

2.3 THE CELT STUDY

Between December 1995 and December 1997, a total of 755 adult patients (aged 16 years or older) with various end-stage liver diseases were assessed for their suitability for transplantation at one of the six DoH liver transplant centres in England and Wales. The six centres were: the Queen Elizabeth Hospital, Birmingham; Addenbrooks Hospital, Cambridge; The Freeman Hospital, Newcastle; St. James' Hospital, Leeds; Kings College Hospital, London; and The Royal Free Hospital, London.

The study focused on reporting the effectiveness, costs and cost-effectiveness of liver transplantation on a per patient basis. Patients were excluded from the study if they had received a previous liver transplant prior to the beginning of the study[4]. However, patients who went on to receive a second or subsequent transplant during the study period, were included in the study.

Patients entered the study at the point they were admitted to the transplant centre for assessment. As a result of the assessment a patient was either:

- listed as a suitable transplant candidate and placed on the transplant waiting list until a suitable liver became available
- deferred and assessed again at a later date
- not listed as a suitable liver transplant candidate

Listed patients were classified as either an emergency or elective case.

Of the 755 transplanted patients, 550 were listed for transplantation, of whom 477 went on to receive a liver transplant. Patients who were listed but did not receive a transplant either died on the waiting list or were removed from the waiting list. Reasons for removal included the patient's condition having improved or having deteriorated, medical reasons, psychological reasons, or the patient's own choice. The patients who received a transplant were followed from point of transplant for two years or until time of death if sooner, 383 patients were alive at two years post transplant, 86% of the transplanted cohort. Data collection for the study ended in December 1999.

Five reports were produced from the original study results; each addressed different aspects of the evaluation. The focus of these reports was: survival [Young *et al*, 2001], HRQL [Ratcliffe *et al*, 2001], costs [Longworth *et al*, 2001], patient costs [Mistry *et al*, 2003] and cost-effectiveness [Longworth *et al*, 2002]. All reports present the results of the analysis across all end-stage liver disease groups, the exception being the cost-effectiveness report which focused on the cost-effectiveness of liver transplantation for

[4] The CELT study team took the decision to exclude these patients as it was difficult to identify the point that assessment first started, thus potentially underestimating these patients' costs. The CELT study team also felt that the clinical data would differ from those patients seen for a first liver transplant. However, exploration of the re-transplant costs for CELT study patients (those patients who underwent two or more transplants within the CELT study) suggests that the average re-transplantation costs £52K, which is similar to the transplant costs of a first transplant (mean = £53K). Therefore excluding these patients should not unduly effect the estimation of transplantation.

three major liver disease groups: alcoholic liver disease (ALD), PBC, and primary sclerosing cholangitis (PSC). Sections 2.4 and 2.5 describe the main cost-effectiveness study, that is the cost-effectiveness of liver transplantation for ALD, PBC and PSC patients, in greater detail.

2.4 CELT STUDY METHODOLOGY

The perspective taken in the main cost-effectiveness study was that of the NHS transplant centres. The CELT study time period was 2.25 years (27 months), which represented the two-year follow-up period post transplant plus the average time a patient spent waiting for a transplant (three months).

An electronic Microsoft® Access database was used for data collection. Databases were installed at each of the six centres participating in the study. A research nurse was employed at each centre to collect prospective information on patient demographics, disease characteristics and resource use attributed to the liver transplant programme.

The research nurse also administered the study questionnaire to patients. The questionnaire contained some demographic questions and two instruments used for measuring HRQL; the short form (SF) 36 [Ware *et al,* 1993] and EuroQOL EQ-5D [Brooks, 1996]. The study questionnaire was administered to all English speaking patients listed for a liver transplant at point of listing, and at three monthly intervals until point of transplant, then at 3, 6, 12 and 24 months post transplant. Questionnaires were not sent to patients who, in the opinion of the research nurse based at each centre, were too ill to complete the questionnaire. Reminders were sent to non-responders approximately three weeks after the first questionnaire was sent.

The costs and effects for each patient who received a transplant were compared with the estimated costs and effects for that patient had they not received a transplant. The only difference in costs and benefits, with and without the transplant programme, for patients who died or were removed from the waiting list, were the costs of assessment. Therefore, patients who were listed for transplantation but removed from the waiting list, or who died before transplantation, were excluded from the cost-effectiveness cohort. However, the assessment costs for the listed patients who did not receive a liver transplant were incorporated in to the cost and cost-effectiveness results for the transplant cohort (See Appendix A2.1 for further details).

2.4.1 The Intervention Group – Liver Transplant Cohort

A total of 347 patients with end-stage ALD (N = 155), PBC (N = 122) or PSC (N = 70) were assessed for their suitability as liver transplant candidates at one of the six centres participating in the study. Patients were followed from assessment, through candidacy (the time between listing and transplantation), their transplant hospital stay and a two-year period post liver transplantation. 247 patients were listed for liver transplantation of whom 208 received a liver transplant during the CELT study period (ALD: N = 82, PBC: N = 81, PSC: N = 45).

Methods for estimating survival, HRQL, QALY and costs for the transplant cohort are detailed below.

2.4.1.1 Survival

Observed information on survival was available for the majority of ALD, PBC and PSC patients from point of assessment up to 2.25 years post-assessment. Patients who did not die during the study were censored 2.25 years after the date of their liver transplant assessment or at the end of the study period for those patients who did not have complete 2.25 year follow-up. Follow-up was incomplete (censored) for 13 out of 208 transplanted cases (6%) – in all 13 cases the study ended before 2.25 years of data had been collected. The mean survival time for the cohort over the 2.25-year study period was estimated from the area under the Kaplan-Meier survival curve [Kaplan & Meier, 1958; Collett, 1994].

In this thesis individual patient survival lengths are calculated in years.

2.4.1.2 HRQL

The EuroQOL EQ-5D is a non-disease specific HRQL questionnaire that asks a series of five questions concerning HRQL (mobility, self care, ability to perform usual activities, pain or discomfort, and anxiety or depression) [Brooks, 1996]. Questionnaire responses can be transformed in to a single index score where a score of one indicates full health and a score of zero indicates death.

The results from the EQ-5D were used in the main CELT analysis as a measure of HRQL. The scoring tariff applied to the CELT cohort is known as the York tariff, where the scoring tariff is based on a representative sample from the UK population [Dolan, 1997].

Using CELT data Ratcliffe *et al* showed that there was no significant change in HRQL over time on the waiting list for liver transplantation [Ratcliffe *et al.* 2005]. As HRQL was not measured at the time of transplantation it was assumed that each patient's HRQL at this point would be the same as it was at their last known pre-transplant response. Similarly, each patient's HRQL score at assessment was assumed to be the same as the score observed at listing. HRQL was assumed to increase or decrease linearly between all other time points.

In a dataset such as this, it is inevitable that HRQL data are not be available for all patients' across all time points. For example, data is regarded as missing for non-English speaking patients and those considered too ill to complete the questionnaire. Mean HRQL was estimated for individuals who had missing data, but had available responses either side of the missing point. Where values either side of the missing value were unavailable the prior or subsequent value was carried forward or backward. A Monte Carlo simulation technique, known as multiple imputation, was used to impute missing values when no EQ-5D data were available[5] [Little & Rubin, 1987; Schafer, 1999].

2.4.1.3 Quality Adjusted Life Years (QALYs)

Individual patient QALYs were estimated by plotting each patient's EQ-5D values over time from point of assessment to 2.25 years post assessment. The number of QALYs for each liver transplant patient was then estimated from the area under the curve.

2.4.1.4 Costs

Comprehensive resource use information was collected prospectively for each patient in the study beginning at the time each patient was assessed for listing. Resource use details were collected on inpatient stay (categorised as ward or intensive therapy unit stay), out-patient visits, the length of the transplant operation, high cost/high volume

[5] Patient's with ALD, PBC or PSC did not complete the EQ-5D at any time point during the study in 13% of cases and multiple imputations were used to impute these missing cases. Multiple imputation is a Markov chain Monte Carlo simulation technique, in which a number of possible values, which are drawn from the predictive distribution of the missing data, are imputed for each missing observation. Multiple imputation assumes that data are missing at random (MAR), which means that the values of the variable(s) which contain the missing data are not a sub-sample of the sampled observations but are a random sample of variables that depend on the values of observed variable(s). In the CELT study it is possible that patient's did not respond to the QOL questionnaire for reasons related to the severity of their underlying liver disease, thus violating the MAR assumption. However, Schafer states that multiple imputation methods do not "require or assume that nonresponse is ignorable. Imputations may in principle be created under any kind of assumptions or model for the missing-data mechanism, and the resulting inferences will be valid under that mechanism" [Schafer, 2007]. Therefore, in the CELT study, although the missing data pattern may not be MAR, imputations were carried out mainly because ignoring the incomplete case would have reduced, an already small, sample size for the study.

drugs, blood products, tests and treatments, dietician sessions, physiotherapy sessions and nutritional support. For each item of resource use the dates of administration were collected. Quantities of blood products, drugs and nutritional support were also collected for costing purposes.

Unit costs for the 1998/9 financial year were obtained from the six centres participating in the study. Unit cost information was not available for all items of resource use from all centres, therefore, a weighted mean cost was calculated, weighted by the number of transplants performed at each centre. Weighted mean costs were applied to each item of resource use.

The British National Formulary [BNF 38, 1999] was used to obtain drug costs. Staff costs for medical and inpatient staff were attributed over the transplant programme activity.

2.4.2 The Control Group – Medical Management of End Stage Liver Disease

Due to ethical and practical considerations it was not feasible to use a RCT design to evaluate the cost-effectiveness of liver transplantation (Section 2.2). Therefore, the CELT study team considered the possibility of collecting data for a non-RCT based control group. Quasi-experimental control groups, intervention delay groups, historical control groups and expert opinions were considered as possible sources of information on non-transplant survival, costs and HRQL.

The CELT study team considered the possibility of collecting information on a cohort of UK patients with end-stage liver disease treated at NHS centres that did not refer patients for transplantation (quasi-experimental group). However, this group of non-transplant patients was rejected because the number of patients who fell in to this category was not deemed large enough to meaningfully compare survival, costs and QALY outcomes with a concurrent transplant group. An intervention delay group, consisting of patients on the waiting list for liver transplantation, was dismissed owing to concerns over selection bias. If the intervention delay group contained patients who died on the waiting list these patients could be sicker than patients who survive to the point of transplant, exaggerating the survival benefit from transplantation [Gail, 1972]. Conversely, if the control group included the non-transplant (waiting list) experience of transplanted patients then the survival affect from transplantation will be underestimated. This later situation is comparable to sicker patients being given priority

to transplantation. Chapter 3 Section 3.3.2.1 discusses weaknesses of the intervention delay group in further detail.

The use of a simple historical control group was also rejected by the CELT study team. It was considered impractical to retrospectively collect clinical and resource use information from patient records, problems were envisaged in locating the records for all eligible patients and in the completeness of the records at the level of detail required for a cost-effectiveness study. Finally, expert opinions were discarded due to concerns over the reliability of collecting information from experts [Black *et al,* 1999].

After discarding these four possible control groups the CELT study team chose to model the expected non-transplant survival, HRQL and costs of the transplant patients, as if these patients continued with the medical management of end-stage liver failure. Prognostic models were used to estimate non-transplant survival; these models were based on historical cohorts of patients with end-stage liver diseases. HRQL and costs, in the absence of transplantation, were estimated from the waiting list experience of the transplant patients. Details of the methods used to estimate non-transplant survival, HRQL and costs are outlined below.

2.4.2.1 Survival

A literature search revealed that several prognostic models based on historical cohorts of patients with end stage liver diseases existed and could potentially be used to estimate survival in the absence of transplantation for the CELT study. In each case, a prognostic model had been fitted to a historical cohort of patients undergoing medical treatments for liver disease. All prognostic models were based on a Cox Proportional Hazards (PH) model. The concept of the Cox PH model is to estimate the hazard, which is the risk of death, at any time among patients who remain alive and under follow-up. Using baseline covariates the model can incorporate several explanatory variables that influence patient survival, for example patient age at time of treatment, or levels of serum bilirubin. Further details of the Cox PH model will be presented in Chapter 5 where prognostic models are considered in more detail.

Prognostic models can be applied to concurrent data using information on the clinical and patient characteristics for the current cohort to give an estimate of survival. The minimum amount of information required from a prognostic model to predict survival in another cohort of patients are the regression coefficients and hazard function from the

prognostic model and patient specific clinical and demographic data for the cohort the model is to be applied to. Given this information, it is possible to compare the mean observed survival for patients who have received the new treatment with their expected survival if they had not. Thus, the survival gain of the new treatment may be estimated.

A key assumption that is made when fitting a prognostic model is that it adequately represents the cohort of patients to whom it is fitted. With this in mind, an appropriate prognostic model must be selected. A choice of prognostic models existed for each of the three liver disease groups evaluated in the main cost-effectiveness study (ALD – two models, PBC – three models; PSC – two models). Table 2.1 outlines each of the prognostic models that could be applied to the CELT cohort to predict non-transplant survival.

A three step process was used to select the most appropriate model for each liver disease group. Firstly, models were chosen if they were shown to be statistically superior to alternative models using criteria set out by Altman and Royston [Altman & Royston, 2000]. These criteria state that models should:

- be validated on both internal and external data sets
- be based on adequate samples of patients
- for each variable fitted in the model the number of events, in this case deaths, should be in the ratio of one explanatory variable for every 10 to 20 deaths
- make clinical sense
- be applied to appropriate cohorts

The second step, for those models that appeared to be statistically valid, was to incorporate models where the historical cohort included UK patients. These patients would be more likely to have similar characteristics and be treated in a similar manner to those in the CELT study, in the absence of transplantation, thus minimising this potential source of bias. Finally, if no disease specific model was superior to the other models after applying the first two steps the final step was to take an average estimate of survival from the remaining models.

At step one, only the Beclere ALD model showed statistical superiority over the Birmingham model. The Birmingham model was not validated on other internal data sets nor was it validated externally. Additionally, the Birmingham model was based on

a relatively small sample of patients (N = 76) and the number of deaths (N ≈ 38)[6] for each of the five variables included in the models was less than the 50 deaths recommended for a five parameter model. Therefore, the Beclere model was chosen as the best model for estimating non-transplant survival in patients with ALD.

Step two resulted in the elimination of the PBC Mayo model and the PSC Mayo model, which did not include UK patients and left the PSC International model and two PBC models, the Royal Free and European models. Given that two PBC models remained, the average survival estimates across these models were applied in the main CELT analysis of PBC patients.

The three models that were not selected for the main CELT analysis (ALD: Birmingham model, PBC Mayo and PSC Mayo) were applied in deterministic sensitivity analysis in order to assess the reliability of the results to model choice.

Survival in the absence of transplantation is observed from point of assessment until point of transplant. Survival in the absence of transplantation is unobserved from point of transplant onwards and needs to be estimated from this point. Clinical information was collected in the CELT cohort at one time point only, immediately prior to transplantation[7]. This information was used to estimate, what would have been, survival in the absence of transplantation using the four disease specific published prognostic models chosen above.

The prognostic models were used to obtain individual patient survival probabilities over time. The survival probabilities can be plotted to obtain individual patient survival curves, where the area under the curve gives an estimate of survival in the absence of transplantation. Further details of this process are given in Chapter 5.

[6] The number of deaths for the Anand model cohort is not stated in the paper, though, from other information that is given, we estimate the proportion to be approximately 49% (38 deaths), which is lower than the 50 required when fitting a 5 variable model.

[7] In 21% of cases clinical information needed to predict survival in the absence of transplantation was missing and ignoring patients with missing information would reduce the sample size of the non-transplant cohort and potentially bias the comparison with the transplant cohort. Therefore, multiple imputations were used to impute missing clinical values [Rubin, 1997; Schafer, 1997].

Table 2.1 Summary of prognostic models that could be applied to the CELT cohort to estimate non-transplant survival

Disease Group	Model name	Publication details	Sample Size	Population	Predictors of survival
ALD	Beclere	Poynard *et al,* 1994 Poynard *et al,* 1999	787	France	Serum bilirubin, serum albumin, the presence or absence of encephalopathy, age
ALD	Birmingham	Anand *et al,*1997	76	UK	Serum bilirubin, serum albumin, blood urea, the presence or absence of spontaneous bacterial peritonitis (SBP)
PBC	European	Christensen *et al,* 1985 Christensen *et al,* 1993	248	Australia, Belgium, Denmark, France, Spain, UK, USA	Serum bilirubin, serum albumin, age, the presence or absence of ascities, the presence of gastrointestinal bleeding
PBC	Royal Free	Hughes *et al,* 1992	289	UK	Serum bilirubin, serum albumin, age, the presence or absence of ascities
PBC	PBC Mayo	Dickson *et al,* 1989 Murtaugh *et al,* 1994	312	USA	Serum bilirubin, serum albumin, age, prothrombin time, oedema score*
PSC	International	Dickson *et al,* 1992	392	UK & USA	Serum bilirubin, age, histological stage, the presence or absence of splenomegaly
PSC	PSC Mayo	Wiesner *et al,* 1989	174	USA	Serum, bilirubin, age, haemoglobin levels, the presence or absence of inflammatory bowel disease, histological stage

* Oedema score: 0 = oedema and no diuretic therapy for oedema, 0.5 = oedema without diuretic therapy, 1 = oedema resolved by diuretic therapy

2.4.2.2 HRQL

Each patient's HRQL score at assessment was assumed to be the same as the score observed at listing. Observed EQ-5D scores were used to estimate HRQL in the absence of transplantation from point of listing to point of transplant. Each patient's last known pre-transplant score was assumed constant over time from point of transplant until death or 2.25 years post-assessment.

2.4.2.3 QALYs

QALYs were estimated in exactly the same way as done for the transplant cohort. EQ-5D values were plotted over time up to 2.25 years post assessment and the area under the curve was calculated in order to measure the number of non-transplant QALYs for each patient.

2.4.2.4 Costs

Costs in the absence of transplantation were estimated by multiplying the average cost per patient per day on the waiting list by each patients estimated survival length, where patient non-transplant survival lengths were estimated from prognostic models. An examination of the cost data for all patients who died on the waiting list revealed that costs increased in the month prior to death (Appendix A2.2). Therefore, an adjustment was made to the cost estimate for those patients who were not expected to survive the full 2.25 year study period.

Equation 2.1 presents the formula used to estimate non-transplant costs for the 2.25 year study period.

Expected costs without transplantation ***Equation 2.1***

$$= A_i*(C_i*S_i) + (1 - A_i)*[C_i*(S_i - 30) + (X_i)]$$

where i denotes the i^{th} patient in the cohort, where i = {1,...,N}

A_i denotes whether patient i was expected to be alive or dead at the end of the study period (1 = alive, 0=dead)

C_i is the average daily cost on the waiting list for patient i

S_i is the predicted survival length (in days) of patient i in the absence of transplantation (time on the waiting list + expected survival time predicted by the prognostic model)

X_i is the predicted cost of patient i for the 30 days prior to death (if the patient was predicted to die), as estimated by the regression model for death costs [Appendix A2.2 Table A2.3.2]

2.4.3 Incremental Cost-Effectiveness of Liver Transplantation

QALYs were discounted at 1.5% and costs at 6% in accordance with NHS guidelines available at the time of analysis [NHS National Institute for Clinical Excellence, 2003]. Given that both cost and QALY data had skewed distributions, bias adjusted non-parametric bootstrapping was performed in order to generate 95% CI around mean cost and QALY estimates[8] [Manly, 1997]. A total of 1,000 repeated samples were taken to estimate 95% CI around mean costs, QALYs and incremental cost effectiveness ratios.

The incremental costs were calculated by subtracting each patient's expected costs in the absence of transplantation from their observed costs with transplantation. Similarly, incremental QALYs were estimated by subtracting each patient's predicted non-transplant QALY from their observed transplant QALY. The individual patient incremental cost per QALY, also known as the ICER, was then calculated by dividing the incremental costs by the incremental QALYs for each patient. Bootstrapping was again performed to obtain 95% CI around the mean cohort incremental costs, QALYs and ICERs.

2.4.4 Sensitivity Analysis

To assess the various assumptions that were made during the CELT study a series of one-way sensitivity analysis were performed across all three liver disease groups; these assumptions are listed in Table 2.2 below.

[8] In the absence of information from the population it is reasonable to make inferences using information obtained from a random sample. In the absence of being able to sample from the population it is reasonable to resample from the random sample and make inferences about the population from this [Manly, 1997]. Bootstrapping takes repeated samples of size N from a cohort of the same size with replacement. For each random sample the parameter/s of interest, for example mean transplant costs, are estimated.

Table 2.2 Details of the one-way sensitivity analysis undertaken in the CELT study

One-way Sensitivity Analysis	Reason the analysis was undertaken
Alternative prognostic models	The three models that were discarded in the main analysis could predict non-transplant survival as accurately or more accurately than the models used in the main analysis
Assume HRQL deteriorated linearly from last known pre-transplant score to zero at point of death in the non-transplant group	In the main analysis HRQL was assumed constant from last known pre-transplant observation until death and this was felt to overestimate QALYs
Organ retrieval costs (estimated at £7,200)*	These were not included in the transplant group costs in the main analysis, owing to the difficulty in obtaining a reliable estimate
Lowest set of unit costs used for key items of resource use (inpatient stay, outpatient visits, transplant operation)	Unit cost estimates varied considerably across the six transplant centres
Highest set of unit costs used for key items of resource use (inpatient stay, outpatient visits, transplant operation)	Unit cost estimates varied considerably across the six transplant centres
Increase daily non-transplant costs (50%)	Assumed that the daily costs in the absence of transplantation would be the same as the mean daily cost on the waiting list, but may be an underestimate as any care provided locally to patients is excluded
Decrease daily non-transplant costs (50%)	Assumed that the daily costs in the absence of transplantation would be the same as the mean daily cost on the waiting list, but may be an overestimate as patients might be treated more resource intensively on the waiting list

* Data from Royal College of Surgeons and Engineering and Physical Sciences Research Council [Engineering and Physical Sciences Research Council, 2002; The Royal College of Surgeons, 2002]

2.5 SUMMARY OF THE MAIN CELT RESULTS AND CONCLUSIONS

Table 2.3 presents key demographic details for the transplant cohort. Although the percentage of male patients varies by disease group, the percentage is typical for each disease, where the incidence of ALD and PSC is higher in males and the incidence of PBC higher in females. Only one patient, a PBC case, was classified as an emergency

case. Survival at two years post transplant was 80% or more for all three disease groups.

Table 2.3 Demographic details of transplanted patients in the main CELT study

	ALD (N = 82)	PBC (N = 81)	PSC (N = 45)
Median Age (IQR)*	50 (45 to 57)	56 (51 to 62)	49 (38 to 56)
Males (%)	67 (82%)	8 (10%)	31 (69%)
Emergency cases (%)	0 (0%)	1 (1%)	0 (0%)
Retransplantations (%)	6 (7%)	9 (11%)	7 (16%)
Survival to 2-years post transplant (%)	67 (82%)	69 (85%)	36 (80%)

* IQR – Inter-quartile range

The main cost-effectiveness results for the CELT study are presented below for the three liver disease groups; ALD, PBC and PSC (Table 2.4). Mean incremental survival differs across disease groups and is greatest for ALD patients. However, once HRQL is adjusted for, there is little difference in QALYs across disease groups, where the quality adjusted survival gain for transplanted patients over a 2.25 year period is approximately six months in each group. The observed costs of the liver transplant programme for ALD patients (£66K) were higher than the costs for PBC (£52K) or PSC (£61K) and the incremental costs of transplantation were over £10K more for ALD patients than PBC or PSC patients.

Table 2.4 Summary of main CELT results (mean with 95% Bootstrapped CI) [Longworth et al, 2003]

	ALD (N = 82)	PBC (N = 81)	PSC (N = 45)
Mean Incremental Survival in years (95% CI)	0.59 (0.38 to 0.74)	0.37 (0.18 to 0.55)	0.23 (0.01 to 0.41)
Mean Incremental QALY (95% CI)	0.55 (0.40 to 0.69)	0.54 (0.39 to 0.69)	0.58 (0.40 to 0.75)
Mean Incremental Cost (95% CI)	£25,712 (£7K to £41K)	£15,224 (£0 to £28K)	£12,182 (-£13K to £33K)
Mean ICER (95% CI)	£48,355 (£12K to £83K)	£28,716 (£1K to £59K)	£21,332 (-£23K to £60K)

Figure 2.1 shows the distribution of bootstrapped net costs and QALY differences on the incremental cost effectiveness plane. There are no negative incremental QALY gains. However, there is some evidence to suggest that liver transplantation is cost saving over a 2.25 year period (negative incremental cost estimates) for PSC patients, where over 15% of the bootstrapped incremental costs fall in the south east quadrant of the cost-effectiveness plane.

The mean ICER was highest for the ALD group (£48K) and lowest for the PSC group (£21K). If an NHS bench mark of £30,000 were assumed to be what the NHS could afford to pay for additional QALYs, then the liver transplant programme would be an acceptable cost-effective intervention for PBC and PSC patients but not for ALD patients. However, the time frame of this study was relatively short at 2.25 years and the cost per QALY for ALD patients would almost certainly decrease over an extended time period.

Figure 2.1 Incremental cost-effectiveness plane for 1,000 bootstrapped cost and QALY estimates for ALD (red), PBC (blue) and PSC (green)

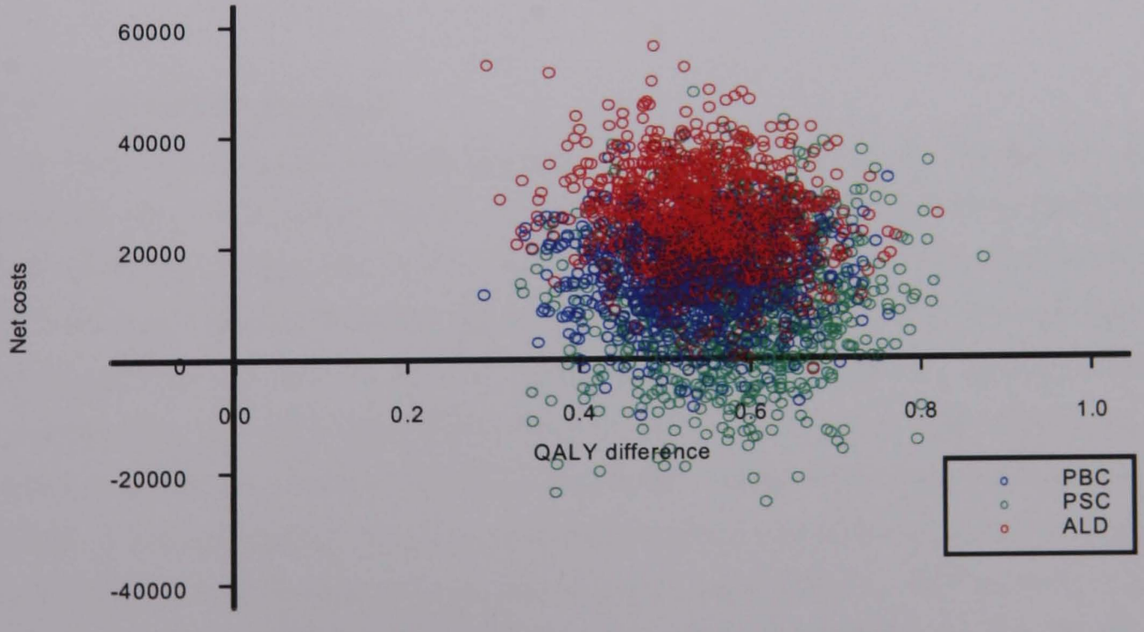
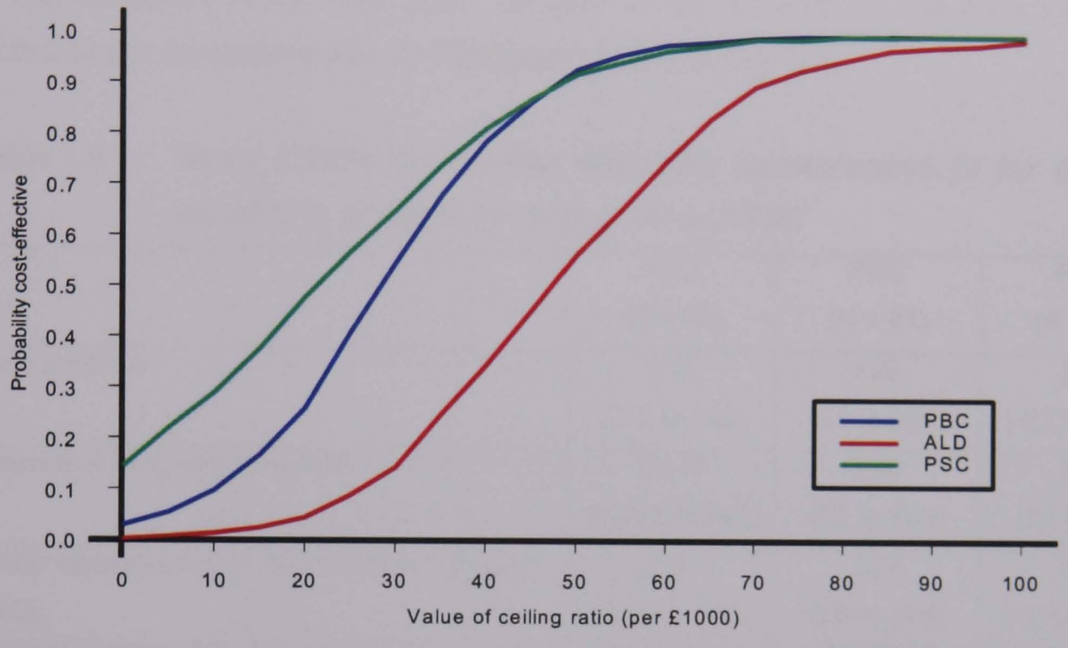


Figure 2.2 shows the cost-effectiveness acceptability curves for the probability of liver transplantation being cost effective for different values of the maximum acceptable incremental cost per QALY for each liver disease group. The curves show that the

cost-effectiveness of liver transplantation is most likely to be acceptable in patients with PSC, then PBC and least likely to be acceptable for ALD patients.

Figure 2.2 Cost-effectiveness acceptability curves for patients with ALD (red line), PBC (blue line) and PSC (green line)



2.5.1 Sensitivity Analysis

The mean incremental cost-effectiveness results for a series of one-way sensitivity analyses are presented in Table 2.5. The incremental cost-effectiveness ratios for the alternative PBC prognostic model (PBC Mayo model) were very similar to the ICER for the averaged Royal Free and European models. However, the ICERs for ALD and PSC were highly sensitive to the choice of prognostic models, where the alternative models increased the average incremental cost per QALY gained. The mean ICER was also sensitive to the assumption that HRQL remained constant over time and assuming that HRQL in the non-transplant group decreased linearly over time resulted in a decrease in the ICER for all disease groups, although the mean ICER for ALD patients was still greater than the £30K NHS benchmark. The reality for patients in the absence of transplant is likely to lie somewhere between the two extremes.

Including an estimated cost for organ procurement increased the ICER across all groups. Estimates of the ICER were sensitive to changes in unit cost estimates for key items of resource use. Using the £30K NHS benchmark liver transplantation might be funded if the lowest sets of unit costs are observed for key resource use items PBC

might not be funded if the highest set of unit costs are observed with an ICER greater than the £30K benchmark. Results were also sensitive to changes in the cost per day on the waiting list.

The impact on the ICERs when varying unit cost estimates in sensitivity analysis was larger than the impact on the ICERs when varying the assumption related to changes in non-transplant HRQL over time. Variation in ICER's was smaller than that when increasing or decreasing the non-transplant daily costs by 50%.

Table 2.5 Mean ICER's (in £1000s) with 95% bootstrapped CI for one-way sensitivity analysis for ALD, PBC and PSC

	ALD (N = 82)	PBC (N = 81)	PSC (N = 45)
Main analysis	£48 (£12 to £83)	£29 (£1 to £59)	£21 (-£23 to £60)
Alternative prognostic models	£119 (£21 to £242)	£30 (£0 to £63)	£35 (£8 to £55)
Linear deterioration of non-transplant groups HRQL	£38 (£10 to £65)	£22 (£0 to £45)	£15 (-£16 to £41)
Organ retrieval costs	£62 (£25 to £101)	£43 (£14 to £76)	£34 (-£15 to £69)
Lowest unit costs	£29 (£1 to £53)	£10 (-£14 to £30)	£6 (-£27 to £32)
Highest unit costs	£56 (£5 to £102)	£32 (-£8 to £73)	£22 (-£42 to £77)
Increase non-transplant cost per day (50%)	£14 (-£40 to £49)	-£4 (-£52 to £29)	-£19 (-£92 to £31)
Decrease non-transplant cost per day (50%)	£82 (£53 to £126)	£61 (£40 to £95)	£62 (£32 to £94)

2.5.2 Summary of CELT Conclusions

By making a series of plausible assumptions, the CELT study illustrates that it is possible to estimate the costs of the medical management of end-stage liver disease and the cost-effectiveness of liver transplantation using published prognostic models as a basis for the non-transplant estimation of survival, HRQL and costs.

The incremental QALY gain of liver transplantation over 2.25 years was approximately six months across all disease groups. However, the cost-effectiveness of liver

transplantation in the UK differed by disease group and was poorest for ALD[9]. The results raised questions concerning the cost-effectiveness of liver transplantation as an intervention for patients with ALD and the authors suggested extending the analysis over a longer time frame so that the mid to long term benefits of liver transplantation could be measured. Liver transplantation might prove to be a cost effective intervention over a longer time frame for all liver disease groups.

2.6 ISSUES OF STATISTICAL UNCERTAINTY IN THE CELT STUDY

The results from the main CELT study for the 2.25 year study period for ALD, PBC and PSC patients have been published and were presented to the sponsors of the study, the DoH. Although there was evidence to suggest that liver transplantation was a cost effective treatment for PBC and PSC patients the DoH were concerned with the relatively short study time period and felt the benefits of transplantation would be more evident over a longer time frame. A five-year time frame was suggested. The proposed extended time frame raised two areas of concern; 1) the selection of an appropriate method for estimating mean study costs in the presence of censoring and 2) measuring uncertainty around the prognostic model, non-transplant estimates in the absence of a control group. These issues will be raised in Section 2.6.2 and explored in the remainder of this thesis. Section 2.6.1 will first consider sources of uncertainty that were addressed within the CELT study when using prognostic models in the absence of an observed non-transplant control group.

2.6.1 Uncertainties Addressed in the Main CELT Study

In the absence of information from an observed control cohort of non-transplant patients, prognostic models were used to estimate, what would have been, the transplant cohort's expected survival in the absence of transplantation. The prognostic models are the pivotal step in estimating the non-transplant survival, HRQL and costs. Non-transplant costs and QALY estimates depend on the length of time a patient with end-stage liver disease is predicted to survive without transplantation. One particular concern was that the patient cohorts on which the prognostic models were derived were not representative of the CELT cohort in terms of waiting list criteria, given that more than one prognostic model existed for each liver disease group. It was therefore important to select the most statistically robust and appropriate models for estimating non-transplant survival. The importance of this choice was evident in the sensitivity results. Whilst it was reassuring that the PBC models gave similar survival and ICER

[9] This was felt to be mainly due to the high cost of assessing a large number of patients with ALD (N = 55) who were ultimately unsuitable liver transplant candidates.

estimates, the alternative ALD Birmingham model and PSC Mayo model showed markedly different ICER estimates from the results of the main analysis. Although the justification of model choice appeared reasonable, one can never be totally sure that the most appropriate model has been selected.

In the absence of an observed non-transplant cohort, information on HRQL and costs in the absence of transplantation were estimated from each transplant patient's HRQL and cost experiences on the waiting list. One-way sensitivity analyses were conducted that increased or decreased the average daily cost of the waiting list by 50%. The rates of cost variation used in the one-way sensitivity analysis were arbitrarily chosen and could be argued as being extreme choices. Cost-effectiveness ratios doubled across all disease groups when daily costs were decreased by 50%, and when daily costs were increased by 50% the cost-effectiveness ratios decreased by £30K or more, in comparison with the main analysis.

The main CELT study was not the only organ transplant study to assume that non-transplant costs remained constant over time [Bonsel *et al.* 1990a; Van Enckevort *et al.* 1997; Anyanwu *et al.* 2002]. Therefore, it appeared reasonable to assume that costs remained constant over time in the CELT study. An alternative approach could be to assume that non-transplant costs vary over time. If data were available on the daily or weekly costs incurred on the waiting list conventional modelling work could explore time dependent changes in non-transplant costs over time. Daily costs were not varied further in the main CELT study and uncertainty in these estimations could perhaps have been explored further.

Unit costs were treated as fixed parameters in the CELT study and varied using the minimum and maximum set of unit costs provided by the six transplant centres for key items of resource use (ward and ITU inpatient stay, out-patient visits and transplant operation). This type of sensitivity analysis is referred to in the literature as extreme scenario analysis [Briggs & Gray, 1999]. The application of the lowest set of unit costs reduced the cost-effectiveness ratio, in favour of liver transplantation, for all three disease groups and the application of the highest set of unit costs increased the cost-effectiveness ratio. An alternative approach to exploring the uncertainty around unit costs would be PSA. In PSA, statistical distributions are assigned to each unit cost estimate and Monte Carlo simulations are run to re-estimate both the outcome of interest and the uncertainty around it [Doubilet *et al,* 1985; Critchfield *et al,* 1986]. The

application of PSA to unit cost uncertainty is well established and will not be explored further in this thesis [See: Lord & Asante, 1999; Briggs, 2001].

Non-transplant HRQL was assumed to be the same as that observed at point of listing for transplantation, this assumption was not varied in sensitivity analysis. However, the assumption that non-transplant HRQL remained constant over time was varied, where one-way sensitivity analysis assumed HRQL deteriorated linearly to point of death, "the best estimate probably lies between these two extremes" [Longworth *et al,* 2003].

2.6.2 Remaining Issues of Uncertainty

For the transplant cohort the potential extension of the study period to five years simply involved the additional collection of survival, HRQL and resource use information for up to five years post assessment. Censoring was anticipated as being an issue within the CELT study if it was extended to five years as a larger proportion of patients would be expected to be lost to follow-up and thus have incomplete data over the study period[10]. Methods for allowing for censoring are well established in time to event (survival) data, however, these methods should not be applied to censored costs and QALY data and a series of alternative methods for estimating mean study costs and QALYs in the presence of censoring exist. Faced with a choice of possible methods, it is not obvious which method will produce the most accurate estimate of mean total costs or QALYs and their uncertainty. This thesis focuses on methods for estimating mean total costs in the presence of censoring, however, the twelve techniques explored in Chapter 4 can also be applied to studies where QALYs are censored.

The second area of uncertainty, results from using prognostic models to estimate non-transplant survival and the subsequent impact of the model estimates on the non-transplant cost and QALY estimates. Chapters 5 to 7 explore methods for estimating prognostic model uncertainties in further detail. Chapter 5 introduces a Monte Carlo simulation technique that incorporates prognostic model parameter uncertainty. Chapter 5 also considers how estimates of prognostic model uncertainty could be incorporated in to selection criteria for choosing an appropriate prognostic model, when more than one model exists.

Although the prognostic models used in the CELT study estimate survival length they do not estimate patient specific outcomes. In the CELT study this information was

[10] Over the 2.25 year study period 6% of the CELT cohort of ALD, PBC and PSC patients dropped out of the study prior to its end, censoring was therefore ignored in the main CELT analysis.

necessary, in order to estimate the individual survival gain of transplantation, and if a patient died during the study period, to estimate the non-transplant costs in the month prior to death (Equation 2.1). PSA is proposed as a technique for estimating individual patient outcomes and the uncertainty around the outcome predictions (Chapter 6).

Finally, Chapter 7 explores the impact of prognostic model uncertainty on non-transplant costs and QALYs over the five-year study period.

The two issues that have been selected for further exploration in the remainder of this thesis were chosen as matters specifically pertinent to a particular cost-effectiveness study in liver transplantation. Therefore, it seemed sensible to conduct a literature review to confirm that the methodology for estimating costs in the presence of censoring and techniques for estimating uncertainty around prognostic model estimates were issues that had been adequately addressed in other solid organ transplant studies. The results of this literature review are presented in Chapter 3.

CHAPTER 3

DIFFICULTIES ARISING IN SOLID ORGAN TRANSPLANT STUDIES

3.1 INTRODUCTION

This thesis takes as its basis a cost-effectiveness study in liver transplantation and explores two issues of uncertainty that are pertinent to this study. The CELT study is an observational study, and Chapter 2 (Section 2.2) explained why it is currently considered unethical to randomise patients to a control group where the alternative to liver transplantation is to withhold treatment. Chapter 2 also illustrated how the randomisation of patients was likely to yield a sample size too small to establish the clinical or cost-effectiveness of liver transplantation. In the UK the practical difficulties in conducting an RCT to evaluate the clinical or cost-effectiveness of solid-organ transplantation arise from the stratification process that would be necessary to account for donor/recipient matching criteria and current donor organ allocation procedures [UK Transplant, 2005a].

The ethical and practical considerations that have prevented the conduct of RCT in liver transplantation are also applicable to solid organ transplantation in general. Therefore, alternative study designs to the RCT are needed to evaluate the effectiveness, costs or cost-effectiveness of solid-organ transplantation.

The CELT study chose a prognostic model approach for estimating, what would have been, survival in the absence of liver transplantation [Longworth *et al,* 2003]. Prognostic models were based on historical cohorts of patients with end-stage liver diseases. Non-transplant costs and HRQL were estimated from the non-transplant experience of patients on the waiting list for transplantation (intervention delay). However, this is just one approach to estimating non-transplant survival, HRQL and costs and alternative approaches exist for observing or estimating a non-transplant cohort which could also produce adequate estimates of non-transplant survival, HRQL or costs. Therefore, the first half of this Chapter (Section 3.2) presents a more complete list of alternative study designs to the RCT that could be used in clinical or cost-effectiveness studies in solid organ transplantation. Section 3.2 also discusses the strengths and limitations of each study design within the context of economic evaluations in solid organ transplantation.

The problems of uncertainty identified in the CELT study are not unique to this one study and are common problems that are characteristic of solid organ transplantation studies. The remainder of the chapter presents the results of a literature review of evaluative studies of solid organ transplantation (Section 3.3). The literature review describes and critically appraises the study designs used when attempting to estimate the clinical or cost-effectiveness of solid organ transplantation. The review was conducted to confirm the methodological issues that were pertinent to the CELT study (accurately estimating mean study costs in the presence of censored costs and prognostic model uncertainty) were relevant to other solid organ transplant studies. The review also sought to establish what approaches had been taken by other researchers to overcome the issues of estimating mean costs in the presence of censoring or to measure prognostic model uncertainty. Section 3.4 discusses the implications of the chapter findings and highlights how this thesis will contribute to the methodology on techniques for estimating statistical uncertainty in evaluation studies.

3.2 ALTERNATIVE STUDY DESIGNS TO THE RCT

Various study designs have been proposed to evaluate a medical intervention. The NHS Centre for Reviews and Disseminations categorised these in to six categories and presented them in a "hierarchy" of scientific credibility [NHS Centre for Reviews and Dissemination, 2006]. These designs are, in order:

- experimental studies with randomisation (RCT)
- experimental studies without randomisation
- cohort observational studies
- case-control observational studies
- observational studies without a control group
- expert opinion

This section explores the five alternative types of study designs that could be used as a substitute to the RCT, i.e. experimental studies without randomisation, cohort observational studies, case-control observational studies, observational studies without a control group and expert opinion. A definition of each type of study design, and the steps that should be considered to minimise potential bias and uncertainty within the study design are presented in Sections 3.2.1 to 3.2.5. The weaknesses of each experimental design are also considered. Section 3.2.6 describes a possible extension to experimental studies without randomisation, observational studies and observational studies without a control group using modelling.

Each approach is presented within the context of solid organ transplantation, where the primary outcome of interest is effectiveness, costs or cost-effectiveness. It is assumed that studies in transplantation for end-stage organ failure consist of an observed consecutive series of transplant patients, with the intention being to compare their outcomes to those of patients treated with non-transplant management of end-stage organ failure over a study period of interest.

3.2.1 Experimental Study without Randomisation

The NHS Centre for Reviews and Dissemination defines an experimental study without randomisation as "the allocation of patients to different interventions [...] managed by the researcher [where] the method of allocation falls short of genuine randomisation e.g. alternative or even-odd allocation" [NHS Centre for Reviews and Dissemination, 2006].

3.2.1.1 Weaknesses

In solid organ transplantation studies the same ethical and practical implications that prevent RCT studies from being conducted also hold true for experimental studies without randomisation. It is still likely to be considered unethical to withhold organ transplantation from a cohort of patients with end-stage organ failure, regardless of what the allocation process is. Further, the practical difficulties from the stratification process that would be necessary to account for donor/recipient matching criteria and current donor organ allocation procedures still apply (see Section 2.2.1 for a brief description of these issues in liver transplantation). Additionally, any non-randomised experimental study is likely to struggle to recruit patients and could have high withdrawal rates, especially in the control cohort (Section 2.2.1).

Therefore, experimental studies without randomisation are not considered as a plausible alternative to the RCT for evaluating the clinical, cost or cost-effectiveness of solid organ transplantation and will not be considered further in this chapter.

3.2.2 Observational Study with Control Group – The Cohort Study

An observational study in solid organ transplantation occurs when the survival, HRQL or costs from an observed cohort of patients receiving solid organ transplantation are compared with survival, HRQL or costs from an observed cohort of non-transplant patients managed for end-stage organ failure. Two possible non-transplant cohorts exist: a concurrent cohort and a historical cohort. The remainder of Section 3.2.2 defines concurrent and historical cohorts, within the context of solid organ transplant studies and considers the potential weaknesses with each type of non-transplant controls.

3.2.2.1 Concurrent Control Group

In an observational study with a concurrent control group, information is collected on a cohort of patients who present with end-stage organ failure during the same time period as the patients who are transplanted. Patients in this cohort are not transplanted and so form the control group against which to assess the outcome of those who are transplanted. Thus, the control group may consist of patients with end-stage organ failure who are not listed for transplantation or the waiting list experience of patients with end-stage organ failure.

There are several reasons why patients might not be transplanted and hence be considered as a "control" patient, including:

- transplantation might not be a referral option at the treatment centre
- the patient has contraindications to transplantation (Box 3.1 presents a list of contraindications for liver transplantation [Neuberger & Lucey, 1994])
- the patient is considered too ill for transplantation
- the patient is considered too well for transplantation
- the patient refuses treatment
- the patient is psychologically unsuitable
- the patient dies whilst being assessed for transplantation
- the patient dies whilst on the waiting list for transplantation
- the patient is removed from the waiting list prior to transplantation[1]

Box 3.1 Contraindications to liver transplantation

- active alcohol abuse
- active sepsis
- advanced cardiac or pulmonary diseases
- cardiac diseases
- current metastatic or extrahepatic cancers or previous malignancies
- diabetes
- diagnosis of AIDS or HIV infection
- hepatitis B
- patient age
- previous biliary surgery
- previous psychiatric illnesses
- severe pulmonary hypertension

In other words, the patient may not be deemed suitable for transplantation (by the clinician) or may themselves choose not to undergo transplantation, or may be unable to access transplantation.

[1] Common reasons for patients being removed from the waiting list include either an improvement or deterioration in health.

To avoid potential differences in outcomes between the treatment and control cohorts the reason for not transplanting a patient should be independent of what their survival outcome might be post-transplant. For example, patients considered too ill for transplantation should be excluded from the control group, as the decision not to transplant is dependent upon (anticipated) survival post transplant. On the other hand, a patient who refuses treatment might be included, so long as the factors that lead to the decision not to transplant would not have influenced the expected outcome post transplant.

There are two possible populations that can be used as control cohorts in transplantation studies: quasi-experimental control groups and intervention delay groups. These two patient groups are discussed in detail below.

Quasi-Experimental Control Group

Within the framework of transplantation, a quasi-experimental control group consists of those patients who are suitable for transplantation but do not receive it for reasons that are independent of what their survival might be post transplant. For example, patients might refuse treatment for personal reasons or be treated at a centre that does not refer patients for transplantation.

Weaknesses

Quasi-experimental control groups were considered by Buxton *et al* in an economic evaluation of a heart transplant programme in two UK hospitals and by Michel *et al* in an economic evaluation of the Dutch liver transplant programme [Buxton *et al,* 1985; Michel *et al,* 1994]. Both sets of authors reject the quasi-experimental method and state the reason for this is because the sample of patients are liable to be too small to provide a suitable sample size to enable the study to detect a significant difference between groups, if one exists. Published results from Anand *et al* suggest that a quasi-experimental study consisting of patients refusing treatment for personal reasons are likely to yield a small sample size, where 7 of 137 ALD cases (5%) assessed for liver transplantation over a seven year period refused to be listed for transplantation for personal reasons [Anand *et al,* 1997].

Moreover, there could also be problems in obtaining adequate information on a quasi-experimental control group. Patients who refuse transplantation or are cared for at a centre where patients are not referred for transplant may return to the care of their local

hospital or GP and would be more difficult to follow-up and might be lost to follow-up altogether.

Intervention Delay Group

There is a general shortage of suitable donors in solid organ transplantation and so organ supply does not meet organ demand [UK Transplant, 2004]. In liver transplantation, for example, there are approximately 270 patients with end-stage liver diseases on the UK waiting list at any one time point and in the United States (US) this figure is in excess of 17,000 patients [UK Transplant, 2005; The organ procurement and transplant network, 2005]. Patients who remain on the waiting list may therefore provide a control group against which to compare the outcomes of the transplanted patients.

There are three possible definitions of the cohort of patients that might be used as an intervention delay group. **Cohort I:** includes all patients on the waiting list who die or are removed from the waiting list before transplantation. **Cohort II:** includes all patients that are transplanted, collecting details on survival and costs both pre-transplantation and post-transplantation. The survival of patients in the pre-transplant cohort would be censored at point of transplant. **Cohort III:** combines cohorts I and II, thus the cohort would include patients who were removed from the waiting list and the waiting list experience of patients who went on to receive a transplant.

Weaknesses

A drawback of Cohort I is that, typically, only a small number of patients are withdrawn, leading to a control group too small for meaningful comparisons to be based on. Furthermore, this group may be an unrepresentative control group, since the subset of patients who improve, deteriorate or die do not mirror the population of potential transplant patients. Therefore, this choice of control group is problematic for both practical and theoretical reasons.

Gail argues that comparing Cohort I to transplanted patients is analogous to comparing transplant patients with non-transplant patients in a situation where transplant priority is given to healthier patients, thus patients who die on the waiting list are sicker than those patients who survive to point of transplant [Gail, 1972]. In this situation, the survival effect from transplantation will be exaggerated: selecting patients that are more likely to survive the transplant operation implies that those who remain on the waiting

list have poorer health and are more likely to die before a suitable donor becomes available.

An attractive appeal of using Cohort II is that no further information need be collected from a second cohort of patients, and within-patient comparisons (in which patients act as their own controls) usually need fewer patients to demonstrate a treatment difference. However, Cohort II is still unlikely to be representative of a population of patients with end-stage liver disease, as patients are censored at point of transplant and their expected survival in the absence of transplantation, from point of transplant, remains unknown.

Using cohort II as a comparison group for transplanted patients would be compromised if Gail's argument above is true, since if low-risk patients are given priority for transplants then their post-transplant survival prognosis may be expected to be good due to their prognosis (rather than the treatment). Conversely, if transplant priority is given to sicker patients over healthier patients then the survival effect from transplantation will be underestimated because the sicker transplant patients will be more likely to die during the transplant operation and the healthier patients will be more likely to remain alive on the waiting list.

Box 3.2 illustrates how the two alternative selection strategies; Strategy 1 giving transplant priority to healthier patients (i.e. non-transplant Cohort I) and Strategy 2 giving transplant priority to sicker patients (i.e. non-transplant Cohort II) could affect the survival outcome of patients with and without transplantation.

Box 3.2 An illustration of the possible outcomes of two alternative transplantation strategies of patients on the transplant waiting list

Let us take two patients, A and B, both of whom have waited three months for a suitable transplant organ. Patient B has a better prognosis than patient A. An organ becomes available and both patient A and B are suitable matches.

Strategy 1: Transplant priority given to healthier patients (Patient B)
Patient B is given the transplant organ and survives to two years post transplant, whereas patient A dies four months after being placed on the waiting list.

The survival gain for transplantation is estimated by the survival of patient B compared to that of patient A, and is equal to 2.25 years (including the time spent on the waiting list) minus 0.33 years, giving 1.92 years.

Strategy 2: Transplant priority given to sicker patients (Patient A)
Patient A is given the transplant organ, but develops post operative complications and dies one month after transplantation. Patient B dies six months after being placed on the waiting list.

The survival gain for transplantation is 0.33 years for the transplanted patient A (including the time spent on the waiting list) minus 0.5 years for patient B, giving -0.17 years.

What this means is that bias may occur when making a comparison between transplant and non-transplant cohorts, depending on whether donated organs are offered as priority to well patients (leading to bias in favour of transplant), sicker patients (leading to bias against transplant), or regardless of prognosis (no bias). In practice it is not known if either of these biases occur, although both have been suggested [Gail, 1972; Longworth *et al*, 2003a]. It is certainly impossible to fully adjust the comparison for the impact of this treatment selection.

In truth, the transplant selection process is likely to include a mixture of these two scenarios i.e. Cohort III a non-transplant cohort combining Cohorts I and II. Thus, the selection effect (and hence the bias) cannot be quantified and removed from the

comparison of transplant and non-transplant patients. Moreover, it is difficult to predict the direction in which the bias operates.

Several authors have proposed alternative survival methods for adjusting for potential differences between an intervention delay group and a concurrent transplant cohort [See: Turnbull *et al,* 1974; Aitkin *et al,* 1983]. These authors argue that Cohort III, which they assume to be homogeneous, will give a more reliable estimate of non-transplant survival than estimates from each of the Cohort I or II (Box 3.2). However, estimates are unlikely to be representative of survival in the absence of transplantation given that those patients who are censored at the point of transplant (Cohort II) have incomplete non-transplant observations (it is unknown how long they would have lived had they not received a transplant), potentially underestimating what these censored patients' non-transplant survival would have been.

3.2.2.2 Historical Cohorts

The historical comparator cohort should consist of patients with the same medical condition as those receiving the new treatment. In studies in transplantation, a historical cohort should consist of non-transplant patients receiving treatment for end-stage organ failure, who would have met the criteria for acceptance on to the transplant waiting list. Data on a clinically defined group of patients can then be obtained from previously collected datasets or collected retrospectively, and a comparison made between patients who are listed for transplantation and the historical cohort who would have been listed for transplant, had it been available at an earlier date.

Weaknesses

The first problem when using a historical control group arises in deciding which patients it should comprise of. In order to illustrate some of the difficulties in selecting an unbiased historical cohort of non-transplant patients, consider the current UK NHS protocol for listing patients with end-stage liver failure and some of the difficulties in identifying similar patients retrospectively. Patients with end-stage liver diseases are accepted on to the liver transplant waiting list if a) their survival in the absence of liver transplantation is expected to be less than one year or b) their HRQL is very poor [UK Transplant, 2005]. Further, a patient should have a reasonable chance of survival after transplantation. There are many contraindications to liver transplantation, some of which are more stringent than others (Box 3.1).

In addition to the above criteria a patient should, in general, be healthy enough to undergo extensive surgery – liver transplantation is a lengthy operation usually taking several hours – and be psychologically capable of undergoing the intense immunosuppressive regime that normally follows transplantation for the remainder of the patient's life. It is easy, retrospectively, to identify whether a patient meets some of the contraindications for liver transplantation (Box 3.1), for example, the presence of pulmonary hypertension might well be available from patient notes. However, other factors, such as HRQL and the ability to undertake an immunosuppressive regime, will be difficult (if not impossible) to ascertain in this way.

An additional but related problem arises when attempting retrospectively to identify whether a patient had illness of sufficient severity for transplantation to be considered. It is problematic to make a comparison of survival or costs post-transplantation between transplanted and non-transplanted patients, since (among other issues) the date of transplant is by definition unknown for patients in the control group. It is more reasonable to compare survival and costs from the time at which the patient is either listed for transplant or not listed. Thus, the time at which the historical patients would have been eligible for listing must be estimated. As noted previously, this is not necessarily the date of diagnosis; patients with end-stage liver disease in the UK would be listed for transplantation when they were expected to survive for less than one-year without liver transplantation. Estimating the point at which to include an historical patient is difficult, and involves ignoring the retrospective knowledge of any death. The expected outcome of a historical cohort of patients might differ from the actual outcome, with some patients surviving longer than predicted and others less. If available, prognostic models[2] could be used to predicted survival to one year, provided that clinical information relating to disease characteristics is also available retrospectively.

The quality of historical are likely to have evolved over time, where today more and more information is held on computers. Missing clinical information could result in the exclusion of patients where case mix factors cannot be adjusted for. In cost and cost-effectiveness studies missing resource information might result in an underestimation of non-transplant costs and thus an over estimation of the incremental costs of transplantation. Worst still, older notes may even have been destroyed or lost. In

[2] Prognostic models are described in greater detail in Chapter 5.

summary, patient notes may not be of sufficient quality to allow a historical comparison to be made.

Moving away from such practical considerations, it should also be remembered that there may be differences between the historical control group and the concurrent treatment group that are due to differences in the experimental environment of the study [Pocock, 1983]. Of particular note are developments in treatment practice over time. For example, in liver transplantation the drug ursodeoxycholic acid has been administered to patients with PBC in order to delay the onset of end-stage liver disease and the consequent need for transplantation [Goulis *et al,* 1999]. Within this context, patients considered for transplantation may differ in terms of characteristics (such as age at time of listing) to those from a historical cohort where the drug was not available, if the drug does in fact postpone the onset of end-stage disease. Other medical advances, such as improvement in palliative care, are also likely to increase both the survival and the cost of transplant patients over and above that which was observed for the historical cohort, potentially exaggerating the effectiveness or cost-effectiveness of transplantation.

The selection criteria for organ transplantation will differ by organ and by country, and ideally the historical non-transplant cohort should come from the same centre as the transplant cohort. Otherwise, potential biases will arise due to differences in the selection of patients, where the type of patient selected is likely to vary due to differences in selection criteria between cohorts [Pocock, 1983]. However, using a historical cohort from another centre or country might be an attractive option if the data are already available, or are more complete than any data that are available for the country or centre in question.

The points mentioned above should be taken in to consideration before collecting retrospective information on a historical cohort or using existing historical datasets. It is possible to adjust for differences in patient or clinical characteristics by matching[3] historical controls with concurrent transplant patients using the characteristics of the transplant patients on the waiting list. Mathematical models can be applied (e.g. regression models) to adjust for differences in the clinical and demographic information between cohorts, provided that this information has been collected in both groups.

[3] Each transplant patient is matched to one or more non-transplant historical controls "such that they are alike with regard to the major prognostic factors" [Pocock, 1983].

However, it is very difficult to adjust for uncertainty caused by experimental differences and selecting a historical comparator group presumes the untestable assumption that there is no effect of time on prognosis. This assumption is unlikely to be met, but any attempt to adjust for a time effect will be speculative, because any time effects on the historical cohort are, by definition, irretrievably confounded with the effect of transplantation and hence not measurable.

3.2.3 Observational Study with Control Group – The Case-Control Study

The case-control study is a retrospective study that seeks to examine factors affecting the health and illness of individuals [Crombie, 1996] The NHS Centre for Reviews and Dissemination define the case-control study as "a comparison of exposure to interventions between participants with the outcome of interest (cases) and those without the outcome (controls)" [NHS Centre for Reviews and Dissemination, 2006]. For example, a case-control study seeking to establish whether smoking is a causal factor in bladder cancer would compare the proportion of smokers amongst a cohort of patients with bladder cancer (cases) with the proportion of smokers amongst a cohort of patients without bladder cancer [See: Kunze *et al,* 1992].

3.2.3.1 Weaknesses

There are two possible ways of defining a case-control study within the context of solid organ transplantation. The first approach is to identify a group of transplant cases and compare these cases with a group on non-transplant controls. Within the definition of the case-control study the focus of the study is to identify whether there are any causal factors that explained why the cases received a transplant and the controls do not. However, the effectiveness, costs or cost-effectiveness of solid organ transplantation can not be evaluated under this definition of the case-control study, unless the aim of the study is to illustrate that a patient's survival, HRQL, costs or cost effectiveness results in transplantation. In this chapter we are considering the appropriateness of alternative non-RCT study designs in the evaluation of the effectiveness or cost-effectiveness of solid organ transplantation, therefore, this approach is not considered further here.

The second approach to the case-control study is to define a population of interest, for example the population of patients who have had end-stage liver disease in England and Wales over the last ten years. The next step is to define the outcome of interest, for example survival, and one or more causal factors, here the causal factor of interest

is liver transplantation. The cases are then defined as those patients with end-stage liver disease who have died during the study period and the controls are defined as patients who are alive. In order to have similar cohorts of cases and controls each case might be matched in terms of patient and disease characteristics to a control patient. The next step is to calculate the proportion of transplants that occurred in the case and control groups and calculate an odds ratio: the odds of having a transplant and subsequently dying compared to the odds of having a transplant and surviving. The practicalities of this second approach will be considered in further detail in the remainder of this section.

The first difficulty in conducting a case-control study in solid organ transplantation is identifying the population. For example, is it possible to identify the population of patients with end-stage liver failure, can this be done from records at transplant centres, hospitals or GPs or is there a proportion of the population that remain undiagnosed and thus will not be captured easily? It is likely that a proportion of the population will be missed, for example patients with end-stage organ failure who are not referred for transplantation as they are not treated at a centre who refer patients for transplantation, leading to a bias which is hard to quantify.

The next concern is how to define a study that will evaluate survival over time, HRQL or costs and how to examine how transplantation affects these outcomes. In principle it would be possible to define the outcome of a case-control study as being either dead within two years of diagnosis of end-stage organ failure (case) or alive two or more years after end-stage organ failure (control), and compare the proportion of transplant operations in each cohort. However, a cohort defined in such a way is likely to be confounded by the health of the patient. The difficulties in making case and control groups similar (but not too similar) with regards to aspects other than the outcomes are well recognised in case-control studies [See: Meirik, 1993]. Even ignoring such confounding, retrospective studies cannot conclude causality: a treatment may have lead to an increase in deaths, or may have been given specifically because the patient was deemed as being at high risk of death (which subsequently occurred). In examining these factors within a case-control study design it would not be known which of these was the real reason for any association.

Further difficulties arise when attempting to use a case-control study to examine the effect of transplant on HRQL or costs. To illustrate this, consider a cohort of patients

with end-stage organ failure where the cost of organ failure for each patient is known. Cases would need to be defined as those who cost more than a cut off value (£30K, for example), and controls would be defined as patients with end-stage organ failure who cost less than this amount. There are two issues here: firstly how the cut off value is meaningfully defined and who decides what this value should be? Secondly, questions are raised as to whether the results actually answer a question pertaining to the effectiveness, cost or cost-effectiveness of organ transplantation in a way that is meaningful to decision makers. It is generally accepted that decisions which take costs in to account should focus on the total actual cost of the treatment and not the proportion of patients whose treatment costs more than some threshold.

The final problem with the case control study relates to the availability of HRQL and cost data. It is unlikely that HRQL data would have been routinely collected at the time, and retrospective collection is likely to be either biased or, in the case of patients who died, practically impossible. For cost studies, resource use data would also need to be gathered retrospectively and this raises concerns about the quality and completeness of retrospective data (as identified with historical control groups in Section 3.2.2.2).

On a more general note, the major weakness of the case-control study is that they cannot "provide information on incidence rates of disease" [Meirik, 1993]. The case-control study will establish whether a causal factor (in this case transplantation) increases or decrease the risk of death in the population, but it will not give enough information to establish how the risk of death might change. For example, a case control study may estimate an odds ratio of 0.33 (suggesting that transplant is associated with a two-thirds reduction in the odds of death), but can not estimate the absolute magnitude of this benefit in the population (e.g. whether the odds have reduced from one in three thousand to one in a thousand, or from three in ten to one in ten). For this reason, case-control studies are principally "first-step" tools to investigate whether there may be any association between a treatment and outcome which, if apparent in the study, should ideally be followed up with more rigorous studies of alternative study designs.

To summarise, in principle it is possible to conduct a case-control study to evaluate the effectiveness, costs or cost-effectiveness of solid organ transplant. However, there are problems obtaining a useful interpretation of the results for decision makers, and

additionally any apparent benefit from transplantation would need investigating further using alternative study designs.

3.2.4 Observational Studies without Control Groups

The fourth type of study design listed by the NHS Centre for Reviews and Disseminations is the observational study without a control group, where three possible studies exist: cross-sectional study, before and after study and the case series [NHS Centre for Reviews and Dissemination, 2006]. The possibilities of using each of these three designs within the context of evaluating the effectiveness, costs or cost effectiveness of solid organ transplantation will be considered in Sections 3.2.4.1 to 3.2.4.3.

3.2.4.1 Cross-Sectional Study

The cross-sectional study is similar to the case-control study where a population of patients with end-stage organ failure are identified and an investigation of the relationship between end-stage organ failure and other variables of interest are carried out. The difference between the cross-sectional study and the case-control study is the case-control study collects information retrospectively whereas the cross-sectional study takes a snap-shot of what is happening at one point in time.

Weaknesses

The cross-sectional study can not be used to evaluate the survival benefits, costs or cost-effectiveness of end-stage organ failure because information is only collected at one time point and this will not capture the longer-term costs or survival attributable to transplantation.

However, a cross-sectional study could be used to compare the HRQL of patients who had received a transplant with patients who had not received a transplant at a particular point in time. Adjustments would need to be made to allow for differences in demographic and clinical characteristics between the pre-transplant cohort and the post-transplant cohort.

3.2.4.2 Before-and-after Study

A before and after study in to the survival, HRQL or costs of solid organ transplantation would compare the findings of study participants before and after solid organ transplantation. Patients would probably be followed from either point of assessment

for transplantation or from point of listing until point of transplant (the "before-transplant" phase) and for a period of time after transplantation ("after-transplant" phase").

Weaknesses

The problems of the before-and-after study are similar to those listed for the intervention delay group using Cohort II described in Section 3.2.2.1. Patients are only included in a before and after study if they survive the "before" phase, and hence this cohort of patients are guaranteed to have survived until point of transplant. The pre-transplant experience would be analogous to the waiting list experience of transplanted patients censored at point of transplant. As outlined in Section 3.2.2.1 this is likely to underestimate the survival and HRQL effect of transplantation and exaggerate the costs of the transplant cohort in comparison with the pre-transplant costs.

3.2.4.3 Case Series/Uncontrolled Trial

A case series (also known as an uncontrolled trial) is a study in which a cohort of patients all receive the new treatment or technology, and there is no control group with which to compare these patients. Thus, in solid organ transplantation, all patients would receive a transplant. A comparison is made (sometimes implicitly) between this case series and a second cohort of patients who have received the alternative treatment(s) in previous studies, i.e. non-transplant management of end-stage organ failure.

Weaknesses

The most obvious weakness of the uncontrolled trial is that there is no control group to compare the costs and effects of organ transplantation with. A comparison with the results from a published study consisting of a second cohort of non-transplant patients may be made[4], but the comparison will be deeply flawed. The case series of transplant patients may or may not be similar to those in any non-transplant comparison cohort, since the two cohorts may be from different countries, different time eras, or different risk groups. The extent to which any of these issues confound the impact of the new therapy is unknown.

[4] The comparison with previous studies would not use the original data from previous studies, only their results.

It is possible that less seriously ill patients are likely to be accepted in to an uncontrolled trial, as these patients are more likely to respond favourably and therefore provide the investigator with more favourable results. This concern was pointed out by Gail in the context of heart transplantation, where less seriously ill patients were felt to achieve favourable outcomes post transplantation [Gail, 1972]. On the other hand, in some instances patients in the most serious need of treatment are enrolled in the study. In either case, these patients are unlikely to be representative of the general population of patients needing organ transplantation. The extent and indeed direction, that these biases will act is unknown.

Since this thesis is concerned with measuring the uncertainty in the comparison between two treatments, and since the case study/uncontrolled trial is unable to provide any such comparison, this thesis will not consider the case series/uncontrolled trial further.

3.2.5 Expert Opinions

An alternative method for collecting information on a non-transplant cohort of patients would be to ask a series of experts to estimate the survival, HRQL and resource usage of each patient had they not received a transplant. These experts should be the same people who assess patients for their suitability for transplantation, i.e. the transplant clinicians. When eliciting expert opinions experts should be blinded to the identity of the patients.

Black *et al,* reviewed methods for obtaining consensus from a series of experts and list three approaches for doing this [Black *et al,* 1999]. The first method is the Delphi technique [Dalkey & Helmer, 1963]. Each expert is presented with a series of disease, clinical and possibly HRQL information. The experts then predict the non-transplant survival, HRQL or resource use should that patient not receive transplantation. Experts may also be asked to place ranges of uncertainty around their estimate. Each expert acts independently of the other experts.

After each expert has returned their initial opinions (known as a round) they may be presented with their own results and the results from the other participants and asked whether they wish to revise their prediction. This process may be repeated until consensus is reached across all experts. Alternatively, the process may be run over a predefined number of rounds, or experts may only be asked once.

The second method is known as the nominal group method, in which a series of experts initially express an independent prediction of non-transplant survival, HRQL or resource use [Delbecq & Van de Ven, 1971]. These opinions are collected, the group of experts would meet, and each expert's predictions are fed back to the group for discussion. The experts are then asked to record a revised prediction, these are aggregated and summarised at the group level

The third technique is the consensus development conference, where a series of experts are brought together and are presented with evidence and opinions from various sources [Fink *et al,* 1984]. The experts then retire as a group to consider the evidence and attempt to reach a consensus.

3.2.5.1 Weaknesses

There are three areas of concern when using expert opinions to simulate a control group of non-transplant patients. Firstly, and most importantly, there is a lack of evidence to show the reliability of consensus methods, where consensus can vary depending on the method used, the way the evidence is phrased or presented, the composition of the group of experts, the number of participants and the meeting environment [Black *et al,* 1999].

Secondly, all human decisions are influenced by emotional factors, which can cause under or over optimistic estimates of the outcome of interest [Poses *et al*, 1991]. It would be difficult to account for the factors that influence these decisions, although it is possible to ask the expert to explain why they came up with a particular estimate and revise it accordingly if consensus argues against reasoning.

Finally, the three consensus techniques demand that the experts estimate survival, HRQL or resource use in the absence of transplantation. This creates a logistical problem for the nominal group method and the consensus development conference technique since it requires co-ordinating the meeting of a team of experts and a discussion of the expected medical needs and outcomes of a (potentially sizeable) number of patients would require several meetings.

3.2.6 Estimating Control Group Endpoints using Modelling

An extension to the observational cohort study with a concurrent or historical control group and the observational study without a control group is to use statistical modelling

to estimate what patients' costs and effectiveness would have been in the absence of transplantation. These models can range from simple regression models [Bonsel *et al,* 1991], Markov models [Sagmeister *et al,* 2002], or even complex discrete event simulation models [Ratcliffe *et al,* 2001]. Any modelling approach used will be based upon one or more sources of data that will originate from:

- intervention delay groups
- quasi-experimental control groups
- historical cohorts
- non-transplant controls in case-control studies or cross-sectional studies
- the pre-transplant experience of transplant patients
- a combination of all of the above listed sources of data

There are arguments as to why modelling is an attractive option to estimate non-transplant survival, HRQL or resource use. Firstly, modelling can be used to allow for uncertainty in estimates by assigning distributions around them. Secondly, modelling can be used in the absence of observed information from a non-transplant cohort of patients (prognostic modelling). Prognostic models based upon historical cohorts of patients can be used to estimate non-transplant outcomes whilst incorporating the characteristics of the current cohort of transplant patients (and estimating their outcomes had they not been transplanted).

3.2.6.1 Weaknesses

Model data will be subject to many of the same cautions, strengths and weaknesses mentioned in sections 3.3.2 to 3.3.5 above. The modelled cohort should be representative of the population of patients with end-stage organ failure who are eligible for transplantation, selective (i.e. patients should not be included in the non-transplant cohort because they are too ill/well for transplantation), and based on an adequate number of patients to achieve reliable non-transplant estimates.

Further, as described in Chapter 1, the modelling of control group data will introduce modelling uncertainties, i.e. model parameter uncertainty, model structure (choice) uncertainty and model process uncertainty. Therefore, any study that uses a modelling approach should carry out extensive deterministic and probabilistic sensitivity analysis to explore the effects of model uncertainties on the study results and conclusions in accordance with health technology guidelines [See: NICE, 2004; CADTH, 2006].

3.2.7 Summary

Section 3.2 has presented alternative study designs to the RCT for estimating the costs or benefits in the absence of solid-organ transplantation. For each choice of design there are potential problems with uncertainty caused by the bias in the selection of a control group of patients. It should however be noted that RCTs are not always the most appropriate design to use, since they too have potential problems, for example ethical issues or inadequate sample sizes due to rare diseases or events.

Studies should be designed so that they minimise the potential biases arising from the study design. As Black points out "there is no such thing as a perfect method; each method has its strengths and weaknesses. The two approaches [RCT and observational studies] should be seen as complementary" [Black, 1996]. In some cases a non-randomised study design might be the most appropriate approach to use for estimating the costs or benefits, or both, of new treatments or technologies.

If every effort is made to minimise study population bias from alternative study designs then it is possible to estimate the effectiveness, costs or cost-effectiveness of a new treatment or technology, within the limitations of the study design.

3.3 A REVIEW OF STUDIES EVALUATING SOLID ORGAN TRANSPLANTATION

Section 3.3 presents the results of a literature review of the evaluation of solid organ transplantation. The aim of the review is to identify solid organ transplantation studies that had compared either the costs, benefits or both the costs and benefits of solid organ transplantation with the medical management of end-stage organ failure. The following issues are critically appraised for each of the studies identified:

- the author's source of evidence for the non-transplant management of end-stage organ failure (e.g. quasi-experimental, intervention delay, historical, case-control, cross-sectional, before-and-after, expert opinion)
- whether censoring was an issue, in cost or cost-effectiveness studies, and if it was which methods were used to allow for censored cost data
- whether studies that used prognostic models to estimate non-transplant survival HRQL or costs measured the uncertainty around prognostic model estimates

- whether other types of modelling, other than prognostic modelling, e.g. Markov models, were used to estimate non-transplant survival, HRQL and costs and whether uncertainty around model estimates was accounted for
- whether other sources of uncertainty were identified or accounted for, and whether deterministic or probabilistic sensitivity analysis were used to account for this uncertainty
- whether studies had a sample size of less than 30 patients per group[5] (these studies are unlikely to detect significant differences between groups). Study sample size was investigated as a further measure of the quality of the study

The review focuses on transplantation studies that compared vital organ transplantation with the non-transplant management of end-stage organ failure. Studies which compared transplantation of the heart, intestine, kidneys, liver, lungs, pancreas or a combination of more than one of these organs with alternative (non-transplantation) management were included in the review.

The search was limited to six databases:

- Ovid – Medline [OVID, 2005]
- Ovid – Cumulative Index to Nursing and Allied Health Limited (CINAHL) [OVID, 2005]
- BIDS – International Bibliography of the Social Sciences (IBSS) [BIDS, 2005]
- NHS Economic Evaluation Database (NHS EED) [Centre for Review and Dissemination, 2005]
- Office of Health Economics Health Economic Evaluations Database (OHE HEED) [OHE HEED, 2005]
- University of Sheffield's online library catalogue (STAR) [University of Sheffield Library, 2005]

To identify articles in solid organ transplantation the search strategy combined the term "transplant" or "graft" with any of the solid organs listed above ("heart", "intestine", "kidney", "liver", "lung", or "pancreas"). These terms were also combined with those

[5] There is no consensus in the statistical literature on the preferred sample size, in studies where a power calculation is inapplicable. A suggested "rule of thumb" is that the sample size should be a minimum of 30 per arm in order for the central limit theorem to apply (i.e. for sample means to follow a normal distribution irrespective of the distribution of the individual measurements). Although arbitrary, this is conventionally a minimum sample size and is the rule adopted here.

relating to economic evaluation studies or cost studies or studies addressing patient benefits (either outcome or HRQL). Search terms included expressions listed on the Medical Subject Headings (MESH) database [National Library of Medicine, 2005]. Full details of terms searched are listed in Appendix 3.1.

Databases were searched for the period January 1st 1980 to May 31st 2005. Exceptions were made for a series of studies examining survival in the Stanford heart transplant series published prior to 1980. These studies were felt to contain important methodological contributions to the literature on allowing for bias in survival for intervention delay groups [Clark *et al,* 1971; Turnbull *et al*, 1974]. Reviews were excluded, though review articles were obtained to identify sources for original studies. One unpublished study (grey literature) known to the author of this thesis was included as it was felt to address important methodological issues in accounting for uncertainties in non-RCT studies [Longworth *et al*, 2003a]. Studies had to be in English to be included in the review.

An initial screen of the articles was performed, in which the author rejected articles if the article title and abstract (when available) clearly indicated that the article did not compare transplantation with the non-transplant management of vital organ failure. If the title or abstract stated that the article compared the effectiveness, costs or cost-effectiveness of transplantation with the non-transplant management of organ failure the original articles were sought and reviewed to obtain further details of the study design and methodology. A total of 678 articles were sought and reviewed to obtain further details. Articles were excluded from the analysis if:

- they did not compare transplantation with medical management
- they were a review article
- the article was unobtainable[6]

Based upon the article title and abstract, a total of 678 potential articles were listed as suitable comparators of transplantation with a non-transplant cohort (Figure 3.1). Nine per cent (N = 62) of these articles were unavailable in the British Library or the University of Sheffield libraries. This left 616 articles which were obtained and read; of these 12% (N = 74) were review articles and 55% (N = 340) did not compare

[6] Articles were sought from the University of Sheffield Libraries, which include access to a number of NHS libraries in the South Yorkshire region, including teaching hospital libraries, and the British Library. If articles were not found after searching these sources they were considered unobtainable.

transplantation with a non-transplant cohort. A total of 202 articles (33% of those read), covering 158 studies were accepted as suitable comparisons of transplantation with a non-transplant cohort.

Figure 3.1 Flowchart depicting the exclusion process for the 678 articles that were reviewed in the literature review

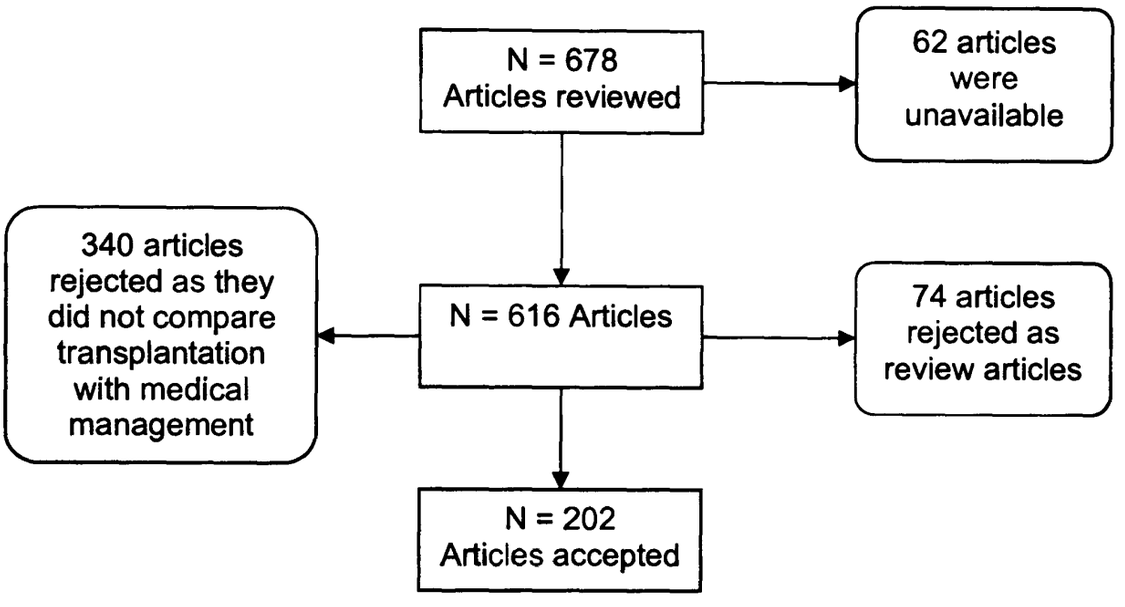


Table 3.1 details the articles included or excluded in the review by organ. Few studies were found in the area of intestinal transplantation and kidney transplant was the most common organ to be studied in this review.

Table 3.1 Summary of 678 accepted and rejected literature review articles by organ

	Accepted	Rejected: Review	Rejected: Not comparing Tx	Rejected: Unable to find	Total
Heart	31	10	49	11	101
Intestinal	4	3	7	2	16
Kidney	75	22	152	23	272
Liver	35	18	92	21	166
Lung	25	6	12	2	45
Pancreas	3	7	11	3	24
Pancreas-Kidney	29	8	17	0	54
Total	**202**	**74**	**340**	**62**	**678**

Tx – Transplantation

Appendix A3.2 gives reference details of the 202 articles included within this review. The remainder of this chapter presents the study results in terms of the number of original studies included in the review (N = 158) and not in terms of the number of articles, since some authors published more than one article describing the same patients.

Figure 3.2 presents a Venn diagram to summarise the number of studies by the database/s the study was identified in. A total of 145 studies (92%) were identified by the Ovid database, 108 from Ovid alone and a further 37 which were identified by Ovid and at least one other database (For example Ovid and CINAHL). Of the other five databases searched in this literature review; NHS EED identified 25 studies (16%), OHE HEED 22 studies (14%), CINAHL 15 studies (9%), STAR 1 study (1%) and BIDS no studies. One unpublished study was also included in this review.

Figure 3.2 **Venn diagram of number of accepted studies by database (N = 158)**

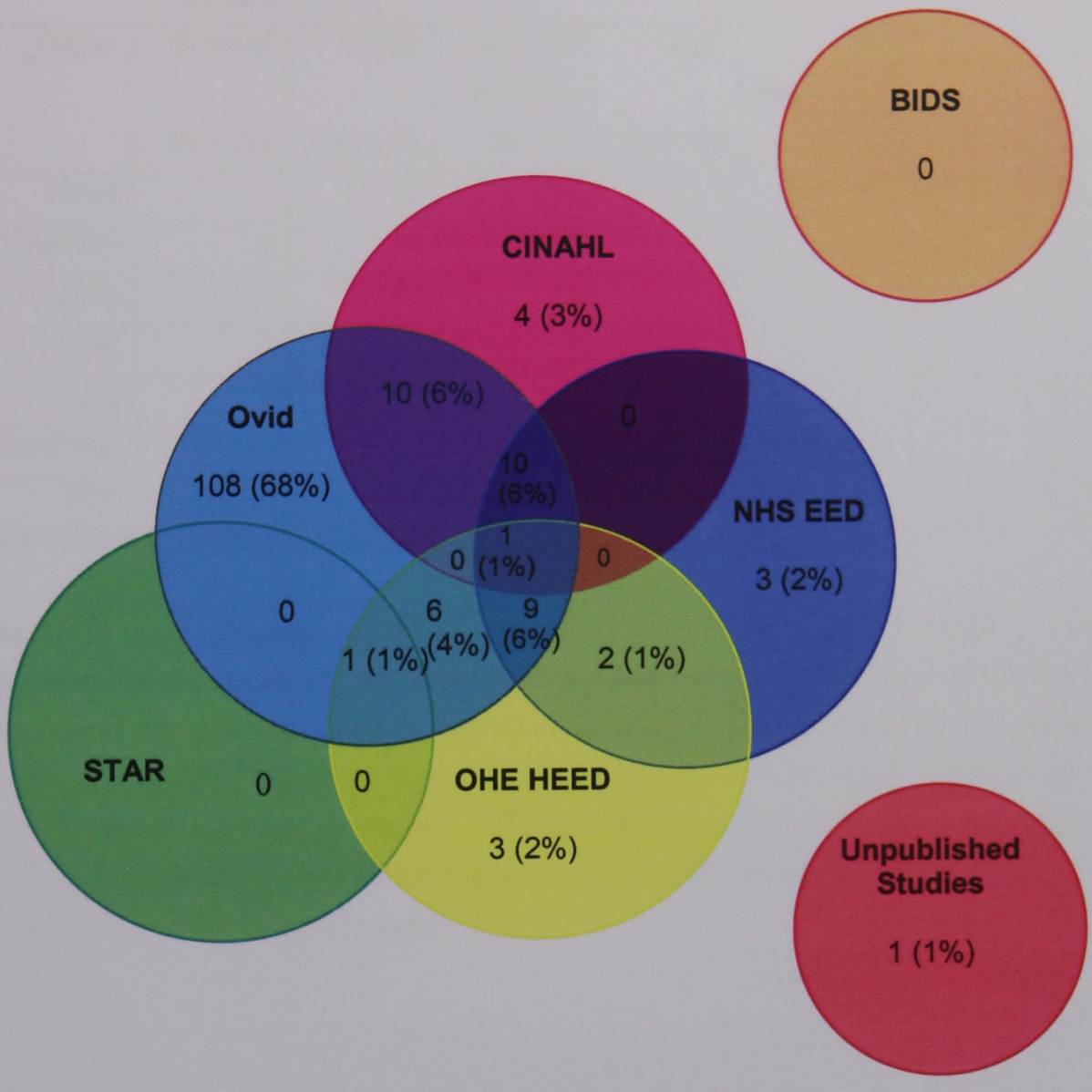


Table 3.2 presents a summary of the literature review results by organ and type of study; survival, HRQL, survival and HRQL, costs or cost-effectiveness studies. Nearly half the studies (47%) compared HRQL only between transplantation and a non-transplant cohort, 15% looked at survival only and seven studies (4%) considered both survival and HRQL. One quarter of the studies presented cost-effectiveness results comparing transplantation with a non-transplant control group, and eight per cent of studies presented costs only.

Table 3.2 Number of studies included in the review by organ and type of study (N = 158)

Organ	Survival (N = 23)	HRQL (N = 75)	Survival & HRQL (N = 7)	Costs (N = 13)	Cost-effectiveness (N = 40)	Total (N = 158)
Heart	5	12	4	1	4	26
Intestine	0	1	0	0	1	2
Kidney	10	25	0	9	20	64
Liver	3	14	0	2	8	27
Lung	2	11	0	0	3	16
Pancreas	0	2	0	0	1	3
SPK	3	10	3	1	3	20

SPK – Simultaneous pancreas kidney transplant

Appendix A3.3 presents a detailed summary of the 158 studies included in this literature review by organ. Sections 3.4.1 to 3.4.4 present a summary of the main findings of the literature review by type of study (survival, HRQL or survival and HRQL, costs or cost-effectiveness). Table 3.3 summarises the different study designs used and Table 3.4 summarises whether the studies dealt with censored costs, modelling, prognostic modelling, used deterministic or probabilistic sensitivity analysis or had a small sample size, by type of study.

Table 3.3 Summary of the study designs used in 158 studies reviewed by type of study (Frequency and percentage reported)

	Survival (N = 23)	HRQL or survival & HRQL (N = 82)	Costs (N = 13)	Cost-effectiveness (N = 40)	Overall (N = 158)
Quasi-experimental cohort	14 (61%)	28 (34%)	11 (85%)	28 (70%)	81 (51%)
Intervention delay cohort	8 (35%)	1 (1%)	1 (8%)	8 (20%)	18 (11%)
Historic cohort	0 (0%)	0 (0%)	0 (0%)	1 (3%)	1 (1%)
Case-control	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Cross-section	0 (0%)	9 (11%)	0 (0%)	0 (0%)	9 (6%)
Before & After	0 (0%)	34 (41%)	1 (8%)	0 (0%)	35 (22%)
Expert opinion	0 (0%)	1 (1%)	0 (0%)	0 (0%)	1 (1%)
Other	1* (4%)	9# (7%)	0 (0%)	3+ (8%)	13 (8%)

* Historical cohort combined with intervention delay group

Five studies used both quasi-experimental and before and after groups, three studies used cross sectional and before and after studies and one study used quasi-experimental and intervention delay groups.

+ Two studies used intervention delay combined with a historical control group and one study used intervention delay combined with a historical control group. Of the three studies, one conducted a further intervention delay group analysis

Table 3.4 **Summary of the issues of statistical uncertainty used in 158 studies reviewed by type of study (Frequency and percentage reported)**

	Survival (N = 23)	HRQL or survival & HRQL (N = 82)	Costs (N = 13)	Cost-effectiveness (N = 40)	Overall (N = 158)
Censoring of costs a potential issue	N/a	N/a	13 (100%)	23 (56%)	33 (21%)
Censored costs adjusted for	N/a	N/a	0 (0%)	3 (8%)	3 (2%)
Prognostic models used	1 (4%)	0 (0%)	0 (0%)	3 (8%)	4 (3%)
Other models used	13 (57%)	3 (4%)	2 (15%)	20 (50%)	38 (24%)
Deterministic sensitivity analysis used	1 (4%)	1 (1%)	1 (8%)	20 (50%)	23 (15%)
Small sample size	1 (4%)	29 (35%)	2 (15%)	9 (23%)	41 (26%)

N/a Not applicable

3.3.1 Survival Only Studies (N = 23)

Failure to adjust for case mix factors (i.e. factors known to affect the survival outcome) which differ between treatment and control groups can bias the survival estimates either in favour of, or against, transplantation. Adjustment for case mix factors can be made using modelling techniques, for example the Cox PH model. Over 60% of the survival studies included in this review made adjustments for case mix factors shown to affect survival. The remaining nine studies failed to adjust for case mix factors, potentially introducing bias in to their survival estimates.

Eight studies used an intervention delay waiting list cohort – this cohort is unlikely to be representative of the population of patients with end-stage organ failure who are eligible for transplantation for the reasons set out in Section 3.2.2.1. Fourteen studies

used a quasi-experimental control group. Only one study used a prognostic model approach to estimate survival in the absence of transplantation, combining information from a historical cohort with an intervention delay group and this study failed to account for prognostic model uncertainties.

3.3.2 HRQL Only (N = 75) and Survival and HRQL Studies (N = 7)

Over half of the studies (52% [82/158]) included in this review compared HRQL with, and in the absence of, transplantation, with seven studies reporting both survival and HRQL results. Twenty-eight HRQL studies (34%) compared transplantation with a quasi-experimental control group, 34 (41%) used a before and after study design, 9 (11%) a cross sectional study design, one study used an intervention delay group and one used expert opinion (the only study in the literature review to use this approach) and the remaining nine studies used more than one type of study design.

The cross sectional studies asked control group and transplant patients about their HRQL at one time point only. These studies would not obtain a representative sample of non-transplant patients since they do not capture the HRQL of patients who died during the study. A similar selection process was true of the before and after studies that profiled HRQL over time; many of these studies compared patients who were well enough to respond to a questionnaire at all study time points. Only a handful of studies included all responders.

Four of the HRQL studies will be subject to recall bias as transplanted patients were asked to recall what their HRQL had been like before transplantation. A further study asked experts to value patients HRQL.

3.3.3 Cost Studies (N = 13)

Eleven of the 13 cost studies compared transplantation with a quasi-experimental control group. Of the remaining two studies, one compared transplantation with an intervention delay group and one used the before and after study design. All studies used observed data, though one kidney transplant study compared observed transplant costs with costs for a hypothetical cohort of dialysis patients, but provided no details regarding the characteristics of this cohort.

None of the cost studies gave details to indicate whether patients were lost to follow-up (censored) during the study, therefore, all thirteen studies were potentially subject to

censored cost data, yet none of the studies accounted for mean total costs in the presence of censoring. Only one study carried out one-way sensitivity analysis to vary assumptions made in the main cost analysis.

3.3.4 Cost-Effectiveness Studies (N = 40)

Eight studies compared transplantation with an intervention delay group. A further three studies combined data from an intervention delay group with historical data. All three studies used prognostic models based on historical cohorts to estimate survival in the absence of transplantation and information from the waiting list to estimate non-transplant utilities and costs. A further 15 studies (38%) used modelling techniques, for example Markov modelling or simulation modelling, to estimate the cost effectiveness of solid organ transplantation.

De By and colleagues compared a cohort of transplant patients whose graft had failed with a cohort of dialysis patients [De By *et al,* 1982]. This study is likely to be biased towards the dialysis cohort as patients with graft failure will tend to have higher mortality and incur more costs than patients with a non-failing kidney graft.

Censoring occurs when a proportion of study patients do not experience an event of interest (e.g. death) during the study period and have incomplete data for the full study period. Ignoring censoring can result in biased estimates of survival, HRQL or costs. Statistical methods exist to allow for censoring in time to event survival studies [Collett, 1994]. However, these techniques are not applicable to cost or HRQL data due to underlying methodological assumptions (See Chapter 4 for further details) [Hallstrom & Sullivan, 1998]. Three (8%) of the forty cost-effectiveness studies adjusted for censored cost data. All three studies used the Kaplan-Meier product limit estimator, also known as Lin's method, where mean costs per time period are weighted by the Kaplan-Meier probability of survival and summed over time [Lin *et al,* 1997]. This method is explained in further detail in Chapter 4.

None of the studies that chose to use prognostic models to estimate non-transplant survival made an adjustment for model parameter uncertainty. One paper, which was part of Longworth and colleagues' liver transplant study [Longworth *et al,* 2003], used discrete event simulation modelling to estimate the cost-effectiveness of liver transplantation [Ratcliffe *et al,* 2001]. In discrete event simulation it is possible to specify distributions, central estimates (e.g. means, medians) and uncertainties around

distributions using ranges and standard errors, this is analogous to PSA. However, although this paper did allow for uncertainty around resource use parameters it assumed that prognostic model parameters were fixed and thus failed to account for prognostic model uncertainties [Personal communication with authors].

The authors of three studies felt that costs increased for a time period immediately prior to death and made an adjustment to estimate non-transplant costs accordingly. Additionally, Anyanwu and colleagues and Van Enckevort *et al,* also adjusted for costs for a period prior to death in the transplant cohort; both studies extrapolated data beyond the observed study period [Van Enckevort *et al,* 1997; Anyanwu *et al,* 2002].

3.4 DISCUSSION

The RCT remains the preferred method for evaluating the effects, cost and cost-effectiveness of medical interventions. Despite this, solid organ transplantation is one therapeutic intervention for which this design is inappropriate, both for ethical and practical reasons. Alternatives to the RCT that could be used in comparator studies of solid organ transplantation with an alternative treatment cohort were presented in Section 3.2. For each study design cautions in the application, strengths and weaknesses were discussed. Section 3.3 followed this with a review of the current literature, giving details on the designs that had been used in practice together with details of other issues that were encountered.

It was disappointing to note that the majority of studies in this review compared transplantation with an inappropriate cohort of non-transplant patients. Half of the studies included in the literature review compared organ transplantation with a quasi-experimental control group. The majority of these studies (N = 53) were kidney transplant studies where dialysis is offered as an alternative treatment for end-stage renal failure. However, few of the studies that included a quasi-experimental control group compared cohorts of patients with end-stage organ failure, and instead compared cohorts undergoing transplantation or an alternative treatment. For example, an obvious alternative treatment to kidney transplantation is dialysis, but dialysis is also given to patients who have other (less chronic) kidney diseases. Including these patients in a comparison cohort for patients with end-stage organ failure creates a control cohort of patients who are likely to be healthier than a cohort of patients undergoing transplantation. A number of studies took this approach and therefore would have underestimated the effectiveness of kidney transplantation, although some

studies did consider this and used concurrent cohorts of dialysis patients listed for kidney transplantation [Chantler *et al,* 1980; Laupacis *et al,* 1993; Laupacis *et al,* 1996; Fujisawa *et al,* 2000].

Only one study used expert opinions to estimate HRQL in the absence of transplantation [Mai *et al,* 1990]. Given the lack of evidence of the reliability of expert opinions, the results from this study are highly likely to be biased [Black *et al,* 1999]. No study used the case-control study design, which is unsurprising given the difficulties of designing and drawing meaningful conclusions from such a study.

Faced with a choice of possible study designs the approach of preference would be the use of a quasi-experimental control group of non-transplant patients with end-stage organ failure who meet the criteria for transplantation. However, the expected sample size of a quasi-experimental control group should be considered with this type of study, and 14 (18%) of 79 concurrent studies used cohorts consisting of less than 30 cases. If the sample size is likely to be small the study is unlikely to detect statistical differences between groups, or to reliably estimate effectiveness, costs or cost-effectiveness.

In the absence of a suitable cohort of quasi-experimental controls I would probably explore the use of a historical control group over an intervention delay group, and five studies in the literature review used this approach [Williams *et al,* 1987; Bonsel *et al,* 1990a; Bonsel *et al,* 1990b; Christensen *et al,* 1999; Liemann Garcia *et* al, 2001; Longworth et al, 2003; Longworth et al, 2003a]. I believe such a cohort is still preferable to an intervention delay group of patients on the waiting list, an intervention delay cohort is unlikely to be representative of the general population of non-transplant patients with end-stage organ failure who meet transplant listing criteria (Section 3.2.2.1).

If the data are available, statistical modelling can be used to adjust for differences between the treatment cohort and either the quasi-experimental, intervention delay or historical control groups. An alternative strategy would be to individually match patients in the treatment group to similar patients in the control arm. A few studies in this review did this. However, matching can reduce the sample size of a study, as patients without a suitable match are excluded from the analysis.

3.4.1 Estimating Mean Total Costs in the Presence of Censoring

None of the HRQL studies tackled the issue of censored data. Only one study considered the HRQL of patients who died, where patients who died during the study period were given a value of zero [Ratcliffe *et al,* 2001; Longworth & Bryan, 2003]. Most of the HRQL studies chose to compare the HRQL of patients who remained alive throughout the study and the HRQL of patients who died was ignored, biasing the results towards healthier patients. A few studies chose to include patients who dropped out of the study and Caine and colleagues used the last value carried forward approach to allow for drop outs [Caine *et al,* 1996].

It has already been highlighted that many of the HRQL studies excluded patients with incomplete (censored) data and it can be assumed that many of the cost and cost-effectiveness studies comparing transplantation with a non-transplant cohort over a fixed time period also ignored the issue of censoring. A total of 33 studies could have been subject to censored cost data and could have adjusted for it. However, only three studies did so, each of which used Lin's method to adjust for the incomplete follow-up of a proportion of their patients [Ohi *et al,* 1986; Garner & Dardis, 1987; Longworth *et al*, 2003a]. Numerous techniques have been proposed in the literature for estimating mean total costs in the presence of censoring and these will be reviewed in Chapter 4. The censored cost methods introduced in Chapter 4 can be applied in any study where censoring is an issue (including RCTs) and can also be applied to censored HRQL data.

3.4.2 Measuring Prognostic Model Parameter Uncertainty

In the absence of a suitable observed control group, four of the studies in this review used prognostic models to estimate survival in the absence of transplantation [Bonsel *et al,* 1990a; Bonsel *et al,* 1990b; Christensen *et al,* 1999; Liemann Garcia *et al,* 2001; Longworth *et al,* 2003; Longworth *et al*, 2003a]. Prognostic models are usually based on cohorts other than the one they are being applied to when estimating control group survival and prognostic model regression coefficients are actually a mean coefficient for a cohort. These model coefficients are therefore estimates and the uncertainty around the mean regression coefficients is represented by the regression coefficient standard error. All the studies included in this literature review assumed that there was no uncertainty around the prognostic model and applied the mean regression coefficients when estimating control group survival. A method is set out in Chapter 5 to incorporate prognostic model parameter uncertainty.

3.4.3 Estimating Individual Patient Outcomes from Prognostic Models and the Uncertainty Surrounding the Outcomes

In this review, three studies assumed that costs increase for a time period immediately prior to death, where it was assumed that patients were likely to be treated more resource intensively and an adjustment was made accordingly [Van Enckevort *et al,* 1997; Anyanwu *et al,* 2002; Longworth *et al*, 2003]. Methods based upon observed data or modelling techniques were applied to estimate costs incurred in the time period prior to death.

Extrapolation techniques must be applied to establish time of death using modelling techniques. If a fixed study time period is observed, uncertainties exist in establishing methods for defining the exact point of death and uncertainty will exist around the estimated time point. Chapter 6 explores a number of techniques for estimating uncertainties when predicting outcomes in a control group of patients over a fixed study period.

3.4.4 Collectively Measuring Multiple Sources of Uncertainty Relating to the use of Prognostic Models Instead of an Observed Non-transplant Cohort

A total of 23 studies conducted deterministic sensitivity analysis to test assumptions made when estimating the costs, effectiveness or cost effectiveness with and without transplantation. In all 23 studies the sensitivity analysis was the basic one-way. Occasionally, extreme scenarios were explored within the one-way analysis. However, none of the studies included in this literature review considered multi-way or PSA to explore the effects of varying more than one assumption at a time. Additionally, no study considered whether varying one assumption would have an impact on other assumptions made within a study.

Although this is a complex issue, techniques exist to enable researchers to conduct multi-way and PSA [Doubilet *et al,* 1985]. Chapter 7 focuses on measuring multiple sources of uncertainty from applying a prognostic model approach to estimate survival, HRQL and costs in the absence of a suitable observed cohort of non-transplant patients. PSA is applied in order to consider the impact of prognostic model uncertainties on survival, HRQL and cost estimates.

3.5 CONCLUSIONS

This chapter has identified some of the problems and uncertainties associated with non-RCT study designs. More specifically, it has focussed on the problems and uncertainties associated with the appropriate selection of a non-transplant cohort of patients in the evaluation of the effectiveness, costs or cost-effectiveness of solid organ transplantation. It is important that any issues associated with the choice of study design are considered at the design stage and failure to do this could result in a non-transplant cohort of patients that is either unrepresentative of the population of interest or not comparable to the transplant cohort (or both).

The solid organ transplant studies identified in the literature review adopted a number of designs and methods to estimate control group survival, HRQL or costs and the uncertainties associated with these estimates. However, it was disappointing to observe that the majority of studies failed to consider or acknowledge potential bias in their studies as a result of the choice of their study design.

Despite this, it should also be pointed out that some of these problems are not specific to the above types of study, since problems with unrepresentative cohorts and low sample sizes in particular may arise even in an RCT. Clearly, no solution is going to be perfect and any of the comparator groups selected are going to have inherent difficulties and so any evaluation will have to make some assumptions in order to estimate the effectiveness, costs or cost-effectiveness of solid organ transplantation. The best studies will give consideration to the criteria mentioned above, acknowledge the difficulties encountered within the study and discuss the possible implications these may have on the results. The plausibility of any assumptions made will raise questions regarding the reliability of the study results, and as a consequence studies should also highlight any assumptions made and perform a deterministic or probabilistic sensitivity analysis around any assumptions that are particularly important or controversial.

The literature review confirmed that estimating costs in the presence of censoring and the use of prognostic models were important issues to be addressed in the CELT study and were also common problems inadequately addressed in many solid organ transplant studies. Therefore, the remainder of this thesis will specifically explore the following issues:

- estimating mean total costs in the presence of censoring (Chapter 4)
- measuring prognostic model parameter uncertainty (Chapter 5)
- estimating individual patient outcomes from prognostic models and the uncertainty surrounding the outcomes (Chapter 6)
- collectively measuring multiple sources of uncertainty related to the use of prognostic models instead of an observed non-transplant cohort (Chapter 7)

CHAPTER 4

ESTIMATING MEAN TOTAL COSTS IN THE PRESENCE OF CENSORING

4.1 INTRODUCTION

The main findings of the CELT study were based on a follow-up period of 2.25 years after point of assessment [Longworth *et al,* 2003]. In the original study, only six per cent of ALD, PBC and PSC transplant patients were lost to follow-up (censored) prior to the end of the 2.25 year study period and censoring was ignored. It was anticipated that the proposed extension to the CELT study, over a five-year post-assessment follow-up period, would increase the proportion of data censored before the end of the study rather than censored at five years. It was felt unwise to ignore censoring for the extended follow-up data, prompting the question: what methods can be used to analyse censored data and which of them work well? This provided the motivation for the research carried out and reported in this chapter.

The simplest method for allowing for censored survival data is the Kaplan-Meier product limit method, which makes no assumptions about the underlying distribution of

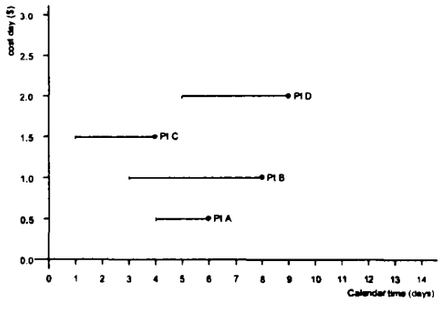
the data [Kaplan & Meier, 1958; Collett, 1994]. However, like most other survival methods, including the Cox PH regression model [Cox, 1972]; it does assume that survival times are independent of the method or reason of censoring. This means that a patient who is censored at a particular time point should be representative of all other patients who have survived to that time point. If this assumption does not hold then censoring is said to be informative and these survival analysis methods become biased, with the degree of bias depending on the proportion of informatively censored patients in the study.

In evaluations of health care interventions the problem of censoring is not unique to survival or other health outcomes; it is also relevant to costs where information may be similarly censored. Unfortunately, the problem of censoring is not solved simply by using methods such as the Kaplan-Meier method and applying it to censored cost data. For the Kaplan-Meier methodology to be appropriately applied costs should be independent of the method or reason for censoring and a censored cost should be representative of all other costs that size or greater. Hallstrom and Sullivan have clearly illustrated the problems in using the Kaplan-Meier estimate on a cost scale [Hallstrom & Sullivan, 1998]. They illustrate the problem by considering the effect of censoring on survival and costs among four hypothetical patients, as demonstrated in Figure 4.1 (reproduced from their paper with kind permission of the journal Medical Care [Medical Care (1998); 36(3): 433-436]).

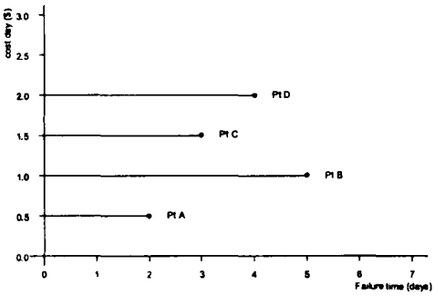
Panel A of Figure 4.1 presents the cost and survival histories for four patients:

- patient A dies after 2 days and incurs costs at a rate of $0.5 per day with a total study cost of $1
- patient B is censored after 5 days and incurs costs at a rate of $1 per day with a total study cost of $5
- patient C is censored after 3 days and incurs costs at a rate of $1.5 per day with a total study cost of $4.5
- patient D is censored after 4 days and incurs costs at a rate of $2 per day with a total study cost of $8

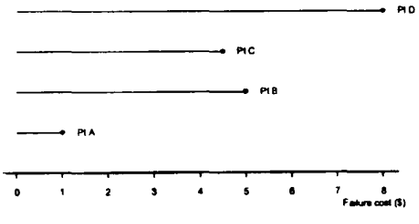
Figure 4.1 **Transcribing the real time costs and survival of four patients (panel A) to the failure time scale (panel B) and failure cost scale (panel C) and computing the Kaplan-Meier failure time estimates (panel B) and the Kaplan-Meier analogue of failure cost estimates (panel C)**



A



	$\hat{S}_t$	1	1	¾	¾	¾
B	$\hat{c}_t$	5/4	5/4	3/2	3/2	1



C	$\hat{S}_t$	1	¾	¾	¾	¾	¾	¾	¾

Panel A of Figure 4.1 depicts the study start time and the time each patient entered and left the study. Panel B shows the total time each patient was in the study (as the length of the horizontal line) and the Kaplan-Meier probability of survival ($\hat{S}_t$) from time zero to five days ($\hat{c}_t$ will be defined later). All patients remain in the study up to day 2, giving a Kaplan-Meier survival probability of one. On day 2 patient A dies, giving a Kaplan-Meier survival probability of 0.75. The final panel of Figure 4.1 shows the total costs incurred by each patient over the study period, where the horizontal line is equal to the total cost (daily cost multiplied by time spent in the study) and $\hat{S}_t$ refers here to

the probability of incurring a cost greater than or equal to the value on the horizontal axis.

If the Kaplan-Meier method is directly applied to the cost scale, the probability of incurring a cost of at least $1 during the study period is one. Thereafter, patient A dies leaving three censored patients in the study with total study costs of at least $2, and with a probability of incurring a cost of at least $2 of 0.75. The mean costs for the study would then be estimated by ($\hat{S}_t$ x daily cost):

$$(1x1) + (0.75x1) + (0.75x1) + (0.75x1) + (0.75x1) + (0.75x1) + (0.75x1) + (0.75x1)$$
$$= \$6.25$$

Panel C also shows why the assumption of independent censoring does not hold for censored cost data. Consider patient C, whose costs are censored at $4.5. This patient is not representative of the two patients, B and D, who are censored after $4.5, because patient B incurs costs at a lower rate than patient C and patient D incurs costs at a higher rate.

Hallstrom and Sullivan pointed out that a more accurate estimate of mean total costs could be obtained by multiplying the average cost per day ($\hat{c}_t$) by the Kaplan-Meier survival probability of being alive on that day (Panel B). Thus, the mean study costs for patients A to D are calculated from time intervals one through to five respectively by:

$$(1x5/4) + (1x5/4) + (3/4x3/2) + (3/4x3/2) + (3/4x1) = \$5.5$$

This method is referred to in the literature as either Lin's method or as the Kaplan-Meier product limit estimator. Note, this method is **not** the same as calculating the Kaplan-Meier estimator on the cost scale directly, which to avoid confusion will hereafter be referred to as the Kaplan-Meier cost method. The method of dividing the time in to intervals will be referred to as Lin's method, and both will be discussed in the subsequent section of this chapter (Section 4.2).

In the remainder of this chapter, costs are defined as censored if they are incomplete for the study period of interest. Costs are defined as complete if either:

- a patient dies during the study period, or

- a patient is alive at the end of a study and has complete costs for the full study period

Otherwise, the cost is deemed censored.

This chapter describes a series of alternative methods for estimating mean total costs in the presence of censoring (Section 4.2), focusing on data that are right censored[1]. An earlier version of this chapter has been published by the author elsewhere [Young, 2005].

Examples of censoring mechanisms include patients who decide that they no longer wish to participate in a study and thus drop out of the study before it ends, or patients who are lost to follow-up as they move away from the region the study is conducted in. Although there are several methods in the literature that can be applied to censored cost data, there are currently no studies that compare multiple causes of censoring and evaluate how different censoring mechanisms may affect the "accuracy" of the estimate of the total average costs. Section 4.3 goes on to explore different reasons why cost data might be censored.

To date no review of all existing methods for estimating mean total costs in the presence of censoring exists and, faced with a series of methods, it is not obvious which one will give the most accurate estimate of mean total costs in the presence of censoring. Therefore, Sections 4.4 to 4.6 explore an issue of methodological uncertainty. The methods identified in Section 4.2 will be compared across four different censoring mechanisms – random censoring, end-of-study censoring, informative censoring and partial censoring – by simulating the censoring mechanisms from a complete cohort of patients included in the CELT study. Section 4.7 will present the results from the simulation study and the findings will be discussed in Section 4.8.

4.2 A REVIEW OF METHODS FOR ESTIMATING MEAN TOTAL COSTS IN THE PRESENCE OF CENSORING

A literature review was conducted in order to identify existing methods for estimating mean total study costs in the presence of censored data. The following databases were searched: Ovid Medline [Ovid, 2005], BIDS Social Science database [BIDS, 2005],

[1] Right censoring occurs when the event of interest occurs at some unknown time point "to the right of the last known survival time" [Collett, 1994]. In the context of cost data, the true (but unobserved) cost is therefore greater than the censored (observed) cost.

NHS EED [Centre for Review and Dissemination, 2005] and OHE HEED [OHE HEED, 2005].

The same identification strategy detailed in Chapter 3 was applied here; titles were assessed and possible relevant articles identified, abstracts from the relevant articles were then reviewed and further inappropriate articles were eliminated. Finally, after evaluating the abstracts, all possibly relevant articles were obtained and read. Articles were included in the review if they proposed original methods for estimating mean study costs in the presence of censoring. Articles were rejected if they:

- reviewed other previously published methods
- treated censoring as a missing data problem rather than specifically allowing for censored data
- presented methods for estimating QALYs in the presence of censoring
- presented methods for estimating cost-effectiveness in the presence of censoring

References from articles related to censored cost data were followed up and included in the review if they proposed original methods for estimating mean study costs in the presence of censoring. Appendix A4.1 presents details of the search terms used and the search results.

A total of 12 different methods for handling censored cost data, including methods that ignore censoring, were identified. These methods were: ignoring censoring, ignoring censored cases, Kaplan-Meier cost method [Fenn, 1995], Cox PH cost method [Fenn, 1996], the partitioned Cox cost method [Lipscomb *et al,* 1998], Lin's methods with either known cost histories (KCH)[2] or unknown cost history (UCH) [Lin *et al,* 1997], the weighted cost method with KCH or UCH [Bang & Tsiatis, 2000], Lin's regression method with KCH or UCH [Lin, 2000], and Carides' regression method [Carides *et al,* 2000]. These methods are outlined in detail below.

4.2.1 Ignoring Censoring (IC)

The average study costs are simply calculated in the same way as one would calculate any mean, by ignoring the fact the data are censored and assuming each patient's cost

[2] Cost data are said to have cost histories when the detail of resource collection is such that information on the time each resource is used is collected alongside information on the quantity of each resource.

was complete. Equation 4.1 presents the mathematical formula for estimating mean study costs ($\overline{C}_{IC}$) when ignoring censoring, where C_i is the total cost for patient i (where i = 1...N), and N denotes the total number of patients in the study.

$$\overline{C}_{IC} = \frac{1}{N}\sum_{i=1}^{N} C_i \qquad \textit{Equation 4.1}$$

4.2.2 Ignoring the Censored Cases (CC)

A further simple technique for estimating mean study costs is to ignore the censored cases. The average cost is calculated across a reduced sample (N_{CC}) of uncensored cases (Equation 4.2).

$$\overline{C}_{CC} = \frac{1}{N_{CC}}\sum_{i=1}^{N_{CC}} C_i \qquad \textit{Equation 4.2}$$

4.2.3 Kaplan-Meier Cost Method (KM)

The Kaplan-Meier cost method was briefly introduced in Section 4.1, and is described in further detail here. The method has been applied by several authors in order to calculate average study costs [See: Hay, 1989, Hiatt *et al*, 1990; Fenn *et al*, 1995]. The Kaplan-Meier cost method uses a cost rather than a time scale and assumes that individual costs are independent of each other and that censored costs are independent of the censoring mechanism.

Costs are sorted in to ascending order and divided in to k intervals, where the first interval starts at zero and ends at the smallest fully observed (i.e. uncensored) cost; the smallest cost is included in this interval. The second interval begins immediately after the first uncensored cost and ends at the second smallest uncensored cost. This process is repeated up to the final interval (k), which begins after the penultimate uncensored cost and ends at the largest uncensored cost. Equation 4.3 presents the formula for calculating the probability of incurring a cost of at least c_k (k = 1 to K), where n_k represents the total number of individuals whose cost is at least c_k, and d_k represents the number of individuals whose cost is known to be between c_{k-1} and c_k who die during this interval. Individuals with censored costs are included in an interval (c_{k-1}, c_k) only if the censored cost is c_k or greater.

$$\hat{K}(c_k) = \prod_{k=1}^{K} \frac{n_k - d_k}{n_k}$$ **Equation 4.3**

The mean cost for the study can then be estimated by multiplying the probability of incurring at least cost c in interval k by the additional cost in interval k by the additional costs occurring in interval k and summing these over all intervals (Equation 4.4), i.e. calculating the area under the Kaplan-Meier cost curve.

$$\overline{C}_{KM} = \sum_{k=1}^{K} \hat{K}(c_k)(c_k - c_{k-1})$$ **Equation 4.4**

4.2.4 Survival Model Method (Cox)

Several authors have suggested using survival modelling techniques to estimate mean total costs in the presence of censoring [Hay, 1989, Dudley *et al,* 1993, Fenn *et al,* 1996, Etzioni *et al,* 1999]. A number of survival models exist for modelling censored data, the most common of which is the Cox PH model. Other choices exist including Weibull or Gompertz models, both of which make more distributional assumptions of the data. In this chapter the Cox PH model is fitted to illustrate the use of modelling techniques when estimating mean total study costs in the presence of censoring.

One of the attractions of the survival model method is that any explanatory variables that are believed to affect survival and costs can be adjusted for within the models. As with the Kaplan-Meier cost method (Section 4.2.3), it is possible to fit a Cox PH model, using a cost rather than time scale. The hazard of accumulating costs at cost level c are related to the explanatory covariates included in the Cox PH model. The probability of a patient incurring at least cost c can then be calculated using Equation 4.5.

$$K(c, X) = \{K_0(c)\}^{\exp(R_C - R_{C0})}$$ **Equation 4.5**

where R_C denotes a patients "risk" of incurring costs and is calculated using Equation 4.6.

$$R_C = \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p$$ **Equation 4.6**

where X_1, X_2, ..., X_p are the explanatory variables (also known as prognostic variables or covariates) included in the model and β_1, β_2, ..., β_p are the regression coefficients.

Returning to Equation 4.5, R_{C0} corresponds to a hypothetical individual with the average values for covariates X_1, ..., X_p and $K_0(c)$ is the cost function for the hypothetical patient with risk cost R_{C0}. An individual patient profile of costs against the probability of incurring cost c can be plotted and the area under the profile calculated to obtain the individual patient's expected total study cost. The average total cost for a cohort can then be calculated.

The following assumptions should hold for the Cox model:

- individual costs should be independent of each other
- censored costs should be independent of the censoring mechanism
- censoring must also be independent of costs within each level of the covariates fitted in the models (e.g. if gender is included as a covariate in the model then censoring must be independent for male patients and female patients)
- the explanatory covariates fitted in the model should be proportional across costs, so the hazard of incurring a cost c in female patients should be the same proportion as the hazard of incurring costs in male patients for all c

4.2.5 Partitioned Cox Method (PC)

Lipscomb and colleagues proposed extending the Cox PH method (Section 4.2.4) using a partitioned Cox PH model using details from patients' cost histories [Lipscomb *et al*, 1998]. The study time period is first divided in to k smaller time intervals, which need not have identical interval lengths. A Cox PH model, using a cost rather than a time scale, is then fitted within each interval. Costs are censored if they are incomplete for the whole interval, but otherwise are complete. The costs of patients who die during the study are excluded from the interval when they die but are included in all prior intervals.

The mean total cost is obtained for each interval by plotting individual patient profiles of each patient's probability of incurring cost c. The area under each profile is calculated to obtain patients expected total interval costs. These costs are summed across intervals and the mean study costs in the presence of censoring are obtained (Equation 4.7).

$$\overline{C}_{PC} = \sum_{k=1}^{K}\left(\int_0^{\infty} K_{k0}(c)^{\exp(R_{kC}-R_{kCo})}\,dk\right) \qquad \textit{Equation 4.7}$$

where $K_{k0}(c) = \exp[-\Lambda_{k0}(c)]$ and $\Lambda_{k0}(c)$ is the baseline cumulative hazard function in interval k, R_{kc} a patients "risk" of incurring costs in interval k (Equation 4.6) and R_{kc0} the average patient's risk of incurring costs for interval k.

4.2.6 Lin's Method – KCH (LK)

Lin's method had previously been used in several studies [See: Hodgson, 1992; Etzioni *et al,* 1996], before Lin *et al* published a paper studying its properties [Lin *et al,* 1997]. The method described here assumes that patient cost histories are known.

The study time period (of length K) is divided in to a series of smaller equal time intervals, $a_1, a_2,...,a_k$, k = 1,...,K. For each of the time intervals the average total interval cost ($\hat{C}_k$) incurred by all patients alive at the start of the interval is calculated. The Kaplan-Meier survival function ($\hat{S}_k$) is also calculated for each interval. The total average cost per interval is then multiplied by the survival estimate, and these are summed over time intervals in order to estimate the average total cost for the study period (Equation 4.8).

$$\overline{C}_{LK} = \sum_{k=1}^{K} \hat{S}_k \hat{C}_k \qquad \textit{Equation 4.8}$$

4.2.7 Lin's Method – UCH (LU)

In some economic evaluations patient cost histories are unknown and therefore Lin's method (KCH) cannot be used to estimate mean study costs in the presence of censoring (Section 4.2.6). Lin *et al* offer an alternative method for estimating mean study costs when patient cost histories are unknown [Lin *et al,* 1997]. The study period is again divided in to k smaller time intervals, k = 1,...,K, where K + 1 = τ denotes the end of the study period. The costs of patients censored before the end of the study are excluded from the mean total cost estimate. Within intervals k = 1 to K, $\hat{A}_k$ are calculated as the average total cost for patients who die in the interval $[a_k, a_{k+1})$. The average total costs ($\hat{A}_{K+1}$) for interval a_{K+1} are calculated for patients who die during the final interval and those with complete study costs. For each interval the average cost is

multiplied by the difference between the Kaplan-Meier survival estimates at the beginning ($\hat{S}_k$) and end ($\hat{S}_{k+1}$) of the interval – this difference is the probability of dying during the interval. The mean study costs are then estimated by summing the weighted costs over the complete study time period (Equation 4.9).

$$\overline{C}_{LU} = \sum_{k=1}^{K+1} \hat{A}_k (\hat{S}_k - \hat{S}_{k+1}) \qquad \textit{Equation 4.9}$$

The accuracy of the method improves by increasing the number of time intervals (a_k). However, there should be a reasonable number of observed deaths per interval in order to obtain a reliable average cost estimate for $\hat{A}_k$ [3].

4.2.8 Bang and Tsiatis' Weighted Cost Method – UCH (WU)

In 1997, Zhao and Tsiatis published methodology for estimating mean QALYs after adjusting for censored QALY data [Zhao & Tsiatis, 1997][4]. Bang and Tsiatis adapt this methodology to censored cost data using a weighted complete case estimator [Bang & Tsiatis, 2000]. The method described here is applied when cost histories are unknown.

The overall study costs (C_i) for patients who died during the study or had complete costs information up to the end of the study are weighted by the reciprocal of the Kaplan-Meier survival probability estimator ($1/\hat{S}(T_i)$), using reverse censoring. In conventional survival analysis censored observations would have an indicator value of zero and deaths a value of one. Here, the censoring pattern is reversed and censored observations have an indicator value of one and deaths a value of zero and the Kaplan-Meier estimator is calculated on the time scale in the usual way, but calculating the time to censoring rather than the time to event.

Weighted costs are then summed across all patients. The total summed costs are divided by the overall study sample size N (where the sample size includes the censored cases) to obtain the mean total cost estimate in the presence of censoring (Equation 4.10).

[3] Based upon simulation work carried out by the authors they recommend that there should be at least five deaths per interval in order to obtain an accurate estimate of mean total costs [Lin *et al*, 1997].
[4] As with cost data, the Kaplan-Meier survival estimator should not be applied directly to a QALY scale because censoring is not independent of the quality adjusted survival times.

$$\overline{C}_{WU} = \frac{1}{N}\sum_{i=1}^{N}\frac{\Delta_i C_i}{\hat{S}(T_i)} \qquad \textit{Equation 4.10}$$

$\Delta_i = I(T_i \leq U_i)$ is an indicator function that equals one if time to death or the end-of-study (T_i) is less than or equal to time to censoring (U_i) and equals zero otherwise.

4.2.9 Bang and Tsiatis' Weighted Cost Method – KCH (WK)

An extension to the weighted cost method (UCH) makes use of patient cost histories over the study period [Bang & Tsiatis, 2000]. As with the partitioned Cox method and Lin's method (KCH) the study time period is divided in to a number of smaller intervals (k = 1...K-1). Within each interval, the interval costs for patients who die during the interval or have complete interval costs are weighted by the Kaplan-Meier survival estimator for the interval[5] using reverse censoring. Weighted estimates are summed across intervals and patients and are divided by the total sample size for the cohort (N) to obtain an estimate of mean study costs in the presence of censoring (Equation 4.11).

$$\overline{C}_{WK} = \frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{K}\frac{\Delta_i^k c_i^k}{\hat{S}(T_i^k)} \qquad \textit{Equation 4.11}$$

where $\hat{S}(T_i^k)$ is the Kaplan-Meier survival estimator for censoring within each interval and c_i^k is the total cost of patient i within interval k.

4.2.10 Lin's Regression Method – UCH (RU)

A further estimator of mean total costs in the presence of censoring proposed by Lin makes use of covariate information related to patient and clinical characteristics [Lin, 2000]. Lin sets out an ordinary least squares regression method to predict patient costs, whilst making adjustments for the covariates that influence these costs using an inverse weighting method. The regression model is fitted to those patients who die or have complete study costs. The weights that are used in the model are equal to the inverse of the survival probability, which is estimated either by the Kaplan-Meier method or by other survival model probability estimates, such as Cox's PH regression

[5] For example, if three monthly interval lengths were used then, for each interval, the Kaplan-Meier probability of being alive during that interval would be calculated as if each interval were an independent study period.

model. A reverse censoring indicator is used in the survival estimates so that deaths are denoted as zero and censored cases as one. The regression model is then applied to the full study cohort of N patients and the mean total study costs obtained (Equation 4.12).

$$\overline{C}_{RU} = \frac{1}{N}\sum_{i=1}^{N}\frac{\beta' Z_i}{\hat{S}(T_i)} \qquad \textit{Equation 4.12}$$

where Z is a vector or covariates thought to influence patient costs, β is a vector of regression coefficients and $\hat{S}(T_i)$ is the Kaplan-Meier survival estimator at time T_i with the roles of the censoring times and survival times reversed, i denotes the i^{th} patient, i = 1 to N.

4.2.11 Lin's Regression Method – KCH (RK)

Lin's regression method can also be adapted to incorporate patient cost histories. The study time period is divided in to a series of k smaller subintervals and a weighted regression model fitted to the costs within each interval. The regression coefficients for the whole study period are then calculated by summing the regression estimates for each of the subintervals. The same regression covariates should be included in all subinterval models. The predicted costs for the study cohort can then be estimated by applying the model and the mean study costs obtained (Equation 4.13).

$$\overline{C}_{RK} = \frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{K}\frac{\beta^{k'} Z_i^k}{\hat{S}(T_i^k)} \qquad \textit{Equation 4.13}$$

where Z is a vector or covariates thought to influence patient costs, β is a vector of regression coefficients and $\hat{S}(T_i)$ is the Kaplan-Meier survival estimator at time T_i with the roles of the censoring times and survival times reversed, i denotes the i^{th} patient, i = 1 to N.

4.2.12 Carides' Regression Method (CR)

The final technique presented in this chapter for estimating mean study costs will be referred to as Carides method [Carides *et al*, 2000]. The mean total cost is calculated in two stages, where the first stage assumes that each patient's total cost C_i can be modelled solely from their survival time. The chosen model should be one that

produces the best fit to the data. Carides and colleagues suggest selecting a multiplicative model ($C_i = g(T_i)Z_i$, where $Z_i = e^{\varepsilon_i}$ is the exponential function for the error term [ε_i]) to estimate costs if the study has a greater potential for expensive events, for example hospitalisations. Otherwise, an additive model ($C_i = g(T_i) + \varepsilon_i$) is assumed. In each case $g(T_i)$ denotes the functional form for modelling survival times (T_i). The model is fitted to the uncensored costs only.

The second stage of calculating the mean total cost involves weighting the cost estimates by the Kaplan-Meier probabilities of death for the cohort (Equation 4.14).

$$\overline{C}_{CR} = \int_0^{T_{MAX}} \hat{g}(t) \, | dS_T(t) | \qquad \textit{Equation 4.14}$$

where T_{MAX} is the longest observed follow-up time.

The study period is divided in to smaller time intervals and within each interval the mean expected total cost for those patients who die in the interval is weighted by the Kaplan-Meier probability of dying in the interval $\hat{S}_k - \hat{S}_{k+1}$ (See Section 4.2.7). To estimate the average lifetime costs, the weighted mean values are summed over the study period. However, if the aim of the study was to estimate costs over a specific time period (L), 2.25 years for example, then the mean estimate for the final interval should include complete costs for the study period of interest (Equation 4.15).

$$\overline{C}_{CR} = \int_0^{L} \hat{g}(t) \, | dS_T(t) | + \overline{C}_{T \geq L} \hat{S}(L) \qquad \textit{Equation 4.15}$$

where $\hat{S}(L)$ is the Kaplan-Meier probability of surviving to time L and $\overline{C}_{T \geq L}$ is the average cost for all patients who have complete costs for the study period of interest.

4.2.13 An Overview of Published Validation Methods for Estimating Mean Total Costs in the Presence of Censoring

Section 4.2.13 briefly summarises the approaches used to validate mean total cost estimates in the presence of censoring by the authors of these proposed methods.

Authors compared their proposed estimator with naïve estimators that ignore censored data [Fenn *et al,* 1995; Lin *et al,* 1997; Lin, 2000], or with estimators proposed by other authors [Bang & Tsaitis, 2000; Carides *et al,* 2000]. Comparisons were also made using both artificially created data sets based on Monte Carlo simulations [Lin *et al,* 1997; Bang & Tsaitis, 2000; Carides *et al,* 2000; Lin, 2000] and real data sets that were subject to censoring [Fenn *et al,*1995; Fenn *et al,* 1996; Lin *et al,* 1997; Lipscomb *et al,* 1998; Bang & Tsaitis, 2000; Carides *et al,* 2000; Lin, 2000]. All authors conclude that their proposed methods produce a better estimate of mean total costs than the methods they are being compared with.

Additionally, there have been a series of papers that have compared Lin's methods with weighted cost methods proposed by Bang and Tsaitis [Lin *et al,* 1997; Bang & Tsaitis, 2000]. Bang and Tsaitis use artificial and real data sets to compare methods for estimating mean total costs and conclude that the methods they propose are as reliable as Lin's methods. Raikou and McGuire also compare these methods but under conditions of heavy censoring (over 80% cases censored) and state a preference for Lin's estimator with KCH and Bang and Tsaitis estimator with KCH [Raikou & McGuire, 2004]. The authors had a slight preference towards Bang and Tsaitis method as it is "not restricted by the censoring pattern." O'Hagan and Stevens and Strawderman consider the mathematical properties of these two sets of estimators and conclude that, under certain conditions and depending upon the choice of interval lengths in to which the study time period has been divided, the methods are equivalent [Strawderman, 2000; O'Hagan & Stevens, 2004].

4.3 FOUR ALTERNATIVE CENSORING MECHANISMS

Censored cost data may arise due to several different reasons. In this section, reasons for censoring are categorised in to four broad groups: random censoring, end-of-study censoring, informative censoring and partial censoring.

4.3.1 Random Censoring

Censoring is said to occur randomly when patients are lost to follow-up at any time point during a study owing to reasons that are unrelated to (i.e. independent of) the event of interest. There are various reasons why random censoring may arise and these include:

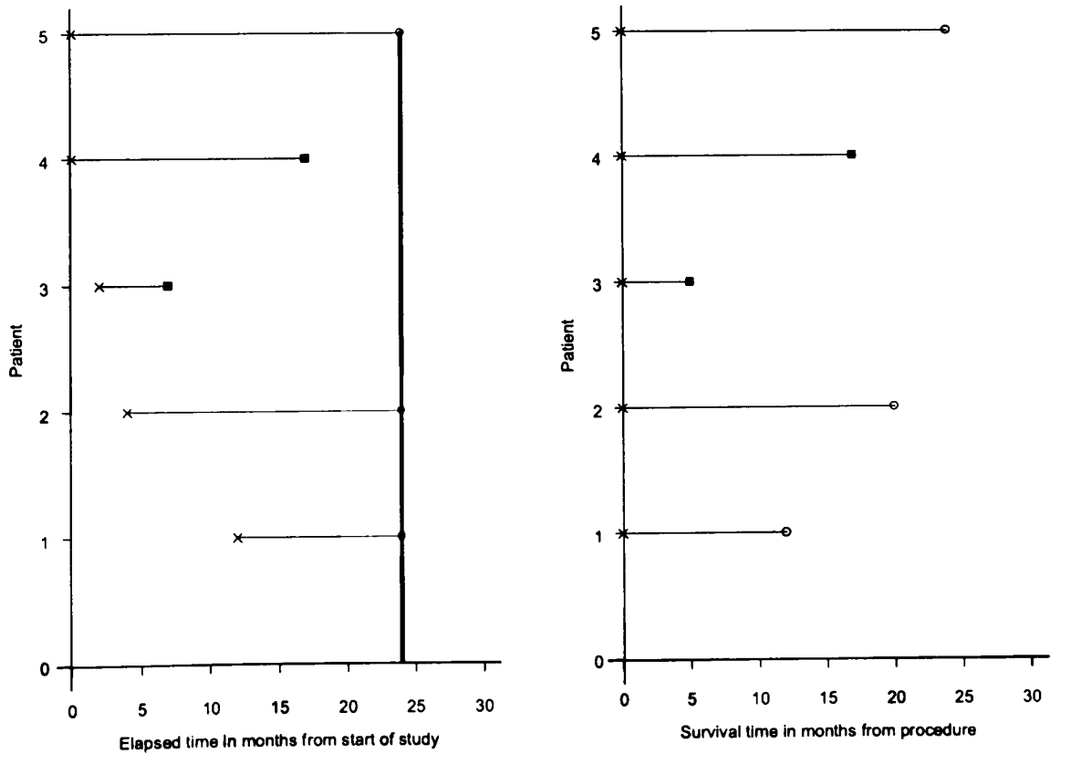
- patients moving away from the area where the study is being conducted

- patients dropping out of the study for reasons unrelated to the treatments they are receiving
- patients dying from causes unrelated to the treatment (e.g. being killed in a road traffic accident whilst in a study looking at deaths related to diabetes)

4.3.2 End-of-Study Censoring

End-of-study censoring is similar to random censoring, except that censoring is restricted to a time frame at the end of a study period. The majority of studies run for pre-specified time periods and often the patient cohort will be recruited over time rather than on the first day of the study. For example, a fixed two-year follow-up study beginning on 1st of January and ending on 31st December the following year would have incomplete follow-up data (that is, for less than two years) for all patients who entered the study after the start date. End-of-study censoring is also known in the literature as type I censoring. Figure 4.2, below, illustrates end-of study censoring.

Figure 4.2 Illustration of end-of-study censoring



X procedure

■ patients who had the event of interest

O patients who are censored

4.3.3 Informative Censoring

Informative censoring occurs when patients withdraw from a study for reasons that are related to the event of interest. For example, patients may withdraw from a study because they are unwell and the reason they are unwell is related to the treatment they are receiving.

4.3.4 Partial Censoring

In economic evaluations there are two situations where partial censoring may arise. Firstly, individual patient survival and outcome(s) may be known but individual costs and HRQL may only be available up to a specific time point. For example, if costs and HRQL data are collected over a two-year study period and survival data over a four-year study period, then cost and HRQL data are censored from the end of year two to the end of year four. Secondly, costs or HRQL information may only be collected at specific time points. For example, costs are collected every six months during a study period, thus a patient dying in month eight will have a six-month cost but censored costs for month seven onwards.

4.4 THE CELT COHORT

The objective of this chapter is to compare 12 methods for estimating mean study costs in the presence of censoring, across four censoring mechanisms. Censoring mechanisms will be simulated from a complete cohort of costs collected from point of assessment over a 2.25 year study period for CELT patients.

In the CELT study *complete* data were available for 726 liver disease cases over a 2.25 year study period. The costs for the 29 patients with *incomplete* study costs were censored 1.83 to 2.17 years after assessment. All 29 patients entered the study towards the end of the recruitment phase and were censored as a result of the study ending before complete 2.25 year data could be observed (i.e. end-of-study censoring). This analysis takes as its basis the 726 complete cases, and for these a known mean total cost and standard error can be calculated.

The data collection process described in Chapter 2, applied in the main CELT analysis for three liver disease groups (ALD, PBC, and PSC), also applies to the complete CELT cohort of 726 patients used here. For those 197 patients (27%) who were not listed as suitable transplant candidates, only the assessment costs are included in the transplant patient costs. Patients listed for transplantation who did not receive a liver

transplant during the study period (N = 73 [10%]) were followed up to the point they were removed from the waiting list.

Covariates that significantly affect costs were incorporated in to the calculation of mean study costs in the presence of censoring for the Cox cost method, partitioned Cox cost method, Lin's regression method KCH and Lin's regression method UCH. Table 4.1 presents demographic patient characteristics for the 726 "complete" CELT cases that were considered in the four methods for estimating mean study costs listed above.

Table 4.1 Demographic patient characteristics for the 726 "complete" CELT patients

Age in years	Mean (SD)	49.3 (12.1)
Gender (%)	Males	334 (46%)
Centre (%)	1	160 (22%)
	2	50 (7%)
	3	118 (16%)
	4	54 (7%)
	5	90 (12%)
	6	254 (35%)
Transplant group (%)	Not listed	197 (27%)
	Elective	396 (55%)
	Emergency	75 (10%)
	Re transplant < 14 days	7 (1%)
	Re transplant > 14 days	51 (7%)
Disease group (%)	Cholestatic	372 (51%)
	Parenchymal	204 (28%)
	Fulminant	78 (11%)
	Other	72 (10%)
Number of listed patients (%)		529 (73%)
Number of transplanted patients (%)		456 (63%)
EQ-5D score at listing	Mean (SD)	0.48 (0.25)

SD: Standard deviation

In each of the four models where covariate information was included, covariates with a p-value less than or equal to 0.05 were maintained in statistical models. Patient age, gender, transplant centre, disease group and whether a patient received a transplant

were significant predictors of costs in the Cox cost model. Transplant centre, transplant group and whether a patient received a transplant were significant predictors of costs in the partitioned Cox cost model. Type of transplant and whether a patient received a liver transplant were significant predictors of costs in linear regression models (Lin's regression methods with KCH and UCH).

4.5 CREATING ARTIFICIAL CENSORED DATASETS FOR THE CELT COHORT

Simulation methods are used to create artificial data sets for each of the four censoring mechanisms introduced in Section 4.3: random, end-of-study, informative and partial censoring. The methods for creating these data sets are described in further detail below.

4.5.1 Random Censoring

Artificial data sets containing randomly censored data were created from the CELT cohort using simulation methods. Random number generators were used to select which patients were to be censored and at what time points the censoring were to occur. Censoring was simulated assuming a binomial distribution and the time of censoring was simulated assuming a uniform distribution over the time period 0 to 2.25 years. Observed CELT study deaths were artificially censored if the randomly generated censoring time was less than the observed time of death. If the censoring time was greater than or equal to the observed time of death, the observed data remained unchanged. Each patient's study costs are censored at the randomly selected censoring time.

A total of five thousand simulated data sets were created[6] for each of three different censoring levels: 10% censoring (representing a low level of censoring), 30% censoring (representing a medium level of censoring) and 50% censoring (representing a high level of censoring). These three alternative censoring levels were simulated in order to investigate how the level of censoring affected the accuracy of the mean total cost estimates for the twelve methods described in Section 4.2.

4.5.2 End-of-study Censoring

As with random censoring, simulation methods were used to create 5,000 artificial data sets for analysis at each of 10%, 30% and 50% end-of-study censoring. The binomial

[6] A total of 5,000 simulations were required to estimate the mean cost estimate to within £60 pounds of the true mean with 95% uncertainty.

distribution was used to create censoring and the uniform distribution to generate censored survival times. To create end-of-study censoring the randomly generated censoring times were restricted to the last six months of the study period. This period was chosen as it reflected the censoring pattern of the 29 patients who were censored in the CELT study (see Section 4.4).

Further data sets were created for the same three censoring levels where the end-of study censoring was restricted to the final year of the study. This choice was made in order to vary the period of time where censoring occurred. The choice of time period was essentially arbitrary.

Thus for end-of study censoring six sets of simulations were run:

- at the 10% censoring level with censoring restricted to the final 0.5 years of the study 5,000 simulated data sets were created
- at the 30% censoring level with censoring restricted to the final 0.5 years of the study 5,000 simulated data sets were created
- at the 50% censoring level with censoring restricted to the final 0.5 years of the study 5,000 simulated data sets were created
- at the 10% censoring level with censoring restricted to the final year of the study 5,000 simulated data sets were created
- at the 30% censoring level with censoring restricted to the final year of the study 5,000 simulated data sets were created
- at the 50% censoring level with censoring restricted to the final year of the study 5,000 simulated data sets were created

4.5.3 Informative Censoring

In the CELT study, utilities were measured using the EQ-5D, which was administered to patients as part of a postal questionnaire at point of listing and at three monthly intervals from point of listing up to point of transplant. After transplantation the questionnaire was administered to patients at 0.25, 0.5, 1 and 2 years post transplant. The responses from the EQ-5D are used to create artificial data sets that simulate informative censoring.

Datasets based on informative censoring were created by considering two alternative scenarios, the first hypothesises that the sickest patients would drop out of the study

due to ill health and the second hypothesises that patients who were well would drop out of the study. The percentile distribution of the EQ-5D was examined across all the time points that the questionnaire was administered and information from the 10th (utility = 0.16), 20th (utility = 0.36), 80th (utility = 0.85) and 90th (utility = 1.00) percentiles were used to create four data sets that simulated informative censoring.

Two data sets were created to simulate informative censoring due to ill health. To create a censored data set simulating informative censoring using the 10th percentile of the EQ-5D distribution, each patient profile of EQ-5D scores over time were examined and patients were censored at the point their utility score first fell below 0.16. For example, patient A might have a profile of utility values 0.38, 0.25, 0.12 and 0.08 at times 0, 0.5, 1 and 2 years for the study period; this patient's utility falls below 0.16 at one–year and they are therefore assumed censored in terms of both costs and survival after this time point. Patients whose profiles never drop below 0.16 remain uncensored. This process is repeated to create a second censored dataset simulating informative censoring using the 20th percentile of the utility distribution, except here patients are censored when their utility score first falls below 0.36, which occurs at 0.5 years for patient A. Using this method to simulate informative censoring due to ill health resulted in 13% censoring at the 10th percentile level, and 31% censoring at the 20th percentile level.

Data sets three and four were created to simulate informative censoring due to good health. In set three patients are censored at the time point where their utility first exceeds the 80th percentile (utility = 0.85) of the EQ-5D distribution. For example, suppose patient B has utilities of 0.42, 0.54, 0.74 and 1.00 at 0, 0.5, 1, and 2 years for the study period; this patient's utility exceeds 0.85 at two years and they are therefore assumed to be censored in terms of survival and costs after this time point. Patients whose profiles never exceed 0.85 remain uncensored. This process is repeated to create the final data set using the 90th percentile of the utility distribution (utility = 1.00). Using this method to simulate informative censoring due to good health resulted in 21% censoring at the 80th percentile level and 14% censoring at the 90th percentile level.

4.5.4 Partial Censoring

Two partially censored cohorts were simulated. In the first data set it was assumed that resource use and costs were collected at one time point only during the CELT study period, 31st March 1998, 2.25 years after data collection for the CELT study first began.

This type of censoring is referred to in this chapter as *one time resource collection* censoring. Any costs incurred after this time point were set to zero and costs were censored as of this date (80% of the data were censored).

In the second data set it was assumed that resource use and cost information were collected up to 2.25 years after the end of a one-year recruitment period in each transplant centre (3.25 years after data collection first began at each centre). The second type of censoring is referred to here as *fixed resource collection* censoring. In the second partially simulated data set 15% of the data were censored.

Survival times were observed over the full study period for partial censoring for both one time resource collection and fixed time resource collection.

4.6 STATISTICAL METHODS FOR EXPLORING THE ACCURACY OF METHODS FOR ESTIMATING MEAN TOTAL COSTS IN THE PRESENCE OF CENSORING

4.6.1 Varying Interval Lengths

The partitioned Cox cost method, Lin's method KCH and UCH, the weighted cost method with KCH, Lin's regression method with KCH and Carides' method all require the division of the study time period in to a series of smaller time intervals. The six methods can then be applied in order to estimate mean total study costs in the presence of censoring. Three monthly interval lengths were used to estimate mean total costs for these methods.

Lin *et al* and O'Hagan and Stevens recommend using the smallest possible choice of interval length to obtain accurate estimates of Lin's methods (KCH and UCH) [Lin *et al*, 1997; O'Hagan & Stevens, 2004]. O'Hagan and Stevens also state that the weighted cost method (with KCH) of Bang and Tsiatis is likely to produce more accurate estimates of mean total costs with smaller interval lengths in comparison with larger ones [Bang & Tsiatis, 2000; O'Hagan & Stevens, 2004]. To observe whether these assumptions were true for the CELT data, the interval lengths were varied for the six methods where it was necessary to divide the study time period in to interval lengths. The accuracy of Lin's method (KCH and UCH), and Carides' method were compared over interval lengths of one month, two months, three months, six months, and twelve months across random, end-of-study, informative and partial censoring. The accuracy

of the partitioned Cox cost method, the weighted cost method KCH and Lin's regression method with KCH were compared over interval lengths of two months, three months, six months and twelve months across random, end-of-study, informative and partial censoring[7].

4.6.2 Estimating Standard Errors

The methods of ignoring costs, the complete case method, Cox cost method, partitioned Cox cost method and Lin's regression methods (KCH and UCH) use the conventional method of estimating the standard error of the mean total cost based on the assumption of normally distributed data (Equation 4.16).

$$\text{Standard Error} = \left(SD/\sqrt{n}\right) \qquad \textit{Equation 4.16}$$

Greenwood's formula was used to estimate the standard error for the Kaplan-Meier method [Collett, 1994]. Non-parametric bootstrapping was applied to estimate the standard errors for the remaining five methods (Lin's methods [KCH and UCH], the weighted cost methods [KCH and UCH] and Carides' method) [Manley, 1997].

4.6.3 Comparing the Accuracy of Methods from the Estimated Means and Standard Errors

The primary focus of any researcher selecting a method to estimate mean total costs in the presence of censoring, is to select a method that accurately measures the mean costs. The accuracy of the 12 methods was assessed by measuring the magnitude (distance) between the estimated means and the true mean for the "complete" data set, prior to censoring.

Standard errors are calculated in order to make inferences about the results to the general population and in statistical tests for comparing potential differences between the costs of two or more treatments. Therefore, it is also important that methods which can estimate the mean total costs accurately also estimate the uncertainty around the mean estimates (standard errors) accurately. The magnitude between the estimated standard errors and the observed standard error are also calculated.

[7] In this thesis CELT survival has been measured using monthly units of time. For the partitioned Cox cost, Lin's regression (KCH) and the weighted cost methods (KCH) the time interval is divided in to a series of smaller intervals and Kaplan-Meier or Cox survival probabilities are estimated for the N patients alive at the beginning of the interval. Thus, it is not possible to estimate survival probabilities for these three methods using one monthly interval lengths.

Methods for estimating mean total costs are ranked in order of magnitude, where the smaller the magnitude the closer the estimate is to the observed mean or standard error. Methods are ranked separately in terms of the accuracy of the mean and the accuracy of the standard error. Rankings are assigned separately for each type of censoring mechanism (random, end-of-study, informative, partial) and at each level of censoring (e.g. 10%, 30%, 50% censoring levels). Kendall's concordance statistic was used to assess whether methods were consistently ranked as being more (or less) accurate in comparison with alternative methods across different censoring mechanisms and for different levels of censoring [Kendall & Gibbons, 1990][8]. For the CELT data a significant p-value ($p \leq 0.05$) was taken to indicate that the ordering of the accuracy of the mean (or standard error) in the presence of censoring does not differ and a non-significant p-value ($p > 0.05$) was taken to indicate that the ranking of methods does differ. In other words, unless significant concordance was demonstrated across the ordering of methods, the accuracy of the mean (or standard error) was deemed to depend either on censoring mechanism or censoring level.

4.6.4 Other Statistical Issues

For Carides' method a non parametric local regression model with smoothing [Cleveland & Devlin, 1988] was fitted to estimate individual patient costs from their survival times.

The S-PLUS statistical computer package was used for all analysis [S-PLUS, 2001].

4.7 A COMPARISON OF MEAN TOTAL COST ESTIMATES ACROSS 12 METHODS FOR ESTIMATING MEAN TOTAL COSTS FOR THE CELT STUDY

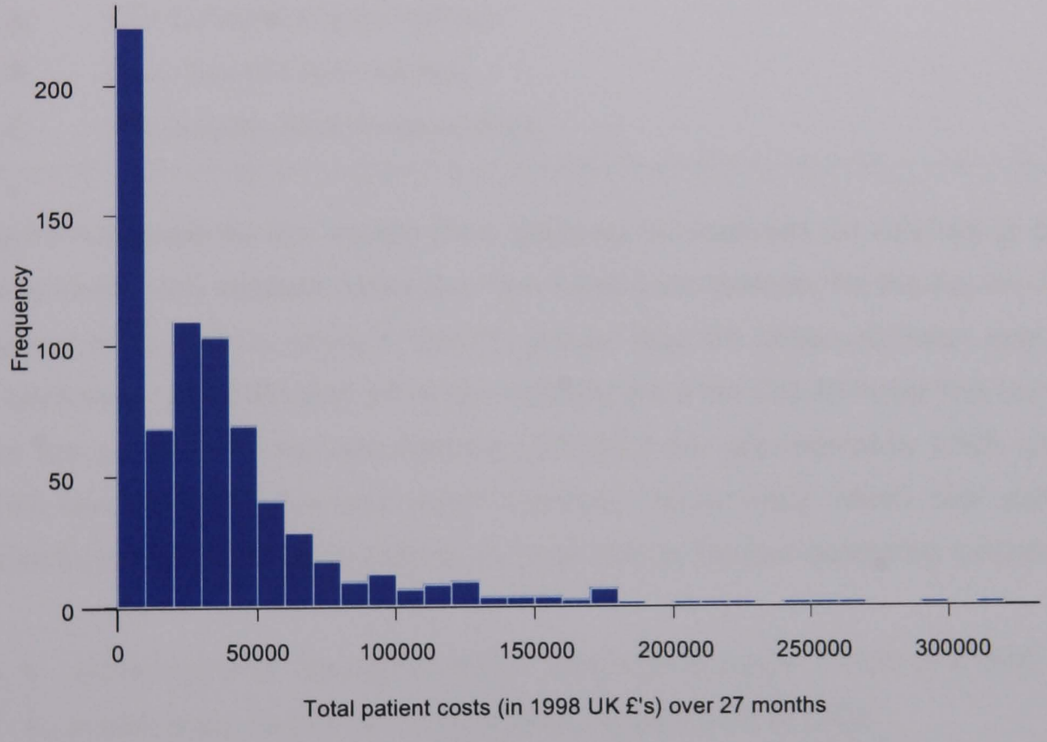
The mean total cost for the 726 uncensored CELT patients was £36,045 and the standard error was £1,517. The median total cost was £27,166 with costs ranging from £393 to £311,873 across the sample. This is a typical example of a positive skewed distribution of cost data (Figure 4.3).

Figures 4.4 to 4.8 present the results from random censoring, end-of-study censoring, informative censoring and partial censoring across different levels of censoring. All graphs are presented as the difference between the estimated mean and standard error and the actual mean (£36,045) and standard error (£1,517) for the CELT cohort

[8] Kendall's concordance statistic measures the agreement between two or more sets of ranked data.

prior to censoring. In Figures 4.4, 4.5 and 4.6 each point on the graphs represents the mean, average total costs and standard errors across 5,000 simulations. In Figures 4.7 and 4.8 each point represents the average difference between the observed and expected mean total cost and the observed and expected standard error for one simulation run. The legend for Figures 4.4 to 4.8 is presented in Box 4.1

Figure 4.3 Distribution of total costs to the liver transplant programme for the CELT cohort over the 27-month study period in 1998 UK £'s (N = 726)



Box 4.1 Legend for Figures 4.4 to 4.9

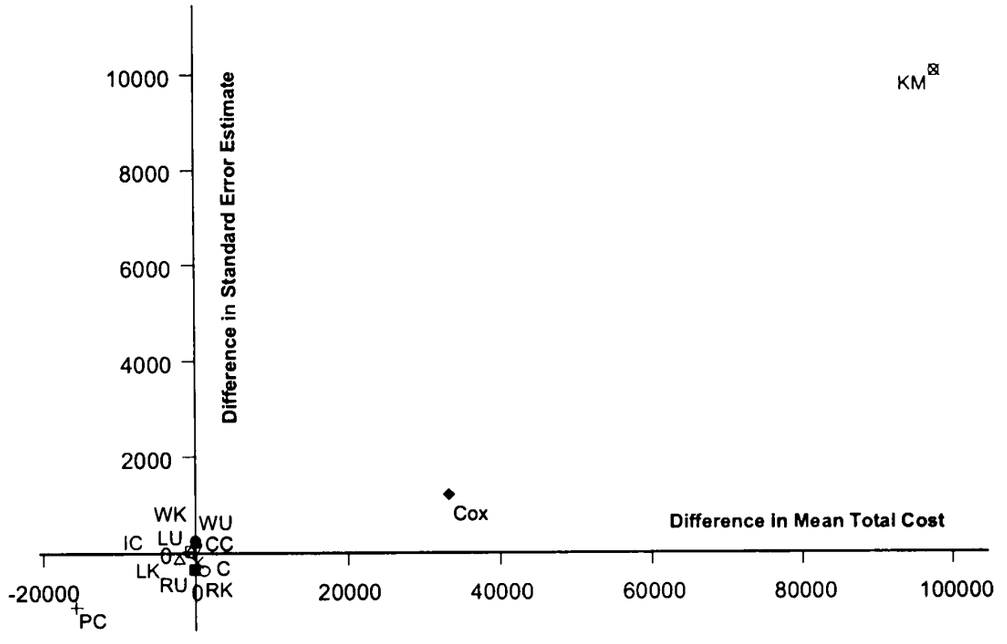
□	WK: Weighted method (KCH)
○	RK: Lin's regression method (KCH)
△	LK: Lin's method (KCH)
+	PC: Partitioned Cox cost method (KCH)
×	C: Carides' method
◇	IC: Ignoring censoring
▽	LU: Lin's method (UCH)
■	RU: Lin's regression method (UCH)
●	WU: Weighted method (UCH)
▲	CC: Complete case method
◆	Cox: Cox PH cost method
⊗	KM: Kaplan-Meier cost method

Figure 4.4 presents the results from applying 12 methods for estimating mean total costs under 10% random censoring. The mean cost estimate for the Kaplan-Meier cost method (£133,752) is nearly £100,000 greater than the observed mean total cost prior to censoring (£36,045) and the mean estimate from the Cox PH cost method (£69,266) and the partitioned Cox cost method (£20,321) are approximately £30K greater and £15K less than the observed mean estimate, respectively. Mean cost estimates for these three methods were consistently poor across the four censoring mechanisms:

- mean costs for the Kaplan-Meier cost method range: £132K to £160K
- mean costs for the Cox cost method range: £68K to £75K
- mean costs for the partitioned Cox cost method range: £19K to £23K

These three methods have been excluded from Figures 4.5 to 4.8 in order to aid the reader in distinguishing between the estimates for the remaining nine methods. Tabular versions of these results, including the results for the three omitted methods are found in Appendix 4.2.

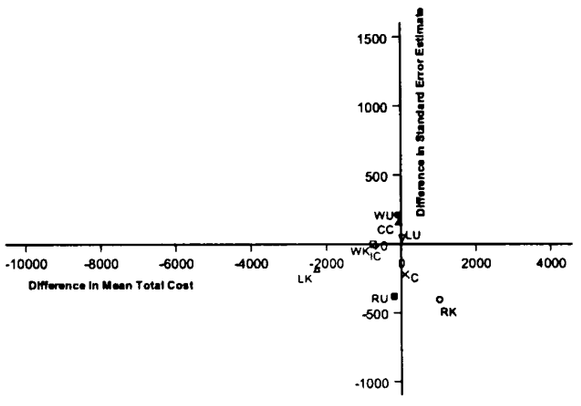
Figure 4.4 Mean total costs and standard errors estimated from 12 censored cost methods, assuming random censoring at 10%



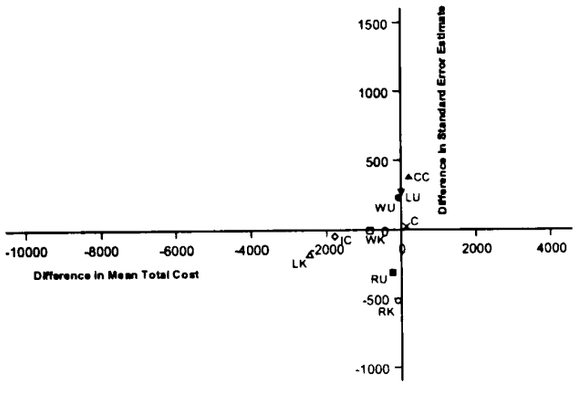
Figures 4.5 to 4.8 show that the majority of methods estimated mean total costs to a reasonable level of accuracy. Carides method, the weighted cost method (KCH and UCH), Lin's regression method with UCH, Lin's method with UCH and the method of ignoring censoring, frequently predicted mean costs to within £1,000. Estimates of mean total costs were worst under informative censoring of sicker patients; these patients incurred higher costs than other patients, and censoring them earlier resulted in a larger under estimation of mean costs in comparison with other censoring mechanisms (the underestimation of the mean costs ranged from £925 to £9,251).

Figure 4.5 Mean total costs and standard errors estimated for nine censored cost methods, assuming random censoring at a) 10%, b) 30% and c) 50%

a) 10% censoring



b) 30% censoring



c) 50% censoring

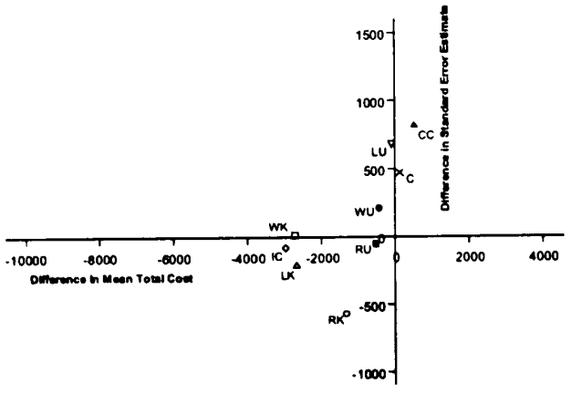


Figure 4.6 **Mean total costs and standard errors estimated for nine censored cost methods, assuming end-of-study censoring at a) 10% from 1.75 years, b) 30% from 1.75 years, c) 50% from 1.75 years, d) 10% from 1.25 years, e) 30% from 1.25 years and f) 50% from 1.25 years**

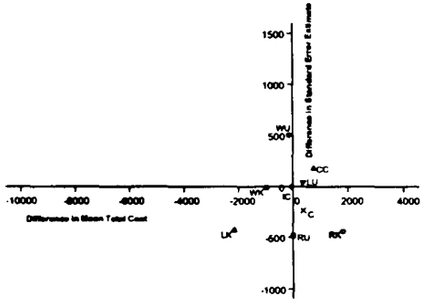
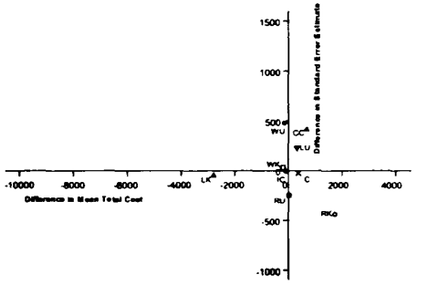
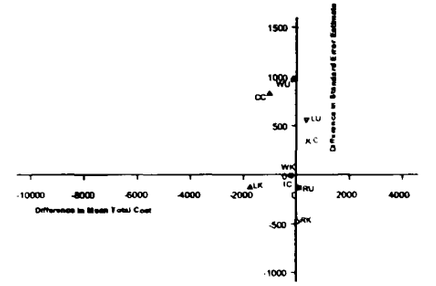
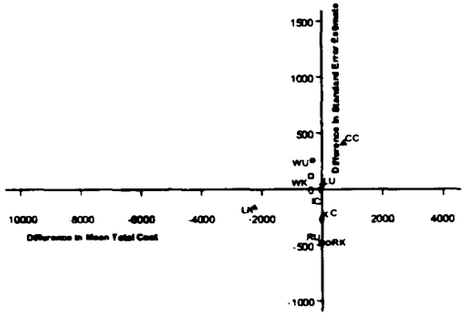
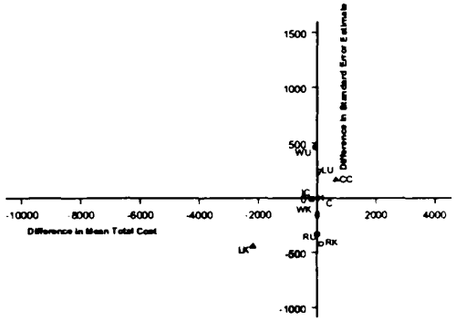
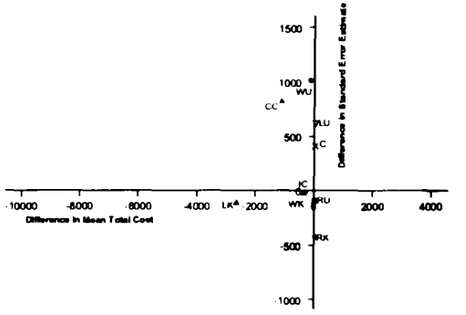
a) 10% from 21 months	b) 30% from 21 months	c) 50% from 21 months
		
d) 10% from 18 months	**e) 30% from 18 months**	**f) 50% from 18 months**
		

Figure 4.7 **Mean total costs and standard errors estimated for nine censored cost methods, assuming informative censoring with EQ-5D scores a) less than the 10th percentile, b) less than the 20th percentile c) greater than the 80th percentile, or d) greater than the 90th percentile of the utility score distribution**

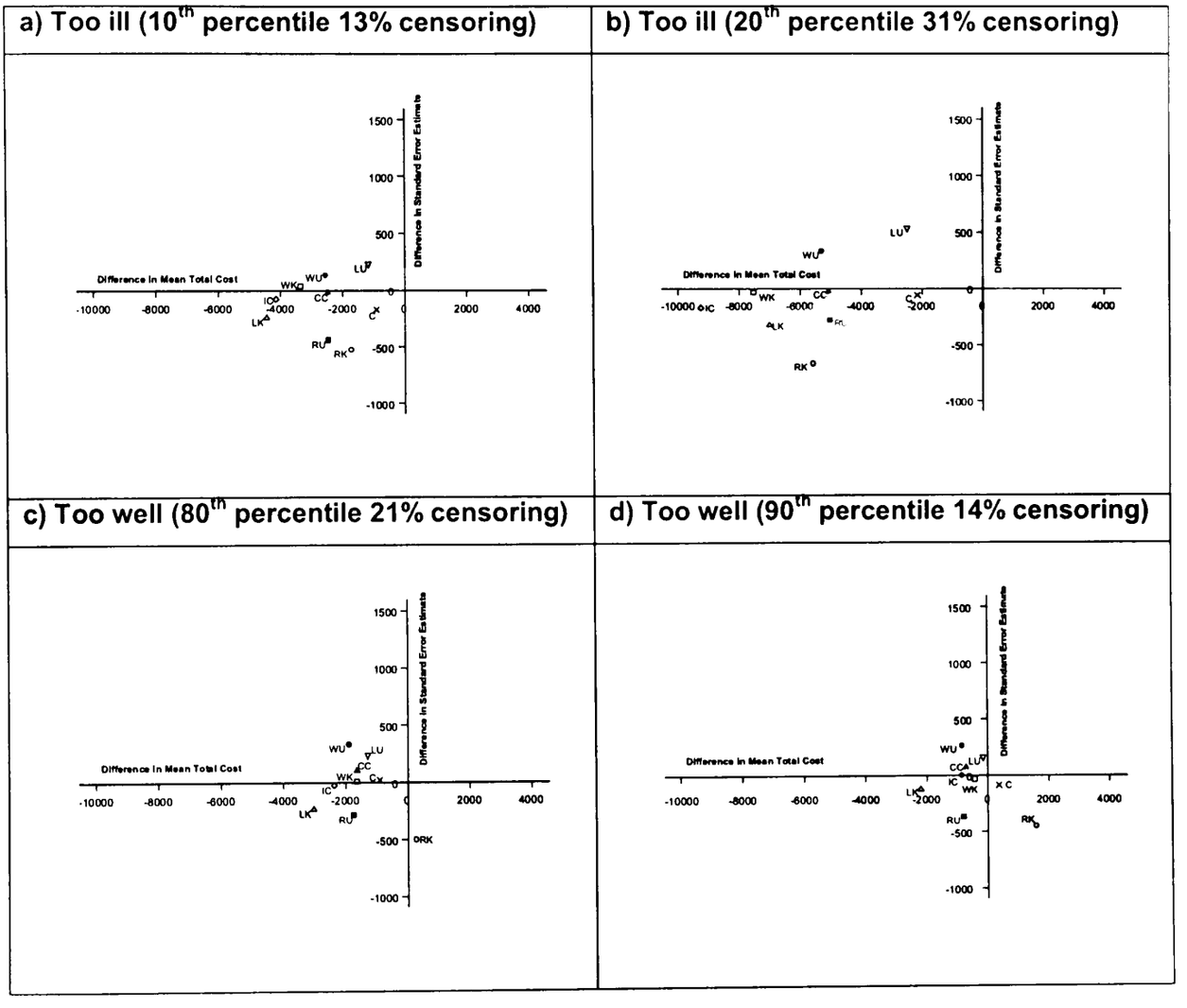
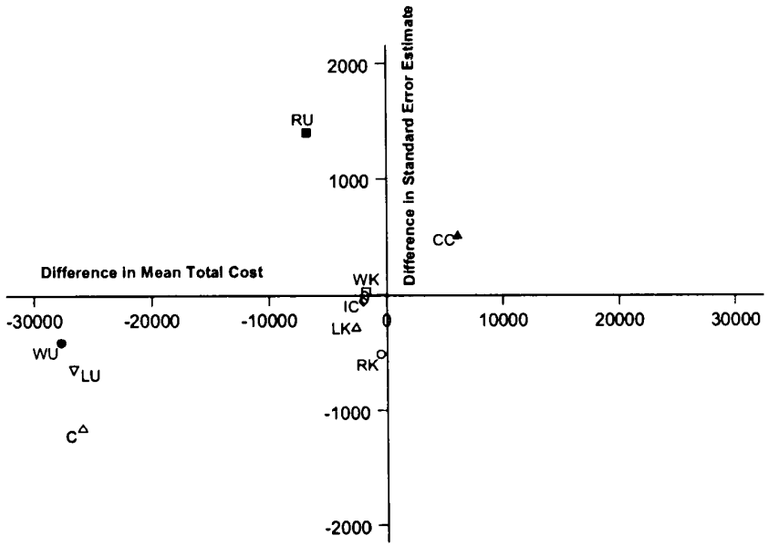
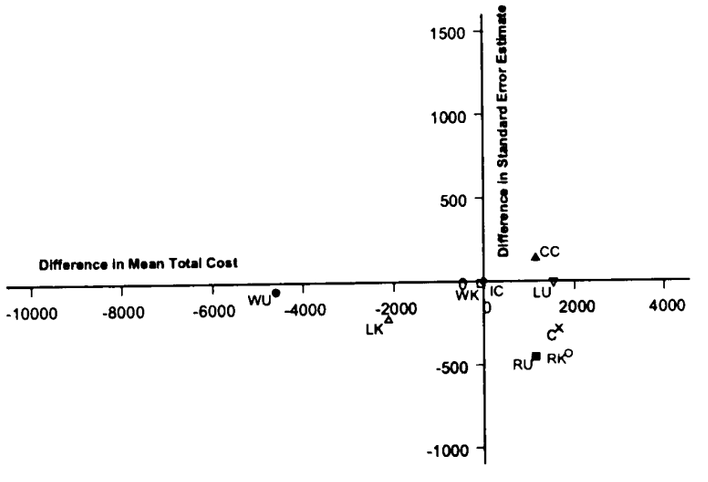


Figure 4.8 **Mean total costs and standard errors estimated for nine censored cost methods, assuming partial censoring where costs are collected a) one time resource collection, or b) fixed resource collection**

a) One time resource collection (80% censoring)



b) Fixed resource collection (15% censoring)



One time resource collection partial censoring (80% censoring) cost estimates for Carides method, Lin's method (UCH) and the weighted cost method (UCH) under

predicted mean costs by over £25K[9]. The mean study costs, estimated from Lin's method (UCH) are based upon the costs of patients who die or have complete costs to the end of the study period (censored costs are ignored). Costs are weighted by the probability of dying within an interval and the largest weights are given to mean costs that occur in the final study interval, which also includes patients who survive the full study period (Table 4.2). Under one time resource collection partial censoring, no patient incurs costs after 19 months, in other words, no patient has complete follow-up data to 2.25 years and the last death occurred at 19 months – 8 months prior to the end of the study. Thus, under Lin's method (UCH), given that all patients are censored prior to the end of the study, no costs can obtain the largest weight observed in the final study interval resulting in a large underestimation in the mean total costs. A similar observation is true for Carides' method where the same probabilities of death (listed in Table 4.2) are assigned to mean cost estimates for patients who die or have complete study costs derived from a non-parametric local regression model.

Table 4.2 Probabilities of death and mean cost per three month period for patients who die during study after simulating partial censoring (one time resource collection 80% censoring)

Month	3	6	9	12	15	18	21	24	27
N	86	18	21	12	4	2	1	0	0
Probability of death per interval	0.12	0.02	0.03	0.02	0.01	0.008	0.008	0.007	0.78
Cost (£s)	22,767	51,691	74,952	97,501	72,808	42,328	67,823	0	0
Cost*Probability of death (£s)	2,697	1,282	2,374	1,880	903	350	561	0	0

For the weighted cost method (UCH) mean study costs are again estimated from the costs for patients who die or survive the full study period and these costs are weighted by the inverse of the Kaplan-Meier estimator, $1/\hat{S}(t)$ using reverse censoring. Under partial censoring the survival estimates are known for all patients for the full study period (only costs are censored). Thus, applying reverse censoring to Kaplan Meier estimates results in survival probabilities equal to one for all patients, because under reverse censoring no patient has the event of interest. Therefore, for one time resource collection partial censoring, we are summing the unweighted uncensored costs for 166

[9] Readers should note the change in scale for Figure 4.8a.

patients, and obtain a total cost of £7,027,378. This total cost is divided by the total number of patients in the CELT sample (N = 726) to obtain a mean estimate of £9,678.

Table 4.3 presents the rankings for the magnitudes of the 12 methods for estimating mean study costs in the presence of censoring. The first four techniques presented in each table are those where cost histories are required to estimate total study costs. Cost histories are not required for the remaining eight methods. Table 4.4 presents the rankings for the magnitudes of the standard error estimates. Appendix 4.2 presents the absolute difference of each mean estimate from the observed mean and the difference of the estimated standard error from the observed standard error.

Lin's method (UCH) consistently gave one of the most accurate estimates of mean costs in the presence of censoring, and predicted the mean to within £7 under random censoring at 30%, and within £3 under end-of-study censoring during the final year of the study at a 10% censoring level. Lin's regression method (UCH) performed well under end-of-study censoring, predicting the mean to within £3 (end-of-study censoring during the final year, 30%) and Carides' method performed well under informative censoring. Both Carides' method and weighted cost method with KCH consistently predicted mean total costs to within £1000 (although were not always the best estimates of the mean) and gave reasonably accurate estimates of the standard errors. The methods of ignoring censoring and the weighted cost method with KCH consistently gave the best estimates of the standard error across all censoring levels and mechanisms.

Surprisingly, the naïve method of ignoring censoring was one of the more accurate techniques for estimating mean total costs in the presence of censoring and frequently predicted mean cost estimates to within £1,000. Under partial censoring (fixed time) ignoring censoring was the best estimator of the mean and standard error estimates (the mean estimate was out by £10 and the standard error was exact). It is likely that this unexpected result is due to the nature of the cumulative costs for a typical patient in the CELT study: costs are high at the beginning of the study period when transplantation occurs and then level off towards the end of the study when patients have stabilised on drugs and typically are seen at outpatient appointments every six to twelve months (Figure 4.9).

Lin's method (UCH) performed poorly under partial censoring when censoring was heavy (80% censoring). The Kaplan-Meier cost method, Cox cost method, the partitioned Cox cost method and Lin's method (KCH) consistently gave the poorest mean estimates across all censoring mechanisms and levels. Lin's method with UCH and Lin's regression method with UCH both gave poor standard error estimates.

Kendall's concordance test statistic was applied to the rankings of means and standard errors to establish whether there were any statistically significant differences in the ordering of methods across different levels of censoring, and different censoring mechanisms. Unsurprisingly, the ordering of the accuracy of methods differed under the two partial censoring mechanisms (one time and fixed time resource collection) for both the mean and standard error estimates (Kendall's concordance = 0.833, $p = 0.074$ and 0.888, $p = 0.052$, respectively). However, the rankings of the accuracy of random, end-of-study and informative censoring did not vary across censoring levels. The rankings for the accuracy of mean cost estimates differed significantly from the rankings of the standard error estimate across all censoring levels and mechanisms, thus a method that gave accurate cost predictions did not necessarily give accurate standard error estimates and visa versa. Appendix A4.3 presents full details of the results for Kendall's concordance statistic.

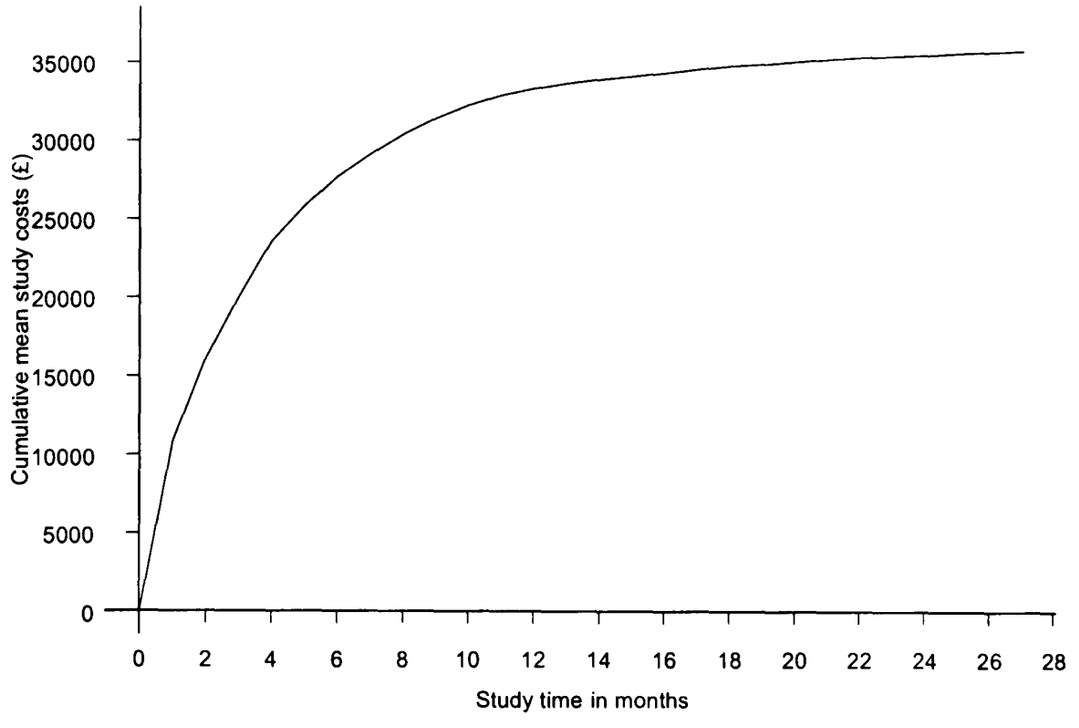
Table 4.3 Ranking the methods for estimating mean total costs; ranked in terms of magnitude from the true mean (£36,045) of the uncensored CELT data

	Random			End-of-study 1.75 years			End-of-study 1.25 years			Informative Low EQ-5D		Informative High EQ-5D		Partial	
	10%	30%	50%	10%	30%	50%	10%	30%	50%	10th (13%)	20th (31%)	80th (21%)	90th (14%)	One Time (80%)	Fixed Time (15%)
Known Cost Histories															
WK: Weighted costs	7	7	8	7	4	5	7	6	7	7	8	4	3	2	2
RK: Lin's regression	8	3	6	8	8	1	5	5	3	3	6	1	8	1	7
LK: Lin's method	9	9	7	9	9	9	9	9	9	9	7	9	9	4	8
PC: Partitioned Cox	10	10	10	10	10	10	10	10	10	10	10	10	10	7	10
Unknown Cost Histories															
C: Carides	4	4	2	5	6	7	4	4	2	1	1	2	2	8	6
IC: Ignoring Censoring	6	8	9	2	2	4	3	7	6	8	9	8	7	3	1
LU: Lin's method	1	1	1	4	5	6	1	2	4	2	2	3	1	9	5
RU: Lin's regression	5	6	4	1	1	2	2	1	1	4	3	6	5	6	=3
WU: Weighted costs	3	2	3	3	3	3	6	3	5	6	5	7	6	10	9
CC: Complete cases	2	5	5	6	7	8	8	8	8	5	4	5	4	5	=3
Cox: Cox regression	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11
KM: Kaplan-Meier	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12

Table 4.4 **Ranking the methods for estimating mean total costs; ranked in terms of magnitude from the true standard error (£1,517) of the uncensored CELT data**

	Random			End-of-study 1.75 years			End-of-study 1.25 years			Informative Low EQ-5D		Informative High EQ-5D		Partial	
	10%	30%	50%	10%	30%	50%	10%	30%	50%	10th (13%)	20th (31%)	80th (21%)	90th (14%)	One Time (80%)	Fixed Time (15%)
Known Cost Histories															
WK: Weighted costs	1	1	1	2	3	2	3	3	2	2	2	1	2	1	3
RK: Lin's regression	9	9	7	7	8	6	9	7	6	9	9	10	9	6	8
LK: Lin's method	5	4	5	6	4	3	4	8	4	7	6	6	5	3	6
PC: Partitioned Cox	10	10	10	10	10	8	10	10	9	10	10	11	11	9	11
Unknown Cost Histories															
C: Carides	7	2	6	5	2	5	5	1	5	5	3	2	4	11	7
IC: Ignoring Censoring	2	3	3	1	1	1	1	2	1	3	4	3	1	2	1
LU: Lin's method	3	6	8	3	5	7	2	5	7	6	8	5	6	7	2
RU: Lin's regression	8	7	2	8	6	4	8	6	3	8	5	7	8	10	9
WU: Weighted costs	6	5	4	9	9	10	6	9	11	4	7	8	7	4	4
CC: Complete cases	4	8	9	4	7	9	7	4	8	1	1	4	3	5	5
Cox: Cox regression	11	11	11	11	11	11	11	11	10	11	11	9	10	8	10
KM: Kaplan-Meier	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12

Figure 4.9 **Cumulative mean total costs of the liver transplant programme over the 27 month study period in 1998 UK £'s**



4.7.1 Summary of Results

The accuracy of the 12 censored cost methods varied by censoring mechanism:

- Lin's method (UCH) gave the most accurate mean estimates under random censoring
- Lin's regression method (UCH) gave the most accurate estimates under end-of-study censoring
- Carides' method gave the most accurate estimates under informative censoring
- the weighted cost method (KCH) or ignoring censoring gave the most accurate estimates under partial censoring

Carides' method and the weighted cost method (KCH) gave accurate predictions of both the mean cost and standard error across censoring levels and censoring mechanisms. Generally, the accuracy of mean estimates did not differ by censoring level, the exception being partial censoring, where Carides' method, Lin's method (UCH) and the weighted cost method (UCH) performed poorly under one time resource

collection censoring. Finally, the accuracy of methods differed for the mean and standard error estimates.

4.7.2 The Impact of Covariates on Mean Total Costs in the Presence of Censoring

Mean cost estimates derived from Lin's regression methods, KCH and UCH, Cox cost method and partitioned Cox cost method adjust for covariates within the cost estimates. In this chapter covariates were selected based upon level of significance ($p < 0.05$) and were not common across the four methods. Thus, Lin's regression methods KCH and UCH adjusted for whether transplanted (yes/no), and transplant group, the partial Cox cost method adjusted for whether transplanted, transplant group and transplant centre, and the Cox cost method adjusted for age, gender, disease group, transplant centre and whether transplanted.

The choice of covariates allowed for in each model might effect mean cost predictions, therefore this section explores how mean total cost estimates are effected by covariate patterns.

4.7.2.1 Cox Cost and Partitioned Cox Cost Models

The Kaplan-Meier method, Cox cost method and partitioned Cox cost method all violate the assumption of independence between censoring and costs, where censored costs should be representative of all costs that value or greater. This chapter has already shown that cost estimates can not be predicted accurately if this assumption is ignored (Section 4.7.1).

When the Kaplan-Meier estimate was applied to the CELT data costs were over estimated by almost £100K across all censoring mechanisms. After adjusting for age, gender, transplant centre, liver disease group and whether receiving a liver transplant (yes/no) mean total cost estimates derived from the Cox cost method were at least £60K less than estimates from the Kaplan-Meier method. However, estimates from the Cox cost model were still twice the observe cost due to the violation of the independence assumption. The partitioned Cox cost method reduced estimates further and underestimated mean costs by over £13K. This section briefly explores the impact of adjusting for covariates upon cost estimates for the Cox cost method and partitioned Cox cost method. Models are explored under 10% random censoring and are generalisable to other levels and mechanisms of censoring for the CELT data set.

Estimates of mean total costs were compared between the Kaplan-Meier model and the Cox cost model under 10% random censoring for each individual covariate in order to understand how the covariate pattern from the Cox cost model affects mean cost estimates. Under 10% random censoring mean cost estimates under the Kaplan-Meier method were approximately £98K greater than the observed mean estimate of £36,045 (Table 4.5). At the individual covariate level gender and disease group are not significant predictors of costs in the Cox cost model but age, centre, transplant group and whether transplanted are. Further, whether transplanted lowered mean costs by over £61K and transplant group lowered mean costs by over £45K whereas age, gender, centre and disease group had little impact on mean cost estimates.

Table 4.5 Estimated mean total costs under 10% random censoring for the Kaplan-Meier method and Cox cost model for individual covariates (univariate models)

	Mean (£)	Overall model likelihood ratio (p-value)	Difference from KM mean estimate (£)
Kaplan-Meier method	134,107		
Cox Cost Method (individual models)			
Age	136,019	5.01 (0.025)	1,912
Gender	137,000	2.11 (0.146)	2,839
Centre	130,766	17.3 (0.004)	-3,341
Disease group	137,923	6.54 (0.088)	3,816
Transplant group	88,629	62.2 (< 0.001)	-45,478
Transplanted (Y/N)	72,882	205.0 (< 0.001)	-61,223

The next step in the investigation was to explore how the five variables included in the Cox cost model affected mean cost estimates within a multivariate Cox cost model. The process is illustrated under 10% random censoring where the mean cost is estimated at £63,383 a £70K reduction in cost estimates under the Kaplan-Meier method. A second multivariate model was fitted which included age, gender, centre and disease group and predicted the mean costs as £134,034 implying that whether transplanted is the only variable that reduces costs significantly in this multivariate model when comparing estimates with those from the Kaplan-Meier method. A final multivariate Cox cost model was fitted that adjusted transplant centre, transplant group and whether transplanted (the three variables adjusted for in the partitioned Cox cost

models). Mean cost estimates after adjusting for these three variables were £68,005 and were similar to estimates after adjusting for age, gender, centre, disease group and whether transplanted.

The partitioned Cox cost method estimated mean costs per three monthly intervals, costs are said to be censored per interval if they are incomplete for the interval. Individual covariate estimates had similar impact on mean estimates to that for the Cox cost model and were most noticeable for transplant group and whether transplanted. Estimates of mean total costs for partitioned Cox cost models that included centre alone had a small impact on mean cost estimates (Table 4.6). However, multivariate partitioned Cox cost models that include variables for whether transplanted and centre reduce mean cost estimates by £16K compared to mean cost estimates from the partitioned Cox model adjusting for centre alone. Total cost estimates are reduced by a further £2K after including transplant group in the multivariate model.

Table 4.6 Mean observed costs under 10% random censoring and partitioned Cox cost estimates of costs per interval for centre, centre + transplanted (Y/N) and centre + transplanted (Y/N) + transplant group under 10% censoring

Months	Observed costs (£)	Partitioned Cox cost estimates (£)		
		Centre	Centre + transplant (y/n)	Centre + transplant (y/n) + transplant group
1 to 3	20,069	19,625	16,138	15,792
4 to 6	7,685	8,401	3,419	2,401
7 to 9	3,807	4,420	1,086	769
10 to 12	1,899	1,984	411	352
13 to 15	838	1,260	329	267
16 to 18	666	1,087	210	216
19 to 21	500	781	223	202
22 to 24	299	409	180	180
25 to 27	282	281	130	142
Total	***36,045***	***38,248***	***22,126***	***20,321***

The proportional hazards assumptions held for all Cox regression models.

Mean cost estimates for the Cox cost method and the partitioned Cox cost method were highly sensitive to the choice of model covariates, where the more significant the covariate the larger the impact on mean cost estimates. Given that these models are sensitive to the selection of covariates the level of impact on mean cost estimates will not be generalisable to other data sets. However, the assumption of independence between censoring and costs are violated for the Kaplan-Meier method, Cox cost method and partitioned Cox cost method and as a consequence all three methods fail to predict mean costs accurately and should not be used when data is censored.

4.7.2.2 Lin's Regression Models, KCH and UCH

Both of Lin's regression methods, KCH and UCH, involve fitting regression models to those patients who died or observe complete costs for either for the full study period (UCH) or per interval (KCH) [See Section 4.2.10 and 4.2.11]. Predicted costs are estimated for all patients (censored and uncensored) for the full study period or per interval. Costs are then predicted for all patients and weighted by the reciprocal of the Kaplan-Meier probability estimator (with reverse censoring) and a mean cost is then estimated. The choice of covariates will affect the predicted costs and the variation around the cost estimates but will not have any impact on the Kaplan-Meier weights. Therefore the remainder of this section focuses on how the selection of covariates in the regression models influences the predicted costs.

Univariate linear regression models were fitted to the complete (uncensored) costs for 726 CELT patients, prior to simulating censoring mechanisms. Age, gender and disease group were not significant predictors of costs at the individual covariate level, however costs differed significantly by centre, transplant group and whether transplanted (Table 4.7). The best fitting univariate model was for transplant group which explained the most variability in cost estimates and had the lowest root mean square error (MSE) in comparison to all other variables[10], costs were highest for patients receiving a transplant within 14 days (mean = £122,258) and lowest for unlisted patients (mean = £5,368). The model for whether transplanted or not also fitted the cost data well in comparison with other models, on average costs were £45,745 greater for those who received at least one liver transplant during the study compared with patients who did not receive one.

[10] The lower the MSE the better the model at predicting costs.

All univariate models accurately predicted the "true" mean costs (£36,045). However, all models failed to measure the "true" variation in the CELT data (Standard error = £1,517), with the best estimate of the standard error after adjusting for transplant group (£895) which is £622 lower than the observed value.

Table 4.7 Model goodness of fit statistics for predicting costs for 726 complete costs: univariate regression models

	F-statistic (p-value)	DF* for F-statistic	R^2	Root MSE	Estimated mean cost (£)	Estimated SE (£)
Age	2.01 (0.156)	1, 724	0.003	40,852	36,045	439
Gender	0.06 (0.813)	1, 724	0.0001	40,908	36,045	13
Centre	7.28 (< 0.001)	5, 720	0.048	40,023	36,045	333
Disease group	0.36 (0.780)	3, 722	0.002	40,935	36,045	59
Transplant group	96.17 (< 0.001)	4, 721	0.348	33,104	36,045	895
Transplanted (Y/N)	299.87 (< 0.001)	1, 724	0.293	34,401	36,045	821

*DF – degrees of freedom

A multivariate linear regression model was fitted to complete costs for all patients and variables that reached statistical significance at $p < 0.05$ were included in the model; transplant group and whether transplanted were the only statistically significant variables in this model[11]. Although this model predicts the mean cost (mean = £36,045) the mean variability around predicted costs i.e. the root MSE is high (31,599). Further, the model predicts high and low costs poorly with the minimum cost being almost £5K greater than the observed minimum cost for the data of £393 and the highest prediction almost £200K lower than the observed maximum cost of £311,873. Thus, the model under predicts the "true" standard error (£1,517) for the CELT data by £549.

[11] These covariates are the same ones as those used to estimate mean costs for Lin's regression method KCH and UCH after simulating censoring in Section 4.7.

In an attempt to improve the estimate of the standard error a further multivariate model was fitted to the 726 complete (uncensored) cases which included all covariates (age, gender, transplant centre, disease group, transplant group and whether transplanted) regardless of level of significance. The hypothesis behind this model was that the greater the number of variables in the model, the larger the variation in the predicted estimates, thus the more accurate the estimate of the standard error. The root MSE for this model remained high at 31,080 and the R^2 was 0.434. However, this model gave negative predictions of costs for 63 (9%) patients, these patients were assumed to have zero costs. After adjusting for negative costs the mean cost estimate was £36,121 and the standard error was still £521 less than the observed standard error. Therefore, for the CELT dataset it was reasonable to include only statistically significant variables in the regression models given that the inclusion of non-significant variables affected mean cost estimates and only marginally improved the standard error estimate.

So far the regression models fitted have been for complete (uncensored) costs for all 726 CELT patients, however, when censoring is present Lin's regression models should be fitted to complete (uncensored) costs. Therefore, the next step is to investigate whether predicted mean costs are still predicted accurately with models under predicting standard errors, after simulating censoring and fitting regression models to a reduced set of uncensored cases. Further, the accuracy of regression model estimates need to be verified for models fitted under UCH and KCH (i.e. fitting regression models per interval when the study is divided into a number of time periods). This hypothesis was tested by simulating censoring for two censoring mechanisms; 30% random censoring and informative censoring too ill (20^{th} percentile) and regression models were fitted with UCH and KCH using three monthly intervals (Table 4.8). As before mean estimates were exact with UCH and underestimated with KCH, presumably due to loss of information when dividing the study period into smaller intervals. Standard error estimates remained lower than the observed standard errors.

Table 4.8 Mean and standard error estimates from linear regression models under 30% random censoring and informative censoring too ill (20th percentile) fitted using models UCH and KCH

	Observed		Predicted estimates UCH		Predicted values KCH	
	Mean	SE	Mean	SE	Mean	SE
Random censoring (N = 504)	38,110	1,934	38,110	1,287	37,797	1,303
Informative censoring too ill (20th percentile) (N = 500)	30,977	1,801	30,977	1,204	27,660	1,111

This section has not estimated mean total costs in the presence of censoring and to complete this process predicted costs should be weighted by Kaplan-Meier estimates with reverse censoring. Instead, this section has focused upon accurately estimating mean costs and standard errors from linear regression models and the impact of covariates upon these estimates. The choice of covariates included in the regression models did not affect mean cost estimates or standard errors. Standard errors were poorly predicted, this can be attributed to the skewed nature of the CELT data (Figure 4.3) which did not follow assumptions of normality that are assumed when fitting linear regression models. The skewed nature of cost distributions is widely recognised and so the under prediction of the standard error estimates from fitting Lin's regression method are likely to occur in other studies. Further, cost data is bounded at zero and poorly fitting models may give negative cost predictions, which when adjusted for will result in an over estimation of mean costs.

4.7.3 Varying Interval Lengths

In this section the estimates of mean total costs in the presence of censoring are compared for a variety of interval lengths, across the six methods where it is necessary to divide the study time period in to smaller interval lengths. Given that rankings of the accuracy of methods did not vary significantly across censoring levels the methods are compared for five censoring mechanisms: 10% random censoring, 10% end-of-study censoring, 13% informative censoring too ill (10th percentile), 14% informative censoring too well (90th percentile) and fixed time resource collection partial censoring. Figure 4.10 presents the difference between the observed mean costs (prior to censoring) and the estimated mean costs for the six censoring mechanisms, for

different interval lengths (See Box 4.2 for legend). Appendix A4.4 presents tabled mean and standard error estimates for each of the six methods.

Box 4.2 Legend for Figure 4.10

■	10% Random censoring
×	10% End-of-study censoring (1.75 years)
▲	13% Informative censoring: too ill (10^{th} percentile)
▽	14% Informative censoring: too well (90^{th} percentile)
●	15% Partial censoring (fixed)

4.7.2.1 Weighted Cost Method with KCH

Mean cost estimates were the most accurate for random censoring and end-of-study censoring when interval lengths of six months were used. For informative censoring too ill, informative censoring too well and partial censoring, estimates were most accurate with three monthly interval lengths.

The least accurate estimates were from two monthly interval lengths, where mean costs were overestimated by between £7K and £12K across the five censoring mechanisms. These results contradict the findings of Bang and Tsiatis and O'Hagan and Stevens who suggest that a smaller interval length will give a more accurate mean cost estimate [Bang & Tsiaits, 2000; O'Hagan & Stevens, 2004]. This apparent contradiction is due to the loss of information from estimating Kaplan-Meier survival probabilities (with reverse censoring) for small interval lengths. With survival measured in monthly units and choosing two monthly interval lengths an event can only occur at one of two time points (month one or month two). This choice of interval lengths thus results in less accurate estimates of survival over time.

4.7.2.2 Lin's Regression Method with KCH

Lin's regression method is subject to the same problem as the weighted cost method and information on survival is lost when interval lengths of two months are used (mean costs were overestimated by £4K to £11K across censoring mechanisms). The best estimates of mean total costs are produced when selecting six monthly intervals and the difference between cost estimates and the true mean cost range between -£1,148 and £1,388 at 6 months.

4.7.2.3 Lin's method with KCH

For Lin's method KCH, the narrower the interval length the more accurate the estimate of mean total costs was, with the most accurate estimate occurring for monthly interval lengths and the least accurate for yearly interval lengths (underestimating mean costs by £6K to £9K).

4.7.2.4 Partitioned Cox Cost Method

The partitioned Cox cost method always underestimated mean study costs. The mean cost estimates were worst for monthly intervals (underestimated by approximately £17K) and best for yearly intervals for the partitioned Cox method. However, even the best estimates were approximately £12K less than the observed mean cost (£36,045).

4.7.2.5 Carides' Method

The accuracy of the mean estimates varied by censoring mechanism and no discernable pattern was observed.

4.7.2.6 Lin's Method with UCH

Under random censoring and end-of-study censoring the most accurate estimates of mean study costs were observed for monthly interval lengths, for both informative censoring mechanisms and partial censoring a three monthly interval length gave the best estimates.

For each of the six methods (weighted [KCH], Lin's regression [KCH], Lin [KCH], partitioned Cox cost, Carides' and Lin [UCH]) the results for the interval length that gave the most accurate estimate of mean total costs were selected. Mean (and standard error) estimates were compared with mean (and standard error) estimates for the six methods where it was unnecessary to divide the study period in to intervals (ignoring censoring, complete cost method, Kaplan-Meier cost method, Cox cost method, weighted cost method (UCH) and Lin's regression method (UCH)). Results were then ranked in terms of the accuracy of the mean (and standard error) estimate for the 12 censoring methods. Table 4.9 and 4.10 present the rankings for accuracy of the mean and standard error estimates.

Lin's method (KCH) is the most accurate method for estimating mean and standard error estimates when monthly interval lengths are selected. Lin's method with KCH (monthly interval lengths) showed the most marked improvements in mean cost

estimates and changed from being one of the poorest predictors of mean study costs to being one of the best predictors of mean study costs. The standard error estimates from Lin's method (KCH) under monthly intervals were also more accurate than when using three monthly intervals. The rankings for all other methods did not alter by more than three places.

Figure 4.10 Difference between observed mean costs (prior to censoring) and estimated mean costs (in the presence of censoring) for six censoring mechanisms by interval length

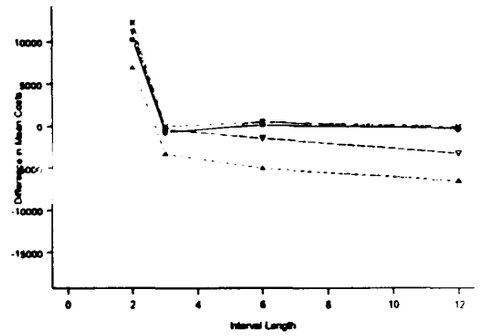
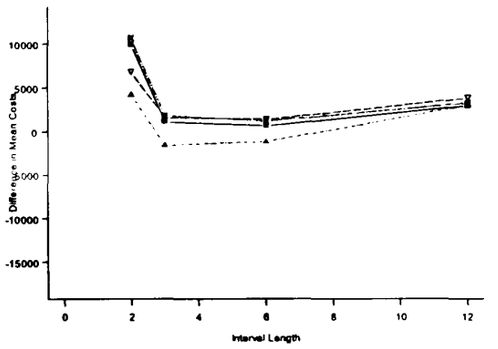
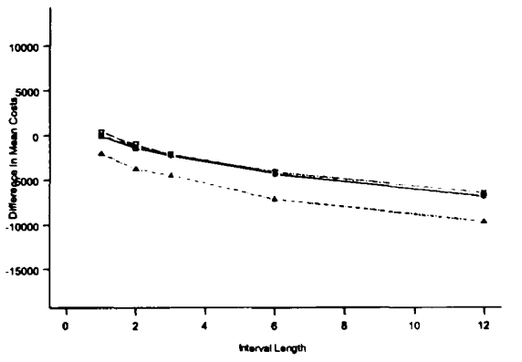
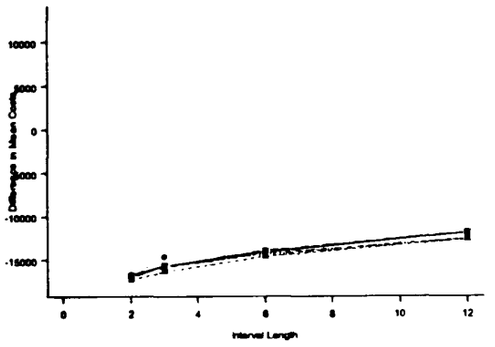
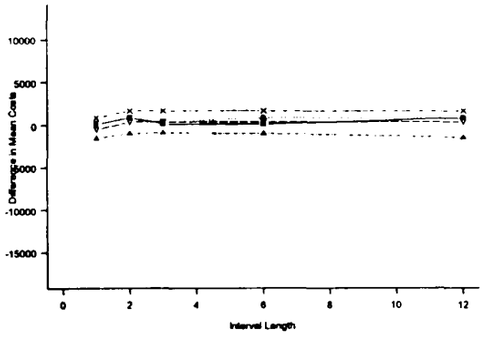
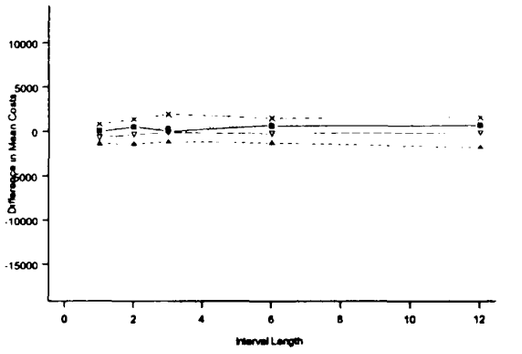
Weighted Cost method (KCH)	Lin's regression method (KCH)	Lin's method (KCH)
		
		
Partitioned Cox method	Carides' method	Lin's method (UCH)

Table 4.9 Ranking the methods for estimating mean total costs; ranked in terms of magnitude from the true mean (£36,045) of the uncensored CELT data

	10% Random Censoring		10% End-of-study Censoring (1.75 years)		13% Informative Censoring Too Ill (10^{th} percentile)		14% Informative Censoring Too Well (90^{th} percentile)		15% Partial Censoring (Fixed)	
	"New" Rank	"Old" Rank	"New" Rank	"Old" Rank	"New" Rank	"Old" Rank	"New" Rank	"Old" Rank	"New" Rank	"Old" Rank
Known cost histories										
WK: Weighted costs	(6 MI*) 6	7	(6 MI) 7	7	(3 MI) 8	7	(3 MI) 4	3	(3 MI) 3	2
RK: Lin's regression	(6 MI) 8	8	(6 MI) 9	8	(6 MI) 2	3	(6 MI) 9	8	(6 MI) 8	7
LK: Lin's method	(1 MI) 2	9	(1 MI) 2	9	(1 MI) 4	9	(1 MI) 3	9	(1 MI) 2	8
PC: Partitioned Cox	(12 MI)10	10	(12 MI) 10	10	(12 MI) 10	10	(12 MI) 10	10	(12 MI) 10	10
Unknown cost histories										
C: Carides	(6 MI) 3	4	(1 MI) 3	5	(3 MI) 1	1	(12 MI) 2	2	(1 MI) 4	6
IC: Ignoring Censoring	9	6	=4	2	9	8	8	7	1	1
LU: Lin's method	(3 MI) 1	1	(1 MI) =4	4	(6 MI) 3	2	(3 MI) 1	1	(1 MI) 5	5
RU: Lin's regression	7	5	1	1	5	4	6	5	=6	=3
WU: Weighted costs	5	3	6	3	7	6	7	6	9	9
CC: Complete cases	4	2	8	6	6	5	5	4	=6	=3
Cox: Cox regression	11	11	11	11	11	11	11	11	11	11
KM: Kaplan-Meier	12	12	12	12	12	12	12	12	12	12

"New" rank – using the interval lengths that give the most accurate mean estimates "Old" rank – using three monthly interval lengths * MI – Monthly intervals (choice of interval lengths used)

Table 4.10 Ranking the methods for estimating mean total costs; ranked in terms of magnitude from the true standard error (£1,517) of the uncensored CELT data

	10% Random Censoring		10% End-of-study Censoring (1.75 years)		13% Informative Censoring Too Ill (10^{th} percentile)		14% Informative Censoring Too Well (90^{th} percentile)		15% Partial Censoring (Fixed)	
	"New" Rank	"Old" Rank	"New" Rank	"Old" Rank	"New" Rank	"Old" Rank	"New" Rank	"Old" Rank	"New" Rank	"Old" Rank
Known cost histories										
WK: Weighted costs	(6 MI*) 5	1	(6 MI) 4	2	(3 MI) 2	2	(3 MI) 3	2	(3 MI) 2	3
RK: Lin's regression	(6 MI) 9	9	(6 MI) 7	7	(6 MI) 9	6	(6 MI) 9	9	(6 MI) 9	8
LK: Lin's method	(1 MI) 1	5	(1 MI) 1	6	(1 MI) 3	4	(1 MI) 2	5	(1 MI) 4	6
PC: Partitioned Cox	(12 MI)11	10	(12 MI) 11	10	(12 MI) 10	10	(12 MI) 11	11	(12 MI) 10	11
Unknown cost histories										
C: Carides	(6 MI) 7	7	(1 MI) 6	5	(3 MI) 7	5	(12 MI) 5	4	(1 MI) 7	7
IC: Ignoring Censoring	2	2	2	1	4	3	1	1	1	1
LU: Lin's method	(3 MI) 3	3	(1 MI) 3	3	(6 MI) 5	6	(3 MI) 6	6	(1 MI) 5	2
RU: Lin's regression	8	8	8	8	8	8	8	8	8	9
WU: Weighted costs	6	6	9	9	6	4	7	7	3	4
CC: Complete cases	4	4	5	4	1	1	4	3	6	5
Cox: Cox regression	10	11	10	11	11	11	10	10	10	10
KM: Kaplan-Meier	12	12	12	12	12	12	12	12	12	12

"New" rank – using the interval lengths that give the most accurate mean estimates "Old" rank – using three monthly interval lengths * MI – choice of interval lengths used (in months)

4.8 DISCUSSION

Of the 53 cost and cost-effectiveness studies included in the literature review in Chapter 3, 20 (38%) modelled the lifetime costs of solid organ transplantation and the remaining 33 studies (62%) estimated the cost-effectiveness or costs of solid organ transplantation over a fixed time period. Only three of the 33 papers gave enough information to establish whether censoring was an issue [Ohi *et al,* 1986; Garner & Dardis, 1987; Longworth *et al,* 2003] and all three studies elected to use Lin's method with KCH. In this chapter Lin's method (KCH) gave one of the most accurate estimates of mean total costs and standard errors in the presence of censoring when the study period was divided in to a series of monthly interval lengths.

The remaining 30 papers that reported the costs or cost-effectiveness of organ transplantation over a fixed time period did not provide enough information to assert whether censoring was an issue. We can only hypothese that censoring is likely to have occurred in at least a proportion of these 30 studies and has been ignored when estimating mean costs. Ignoring censoring will underestimate the mean total cost for a study, since any costs incurred by patients beyond the point they are censored are ignored. The magnitude of the underestimation is likely to depend on the proportion of censored cases, where the larger the proportion of censored cases the higher the degree of underestimation. However, the results in this chapter showed that the technique of ignoring censoring was one of the more accurate methods for estimating mean total costs. Therefore, it is likely that the cost and cost effectiveness estimates presented in the 30 studies that potentially ignored the issue of censoring produced more accurate mean cost estimates than if they had chosen an inappropriate method for accounting for censoring (for example the Kaplan-Meier method).

4.8.1 Inclusion of Naïve Methods

Naïve methods such as ignoring censoring, the complete case method and survival analytic time methods using a cost rather than time scale (Kaplan-Meier, Cox, partitioned Cox) were included in this chapter in order to illustrate the magnitude of the difference between observed and expected costs in comparison with other proposed methods. These methods were included despite the fact that it is now widely acknowledged that they will under or over estimate mean total study costs in the presence of censoring [Lin *et al,* 1997; Hallstrom & Sullivan, 1998].

The technique of ignoring censoring produced fairly accurate estimates of mean total costs and standard errors under some scenarios (end-of-study censoring restricted to the final six months of the study and partial censoring). It could be argued that, rather than applying one of the more complex methods described in this chapter, for example, Lin's method with or without cost histories, that simple methods could be used. However, ignoring censoring gave more accurate estimates of mean total study costs when censoring occurred towards the end of the study period, as was the case with end-of-study and partial censoring and the accuracy of this method could be due to the nature of the CELT dataset, where on average larger costs were incurred earlier in the study period at the point of the transplant operation and stabilised in the later end of the study period. Therefore, the results found here might not be generalisable to other cohorts where different patterns of resource usage are observed, this analysis should be repeated on other datasets to confirm these results.

Despite being widely criticised in the literature my experience is that the Kaplan-Meier cost method and Cox cost methods continue to be applied in practice when estimating mean total costs in the presence of censoring. Estimates of mean total study costs using the Cox cost method were almost double the observed mean cost and estimates from the Kaplan-Meier cost method were almost quadruple the observed mean, offering further indication that these methods should not be applied to censored cost data. However, it should be pointed out that the Kaplan-Meier method performs poorly if applied directly to censored cost data, and when applied to survival data as part of an estimation of mean study costs, such as Lin's methods, Lin's regression methods and weighted cost methods (KCH and UCH), produces more accurate estimates of mean study costs.

4.8.2 Magnitude for Measuring Accuracy

Kendall's concordance statistic was used to measure the agreement of the ranking of the accuracy of methods across different levels of censoring and different censoring mechanisms. The ranking approach was chosen over alternative methods, for example mean squared error, as it measured the consistency of methods across alternative censoring levels and mechanisms rather than measuring the magnitude of the precision of each method in estimating the mean or standard error.

The ranking of mean cost estimates differed significantly from the ranking of standard error estimates. Methods that were accurate at predicting mean costs, for example

Lin's method with UCH and Lin's regression method with UCH did not give the best estimates of the uncertainty surrounding the mean, as measured by the standard error. Researchers are usually interested in the uncertainty surrounding mean cost estimates in addition to a reliable estimate of mean total costs. Therefore, this chapter has focused upon comparing methods for estimating mean cost estimates and standard error estimates independently. In practice one would be more interested in selecting a method that accurately estimated mean total costs, with the accuracy of the standard error estimate becoming important when the mean estimate has been proven to be accurate. Lin's method with KCH with monthly interval lengths and the weighted cost method with KCH with three or six monthly interval lengths produced accurate estimates of both the mean cost and standard error estimates.

4.8.3 Interval Choices

Initially, the twelve methods for estimating mean total costs in the presence of censoring were compared and a common three monthly interval length was applied to the six methods which required the division of the study time period in to smaller interval lengths. Lin's method (UCH), Carides' method and Lin's regression method (UCH) were shown to be the best estimates of mean total costs under these circumstances. Interval lengths were then varied for the six methods where estimates relied upon dividing the study period in to smaller intervals, to investigate how the choice of interval length affected mean cost estimates. The most notable improvement in mean cost estimates was Lin's methods (KCH) using monthly interval lengths, where estimates improved by approximately £2,000 across methods and resulted in Lin's method (KCH) being one of the most accurate methods of estimating mean total costs in the presence of censoring.

O'Hagan and Stevens explored the accuracy of Lin's method (KCH) and the weighted cost method (KCH) under random censoring and concluded that the accuracy of Lin's method improves, the shorter the interval lengths used [O'Hagan & Stevens, 2004]. Results from the CELT study support these results for Lin's method, which gave estimates to within £1 of the mean total cost using monthly interval lengths under random censoring (30%).

Additionally, O'Hagan and Stevens stated that the weighted cost method of Bang and Tsiatis will produce similar estimates to Lin's method, though with KCH Lin's estimate makes use of information on censored costs per interval, whereas the weighted

estimates ignore this data, thus Lin's method will tend to produce more accurate estimates. For this study, when estimates using three monthly intervals were compared the weighted cost method consistently produced more accurate estimates than Lin's method. However, when interval lengths were varied, Lin's method gave a more accurate estimate of mean total costs than the weighted cost method did.

4.8.4 Censoring Mechanisms

Most of the methods listed here assume that censoring is independent of survival times and for informative censoring this assumption does not hold. Estimates of mean total costs varied more for informative censoring than for other censoring mechanisms. However, for Carides' method and Lin's method with UCH (three monthly intervals) estimates were still within £3,000 of the observed mean cost when informative censoring was due to ill health (20^{th} percentile 31% censoring). In reality censoring is unlikely to occur due to one type of censoring mechanism only, but through a combination of mechanisms, and despite all precautions taken during the study it is not always possible to determine why patients are censored during a study period. Future research could use a modelling process to combine different types of censoring and see how they impact on each of the methods for estimating mean total costs.

4.8.5 Censoring Levels

The ordering of the accuracy of methods did not differ significantly across different levels of censoring. This was true of all censoring mechanisms except partial censoring, where the ranking of methods for heavy censoring (one time resource collection censoring 80%) differed from that for partial censoring (fixed time resource collection censoring 15%). The cause of this discrepancy appears to be due to the inaccuracy of the mean predictions from Carides' method, weighted cost method (UCH) and Lin's method (UCH) under heavy censoring. Based on the results from this study it is recommended that alternative methods are used under heavy censoring, for example Lin's regression method (UCH), Lin's method (KCH), or the weighted cost method (KCH). However, these observations should be validated using other cohorts where censoring patterns and patterns of resource use differ from those observed in the CELT study.

4.8.6 Variants on Methods for Estimating Mean Total Costs

Variants to Lin's method (KCH), the weighted cost methods (KCH and UCH) and Lin's regression method (KCH & UCH) have been excluded from this review, despite being

published in the literature. Bang and Tsiatis' weighted cost estimators (KCH and UCH) belong to a wider class of estimators, the properties of which have been studied by several authors [Bang & Tsiatis, 2000; Zhao & Tian, 2001]. These estimators have not been used here as they require estimating the variance and covariance for the original estimator and we were unable to obtain sensible estimates (> 0) for the CELT data. These extensions to the weighted cost methods assume that cost are normally distributed, which is not true in the case for the CELT data, which are positively skewed (Figure 4.3). Owing to this skewness this confirms the observations of Carides and colleagues that "reliance on asymptotic [normality] theory is not recommended" and Raikou and McGuire who show that the extensions to the weighted cost estimator were unstable [Carides *et al,* 2000; Raikou & McGuire, 2004].

Lin and colleagues suggest excluding incomplete costs from the interval in which they are censored, as an alternative method (Lin's method with KCH). This approach assumes that censoring occurs only at the start of intervals, which was not true for the CELT study, therefore this method was ignored.

Huang uses calibration regression, an extension to Lin's regression methods (KCH and UCH) to estimate the lifetime costs of treatments or technologies using information collected over a limited time period [Huang, 2002]. This method was excluded because the aim of this chapter was to compare the accuracy of methods for estimating mean study costs over a fixed time period rather than the lifetime of the treatment. Finally, Jain and Strawderman suggest a generalisation to Lin's regression method with UCH [Jain & Strawderman, 2002]. Lin's regression method assumes that costs are linearly related to covariates (PH) whereas Jain and Strawderman's flexible hazards model makes no such assumption. However, Jain and Strawderman also note that their method is slightly less efficient than Lin's regression method if the PH assumption holds, which it did for the CELT data.

4.8.7 Application of Censored Cost Methods to QALY Data

As with censored cost data, survival analytic methods should not be applied directly to QALYs to estimate mean study QALYs in the presence of censoring. This is because, as with censored cost data, the assumption of independence between QALYs and the underlying censoring mechanism will not hold. However, there is no reason why Lin's method (KCH and UCH), Lin's regression method (KCH and UCH), the weighted cost method (KCH and UCH) and Carides' method cannot be applied to censored QALY

data. In fact, Zhao and Tsiatis originally proposed the weighted cost method to estimates mean QALYs in the presence of censoring [Zhao & Tsiatis, 1997]. Further research should be conducted to explore the accuracy of methods applied to QALY data to establish whether the same methods that produced accurate mean cost estimates produce accurate estimates of mean QALYs.

4.8.8 Simulating Censoring from the CELT data set

The approach taken in this thesis was to simulate censoring from a real data set rather than the alternative approach of creating a completely simulated dataset with known properties, for example with a known pattern or resource use. Although either approach is applicable the one taken was to simulate censoring from a real dataset to explore how mean total cost results were affected by censoring in a real observed situation. The advantage of taking this approach was to gain an understanding of how censoring would affect mean total costs in the extended five year CELT study and thus select an appropriate method for adjusting for censoring over a five year period.

4.8.9 Generalisability

The recommendations and conclusions that are presented in Chapter 4 are based on the data from one cost-effectiveness study in liver transplantation. The CELT study was subject to end-of-study censoring and had a distinct pattern of cumulative costs over time; costs are high at the beginning of the study period when transplantation occurs, and then level off towards the end of the study when patients are stabilised on drugs and typically are seen at outpatient appointments every six to twelve months. The pattern of resource use is likely to be observed in other solid organ transplant studies. Therefore, it is probably reasonable to assume that this pattern of censoring and resource use is generalisable to other cost and cost-effectiveness studies in solid organ transplantation.

However, the methods that produced the most accurate estimate of mean total costs i.e. Lin's method KCH with small intervals, might not give the most accurate estimates of mean costs in other data sets where the pattern of resource use differs to that observed in the CELT study. Each of the non-naïve methods for estimating mean total costs (Lin's methods KCH and UCH, weighted cost methods UCH and KCH, Lin's regression methods UCH and KCH and Carides' method), applies a form of weighting to costs, where for each method the weights are a variation of the Kaplan-Meier survival probabilities. Thus, Lin's method KCH weight mean costs per interval by the

Kaplan-Meier probability of survival in each interval, with mean costs in later intervals weighted lower than mean costs in earlier intervals. If a study observed an increase in resource usage over time, for example in studies of chronic conditions, then lower weights will be applied to the higher costs at the end of the study which could result in an underestimation of mean costs. Lin's method UCH and Carides' methods use a similar process and weight cost estimates by the probability of dying, where the largest weights occur in intervals with the greatest probability of death. Finally, the weighted cost methods and Lin's regression methods weight costs by the inverse of the Kaplan-Meier survival probability, with reverse censoring, thus inflating costs, with a higher level of inflation applied to costs in the later intervals of a study.[12] Therefore, the weighted cost method and Lin's regression method with UCH might overestimate mean cost estimates in studies where costs increase over time by over inflating costs at the end of the study period.

This chapter has explored the accuracy of 12 methods at estimating mean total costs for alternative censoring mechanisms for one pattern of resource use and has not explored the accuracy of methods for alternative resource use patterns, for example increasing costs over time or constant costs over time. Therefore, it is recommended that further work is carried out to establish whether the results here are generalisable to censored data sets outside the field of organ transplantation where alternative resource patterns are likely to be observed.

Additionally, HRQL is typically measured at one or more fixed time points during a study, patients may chose not to respond at particular time points and thus HRQL data could have a higher likelihood of informative censoring than resource use data. Given that the censoring mechanism may differ to that for resource use data caution is advised in generalising the results shown here to QALY. Further work should also be conducted to establish the generalisability of the results to censored QALY and cost-effectiveness data.

4.9 CONCLUSIONS

A number of messages have emerged after exploring an issue of methodological uncertainty by comparing 12 methods for estimating mean total costs in the presence of censoring. The 12 methods were compared across different censoring mechanisms

[12] If cost histories are known the study period is divided into several intervals and costs will be given a larger inflation rate in the later part of each interval in comparison to earlier parts of the interval.

and different levels of censoring, ranging from 10% (light censoring) to 80% (heavy censoring). It is assumed that the findings in Chapter 4 are generalisable to any study in solid organ transplantation where cost censoring is an issue. However, further work should be carried out to establish whether these results are generalisable beyond the field of organ transplantation and to censored HRQL or QALY data.

The accuracy of mean study cost estimates can vary by censoring mechanism and for the CELT study the initial results showed that:

- Lin's method (UCH) gave the most accurate estimate under random censoring
- Lin's regression method (UCH) gave the most accurate estimate under end-of-study censoring
- Carides' method gave the most accurate estimate under informative censoring
- the weighted cost method (KCH) or ignoring censoring gave the most accurate estimate under partial censoring

However, the choice of interval length affects the accuracy of mean study cost estimates for the six methods where the study time period can be divided in to smaller interval lengths (weighted cost method [KCH], Lin's regression method [KCH], Lin's method [KCH], partitioned Cox cost method, Carides' method and Lin's method [UCH]). Selecting the interval length that resulted in the best mean cost estimate and comparing the estimates across all 12 censoring methods showed that Lin's method (KCH) with monthly interval lengths gave the most accurate estimate of mean study costs across all censoring mechanisms and levels. Therefore, based upon the results of the CELT study, it is recommended that in the presence of censored cost data Lin's method with short interval lengths (KCH) is used to estimate mean study costs in the presence of censoring.

The results presented in Chapter 4 have also shown that a method that produces an accurate mean estimate does not necessarily produce an accurate estimate of the uncertainty (standard error) around the mean estimate. For the CELT study, Lin's method with short interval lengths (KCH) gave the most accurate estimates of the mean and its standard error. The weighted cost method (KCH) also produced accurate estimates of both the mean and standard error across all censoring mechanisms and levels.

In addition to the recommendations on the selection of the most appropriate method for estimating mean total costs in the presence of censoring, Chapter 4 has shown that three methods should not be applied when costs are not incurred at the end of the study period of interest. When no patient has complete costs for the full study period Carides' method; Lin's method (UCH) and the weighted cost method (UCH) gave estimates that substantially underestimates mean costs.

Censoring was anticipated as being an issue in the extension of the CELT study from 2.25 years to five years. Therefore, methodological uncertainties in techniques for estimating mean study costs in the presence of censoring were explored further in this thesis. The importance of exploring the methodological uncertainty in methods for estimating mean study costs in the presence of censoring was verified from the literature review conducted in Chapter 3 where 30 studies, potentially, ignored censoring when estimating mean study costs[13]. However, given that Chapter 4 has also shown that ignoring censoring produces one of the more accurate estimates of mean study costs, for the CELT study, the cost estimates from the 30 studies identified in Chapter 3 may have produced more accurate mean cost estimates than if they had selected an alternative method.

In the CELT study the issue of censored cost data was only relevant to the transplant arm of the study. For the non-transplant arm prognostic models were to be applied in order to estimate survival over the five-year study period and patients were predicted as being alive at five years or to have died during the study. However, prognostic models are subject to parameter uncertainty, model selection uncertainty and uncertainty in estimating patient specific outcomes. Therefore, the remaining chapters of this thesis (Chapters 5 to 7) present methods for estimating prognostic model uncertainties.

[13] These 30 studies did not provide enough information to assert whether censoring was an issue.

CHAPTER 5

MEASURING PROGNOSTIC MODEL PARAMETER UNCERTAINTY

5.1 INTRODUCTION

The previous chapter explored the first of two issues that arose when extending the CELT case study from 2.25 to five years: methodological uncertainty pertaining to the accuracy of methods for estimating mean study costs in the presence of censoring. Chapters 5 to 7 explore the second issue to be addressed in this thesis; measuring uncertainty in published prognostic models in the absence of patient specific non-transplant data. This chapter presents methods for estimating prognostic model parameter uncertainty.

Published prognostic models were used in the main CELT study to estimate survival in the absence of liver transplantation over 2.25 years for patients with end-stage ALD, PBC or PSC. It is proposed that the same prognostic models will be used to estimate non-transplant survival over the extended CELT study period of five years, with a view to estimating the longer-term effect of transplantation.

The prognostic models that were applied to the CELT cohort to estimate survival in the absence of transplantation were all Cox PH regression models. The Cox PH model described here is the fixed covariate[1] Cox PH model. The concept of the Cox PH model is to describe the survival of a cohort of patients based on additional information specific to the individual patient. Mathematically, this is achieved by expressing patient survival in terms of the patient's demographic or clinical characteristics, which are referred to as *explanatory variables*. The probability of survival at a particular time point t, written S(t), is modelled in terms of a set of *p* explanatory variables ($x_1,\ldots, x_p$) that have some influence on patient survival (e.g. patient age at time of treatment or levels of bilirubin as a measure of liver function). These factors are linearly combined to create a risk score R, also known as a prognostic index, (Equation 5.1), where larger values of the risk score indicate a greater risk or poorer prognosis.

$$R = \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_p x_{pi} \qquad \textit{Equation 5.1}$$

The terms $\beta_1,\ldots\beta_p$ are known as the *regression coefficients*, where β_j (j = 1,...,p) quantifies the effect that the explanatory variable x_j has on the risk score R. The risk score is then used to model the probability of surviving to time t (Equation 5.2)

$$\text{Probability of surviving to time t} = S(t) = S_0(t)^{\exp(R - R_0)} \qquad \textit{Equation 5.2}$$

In equation 5.2, R_0 is the risk score for a hypothetical patient with average values for x_j (for example average age, bilirubin levels etc.) and $S_0(t)$ is the survival probability for this hypothetical patient, or the probability of the hypothetical patient surviving to time point t. $S_0(t)$ is often referred to as the *baseline survival function*. To illustrate this, suppose that S(t) were expressed in terms of one explanatory variable, the patient's age at transplant and further suppose the average age of the cohort is 50 years. R_0 would then correspond to the risk score of a 50-year-old patient and $S_0(t)$ would correspond to the probability with which a 50-year-old patient survives to a time t. The survival probability for patients of other ages can also be calculated.

[1] In the fixed covariate Cox PH model described here the demographic and clinical characteristics are collected at one time point only and are assumed to remain constant over time.

The minimum amount of information required from a prognostic model in order to predict survival in another cohort, are the estimates of the regression coefficients, β_1, ...,β_p, and the baseline hazard or survival rates at given time points[2].

Given this information, it is possible to estimate the expected survival of patients had they not received the new treatment and compare this with their observed survival for patients who have received the new treatment. Thus, the survival gain of the new treatment may be estimated.

It is important to recognise that the degree to which each explanatory variable influences survival is not known exactly, but rather is an estimate based on the survival of a cohort of patients. As such, there is uncertainty associated with each of the estimates of the regression coefficients and the authors of some published prognostic models acknowledge this by providing their standard errors. This information is usually ignored when estimating control group survival [See: Neuberger *et al,* 1986; Bonsel *et al,* 1990a; Longworth *et al,* 2003]. Incorporating these standard errors in to the calculation of estimated survival gives a more accurate representation of the model uncertainty. Further, these regression coefficients will typically not be independent of each other. Therefore, an even more accurate representation of model uncertainty can be obtained if, in addition to the standard errors, the correlations between the regression coefficients are available.

In the CELT study, the prognostic models were an essential tool in the estimation of the costs, HRQL and QALYs of the medical management of end-stage liver disease, as the estimates of non-transplant costs, HRQL and QALYs were all functions of length of survival in the absence of transplantation. The main focus of this chapter is to introduce methods for estimating model parameter uncertainty; an earlier version of this work has been published by the author elsewhere [Young & Thompson, 2004]. A revised version of the PBC Mayo model[3] is used to illustrate the techniques and methods introduced.

---

[2] The hazard function h(t) and survival function S(t) are mathematically related $\left[h(t) = -\frac{d}{dt}\{\log S(t)\}\right]$.

[3] The Mayo model used in this chapter is based upon unpublished data (provided directly to the author) that differs from that used to fit the published prognostic models [Dickson *et al,* 1989; Murtaugh *et al,* 1994] in the main CELT study due to additional follow-up information and the correction of minor data errors in patient age and prothrombin times.

A secondary issue considered here was whether measures of prognostic model parameter uncertainty should be incorporated in to the selection criteria for an appropriate prognostic model when more than one prognostic model exists.

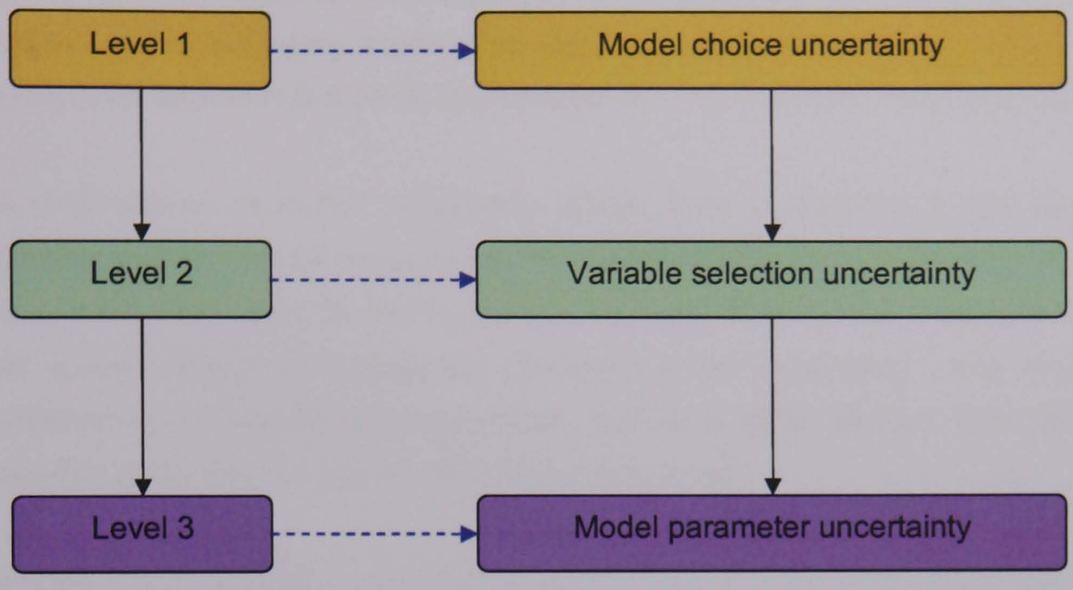
This chapter is divided in to a further nine sections. Section 5.2 presents an overview of the literature on model uncertainty. Section 5.3 then introduces the PBC CELT cohort in greater detail and presents details of the PBC Mayo clinic data. The section also presents results from a direct comparison of survival for two patient cohorts with, and in the absence of, liver transplantation. The chapter then goes on to describe how prognostic models can be applied to estimate survival in another cohort (Section 5.4). Monte Carlo simulation methods for adjusting for model parameter uncertainty are introduced in Section 5.5 and 5.6. Section 5.7 compares the levels of parameter uncertainty across three PBC prognostic models. Section 5.8 considers whether levels of model parameter uncertainty could be used as a model selection criteria, when more than one prognostic model is available, on the basis that the less the uncertainty the better the model. A discussion of some of the advantages and disadvantages of adjusting for model uncertainty are then discussed in Section 5.9.

5.2 AN OVERVIEW OF THE LITERATURE ON STATISTICAL MODEL UNCERTAINTY

In this section a brief overview of possible sources of model uncertainty is presented. Uncertainty is considered at three levels; model choice uncertainty (level 1), variable selection uncertainty (level 2), and model parameter uncertainty due to the underlying data set the model has been fitted to (level 3) [Figure 5.1].

Model uncertainty (also known as structural uncertainty) arises when selecting the mathematical structure of the model. For example, a researcher aiming to model the hazard or risk of death needs to decide between analysing survival at a fixed point in time (using logistic regression or a similar model) and analysing the actual survival times (using Cox regression or a similar model). The choice between these two approaches, and the choice of model with which to accomplish it, may not be obvious, since either may fit a data set equally well. Bayesian methods have been developed to allow for model selection uncertainty [see: Draper, 1995; Kang *et al*, 2000]. These methods involve running "repeated analysis utilising different models and specifying prior probabilities of different models across this model space" [Briggs, 2000].

Figure 5.1 The three levels of statistical model uncertainty



There are several issues that arise when selecting which variables are to be included in a mathematical model. Firstly, the variable selection techniques will depend on the requirements of the model and often on the analyst. In some cases, some or all variables will be included in a model regardless of their statistical significance, whereas in others only variables that are statistically significant will be included. Secondly, the analyst needs to select the significance levels for p-values, where variables with a p-value lower than a chosen level are included in the model. Thirdly, there is the method of variable selection. If *forward selection* is used, each of the variables are fitted to the model one at a time, with the single best fitting variable (the one that is the most significant predictor) being included in the model. The unselected variables that remain are then fitted, one at a time, to the new model, which now includes the best fitting variable. If the model is significantly improved by the inclusion of a second variable then the model is again updated with the most significant of the variables. The process is continued until none of the remaining variables significantly improve the model fit. Conversely, in *backwards selection* all variables are fitted to the model and non-significant variables are removed one at a time until only significant variables remain.

The use of forward and backward selection can result in different combinations of variables in the final models and choice of technique will depend on the preference of the analyst and on the study objectives. Collett states that "there are likely to be a number of equally good models, rather than a single 'best' model" and recommends that analysts consider a number of alternative models based on different combinations

of variables [Collett, 1994]. In contrast, Wang *et al* compared frequentist stepwise variable selection methods (including forward and backward stepwise regression) with Bayesian model averaging approaches using simulated data sets and conclude that the Bayesian approach is a preferable variable selection method [Wang *et al,* 2004].

The final source of model uncertainty arises from uncertainty in the parameter estimates and this can be measured by the standard error of the regression coefficient or the covariance matrix for the regression coefficients. Numerous papers in different topic areas suggest techniques for allowing for this uncertainty using simulation, bootstrapping or Bayesian techniques [See: Gigli *et al*, 2000; Babyak, 2004; Bertrand-Krajewski, 2004; Cox & Popken, 2004; Kawrakow 2004].

The above discussion on model uncertainty is not meant to be a comprehensive review of the literature but seeks to give an overview of possible sources of statistical model uncertainty. The objective of this chapter is to identify techniques for adjusting for uncertainty in published mathematical models applied to another data set and all the techniques described in this section are derived to measure model uncertainty in an observed data set. No literature was found on methods for incorporating model uncertainty when fitting an existing model to another data set. Therefore, Sections 5.5 and 5.6 set out a simulation method that allows for this.

5.3 A COMPARISON OF OBSERVED SURVIVAL WITH AND IN THE ABSENCE OF LIVER TRANSPLANTATION

This section introduces the PBC Mayo clinic cohort, from whom a prognostic model is derived for estimating survival in the absence of liver transplantation. The model is applied to 81 CELT patients with end-stage PBC to estimate, what would have been, their survival in the absence of transplantation over the extended five year CELT study period. This section also introduces the PBC CELT cohort and compares demographic and clinical characteristics for the CELT cohort with the PBC Mayo cohort over five years. The remainder of Section 5.3 describes methods for estimating survival with and without transplantation and the survival gain from liver transplantation using a historical control group to estimate non-transplant survival.

5.3.1 PBC Mayo Cohort – The Non-Transplant Group

Between January 1974 and April 1984 a total of 312 patients with PBC participated in one of two RCTs at the Mayo Clinic, Rochester, Minnesota. The objective of both trials

was to evaluate the therapeutic effect of D-penicillamine as a treatment for PBC in comparison with a placebo drug [Dickson *et al,* 1985]. The two trials differed in that one compared the drug in patients with histological stage 1 or 2 PBC[4], and in the other, patients had stage 3 or 4 PBC. Neither trial found a therapeutic difference between the treatment and control arms of the trial.

In both trials, detailed demographic, clinical and biochemical information were collected for each patient every time they visited the Mayo clinic for treatment for PBC. The patient and clinical characteristics provided in the full Mayo data set consisted of: patient age, gender, presence or absence of ascities, presence or absence of hepatomegaly, presence or absence of spiders, oedema score[5], albumin, serum bilirubin, serum cholesterol, alkaline phosphate, serum glutamic-oxaloacetic transaminase, platelets, prothrombin time, histological stage, survival length and patient outcome.

Patients were followed up until April 1988, yielding a median follow-up time of 6.3 years (IQR: 3.7 years to 8.9 years). Patients had between one and sixteen visits to the Mayo clinic during the study period.

A total of 29 (9.3%) patients received a liver transplant during the two trials. These patients were censored in the original studies at point of transplant and no further clinical information was collected on these patients.

5.3.2 PBC CELT Cohort – The Transplant Group

Eighty-one patients with end-stage PBC underwent liver transplantation during the CELT study period. The main CELT study collected survival information up-to two years post transplantation. Survival outcome was updated in September 2001 when all six transplant centres who participated in the study provided information on survival and date of death, if applicable, for all transplanted patients. The median study follow-up time for the extended study period was 4.8 years (IQR: 4.4 to 5.3 years).

Information was available on the age and gender of all patients. Clinical information was collected on all patients immediately prior to transplantation and included: serum

[4] Histological stage is a categorisation of disease severity based on liver biopsy results; stages 1 and 2 are less severe than 3 and 4 where patients have fibrosis or cirrhosis of the liver.
[5] Oedema score: 0 = no oedema and no diuretic therapy for oedema, 0.5 = oedema present without diuretic therapy or oedema resolved by diuretics, 1 = oedema present despite diuretic therapy.

bilirubin levels, presence or absence of ascities, oedema score, serum albumin levels, prothrombin time, urea, creatinine levels, encephalopathy and blood group.

5.3.3 Estimating the Survival Gain of Liver Transplantation – Kaplan-Meier Method

Given that data were available for a non-transplant cohort of patients (PBC Mayo cohort) an intuitive first step in estimating the survival gain of liver transplantation is to use the available data. Therefore, Sections 5.3.3 and 5.3.4 explore a series of methods for estimating non-transplant survival and the survival gain from transplantation using the original Mayo data to estimate non-transplant survival, these results will be compared with non-transplant survival estimates using prognostic models.

The simplest comparison of survival with, and in the absence of, transplantation is to estimate the survival for each cohort by the Kaplan-Meier method.

5.3.3.1 Kaplan-Meier Analysis 1: Comparing PBC CELT survival from date of transplant with PBC Mayo patient survival from date of first visit

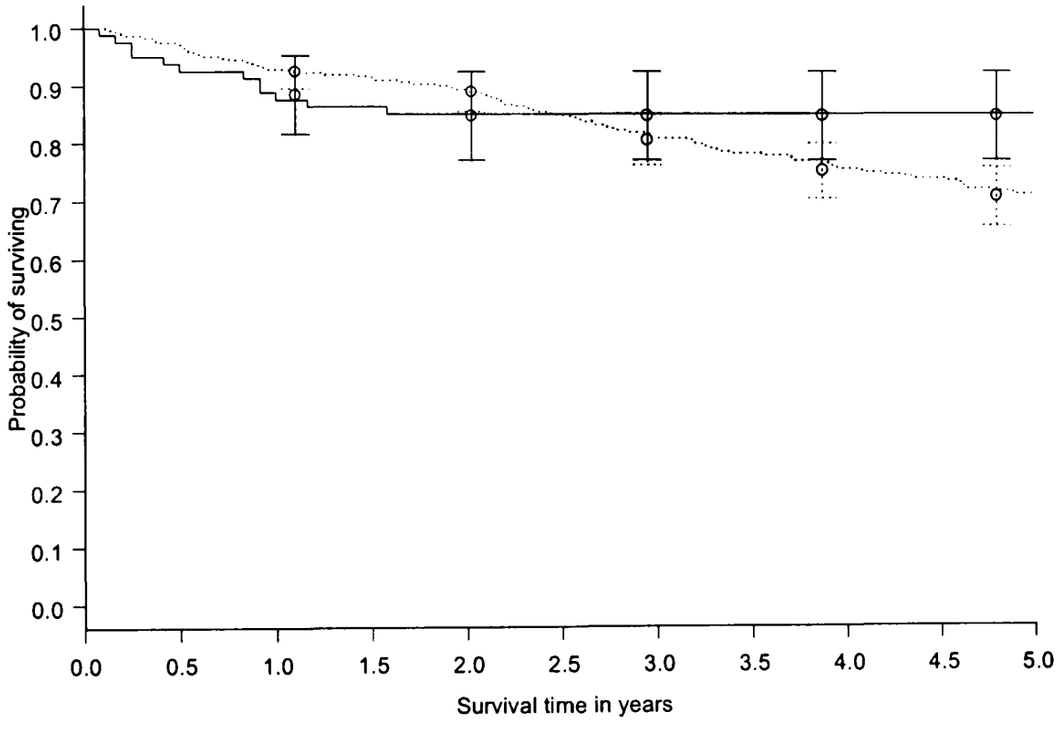
Figure 5.2 presents the Kaplan-Meier survival curves for the PBC Mayo cohort (non-transplant group) and the PBC CELT transplant group. The survival curves cross at approximately 2.5 years, suggesting that the apparent early survival advantage in the absence of transplantation ends at this point, and the transplant patients have better survival over the remainder of the five-year period. This result is not entirely unexpected. The most likely time period for mortality post-liver transplantation is in the six-month period that follows transplantation. During this time, patients are more likely to suffer from post-operative complications, infections and rejection episodes. However, once this period has elapsed, prognosis tends to stabilise. By contrast, mortality will occur at a more constant rate over time in the non-transplant group.

The mean survival for each cohort was calculated from the area under the Kaplan-Meier survival curves [Collett, 1994]. Greenwoods formula was applied to estimate 95% CI around the two means. The mean survival gain from transplantation is the difference between the areas for the non-transplant and transplant survival curves.

The mean survival time for the transplant group was 4.4 years (95% CI: 4.0 to 4.7 years) which was similar to the mean survival for the non-transplant group of 4.3 years (95% CI: 4.1 to 4.4 years). The proportion of deaths over five years post-transplant for

the CELT cohort (12 [15%]) was lower than the number of deaths over five years for the Mayo cohort (88 [28%]). The mean survival gain from transplantation over five years suggested that there was no significant difference in survival between the two groups (Mean transplant survival gain = 0.11 years, 95% CI: -0.26 to 0.48 years; Wilcoxon χ_1^2 test = 3.24, p = 0.072).

Figure 5.2 Kaplan-Meier Survival Curves with 95% CI at 1, 2, 3, 4 and 5 years for the PBC Mayo Cohort (non-transplant group; N = 312, dotted line) and the CELT Cohort (transplant group; N = 81, solid line)



5.3.3.2 Is Kaplan-Meier Analysis 1 providing a fair comparison?

The above analysis compared survival between transplanted and non-transplanted patients. However, it is only valid to interpret this as a comparison between transplantation and non-transplantation if the two patient cohorts are themselves comparable. To understand the difficulties in making this inference, it is necessary to take a step back from the data and to consider the clinical characteristics of PBC and the criterion by which the cohorts were defined.

PBC is a chronic progressive disease consisting of a pre-symptomatic stage that lasts approximately twenty years, a symptomatic stage that lasts between five and ten years, and a pre-terminal/accelerated phase that lasts approximately two years [Pasha &

Dickson, 1997]. Patients in the CELT study have end-stage liver failure and are highly likely to be in the pre-terminal phase of PBC. On the other hand, patients in the Mayo cohort may be in any of the three stages, though most likely to be in the symptomatic or pre-terminal stage. Since the objective of the CELT study was to estimate the impact of liver transplantation among patients with end-stage PBC, the study population (and the patients to whom this research is directed) are patients with end-stage PBC. An appropriate comparator cohort should therefore comprise of PBC patients whose disease has progressed towards or in to the pre-terminal stage.

The Mayo study collected clinical information on study patients over a series of time points. Table 5.1 presents the clinical and demographic characteristics of the Mayo group at first and final study visit[6]. The median time between first and final visits was 3.7 years (IQR: 1.5 to 7.2 years). During the study the patient's health deteriorated due to the progression of the liver disease[7]; bilirubin levels and prothrombin times increased ($t = -10.04$, $p < 0.001$, and $t = -9.71$, $p < 0.001$, respectively), albumin levels decreased slightly ($t = 13.13$, $p < 0.001$) and the number of patients with oedema (with or without diuretics) increased as did the proportion of patients with ascities ($\chi^2_2 = 43.50$, $p < 0.001$, and $\chi^2_1 = 40.31$, $p < 0.001$, respectively). Furthermore, Table 5.1 shows the Mayo patient cohort at their first visit are at less risk of death than the CELT patients, but are more similar to the CELT cohort at the time of their final visit.

[6] The clinical characteristics presented in Table 5.1 were commonly collected in both the Mayo and CELT studies.

[7] Higher bilirubin levels, longer prothrombin times, lower albumin levels, higher oedema scores and the presence of ascities suggest more severe disease levels.

Table 5.1 Demographic and clinical characteristics for the Mayo non-transplant cohort at the time of first and final study visits and the PBC CELT cohort immediately prior to transplantation

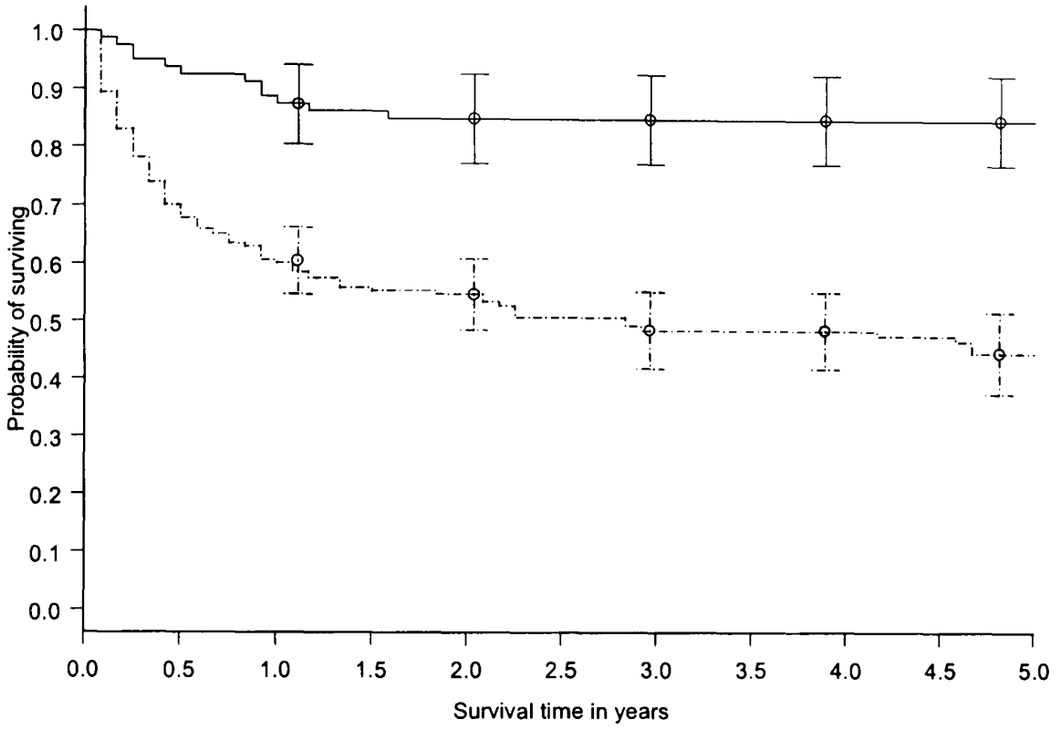
	Mayo cohort First visit (N = 312)	Mayo cohort Final visit (N = 312)	CELT cohort Point of transplant (N = 81)
Age in years (SD*)	50.0 (10.6)	54.7 (10.9)	55.2 (8.1)
Serum bilirubin in mg/dl (SD)	3.2 (4.5)	7.2 (8.2)	8.1 (8.7)
Serum albumin in g/dl (SD)	3.5 (0.4)	3.1 (0.6)	3.1 (0.7)
Prothrombin time in seconds (SD)	10.7 (0.9)	12.0 (2.4)	16.0 (4.4)
Females	296 (89%)	296 (89%)	73 (90%)
Oedema score 0	247 (79%)	165 (53%)	50 (62%)
Oedema score 0.5	44 (14%)	79 (25%)	11 (14%)
Oedema score 1	21 (7%)	68 (22%)	20 (26%)
Ascities present	24 (8%)	84 (27%)	41 (51%)

5.3.3.3 Kaplan-Meier Analysis 2: Comparing PBC CELT survival from date of transplant with Mayo patient survival from date of last visit

It was decided that a more appropriate comparison group for the CELT survival post transplant was the survival in the Mayo cohort from final study visit onwards (Figure 5.3). The remainder of the non-transplant analysis in this chapter uses information from each patient's final Mayo study visit.

In the Mayo cohort, a total of 136 (44%) patients died over the five-year period following their final study visit, which is almost triple the proportion that died post-transplant in the CELT study (12 [15%]). The Kaplan-Meier survival curve shows that survival without transplantation is consistently lower than survival post-transplant over the five-year period (Wilcoxon χ_1^2 test = 31.4, $p < 0.001$). The mean survival post transplant was 4.4 years (95% CI: 4.0 to 4.7 years) and in the absence of transplantation was 2.8 years (95% CI: 2.5 to 3.0 years) with a mean transplant survival gain of 1.59 years (95% CI: 1.15 to 2.03 years). In contrast to Kaplan-Meier Analysis 1, these results imply that transplantation is an effective treatment of choice for patients with end-stage PBC.

Figure 5.3 **Kaplan-Meier Survival Curves with 95% CI at 1, 2, 3, 4 and 5 years for the PBC Mayo Cohort from final study visit (non-transplant group; N = 312, dashed line) and the PBC CELT Cohort (transplant group; N = 81, solid line)**



5.3.3.4 Is Kaplan-Meier Analysis 2 providing a fair comparison?

By comparison to Analysis 1, the above analysis is more realistic, since the patients are more alike at the outset. Despite this, Table 5.1 still suggests that there are differences between the demographic and clinical characteristics of the PBC CELT cohort at point of transplant and the PBC Mayo cohort at final study visit. The two cohorts were similar in mean age and the proportion of female patients. However, the CELT cohort had significantly higher serum bilirubin levels, prothrombin times and a higher proportion of patients had ascities than the Mayo cohort ($t = 2.37$, $p = 0.018$; $t = 12.14$, $p < 0.001$; $\chi_1^2 = 16.64$, $p < 0.001$, respectively), suggesting that at time of transplant the CELT patients were at worse prognosis than the Mayo cohort, even at their final study visit.

The survival curves, as estimated by the Kaplan-Meier method, do not take in to account the differences in patient characteristics between the Mayo and CELT cohorts. Consequently, it is quite likely that even the second analysis is underestimating the true

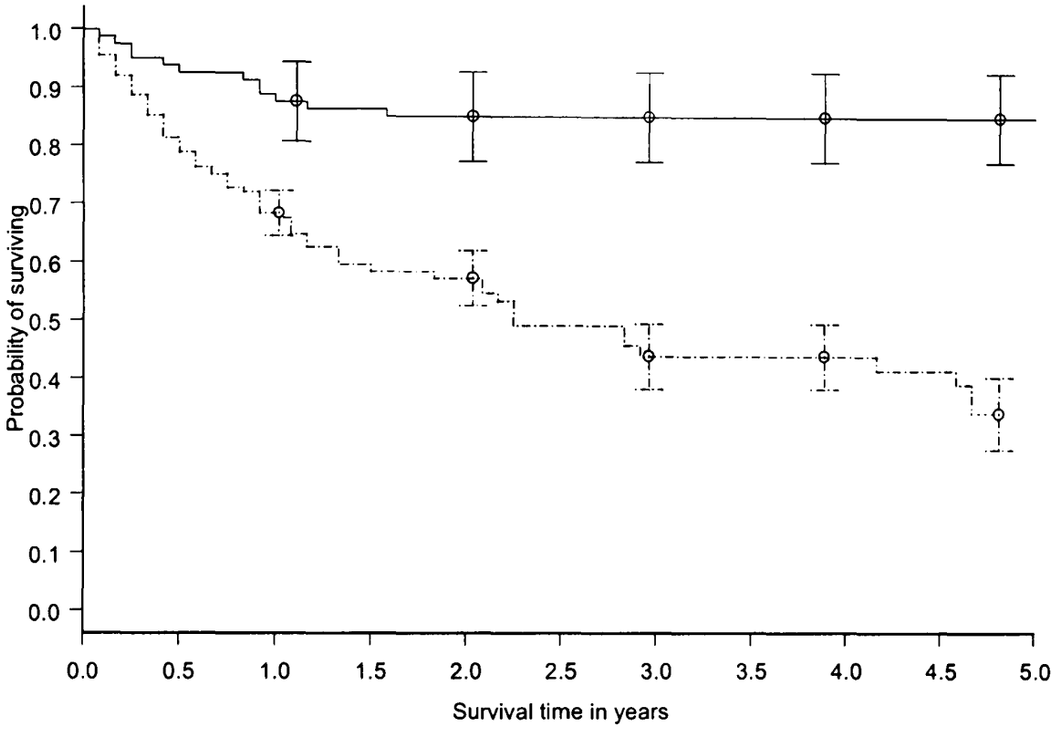
benefit of transplantation over the medical management of end-stage liver failure. In Chapter 3, we discussed how differences between historical and concurrent cohort populations can cause biases in the results and that it is possible (to some extent) to adjust for these differences provided that the information on the differences is collected. There appears to be a difference in disease severity between the CELT and Mayo cohorts and it is possible to adjust for such differences using approaches such as Cox PH regression analysis.

5.3.4 Estimating the Survival Gain of Liver Transplantation – Cox Method

The patient specific data from the PBC CELT and PBC Mayo cohorts were combined in to one data set to adjust for the differences in patient characteristics between cohorts. A Cox PH regression model was employed to obtain an adjusted estimate of survival in the absence of transplantation and this estimate was used to calculate the survival gain from transplantation. Variables that were common to both data sets were adjusted for in the Cox model: patient age, gender, serum bilirubin levels, prothrombin time, serum albumin, oedema score and presence or absence of ascities. The assumptions of proportionality were checked and held for each variable. Appendix A5.1 presents the Cox PH model results.

Figure 5.4 below, presents the unadjusted survival curve for the PBC CELT cohort (transplant group) and the predicted survival curve for the PBC Mayo cohort (non-transplant group) from date of final study visit, adjusted for the demographic and clinical characteristics using the Cox PH model described above. The mean survival for transplant patients was 4.4 years (95% CI: 4.0 to 4.7 years). The mean estimate of survival over five years in the absence of transplantation was 2.6 years (95% CI: 2.4 to 2.9 years), slightly lower than the unadjusted estimate presented in Section 5.3.3.3 (2.8 years). This resulted in an estimated increase in the mean survival gain over five years from transplantation (Mean = 1.72 [95% CI: 1.68 to 1.76 years]).

Figure 5.4 **Adjusted Kaplan-Meier Survival Curves with 95% CI at 1, 2, 3, 4 and 5 years for the Mayo Cohort from final study visit (non-transplant group; N = 312, dashed line) and the CELT Cohort (transplant group; N = 81, solid line)**



5.3.5 Summary of Results

No significant gain in survival after transplantation was evident over five years when using non-transplant Mayo data from first study visit. However, there was evidence of a mean survival gain of 1.59 to 1.72 years when using Mayo data from final study visit, both with and without adjustment for clinical and demographic patient characteristics. The most reliable of the three estimates of the survival gain was 1.72 years, since this estimates the difference in survival between transplantation and medical management after adjusting for differences in clinical and demographic data between the transplant and non-transplant cohorts.

However, it is not always possible to collect person-specific data from a historical cohort of patients and this was the case in the CELT study, where historical non-transplant data became available for the PBC group (Mayo non-transplant cohort) only after the main CELT analysis had been published. The CELT study therefore used a different approach, although one that still relies on the Cox PH regression model. In this, published prognostic models were used to estimate the non-transplant survival of

patients with end-stage liver disease and consequently QALYs and costs in the absence of transplantation. The use of prognostic models to estimate non-transplant survival is described in detail below.

5.4 USING PROGNOSTIC MODELS TO PREDICT NON-TRANSPLANT SURVIVAL

A prognostic model is essentially a mathematical formula that attempts to predict an outcome such as length of survival. This formula is derived on the basis of observed data from previous patients, with the intention of providing a prognosis for future patients. It is possible to use the information from a prognostic model to estimate the survival for a particular treatment or condition. It is also possible to compare survival estimates from a prognostic model with observed survival for patients from a comparator cohort, for example, comparing survival with, and in the absence of, liver transplantation. Thus, assuming the patients on whom the prognostic model was derived can be generalised to the present cohort (other than in their therapeutic treatment), the survival gain of a new treatment or technology may be estimated.

The remainder of this section illustrates this process in detail. The prognostic model described here was derived from fitting a fixed covariate Cox PH model to the person specific PBC Mayo data, from point of final study visit. Patient age, gender, bilirubin levels, albumin levels, prothrombin time, oedema score and the presence or absence of ascities were included in the model. The assumptions of proportionality were checked and held for each variable (Appendix A5.1). Table 5.2 presents the prognostic model regression coefficients and standard errors and Table 5.3 presents the baseline survival [$S_0(t)$] over five years in 0.25 year increments.

Bilirubin levels, age, albumin levels, oedema scores and gender were significant predictors of non-transplant survival in the Mayo cohort, whereas prothrombin time and ascities were not significant. The probability of survival without transplantation decreases with increasing age, increasing bilirubin levels and decreasing albumin levels. Patients with an oedema score of 0.5 or 1 were less likely to survive than patients with a score of 0 and female patients had a greater probability of surviving than male patients. The variables were included in the model regardless of the level of statistical significance.

Table 5.2 Regression coefficients and their standard errors for a fixed covariate Cox PH model fitted to the PBC Mayo cohort (N = 312)

	Regression coefficient	Standard error	Z	p-value
Log_e(bilirubin in mg/dl)	0.87	0.11	7.65	< 0.001
Albumin in g/dl	-0.93	0.22	-4.31	< 0.001
Oedema score 0.5*	0.45	0.23	1.93	0.053
Oedema score 1	0.52	0.25	2.06	0.040
Ascities present	0.19	0.22	0.86	0.390
Gender: female	-0.64	0.25	-2.59	0.010
Age in years	0.03	0.01	2.96	0.003
Log_e(prothrombin time in seconds)	1.10	0.66	1.67	0.096

* Oedema scores: 0 – no oedema and no diuretic therapy for oedema; 0.5 – oedema present without diuretics or oedema resolved by diuretics; 1 – oedema present despite diuretics (score 0 used as a base score in the model)

The baseline risk score (R_0) for an average patient from the PBC Mayo cohort (a 54 year old female, with a serum bilirubin level of 7.2, a serum albumin level of 3.07 and a prothrombin time of 12 seconds with no oedema [score 0] and no ascities) is 2.41.

Table 5.3 **Baseline survival estimates $S_0(t)$ up to five years for a fixed covariate Cox PH model fitted to the PBC Mayo cohort (N = 312)**

Time in years	$S_0(t)$
0.25	0.917
0.50	0.835
0.75	0.782
1.00	0.734
1.25	0.686
1.50	0.644
1.75	0.644
2.00	0.632
2.25	0.547
2.50	0.547
2.75	0.547
3.00	0.489
3.25	0.489
3.50	0.489
3.75	0.489
4.00	0.489
4.25	0.462
4.50	0.462
4.75	0.435
5.00	0.379

This prognostic model, detailed in Table 5.2 and 5.3, can now be fitted to the PBC CELT cohort in order to estimate their survival in the absence of liver transplantation. The model is applied to each patient in the CELT cohort to obtain individual prognostic scores. For example, a 55-year-old female patient with a serum bilirubin level of 8.1mg/dl, a serum albumin level of 3.1g/dl, with a prothrombin time of 16 seconds with no oedema and ascities has a risk score of 3.02 (Equation 5.3) and an adjusted risk score ($R - R_0$) of 0.61 (3.02 – 2.41).

$$R = (0.03 \times 55) - 0.64 + (0.87 \times \log_e(8.1)) - (0.93 \times 3.10) + (1.10 \times \log_e(16.0)) + 0.19 = 3.02 \qquad \textit{Equation 5.3}$$

This score can be transferred in to the expected probability of surviving to time point t, using Equation 5.2 and the baseline survival results presented in Table 5.3. Table 5.4 presents the 1, 2, 3, 4 and 5-year survival probabilities for the example patient. Here, the patient has a 57% probability of surviving one-year without a transplant, which decreases to 17% at five years.

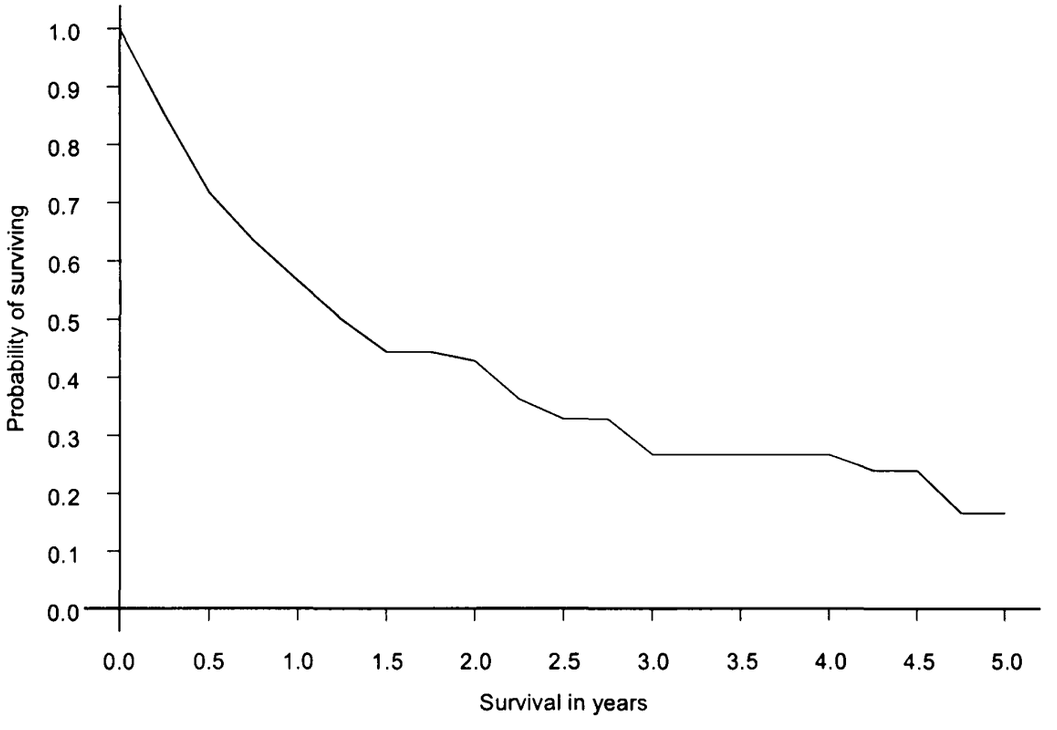
Table 5.4 Survival probabilities at 1, 2, 3, 4 and 5 years for a 55-year-old female patient with a bilirubin level of 8.1mg/dl, albumin level of 3.1g/dl and prothrombin time of 16 seconds, no oedema, with ascities

	1 year	2 years	3 years	4 years	5 years
Probability of surviving	0.566	0.430	0.268	0.268	0.168

These probabilities can be plotted over time to give an individual profile of expected survival in the absence of transplantation and the area under the curve denotes the patients expected survival length (Figure 5.5). The example patient has a predicted non-transplant survival of 2.10 years over the five-year time period.

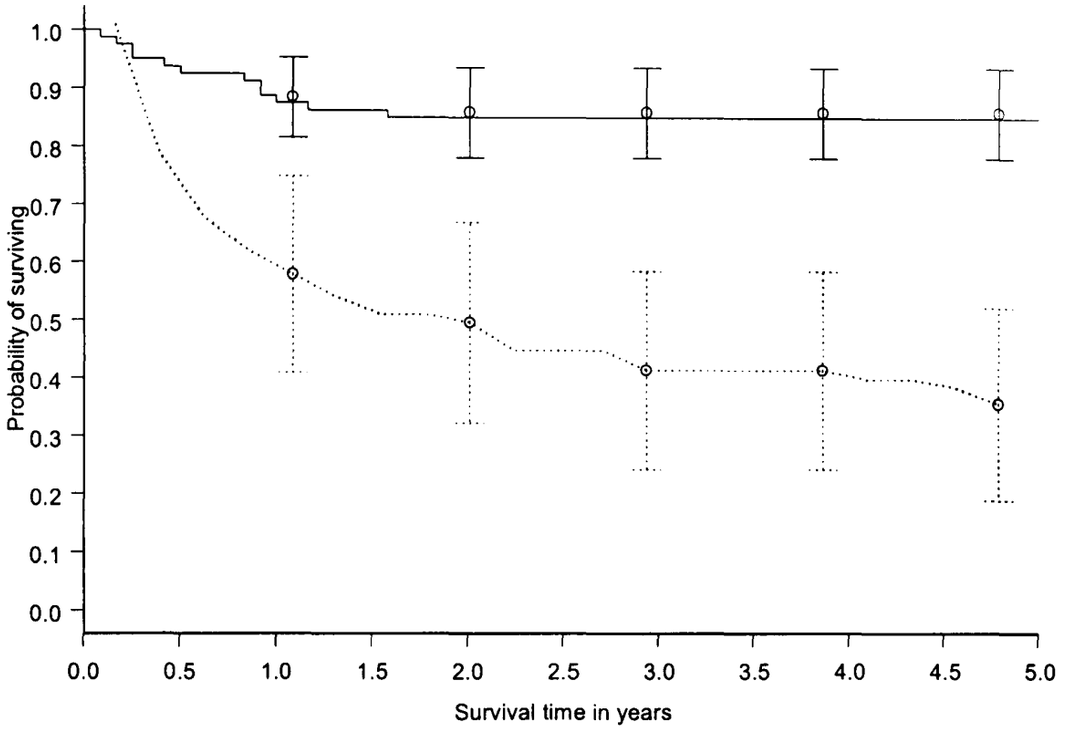
Using this approach, non-transplant survival can be calculated for each patient in the cohort. Overall, the estimated average length of survival in the absence of transplantation for the CELT cohort is 2.5 years (95% CI: 2.2 to 2.8 years) over the five-year period. More importantly, each individual patient in the cohort now has a survival duration estimated in the absence of transplantation with which to compare to his or her observed survival time following transplantation. It is possible to estimate each patient's individual survival gain by subtracting the estimated non-transplant survival from each patients observed transplant survival. The mean survival gain over five years from transplantation is 1.69 years (95% CI: 1.24 to 2.13 years).

Figure 5.5 **Estimated survival in the absence of transplantation for a 55-year-old female patient with a bilirubin level of 8.1mg/dl, albumin level of 3.1g/dl and prothrombin time of 16 seconds, no oedema, with ascities**



The Kaplan-Meier transplant survival curve for the PBC CELT cohort and the mean non-transplant survival estimates, derived from the PBC Mayo prognostic model, are presented in Figure 5.6 with 95% CI to show cohort uncertainty at 1, 2, 3, 4 and 5 years. It is important to note that the CI represent cohort uncertainty, but not model parameter uncertainty. In other words, the variability between patients is accounted for by the CI, but the non-transplant survival estimates are assumed to be a fixed quantity. The CI does not incorporate any of the uncertainty in the prognostic models estimates.

Figure 5.6 **Kaplan-Meier transplant survival (solid line) and non-transplant survival estimated from the PBC Mayo prognostic model (dotted line) for 81 PBC CELT patients (with 95% CI for cohort uncertainty)**



So far this chapter has presented summary statistics such as the mean sample estimates of survival with and without transplantation, or the mean survival gain from transplantation. These are supplemented with a 95% CI, which represents the uncertainty around the estimate and allows inferences to be drawn as to how close the sample mean is likely to be to the population mean [Bland, 2000]. The basic logic is as follows: whenever a sample mean is estimated, it is highly unlikely that this number will be exactly equal to the true population mean value. The CI allows a plausible range to be proposed within which the true population mean is expected to lie. The same process and logic is true when fitting statistical models. Just as the sample mean is an estimate of the population mean, the regression coefficient used in a prognostic model is not the true effect of that prognostic factor, but is an estimate of it. It is possible (and wise) to derive CI in order to represent uncertainty for each of the estimated regression coefficients ($\hat{\beta}_j$) that are included in a statistical model. For example, using the standard errors for the Mayo prognostic model for bilirubin (Standard error = 0.11 see Table 5.2) a 95% CI can be constructed around the estimated regression coefficient of 0.87 (95% CI: 0.65 to 1.09).

The authors of some published prognostic models provide information on the standard errors of the regression coefficients, which can be used to estimate CI around the regression coefficients [See: Dickson *et al,* 1989; Christensen *et al,* 1993, Poynard *et al,* 1999]. Typically, authors applying models for estimating control group survival have ignored this additional information.

5.5 MEASURING MODEL PARAMETER UNCERTANTY – USING INFORMATION FROM STANDARD ERRORS

This section presents a simple Monte Carlo simulation method to adjust for model parameter uncertainty. This technique is applied to allow for uncertainty in the PBC Mayo model's parameter estimates over a series of five steps. The computer syntax for this method can be found in Appendix A5.2.

In Step 1, 3,000 sets of regression coefficients are randomly generated from a normal distribution. For each coefficient, the mean across 3,000 simulations, is equal to the value of the coefficient estimated for the PBC Mayo cohort (see Table 5.2), and the standard deviation is equal to its standard error. The simulations are performed using the `rmvnorm` statistical function in the S-PLUS statistical computer package [S-PLUS 6, 2001]. The simulations provide an empirical distribution of regression coefficients that would have been observed had the original model been derived from 3,000 random patient samples, instead of just one.

In Step 2, each of the 3,000 sets of regression coefficients is applied to the PBC CELT data set and used to derive 3,000 sets of prognostic risk scores (Equation 5.1). Each set of scores is adjusted for the base line risk score of the average patient in the Mayo cohort (R_0 = 2.41). Step 3 uses these adjusted scores to calculate the individual patients' probabilities of surviving to time t for all CELT patients (Equation 5.2), generating 3,000 sets of individual estimated survival probabilities in the absence of liver transplantation at three monthly intervals. In other words, this step provides insight to how the predicted non-transplant survival of CELT patients would have varied had the original prognostic model been derived on different patient samples.

In Step 4, each set of probabilities is plotted against time to obtain the patient's overall expected survival. The area under each profile is the expected non-transplant survival time to five years for a patient in the CELT cohort. Finally, in Step 5 the average survival gain for each of the 3,000 simulated data sets is calculated, by subtracting the

expected non-transplant survival over five years from the average observed post transplant survival over five years, obtained from the observed post-transplant survival for the CELT cohort (4.4 years [53 months]).

Box 5.1 summarises the simulation process.

Box 5.1 Summary of the Monte Carlo simulation method for incorporating prognostic model parameter uncertainty in to survival estimates

The uncertainty in the regression model's predictive accuracy was estimated using the following algorithm:

Step 1: Simulate 3,000 sets of regression coefficients
Step 2: Calculate the risk score for each patient in each of the 3,000 simulations
Step 3: From the risk scores, calculate the predicted survival for each patient in each of the 3,000 simulations
Step 4: Calculate the expected non-transplant survival over five years for each patient in each of the 3,000 simulations
Step 5: For each of the 3,000 simulations, calculate the average survival gain from transplantation

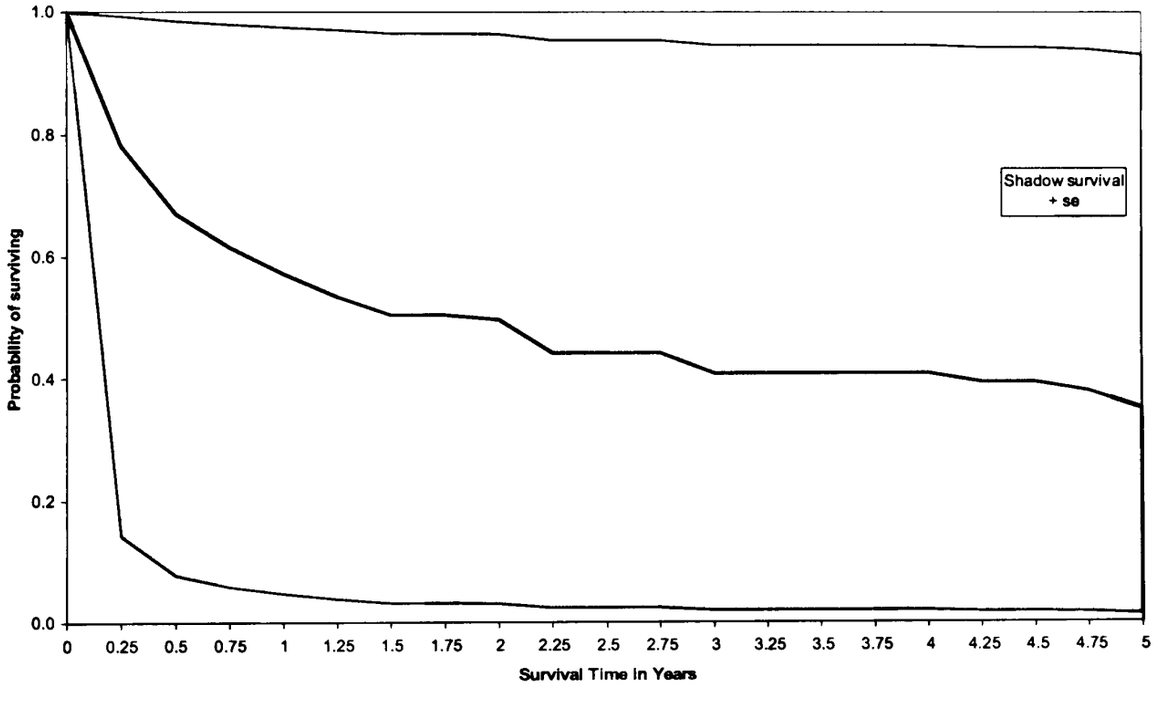
These simulations allow the incorporation of the PBC Mayo model parameter uncertainty in to the estimates of the survival gain from liver transplantation. A 95% CI for the survival gain from transplant can be calculated from the 2.5^{th} and 97.5^{th} percentiles from the simulated mean non-transplant survival and the simulated mean survival gain from liver transplantation. The point estimate (mean) is defined analogously as the average of the 1500^{th} and 1501^{st} largest values.

The expected non-transplant survival was 2.5 years (95% CI: 0.3 to 4.8 years) over five years, after allowing for model parameter uncertainty. This CI is much wider than the CI for cohort uncertainty presented in Section 5.4 (average non-transplant survival 2.5 years: 95% CI: 2.2 to 2.8 years).

Figure 5.7 presents the prognostic model uncertainty graphically, having applied the Mayo model to the CELT cohort, after allowing for the uncertainty from the regression coefficients of the prognostic model. The estimated survival gain over five years from transplantation is 1.9 years (95% CI for model uncertainty: -0.4 to 4.1 years). Note that

these CI are much wider that the intervals calculated for cohort uncertainty only (i.e. assuming perfect model predictions) and imply that the mean survival gain after liver transplantation is not statistically significant.

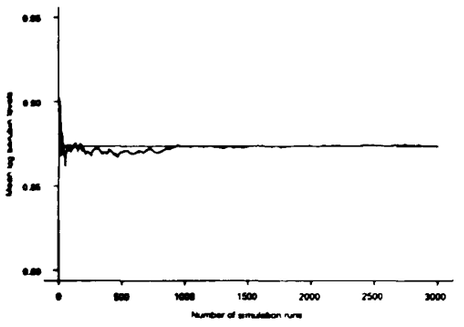
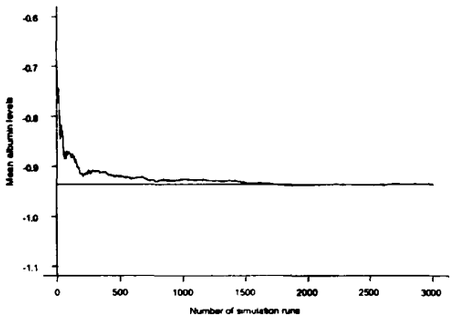
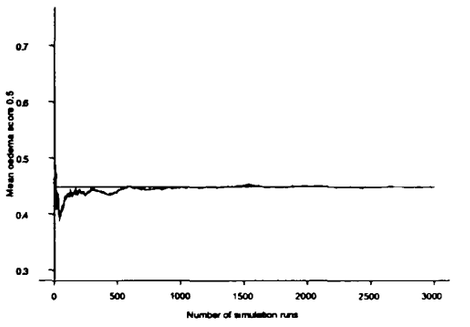
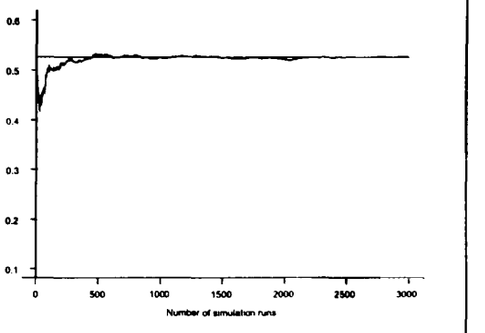
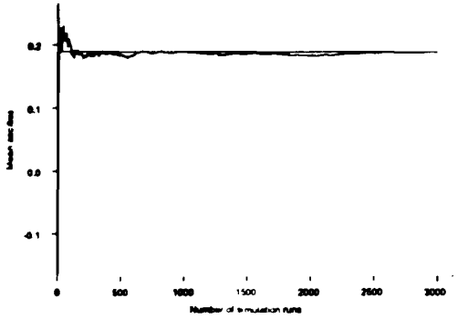
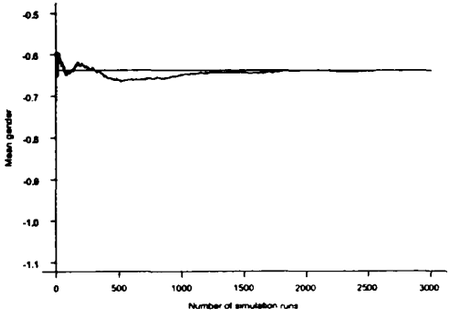
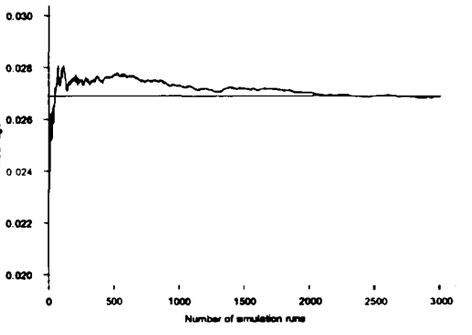
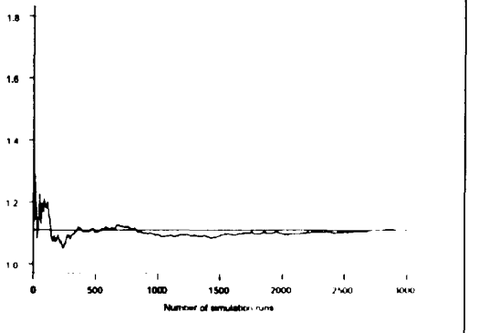
Figure 5.7 Mean non-transplant (shadow) survival (solid line) with 95% CI (blue shaded area) estimated from the PBC Mayo model using Monte Carlo simulations for 81 PBC CELT patients – using the standard errors (se) of the regression coefficients to account for model parameter uncertainty



5.5.1 Computational Considerations

When adjusting for any uncertainty, whether it is in parameter estimates or in estimating individual patient outcomes, the number of simulations run should be specific to the type of uncertainty and the convergence of the mean estimates. Figure 5.8 illustrates that mean regression coefficient estimates converge to the mean, before the end of the 3,000 simulation runs. Thus, it appears that 3,000 simulation runs is a reasonable number of runs for obtaining convergence of mean regression coefficients for prognostic model parameter uncertainty.

Figure 5.8 **Mean regression coefficients adjusting for model parameter uncertainty over 3,000 simulations runs for a) $\log_e$ bilirubin levels, b) albumin levels, c) oedema score 0.5, d) oedema score 1, e) ascities, f) gender, g) age and h) $\log_e$ prothrombin time**

a) Log_e(Bilirubin levels)	b) Albumin levels	c) Oedema score 0.5	d) Oedema score 1
			
			
e) Ascities present	f) Gender: female	g) Age	h) Log_e(Prothrombin time)

5.6 MEASURING MODEL PARAMETER UNCERTAINTY – USING INFORMATION FROM THE COVARIANCE MATRIX

In Section 5.5 it was assumed that the prognostic model parameters were fully independent of each other (for example, the coefficient for patient age is unrelated to the coefficient for bilirubin levels). This assumption is relaxed in Section 5.6 and the inclusion of extra information on the covariance matrix of the regression coefficients is included in the simulation runs for measuring prognostic model parameter uncertainty. The extra information on the covariance matrix of the regression coefficients could narrow the CI for prognostic model uncertainty, albeit only marginally. This would enable coefficients to be sampled from a multivariate normal distribution in which the estimated effect of one prognostic factor could take in to account the estimates of others, rather than it being assumed that the effects would be estimated independently of each other.

The majority of published prognostic models provide the standard errors of the regression coefficients, and to date, none that I am aware of, have provided details of the correlation (or covariance) matrix. In order to obtain information on the correlation between regression coefficients the authors of each of the three prognostic models used in the PBC CELT study; PBC Mayo, Royal Free and European models, were contacted and asked if they would be willing to provide this extra information. The Mayo Clinic provided the person specific data used to fit their PBC prognostic model, enabling the covariance matrix to be estimated.

A second set of 3,000 Monte Carlo simulations are subsequently run, accounting for both the standard errors of, and the correlations between, the estimated coefficients in the Mayo prognostic model[8]. The convergence of the mean regression coefficients was verified over 3,000 simulations for each of the variables included in the Mayo model using the same convergence checks detailed in Section 5.5.1. All variables converged prior to 3,000 simulations.

The average non-transplant survival probability to time t are plotted over time, and the average non-transplant survival and the transplant survival gain over five years is calculated, in the same way as described in Steps 1 to 5 of Section 5.5 (Box 5.1).

[8] The covariance matrix is derived by multiplying the correlation matrix by the standard errors.

Table 5.5 presents the correlations between regression coefficients for the PBC Mayo model. There is a strong correlation between age and gender, and between bilirubin and prothrombin time; and there is moderate correlation between prothrombin time and albumin, oedema and ascities; between age and bilirubin, oedema and ascities, between oedema and gender, and between oedema and ascities. There is poor to slight correlation between the remaining variables included in the model.

After accounting for both the standard errors and the correlations between the regression coefficients, the expected non-transplant survival over five years was 2.5 years (95% CI: 0.3 to 4.8 years), and the survival gain from transplantation was 1.9 years (95% CI: -0.4 to 4.2 years). Adjusting for the correlation between regression coefficients very slightly increases the model parameter uncertainty, in this case, as can be seen in Figure 5.9.

Table 5.5 Correlation between regression coefficients for the PBC Mayo prognostic model presented in Table 5.2

	Log_e(bilirubin)	Albumin	Oedema: 0.5	Oedema: 1	Ascities	Gender	Age	Log_e(prothrombin)
Log_e(bilirubin)	1.00	-0.19	-0.03	0.05	0.09	0.19	0.39	-0.79
Albumin		1.00	-0.07	-0.09	0.10	-0.19	0.01	0.33
Oedema: 0.5			1.00	0.85	-0.38	-0.45	-0.36	-0.24
Oedema: 1				1.00	-0.51	-0.52	-0.31	-0.30
Ascities					1.00	-0.00	-0.38	-0.30
Gender						1.00	0.82	-0.04
Age							1.00	-0.01
Log_e(prothrombin)								1.00

Figure 5.9 **Mean non-transplant (shadow) survival (solid line) with 95% CI (blue and orange shaded areas) estimated from the PBC Mayo model using Monte Carlo simulations for 81 PBC CELT patients – using the standard errors (se) and the correlations (corr) between the regression coefficients to account for model parameter uncertainty**

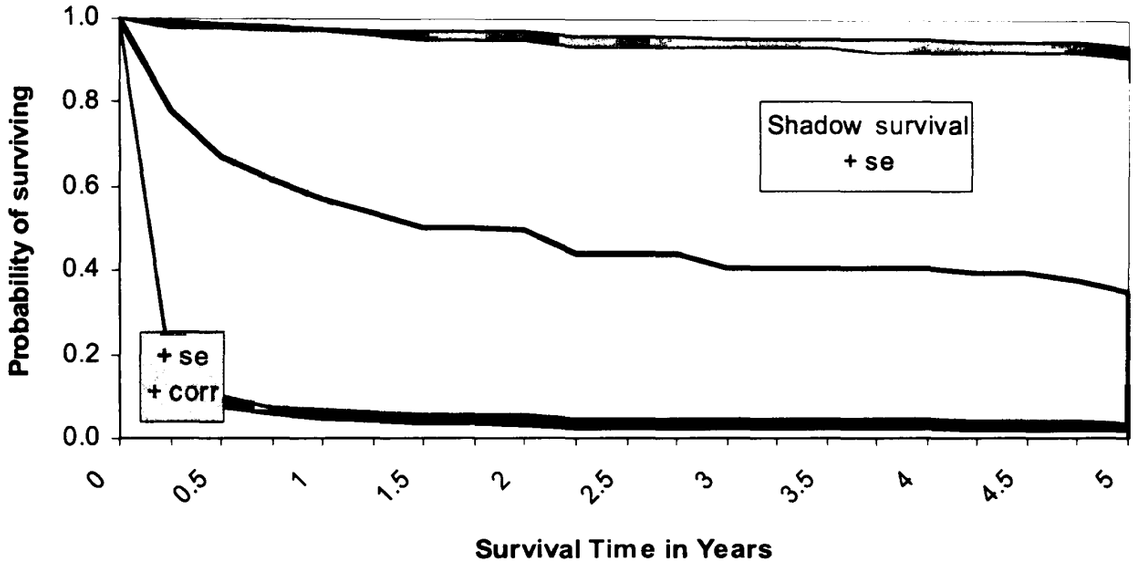


Table 5.6 summarises the results from using Monte Carlo simulation methods to predict model uncertainty using information from the standard errors of the regression coefficients, and the standard errors and correlation between regression coefficients, from a prognostic model fitted to a cohort of patients with PBC treated at the Mayo Clinic.

Table 5.6 Summary of results: Average survival estimates (with 95% CI) over five years in PBC CELT patients and estimated survival gain from liver transplantation

	Average survival (years)	Estimated survival gain (years)
Observed post-transplant survival (CELT patients)	4.4 (4.0 to 4.7)	-
Observed non-transplant survival (Mayo patients)	2.8 (2.5 to 3.0)	1.6 (1.2 to 2.0)
Observed adjusted non-transplant survival (Mayo patients)	2.6 (2.4 to 2.9)	1.7 (1.7 to 1.8)
Predicted non-transplant survival in CELT patients from Mayo data prognostic model		
Coefficients alone	2.5 (2.2 to 2.8)	1.9
Coefficients and standard errors	2.5 (0.3 to 4.8)	1.9 (-0.4 to 4.1)
Coefficients, standard errors and correlations	2.5 (0.3 to 4.8)	1.9 (-0.4 to 4.2)

5.7 COMPARING MODEL PARAMETER UNCERTAINTY ACROSS THREE PBC PROGNOSTIC MODELS

In their analysis of the CELT data Longworth *et al* used three published prognostic models to estimate non-transplant survival for patients with end-stage PBC: the PBC Mayo model (detailed above), the Royal Free model, and the European model [Longworth *et al*, 2003]. The Royal Free model predicts survival using information on patient age, serum bilirubin, serum albumin and the presence or absence of ascities [Hughes *et al*, 1992]. This model applies Cox PH assumptions to a cohort of 289 patients referred to the Royal Free Hospital, London UK, with PBC between 1977 and 1989.

The European prognostic model is based upon a cohort of patients with PBC who took part in a multi-centre RCT between 1971 and 1983 (56.9% of patients were from UK centres), in which 248 patients were randomised to receive either azathioprine or placebo [Christensen *et al*, 1985; Christensen *et al*, 1993]. Clinical data were collected every six months and were included in a time dependent covariate Cox PH model. Serum bilirubin, serum albumin, age, the presence of ascities and the presence of gastrointestinal bleeding were found to be significant predictors of survival. All three

models result in similar predictions of non-transplant survival estimates when applied to the PBC CELT cohort (Figure 5.10).

Figure 5.10 Estimated non-transplant survival for 81 PBC CELT patients from the Mayo model (solid line), Royal Free model (dotted line) and European model (dashed line) over five years

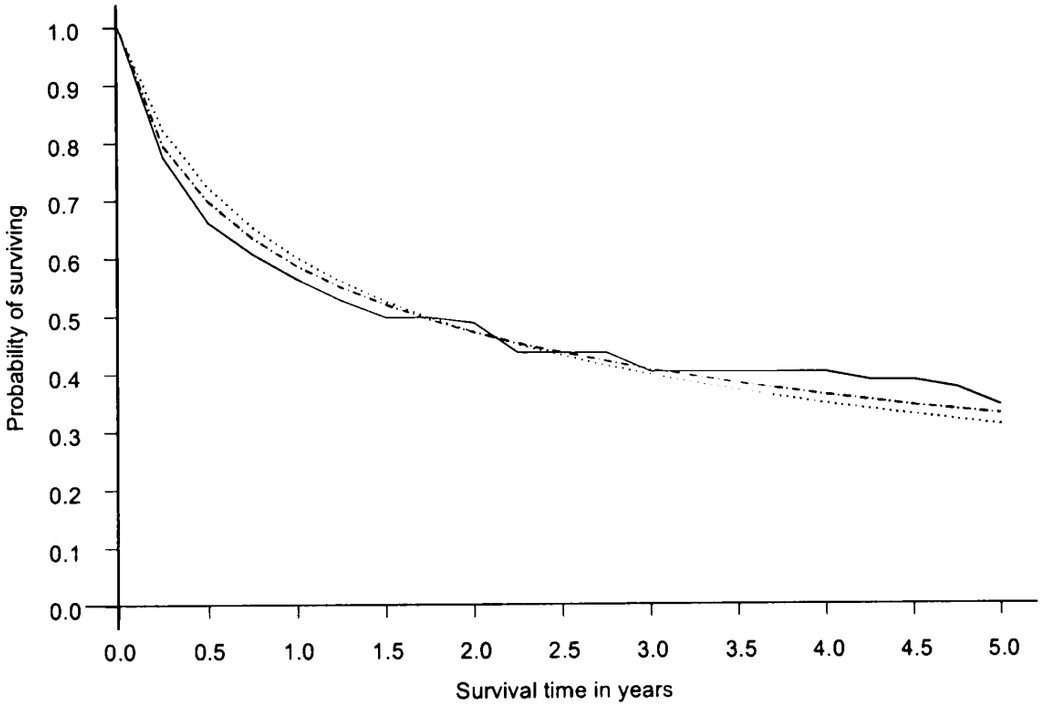


Table 5.7 summarises the prognostic factors included in each on the three PBC prognostic models[9]. All three models use information on patient age, bilirubin levels, albumin levels and the presence or absence of ascities, to predict survival. In addition to these, the Mayo model uses additional information on gender, prothrombin time and oedema score, whereas the European model uses information on the presence or absence of gastrointestinal bleeding.

In each of the three models a different modelling approach has been used. The Mayo model uses data from one time point (the last study visit) to predict survival in the absence of transplantation, the objective of this model was to estimate survival in the absence of transplantation in patients with end-stage PBC. The Royal Free and

[9] The PBC Mayo model variables are the ones included in the model presented in this thesis (Table 5.2 and 5.3) and not the ones included in the published models.

European models differed from the Mayo model in their objectives: they aimed to predict short-term prognosis for PBC patients at any stage of their disease via updated covariates, both models utilising information from several visits in a person-interval model and a time dependent Cox model, respectively.

Table 5.7 Variables used in the PBC Mayo, Royal Free and European prognostic models

	PBC Mayo	Royal Free	European
Age	✓	✓	✓
Bilirubin	✓	✓	✓
Albumin	✓	✓	✓
Ascities	✓	✓	✓
Gender	✓		
Prothrombin time	✓		
Oedema	✓		
Gastrointestinal bleeding			✓

The correlation between regression coefficients was not available from the authors of either the Royal Free or European prognostic models; therefore it was not possible to compare model uncertainty across models after accounting for correlations. However, the standard errors for the regression coefficients were available for all three prognostic models. The methods described in Section 5.5 above are applied to the Royal Free and European models in order to estimate model parameter uncertainty and compare the degree of uncertainty with that from the PBC Mayo model. 3,000 simulations are run for each model, regression coefficients converged to the mean prior to 3,000 simulations. The results are shown in Figure 5.11 below.

Figure 5.11 Mean non-transplant (shadow) survival (solid line) with 95% CI (blue shaded areas) estimated from a) the Royal Free model and b) the European model using Monte Carlo simulations for 81 PBC CELT patients – using the standard errors (se) of the regression coefficients to account for model parameter uncertainty

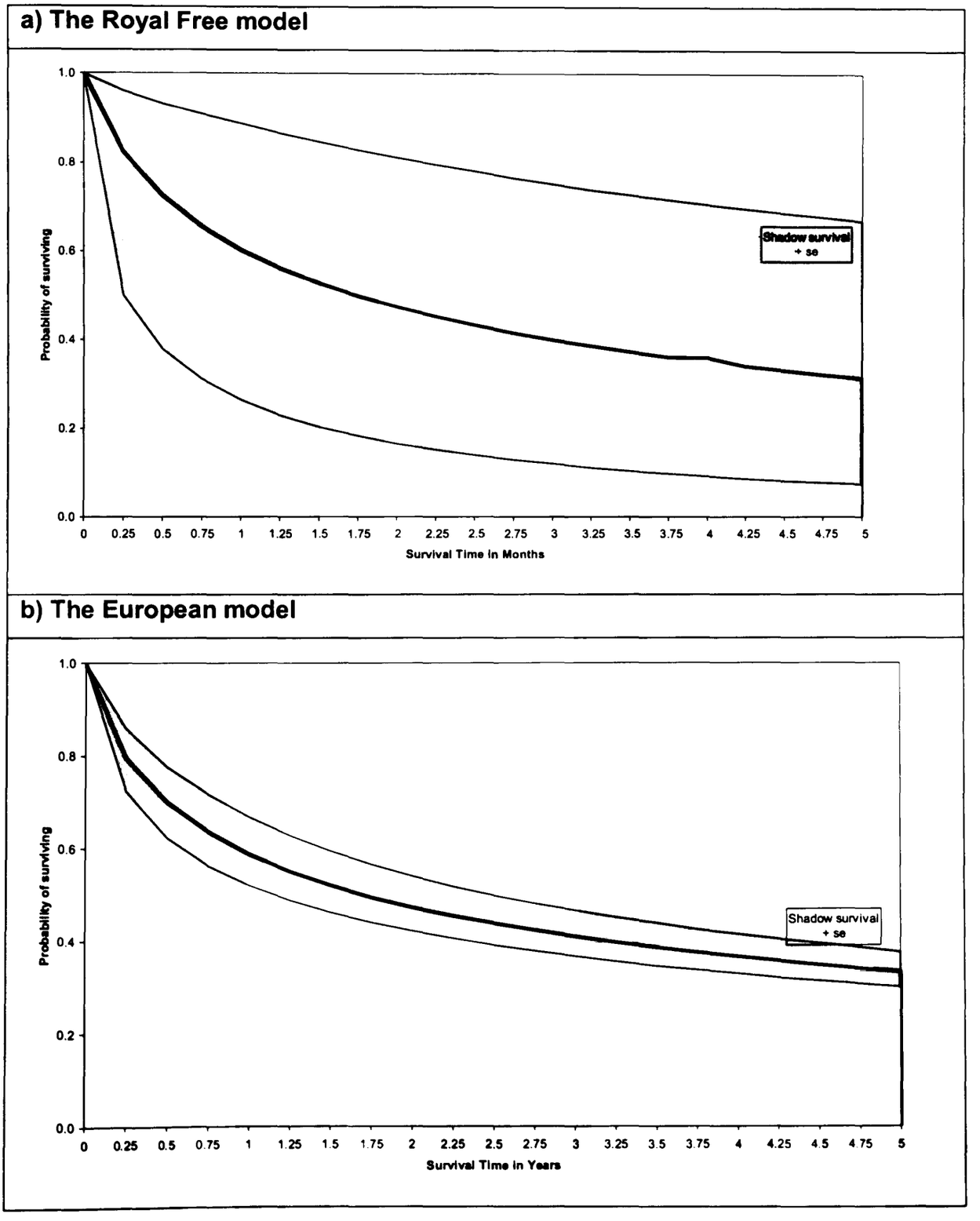


Table 5.8 presents a summary of the estimates of non-transplant survival, the expected survival gain over five years and the 95% CI represent the respective model parameter uncertainty for PBC CELT patients using the PBC Mayo, Royal Free and European prognostic models. It can be seen by comparing Figures 5.7, 5.11a and 5.11b and from Table 5.8 that the model uncertainty is much larger for the PBC Mayo model than for the Royal Free and European models. When the latter two models are used, there is evidence of a survival gain after transplantation that is not apparent from the Mayo model. Despite this, the average survival gain after transplantation is remarkably similar for all three models, with gains ranging slightly between 23 and 24 months. The choice of model from which non-transplant survival is estimated does not therefore give rise to inconsistent estimates of the effect of transplantation, but does produce large differences in the amount of uncertainty surrounding these.

Table 5.8 Summary of model parameter uncertainty for non-transplant survival over five years and the survival gain after liver transplantation for 81 PBC CELT patients from three prognostic models (PBC Mayo, Royal Free and European)

	Non-transplant survival (95% CI)	Survival gain (95% CI)
PBC Mayo Model	2.47 (0.27 to 4.81)	1.93 (-0.38 to 4.13)
Royal Free Model	2.43 (0.98 to 3.99)	1.97 (0.41 to 3.42)
European Model	2.45 (2.21 to 2.74)	1.95 (1.66 to 2.19)

The variation in the degree of model uncertainty for the three prognostic models is surprising and could have arisen for many reasons including: the number of patients on which each prognostic model was derived, the amount of uncertainty explained by each model, the number and nature of the prognostic factors selected for each model, or the type of model fitted. Each of these possible reasons is explored in more detail below.

5.7.1 Number of Patients on which each Prognostic Model was Derived

When prognostic models are derived on larger patient cohorts, less uncertainty is likely to arise in the model estimates, in comparison with models derived from smaller cohorts. More precisely, if a large number of patients die during follow-up, the coefficients of the prognostic model are more accurately estimated (that is, their standard errors are smaller) than if the model had been evaluated on a smaller cohort, or in a cohort where few deaths had occurred.

A published prognostic model derived from a small sample would result in outcomes being estimated with less precision and this would be reflected by the standard errors of the model regression coefficients. A published prognostic model applied to an alternative cohort would use the simulation techniques described in this chapter to account for prognostic model uncertainty.

The sample sizes for the three PBC prognostic models and the proportion of deaths per cohort were similar (European model: 248 patients – 49% died, Royal Free model: 289 patients – 40% died, Mayo model: 312 patients – 44% died). Given that the confidence limits for model uncertainty were largest for the PBC Mayo model and smallest for the European model, the results do not conform to this hypothesis, meaning that sample size is not the reason for the variation in model uncertainty.

It is possible that other characteristics of the patient cohort affected the accuracy. For instance, if the patient population is a relatively homogeneous sub-sample of PBC patients, the estimated coefficients may be more precisely estimated than had they been estimated on the full (i.e. more heterogeneous) PBC population. Whether such estimates can be validly generalised to the general population of patients with PBC is unclear, but this consideration is independent of the calculated standard errors.

5.7.2 The Amount of Uncertainty Explained by each Prognostic Model

The standard errors of the coefficients included in the Mayo model – where uncertainty is greatest, may be comparatively larger than the standard errors for the other two models. To explore the variability in uncertainty levels between models, the ratio of the regression coefficient to the standard error (Regression coefficient/Standard error) was calculated for each variable in the PBC Mayo, Royal Free and European models. The ratios, which are equivalent to the z-statistic, are presented in Table 5.9, the larger the

ratio, the smaller the amount of uncertainty for a specific variable, and the greater the association between the variable and survival.

Table 5.9 Ratio of regression coefficients to standard errors (regression coefficient/standard error) for prognostic model variables: PBC Mayo, Royal Free and European models

	PBC Mayo	Royal Free	European
Age	2.96	5.00	3.64
Bilirubin	7.65	7.92	9.73
Albumin	4.31	5.26	4.47
Ascities	0.86	3.61	6.62
Gender	2.59		
Prothrombin	1.67		
Oedema: 0.5	1.93		
Oedema: 1	2.06		
Gastrointestinal bleeding			3.10

Generally, the ratios of regression coefficient to standard errors were much higher for the variables included in the European Model and Royal Free models than for the PBC Mayo model; ratios were less than 3.5 for six of the variables in the Mayo model, whereas no ratios were less than 3.5 for the other two models. Additionally, the ratios were much lower for the PBC Mayo models than that for the Royal Free and European models, for each of the four variables that were common to all four models (bilirubin, albumin, age and ascities). Bilirubin levels were the most significant predictors of survival across all three models and had a higher level of significance for the European model in comparison with the PBC Mayo model and Royal Free model.

The results presented in Table 5.9 support the evidence presented in Figures 5.7 and 5.11 that there is more parameter uncertainty in the PBC Mayo model than the European model and Royal Free model. However, based upon the results presented in the table, it is not possible to conclude that survival estimates from the European model will show less uncertainty than estimates from the Royal Free model, where uncertainty is least for bilirubin and ascities in the European model and for age and albumin in the Royal Free model.

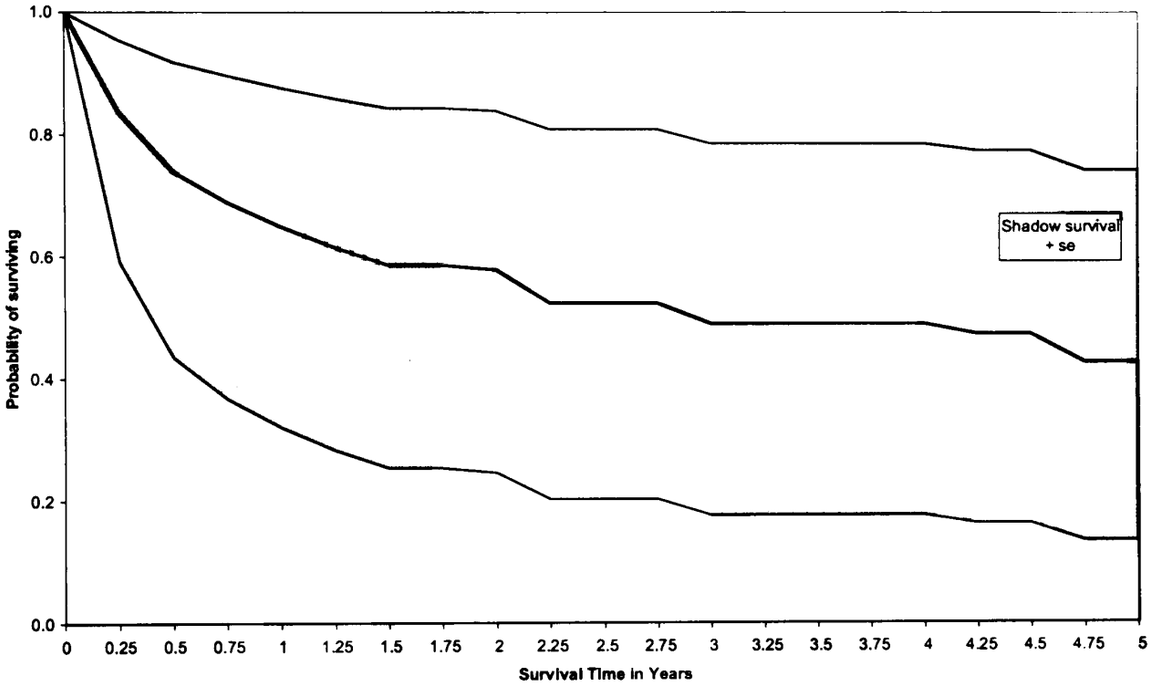
5.7.3 The Number and Nature of the Prognostic Factors

A total of seven variables were fitted in the PBC Mayo model and variables were included regardless of level of statistical significance. In contrast, the European and Royal Free models only included variables that were statistically significant. This is an important consideration, since the standard error of the predicted risk score is directly related to the square root of the sum of the squared standard errors for each coefficient[10]. Adding more factors may increase or decrease this model uncertainty, dependent on how predictive the prognostic factor is. However, non-significant predictive factors would be expected to have larger standard errors (and hence greater variation) than significant factors, and including only significant terms can lead to narrow (and quite probably, artificially narrow) CI.

The PBC Mayo model was refitted in order to compare uncertainty across models when only significant variables were included, using the 5% significance level as a cut off for significant variables. Figure 5.12 shows the non-transplant survival curve for the refitted Mayo model if non-significant variables (ascities and prothrombin time) are excluded from the model. The 95% CI is smaller than that when prothrombin time and ascities are included in the model (95% CI: 0.16 to 2.78 years excluding prothrombin time and ascities compared with -0.36 to 4.13 years when these variables were included), and now suggests that transplantation is a beneficial treatment of choice for patients with end-stage PBC. The mean transplant survival gain over five years is slightly shorter (1.47 years) than when prothrombin time and ascities are included in the model (1.93 years). Further model details are presented in Appendix A5.1.

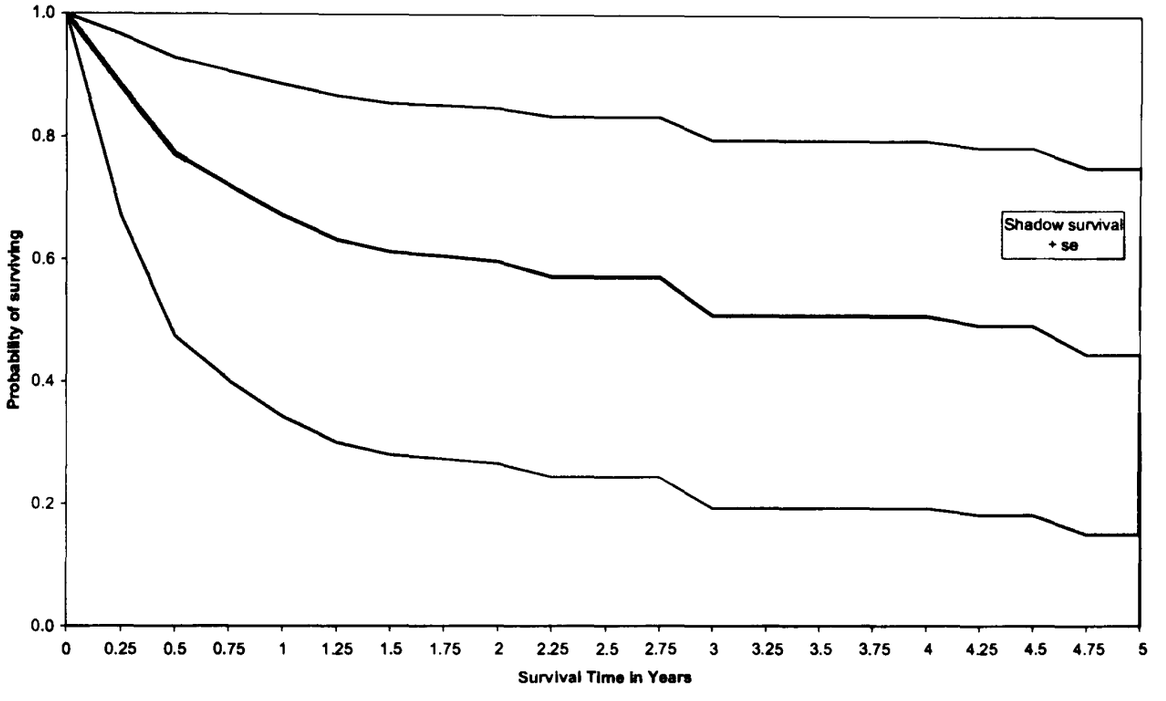
[10] The relationship is: Standard Error(predicted risk score) = √ sum of squared model regression coefficient standard error + 2*sum of all model regression coefficient covariances. Since the latter term is not available or not included here the Standard Error (predicted risk score) =√ sum of squared model regression coefficient standard error.

Figure 5.12 **Mean non-transplant (shadow) survival (solid line) with 95% CI (blue shaded areas) estimated from the PBC Mayo model (excluding ascities and prothrombin time) using Monte Carlo simulations for 81 PBC CELT patients – using the standard errors (se) of the regression coefficients to account for model parameter uncertainty**



Given that the number of variables was higher in the Mayo model than both the European and Royal Free models, a third version of the PBC Mayo model was fitted where the variables included in the model were limited to the four variables common to all three PBC prognostic models: bilirubin, age, albumin and ascities. Variables were included in the model regardless of level of significance, though only ascities was non-significant at the 5% level ($p = 0.064$, Appendix A5.1). Figure 5.13 shows that, in comparison with the original Mayo model, model uncertainty has decreased, the survival gain after transplantation is now 1.39 years (95% CI: 0.18 to 2.95 years) – the gain is slightly less than when including prothrombin time, gender and oedema (1.93 years [95% CI: -0.36 to 4.13 years]).

Figure 5.13 **Mean non-transplant (shadow) survival (solid line) with 95% CI (blue shaded areas) estimated from the PBC Mayo model (including bilirubin, albumin, age and ascities) using Monte Carlo simulations for 81 PBC CELT patients – using the standard errors (se) of the regression coefficients to account for model parameter uncertainty**



The results from the two alternative PBC Mayo prognostic models demonstrate how the number of variables included in the model, and the way that variables are selected, can influence the level of model parameter uncertainty. Researchers should be aware that if variables are included in a model, regardless of level of significance, it is likely to increase the level of model uncertainty. However, just because a variable is not statistically significant does not mean that it is not clinically meaningful, and several studies have shown that there is a relationship between the variables included in the PBC Mayo model (bilirubin, albumin, prothrombin time, age, gender, oedema scores and ascities) and liver disease severity/survival [See: Hughes *et al,* 1992; Kamath *et al,* 2001]. It is therefore defendable to include non-significant variables in our PBC Mayo prognostic model, as although they are insignificant here, they have been shown to be statistically significant in other studies.

5.7.4 The Type of Model Fitted

Both the Royal Free and European models make use of repeated measures of patient and clinical characteristics over time and these models allow for these changes over

the study period. Hughes *et al,* fit a person interval model in which a series of intervals are defined [Hughes *et al,* 1992]. The first interval begins at the point where a patient visits the Royal Free hospital for treatment for end stage liver disease and ends two years after that time point or at time of death if sooner. The second interval begins at the hospital visit that follows the end of the first interval and ends two years after that time point or at point of death if sooner. This process continues for the duration of the study, and each patient's follow up is defined by the number of intervals for which there are available data. By contrast, the authors of the European model use a time dependent Cox PH model in which the first interval starts at first study visit and ends immediately prior to the second study visit, at which time the next interval begins and so on. Unlike the Royal Free model, no follow-up information is lost as interval lengths are not fixed. The authors of the Royal Free model state that 81% of total follow-up information and 93% of deaths are captured in their model, with the remainder occurring outside of the defined two-year intervals. This is likely to be a contributing factor to the increased model uncertainty in the Royal Free model in comparison with the European model.

In contrast, the PBC Mayo prognostic model includes data only on the final study visit. This type of model was chosen because the ultimate aim of the CELT study was to compare transplant survival with non-transplant survival and clinical information was only collected at one pre-transplant time point in the CELT study.

5.7.5 Summary of Possible Contributors to Uncertainty Levels

In summary, sample size and the number of deaths per study do not appear to be contributing factors that influence the levels of model parameter uncertainty across the three PBC prognostic models. However, the level of uncertainty of individual model variables and the number of variables included in the model contribute to the increased uncertainty levels for the PBC Mayo model over the Royal Free and European models. Additionally, the person specific interval model fitted to the Royal Free cohort appears to be a contributing factor to the additional uncertainty in this model over the European model.

5.8 PROGNOSTIC MODEL CRITERIA – CHOOSING A PROGNOSTIC MODEL

In the CELT study a series of prognostic models existed that could be used to estimate non-transplant survival: ALD (2 models), PBC (3 models), PSC (2 models). In Chapter

2, six selection criteria were outlined (Box 5.2) and these criteria were applied in the selection of prognostic models in the main CELT analysis.

Box 5.2 Selection Criteria applied in the main CELT study when selecting a disease specific prognostic model to estimate non-transplant survival

These criteria state that models should:

- be validated on both internal and external data sets **(Criterion 1)**
- be based on adequate samples of patients **(Criterion 2)**
- for each variable fitted in the model the number of events, in this case deaths, should be in the ratio of one explanatory variable for every 10 or more deaths **(Criterion 3)**
- the variables included in the model should make sense clinically **(Criterion 4)**
- the model should be applied to appropriate cohorts **(Criterion 5)**
- the cohort should include UK patients **(Criterion 6)**

Criteria 1 to 5 were originally suggested by Altman and Royston [Altman & Royston, 2000] and formed the basis of the model selection process. The most appropriate prognostic models were then selected in the main CELT analysis across a series of steps. In the first step, any model failing three of the first five criteria was defined as being statistically invalid, and was then discarded. The second step, for those models that appeared to be statistically valid, was to incorporate models where the historical cohort included UK patients **(Criterion 6)**. Longworth and colleagues applied Criterion 1 to 6 in their CELT analysis. This resulted in the selection of the Beclere model for ALD patients, an average of the European and the Royal Free models for PBC patients and the International model for PSC patients [Longworth *et al,* 2003].

Based on the work on prognostic model parameter uncertainty, detailed in this chapter, it does not seem unreasonable to propose a seventh selection criterion, where the lower the level of uncertainty the better the model at predicting survival. Thus, in addition to Criteria 1 to 6, it could be proposed to add Criterion 7, which states that a model should:

- display the least amount of parameter uncertainty (in comparison with other appropriate models) **(Criterion 7)**

Table 5.10 sets out the seven selection criteria for the three PBC prognostic models. The first six Criteria are either met ("Yes") or are not met ("No"), whereas Criterion 7 ranks the models by their amount of parameter uncertainty ("1" being the least amount of uncertainty, "3" the most uncertainty). Clearly, the ideal model would meet the first six Criteria and contain the least amount of uncertainty. None of the PBC prognostic models displayed clear statistical superiority over the other models (Criteria 1 to 5), although the Royal Free model was not validated on an external data set. The PBC Mayo model was not based on UK patients and is thus eliminated from the model selection process.

Table 5.10 Seven prognostic model selection criteria applied to the European, Royal Free and PBC Mayo prognostic models

	European Model	Royal Free Model	PBC Mayo Model
Criterion 1	Yes	No	Yes
Criterion 2	Yes	Yes	Yes
Criterion 3	Yes	Yes	Yes
Criterion 4	Yes	Yes	Yes
Criterion 5	Yes	Yes	Yes
Criterion 6	Yes	Yes	No
Criterion 7*	1	2	3

* See Figure 5.7 and 5.11

For the remaining two prognostic models it can be seen clearly from Figure 5.11 that the European model displays the least amount of model parameter uncertainty (Criterion 7) and would therefore be selected as the most appropriate model for predicting non-transplant survival. However, just because a model, such as the PBC Mayo prognostic model displays a large amount of uncertainty, does not necessarily mean that it is a bad model. This chapter has already discussed possible causes of increased parameter uncertainty. In the case of the inclusion of non-significant variables in prognostic models (increasing uncertainty), this chapter argued for their inclusion in prognostic models – a statistically insignificant variable may be clinically meaningful. Although it may be intuitive to select a model with the least amount of uncertainty, explanations may exist that explain that additional uncertainty. A

prognostic model should not be excluded from the selection process if the additional parameter uncertainty could be attributed to the inclusion of clinically meaningful variables that are not necessarily statistically significant. A model displaying a large amount of parameter uncertainty will not necessarily give inaccurate survival estimates; all three PBC models give similar estimates of non-transplant survival (Table 5.8).

5.9 DISCUSSION

5.9.1 Historical Cohorts or Prognostic Models – Which are Preferable when Person Specific Historical Data are Available?

In the main CELT analysis no patient specific non-transplant data were available, for comparison with data from a transplant cohort; therefore, disease specific published prognostic models were applied to estimate survival in the absence of transplantation. The patient specific historical data from the Mayo cohort became available at a later date.

But what if historical data had been available to use in the CELT study? If this was the case, the authors of the main CELT analysis would probably have used the approach described in Section 5.3 to estimate non-transplant survival and the survival gain from transplantation. Cox PH regression models would have been used to adjust for differences between cohorts. This is a more conventional approach to use. Yet, why not fit a prognostic model to the historical data, and use the methods described in Sections 5.5 and 5.6 to adjust for model uncertainties? Again I suspect convention would be the reason that the technique of adjusting for model parameter uncertainty might not be used. However the Kaplan-Meier survival results presented in Section 5.3.4 assume that the estimated regression coefficients are accurate and thus fail to account for any uncertainty around these estimates (The 95% CI in Figure 5.4 depict cohort uncertainty alone). Information is available on the regression coefficients and their standard errors for the Cox PH model applied in Section 5.3, and model uncertainty still exists, and so it should be allowed for, this would increase the uncertainty around mean estimates[11].

[11] Allowing for parameter uncertainty in the Cox PH model applied to the Mayo cohort (Section 5.3) increase CI from 2.4 to 2.9 years to 0.6 to 4.8 years, in the same way as allowing for parameter uncertainty increased CI when the PBC Mayo model is applied to the CELT cohort (CI increased from 2.2 to 2.8 years to 0.3 to 4.8 years (Table 5.6).

5.9.2 Problems with Historical Data

Uncertainties in model predictions arise in part due to the variability of patient characteristics between the prognostic model cohort and the cohort the prognostic model is fitted to. This chapter has described an approach for adjusting for this uncertainty. However, there are other considerations that need to be acknowledged when applying a historical model to a concurrent data set that will affect the accuracy of the prognostic model predictions. Firstly, there will be changes in clinical practice over time, which may improve patient prognosis in the absence of transplantation. Secondly, there may be advances in medicine over time, which may again improve patient prognosis or prolong the occurrence of end-stage liver disease. Thirdly, there is uncertainty in the model estimates due to differences in practice between countries; patients with PBC in the UK may be treated differently to those in the US. The first two aspects would result in an underestimation of survival in the absence of transplantation, whereas the third could result in either an under or overestimation of survival in the absence of transplantation.

It is impossible to measure how these sources of uncertainty could effect model predictions. However, it is important to acknowledge their existence as they add to the degree of uncertainty in the model predictions.

5.9.3 Model Parameter Uncertainty

In the liver transplant example used here, there was no evidence of a statistically significant survival gain from transplantation after adjusting for the standard errors and correlation coefficients of the regression coefficients for the PBC Mayo model. Had the standard errors, in this example, been smaller then the uncertainty surrounding the estimates of survival in the absence of transplantation would have been smaller, and it is possible that there would have been some evidence of a survival benefit.

In Section 5.6, Monte Carlo simulations incorporated parameter uncertainty using information on the standard errors and the correlation between regression coefficients. Overall, there was little correlation between variables, although some variables were highly correlated, for example age and sex (Table 5.5). Had the correlation between variables been larger, the change in the amount of uncertainty after allowing for the correlations would have been more noticeable. In contrast, if the regression variables had been independent of each other, i.e. the correlation between coefficients was zero;

the uncertainty surrounding the survival estimates would be the same as when the standard errors alone are adjusted for.

5.9.4 Variation in the Amount of Model Parameter Uncertainty

Often, adding non-significant factors in to the model will increase the model uncertainty, since in general, a greater uncertainty occurs around non-significant variables than significant ones. The inclusion of clinically meaningful variables could be more justifiable if the variables included had been shown to be important in other studies within that country. Thus, for the US Mayo model the inclusion of clinically significant variables within the US setting could be justified. However, if several studies from several other countries had found a particular covariate to be clinically meaningful but there was no evidence from the US, in the case of the Mayo model, there could be argument for inclusion of the clinically meaningful variable. However, the inclusion of clinically meaningful variables in prognostic models should be considered carefully and involve discussions with clinical experts before making decision on whether to include or exclude a variable.

There are no clear recommendations as to what the best approach is for the inclusion of variables in mathematical models, although it is generally recognised that statistical significance should not be the sole basis of this [Collett, 1994; Altman & Royston, 2000; Bradburn *et al,* 2003]. Therefore, variables were included in the PBC Mayo model that shared a commonality between both the Mayo and CELT cohorts and were included in the model regardless of the significance level.

5.9.5 Further Model Parameter Uncertainty

A further area of uncertainty that is specific to the Cox PH model has not been accounted for here, namely the uncertainty in the baseline survival estimator. For any time dependent Cox PH regression model the baseline survival is assumed fixed and therefore constant over the period of the model [Collett, 1994]. Therefore authors of published prognostic models provide no information on the uncertainty in baseline survival estimates. In fact baseline survival estimates are correlated with the regression coefficients and when modelling uncertainty in these estimates it is assumed that a change in the coefficients would alter the baseline survival estimates. Further exploration of model parameter uncertainty that accounts for uncertainties in the baseline survival coefficients could be the basis of future work.

5.9.6 Generalisability

In this chapter a Cox PH model was used to illustrate methods to adjust for model parameter uncertainty using Monte Carlo techniques. These techniques can easily be adapted to other model forms, from the basic linear regression model to the Weibull model, using the same process described above and detailed in Appendix A5.2.

The simulation techniques presented in this chapter to adjust for parameter uncertainty are relatively simple and easy to apply. In fact, the command to generate the simulations in the statistical computer package S-PLUS meant that no additional computer programming was required for this analysis [S-PLUS 6, 2001]. As demonstrated in this chapter, a simple pre-existing function for the S-PLUS computer package (`rmvnorm`) was used to generate 3,000 sets of regression model coefficients that were then applied to the CELT cohort in order to estimate non-transplant survival. Using this one function, it was possible to incorporate uncertainty from the standard errors and correlation between regression coefficients. Alternative computer packages could also have been used for this analysis, for example EXCEL/Crystal Ball software, or the Bayesian software package BUGS [Decisioneering, Inc, 1994; Spiegelhalter *et al,* 1999].

What is required by authors of future prognostic models is an additional table providing the reader with the correlation coefficients of the regression coefficients. Providing the reader with this additional information would give potential users the opportunity to obtain more realistic estimates of uncertainty and journal editors ought to ensure that this is done. The limits imposed by journals on overall reporting space, tables and figures will inevitably prove an obstacle to providing potential users with this information, but since many journals now have internet sites, the additional information could be obtained via these sites.

It is accepted that the likely reason for variation in the levels of uncertainty across three prognostic models used in this chapter, i.e. number of covariates and type of model fitted, could be specific to this example. Potential variation in other prognostic models might be explained by variations in sample size or the proportion of events per model. Therefore, reasons for the variation in uncertainty between models should be explored on a case by case basis.

It is important that the prognostic model is representative of the cohort it is being applied to; if it is not representative then it will not estimate the desired outcome. An extreme example of this would be fitting a published prognostic model for estimating survival derived from a cohort of heart transplant patients to predict survival in a cohort of liver transplant patients. Therefore, it is much more important to apply a realistic prognostic model with a large amount of uncertainty than a less realistic model with a small amount of uncertainty.

5.10 CONCLUSIONS

In this chapter a Monte Carlo simulation technique has been presented and this technique may be used to incorporate model parameter uncertainty in to outcome estimates, for example survival estimates. It is recommended that future studies of health care technologies should incorporate model parameter uncertainty in to estimates when using prognostic modelling techniques to estimate outcomes.

Based upon the results from one model, it is unclear whether additional information on the correlations between regression coefficients will have any significant impact on model parameter uncertainty and it is recommended that future work is carried out to establish this. Guidelines set by the National Institute for Health and Clinical Excellence (NICE) state that decision makers should "know about the uncertainty associated with clinical and cost-effectiveness information" [NICE, 2004]. Therefore, it is recommended that the additional uncertainty from the correlation between regression coefficients should be incorporated in to model parameter uncertainty, in order to provided further knowledge about uncertainty.

A large amount of model parameter uncertainty was demonstrated around the PBC Mayo prognostic model and the survival gain from transplantation over a five-year period was shown to be statistically non-significant. However, a survival gain from transplantation was evident if model uncertainty was excluded from the estimates of the survival gain post transplantation, or when alternative PBC prognostic models were applied to the CELT data.

Estimates of non-transplant survival and the survival gain up to five years after transplantation were compared across three PBC prognostic models, after allowing for model parameter uncertainty. All three models gave a similar average non-transplant survival estimate; however, the model uncertainty around these estimates differed

across models, was greatest for the PBC Mayo model and smallest for the European model.

Chapter 5 also considered whether levels of uncertainty around prognostic models could be used to influence the choice of prognostic model, when more than one model exists. If all other model choice considerations are equal, models with a smaller amount of uncertainty are preferable to models with a large amount of uncertainty. However, it is recommended that when using levels of uncertainty to choose between prognostic models consideration should be given as to why one model may display more uncertainty to another model. If obvious explanations exist, for example a model including non-statistically significant variables, though clinically meaningful ones, is likely to display more uncertainty than one that includes only statistically significant variables; this model should not necessarily be excluded from the model selection process. It is important that decision makers are presented with information on the survival gain that incorporates the uncertainty around prognostic model parameter estimates in order to make an informed decision about new or existing treatments or technologies, for example liver transplantation.

This chapter has begun to apply Monte Carlo simulation methods to estimate some of the uncertainty around prognostic model estimates. Chapter 6 moves on to describe methods for estimating a further area of prognostic model uncertainty when estimating individual patient outcomes from a prognostic model.

CHAPTER 6

ESTIMATING INDIVIDUAL PATIENT OUTCOMES FROM PROGNOSTIC MODELS AND THE UNCERTAINTY SURROUNDING THE OUTCOMES

6.1 INTRODUCTION

In the absence of observed data on non-transplant survival, it is proposed that the prognostic models applied in the main CELT study [Longworth *et al,* 2003] are used to estimate non-transplant survival over the extended CELT study period of five years. All the published prognostic models applied to the CELT cohort are Cox PH models and are specific to one of three liver disease groups: ALD, PBC or PSC. The prognostic models are used to obtain patient specific estimates of non-transplant survival, QALYs and costs and subsequently patient level estimates of the survival, QALY and cost gain from transplantation. Non-transplant QALY and cost estimates clearly depend on a patient's expected survival time.

A particular feature of the CELT non-transplant cost estimates is that each patient's study cost will depend not only on the time for which the patient is predicted to survive, but whether or not the patient was predicted to survive for the complete five-year study period. Empirical evidence suggests that patients who die require additional resources

(and thereby incur additional cost) toward the end of their lives. Consequently, an adjustment was made to the non-transplant costs for patients who were predicted to die during the study period. In the main CELT study the costs in the month prior to death were estimated at patient level using a linear regression model, based upon patient age and liver disease group. However, if these costs are to be applied, then it is necessary to know which patients are going to die during the study period.

Altman and Royston state that "our ability to provide informative prognosis at the individual level ... is almost always limited" [Altman & Royston, 2000]. For the majority of studies it is not necessary to estimate individual patient outcomes, because outcomes are observed for all individuals participating in the study. However, in some situations, for example, the prognosis of cancer patients or patients with end-stage organ failure, the estimated outcome can influence the choice of treatment or help families and patients come to terms with their illnesses and it becomes necessary to predict individual patient outcomes [Henderson *et al,* 2001]. Additionally, individual estimates of patient outcomes may aid in allocating "the effective use of limited health care resources" [Henderson *et al,* 2001]. Moreover, the estimation of individual patient outcomes only arises when survival is estimated over a fixed time period, such as a five-year period. If the study objective is to estimate lifetime costs, effects, or costs and effects, it would not be necessary to estimate outcomes at an individual level, as all patients will die.

To summarise, the analytical need to predict individual patient outcomes can be generalised to studies where the following conditions apply:

- observed information on patient outcomes are unavailable and prognostic models are utilised to estimate patient outcomes and
- outcomes are to be estimated over a fixed study period and
- individual outcome estimates are needed, for example to:
  - estimate the survival, QALY or costs at the patient level
  - estimate the survival, QALY or cost gain at the patient level
  - make an adjustment to costs, or HRQL data for a time period prior to death

If it were possible to predict the lifetime survival of patients from published prognostic models, then it would also be possible to state, at any particular time point, which

patients were alive and which were dead. However, for some diseases or treatments (including liver transplantation) the authors of prognostic models tend to provide information to estimate survival over a fixed time period. For example, the original versions of the European and PBC Mayo prognostic models for predicting survival in patients with PBC publish information for estimating survival to eight years and seven years, respectively [Christensen *et al,* 1985; Dickson *et al,* 1985].

One reason that authors of published prognostic models do not always give enough information to predict the time of later deaths is because survival predictions at later time points have more uncertainty around them than at earlier time points. This is a well-known limitation of the Cox PH regression model and arises through the manner in which the estimation is formulated. The regression equation bases the survival estimate at any time on the estimated hazard, which is the probability of death among patients who remain alive and under observation. Since few patients remain under observation at the later time points (the majority having previously died or having been censored), the estimate of the hazard (and consequently survival) at later times is imprecise. In these situations, methods are needed to predict individual patient outcomes.

Suppose that we are conducting a study where we need to use prognostic models to estimate individual patient outcomes, and we have chosen a prognostic model, which we then apply to a cohort of patients to estimate their survival over a fixed time period. For example, in the CELT study we might choose to apply the PBC Mayo prognostic model (Chapter 5, Table 5.2 & 5.3) to a cohort of patients with end-stage PBC to estimate their survival in the absence of transplantation over a five-year period. Assuming that the model we are applying is a Cox PH model, it is possible to obtain individual estimates of the probability of surviving over a series of time points, where the probability of surviving to any one time point will range between zero and one. These probabilities can then be plotted over time and an individual's mean survival time can be calculated from the area under their resultant survival curve (Chapter 5, Section 5.4). The predicted survival over the duration of the five-year study period, for the PBC CELT patients, ranged from 0.13 years to 4.93 years, depending on the individual's prognosis. Given that each of the patients has a predicted mean survival time of less than five years; one might (naively) assume that all patients die during the study period. If this were true, a death cost would then need to be added to all patients if the study was also estimating costs.

However, the nature of the Cox PH prognostic model is such that survival estimates will always range from slightly greater than zero to slightly less than the final study time point, e.g. slightly less than five years. Thus, no single patient can have a predicted survival time greater than the last time point of interest (e.g. five years) when applying a Cox PH prognostic model to a cohort of patients over a fixed time period. In addition to survival estimates over the fixed time period, information is available on the probability of survival at the last time point of interest (e.g. five years) and it is possible to estimate the proportion of patients who survive to the end of the study by taking the average survival probability for the cohort at the final time point. The average survival probability, when applying the PBC Mayo model to the PBC CELT cohort over a five-year period is 0.315, thus 31.5% of PBC CELT patients are expected to survive for at least five years.

It is therefore clear that the predicted survival (calculated from the area under the survival curve) over a fixed time period does not itself tell us whether the patient would survive the study period or not. Therefore, survival probabilities at the fixed time point of interest can be used to estimate patient outcomes, rather than assuming the death of all patients within a fixed study period.

Suppose that a patient has an expected survival length of 4.93 years and a 0.98 probability of survival at five years. Given this information, we might revise our belief that this patient died at 4.93 years, and instead assume that they survived the full five-year study period. It is reasonable to prefer this logic over the above argument, as the patient had only a two per cent chance of death at five years. We might next consider another patient who had a survival probability of 0.85 at five years (15% probability of death at five years) and an estimated survival length of 4.58 years and decide that this patient probably survived to five years, and alter their estimated survival accordingly. This process can be repeated for the whole cohort of patients and will result in the selection of a proportion of patients who, based on their survival probabilities at five years, would probably have survived the full study period. Altering these patients' survival lengths would increase the mean survival time for the cohort. Any additional outcome estimates that were dependent upon survival, for example QALYs and costs would also change for those patients assumed to survive the full study period.

At some point it will become necessary to decide what the cut off survival probability is, when using the approach described above. In other words, at what probability X do we

say, patients with a probability less than X will die, and those with a probability greater than X will live? The choice of this probability will typically be based on empirical evidence.

This chapter focuses on introducing a series of methods for estimating patient outcomes, and the uncertainty around the patient outcome estimates, after using prognostic models to estimate survival. At the fixed time point of interest we wish to know, not only the proportion of patients surviving, but which patients survive. In the particular instance of the CELT study, this information is needed in order to assign individual patient costs to patients who die in the month prior to death and to estimate the patient specific survival, QALY and cost gains from liver transplantation over a five-year period.

This chapter is divided in to a further five sections. Section 6.2 presents an overview of methods for estimating individual patient outcomes, identified in the transplant literature review, previously presented in Chapter 3. Section 6.3 introduces the PBC CELT cohort used throughout this chapter to estimate individual patient outcomes in the absence of transplantation. Methods for estimating patient outcomes and the uncertainty surrounding the outcome estimates are discussed in Section 6.4, where a preferred method is recommended. In the CELT study patient outcome estimates impacted on non-transplant QALY and cost estimates, therefore, Section 6.5 investigates how the prediction of individual patient outcomes, and outcome uncertainty, effects non-transplant QALY and cost estimates. The final section of this chapter (Section 6.6) will discuss the implications of the methods introduced here and consider their implementation outside the field of transplantation.

6.2 A BRIEF OVERVIEW OF THE ESTIMATION OF INDIVIDUAL PATIENT OUTCOMES IN SOLID-ORGAN TRANSPLANTATION STUDIES

In this section, we return to the results of the solid organ transplant literature review of cost, effectiveness and cost-effectiveness studies, first presented in Chapter 3. The review considered how issues of uncertainty were dealt with in non-RCT studies and identified further issues of statistical uncertainty, one of which was estimating individual patient outcomes in the absence of observed data. Four of the 158 studies included in the review used Cox PH prognostic models to estimate survival in the absence of observed data [Bonsel *et al,* 1990; Christensen *et al,* 1999; Longworth, *et al,* 2003; Longworth *et al*, 2003a]. Of these studies, two compared survival, with and in the

absence of transplantation, at patient level [Christensen *et al,* 1999; Longworth *et al,* 2003]. Additionally, three of the 158 studies (one of which also used prognostic models) included in the review made an adjustment to costs for a time period prior to death [Van Enckevort *et al,* 1997; Anyanwu *et al,* 2002; Longworth *et al,* 2003]. However, Van Enckevort *et al* and Anyanwu *et al* estimated the lifetime costs of lung transplantation, where all patients will eventually die [Van Enckevort *et al,* 1997; Anyanwu *et al,* 2002].

The third of the three papers that adjusted for costs in the final period of life was the main CELT study, which estimated the cost-effectiveness of liver transplantation over a 2.25 year study period [Longworth *et al,* 2003]. The CELT study also estimated the survival, QALY and cost gain of liver transplantation on a per patient basis. In the main CELT study, non-transplant costs were assumed to be constant over time with an adjustment made for costs in the month prior to death for patients assumed to die within the 2.25 year study period[1]. The published CELT paper fails to present details of how individual patients were estimated to survive or die within the 2.25 year study period [Longworth *et al,* 2003]. However, from personal involvement in the CELT study, I can state that the cut off length for survivors were selected by examining the expected survival times for each prognostic model and assuming that patients surviving for two years or more survived the full 2.25 year study period. For some of the prognostic models in the CELT study no patients' expected non-transplant survival time was greater than two years, in these situations the cut-off time was decreased in 0.25 year increments until at least one patient survived until the end of the study. For example, in the main ALD CELT analysis the survivors were selected according to whether their predicted survival, estimated by applying the Beclere prognostic model to the ALD CELT patients, was greater than 1.75 years or not. Yet the ALD model used in the sensitivity analysis selected survivors with an expected survival time greater than two years. The selection process was not based on information available from the prognostic model, nor from any other source of data. Given that this method is essentially arbitrary it will not be considered further in this chapter.

Christensen and colleagues also used Cox PH prognostic models to estimate survival, with and in the absence of liver transplantation and computed the patient level survival gain from transplantation over a six-month period [Christensen *et al,* 1999]. Two

[1] The choice of a one month adjustment period was based on an examination of costs on the waiting list for liver transplantation observed in the main CELT study (see Appendix A2.2).

prognostic models were applied to a series of patients; the first estimated their survival over six months without transplantation and the second estimated their survival with transplantation. The authors did not estimate patient level outcomes at six months.

Although they do not estimate individual patient outcomes, O'Brien *et al* use extrapolation techniques to estimate cohort survival over a 20-year period, both with and in the absence of heart transplantation [O'Brien *et al,* 1987]. The authors used three different extrapolation models; linear extrapolation, Weibull survival model and the exponential survival model, to extrapolate transplant survival beyond a five-year observed period. However, this is extrapolating survival at the cohort level and not the individual patient level. Additionally, it is not possible to apply a survival model approach, such as fitting a Weibull model, to the estimated survival from the Cox PH prognostic model without making some inference about the survival outcome of, at least a proportion of, patients. Fitting a survival model requires a proportion of patients to have the event of interest and this returns to the problem of how to allocate the proportion who survive and the proportion who die.

It is concluded that none of the studies that were identified in the solid organ transplant literature review use adequate methods for estimating individual patient outcomes. Therefore, after presenting details of the PBC CELT data set, this chapter will present three possible methods for estimating individual patient outcomes over a fixed study time period. Each of the three methods will utilise the limited information available from three PBC prognostic models for estimating non-transplant survival.

6.3 ESTIMATES OF NON-TRANSPLANT SURVIVAL FOR THE PBC CELT COHORT

As in the previous chapter, a sub-cohort of 81 CELT patients with end-stage PBC will be used to illustrate a series of three proposed techniques for estimating individual patient outcomes in the absence of liver transplantation. All three PBC prognostic models; European model, PBC Mayo model and Royal Free model, which were described in Chapter 5, will be applied to the 81 PBC patients to estimate their survival in the absence of transplantation over a five-year period. The three techniques for estimating individual patient outcomes will be set out in full for the PBC Mayo model and summarised for the European and Royal Free models.

After applying the three prognostic models, the mean non-transplant survival over the five-year study period, prior to estimating individual patient outcome was:

- 2.47 years – PBC Mayo model
- 2.45 years – European model
- 2.50 years – Royal Free model

Mean survival estimates are similar across the three models, differing by approximately three weeks between the European model and the Royal Free model.

6.4 METHODS FOR PREDICTING INDIVIDUAL PATIENT OUTCOMES

In this section, three methods for estimating individual patient outcomes are considered. The average non-transplant survival and the average number of survivors, together with their 95% CI, are presented for each method.

All three methods utilise information on the probability of survival at the final study time point, in the CELT study this is the expected non-transplant survival probability at five years. The first method described in this section uses information on the mean probability of survival for a cohort of patients at the final time point of a study. Subsequent techniques account for patient specific information available after applying a prognostic model to a cohort of patients.

6.4.1 Predicting Individual Patient Outcomes: Probability of Survival Equivalent for all Patients, Random Selection of Survivors (Method 6.4.1)

In the absence of any information on patient mortality, a reasonable starting point would be to assume that each patient has an equal chance of survival or death:

(Probability or survival = Probability of death = 0.5)

However, it is unlikely that the researcher has no expectations about the proportion of survivors in their cohort. Possible sources of information on the likely outcome of patients may come from expert opinions, published information, or alternative data sources. In this chapter, information from published prognostic models is utilised, in order to estimate survival and patient outcomes over a five-year period. Information on the mean probability of survival at five years, for the PBC CELT cohort, after applying

the PBC Mayo prognostic model is used as a starting point. The mean probability of survival at five years is 0.315 for the 81 PBC CELT patients.

The mean survival probability can be expressed statistically by assuming that the probability of survival follows a binomial distribution, where the probability of an individual patient surviving to five years is 0.315. Formally, this probability is written as $P_{Si} = 0.315$, where i = 1 to 81 refers to the patient identifier. A random number generator is used to simulate the expected outcome for each individual patient in a cohort. This random number generator produces a number, either zero or one, for each patient. A survival prediction can be generated for each of the 81 cases in the PBC CELT cohort by assuming that the occurrence of a one denotes that the individual will survive the full study period without transplantation and a zero denotes a patient who will die within the five-year study period.

The following computer syntax is used in S-PLUS [S-PLUS 6, 2001] to simulate this process:

```
rbinom(81, 1, 0.315)
```

For one simulation run, a series of estimated events for each of the 81 PBC patients are obtained, as illustrated below:

0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 1, 1, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 1, 1, 0, 0, 1, 0, 1, 0, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 1, 0, 1, 0, 0, 0, 1, 1, 0, 0, 0, 1, 0, 1, 0, 0, 0, 0

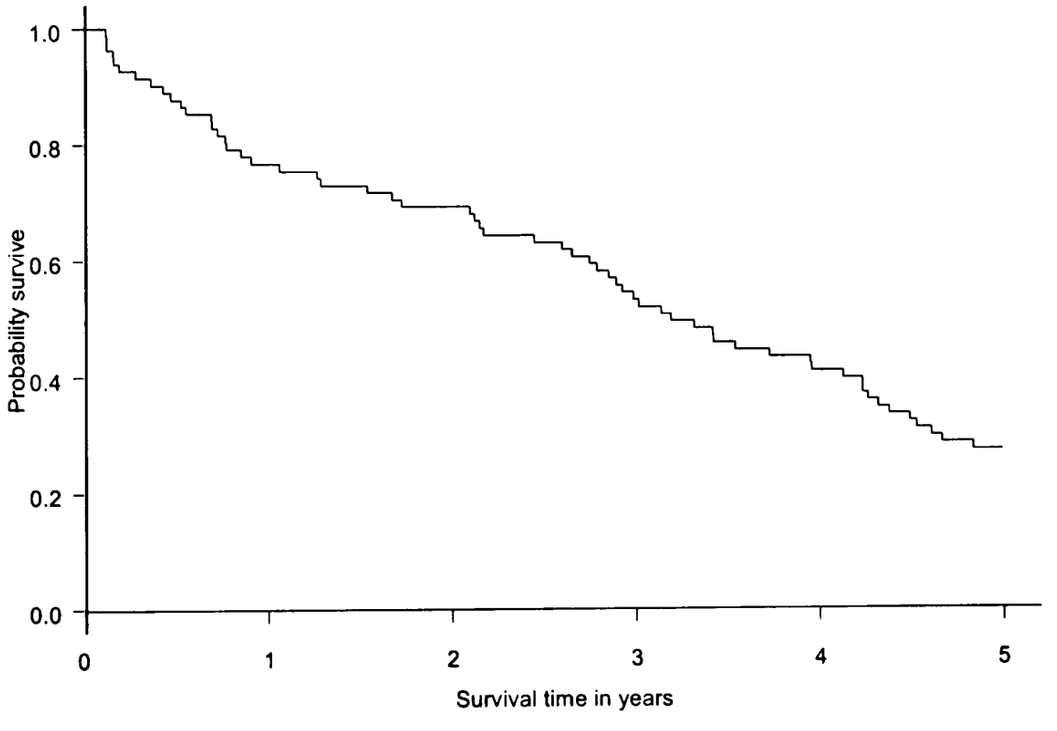
The number and proportion of survivors in the cohort can then be calculated. For the simulation run presented above, the number of survivors equals 22, giving an estimated five-year survival of 27% (for this simulation run the percentage of survivors is lower than the "true" percentage of 31.5%).

It is assumed that all estimated survivors survive the full study period and have a survival length of five years. The survival length for patients predicted to die is estimated by applying the PBC Mayo prognostic model to the PBC CELT cohort. The mean non-transplant survival length for the PBC CELT cohort can then be estimated from the area under the Kaplan-Meier survival curve.

Method 6.4.1 is illustrated in Appendix A6.1 for the PBC CELT cohort for one simulation run. Table A6.1 shows the expected survival lengths for the PBC CELT cohort, the predicted survival outcomes for one simulation run and the corresponding expected survival times.

The 22 predicted survivors were assumed to have an expected survival length of five years and the 59 predicted deaths an estimated non-transplant survival length calculated using the PBC Mayo prognostic model. The Kaplan-Meier survival curve for 81 patients is shown in Figure 6.1, using the results from one simulation run. The mean survival length, estimated from the area under the curve is 3.0 years.

Figure 6.1 Kaplan-Meier survival curve for 81 CELT PBC patients where expected individual patient survival is predicted assuming the probability of survival is equivalent for all patients, random selection of survivors (p = 0.315)[1]



[1] Results from one simulation run

To generate uncertainty surrounding the estimates of the percentage of expected non-transplant survivors, or the mean non-transplant survival length, more than one simulation is necessary. A total of 1,000 simulations were run in order to measure the

uncertainty around the predicted percentage of survivors in the cohort, and the mean survival time for the cohort. One thousand simulations were found to be sufficient to obtain convergence to the mean percentage of survivors and mean survival lengths.

The proportion of patients surviving, and the mean survival times, were recorded for each of the 1,000 simulations. The 50^{th} percentile of the 1,000 simulated outcomes represents the mean percentage of survivors, and the 50^{th} percentile of the estimated survival times represents the mean survival length for the cohort across all simulations. The 2.5^{th} and 97.5^{th} percentiles represent the 95% CI.

Method 6.4.1 measures two areas of uncertainty. Firstly, the patients who are assumed to survive to five years are varied between simulation runs. This allows many possible permutations of patients selected as survivors and patients selected as deaths. Secondly, the proportion of patients expected to survive will vary by simulation run, although across the 1,000 runs it will average out at approximately 31.5%. In fact, the mean proportion of survivors over the 1,000 simulation runs was 31% (95% CI: 17% to 34%). If a larger number of simulations had been run, say 10,000 simulations, the mean proportion would have converged further to the 31.5%, set by using a binomial distribution with $p = 0.315$.

The mean survival time across 1,000 simulations is 3.08 years (95% CI: 2.74 to 3.39 years). This is approximately nine months greater than the mean survival time from the PBC Mayo prognostic model prior to estimating patient outcomes (Mean = 2.47 years). This result is expected, because approximately 31 per cent of the simulated patients have their expected survival time extended to five years.

However, the method described above ignores information related to individual patient's survival lengths. To illustrate what I mean by this, let us take five patients for whom information is available on expected survival times from a prognostic model (Table 6.1). Method 6.4.1 has been applied to these five patients to simulate their survival outcomes and their revised expected survival lengths, over a five-year study period.

Table 6.1 Results from one simulation run for predicting patient outcome for five hypothetical patients (Applying Method 6.4.1)

Patient	Predicted survival time (estimated from a prognostic model)	Simulated patient outcome	Revised survival time after simulation
A	0.13 years	Died	0.13 years
B	4.93 years	Survived	5.00 years
C	0.25 years	Survived	5.00 years
D	4.88 years	Died	4.88 years
E	2.50 years	Died	2.50 years

Patient A had a prognostic model survival time, prior to simulation of expected survivors, of 0.13 and for this simulation run was expected to die. This appears to be a reasonable estimate, based on patient A's prognostic model survival time. The outcome estimate of survival for patient B also appears to be a reasonable estimate, as their prognostic model survival estimate, prior to simulation, is 4.93 years. However, patient C has a pre-simulation survival estimate of 0.25 years yet is simulated as expecting to survive five years in the absence of transplantation and patient D a survival length of nearly five years is expected to die. Patient E survival estimate is 2.5 years and is an expected death.

This anomaly has arisen as a result of attempting to apply a group-level probability of 0.315 to each of the individual patients. On average each patient will be predicted to survive in 315 of the simulations and be predicted to die in 685 simulations, when performing 1,000 simulations and assuming an overall cohort survival probability of 0.315. However, from examining the expected survival times prior to simulation it appears unrealistic to assume that patients B and D will only survive in 31.5% of simulations; it is expected that they will survive in a higher proportion of simulation runs than patients A, C and E. Likewise, based on their expected survival lengths, it is expected that patients A and C die in more simulation runs than patients B, D and E.

Method 6.4.1 will give an accurate estimate of the mean proportion of survivors because the five-year survival probability is based on available prior information relating to survival outcome. However, Method 6.4.1 will overestimate the mean survival length. To understand why, let us return to the five example patients detailed in Table 6.1. Prior to simulating patient outcomes, the mean survival time for these five

patients is 2.54 years. Suppose that in the first simulation run patient B is selected as the only survivor, meaning that their survival time over a five-year study period is now assumed to be five years. The mean survival for the cohort increases slightly to 2.55 years. In a second simulation run patient A is predicted as the only survivor and their survival time is assumed to be five years, increasing the mean survival time for the cohort by approximately one-year, to 3.51 years.

Thus for any one simulation run, if a survival rate of 31.5% is assumed to apply to any individual patient (regardless of their prognosis), then on average 31.5% of the cohort will be randomly selected as survivors. Further, a proportion of these patients will have a poor prognosis with a short predicted non-transplant survival time (e.g. one-year). The individual patient survival times for the simulated survivors are set to five years and the randomly selected patients with short survival times will increase the mean survival time to being longer than what it might have been, had the original prognostic model predicted survival times been accurate.

In summary, we have assumed that the probability of any individual patient surviving for five years is 0.315. We have then used a random number generator to define whether each person survived until five years or not and estimated the individual's survival time to be equal to either five years (if they were defined as having lived), or to their originally predicted survival time (if they were defined as having died). This simulation process was repeated to allow the uncertainty in the process to be estimated. Method 6.4.2 illustrates how extra available information can be incorporated in to the estimation of patient outcomes.

6.4.2 Predicting Individual Patient Outcomes: Selecting the N^2 Patients with the Longest Survival Times as Survivors (Method 6.4.2)

The second method proposed in this chapter for estimating patient level outcomes attempts to customise the predicted outcome to the patient, rather than to assume the chances of surviving five years are the same for every individual. The method that is set out here assumes that information is available regarding the average proportion of survivors for a cohort.

The mean probability of survival, in the absence of transplantation, at five years can be converted in to the average number of survivors by multiplying the study sample size

[2] N is the average number of survivors at the end of the fixed study period.

by the mean survival probability. The mean probability of survival at five years for the 81 PBC CELT patients, estimated using the PBC Mayo model, is 0.315 (95% CI: 0.247 to 0.383). The mean expected number of survivors at five years is therefore:

$$81 \times 0.315 = 26 \text{ patients}^3$$

The 95% CI for the expected number of survivors is calculated in a similar way and ranges from 20 patients (81x0.247 = 20) to 31 patients (81x0.383 = 31).

Expected non-transplant survival estimates are available for each patient after applying the PBC Mayo prognostic model to the PBC CELT cohort. It seems reasonable to use this information when selecting the survivors for the cohort. On average, 26 patients survive, and it is assumed that the 26 patients with the longest survival times, over the five-year study period, survive to five years and the remaining 55 patients with shorter survival times die during the study period. Appendix A6.1 Table A6.2 illustrates the selection process for the PBC CELT cohort assuming the 26 patients with the longest survival times survive. The mean non-transplant survival time for the PBC CELT cohort, estimated from the area under the Kaplan-Meier survival curve, is 2.59 years.

To allow for uncertainty in the estimated number of survivors, the analysis is repeated using one-way sensitivity analysis. Two sensitivity analysis are performed, firstly using the lower 95% confidence limit for the probability of surviving to five years (N = 20) and secondly the upper limit (N = 31), giving a 95% CI for survival of 2.48 years to 2.70 years.

One of the drawbacks of this method is that its predicted outcomes are deterministic, and do not allow any uncertainty in the selection of cases estimated to survive or die. Survival outcome priority is given to the cases with the highest survival times, and although this is not an unreasonable assumption to make, it does not allow for a random element, where cases with a poor survival probability survive longer than expected, or those with a good survival probability suffer some form of complication and die unexpectedly. Thus, although the mean survival estimates from this method will be more accurate than those from Method 6.4.1, there is likely to be some extra

[3] Given that is not feasible to have 0.1 of a patient numbers have been rounded up to the next whole number.

uncertainty due to chance, which if not allowed for, would result in artificially narrow CI. The final method in this section uses techniques to adjust for this uncertainty.

6.4.3 Predicting Individual Patient Outcomes: Probabilistic Sensitivity Analysis (Method 6.4.3)

The final method considered in this section uses probabilistic sensitivity analysis (PSA) to incorporate uncertainty in individual outcome estimates. In PSA, statistical distributions are assigned to parameters of interest and Monte Carlo simulations subsequently run to re-estimate both the outcome of interest and the uncertainty around it [Doubilet *et al,* 1985; Critchfield *et al,* 1986]. PSA is a useful method to use in economic evaluations, where there is often significant parameter uncertainty behind generated outcomes.

In this section, each patient's expected non-transplant survival probability at five years, estimated from the PBC Mayo model, will be utilised in PSA to assign an individual patient outcome distribution. Each patient's outcome will be assumed to follow a binomial distribution.

Table 6.2 presents expected non-transplant survival probabilities, at five years, for three hypothetical patients. Each patient's outcome is assumed to follow a binomial distribution and 1,000 simulations are run for each of the three patients to predict individual patient outcomes. Over the course of 1,000 simulations patient A, who has a predicted survival probability of 0.09, will on average survive in 90 of the simulations and will die in the remaining 910. Patient B on the other hand will, on average, be expected to survive in 500 out of 1,000 simulations, and likewise patient C is expected to survive in 980 of the simulations.

Table 6.2 Illustration of the expected number of survival outcomes and death outcomes for three patients over a series of 1,000 simulations

Patient	Expected survival probability	Average number of times predicted to survive	Average number of times predicted to die
A	0.09	90	910
B	0.50	500	500
C	0.98	980	20

PSA can be run in the S-PLUS statistical package, and the S-PLUS syntax, below, will generate a set of survival outcomes for 81 PBC CELT patients. Each patient's five-year non-transplant survival probability is used to estimate patient outcomes, where survival probabilities are based on estimates from the PBC Mayo prognostic model at five years.

```
res <- numeric(81)
for(i in 1:81) res[i] <- rbind(1, 1, Prob5y[i])
```

Prob5y is the data field which contains the predicted survival probabilities for each of the 81 individual PBC patients at five years. Appendix A6.1 presents an illustration of the PSA method for one simulation run in further detail.

A total of 1,000 simulations are run in order to measure the uncertainty in the predicted patient outcomes and the proportion of patients surviving. One thousand simulations were found to be sufficient to obtain convergence to the mean percentage of survivors and mean survival lengths. The mean survival length from 1,000 simulation runs was 2.60 years (95% CI: 2.46 to 2.75 years) and the mean proportion of survivors was 32% (95% CI: 25% to 40%).

The 95% CI for the mean proportion of survivors and mean survival lengths are virtually identical to those for Method 6.4.2. The slight difference in mean outcome and survival estimates between Methods 6.4.2 and 6.4.3 are due to the PSA method allowing for the genuine uncertainty in the selection of patients who survive the full study period or die during the study period.

6.4.4 Summary of Three Methods for Estimating Individual Patient Outcomes: PBC CELT Cohort – PBC Mayo Model

Table 6.3 summarises the mean proportion of survivors, and the mean survival estimates, for the three alternative methods for predicting individual patient outcomes, for the PBC CELT cohort after applying the PBC Mayo prognostic model to estimate non-transplant survival. Method 6.4.1 assumes that each patient has the same survival probability, set as the mean probability of survival at five years ($p = 0.315$ for the PBC Mayo model). The method of assuming that the probability of survival was equivalent for all patients (random selection of survivors) gave a mean survival estimate of 3.06 years. Method 6.4.1 will always overestimate the mean survival time because this

method does not account for individual patient information on survival and randomly allocates the proportion of survivors over the entire cohort.

Method 6.4.2 utilises information on the probability of survival at five years and selects the N patients with the longest survival times as survivors. Method 6.4.2 will give a more accurate estimate of the mean survival than Method 6.4.1 but does not allow for the random selection of survivors by chance (the patients with the longest survival times are always selected as survivors). Whereas, Method 6.4.3 allows for the random selection of survivors using a PSA approach. Method 6.4.3 and 6.4.2 result in similar mean survival estimates, but the 95% CI for Method 6.4.3 are slightly wider, reflecting the genuine uncertainty allowed for using this method.

Table 6.3 Mean percentage of PBC CELT non-transplant survivors and mean survival (95% CI) after applying three methods for predicting patient outcome – PBC Mayo model

Method	Mean percentage of survivors	Mean survival time in years
Method 6.4.1: Probability of survival equivalent for all patients, random selection of survivors (p = 0.315)	31% (17% to 34%)	3.06 (2.74 to 3.39)
Method 6.4.2: Selecting the N patients with the longest survival times as survivors (N = 26)	32% (25% to 38%)	2.59 (2.48 to 2.70)
Method 6.4.3: Probabilistic Sensitivity Analysis	32% (25% to 40%)	2.59 (2.46 to 2.75)

6.4.5 Summary of Three Methods for Estimating Individual Patient Outcomes: PBC CELT Cohort – European and Royal Free Models

As a further demonstration of the proposed methodology for estimating individual patient outcomes, the same three methods, described above, were used to predict individual patient outcomes, in the absence of transplantation, for 81 PBC CELT patients using two alternative prognostic models. The European prognostic model and the Royal Free prognostic model are applied to the PBC CELT cohort to predict their non-transplant survival over a five-year study period.

To apply Method 6.4.1, the mean non-transplant survival probability at five years was calculated. The mean survival probability at five years was 0.333 for the European model and 0.323 for the Royal Free model. Therefore, for Method 6.4.1, where the probability of survival was equally likely for each of the 81 PBC CELT patients, the mean survival was 3.19 years when applying the European model and 3.20 years when applying the Royal Free model (Table 6.4 and 6.5).

The 95% CI around mean survival probabilities at five years were 0.26 to 0.40 for the European model and 0.25 to 0.39 for the Royal Free model. To estimate the expected number of survivors at five years (Method 6.4.2), the mean proportion of survivors at five years was multiplied by the total sample size and rounded to the nearest whole number. The expected number of survivors was similar for the European and Royal Free models; 27 (95% CI: 21 to 33 patients) and 27 (95% CI: 21 to 32 patients), respectively. After applying Method 6.4.2 to predict individual patient outcomes the mean survival times were 2.64 (European model) and 2.74 (Royal Free model). Survival estimates were similar across the three prognostic models, with mean non-transplant survival estimates differing by only 0.15 years between the PBC Mayo model and the Royal Free model.

The same pattern in mean survival estimates can be seen across all three prognostic models, with Method 6.4.1 producing mean survival estimates that are approximately 0.5 years greater than estimates for Methods 6.4.2 and 6.4.3 for all three models. Method 6.4.2 and Method 6.4.3 produce similar mean survival estimates, but with wider CI for the PSA method, which captures outcome uncertainty at the individual level, across all three prognostic models (Table 6.3 to 6.5).

Table 6.4 **Mean percentage of PBC CELT non-transplant survivors and mean survival (95% CI) after applying three methods for predicting patient outcome – European model**

Method	Mean percentage of survivors	Mean survival time in years
Method 6.4.1: Probability of survival equivalent for all patients, random selection of survivors (p = 0.333)	33% (22% to 43%)	3.19 (2.88 to 3.52)
Method 6.4.2: Selecting the N patients with the longest survival times as survivors (N = 27)	33% (26% to 41%)	2.64 (2.56 to 2.74)
Method 6.4.3: Probabilistic Sensitivity Analysis	33% (27% to 40%)	2.63 (2.52 to 2.75)

Table 6.5 **Mean percentage of PBC CELT non-transplant survivors and mean survival (95% CI) after applying three methods for predicting patient outcome – Royal Free model**

Method	Mean percentage of survivors	Mean survival time in years
Method 6.4.1: Probability of survival equivalent for all patients, random selection of survivors (p = 0.323)	32% (21% to 42%)	3.20 (2.92 to 3.50)
Method 6.4.2: Selecting the N patients with the longest survival times as survivors (N = 27)	32% (25% to 39%)	2.74 (2.64 to 2.84)
Method 6.4.3: Probabilistic Sensitivity Analysis	32% (26% to 40%)	2.72 (2.61 to 2.85)

6.4.6 Overall Summary

Method 6.4.1 uses Monte Carlo simulations to estimate individual patient outcomes and assumes the probability of survival is equivalent for all patients. This method should not be used to estimate individual patient outcomes because it does not utilise patient specific information and will overestimate survival. Applying Method 6.4.1 within the CELT study would, in turn, result in an underestimation of the survival gain from liver transplantation over five years.

Method 6.4.2 selects the N patients with the longest estimated survival times as survivors, where N is the estimate of the average number of survivors at the end of the study period. Method 6.4.3 uses individual patient survival probabilities in a PSA to estimate survival outcomes. Both Method 6.4.2 and 6.4.3 allow for individual patient information on the probability of survival at five years, and both produce similar mean estimates. Method 6.4.3 is recommended here as it allows for the additional uncertainty in the selection of the patients who survive, whereas Method 6.4.2 does not allow for this uncertainty.

However, some researchers may prefer Method 6.4.2 to the PSA method due to the potential additional work load when conducting a PSA analysis for a similar mean estimate with slightly wider CI. The conclusions from comparing the difference in survival between two cohorts of patients are unlikely to differ in the majority of studies, were Method 6.4.2 selected rather than Method 6.4.3, when estimating individual patient outcomes. However, if the difference in survival approached statistical significance then researchers might draw different conclusions had they selected PSA over Method 6.4.2 or visa versa. It should be pointed out that the PSA analysis that incorporated outcome uncertainty took less than five minutes to run. Additionally, one of the advantages of PSA is that, alongside individual patient outcome uncertainty, other sources of uncertainty can also be incorporated (for example, prognostic model parameter uncertainty and outcome uncertainty can be incorporated within the same analysis); this topic will be returned to briefly in Section 6.7 and explored further in Chapter 7.

In conclusion, PSA is recommended here as the preferred method for estimating individual patient outcomes from a prognostic model, over a fixed time period because it allows for patient level uncertainty and can combine other sources of uncertainty within a single PSA run.

6.5 THE IMPACT OF OUTCOME UNCERTAINTY ON NON-TRANSPLANT SURVIVAL, QALY AND COST ESTIMATES

So far, this chapter has investigated the impact of estimating individual patient outcomes and outcome uncertainty on survival. In many studies, survival is not the sole outcome of interest. For example, the CELT study was concerned with estimating survival, QALYs and costs with, and in the absence of, transplantation and non-transplant QALY and cost estimates were dependent upon non-transplant survival

estimates. Therefore, this section investigates how allowing for the uncertainty in outcome estimates impacts upon non-transplant QALY and cost estimates.

As in the previous section, the PBC Mayo, European and Royal Free prognostic models were each applied to 81 PBC CELT patients to estimate their survival in the absence of liver transplantation.

For each of the three PBC prognostic models (PBC Mayo, European and Royal Free) 1,000 PSA simulations (Method 6.4.3) were run to allow for prognostic model outcome uncertainty. For each simulation run the mean survival time, expected number of survivors, QALYs and costs in the absence of transplantation are calculated. One thousand simulations were found to be sufficient to obtain convergence to the mean percentage of survivors and mean survival lengths.

Non-transplant HRQL was measured using the EQ-5D and each PBC patient's last observed HRQL score pre-transplant was assumed to remain constant over time, until time of death. At time of death HRQL was assumed to be zero. Non-transplant QALYs were estimated by multiplying the expected non-transplant survival time by the last observed pre-transplant EQ-5D score.

Individual patient average daily costs were estimated from cost data collected on the CELT patients prior to transplant and predicted for the same patients in the absence of transplantation using a three-part model[4]. The first part of the model predicts whether each patient incurs a (non-zero) daily cost (Equations 6.1 and 6.2). If a non-zero cost is predicted the second part of the model is used to predict what that cost would be (Equation 6.3); otherwise, a daily cost of zero is assigned to the patient. Finally, a death cost is added on for patients who are predicted to die during the study (Equation 6.5).

[4] The method of estimating non-transplant costs, described here, models costs more completely than the approach used in the main CELT analysis [Longworth *et al*, 2003]. In the main CELT analysis individual patient average daily costs on the waiting were used to estimate non-transplant costs and these were assumed constant over time until the month prior to death, Equation 6.5 was then applied to predict costs in the month prior to death.

The model used to predict whether a cost arises is:

$$\text{Logit(Probability of incurring daily costs)} = lop = 1.70 \times \log_{10}(\text{bilirubin}) + 0.082 \times \text{age} - 1.51 \times \text{ascities present} - 4.78 \qquad \textit{Equation 6.1}$$

The probability of incurring daily costs is calculated by applying Equation 6.2

$$\text{Probability of incurring daily costs} = \frac{\exp(lop)}{1+\exp(lop)} \qquad \textit{Equation 6.2}$$

For patients who are predicted to have a cost, the value of this cost in UK pounds is predicted by the model:

$$\text{Expected daily } \log_e \text{ cost} = E[\log_e(\overline{C})] = 1.84 + 1.31 \times \text{female} + 0.68 \times \text{ascities present} \qquad \textit{Equation 6.3}$$

Finally, the probability of incurring a daily cost is multiplied by the expected daily cost to obtain the predicted cost per day in the absence of transplantation (Equation 6.4).

$$\text{Predicted cost per day} = \left[\frac{\exp(lop)}{1+\exp(lop)}\right] \times (\{\exp(E[\log_e(\overline{C})])\} \times 3.51) \qquad \textit{Equation 6.4}$$

The retransformation of log transformed data results in an estimation of the geometric mean rather than the arithmetic mean [Graham *et al,* 1988]. Duan proposed the smearing estimate which corrects for bias in taking the exponential the log transformed costs to obtain the arithmetic mean cost rather than the geometric mean cost [Duan, 1983]. The smearing estimate is calculated by summing the exponential of the residuals for the log transformed ordinary least squares model (Equation 6.3) and taking the average over the total number of observations in the sample (N = 81 in this chapter). The smearing estimate for Equation 6.3 is 3.51, and is applied when retransforming the log daily cost estimates back in to UK pounds in order to obtain an unbiased estimate of expected daily costs.

The five year costs in the absence of transplantation for survivors were calculated by multiplying the predicted daily cost on the waiting list by 1,826 days (five years). For patients who were predicted to die within the five-year period, an adjustment was made to the non-transplant costs by adding on additional costs in the final month of life:

Costs in the month prior to death (£) = (101.6 x age) – 1,676.4 ***Equation 6.5***

The total non-transplant costs for each patient, was then calculated using Equation 6.6.

Equation 6.6

Total non-transplant costs (£)
= (Predicted daily cost on waiting list x [Predicted survival time (in days) – 30])
+ Cost in the month prior to death
(if patient predicted to die within five years)
OR
= Predicted daily cost on the waiting list x 1,826
(if patient predicted to survive five years)

The mean total non-transplant costs for the PBC CELT cohort were then estimated for each of the three PBC prognostic models and are presented in Table 6.6 with the mean percentage of survivors at five years, the mean survival estimates and mean QALYs for the three PBC prognostic models. The means and 95% CI were derived from the 50th, 2.5th and 97.5th percentiles of the one thousand simulations.

Mean survival, QALY and cost estimates and the widths of the 95% CI were similar across the three prognostic models, with survival and QALY estimates being lowest for the Mayo model and highest for the Royal Free model.

To compare the impact of accounting for outcome uncertainties at the individual level, non-transplant survival, QALY and cost were also estimated for Method 6.4.2 which selected the N patients with the longest estimated survival times (Table 6.7)[5].

[5] When comparing the results from Table 6.6 with Table 6.7 it should be born in mind that Method 6.4.2 and Method 6.4.3 (PSA) are presenting slightly different summary measures which could explain the disparity in mean costs between methods. Method 6.4.2 presents cohort mean survival, QALY and cost estimates with confidence intervals estimated from cohort mean from two sensitivity analysis. PSA produces 1,000 estimates of the cohort mean survival, QALY and cost and summarises the mean and 95% CI from the 50th, 2.5th and 97.5th percentiles of the simulated means.

A comparison of the QALY and cost estimate between Method 6.4.2 and PSA (Method 6.4.3) shows that, like survival estimates, allowing for prognostic model outcome uncertainty slightly increases the uncertainty around non-transplant QALY and cost estimates (Table 6.8)

Table 6.6 Mean percentage of survivors, survival, QALYs and total non-transplant costs (95% CIs) over five years for 81 PBC CELT patients estimated by the PBC Mayo, European or Royal Free prognostic models: using PSA (Method 6.4.3) to incorporate prognostic model outcome uncertainty

	PBC Mayo Model	European Model	Royal Free Model
Mean percentage of survivors	32 (25 to 40)	33 (27 to 40)	32 (26 to 40)
Mean survival in years	2.59 (2.46 to 2.75)	2.63 (2.52 to 2.75)	2.72 (2.61 to 2.85)
Mean QALYs in years	1.34 (1.26 to 1.44)	1.40 (1.33 to 1.48)	1.45 (1.38 to 1.54)
Mean non-transplant costs (£'s)	£82,178 (£77,512 to £88,358)	£80,951 (£77,057 to £85,551)	£81,631 (£77,442 to £86,608)

Table 6.7 Mean percentage of survivors, survival, QALYs and total non-transplant costs (95% CIs) over five years for 81 PBC CELT patients estimated by the PBC Mayo, European or Royal Free prognostic models: using Method 6.4.2 to incorporate prognostic model outcome uncertainty

	PBC Mayo Model	European Model	Royal Free Model
Mean percentage of survivors	32 (25 to 38)	33 (26 to 41)	32 (25 to 39)
Mean survival in years	2.59 (2.48 to 2.70)	2.64 (2.56 to 2.74)	2.74 (2.64 to 2.84)
Mean QALYs in years	1.28 (1.22 to 1.33)	1.35 (1.30 to 1.42)	1.41 (1.35 to 1.48)
Mean non-transplant costs (£'s)	£78,929 (£75,369 to £80,990)	£74,545 (£72,079 to £77,254)	£80,118 (£77,734 to £83,178)

Table 6.8 A Comparison of 95% CI Widths for Non-Transplant QALY and Cost Estimates for the PBC Mayo model, European Model and Royal Free Model after applying Method 6.4.3 (PSA) and Method 6.4.2

		95% CI Widths		
		PBC Mayo Model	European Model	Royal Free Model
Non-transplant QALYs	Method 6.4.3 (PSA)	0.18	0.15	0.16
	Method 6.4.2	0.11	0.12	0.13
Non-transplant costs	Method 6.4.3 (PSA)	£10,846	£8,494	£9,166
	Method 6.4.2	£5,621	£5,175	£5,444

6.6 DISCUSSION

The published prognostic models used in the CELT study did not provide enough information to estimate "lifetime" patient survival. Additionally, the prognostic models did not provide information as to who was alive or dead at a fixed prior time point (five years). Therefore, this chapter has presented methodology that can be used to account for uncertainties that arise from applying Cox PH prognostic models to estimate survival, QALYs and costs in the absence of observed data for a cohort of patients over a fixed time period.

6.6.1 Alternative Approaches to Estimating Individual Patient Outcomes

Three alternative approaches have been presented in this chapter that can be applied to survival estimates in order to estimate individual patient outcomes. An alternative approach that was excluded from Section 6.4 was to fit a mathematical model, for example, a linear regression model, to the prognostic model estimates at either the cohort or patient level, in order to extrapolate survival beyond the study time period. However, this involves fitting prognostic models to prognostic models. This will inevitably increase the amount of uncertainty in the estimation process. Therefore, methods that utilised information from prognostic models were focused upon.

A further method for estimating individual patient outcomes, excluded from this chapter, was to treat the unknown outcome as missing data. This approach was used recently by Oostenbrink and colleagues who apply missing data techniques to cost data, where costs information is incomplete for a proportion of patients [Oostenbrink *et al*, 2003; Oostenbrink & Al, 2005].

Missing data techniques assume that data are either missing completely at random (MCAR) or MAR [Little & Rubin, 1987; Little, 1993]. MCAR occurs when the missing data are independent of observed and unobserved data, in other words the missing data occurs in an entirely random fashion. Data are said to be MAR if the values of the variable(s) which contain the missing data are not a sub-sample of the sampled observations but are a random sample of variables that depend on the values of observed variable(s).

It would be inappropriate to apply missing data techniques to estimate individual patient outcomes in the absence of transplantation in the CELT study. To use imputation techniques, a proportion of patients in the cohort should be known to have had an observed death and a proportion known to survive to the end of the study. This was not the case when predicting non-transplant survival, where outcomes were missing for all patients in the cohort. It was considered inappropriate to combine the non-transplant cohort with the transplant cohort as if they were two separate data sets, or to combine the non-transplant cohort with patients who were listed for transplant within the CELT study but who did not receive transplantation. The reasons in each case are that the MAR assumption is violated. In the first scenario, the transplant cohort are a sub-sample of the combined cohort of transplant and non-transplant patients. In the second scenario, the "non-transplant" patients were on the waiting list long enough for a suitable organ to become available and are a sub-sample of patients on the waiting list.

6.7 CONCLUSIONS

This chapter has addressed an issue of methodological uncertainty by presenting and comparing three alternative methods for estimating individual patient outcomes. PSA uses individual patient survival probabilities to estimate survival outcomes and is the recommended method for estimating the uncertainty around individual patient outcome estimates derived from prognostic models. This approach is selected over a deterministic approach where the N patients with the longest survival times are selected as survivors (Method 6.4.2). Both PSA and Method 6.4.2 gave a similar mean percentage of survivors and survival estimates, however, PSA gave wider (but more realistic) CI owing to the extra uncertainty it allowed for.

In the CELT study, outcome uncertainty had little impact on survival, cost and QALY estimates and increased the uncertainty around non-transplant estimates slightly and did not differ noticeably for the three PBC prognostic models. Each patient's expected

outcome is dependent on their expected survival time over the fixed study period, therefore, it follows that any estimates of patient outcomes and survival times should incorporate both prognostic model parameter uncertainty, using the methods presented in Chapter 5 and uncertainty around the outcome estimates, using PSA. The impact of both prognostic model parameter uncertainty and outcome uncertainty on survival, QALY and cost estimates can then be investigated. A further source of prognostic model uncertainty occurs when a series of equally appropriate prognostic models can be applied to a cohort of patients.

Additionally, CELT non-transplant costs are also estimated from a three-part prognostic model and the parameters from each part of the prognostic model are also subject to parameter uncertainty. Therefore, the impact of parameter uncertainty, in the three-part cost model, on cost estimates should also be investigated. Thus, Chapter 7 goes on to combine prognostic model parameter (both survival models and cost models), outcome and model selection uncertainties with cohort uncertainties using PSA analysis.

CHAPTER 7

COLLECTIVELY MEASURING MULTIPLE SOURCES OF UNCERTAINTY RELATED TO THE USE OF A PROGNOSTIC MODEL INSTEAD OF AN OBSERVED NON-TRANSPLANT COHORT

7.1 INTRODUCTION

Prognostic models can be applied to estimate control group survival, and subsequently QALYs and costs, in the absence of information from an observed control group. However, it is inevitable that uncertainty in survival, QALY and cost estimates will exist when applying a prognostic model to a cohort of patients. In Chapters 5 and 6, simulation methods have been presented that can be applied in order to incorporate three sources of uncertainty in prognostic model estimates. Firstly, when published prognostic models are used to estimate control group survival, the prognostic model's parameters are themselves estimates and the uncertainty that exists therein should be incorporated in to survival, QALY and cost estimates. Secondly, when more than one prognostic model is available as a basis for estimating survival, uncertainty exists as to whether the selected model is the most appropriate choice. Finally, uncertainty occurs when estimating the individual patient outcomes in the control group.

This thesis has not yet examined the impact of multiple sources of prognostic model uncertainty on survival, QALY and cost estimates in the absence of observed information from a control group. Nor has this thesis incorporated parameter uncertainties from the three-part cost model applied to estimate non-transplant costs (Chapter 6, Section 6.5).

In health economics it has historically been recommended that analysts allow for uncertainty by selecting a base case scenario for the main analysis, which is based on the most plausible parameter values. These values should then be varied in a series of sensitivity analyses "to examine the robustness over a range of alternative values" [Drummond *et al,* 1997].

Only 20 of the 40 solid organ transplantation studies reviewed in Chapter 3 varied parameter values, all of which used deterministic one-way sensitivity analysis. In one-way analysis, parameter values are varied one at a time and the effect of varying the values on the overall analysis is examined. However, one-way sensitivity analysis does not allow for possible interdependences between variables, where varying one parameter has an impact on the values of other parameters.

One-way sensitivity analysis was used in the main CELT analysis to vary the choice of prognostic model for estimating non-transplant survival, QALYs and costs [Longworth *et al,* 2003]. However, intuitively, this does not appear to be the most appropriate technique to allow for prognostic model parameter uncertainty, or to allow for the uncertainty in the individual patient outcome estimates. Additionally, non-transplant cost and QALY estimates will be dependent upon individual patient outcome predictions and estimated survival in the absence of transplantation and a series of one-way sensitivity analyses would not allow for these interdependencies between parameter values.

It is inappropriate to use one-way or multi-way sensitivity analysis to examine prognostic model parameter uncertainty. Briggs recommends using PSA to allow for parameter uncertainty in economic evaluations [Briggs, 2001]. Further, NICE guidelines on economic evaluations in health technology assessments recommend PSA to allow for parameter uncertainties in model estimates [NICE, 2004]. PSA was first introduced in Chapter 6 as the preferred method for estimating individual patient outcomes and the uncertainty around outcome estimates.

PSA will be used in Chapter 7 to examine the impact of prognostic model uncertainties on estimates of non-transplant survival, QALYs and costs over a five-year study period, for a series of CELT study patients. An advantage of applying PSA to the CELT study will be to incorporate the interdependencies between survival and outcome estimates, between survival and QALY estimates and between survival and cost estimates. Whilst the PSA technique presented in this chapter is not original, its application to organ transplantation and to these specific sources of uncertainty is.

Section 7.2 summarises the three sources of prognostic model uncertainty mentioned above within the context of estimating non-transplant survival in the CELT study. Section 7.3 presents non-transplant survival, QALY and cost estimates prior to incorporating uncertainties from prognostic models. Following this, the estimated non-transplant survival, QALYs and costs over five years are presented for a cohort of 81 PBC patients using the European prognostic model to estimate survival. Section 7.4 examines the effects of prognostic model uncertainties using PSA on non-transplant survival, QALY and cost estimates for the PBC CELT cohort after applying the European prognostic model. Section 7.5 discusses techniques for combining a series of prognostic models within PSA, as a solution to incorporating uncertainty in the choice of prognostic models, when more than one model exists. The section then compares non-transplant survival, QALY and cost estimates for three PBC prognostic models (European, Royal Free and PBC Mayo models) after allowing for prognostic model uncertainties. The difference between four sets of results and the implications of the results for decision makers and the transplant community will be discussed in Section 7.6.

7.2 SOURCES OF PROGNOSTIC MODEL UNCERTAINTY IN THE CELT STUDY

In the CELT study, non-transplant survival was estimated by applying published Cox PH prognostic models to patient and clinical information from transplant patients; this information was collected at time of transplantation. Using the information at point of transplant, it was possible to estimate the transplant patients' survival had they not received a liver transplant. The three main sources of prognostic model uncertainty when estimating non-transplant survival in the absence of an observed control group are presented below.

7.2.1 Prognostic Model Parameter Uncertainty

Incorporating the standard errors of regression coefficients in to the estimate of non-transplant survival gives a more accurate estimate of model uncertainty (Chapter 5). Of the seven published prognostic models considered in the main CELT study, six provided the standard errors of the regression coefficients. Anand *et al* did not provide information on the standard errors of the regression coefficients for the Birmingham ALD model [Anand *et al,* 1987]. Prognostic model parameter uncertainty was not incorporated in to the published CELT analysis [Longworth *et al,* 2003].

In Chapter 5 it was shown that allowing for prognostic model parameter uncertainty increased the amount of uncertainty around survival estimates. Therefore, it is expected that allowing for parameter uncertainty will also increase the uncertainty around non-transplant QALY and cost estimates.

7.2.2 Selection Uncertainty: Choosing a Prognostic Model

A series of prognostic models existed for each of the three liver disease groups included in the main CELT analysis: ALD (2 models), PBC (3 models) and PSC (2 models). Chapter 5 set out a series of seven selection criteria, to be applied to the alternative prognostic models (Section 5.8). Criteria 1 to 6 were applied in the main CELT paper, which resulted in the selection of the ALD Beclere model, an average of the PBC European model and PBC Royal Free model and the PSC International model in the main analysis [Longworth *et al,* 2003]. As pointed out in Chapter 2, although the justification of model choice appeared reasonable, one can never be totally sure that the best model has been selected; therefore, the alternative models (ALD Birmingham, PBC Mayo, and PSC Mayo) were applied in one-way sensitivity analysis.

The 7th criterion states that a prognostic model should display the least amount of parameter uncertainty (in comparison with other models) and was first proposed in Chapter 5 as an additional selection criterion for a reliable prognostic model. A model displaying the least amount of uncertainty in the survival estimates should, consequently, show less uncertainty in the subsequent cost and QALY estimates. Additionally, Chapter 5 also recommended that researchers explore whether there is an explanation as to why a prognostic model displays more uncertainty than a comparable model. For example, the inclusion of non-significant clinically meaningful variables in a model is likely to increase model uncertainty in comparison with a model that excluded non-significant variables. Therefore, it was recommended that criterion 7

should not be included as an additional selection criterion and any explanation for the increase in model parameter uncertainty should be considered when choosing an appropriate model.

Chapter 5 discussed methods for selecting an appropriate model, rather than combining alternative models. Therefore, Section 7.7 revisits this problem to discuss whether it is possible to incorporate prognostic model choice uncertainty in a PSA.

7.2.3 Individual Patient Outcome Uncertainty

In the CELT study, prognostic models are applied to the CELT transplant cohort in order to estimate each patient's survival, QALY and costs in the absence of transplantation. Survival, QALY and cost gains from transplantation are estimated at the patient level, thus making it necessary to estimate individual patient outcomes over the fixed study period. Additionally, it was assumed that non-transplant costs would vary in the month prior to death and that these costs could be estimated individually for patients based on their age at listing and their type of liver disease.

In Chapter 6, a PSA method was introduced for estimating individual patient outcomes and the uncertainty around the outcome estimates, this method was also omitted in the original CELT study. Allowing for individual outcome uncertainty resulted in an increase in the uncertainty around survival estimates. Chapter 6 showed that allowing solely for individual patient outcome uncertainty increased the uncertainty around non-transplant QALY and cost estimates.

7.2.4 Summary

All three sources of prognostic model uncertainty mentioned above have an impact on survival, QALY and cost estimates. For example, the choice of prognostic model could influence any decisions made regarding the treatment of end-stage liver disease. In addition to this, it is important to allow for any interdependencies between survival, QALY and cost estimates and between estimated survival and survival outcomes, highlighting the importance of incorporating all prognostic model uncertainties within a PSA. Further, the three-part model for predicting non-transplant costs is subject to model parameter uncertainty, and allowing for this will further increase the uncertainty around cost estimates. Thus, cost model parameter uncertainty should also be incorporated within the PSA.

7.3 ESTIMATING NON-TRANSPLANT SURVIVAL, QALY AND COSTS FOR THE PBC CELT COHORT – IGNORING PROGNOSTIC MODEL UNCERTAINTIES

As in Chapters 5 and 6, a cohort of 81 patients with end-stage PBC who received a liver transplant will be used to illustrate the impact of prognostic model uncertainty on non-transplant survival, QALY and cost estimates over a five-year study period. Section 7.3 presents details of the non-transplant survival, QALY and cost estimates prior to incorporating prognostic model uncertainties.

7.3.1 Prognostic Models Considered for the PBC CELT Cohort

Three alternative prognostic models were available for estimating non-transplant survival over five years in PBC patients; the European model, the Royal Free model, and the PBC Mayo model. These models have been introduced previously in Chapter 5, and Table 5.4 summarised the patient and clinical characteristics needed to predict survival in the absence of transplantation for each of the three prognostic models. All three models use patient age, bilirubin levels, albumin and the presence or absence of ascities to predict survival. However, the European model uses additional information on the presence or absence of gastrointestinal bleeding and the PBC Mayo model uses additional information on gender, prothrombin time and oedema scores.

7.3.2 Estimating Non-Transplant Outcomes (Survival, QALYs and Costs) at the Patient Level

Once a suitable prognostic model has been identified, the predicted survival time, during the study period, is estimated for each individual patient by plotting the patient specific survival probabilities over time and calculating the area under the curve. Non-transplant QALYs are estimated by multiplying the expected survival time by the last observed pre-transplant EQ-5D score.

Individual patient average daily costs are estimated from cost data collected on the CELT patients prior to transplant and predicted for the same patients in the absence of transplantation using a three-part model (Chapter 6, Section 6.5). The total non-transplant costs for patients who died were calculated by multiplying their estimated survival in days (minus 30 days)[1], by their estimated daily cost on the waiting list and adding this to their expected cost in the month prior to death. For patients who survived

[1] An adjustment was made to the expected non-transplant costs of patients predicted to die for the 30 day period prior to death.

the full five-year study period their expected daily cost was multiplied by 1,826 days (five years).

7.3.3 Estimating the Overall Mean Non-Transplant Outcomes (Survival, QALYs and Costs) with no Allowance for Model Uncertainty (Step 1)

In order to estimate non-transplant survival, QALYs and costs we need to select a prognostic model that can be used as a base case. This model should be the most appropriate to apply to the PBC CELT cohort and will be selected using the six selection criteria set out in Chapter 5, Section 5.8 (Box 5.2). The PBC Mayo model is not based on UK patients (Criterion 6) and is thus eliminated from the selection process.

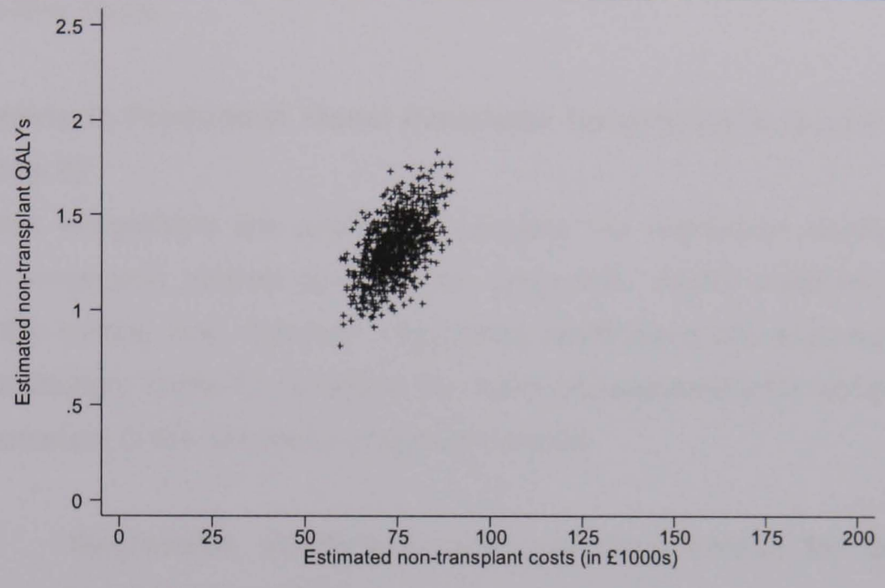
After the exclusion of the PBC Mayo model two prognostic models remained: the Royal Free model and the European model. The European model met all six selection criteria, whereas the Royal Free model had not been validated on internally and external data sets (Chapter 5: Table 5.10). Therefore, in this chapter the European model is selected as the base case in preference to the Royal Free model because it meets all six selection criteria.

The European prognostic model is applied to the CELT cohort to estimate non-transplant survival over five years. The mean probability of survival to five years in the absence of transplantation for the CELT cohort as a whole was 0.333. Thus, the expected number of non-transplant survivors at five years is 0.333x81 = 27 patients. Each patient's expected survival is then calculated from the area under the predicted survival curve. The survival times are extended to five years for the 27 patients with the longest survival.

The expected non-transplant survival, QALYs and costs are presented in Figure 7.1, with 95% bootstrapped CIs to represent cohort uncertainty. One thousand bootstrap replicates were sufficient for convergence of non-transplant survival, QALY and cost to the mean estimate. On average, patients are expected to survive 2.63 years in the absence of transplantation. After adjusting for HRQL, the quality adjusted survival decreases to approximately 16 months (1.34 years). The mean predicted costs over five years in the absence of transplantation are £74,420. Figure 7.1 also presents the mean QALY and cost estimates for each of the 1,000 bootstrap replicates.

The non-transplant survival, QALYs and cost estimates will form the base case scenario, prior to adjusting for prognostic model uncertainties.

Figure 7.1 Mean predicted non-transplant survival, QALYs and costs over five years for 81 PBC patients: applying the European prognostic model – Incorporating cohort uncertainty



Survival in years (95% CI)
2.63 (2.22 to 3.05)
QALYs (95% CI)
1.34 (1.06 to 1.65)
Costs (95% CI)
£74,420 (£63,872 to £85,067)

7.4 COMBINING PROGNOSTIC MODEL UNCERTAINTIES FOR THE EUROPEAN MODEL USING PSA – PARAMETER AND PATIENT OUTCOME UNCERTAINTY

This thesis has now presented methods to estimate individual patient survival on the basis of prognostic models, and this, together with the approach set out in Section 7.3.2 allows the estimation of total patient costs and QALYs. The PSA approach provides a framework within which prognostic model parameter uncertainty and individual outcome uncertainty can be incorporated in to survival estimates. This chapter will now proceed to incorporate this uncertainty, thus enabling the comparison of non-transplant survival, QALY and cost results with and without the inclusion of prognostic model uncertainties.

The European prognostic model has been selected as the base case model for estimating non-transplant survival, QALYs and costs over five years for a cohort of PBC patients (Step 1). Figure 7.1 presented mean non-transplant survival, QALY and cost estimates over five years with 95% bootstrapped CI showing cohort uncertainty prior to incorporating prognostic model uncertainties. In this section, prognostic model uncertainties will be incorporated in to survival, QALY and cost estimates over a series of three further steps.

7.4.1 Adding in Prognostic Model Parameter Uncertainty to Cohort Uncertainty (Step 2)

Monte Carlo simulations are applied to simulate the regression coefficients for the European prognostic models to allow for prognostic model parameter uncertainty. Through the central limit theorem, regression coefficients are assumed to follow a normal distribution; Table 7.1 presents the mean and standard error estimates for each of the parameters in the European prognostic model.

Table 7.1 Regression coefficients (and standard errors) for the European prognostic model

	European Model*
Age in years	0.04 (0.011)
Log_{10}(Bilirubin)	2.53 (0.260)
Albumin	-0.09 (0.019)
Ascities present	1.39 (0.210)
Gastrointestinal bleeding	0.65 (0.210)

* Variables should be standardised before coefficients are applied (Age in years – 55; log_{10} bilirubin – 1.53; Albumin – 34.3)

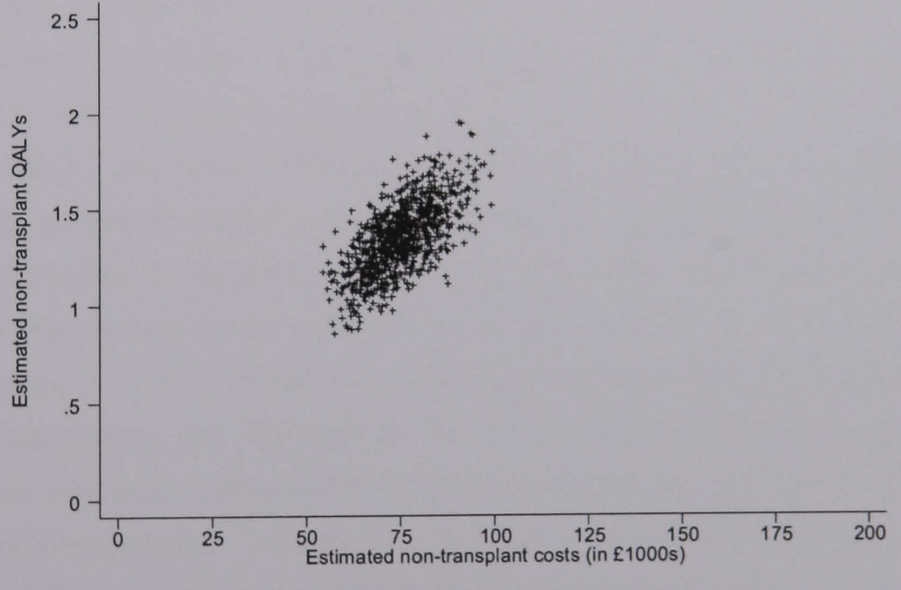
One thousand Monte Carlo simulations are run to allow for prognostic model parameter uncertainty[2] and for each of the 1,000 simulation runs 1,000 bootstrap replicates are run to account for cohort uncertainty. This resulted in a total of one million simulations. After each simulation run, the cohort mean non-transplant survival, QALYs and costs are estimated. The 50th percentile of the 1,000,000 simulated results represents the

[2] The number of simulations for parameter uncertainty was reduced from 3,000 (Chapter 5) owing to computational constraints. For each simulation run 1,000 bootstrap replicates were run to allow for cohort uncertainty. Figure 5.6 (Chapter 5) shows that the majority of parameters converge to the mean prior to 1,000 runs.

mean survival time, QALYs or costs across all simulations and the 2.5th and 97.5th percentiles represent the 95% confidence limits.

Figure 7.2 presents non-transplant survival, QALY and cost estimates with 95% CI over five years, after incorporating both prognostic model parameter uncertainty and cohort uncertainty. The mean survival, QALY and cost estimates are approximately the same as those presented in Figure 7.1 – when prognostic model parameter uncertainty is ignored. Mean survival estimates differ by approximately one week, QALY estimates are identical and cost estimates differ by approximately £600. Mean survival, QALY and cost estimates would tend further towards the mean estimates presented in Figure 7.1 if the number of simulations were increased.

Figure 7.2 Mean predicted non-transplant survival, QALYs and costs over five years for 81 PBC patients: applying the European prognostic model – Incorporating cohort and prognostic model parameter uncertainty*



Survival in years (95% CI)
2.65 (2.13 to 3.18)
QALYs (95% CI)
1.35 (1.03 to 1.71)
Costs (95% CI)
£75,019 (£59,677 to £91,649)

* Figure shows results from 1,000 randomly selected simulations only

As expected, incorporating prognostic model parameter uncertainty in to non-transplant survival, QALY and cost estimates increased the uncertainty around mean survival, QALY and costs, shown by the increase in the width of the CI between Figure 7.2 and Figure 7.1 where only cohort uncertainty was allowed for.

Figure 7.2 also presents a plot of the mean non-transplant QALY and cost estimates[3]. The figure shows that mean non-transplant costs increase linearly with increasing QALYs.

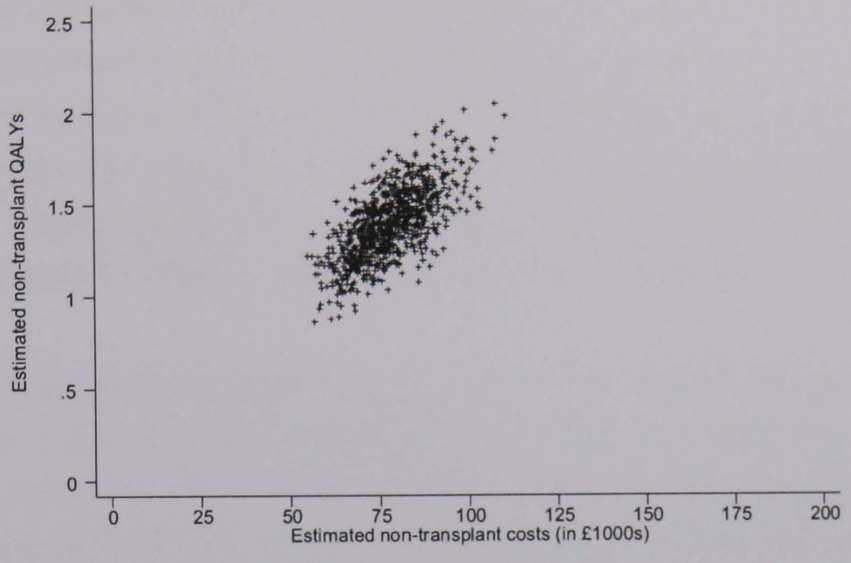
7.4.2 Adding in Individual Patient Outcome Uncertainty to Cohort Uncertainty and Prognostic Model Parameter Uncertainty (Step 3)

The third step expands on Step 2 and incorporates cohort uncertainty, prognostic model parameter uncertainty and individual patient outcome prediction uncertainty. Individual patient outcomes are assumed to follow binomial distributions with expected survival probabilities at five years ranging between zero and one.

As with Step 2, one million simulations are run to account for cohort uncertainty, prognostic model parameter and individual patient outcome uncertainty and the results (Figure 7.3) are compared with those from Steps 1 and 2. Mean non-transplant survival, QALY and cost estimates are slightly higher after incorporating cohort uncertainty, prognostic model parameter uncertainty and outcome uncertainty in to the prognostic models, in comparison with mean estimates after adjusting for prognostic cohort uncertainty and model parameter uncertainty (Figure 7.2) or cohort uncertainty alone (Figure 7.1). As explained in Chapter 6, by incorporating individual patient outcome uncertainty, the 27 patients with the longest survival times are no longer selected as survivors, rather individual patient survival probabilities at five years are used to predict survival. This random selection process will result in patients with shorter predicted survival times selected as surviving to five years and this will result in an increased mean survival for the cohort (Chapter 6 Section 6.4.1). The increased survival has an impact on the mean costs and QALY estimates, which also increase because patients are surviving slightly longer.

[3] Figures 7.2 to 7.10 presents mean QALY and cost estimates for a random selection of 1,000 simulations of the 1,000,000 simulations run (one estimate is randomly selected from each of the simulation runs that allow for parameter uncertainty. One million points were not plotted owing to computational constraints).

Figure 7.3 Mean predicted non-transplant survival, QALYs and costs over five years for 81 PBC patients: applying the European prognostic model – Incorporating cohort, prognostic model parameter and outcome uncertainty*



Survival in years (95% CI)
2.73 (2.19 to 3.30)
QALYs (95% CI)
1.40 (1.06 to 1.77)
Costs in £ (95% CI)
£77,833 (£61,368 to £96,982)

* Figure shows results from 1,000 randomly selected simulations only

Incorporating outcome uncertainty in to non-transplant survival, QALY and cost estimates increases the uncertainty around mean estimates slightly, in comparison with the uncertainty when incorporating cohort and prognostic model parameter uncertainty. After incorporating outcome uncertainty the CI widths for survival, QALY and costs are 1.19 years, 0.71 years and £35,614 in comparison with widths of 1.05 years, 0.68 years and £31,972 when incorporating cohort uncertainty and model parameter uncertainty.

7.4.3 Adding in Uncertainty in Non-Transplant Cost Estimates to Cohort Uncertainty, Prognostic Model Parameter Uncertainty and Individual Patient Outcome Uncertainty (Step 4)

It has already been stated that individual patient estimates of QALYs and costs depend on the predicted survival times. Furthermore, cost estimates depend upon outcome

predictions, where additional costs in the month prior to death are incorporated in to non-transplant cost estimates. Both death costs and daily costs on the waiting list are estimated at the patient level using regression models and, as with the prognostic models, these models are subject to parameter uncertainty. Therefore, the final step incorporates prognostic model uncertainties (cohort, parameter and outcome) and parameter uncertainties in the individual patient cost predictions. The impact of these uncertainties on the survival, cost and QALY predictions is examined and compared with the results from Step 1 to 3. In this final step the incorporation of cost model[4] uncertainty affects only the non-transplant cost estimates; survival and QALY estimates are independent of costs and remain unchanged from those in Step 3.

As with the prognostic model parameters, regression coefficients for cost estimates are assumed to follow a normal distribution. The means and standard errors for the regression coefficients for whether daily costs are incurred (yes/no), predicted daily costs (if incurred) and costs in the month prior to death (where the patient has been predicted to die) are presented in Table 7.2.

Table 7.2 Regression coefficients (and standard errors) for the three models used to estimate non-transplant costs

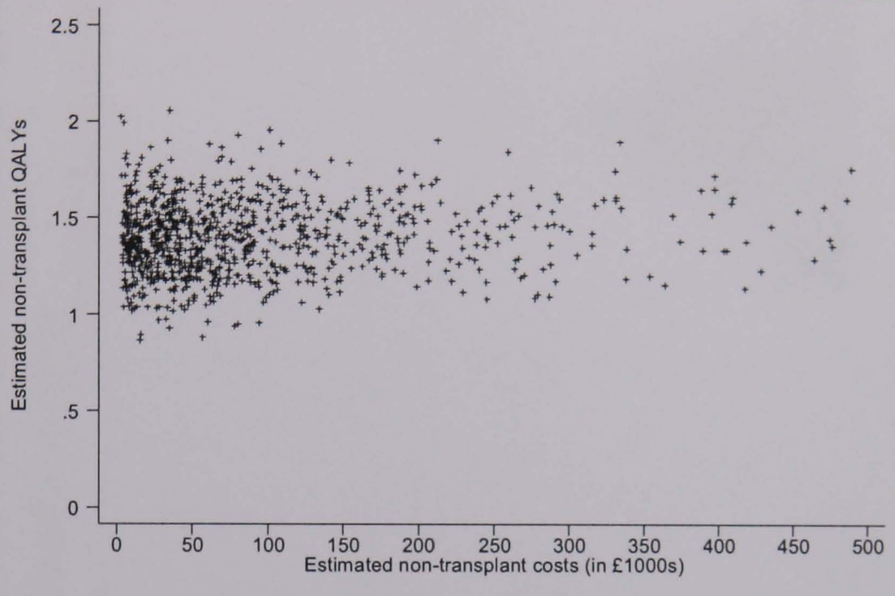
	Daily Cost Incurred (Y/N)	Log_e Daily Cost	Cost in the 30 days prior to death
Constant	-4.78 (3.32)	1.84 (0.66)	-1,676.4 (1,493.41)
Age	0.08 (0.05)	N/a	101.6 (18.65)
Ascities present	-1.51 (0.80)	0.68 (0.44)	N/a
Gender: female	N/a	1.31 (0.69)	N/a
Log_{10}(Bilirubin)	1.70 (0.85)	N/a	N/a

N/a – not applicable

Figure 7.4 presents mean non-transplant survival, QALY and cost estimates, with 95% CI, over five years for the PBC CELT cohort after adding cost model parameter uncertainties to cohort, prognostic model parameter and outcome uncertainties. As mentioned above, Figure 7.4 shows that the non-transplant survival and QALY estimates and their 95% CIs are the same as those presented in Figure 7.3 (Step 3) as the additional uncertainty allowed for only impacts on costs.

[4] The cost model applied in this chapter is the same model as described in Chapter 6 (Equations 6.1 to 6.6).

Figure 7.4 **Mean predicted non-transplant survival, QALYs and costs over five years for 81 PBC patients: applying the European prognostic model – Incorporating cohort, prognostic model parameter, outcome and cost model parameter uncertainty***



Survival in years (95% CI)
2.73 (2.19 to 3.30)
QALYs (95% CI)
1.40 (1.06 to 1.77)
Costs in £ (95% CI)
£58,696 (£5,509 to £484,257)

* Figure shows results from 1,000 randomly selected simulations only

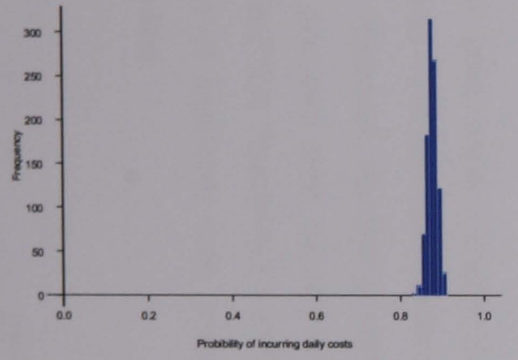
The first thing to note about the cost estimates is that, as with the survival, QALY and cost estimates after allowing for prognostic model parameter uncertainty, after allowing for cost model parameter uncertainty the CIs around cost estimates have increased substantially. The CIs around cost estimates are much wider than those presented for mean costs in Step 3 (CI width: £35,614), where the CI is now over £440,000 wider.

Figure 7.5 illustrates how the uncertainty in each part of the three part cost model has increased after allowing for model parameter uncertainty, uncertainty estimates are compared with those after allowing for cohort uncertainty alone, for each of the three parts of the prognostic model. The increase in the uncertainty around each of the three parts of the cost model has increased the overall uncertainty in the total non-transplant cost estimates (Table 7.3)

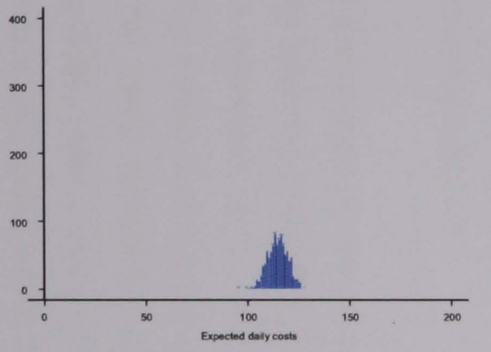
Figure 7.5 Cost distributions, incorporating cohort uncertainties (a to c) and predictive cost model parameter uncertainty (d to f)

Distributions after accounting for cohort uncertainties

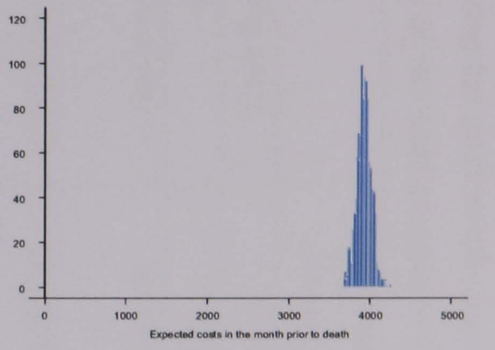
a) Probability of incurring daily costs

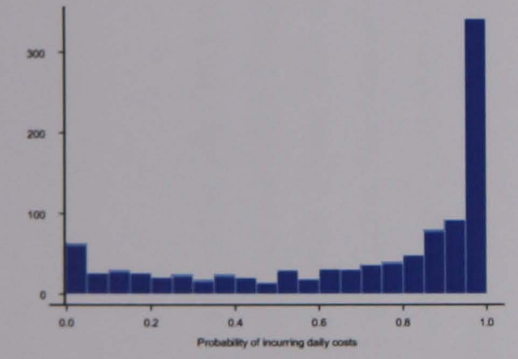


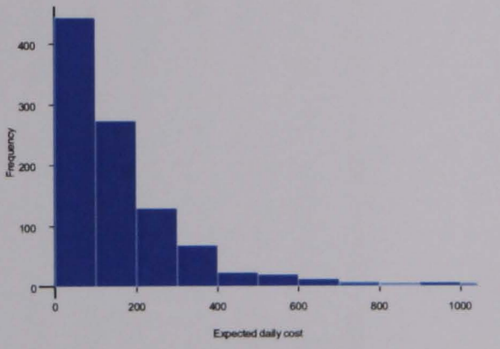
b) Expected daily cost

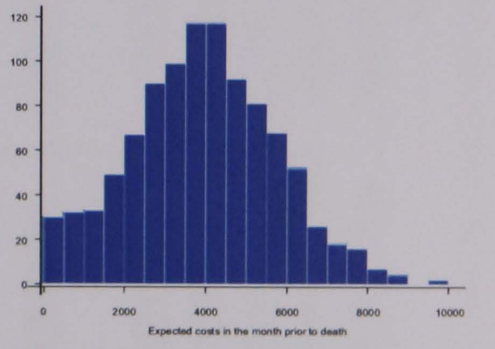


c) Costs in the month prior to death









d) Probability of incurring daily costs

e) Expected daily cost*

f) Costs in the month prior to death*

Distributions after accounting for model parameter uncertainties

* Scales for plots b and e, and c and f differ owing to computational constraints (histograms b and c "disappear" if scales are increased)

Table 7.3 95% bootstrapped CI for cohort uncertainty and 95% CI after allowing for cost model parameter uncertainty for each of the three parts of the cost prediction model

	95% bootstrap CI for cohort uncertainty	95% CI for cost model parameter uncertainty
Probability of incurring costs	0.85 to 0.90	0.01 to 1.00
Expected daily costs	£102 to £124	£15 to £784
Expected costs in the month prior to death	£3,766 to £4,105	£294 to £7,575
Expected total costs over five years	£63,872 to £85,067	£5,509 to £484,257

The second thing to note is that the mean non-transplant cost estimates for each prognostic model are approximately £20K less than the estimates when cost model uncertainty remained unaccounted for.

The main cause of the change in mean non-transplant costs is the decrease in the probability of patients incurring a daily cost. The probability of incurring a daily cost on the transplant list is estimated using a binary logistic regression model, where the probability of incurring a daily cost ranges between zero and one. Prior to allowing for cost uncertainties the mean probability of incurring a daily cost was 0.88 (95% bootstrap CI: 0.85 to 0.90). After incorporating cost model parameter uncertainties the probability of incurring a daily cost ranged between zero and one across simulations, with a mean estimate of 0.82. Figures 7.5a and 7.5d illustrate how the probability distribution of incurring costs changes after allowing for cost model parameter uncertainty.

The expected mean daily costs (Chapter 6, Equation 6.3) prior to allowing for cost model uncertainty was £114 and was similar to the mean costs (£112) after uncertainty was accounted for. However, like the probability of incurring costs, distributions were markedly different pre and post adjustment for cost model uncertainty (Pre 95% bootstrap CI: £102 to £124; Post 95% CI: £15 to £784). Figures 7.5b and 7.5e illustrate how the distribution of expected daily costs changes after allowing for cost model parameter uncertainties.

Prior to allowing for prognostic model parameter uncertainty the mean daily cost estimates, derived by multiplying the probability of incurring a cost by the estimated daily cost, was £98 (95% bootstrap CI: £89 to £106). After allowing for cost model parameter uncertainty the average probability of incurring a cost has decreased and the expected daily cost has remained approximately the same. Thus, the mean daily cost (the probability of incurring a cost multiplied by the estimated daily cost) decreases to £59 (95% CI: £0 to £637). This results in the mean daily cost estimate being lower, by approximately £40 per day, after accounting for cost model uncertainty, and this in turn reduces the mean total non-transplant costs over five years[5].

7.4.4 Summary

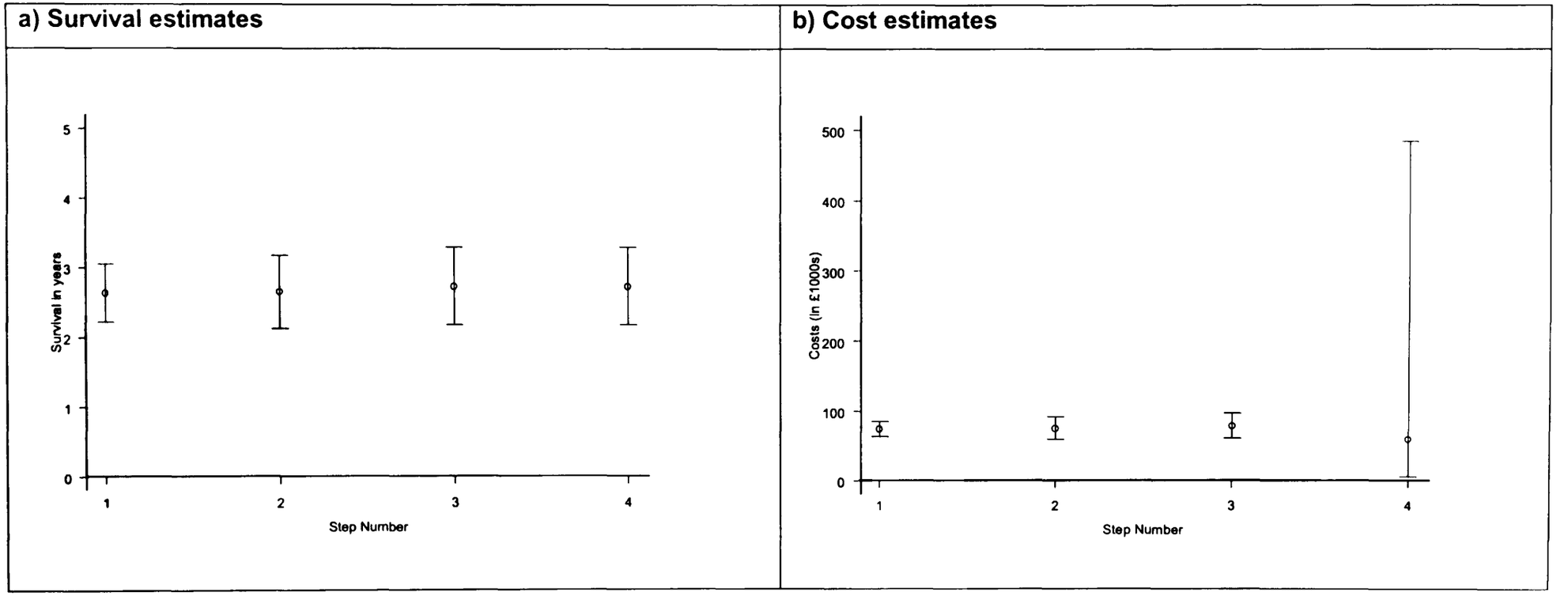
Figure 7.6 illustrates the effects of accounting for cohort uncertainty, prognostic model parameter uncertainty, outcome uncertainty and cost model parameter uncertainty (Steps 1 to 4) on mean survival and cost estimates[6]. The incorporation of prognostic model parameter and outcome uncertainty increases the uncertainty around non-transplant survival, QALY and cost estimates. Additionally, the incorporation of cost model parameter uncertainty markedly increases the uncertainty around non-transplant cost estimates, because the method used to estimate costs in the absence of transplantation proved to be subject to a large amount of uncertainty.

The three-part model that was chosen to predict non-transplant costs was the best fitting model for the PBC CELT data, yet the model failed to accurately predict non-transplant costs. All three parts of the model were derived from CELT patients and the variables selected for inclusion in the model were statistically significant predictors of costs in the CELT cohort ($p \leq 0.05$). Had we used an external model (had one existed) or included non-significant variables in the model, then the uncertainty around the non-transplant cost estimates could have been even greater than that presented in Figure 7.4.

[5] For example, a patient surviving for 365 days will have estimated total costs of £35,770, assuming costs are incurred at a constant rate of £98 a day (excluding death costs) when cost model uncertainty is not accounted for and a total cost of £21,535, assuming costs are incurred at a constant rate of £59 a day (excluding death costs) when cost model uncertainty is accounted for.

[6] QALY estimates are not shown; the only uncertainty that impacted on the QALY estimates was that from the prognostic models. Therefore, the pattern of uncertainty is exactly the same pattern as for survival estimates.

Figure 7.6 Mean non-transplant a) survival and b) cost estimates with 95% CIs for the four stages of prognostic model uncertainty for 81 PBC CELT patients using the European model



Decision makers should be presented with non-transplant estimates that incorporate model parameter uncertainty (prognostic model and cost model) and outcome estimation uncertainty in addition to cohort uncertainty, in order to make an informed decision regarding non-transplant cost and effectiveness estimates.

7.5 SELECTION UNCERTAINTY: CHOOSING A PROGNOSTIC MODEL

Thus far, this chapter has examined the effects of patient cohort uncertainty, prognostic model parameter uncertainty, prognostic model outcome uncertainty and cost model parameter uncertainty on non-transplant survival, QALY and costs, for one of three prognostic models. This section considers methods for allowing for uncertainties in the choice of prognostic models within PSA.

The most obvious way of combining the results from a series of prognostic models is to combine the original data sets for the three models and refit a prognostic model based upon the combined data [Sutton *et al,* 1998]. The original data were sought for the three prognostic models, though were only provided for the PBC Mayo model, therefore this approach could not be used.

Another alternative would be to apply the Bayesian methods referred to in Chapter 5 for dealing with model selection/structural uncertainty [Draper, 1995, Kang, 2000], which is the uncertainty that arises when selecting the mathematical structure of the model. These methods involve running "repeated analysis utilising different models and specifying prior probabilities of different models across this model space" [Briggs, 2000]. These methods also require the original data and therefore, can not be used here to account for model choice uncertainty.

In the absence of the original data, a meta-analytical approach, pooling the estimates from the regression coefficients and baseline survival/hazards to obtain an "averaged" prognostic model, might be considered. However, this is an inappropriate method to use because the choice of mathematical structure of the variables may differ between models. For example, the European model's variables are standardised prior to model fit and the PBC model and Royal Free model are not standardised. Further, the "averaged" coefficients and baseline survival/hazard estimates do not themselves produce meaningful survival estimates. Each individual prognostic model may contain a different combination of variables, as is the case for the three PBC models. Variables that are common to the prognostic models are not necessarily collected using the same

units, for example bilirubin is collected in mg/dl for the PBC Mayo model and μmol/L for the European and Royal Free data sets. It is also important to recognise that some models may transform variables on to another scale, for example, a $\log_e$ scale, whereas other models may transform variables on to a $\log_{10}$ scale or not transform the data at all.

Therefore, rather than combining the results from the three models within a PSA, it is felt more appropriate to present separate estimates of non-transplant survival, QALYs and costs for each of the three prognostic models. Therefore, this section presents survival, QALY and cost estimates over five years for the PBC Mayo and Royal Free models after adjusting for cohort uncertainty, (prognostic and cost) model parameter uncertainty and prognostic model outcome uncertainty (Steps 1 to 4). The results from the European prognostic model are repeated in this section to aid the comparability of the uncertainty in non-transplant survival, QALY and cost estimates across the three PBC prognostic models.

Figures 7.7 to 7.10 present the mean survival, QALY and cost estimates over five years for 81 PBC CELT patients, for the PBC Mayo and Royal Free models, after accounting for prognostic model and cost model uncertainties (Steps 1 to 4). The first thing to note is that the pattern of increasing uncertainty is the same across all three prognostic models, where CI for survival, QALY and cost estimates:

- increase noticeably after allowing for Cox PH model parameter uncertainty (Step 2) and
- increase marginally after the additional incorporation of outcome uncertainty (Step 3) and
- cost estimates increase substantially after the additional incorporation of cost model parameter uncertainty (Step 4)

Uncertainty in greatest when the PBC Mayo model is used to estimate non-transplant survival and is least when the European model is used. Although the PBC Mayo model displays more uncertainty than other models, this does not mean that it is a poor model: all three PBC prognostic models give similar mean non-transplant survival estimates over five years.

Figure 7.7 **Mean predicted non-transplant survival, QALYs and costs over five years for 81 PBC patients: applying a) the European b) the PBC Mayo model and c) the Royal Free model – Incorporating cohort uncertainty**

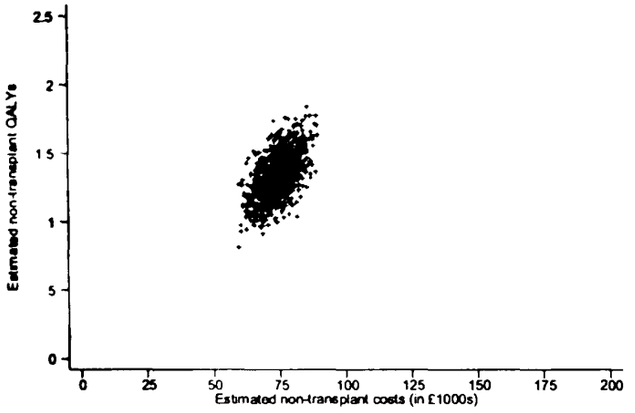
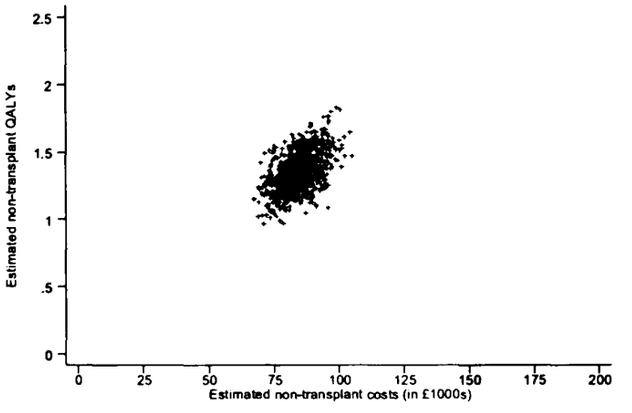
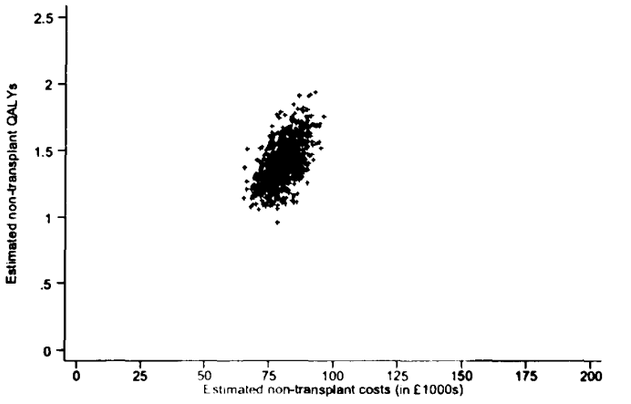
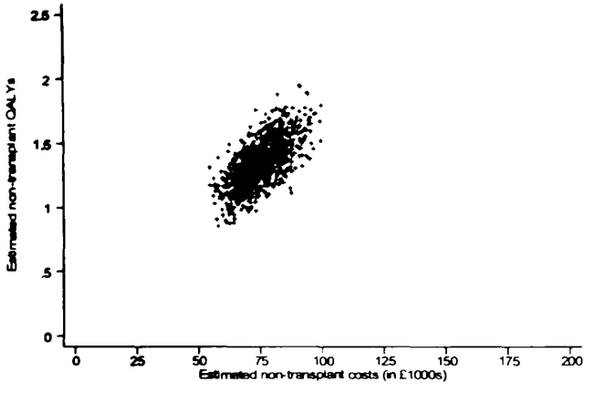
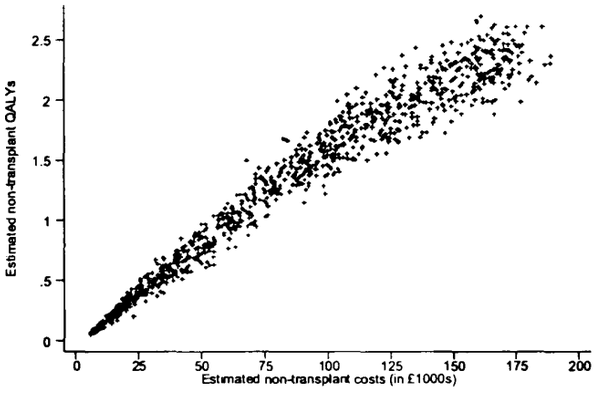
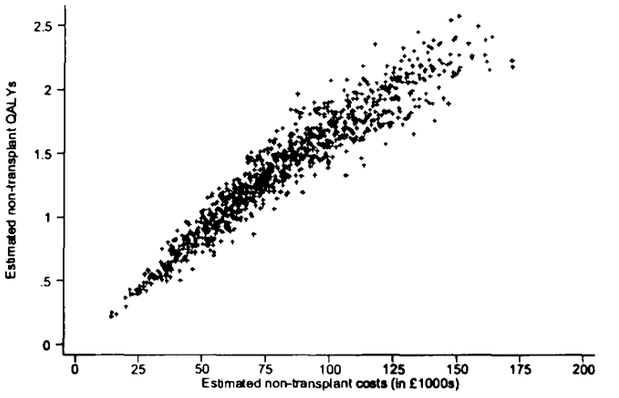
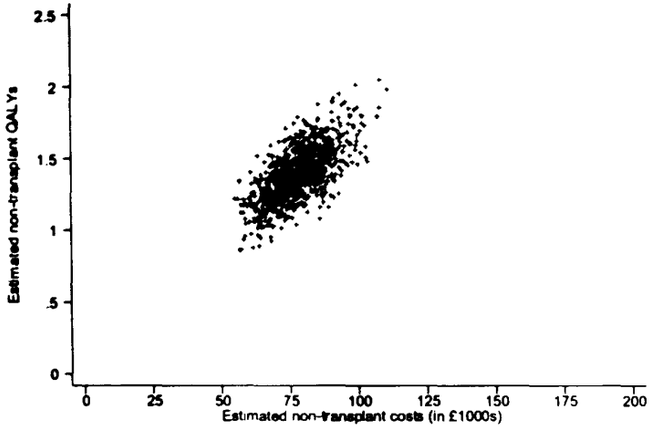
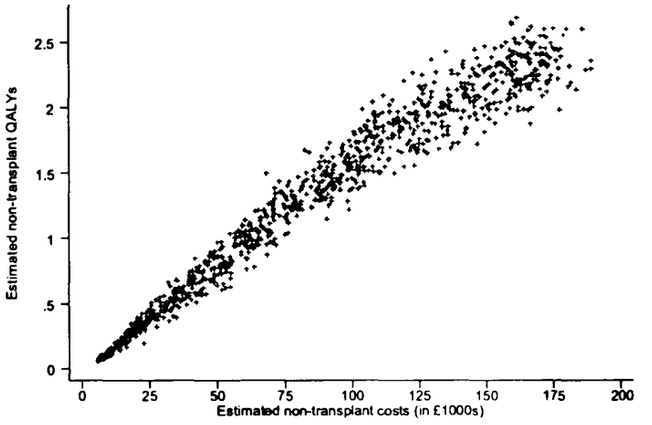
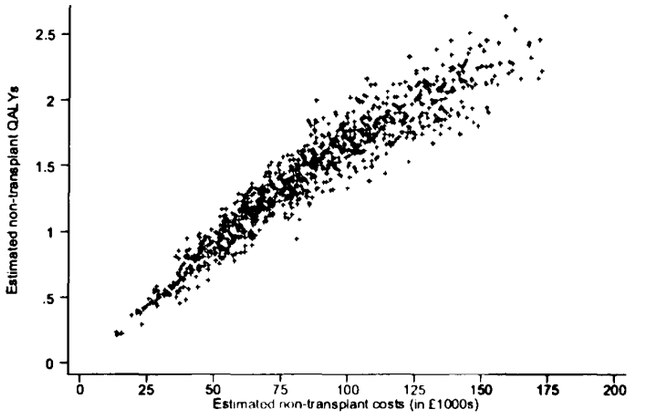
a) European Model	b) PBC Mayo Model	b) Royal Free Model
		
Survival in years (95% CI) 2.63 (2.22 to 3.05)	**Survival in years (95% CI)** 2.60 (2.24 to 3.05)	**Survival in years (95% CI)** 2.74 (2.34 to 3.13)
QALYs (95% CI) 1.34 (1.06 to 1.65)	**QALYs (95% CI)** 1.29 (1.04 to 1.60)	**QALYs (95% CI)** 1.41 (1.13 to 1.71)
Costs (95% CI) £74,420 (£63,872 to £85,067)	**Costs in £ (95% CI)** £78,849 (£67,277 to £90,718)	**Costs in £ (95% CI)** £80,107 (£69,831 to £90,387)

Figure 7.8 **Mean predicted non-transplant survival, QALYs and costs over five years for 81 PBC patients: applying a) the European b) the PBC Mayo model and c) the Royal Free model – Incorporating cohort and prognostic model parameter uncertainty***

a) European Model	b) PBC Mayo Model	c) Royal Free Model
		
Survival in years (95% CI)	**Survival in years (95% CI)**	**Survival in years (95% CI)**
2.65 (2.13 to 3.18)	2.61 (0.25 to 4.91)	2.74 (1.08 to 4.32)
QALYs (95% CI)	**QALYs (95% CI)**	**QALYs (95% CI)**
1.35 (1.03 to 1.71)	1.29 (0.10 to 2.50)	1.40 (0.52 to 2.23)
Costs (95% CI)	**Costs in £ (95% CI)**	**Costs in £ (95% CI)**
£75,019 (£59,677 to £91,649)	£79,271 (£8,385 to £176,348)	£80,423 (£29,550 to £144,550)

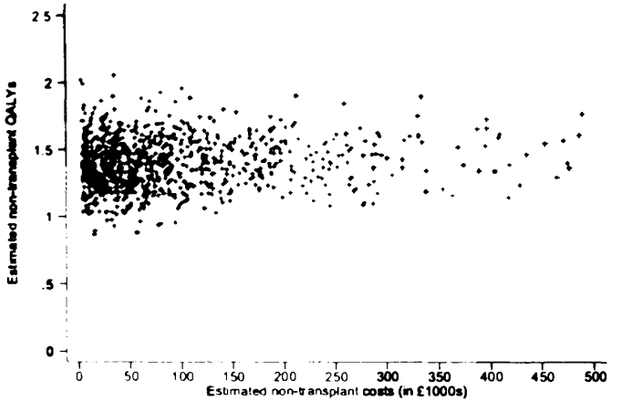
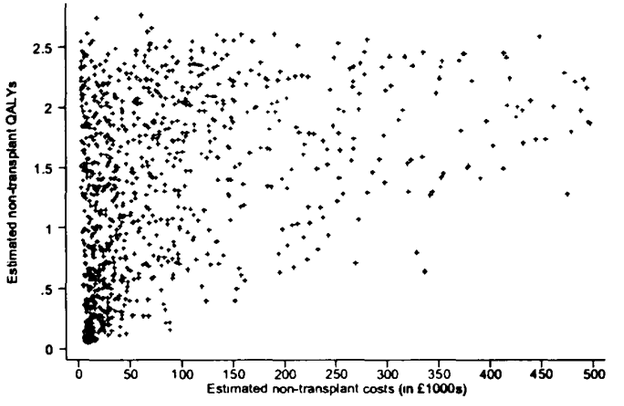
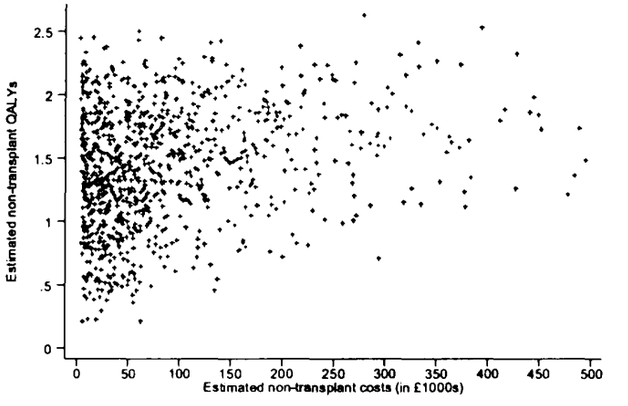
* Figure shows results from 1,000 randomly selected simulations only

Figure 7.9 **Mean predicted non-transplant survival, QALYs and costs over five years for 81 PBC patients: applying a) the European b) the PBC Mayo model and c) the Royal Free model – Incorporating cohort and prognostic model parameter and outcome uncertainty***

a) European Model	b) PBC Mayo Model	c) Royal Free Model
		
Survival in years (95% CI) 2.73 (2.19 to 3.30)	**Survival in years (95% CI)** 2.72 (0.23 to 4.94)	**Survival in years (95% CI)** 2.84 (1.10 to 4.42)
QALYs (95% CI) 1.40 (1.06 to 1.77)	**QALYs (95% CI)** 1.36 (0.09 to 2.52)	**QALYs (95% CI)** 1.45 (0.52 to 2.28)
Costs in £ (95% CI) £77,833 (£61,368 to £96,982)	**Costs in £ (95% CI)** £83,836 (£78,161 to £178,730)	**Costs in £ (95% CI)** £84,337 (£30,104 to £150,919)

* Figure shows results from 1,000 randomly selected simulations only

Figure 7.10 **Mean predicted non-transplant survival, QALYs and costs over five years for 81 PBC patients: applying a) the European b) the PBC Mayo model and c) the Royal Free model – Incorporating cohort and prognostic model parameter and outcome uncertainty and cost model parameter uncertainty***

a) European Model	b) PBC Mayo Model	c) Royal Free Model
		
Survival in years (95% CI)	**Survival in years (95% CI)**	**Survival in years (95% CI)**
2.73 (2.19 to 3.30)	2.72 (0.23 to 4.94)	2.84 (1.10 to 4.42)
QALYs (95% CI)	**QALYs (95% CI)**	**QALYs (95% CI)**
1.40 (1.06 to 1.77)	1.36 (0.09 to 2.52)	1.45 (0.52 to 2.28)
Costs in £ (95% CI)	**Costs in £ (95% CI)**	**Costs in £ (95% CI)**
£58,696 (£5,509 to £484,257)	£48,570 (£4,580 to £665,802)	£58,812 (£4,988 to £566,203)

* Figure shows results from 1,000 randomly selected simulations only

7.6 DISCUSSION

In Chapter 3 the difficulties in selecting an appropriate observed non-transplant control group with which to compare the effectiveness, costs or cost-effectiveness of solid organ transplantation were discussed. In the absence of observed data from a control group, the approach adopted by several transplant studies, including the CELT study, was to use a prognostic model to estimate survival, QALYs or costs in the absence of transplantation. Given that it is necessary to estimate non-transplant outcomes, it is natural that there will be some degree of uncertainty relating to these estimates in comparison with predictions from an observed cohort of non-transplant patients. This chapter has shown that it is possible to allow for this uncertainty using PSA.

7.6.1 Cost-Effectiveness Studies in Liver Transplantation: Choice of Control Group

A total of eight studies in the solid organ transplant literature review (including the main CELT study) attempted to measure the cost-effectiveness of liver transplantation. Three of these eight studies used a current comparator group. Burroughs *et al* used a cohort of patients with end-stage cirrhosis who were treated for complications at the same centre as the transplant patients [Burroughs *et al,* 1992]. The comparator cohort is likely to be of a different case mix to the transplant cohort and is likely to contain patients who are not eligible for transplantation. It is not clear how these potential differences between groups could affect the study results, as patient's clinical details are not supplied in the paper. Caution should also be applied to the interpretation of the paper by Manjo and colleagues, where it would be difficult to generalise the results, as the transplant cohort consists of only four patients [Manjo *et al,* 2000]. Finally, Williams *et al's* study is likely to exaggerate the effects of transplantation [Williams *et al,* 1987]. The transplant cohort is selective and contains patients who survive for at least six months post transplant, although their non-transplant cohort met listing criteria for transplantation, they consist only of patients who died prior to transplant. None of these three studies acknowledge the likely biases from their choice of study design.

The remaining five studies used modelling techniques to estimate the cost-effectiveness of liver transplantation [Bonsel *et al,* 1990; Sarasin *et al,* 1998; Farinati *et al,* 2001; Sagmeister *et al,* 2002; Longworth *et al,* 2003], and two of these five studies used prognostic models to estimate, what would have been, non-transplant survival [Bonsel *et al,* 1990; Longworth *et al,* 2003]. However, none of the five studies incorporated model parameter uncertainty in to cost-effectiveness estimates.

Thus, to date, decision makers have been unable to make a completely informed decision about the cost-effectiveness of liver transplantation, as the selection of the non-transplant control cohort is likely to give biased estimates or the decision makers have not been presented with full information on the uncertainty around cost-effectiveness estimates.

7.6.2 Reliability of Cost Estimates

Non-transplant cost estimates were heavily dependent upon the survival estimates from the prognostic model. Cost estimates were influenced by both survival length, as estimated from the prognostic model and estimated outcome, where costs increased in the month prior to death. Therefore, this chapter also explored the impact of both prognostic model uncertainties and non-transplant cost model parameter uncertainties, predicted from a three-part model.

The most striking result to be raised from this piece of research is the amount of uncertainty that should properly be attributed to the method for estimating non-transplant costs. In both the main CELT study and here it was assumed that non-transplant costs could be estimated based on the costs of patients on the waiting list. Non-transplant costs were estimated on a per patient basis and assumed to be constant over time until the month prior to death. In the main CELT analysis each patient's average daily cost on the waiting list was used to estimate non-transplant costs over time. However, this approach did not allow for uncertainty in the cost predictions, therefore the method for estimating non-transplant costs adapted here modelled costs more completely than the approach used in the main CELT analysis [Longworth *et al,* 2003].

The three-part modelling approach for estimating non-transplant costs adopted in this chapter is not necessarily the only model solution for estimating costs in the absence of transplantation and other approaches may produce more reliable estimates. For example, a model that incorporates more detailed information relating to the individual change in costs over time on the waiting list, may give more accurate predictions than the model used here, which uses the individual average daily cost on the waiting list. Alternative approaches to modelling daily costs could be explored by using a Bayesian approach to model structural uncertainty [Draper, 1995]. However, the cost models that have been used here are ones that best fit the CELT data, and the variables selected in the models are ones that were found to be significant predictors of costs for the

CELT cohort. Despite this the model estimates were subject to a large amount of model parameter uncertainty. Decision makers should be presented with mean non-transplant cost estimates with 95% CI that include the parameter uncertainty around the cost model in order to highlight how poor the estimates of non-transplant costs are.

The final area of concern relating to cost estimates was how allowing for cost model parameter uncertainty changed the mean non-transplant cost estimates. It is generally assumed that mean estimates remain constant and the degree of uncertainty around them varies. Therefore, decision makers make policy implications based on mean estimates and the information on uncertainty is used to make decisions about whether further research is needed (value of information analysis) [Claxton, 1999]. Given that the best estimate of a statistic (e.g. mean cost) is the observed mean cost from the original data, it can be seen that the best estimate of the mean cost estimated from a prognostic model should be derived using the mean regression coefficients. Therefore, it is recommended that decision makers make choices about the clinical or cost-effectiveness of treatments or technologies using a base case scenario based upon mean survival, QALY and cost estimates prior to allowing for any uncertainties. Mean survival, QALY, cost or cost-effectiveness estimates that incorporate uncertainties should then be presented separately and these can be used to inform decisions regarding whether further research should be undertaken.

7.6.3 Cost-Effectiveness of Liver Transplantation

The aim of Chapters 5 to 7 was to propose methods for estimating uncertainty around prognostic model estimates and to demonstrate the effects of this uncertainty on survival, QALY and cost estimates. Therefore, the impact of prognostic model uncertainties on the cost-effectiveness of liver transplantation in the UK was not examined.

Had one of the objectives of this thesis been to estimate the cost effectiveness of liver-transplantation over a five-year period after incorporating prognostic model uncertainties it would have been necessary to make assumptions about transplant resource use and HRQL. Survival information post transplant, was routinely collected and available at each of the six centres participating in the CELT study, and was obtained from each centre for the extended five-year study period. However, resources were unavailable to the CELT study for obtaining information on HRQL and resource use information from 2.25 years post assessment to five years post assessment.

Hence, estimation techniques are needed in order to estimate transplant HRQL and costs for the extended study period, drawing upon information from the CELT study.

A full PSA analysis would incorporate all uncertainties in to the estimation of the cost-effectiveness of liver transplantation in the UK over a five-year period, but was beyond the scope and focus of this thesis.

7.6.4 Incorporating Other Sources of Uncertainty in to PSA

This chapter has focused on presenting methods for measuring the impact of prognostic model uncertainty on non-transplant survival, QALY and cost estimates in the absence of information from an observed control group. Uncertainty from prognostic models occurs in addition to standard sources of uncertainty, for example assigning distributions around unit costs or HRQL estimates, and this analysis can be extended further to incorporate these sources of uncertainty within a full PSA.

One source of prognostic model uncertainty we were unable to incorporate in to the PSA analysis was model choice uncertainty. In principle, it should be possible to combine the regression coefficients and baseline survival/hazard estimates in to one "average" prognostic model, providing that the original data are available for all three models. Insufficient information was available for all three PBC models to produce an average model. Therefore, a series of three deterministic one-way sensitivity analyses (scenario analysis) were presented, one for each prognostic model, which incorporated prognostic model uncertainties. It is recommended that the results for all three scenarios be presented to decision makers.

7.7 CONCLUSIONS

This is the first time that PSA has been used to describe uncertainty around prognostic model estimates and, more specifically, uncertainty in non-transplant survival, QALY and cost estimates in liver transplantation. Although mean survival, QALY and cost estimates were robust across three PBC prognostic models the uncertainty around these estimates varied by prognostic model.

Allowing for prognostic model uncertainty increased the amount of uncertainty around mean survival, QALY and cost estimates, in comparison to estimates where only cohort uncertainty is accounted for. Individual patient outcome uncertainty had only a small impact on non-transplant survival, QALY and cost estimates. In contrast parameter

uncertainty from the three-part cost model only impacted on non-transplant cost estimates and greatly increased the uncertainty around the mean cost estimates. Additionally, cost model parameter uncertainty changed the mean cost estimates by over £20K in comparison with mean cost estimates prior to accounting for cost model uncertainties. It is recommended that further research in to the repeatability of these results to other data sets and models be explored in further detail. Furthermore, decision makers should be presented with mean survival, QALY, cost or cost-effectiveness estimates prior to allowing for any uncertainties (base case scenario) and subsequently presented with estimates that allow for uncertainties such as model parameter uncertainties (and the uncertainty around them).

Intuitively, a prognostic model displaying the least amount of uncertainty might be an appealing choice of model, in comparison with models that display a greater amount of uncertainty. However, this thesis has already argued that a model displaying a large amount of parameter uncertainty will not necessarily give inaccurate survival estimates, with uncertainty being attributed to a number of factors including; the number of parameters in the model, the (justifiable) inclusion of non-significant variables, the type of model fitted (structural form), sample size and the number of events. The only "justification" for excluding one of the three PBC prognostic models, the PBC Mayo model, was that it was not based on a cohort of UK patients.

Decision makers should be presented with survival, QALY and cost estimates that account for the uncertainty that arises from using prognostic models, in the absence of information from an observed cohort. Additionally, when more than one prognostic model exists, scenario analysis should be undertaken and estimates provided for each of the prognostic models and each set of estimates should allow for prognostic model uncertainties.

This thesis has identified and explored two areas of uncertainty identified as issues important within an economic evaluation in liver transplantation and generalisable to other economic evaluations in health care (selecting an appropriate method for estimating mean total costs in the presence of censoring and measuring uncertainties from prognostic models). Chapter 8 will summarise the findings of this thesis and explore the wider implications of these findings within the transplant community, for researchers and decision makers.

CHAPTER 8

CONCLUSIONS

8.1 INTRODUCTION

When conducting an economic evaluation it is important to identify all possible sources of uncertainty (sampling variation, methodological uncertainty, extrapolation and generalisability, model parameter, model structure and model process) and present information about these uncertainties to decision makers. Providing full details of these uncertainties and their effects on the health care technology under evaluation will enable decision makers to reach an informed decision and make informed recommendations on the need for further research.

The aim of this thesis was to explore some methodological issues for measuring uncertainties in economic evaluations. The issues explored here were identified as important concerns in the potential extension of an economic evaluation in liver transplantation over a five-year study period. A review of the literature of solid-organ

transplant studies confirmed that the uncertainties identified in the CELT study were issues that were not addressed adequately in other solid organ transplant studies.

This thesis has focused on two separate issues of methodological, model parameter and model uncertainties, namely:

- the selection of an appropriate method for estimating mean total costs in the presence of censoring (methodological uncertainties)
- the measurement of uncertainty around survival, QALY and cost estimates derived from prognostic models in the absence of observed data (model parameter, model selection and methodological uncertainties)

This final chapter will summarise the main findings of this thesis and discuss the implications of these findings within solid-organ transplantation. This chapter will also discuss how the methods and guidance given in this thesis contribute to health services research and its value to decision makers.

8.2 ISSUES OF UNCERTAINTY IN SOLID-ORGAN TRANSPLANTATION STUDIES

Chapter 2 presented an overview of the results of an economic evaluation in liver transplantation in the UK over a 2.25 year study period (the CELT study) [Longworth *et al,* 2003]. The cost-effectiveness of liver transplantation was evaluated for three liver disease groups: ALD, PBC and PSC.

Survival, HRQL and resource usage were observed for the transplant cohort from point of assessment to 2.25 years post assessment. In the absence of data from an observed cohort of non-transplant patients with end-stage liver failure, prognostic models were used to estimate the transplant patients' non-transplant survival, had they not received a transplant.

The incremental QALY gain from liver transplantation over 2.25 years was approximately six months across all disease groups. However, the cost-effectiveness of liver transplantation in the UK differed by disease groups. The cost-effectiveness of liver transplantation was most likely to be acceptable in patients with PSC (Mean ICER: £21K), then PBC (Mean ICER: £29K) and least likely to be acceptable for ALD patients (Mean ICER: £48K).

Longworth *et al* suggested extending the analysis over a longer time frame so that the mid to long term benefits of liver transplantation could be measured. Therefore, a proposed extension to the CELT study time frame from 2.25 years to five years was considered. The proposed extended time frame raised two areas of concern: selecting an appropriate method for estimating mean total costs in the presence of censoring and measuring uncertainties around prognostic model estimates.

Solid organ transplantation is currently considered to be the treatment of choice for patients with end-stage organ failure, despite having never been the subject of an RCT. In the absence of this gold standard evaluation (a situation which seems likely to remain), researchers seeking to evaluate the effectiveness, costs or cost-effectiveness of solid organ transplantation need to be especially careful to select a comparator group for transplanted patients that would appear to give the most reliable approach and that considers representativeness, comparability and sample size in particular. Chapter 3 described alternatives to the RCT study design which could be applied to studies in solid organ transplantation in order to assess their effectiveness, costs or cost-effectiveness. The use of quasi-experimental control groups, intervention delay cohorts, historical cohorts, case-control studies, before and after studies, expert opinions and modelling techniques were considered as possible methods to create comparisons for evaluating the medical management of end-stage organ failure in comparison with transplantation. However, all of the comparator groups listed above were either unrepresentative of the general population, not comparable to the transplant cohort, or were unlikely to be of sufficiently large sample size to draw conclusions from.

Chapter 3 confirmed that the two uncertainty issues identified in the CELT study were not adequately addressed within other effectiveness, cost and cost-effectiveness studies in solid organ transplantation. Virtually all of the 158 studies included in the literature review in Chapter 3 failed to acknowledge any of the inherent problems relating to the choice of an appropriate non-transplant control group or to perform any form of sensitivity analysis to allow for any sources of uncertainty in the study. The literature review also confirmed the importance of providing guidance for estimating mean study costs, where only three out of 33 studies attempted to adjust for censoring when estimating mean study costs. The remaining 30 studies failed to give sufficient detail to establish whether censoring was an issue and it was assumed that it would be for a proportion of the thirty studies. Further, none of the studies that used prognostic

models to estimate non-transplant survival accounted for uncertainty in prognostic model estimates. These two concerns were not unique to the CELT study and were generalisable to other solid organ transplant studies.

8.3 GUIDANCE ON METHODS FOR ESTIMATING MEAN TOTAL COSTS IN THE PRESENCE OF CENSORING – METHODOLOGICAL UNCERTAINTY

Censored data can be a problem in all health care research studies, it occurs when a proportion of patients in a cohort do not have the event of interest (e.g. death) and have incomplete follow up data for the study period. Although well established methods exist to account for censored survival data it is inappropriate to directly apply these methods to censored cost or QALY estimates.

A review of the literature identified 12 methods that can be applied to censored cost data to estimate mean study costs in the presence of censoring: ignoring censoring, ignoring censored cases, Kaplan-Meier cost method, Cox PH cost method, the partitioned Cox cost method [Lipscomb *et al,* 1998], Lin's methods with either KCH or unknown cost history (UCH) [Lin *et al,* 1997], the weighted cost method with KCH or UCH [Bang & Tsiatis, 2000], Lin's regression method with KCH or UCH [Lin, 2000] and Carides' regression method [Carides *et al*, 2000]. To date no review of all existing methods for estimating mean total costs in the presence of censoring exists. Therefore, this thesis sought to review these methods and to offer guidance in the selection of appropriate methods for estimating mean study costs in the presence of censoring.

The 12 methods identified in the literature review were compared across four different censoring mechanisms – random censoring, end-of-study censoring, informative censoring and partial censoring – by simulating the censoring mechanisms from a complete cohort of patients included in the CELT study. Methods were also compared across differing levels of censoring ranging from light censoring (10% of data censored) to heavy censoring (80% censored). Finally, the accuracy of mean cost estimates was compared across different interval lengths for the six methods where the study time period can be divided in to smaller interval lengths (weighted cost method [KCH], Lin's regression method [KCH], Lin's method [KCH], partitioned Cox cost method, Carides' method and Lin's method [UCH]). The aim was to establish whether the choice of interval length affected the accuracy of the mean cost estimates for each of the six methods.

Chapter 4 included three methods that have serious shortcomings in estimating mean total costs in the presence of censoring: the Kaplan-Meier method (estimates were almost four times greater than the observed mean cost for the "complete" data set prior to censoring [£36,045]), Cox cost method (estimates approximately twice as high as the observed estimate) and partitioned Cox cost method (estimates at least £10K lower than the observed estimate). Yet anecdotal evidence suggests that these methods are still commonly used. As demonstrated here and elsewhere, application of the Kaplan-Meier method, Cox cost method or partitioned Cox cost method will result in serious bias when estimating mean costs affected by censoring. For this reason, it was felt important to investigate and address this issue in depth, in order to ascertain how well 12 proposed methods work in various circumstances.

Chapter 4 identified a further three methods that gave poor estimates of mean study costs when resources were not incurred in the final period of the study: Lin's method (UCH), Carides' method and the weighted cost method (UCH). It is recommended that these three methods should not be applied when costs are not incurred at the end of the study period of interest. This scenario is most likely to occur when economic evaluations are subject to partial censoring and is therefore more likely in economic evaluations designed to collect costs at specific time points during the study or when costs are collected at one time point during the study period. For example, a study may be designed to collect costs every six months, thus a patient dying in month eight will have a six-month cost but censored costs for month seven onwards or a study may be designed to collect cost data over a two year period and survival data over a four year period, therefore cost data are censored from the end of year two to the end of year four. The scenario is less likely to occur for treatment specific interventions, though potentially costs might not be incurred at the end of the study period when the majority of resource use is incurred at the beginning of the study period with few resources occurring later on in the study i.e. liver transplantation, surgical interventions and service interventions.

It was surprising that the method of ignoring censoring was one of the more accurate methods for estimating mean total costs. The technique of ignoring censoring produced fairly accurate estimates of mean total costs and standard errors when censoring was restricted to the end of the study period (end-of-study censoring restricted to the final six months of the study and partial censoring). The accuracy of this method could be due to the nature of the CELT dataset, where on average larger costs were incurred

earlier in the study period at the point of the transplant operation and stabilised in the later end of the study period. Assuming that these results are generalisable to other solid organ transplantation studies, the cost and cost effectiveness estimates presented by the 30 studies identified in Chapter 3 (that potentially ignored the issue of censoring) are likely to result in more accurate mean estimates than if they had chosen an inappropriate method for accounting for censoring (For example the Kaplan-Meier method).

Selecting the interval length that resulted in the best mean cost estimate and comparing the estimates across all 12 censoring methods showed that Lin's method (KCH) with small interval lengths gave the most accurate estimate of mean study costs across all censoring mechanisms and levels. Therefore, based upon the results of the CELT study, it is recommended that in the presence of censored cost data Lin's method with short interval lengths (KCH) is used to estimate mean study costs in the presence of censoring.

The majority of methods that gave reliable mean cost estimates were poor at estimating the standard error. In order to obtain accurate estimates of both the mean and standard error, the results of Chapter 4 suggest Lin's method (KCH) with small interval lengths and the weighted cost method (KCH) give the most accurate estimates of both. Lin's method (KCH) and the weighted cost method (KCH) performed well across all censoring mechanisms and levels.

8.3.1 Implications – Censoring Methodology

Bodies issuing guidance on the evaluation of health care technologies do not currently offer advice on methods for estimating mean total costs in the presence of censoring [See: Drummod & Jefferson, 1996; NICE, 2004; CADTH, 2006]. Based upon the results of one study it is recommended that Lin's method (KCH) with small interval lengths be used to estimate mean total costs in the presence of censoring. Lin's method (KCH) gave accurate estimates of both the mean study costs and their uncertainty.

The weighted cost method (KCH) also gave accurate estimates of the mean study costs and its standard error, but when dividing the study period up in to smaller intervals it is not obvious how to select the interval length that achieves the most accurate cost and standard error estimates. Comparing estimates across interval

lengths, the weighted cost method gave the most accurate estimates using interval lengths of six months for random and end-of-study censoring and three months for partial and informative censoring. However, for two monthly interval lengths there was a loss of information when estimating the Kaplan-Meier survival probabilities. With CELT survival measured in monthly units and choosing two monthly interval lengths an event can only occur at one of two time points (month one or month two). This choice of interval lengths thus results in less accurate estimates of survival over time. Therefore, it is suggested that further research is undertaken to establish whether it is possible to create a formula for selecting interval lengths for the weighted cost method (KCH) that result in the most accurate estimate of mean study costs.

Under random, end of study and informative censoring due to good health the accuracy of methods for estimating mean total costs in the presence of censoring (excluding naïve methods: Kaplan-Meier, Cox cost, partitioned Cox cost, ignoring censoring and complete costs) varied by approximately £3K. It could be argued that a £3K difference around the "true" mean estimate is a small difference and it is immaterial which method is used to estimate mean costs (Observed mean CELT study cost = £36,045). Variation increased under informative censoring due to ill health and partial censoring (ignoring estimates for Lin's method (UCH), the weighted cost (UCH) and Carides' method under 80% partial censoring) to between £4K to £8K and could now be argued to be a meaningful difference in variation across methods. These differences are specific to the CELT study and the variation across methods depends upon the type of censoring mechanism. If the mean cost of transplantation had been lower, for example £20K, a £3K difference between methods might be regarded as meaningful. Given that the variation across methods is influenced by censoring mechanism and the expected "true" mean cost would impact on the importance given to the magnitude of the variation across methods it is recommended that Lin's method KCH with small interval length be used to estimate mean total costs in the presence of censoring.

The recommendations on the selection of an appropriate method for estimating mean total costs in the presence of censoring are based on observations from one cost-effectiveness study. The CELT study was most likely to be subject to end-of-study censoring and had a distinct pattern of cumulative costs over time. Other solid organ transplant studies are likely to observe a similar resource usage pattern over time, it is assumed that the recommendations given in this thesis are generalisable to other solid organ transplant studies estimating mean costs in the presence of censoring.

However, the methods that produced the most accurate estimate of mean total costs i.e. Lin's method KCH with small intervals, might not give the most accurate estimates of mean costs in other data sets where the pattern of resource use differs to that observed in the CELT study. Each of the non-naïve methods for estimating mean total costs (Lin's methods KCH and UCH, weighted cost methods UCH and KCH, Lin's regression methods UCH and KCH and Carides' method), apply Kaplan-Meier survival probability weights to costs.

Lin's method KCH weights mean costs per interval by the Kaplan-Meier probability of survival in each interval, with mean costs in later intervals weighted lower than mean costs in earlier intervals. If a study observed an increase in resource usage over time, for example in studies of chronic conditions, then lower weights will be applied to the higher costs at the end of the study which could result in an underestimation of mean costs. Similarly, cost estimates from a study observing a constant resource use pattern over time, for example a drug treatment, may also be underestimated.

Lin's method UCH and Carides' methods use a similar process and weight cost estimates by the probability of dying, where the largest weights occur in intervals with the greatest probability of death. Finally, the weighted cost methods and Lin's regression methods weight costs by the inverse of the Kaplan-Meier survival probability, with reverse censoring, thus inflating costs, with a higher level of inflation applied to costs in the later intervals of a study.[1] Therefore, the weighted cost method and Lin's regression method with UCH might overestimate mean cost estimates in studies where costs increase over time (chronic conditions) or studies with constant resource patterns over time (drug treatments) by over inflating costs at the end of the study period.

The above discussion speculates as to how mean total cost estimates could be affected by alternative patterns of resource use to those observed for in the CELT study, i.e. resource use patterns for chronic conditions and drug treatments, and it is important to investigate these hypotheses with further research.

[1] If cost histories are known the study period is divided into several intervals and costs will be given a larger inflation rate in the later part of each interval in comparison to earlier parts of the interval.

8.4 PROGNOSTIC MODEL UNCERTAINTIES – MEASURING MODEL PARAMETER, MODEL STRUCTURE AND METHODOLOGICAL UNCERTANITIES

To date none of the published solid organ transplant studies that applied prognostic models to estimate non-transplant survival, and subsequently QALYs or costs, incorporated uncertainties in the prognostic model estimates. Chapters 5 to 7 presented simulation methods for estimating the uncertainty around these estimates, demonstrating that it is possible to account for uncertainties in prognostic models.

Published prognostic models can be applied in studies to predict patient outcomes (i.e. survival, QALYs and costs) in the absence of observed data or in situations where researchers wish to model the likely implications of a change in a current treatment policy or the introduction of a new policy. Prognostic models were applied to CELT study patients in order to estimate, what would have been, their survival in the absence of liver transplantation.

Prognostic models are subject to various types of model uncertainty (parameter, structure and methodological process) and Chapters 5 to 7 present PSA techniques that account for prognostic model parameter and outcome uncertainties. Chapters 5 and 7 also considered ways of selecting the appropriate prognostic model, when more than one model exists. The remainder of this section considers the methodological contributions of the simulation techniques proposed in this thesis to account for model parameter uncertainty, selection (structural) uncertainty and methodological uncertainty.

8.4.1 Model Parameter Uncertainty

Prognostic models are subject to model parameter uncertainty and Chapter 5 proposed a Monte Carlo simulation technique to allow for this uncertainty when estimating survival from Cox PH models. The Monte Carlo method uses information from prognostic model regression coefficients and their standard errors to measure model parameter uncertainty. The method can be applied to any mathematical model to estimate parameter uncertainty, provided that information on the standard errors of (and correlation between) regression coefficients are given. In Chapter 7 the Monte Carlo technique was also applied to a three-part regression model for estimating non-transplant costs to measure the parameter uncertainty around cost estimates, alongside prognostic model parameter uncertainty.

Further work should be carried out to establish whether it is important to adjust for correlations between regression coefficients when accounting for prognostic model parameter uncertainties. The only prognostic model where it was possible to incorporate correlations between regression coefficients was the PBC Mayo model, where the original Mayo model data were available. The results showed that allowing for correlations between regression coefficients for the PBC Mayo model had little noticeable impact on the uncertainty around prognostic model estimates in comparison with estimates where only information on the standard errors of regression coefficients was used. However, guidelines set by NICE state that decision makers should "know about the uncertainty associated with clinical and cost-effectiveness information" [NICE, 2004]. Therefore, it is recommended that the additional uncertainty from the correlation between regression coefficients should be incorporated in to model parameter uncertainty, in order to provided further knowledge about uncertainty, provided that this information is available.

None of the published prognostic models used to estimate non-transplant survival in the CELT study provided information on the correlation or covariance matrix for prognostic model regression coefficients. Therefore, in order for this extra uncertainty to be incorporated in to survival estimates it is advised that authors of future prognostic models provide an additional table with the correlation coefficients of the regression coefficients. Providing the reader with this additional information would give potential users the opportunity to obtain more realistic estimates of uncertainty and journal editors ought to ensure that this is done.

Both Chapter 5 and 7 demonstrated that the incorporation of prognostic model parameter uncertainty impacts on the CI around mean outcome estimates, where the greater the amount of uncertainty in the prognostic model, the larger the amount of uncertainty around mean estimates. This was demonstrated in Chapter 5, where the PBC Mayo model, which had a larger amount of uncertainty in its parameters than the Royal Free and European models, displayed a larger amount of uncertainty around mean non-transplant survival estimates than the other two models. All three prognostic models gave similar mean non-transplant survival estimates.

Additionally, this thesis has raised three issues regarding the accuracy of non-transplant cost estimates. Firstly, it raises questions about the appropriateness of assuming non-transplant costs remain constant over time, except for a period prior to

death. The main CELT study was not the only organ transplant study to assume that non-transplant costs remained constant over time [Bonsel *et al.* 1990a; Van Enckevort *et al.* 1997; Anyanwu *et al.* 2002]. Therefore, it appeared reasonable to assume that costs remained constant over time in the CELT study. It is perhaps more realistic to assume that non-transplant costs vary over time, and further work should be carried within the area of transplantation to explore the pattern of costs over time using time dependent modelling techniques. For example, resource use information was collected for the CELT study on a daily basis, making it possible to calculate CELT study costs on a daily basis. Modelling daily costs over time would establish whether assumptions about alternative rates of change in costs reflect non-transplant resource use more accurately. It is therefore suggested that future economic evaluations in solid organ transplantation should investigate these issues further.

Secondly, the large amount of uncertainty around modelled non-transplant cost estimates raises concerns about whether other models might fit the data better than the three-part model applied in Chapters 6 and 7. Alternative approaches to modelling daily costs could be explored using methods that allow for model structure uncertainty [See: Draper, 1995].

The final area of concern was the effect of cost model parameter uncertainty on mean estimates, where mean non-transplant cost estimates decreased by £20K after allowing for cost model parameter uncertainty. It is generally assumed that mean estimates remain constant and only the degree of uncertainty around them varies, however this thesis showed that this assumption does not always hold true. It is recommended that decision makers make choices about the clinical or cost-effectiveness of treatments or technologies using a base case scenario based upon mean survival, QALY and cost estimates prior to allowing for any uncertainties. Mean survival, QALY, cost or cost-effectiveness estimates that incorporate uncertainties should then be presented separately and these can be used to inform decisions regarding whether further research should be undertaken. This supports recommendations made by Claxton that decision makers make policy implications based on mean estimates and the information on uncertainty is used to make decision about whether further research is needed using value of information analysis [Claxton, 1999].

8.4.2 Model Selection Uncertainty

Three prognostic models were available for estimating non-transplant survival in patients with end-stage liver failure. The derivation of these three prognostic models is subject to model process uncertainty, where three independent research teams chose alternative modelling approaches (fixed time or time dependent models), and had different variable selection criteria (Chapter 5). This resulted in three alternative prognostic models that contained different combinations of variables for estimating survival. All three prognostic models gave similar mean estimates when applied to the CELT cohort. Chapter 7 considered whether it was possible to merge the regression parameters from the three prognostic models in to one model and concluded that it was more appropriate to present separate estimates of non-transplant survival, QALYs and costs for each prognostic model. Therefore, in the absence of original prognostic model data, when more than one appropriate prognostic model exists, scenario analysis that presents the results from all alternative prognostic models is recommended.

8.4.3 Methodological Uncertainty

The analytic need to estimate individual patient outcomes will generally be specific to studies where:

- observed information on patient outcomes are unavailable and prognostic models are utilised to estimate patient outcomes;
- outcomes are to be estimated over a fixed study period;
- individual outcome estimates are needed, for example to:
  - estimate the survival, QALY or costs at the patient level;
  - estimate the survival, QALY or cost gain at the patient level; and,
  - make an adjustment to costs, or HRQL data for a time period prior to death.

In the CELT study, the published prognostic models did not provide enough information to estimate "lifetime" patient survival. Additionally, they did not provide information as to who was alive or dead at a fixed prior time point (five years). Therefore, Chapter 6 presented three alternative methods for estimating individual patient outcomes and the uncertainty around them:

- assumes the probability of survival is equivalent for all patients, with survivors selected randomly

- assumes the N patients with the longest expected survival times are the survivors, where N is the expected number of survivors in the cohort
- PSA

PSA was recommended for estimating individual patient outcomes, as it accounted for patient level information relating to expected outcome and allowed for uncertainty in individual patient outcome estimates.

As outlined above, there are circumstances in which a researcher will want to estimate individual patient outcomes from a Cox PH prognostic model, and to date, no study has attempted to formalise these methods. Therefore this thesis has made a contribution to the methodology by providing three alternative methods for estimating individual patient outcomes, with PSA being the preferred method.

8.4.4 Implications – Measuring Prognostic Model Uncertainties

PSA often purports to reflect the true uncertainty in clinical and cost-effectiveness studies. This thesis highlights that in the absence of an observed control group there are more uncertainties than are normally recognised in traditional cost-effectiveness studies, in which data from a control cohort are traditionally observed. PSA is able to incorporate the additional uncertainty from using prognostic models to estimate control group outcomes in the absence of observed data alongside more traditional forms of uncertainty, for example uncertainty in unit cost or HRQL estimates.

8.5 FURTHER RESEARCH IN TO UNCERTAINTY

This thesis has focused on providing guidance in the selection of a method for accurately estimating mean total costs in the presence of censoring and presenting methods for measuring uncertainty around prognostic models. In the course of exploring these issues other areas of uncertainty in censoring, prognostic modelling and within the CELT study were highlighted. This section presents these issues in further detail.

8.5.1 Further Issues of Censoring in QALY, Cost or Cost-effectiveness Studies

It is assumed that the recommendation of using Lin's method (KCH) with small interval lengths is generalisable to all solid organ transplantation studies that are subject to censored cost data. However, it is recommended that further research is undertaken to

establish whether these recommendations can be generalised to health care research studies.

HRQL is typically measured at one or more fixed time points during a study, patients may chose not to respond at particular time points and thus HRQL data could have a higher likelihood of informative censoring than resource use data. Given that the censoring mechanism may differ to that for resource use data caution is advised in generalising the results shown here to QALY. Further work should also be conducted to establish the generalisability of the results to censored QALY and cost-effectiveness data. Additional, future work should focus on accurate estimates of both the mean and the standard error because decision makers are advised to give advice on the necessity of further research, based upon the uncertainty around clinical and cost-effectiveness estimates.

Any method that accounts for censoring in cost-effectiveness studies should allow for the correlation between QALY and cost data and a number of researchers have proposed methods that account for this when estimating the cost-effectiveness of treatments in the presence of censoring [Willan & Lin, 2001; Backhouse *et al*, 2002; Willan *et al,* 2002; Willan *et al,* 2003; Heitjan *et al,* 2004; Willan *et al,* 2004]. Therefore, any future research that compares the accuracy of methods for estimating the cost-effectiveness of health technologies should allow for these methods.

8.5.2 Further Issues of Model Uncertainty

This thesis focused upon methodology for estimating the uncertainty around the Cox PH prognostic models with the parameter uncertainty around non-transplant cost estimates presented in Chapter 7. A further area of model uncertainty that could also be explored in more detail is the uncertainty around HRQL estimates. It was decided to focus on the uncertainty around Cox PH and cost models because conventional PSA typically allows for uncertainty in HRQL estimates. However, it is important to bring to the readers attention that utility estimates are also subject to model parameter and model structure uncertainty.

The EQ-5D was used in the main CELT analysis to estimate non-transplant HRQL, and subsequently, utility values and QALYs. Utility estimates for the EQ-5D were derived by applying an algorithm to each patient's EQ-5D responses. The algorithm for estimating utilities from EQ-5D responses was originally derived by Dolan by fitting a general least

squares regression model, i.e. a prognostic model, to a cohort of responders to the EQ-5D [Dolan, 1997]. Thus, the EQ-5D algorithm consists of a set of regression coefficients derived from a prognostic model and these regression coefficients are subject to model parameter uncertainty. The Monte Carlo simulation techniques, presented in Chapter 5, could have been applied to the CELT cohort to account for model parameter uncertainty using information on the standard errors (and correlation) from the regression coefficients. Any future PSA analysis of the CELT data that attempts to account for all sources of uncertainty should allow for the model uncertainties in EQ-5D utility estimates. Further, all clinical and cost-effectiveness studies of health care technologies that wish fully incorporate uncertainties and who have used the EQ-5D should consider applying Monte Carlo simulation methods to EQ-5D model parameters to allow for EQ-5D model parameter uncertainty.

Further, utility estimates can also be subject to model structure/selection uncertainty. For example, the CELT study patients complete two HRQL questionnaires the EQ-5D and the SF-36. It is possible to derive patient utilities from both of these instruments[2]. Several authors have compared the utility values of the SF-6D with the EQ-5D and concluded that the instruments give different utility values [See: McDonough *et al,* 2005; Stavern *et al,* 2005]. The SF-6D is ineffective at describing health states at the lower end of the utility scale (close to zero) and the EQ-5D is ineffective at the upper end of the scale (close to one) [Longworth & Bryan, 2003; Brazier *et al.* 2004]. Instrument selection was not considered in this thesis, which focused on prognostic model selection process, an area of uncertainty that has not previously been considered by other researchers. However, the main CELT analysis estimated the cost-effectiveness of liver transplantation over 2.25 years using both the EQ-5D (main analysis) and the SF-6D (sensitivity analysis) [Longworth *et al,* 2003].

8.5.3 Other Issues of Uncertainty

This thesis has approached the issue of statistical uncertainty from a standard frequentist point of view, thus all the methods and issues dealt with, within this thesis, use frequentist methods. An alternative approach would have been to consider the use of Bayesian methods in the estimation of uncertainty within the CELT study and methods exist for accounting for uncertainty in health care research studies within a Bayesian framework [See: Spiegelhalter *et al,* 2004]. None of the methods identified for

[2] Brazier *et al* derived a preference based measure from the SF-36, known as the SF-6D [Brazier *et al,* 2002]. A scoring system is assigned to six of the questions in the 36 item instrument and, as with the EQ-5D a value of one represents full health and zero death.

estimating mean costs in the presence of censoring used Bayesian analysis, therefore it was not applicable to focus on Bayesian methods here. However, it would have been possible to develop Bayesian techniques when estimating prognostic model uncertainty and these techniques could be developed in future research.

The focus of this thesis was to identify an appropriate method for estimating average costs in the presence of censoring and presenting methodology for estimating uncertainty around prognostic model estimates. Therefore, this thesis did not focus on the overall cost-effectiveness of liver transplantation for the CELT study over a five-year study period, nor did it conduct a full PSA that allowed for all possible sources of uncertainty that arose within the CELT study.

8.6 IMPLICATIONS FOR THE COST-EFFECTIVENESS OF LIVER TRANSPLANTATION

As already stated in this chapter, the aim of this thesis was to explore the methodological issues of measuring uncertainties in economic evaluations. Therefore, the impact of prognostic model uncertainties on the cost-effectiveness of liver transplantation in the UK was not estimated. However, this section examines how allowing for prognostic model uncertainties might affect cost-effectiveness estimates for liver transplantation and the likely impact on policy decisions.

Survival information post transplant was routinely collected and available at each of the six centres participating in the CELT study, and was obtained from each centre for the extended five-year study period. Thus, it was possible to estimate the survival gain from transplantation over five-years after adjusting for model parameter uncertainty in the non-transplant estimates. In Chapter 5 the survival gain from transplantation over five years was estimated at 1.9 years for PBC patients, and after adjusting for prognostic model parameter uncertainty varied from -0.4 years to 4.1 years across the three prognostic models. This analysis could be expanded further to include the uncertainty in estimating non-transplant outcomes over five years. Chapter 7 showed a slight increase in the uncertainty in non-transplant survival estimates after incorporating prognostic model parameter uncertainty and outcome uncertainty into survival estimates in comparison to uncertainty that incorporated prognostic model parameter uncertainty. Therefore, it is expected that after allowing for this the mean survival gain would remain at approximately 1.9 years and the uncertainty around the survival gain would increase slightly compared to the results shown in Chapter 5.

Although survival information was routinely collected, no resources were available (to the CELT study) to obtain information on HRQL or resource use for 2.25 to five years post assessment. Hence estimation techniques were needed to estimate transplant HRQL and costs for the extended study period, drawing upon information from the CELT study. It is likely that the assumption of constant HRQL over time assumed in the non-transplant HRQL estimates would also be applied to transplant HRQL between 2.25 and five years. Therefore, I would expect to observe an increase in transplant QALY over those observed to 2.25 years in the CELT study with a smaller increase in non-transplant QALYs at five years. Thus, there is likely to be a QALY gain in favour of transplantation at five years. After adjusting for prognostic model uncertainties there is likely to be an increase in the uncertainty around the incremental QALYs compared to estimates at 2.25 years. I would expect a similar range of uncertainty in the QALY gain from transplantation to that observed for the survival gain over five years for PBC patients (-0.4 to 4.1 years) with possible evidence of a negative QALY gain in favour of non-transplant patients to a large positive gain of 4 years or more in favour of transplantation.

Figure 4.9 illustrated how cumulative CELT study costs remained fairly constant in the later part of the 2.25 year study period. This pattern is likely to continue over a five year period where patients are stable and typically seen at outpatient appointments every six to twelve months, therefore the mean costs of transplantation are not expected to increase very much compared to those observed at 2.25 years. In contrast, individual patient non-transplant costs are expected to increase cumulatively over time. Therefore, I would expect a reduction in the incremental costs of transplantation, which could be cost-saving, and a large increase in the level of uncertainty around cost estimates, which would range from cost saving in favour of non-transplantation to cost saving in favour of transplantation, after allowing for prognostic model and cost model uncertainty.

An increase in the QALY gain and a reduction in the incremental costs would result in a decrease in the incremental cost effectiveness of transplantation over five years. Thus, the mean ICER for ALD patients could drop below the £30K NHS benchmark indicating transplantation as a beneficial treatment for ALD patients over five years. The mean ICER for PBC and PSC patients were both less than £30K at 2.5 years (Chapter 2) and it is expected that they would drop further below the £30K benchmark. Furthermore, allowing for prognostic model and cost model uncertainties would increase the amount

of variation around ICERs considerably compared to the variation estimated at 2.25 years in the original CELT study. After allowing for prognostic model and cost model uncertainties I would expect that bootstrapped estimates of the ICERs to cover all four quadrants of the incremental cost-effectiveness plane, with the majority of ICERs showing liver transplantation as a cost-effective treatment over five years and a very small proportion showing non-transplantation to be less costly and more beneficial than transplantation.

Thus, the methods presented in this thesis for measuring uncertainty in cost-effectiveness studies are likely to substantially increase the degree of uncertainty in cost-effectiveness estimates. However, I would not expect these results to alter the current decision to fund liver transplantation. Liver transplantation is accepted as the treatment of choice for patients with end stage liver failure and it is believed to be beneficial and to improve patients HRQL, although acknowledged as a costly procedure [Burton & Heyse, 1985]. A cost-effectiveness study of lung-transplantation in The Netherlands suggested that lung transplantation was beneficial and improved HRQL but was very expensive [van Enckevort, TenVergert *et al.* 1998]. As a result of the Dutch study the Dutch Health Care insurance board and Dutch Minister of Health Affairs chose to include lung-transplantation in the countries health benefit package [Ouwens *et al.* 2003].

8.7 CONCLUDING COMMENTS

This thesis focused upon two issues of statistical uncertainty that arose as a result of the potential extension of the study period in an economic evaluation in liver transplantation. By addressing these two issues of uncertainty this thesis has aided the transplant community by enabling future studies in solid organ transplantation to obtain more accurate estimates of survival, QALY and cost outcomes. When study observations are subject to censoring, the guidance provided in this thesis on censoring methods will enable more accurate estimates of transplant costs, and the uncertainty around them. This thesis has also highlighted the importance in acknowledging the difficulties in selecting an appropriate non-transplant control group.

The methodology presented in this thesis for estimating uncertainty around non-transplant survival, QALY and costs estimated from prognostic models should be applied to any solid organ transplant study that uses prognostic models in the absence of observed non-transplant data. In presenting the uncertainty around non-transplant

estimates the transplant community will be presented with a truer picture of the uncertainty around non-transplant survival, QALY and cost estimates. This, in turn, will enable a more realistic picture of the uncertainty around the survival and QALY gains and the cost differences and aid decision makers in implementing policy decisions regarding solid organ transplantation.

This thesis has also raised questions about the credibility of the results from previous studies that assumed non-transplant costs could be estimated from the transplant experience of patients on the waiting list. Chapter 7 showed that there was a large amount of uncertainty around non-transplant cost estimates and raised questions about the accuracy of estimating non-transplant costs in this way. It is important that future economic evaluations in solid organ transplantation address this issue and this should result in more credible estimates of the cost-effectiveness of solid organ transplantation in the future.

This thesis has contributed to the methodology and guidance in health care research studies by offering guidance on selection of appropriate method for estimating mean study costs in the presence of censoring. Further, this thesis has contributed to the research field by presenting simulation methods for estimating uncertainty around prognostic model estimates and presenting a PSA approach for estimating individual patient outcomes (and uncertainty) over a fixed study time period estimated from Cox PH prognostic models.

This thesis identified two previously unexplored issues of uncertainty that can arise in clinical and cost-effectiveness studies in health care technologies. By exploring these issues and offering guidance on techniques and presenting methodology for estimating uncertainties this thesis will enable decision makers to have more confidence in clinical and cost-effectiveness estimates and present them with a more complete picture of uncertainty around these estimates. A more accurate picture of the uncertainty around effectiveness, cost and cost-effectiveness estimates can, in turn, inform value of information analysis, which can be used to inform decision makers of the need for further investigation in to health care services [Claxton, 1999]. Thus, decisions in to the need for future research depend upon accurate estimates of the uncertainty around clinical or cost-effectiveness estimates of health care technologies and this thesis has provided methodology that contributes to a more precise estimate of this uncertainty.

References

Adang EM, Engel GL, *et al,* (1996) Comparison before and after transplantation of pancreas-kidney and pancreas-kidney with loss of pancreas--a prospective controlled quality of life study. Transplantation; 62(6): 754-759

Aitkin M, Laird N, Francis B, (1983) A reanalysis of the Stanford heart transplant data. Journal of the American Statistical Association; 78 (382): 264-274

Al MJ, Koopmanschap M, *et al, (*1998) Cost-effectiveness of lung transplantation in the Netherlands: a scenario analysis. Chest; 113(1): 124-130

Altman DG, Royston P, (2000) What do we mean by validating a prognostic model? Statistics in Medicine; 19: 453-473

Anand AC, Ferraz-Neto BH, *et al,* (1997) Liver transplantation for alcoholic liver disease: evaluation of a selection protocol. Hepatology; 25(6): 1478-1484

Anyanwu AC, McGuire A, *et al, (*2001) Assessment of quality of life in lung transplantation using a simple generic tool. Thorax; 56(3): 218-222

Anyanwu AC, McGuire A, *et al,* (2002) An economic evaluation of lung transplantation. The Journal of Thoracic and Cardiovascular Surgery; 123 (3): 411-420

Aranzabal J, Perdigo L, *et al, (*1991) Renal transplantation costs: an economic analysis and comparison with dialysis costs. Transplantation Proceedings; 23(5): 2574

Babyak MA, (2004) What you see may not be what you get: a brief, nontechnical introduction to overfitting in regression-type models. Psychosomatic Medicine; 66(3): 411-21

Blackhouse G, Briggs AH, O'Brien BJ, (2002) A note on the estimation of confidence intervals for cost-effectiveness when costs and effects are censored. Medical Decision Making; 22: 173-177

Baltzan MA, Ahmed S, *et al, (*1997) Variations in living donor graft rates by dialysis clinic: effect on outcome and cost of chronic renal failure therapy. Clinical Nephrology; 47(6): 351-355

Baltzan MA, Shoker AS, *et al, (*1996) HLA-identity-long-term renal graft survival, acute vascular, chronic vascular, and acute interstitial rejection. Transplantation; 61(6): 881-885

Bang H, Tsiatis AA, (2000) Estimating medical costs with censored data. Biometrika; 87(2): 329-343

Batra N, (2001) Hepatitis C screening and treatment versus liver transplantation: a financial option appraisal and commissioning model for purchasers. Disease Management & Health Outcomes; 9(7): 371-384

Belle SH, Porayko MK, *et al, (*1997) Changes in quality of life after liver transplantation among adults. National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) Liver Transplantation Database (LTD). Liver Transplantation and Surgery; 3(2): 93-104

Bertrand-Krajewski J-L, (2004) TSS concentration in sewers estimated from turbidity measurements by means of linear regression accounting for uncertainties in both varialbels. Water Science and Technology; 50(11): 81-88

BIDS, (2005) Available from: http://www.bids.ac.uk/. Accessed July 2005

Black N, (1996) Why we need observational studies to evaluate effectiveness in health care. British Medical Journal; 312 (7041): 1215-1218

Black N, Murphy M, *et al,* (1999) Consensus development methods: a review of best practice in creating clinical guidelines. Journal of *Health Services* & Research Policy; 4(4): 236-248

Bland M, (2000) *An Introduction to Medical Statistics.* Oxford Medical Publications, Oxford UK

Blommers TJ, Schanbacher B, Corry RJ, (1984) Transplant and dialysis: the cost/benefit question. Iowa Medicine; 74(1): 15-17

BNF 38, (1999) British Medical Association and the Royal Pharmaceutical Society of Great Britain

Bocchi EA, Bellotti G, *et al,* (1996) Mid-term results of heart transplantation, cardiomyoplasty, and medical treatment of refractory heart failure caused by idiopathic dilated cardiomyopathy. Journal of Heart and Lung Transplantation; 15(7): 736-745

Bonal J, Cleries M, Vela E, (1997) Transplantation versus haemodialysis in elderly patients. Renal Registry Committee. Nephrology Dialysis Transplantation; 12(2): 261-264

Bonsel GJ, Essink-Bot ML, *et al,* (1990) Orthotopic liver transplantation in The Netherlands. The results and impact of a medical technology assessment. Health Policy; 16(2):147-161

Bonsel GJ, Klompmaker IJ, *et al,* (1990a) Use of prognostic models for assessment of value of liver transplantation in primary biliary cirrhosis. The Lancet; 335:493-497

Bonsel GJ, Klompmaker IJ, *et al, (*1990b) Cost-effectiveness analysis of the Dutch liver transplantation programme. Transplantation Proceedings; 22(4): 1481-1484

Bonsel GJ, Essink-Bot ML, *et al, (*1992) Assessment of the quality of life before and following liver transplantation. First results. Transplantation; 53(4): 796-780

Bortman G, Delgado D, *et al,* (1999) Analysis of quality of life before and after heart transplantation. Transplant Proceedings 31(6):2555

Bradburn MJ, Clark TG, *et al,* (2003) Survival analysis part II: multivariate data analysis--an introduction to concepts and methods. British Journal of Cancer; 89(3):431-436

Brazier J, Roberts J, Deverill M, (2002) The estimation of a preference based measure of health from the SF-36. Journal of Health Economics; 21 (2): 271-292

Brazier J, Roberts J, *et al,* (2004) A comparison of the EQ-5D and the SF-6D across seven patient groups, Health Economics; 13(9): 873-884

Briggs AH, (2000) Handling uncertainty in cost-effectiveness models. Pharmacoeconomics; 17(5): 479-500

Briggs AH, (2001) Handling uncertainty in economic evaluation and presenting the results. In E*conomic Evaluation in Health Care: merging theory with practice.* Eds. Drummond M, McGuire A. Oxford University Press, Oxford

Briggs AH, Fenn P, (1997) Trying to do better than average: a commentary on "statistical inference for cost-effectiveness ratios." Health Economics; 6(5): 491-495

Briggs AH, Gray D, (1999) Handling uncertainty when performing economic evaluation of health care interventions. Health Technology Assessment; 3(2)

Briggs A, Sculpher M, Buxton M, (1994) Uncertainty in the economic evaluation of health care technologies: the role of sensitivity analysis. Health Economics; 3(2): 95-104

Briggs AH, Wonderling DE, Mooney CZ, (1997) Pulling cost-effectiveness analysis up by its bootstraps: a non-parametric approach to confidence interval estimation. Health Economics; 6(4): 327-340

Brooks R, (1996) EQ-5D, the current state of play. Health Policy; 37: 53-72

Burroughs A, Blake J, *et al,* (1992) Comparative hospital costs of liver transplantation and the treatment of complications of cirrhosis. European Journal of Gastroenterology and Hepatology; 4: 123-128

Burton BT, Heyse SP, (1985) Technology assessment in biomedical research: magnetic resonance imaging as a case in point. Journal of Nuclear Medicine and Allied Sciences. 29(4): 331-335

Busschbach JJ, Horikx PE, *et al,* (1994) Measuring the quality of life before and after bilateral lung transplantation in patients with cystic fibrosis. Chest; 105(3): 911-917

Buxton M, Acheson R, *et al,* (1985) Costs and benefits of the heart transplant programmes at Harefield and Papworth hospitals. Department of Health and Social Security Research Report No. 12

Caine N, Sharples LD, *et al,* (1990) Prospective study comparing quality of life before and after heart transplantation. Transplantation Proceedings; 22 (4): 1437-1439

Caine N, Sharples LD, *et al,* (1996) Measurement of health-related quality of life before and after heart-lung transplantation. Journal of Heart and Lung Transplantation; 15 (10): 1047-1058

Calne RY, Williams R, (1968) Liver transplantation in man I. Observations on technique and organisation in five cases. British Medical Journal; 4(630): 535-540

Canadian Agency for Drugs and Technologies in Health, (2006) *Guidelines for the economic evaluation of health technologies: Canada* [3rd Edition]. Ottawa

Cardies GW, Heyse JF, Iglewicz B, (2000) A regression-based method for estimating mean treatment cost in the presence of right-censoring. Biostatistics; 1(3): 299-313

Centre for Review and Dissemination, (2005) NHS Economic Evaluation Database (NHS EED) Available from: http://www.york.ac.uk/inst/crd/nhsdhp.htm. Accessed July 2005

Chantler C, Carter JE, *et al,* (1980) 10 years' experience with regular haemodialysis and renal transplantation. Archives of Disease in Childhood; 55(6):435-445.

Cheung AH, Sutherland DE, *et al,* (1992) Simultaneous pancreas-kidney transplant versus kidney transplant alone in diabetic patients. Kidney International; 41(4): 924-929

Christensen E, Altman DG, *et al,* (1993) Updating prognosis in primary biliary cirrhosis using a time dependent Cox regression model. Gastroenterology; 105: 1865-1876

Christensen E, Gunson B, Neuberger J, (1999) Optimal timing for liver transplantation for patients with primary biliary cirrhosis: use of prognostic modelling. Journal of Hepatology; 30: 285-292

Christensen E, Neuberger J, *et al,* (1985) Benificial effect of azathioprine and prediction of prognosis in primary biliary cirrhosis: Final results of an international trial. Gastroenterology; 89: 1084-1091

Churchill DN, Torrance GW, *et al,* (1987) Measurement of quality of life in end-stage renal disease: the time-trade off approach. Clinical and Investigative Medicine - Medicine Clinique et Experimentale; 10(1): 14-20

Claxton K, (1999) Bayesian approaches to the value of information: implications for the regulations of new pharmaceuticals. Health Economics; 8(3): 269-274

Clark DA, Stinson EB, *et al,* (1971) Cardiac Transplantation in Man: VI. Prognosis of patients selected for cardiac transplantation. Annals of Internal Medicine; 75: 15-21

Cleveland W, Devlin S, (1988) Locally weighted regression: An approach to regression analysis by local fitting. Journal of the American Statistical Association; 83(403): 596-610

Cohen L, Littlefield C, *et al,* (1998) Predictors of quality of life and adjustment after lung transplantation. Chest; 113(3): 633-644

Cole CR, Bucuvalas JC, *et al,* (2004) Impact of liver transplantation on HRQOL in children less than 5 years old. Pediatric Transplantation; 8(3): 222-227

Collett D, (1994) *Modelling survival data in medical research.* Chapman & Hall, London

Cope JT, Kaza AK, *et al,* (2001) A cost comparison of heart transplantation versus alternative operations for cardiomyopathy. Annals of Thoracic Surgery; 72(4):1298-1305

Cosimi AB, Auchincloss H Jr, *et al,* (1998) Combined kidney and pancreas transplantation in diabetics. Archives of Surgery; 123(5): 621-625

Cotrufo M, Romano G, *et al,* (2005) Treatment of extensive ischemic cardiomyopathy: quality of life following two different surgical strategies. European Journal of Cardio-Thoracic Surgery; 27(3): 481-487

Cox D, (1972) Regression models and life tables (with discussion. Journal of the Royal Statistical Society B; 74: 187-220

Cox LA, Popken DA, (2004) Bayesian Monte Carlo uncertainty analysis of human health risk from animal antimicrobial use in a dynamic model of emerging resistance. Risk Analysis; 24(5): 1153-1164

Coyle D, Buxton MJ, O'Brien BJ, (2003) Measures of importance for economic analysis based on decision modelling. Journal of Clinical Epidemiology; 56(10): 989-997

Critchfield GC, Willard KE, Connelly DP, (1986) Probabilistic sensitivity analysis methods for general decision models. Computers and Biomedical Research; 19: 254-265

Crombie IK, Davies HTO, (1996) Research in health care: Design, conduct and interpretation of health services research. Oxford University Press, Oxford

Croxson BE, Ashton T. (1990) A cost-effectiveness analysis of the treatment of end-stage renal failure. New Zealand Medical Journal; 103(888): 171-174

Dalkey NC, Helmer O, (1963) An experimental application of the Delphi method to the use of experts. Management Science; 9: 458-467

De Bona M, Ponton P, *et al,* (2000) The impact of liver disease and medical complications on quality of life and psychological distress before and after liver transplantation. Journal of Hepatology; 33(4): 609-615

De By TM, D'Amaro J, *et al,* (1982) A retrospective analysis of hospital haemodialysis and kidney transplantation costs in The Netherlands. Netherlands Journal of Medicine; 25(3): 83-87

de Wit AG, Ramsteijn PG, de Charro F, (1998) Economic evaluation of end stage renal disease treatment. Health Policy; 44(3): 215-232

Decisioneering, Inc. (1994) *Crystal ball: User's guide.* Denver, CO; Decisioneering, Inc

Delbecq AL, Van de Ven AH, (1972) The nominal group as a research instrument for exploratory health studies. American Journal of Public Health; 62(3): 337-342

Devlins GM, Mandin H, *et al,* (1990) Illness intrusiveness and quality of life in end-stage renal disease: comparison and stability across treatment modalities. Health Psychology; 9(2): 117-142

Dickson ER, Fleming TR, *et al,* (1985) Trials of penicilamine in advanced primary biliary cirrhosis. New England Journal of Medicine; 312: 1011-1015

Dickson ER, Grambsch PM, *et al,* (1989) Prognosis in primary biliary cirrhosis: model for decision making Hepatology; 10: 1-7

Dickson ER, Murtaugh PA, *et al,* (1992) Primary sclerosing cholangitis: Refinement and validation of survival models. Gastroenterology; 103: 1893-1901

DiMartini A, Rovera GM, *et al,* (1998) Quality of life after small intestinal transplantation and among home parenteral nutrition patients. Journal of Parenteral and Enteral Nutrition; 22(6): 357-362

Djamali A, Kendziorski C, *et al,* (2003) Disease progression and outcomes in chronic kidney disease and renal transplantation. Kidney International; 64(5): 1800-1807

Dolan P, (1997) Modelling valuations for EuroQol health states. Medical Care; 35 (11): 1095-1108

Doubilet P, Begg CB, *et al,* (1985) Probabilistic sensitivity analysis using Monte Carlo simulations. A practical approach. Medical Decision Making; 5: 157-177

Doubilet P, Weinstein MC, McNeil BJ, (1986) Use and misuse of the term "cost effective" in medicine. New England Journal of Medicine; 314(4):253-256

Douzdjian V, Ferrara D, Silvestri G, (1998) Treatment strategies for insulin-dependent diabetics with ESRD: a cost-effectiveness decision analysis model. American Journal of Kidney Disease; 31(5): 794-802

Douzdjian V, Ferrara D, Silvestri G, (1998) Cost-utility analysis of pancreas transplantation compared to other treatment options for type I diabetics with end-stage renal disease. Transplantation Proceedings; 30: 278

Draper D, (1995) Assessment and propagation of model uncertainty (with discussion) Journal of the Royal Statistical Society (Series B) 57(1); 45-97

Drummond MF, Jefferson TO, (1996) Guidelines for authors and peer reviewers of economic submissions to the BMJ. *British Medical Journal;* 313(7025): 275-283

Drummond MF, O'Brien B, *et al,* (1997) *Methods for the economic evaluation of health care programmes (2nd Edition).* Oxford University Press, Oxford

Drummond MF, Sculpher MJ, *et al,* (2005) *Methods for the economic evaluation of health care programmes (3rd Edition).* Oxford University Press, Oxford

Duan, N, (1983) Smearing Estimate: A Nonparametric Retransformation Method. Journal of the American Statistical Association; 78(383): 602-610

Dudley AR, Harrell FE, *et al,* (1993) Comparison of analytic models for estimating the effect of clinical factors on the cost of coronary artery bypass graft surgery. Journal of Clinical Epidemiology; 46 (3); 261-271

Efron B, Tibshirani R, (1993) An introduction to the bootstrap. New York: Chapman and Hall

Eggers P, (1992) Comparison of treatment costs between dialysis and transplantation. Seminars in Nephrology; 12(3): 284-289

Engineering and Physical Sciences Research Council (EPSRC), (2002) Available at: www.epsrc.ac.uk/documents/info_pub/researeh_highlights/liver.htm Accessed August 2002

Esmatjes E, Ricart MJ, *et al,* (1994) Quality of life after successful pancreas-kidney transplantation. Clinical Transplantation; 8(2 Part 1): 75-78

Etzioni R, Urban N, Baker M, (1996) Estimating the costs attributed to a disease with application to ovarian cancer. Journal of Clinical Epidemiology; 49(1): 95-103

Etzioni RD, Feuer EJ, *et al,* (1999) On the use of survival analysis techniques to estimate medical care costs. Journal of Health Economics; 18: 365-380

European Liver Transplant Registry, www.eltr.org/results.htm Accessed January 2006

Evangalista LS, Moser D, *et al,* (2004) Functional status and perceived control influence quality of life in female heart transplant recipients. Journal of Heart and Lung Transplantation; 23(3): 360-367

Evans RW, Manninen DL, *et al, (*1985) The quality of life of patients with end-stage renal disease. New England Journal of Medicine; 312(9): 553-559

Farinati F, Gianni S, *et al, (*2001) Does the choice of treatment influence survival of patients with small hepatocellular carcinoma in compensated cirrhosis? European Journal of Gastroenterology and Hepatology; 13(10): 1217-1224

Fenn P, McGuire A, *et al,* (1995) The analysis of censored treatment cost data in economic evaluation. Medical Care; 33: 851-863

Fenn P, McGuire A, *et al,* (1996) Modelling programme costs in economic evaluation. Journal of Health Economics; 15: 115-125

Fieller EC, (1932) The distribution of an index in a normal bivariate population. Biometrika; 56: 635-639

Fieller EC, (1954) Some problems in interval estimation. Journal of the Royal Statistical Society Series B; 16(2): 175-185

Fink A, Kosecoff J, *et* al, (1984) Consensus methods: characteristics and guidelines for use. American Journal of Public Health; 74: 979-983

Fisher DC, Lake KD, *et al,* (1995) Changes in health-related quality of life and depression in heart transplant recipients. Journal of Heart and Lung Transplantation; 14(2): 373-381

Foster WR, Burton BT, (1989) Technology assessment applied to liver transplantation in adults. International Journal of Technology Assessment in Health Care; 5: 173-182

Frank A, Deng S, *et al,* (2004) Transplantation for type I diabetes: comparison of vascularized whole-organ pancreas with isolated pancreatic islets. Annals of Surgery; 240(4): 631-643

Fujisawa M, Ichikama Y, *et al,* (2000) Assessment of health-related quality of life in renal transplant and hemodialysis patients using the SF-36 health survey. Urology; 56(2): 201-206

Gaber AO, Hathaway DK, *et al,* (1994) Improved autonomic and gastric function in pancreas-kidney vs kidney-alone transplantation contributes to quality of life. Transplantation Proceedings; 26(2): 515-516

Gail MH, (1972) Does cardiac transplantation prolong life?: A reassessment. Annals of Internal Medicine; 76: 815-817

Gajarski RJ, Towbin JA, Garson A, (1996) Fontan palliation versus heart transplantation: a comparison of charges. American Heart Journal; 131(6): 1169-1174

Garner TI, Dardis R, (1987) Cost-effectiveness analysis of end-stage renal disease treatments. Medical Care; 25(1): 25-34

Gigli A, Verdecchia A, (2000) Uncertainty of AIDS incubation time and its effects on back-calculation estimates. Statistics in Medicine 19(2); 175-89

Goulis J, Leandro G, Burroughs AK, (1999) Randomised controlled trials of ursodeoxycholic acid therapy for primary biliary cirrhosis: a meta-analysis. The Lancet; 354: 1053-1060

Grady KL, Jalowiec A, White-Williams C, (1996) Improvement in quality of life in patients with heart failure who undergo transplantation. Journal of Heart and Lung Transplantation; 15(8): 749-757

Graham D, Graham C, Whitcombe A, (1988) Mathematics A-Level course companion (2nd Edition) Letts study aids, Charles Letts & Co Ltd. London, Edinburgh and New York

Groen H, van der Bij W, *et al, (*2004) Cost-effectiveness of lung transplantation in relation to type of end-stage pulmonary disease. American Journal of Transplantation; 4(7) 1155-1162

Gross CR, Kangas JR, *et al,* (1995) One-year change in quality-of-life profiles in patients receiving pancreas and kidney transplants. Transplantation Proceedings; 27(6): 3067-3068

Gross CR, Savik K, *et al, (*1995) Long-term health status and quality of life outcomes of lung transplant recipients. Chest; 108(6): 1587-1593

Gross CR, Malinchoc M, *et al, (*1999) Quality of life before and after liver transplantation for cholestatic liver disease. Hepatology; 29(2): 356-364

Gross CR, Zehrer CL, (1993) Impact of the addition of a pancreas to quality of life in uremic diabetic recipients of kidney transplants. Transplantation Proceedings; 25(1 Part 2): 1293-1295

Gudex CM, (1995) Health-related quality of life in endstage renal failure. Quality of Life Research; 4(4): 359-366

Haberman S, (1980) Heart transplants: putting a price on life. Health and Social Service Journal; 90(4700): 877-879

Hallstrom AP, Sullivan SD, (1998) On estimating costs for economic evaluation in failure time studies. Medical Care; 36(3): 433-436

Hathaway DK, Hartwig MS, *et al,* (1994) A prospective study of changes in quality of life reported by diabetic recipients of kidney-only and pancreas-kidney allografts. Journal of Transplant Coordination; 4(1): 12-17

Hathaway DK, Hartwig MS, *et al,* (1994a) Improvement in quality of life reported by diabetic recipients of kidney-only and pancreas-kidney allografts. Transplantation Proceedings; 26(2): 512-514

Hathaway DK, Winsett RP, *et al,* (1998) Post kidney transplant quality of life prediction models. Clinical Transplantation; 12(3): 168-174

Hay JW, (1989) Econometric issues in modelling the costs of AIDS. Health Policy; 11: 125-145

Haycox A, Jones D, (1996) The cost effectiveness of renal provision in the UK. Journal of management in medicine; 10(1): 6-15

Heitjan CF, Kim CY, Li H, (2004) Bayesian estimation of cost-effectiveness from censored data. Statistics in Medicine; 23(8): 1297-1309

Hellinger FJ, (1982) An analysis of a public programme for heart transplantation; Journal of Human Resources; 17(2): 307-313

Henderson R, Jones M, Stare J, (2001) Accuracy of point predictions in survival analysis. Statistics in Medicine; 20: 3083-3096

Hiatt RA, Quesenberry CP Jr, *et al,* (1990) The cost of acquired immunodeficiency syndrome in northern California: The experience of a large prepaid health plan. Archives of Internal Medicine; 150: 833-838

Hockerstedt K, (1990) Liver Transplantation Today; Scandinavian Journal of Gastroenterology 25(1): 1-10

Hockerstedt K. Kankaanpaa J, (1986) Liver transplantation in Europe--present status. International Journal of Technology Assessment in Health Care; 2(3):451-463

Hodgson TA, (1992) Cigarette smoking and lifetime medical expenditures. The Milbank Quarterly; 70 (1): 81-125

Holohan TV, (1996) Cost-effectiveness modelling of simultaneous pancreas-kidney transplantation. International Journal of Technology Assessment in Health Care; 12(3): 416-424

Hosenpud JD, Bennett LE, *et al, (*1998) Effect of diagnosis on survival benefit of lung transplantation for end-stage lung disease. The Lancet; 351(9095): 24-27

Huang Y, (2002) Calibration regression of censored lifetime medical cost. Journal of the American Statistical Society; 97(457): 318-327

Hughes MD, Raskino CL, *et al,* (1992) Prediction of shrt-term survival with an application in primary biliary cirrhosis. Statistics in Medicine; 11: 1731-1745

Hummel *et al,* (2000) How expensive is heart transplantationn (HTX) when compared with treatment of end stage heart failure (ESHF) before HTx? A single centre analysis of 350 patients. Journal of Heart and Lung Transplantation; 19(1): 44

Jacobson SH, Fryd D, *et al,* (1988) Transplantation, hemodialysis, and continuous ambulatory peritoneal dialysis for end-stage renal disease in diabetic patients. Journal of Diabetic Complications; 2(3): 150-157

Jain AK, Strawderman RL, (2002) Flexible hazard regression modelling for medical cost data. Biostatistics; 3: 101–118

Jassal SV, Krahn MD, *et al,* (2003) Kidney transplantation in the elderly: a decision analysis. Journal of the American Society of Nephrology; 14(1): 187-196

Jofre R, Lopez-Gomez JM, *et al, (*1998) Changes in quality of life after renal transplantation. American Journal of Kidney Disease; 32(1): 93-100

Johnson JL, Schellberg J, *et al,* (1990) Does pancreas transplantation really improve the patient's quality of life? Transplantation Proceedings; 22(2): 575-576

Johnson JP, McCauley CR, Copley JB, (1982) The quality of life of haemodialysis and transplant patients. Kidney International; 22(3): 286-291

Jones BM, Taylor F, *et al,* (1992) Longitudinal study of quality of life and psychological adjustment after cardiac transplantation. Medical Journal of Australia; 157(1): 24-26

Kalo Z, Jaray J, Nagy J, (2001) Economic evaluation of kidney transplantation versus hemodialysis in patients with end-stage renal disease in Hungary. Progress in Transplantation; 11(3): 188-193

Kamath P, Wiesner RH, *et al,* (2001) A model to predict survival in patients withend-stage liver disease. Hepatology; 33: 464-470

Kaminota M, (2001) Cost-effectiveness analysis of dialysis and kidney transplants in Japan. Keio Journal of Medicine; 50(2): 100-108

Kang SH, Kodell RL, Chen JJ, (2000) Incorporating model uncertainties along with data uncertainties in microbial risk assessment. Regulatory Toxicology and Pharmacology 32(1); 68-72

Kaplan EL, Meier P, (1958) Nonparametric estimation from incomplete observations. Journal of the American Statistical Association; 53: 457-481

Karam V, Castaing D, *et al,* (2003) Longitudinal prospective evaluation of quality of life in adult patients before and one year after liver transplantation. Liver Transplantation; 9(7): 703-711

Karlberg I. (1992) Cost analysis of alternative treatments in end-stage renal disease. Transplantation Proceedings; 24(1): 335

Karlberg I, Nyberg G. (1995) Cost-effectiveness studies of renal transplantation. International Journal of Technology Assessment in Health Care; 11(3): 611-622

Kawrakow I, (2004) The effect of Monte Carlo statistical uncertainties on the evaluation of dose distributions in radiation treatment planning. Physics in Medicine and Biology; 49: 1549-1556

Kendall M, Gibbons JD, (1990) *Rank correlation methods.* Oxford: Oxford University Press

Kiberd BA, Larson T, (2000) Estimating the benefits of solitary pancreas transplantation in nonuremic patients with type I diabeties mellitus. Transplantation; 70(7): 1121-1127

Kiebert GM, van Oosterhout EC, *et al,* (1994) Quality of life after combined kidney-pancreas or kidney transplantation in diabetic patients with end-stage renal disease. Clinical Transplantation; 8(3 Part 1): 239-245

Kober B, Kuchler T, *et al, (*1991) A psychological support concept and quality of life research in a liver transplantation program: an interdisciplinary multicenter study. Psychotherapy & Psychosomatics; 54(2 to 3): 117-131

Koch U, Muthny FA, (1991) Quality of life in patients with end-stage renal disease in relation to the method of treatment. Psychotherapy & Psychosomatics; 54(2 to 3): 161-171

Krakauer H, (1986) Assessment of alternative technologies for the treatment of end-stage renal disease. Israel Journal of Medical Sciences; 22(3 to 4): 245-259

Kugler C, Strueber M, *et al, (*2004) Quality of life 1-year after lung transplantation. Progress in Transplantation; 14(4): 331-336

Kunze E, Chang-Claude J, Frentzel-Beyme R, (1992) Life style and occupational risk factors for bladder cancer in Germany. A case-control study. Cancer; 69(7): 1776-1790

Kutner NG, Brogan D, Kutner MH, (1986) End-stage renal disease treatment modality and patients' quality of life. Longitudinal assessment. American Journal of Nephrology; 6(5): 396-402

Lanuza DM, Lefaiver C, *et al, (*2000) Prospective study of functional status and quality of life before and after lung transplantation Chest; 118(1): 115-122

Laupacis A, Keown P, *et al, (*1996) A study of the quality of life and cost-utility of renal transplantation. Kidney International; 50(1): 235-242

Laupacis A, Pus N, *et al, (*1993) Disease-specific questionnaire for patients with a renal transplant. Nephron; 64(2): 226-231

Levy MF, Jennings L, *et al, (*1995) Quality of life improvements at one, two, and five years after liver transplantation. Transplantation; 59(4): 515-518

Liermann Garcia RF, Evangelista Garcia C, *et al, (*2001) Transplantation for primary biliary cirrhosis: Retrospective analysis of 400 patients in a single centre. Hepatology; 33 (1): 22-27

Lietz K, John R, *et al,* (2004) Outcomes in cardiac transplant recipients using allografts from older donors versus mortality on the transplant waiting list: implications for donor selection criteria. Journal of the American College of Cardiology; 43(9): 1553-1561

Limbos MM, Chan CK, Kesten S, (1997) Quality of life in female lung transplant candidates and recipients. Chest; 112(5): 1165-1174

Limbos MM, Joyce DP, *et al,* (2000) Psychological functioning and quality of life in lung transplant candidates and recipients. Chest; 118(2): 408-416

Lin DY, Feuer EJ, *et al,* (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 (1): 35-47

Liou TG, Adler FR, *et al,* (2001) Survival effect of lung transplantation among patients with cystic fibrosis. JAMA; 286(21): 2683-2689

Lipscomb J, Ancukiewicz M, *et al,* (1998) Predicting the cost of illness: A comparison of alternative models applied to stroke. Medical Decision Making; 18 (Supplement): S39-S56

Little RJA (1993) Pattern-mixture models for multivariate incomplete data. Journal of the American Statistical Association; 88: 125-134

Little RJA, Rubin DB, (1987) *Statistical Analysis with Missing Data.* J Wiley & Sons, New York

Llovet JM, Fuster J, Bruix J, (1999) Intention to treat analysis of surgical treatment for early hepatocellular carcinoma: resection versus transplantation. Hepatology; 30(6): 1434-1440

Longworth L, Bryan S, (2003) An empirical comparison of the EQ-5D and the SF-6D in liver transplant patients. Health Economic Letters; 12: 1061-1067

Longworth L, Young T, Ratcliffe J. Bryan S. Buxton M. on behalf of the Cost-Effectiveness in Liver Transplantation (CELT) Team, (2001) Economic evaluation of the liver transplant programme in England and Wales: An assessment of the costs of liver transplantation. Final Report to Department of Health, Brunel University, Uxbridge

Longworth L, Young T, Ratcliffe J. Bryan S. Buxton M. on behalf of the Cost-Effectiveness in Liver Transplantation (CELT) Team (2002) Economic evaluation of the liver transplant programme in England and Wales: An assessment of the cost-effectiveness of liver transplantation for three liver disease groups. Final Report to Department of Health, Brunel University, Uxbridge

Longworth L, Young T, Buxton MJ. Ratcliffe J. Neuberger J. Burroughs A. Bryan S. CELT Project Team, (2003) Midterm cost-effectiveness of the liver transplantation program of England and Wales for three disease groups. Liver Transplantation; 9(12): 1295-307

Longworth L, Young T, Mistry H. Ratcliffe J. Buxton M, (2003a) An economic evaluation of small bowel transplantation for paediatric patients. Final Report to Department of Health, Brunel University, Uxbridge

Lopez-Neblina F, Alvarez Jimenez H, Finkelstein London I, (1999) High efficiency kidney transplantation: concept, technique, results, and cost analysis. Transplantation Proceedings; 31(7): 3025

Loubeau PR, Loubeau JM, *et al,* (2001) The economics of kidney transplantation versus hemodialysis Progress in Transplantation; 11(4): 291-297

Ludbrook A, (1981) A cost-effectiveness analysis of the treatment of chronic renal failure. Applied Economics; 13: 337-350

MacNaughton KL, Rodrigue JR, *et al,* (1998) Health-related quality of life and symptom frequency before and after lung transplantation. Clinical Transplantation; 12(4): 320-323

Madrigal G, (1994) Cost estimate of kidney transplants in Costa Rica: comparison to chronic dialysis. Transplantation Proceedings; 26(1): 121

Mai FM, McKenzie FM, Kostuk WJ, (1990) Psychosocial adjustment and quality of life following heart transplantation. Canadian Journal of Psychiatry - Revue Canadienne de Psychiatrie; 35(3): 223-227

Majno PE, Sarasin FP, *et al,* (2000) Primary liver resection and salvage transplantation or primary liver transplantation in patients with single, small hepatocellular carcinoma and preserved liver function: an outcome-oriented decision analysis. Hepatology; 31(4): 899-906
Manly BFJ, (1997) *Randomization, bootstrap and Monte Carlo methods in biology (2nd Edition).* Chapman & Hall, London, UK

Manning WG, Fryback DG, Weinstein MC, (1996) Reflecting uncertainty in cost-effectiveness analysis In *Cost-effectiveness in health and medicine.* Eds. Gold MR, Siegle JE, Russell LB, Weinstein MC. Oxford University Press, Oxford

Matas AJ, McHugh L, *et al,* (1998) Long-term quality of life after kidney and simultaneous pancreas-kidney transplantation. Clinical Transplantation; 12(3): 233-242

Matas AJ, Schnitzler MA, (2003) Payment for living donor (vendor) kidneys: a cost-effectiveness analysis. American Journal of Transplantation; 4(8): 216-221

McDonald SP, Craig JC, (2004) Long-term survival of children with end-stage renal disease. New England Journal of Medicine; 350(26): 2654-2662

McDonald SP, Russ GR, (2002) Survival of recipients of cadaveric kidney transplants compared with those receiving dialysis treatment in Australia and New Zealand, 1991-2001. Nephrology Dialysis Transplantation; 17(12): 2212-2219

McDonough CM, Grove MR, et al, (2005) Comparison of the EQ-5D, HUI, and SF-36 derived social health status values among spine patient outcomes research trial (SPORT) participants. Quality of Life Research; 14 (5): 1321-1332

Meirik O, (1993) Cohort and Case-control studies. http://www.gfmer.ch/Books/Reproductive_health/Cohort_and_case_control_studies.htmlAccessed June 2006

Mendez R, Asward S, *et al,* (1992) Cost and financing of kidney transplantation in the United States. Transplantation Proceedings; 24(5):2127-2128

Michel BC, van Hout BA, Bonsel GJ, (1994) Assessing the benefits of transplant services. Baillieres Clinical Gastroenterology; 8(3): 411-423

Milde FK, Hart LK, Zehr PS, (1995) Pancreatic transplantation. Impact on the quality of life of diabetic renal transplant recipients. Diabetes Care; 18(1): 93-95

Mistry H, Longworth L, Young T, Buxton MJ, (2003) Economic evaluation of the liver transplant programme in England and Wales: Costs for patients listed for a liver transplant attending outpatient appointments. Final report to the Department of Health, Brunel University, Uxbridge

Morris PL, Jones B, (1988) Transplantation versus dialysis: a study of quality of life. Transplantation Proceedings; 20(1): 23-26

Morris PL, Jones B, (1989) Life satisfaction across treatment methods for patients with end-stage renal failure. Medical Journal of Australia; 150(8): 428-432

Mullahy J. Manning W, (1995) Statistical issues in cost-effectiveness analysis. In *Valuing Health Care*, Editor: Sloan, F. Cambridge University Press, Cambridge,

Murtaugh PA, Dickson ER, *et al,* (1994) Primary biliary cirrhosis: prediction of short-term survival based on repeated patient visits. Hepatology; 20(1): 126-134

Muthny FA, Koch U, (1991) Quality of life of patients with end-stage renal failure. A comparison of hemodialysis, CAPD, and transplantation. Contributions to Nephrology; 89: 265-273

Nakache R, Tyden G, Groth CG, (1989) Quality of life in diabetic patients after combined pancreas-kidney or kidney transplantation. Diabetes; 38(Supplement 1): 40-42

Nakache R, Tyden G, Groth CG, (1994) Long-term quality of life in diabetic patients after combined pancreas-kidney transplantation or kidney transplantation. Transplantation Proceedings; 26(2): 510-511

Nathan DM, Fogel H, *et al,* (1991) Long-term metabolic and quality of life results with pancreatic/renal transplantation in insulin-dependent diabetes mellitus. Transplantation; 52(1): 85-91

National Institute for Clinical Excellence (NICE), (2004) Guide to the Methods of Technology Appraisal. NICE: London

National Institute of Health, (1984) Consensus Development Conference on Liver Transplantation – June 20-23, 1983 Hepatology; 4 (1 Supplement): 1S-110S.

National Library of Medicine, (2005) Medical Subject Headings (MeSH) Browser. Available from: http://www.nlm.nih.gov/mesh/MBrowser.html. Accessed August 2005

Neuberger J, Altman DG, *et al, (*1986) Use of a prognostic index in evaluation of liver transplantation for primary biliary cirrhosis. Transplantation; 4: 713-716

Neuberger J, Lucey M, Editors. (1994) *Liver Transplantation: Practice and Management.* BMJ Publishing Group, London

NHS Centre for Reviews and Dissemination. Khan KS, ter Riet G, *et al,* (2006) Stage II: Conducting the review - Phase 5: Study quality assessment. Available from: http://www.york.ac.uk/inst/crd/pdf/crd4_ph5.pdf Accessed April 2006

NHS National Institute for Clinical Excellence, (2003) Guidance for manufacturers and sponsors. Technology appraisals process series No 5. London 2001. Available at: http://www.nice.org.uk . Accessed October 2003

Nishimura R, Dorman JS, *et al,* (2003) Incidence of ESRD and survival after renal replacement therapy in patients with type 1 diabetes: a report from the Allegheny County Registry. American Journal of Kidney Disease; 42(1): 117-124

O'Brien B, Buxton MJ, Ferguson B, (1987) Measuring the effectiveness of heart transplant programmes: quality of life data and their relationship to survival analysis. Journal of Chronic Diseases; 40(Supplement 1): 137S-158S

O'Brien B, Banner NR, *et al,* (1988) The Nottingham Health Profile as a measure of quality of life following combined heart and lung transplantation. Journal of Epidemiology and Community Health; 42(3): 232-234

O'Brien BJ, Drummond MF, *et al,* (1994) In search of power and significance: issues in the design and analysis of stochastic cost-effectiveness studies in health care. Medical Care; 32(2): 150-163

O'Carroll RE, Smith K, *et al, (*2000) A comparison of the WHOQOL-100 and the WHOQOL-BREF in detecting change in quality of life following liver transplantation. Quality of Life Research; 9(1): 121-124

O'Grady JG, Williams R, (1986) An appraisal of liver transplantation in Great Britain. International Journal of Technology Assessment in Health Care; 2: 465-469

O'Hagan A, Stevens JW, (2004) On estimators of medical costs with censored data. Journal of Health Economics; 23(3): 615-625

Office of Health Economics: Health Economic Evaluation Database (OHE HEED), (2005) Available from: http://www.ohe-heed.com/about.htm. Accessed July 2005

Ohi G, Hasegawa T, Kumano H, (1986) Why are cadaveric renal transplants so hard to find in Japan? An analysis of economic and attitudinal aspects. Health Policy; 6(3):269-278

Oostenbrink JB, Al MJ, Rutten-van Molken MPMH, (2003) Methods to analyse cost data of patients who withdraw in a clinical trial setting. Pharmacoeconomics; 21(15): 1103-1112

Oostenbrink JB, Al MJ, (2005) The analysis of incomplete cost data due to dropout. Health Economics; 14: 763-776

Ost L, Groth CB, *et al,* (1980) Cadaveric renal transplantation in patients of 60 years and above. Transplantation; 30(5): 339-340

Ouwens JP, van Enckevort PJ, *et al.* (2003) The cost effectiveness of lung transplantation compared with that of heart and liver transplantation in the Netherlands. Transplant International; 16(2): 123-127

OVID, (2005) Available from: http://gateway.uk.ovid.com/gw2/ovidweb.cgi. Accessed July 2005

Packa DR, (1989) Quality of life of adults after a heart transplant. Journal of Cardiovascular Nursing; 3(2):12-22

Pasha TM, Dickson ER, (1997) Survival algorithms and outcome analysis in primary biliary cirrhosis. Seminars in Liver Disease; 17(2):147-158

Parfrey PS, Vavasour HM, *et al,* (1987) Symptoms in end-stage renal disease: dialysis v transplantation. Transplantation Proceedings; 19(4): 3407-3409

Parfrey PS, Vavasour HM, *et al,* (1988) Clinical features and severity of nonspecific symptoms in dialysis patients. Nephron; 50(2): 121-128

Parfrey PS, Vavasour HM, Gault MH. (1988a) A prospective study of health status in dialysis and transplant patients. Transplantation Proceedings; 20(6): 1231-1232

Parfrey PS, Vavasour HM, *et al,* (1989) Development of a health questionnaire specific for end-stage renal disease. Nephron; 52(1): 20-28

Park H, Bang WR, *et al,* (1992) Quality of life of ESRD patients: development of a tool and comparison between transplant and dialysis patients. Transplantation Proceedings; 24(4): 1435-1437

Park IH, Yoo HJ, *et al,* (1996) Changes in the quality of life before and after renal transplantation and comparison of the quality of life between kidney transplant recipients, dialysis patients, and normal controls. Transplantation Proceedings; 28(3): 1937-1938

Payne JL, McCarty KR, *et al, (*1996) Outcomes analysis for 50 liver transplant recipients: the Vanderbilt experience. American Surgeon; 62(4): 320-325

Piehlmeier W, Bullinger M, *et al,* (1991) Quality of life in type 1 (insulin-dependent) diabetic patients prior to and after pancreas and kidney transplantation in relation to organ function. Diabetologia; 34(Supplement 1): S150-S157

Piehlmeier W, Bullinger M, *et al,* (1992) Quality of life in diabetic patients prior to or after pancreas transplantation in relation to organ function. Transplantation Proceedings; 24(3): 871-873

Piehlmeier W, Bullinger M, *et al,* (1996) Evaluation of the quality of life of patients with insulin-dependent diabetes mellitus before and after organ transplantation with the SF 36 health survey. European Journal of Surgery; 162(12): 933-940

Pietrabissa A, Ciaramella A, *et al,* (1992) Effect of kidney transplantation on quality of life measures. Transplant International; 5(Supplement 1): S708-S710

Pocock SJ, (1983) *Clinical Trials: A practical approach.* John Wiley & Sons, Chichester, UK

Polsky D, Glick HA, *et al,* (1997). Confidence intervals for cost-effectiveness ratios: a comparison of four methods. Health Economics; 6: 243-52

Poses RM, McClish DK, *et al,* (1991) Ego bias, reverse ego bias, and physicians' prognostic. Critical Care Medicine; 19(12): 1533-1539

Poynard T, Barthelemy P, *et al,* (1994) Evaluation of efficacy of liver transplantation in alcoholic cirrhosis by a case-control study and simulated controls. The Lancet; 344: 502-507
Poynard T, Naveau S, Doffoel M, (1999) Evaluation of efficacy of liver transplantation in alcoholic cirrhosis using matched and simulated controls. Journal of Hepatology; 30: 1130-1137

Price CE, Lowe D, *et al, (*1995) Prospective study of the quality of life in patients assessed for liver transplantation: outcome in transplanted and not transplanted groups. Journal of the Royal Society of Medicine; 88(3): 130-135

Raikou, M, McGuire A, (2004) Estimating medical care costs under conditions of censoring. Journal of Health Economics; 23(3): 443-470

Ramsey S, Larson P, (1998) Lung Transplantation. The Lancet; 351(9111) 1285

Ramsey S, Patrick DL, *et al, (*1995a) The cost-effectiveness of lung transplantation. A pilot study. Chest; 108(6): 1594-1601

Ramsey S, Patrick DL, *et al, (*1995b) Improvement in quality of life after lung transplantation: a preliminary study. The University of Washington Medical Center Lung Transplant Study Group. Journal of heart and lung transplantation; 14(5): 870-877

Ratcliffe J, Young T, Buxton M. Eldabi T. Paul R. Burroughs A. Papatheodoridis G. Rolles K, (2001) A simulation modelling approach to evaluating alternative policies for the management of the waiting list for liver transplantation. Health Care Management Science; 4: 117-124

Ratcliffe J, Longworth L, Young T. on behalf of the Cost-Effectiveness in Liver Transplantation (CELT) Team, (2001) Economic evaluation of the liver transplant programme in England and Wales: An assessment of quality of life pre and post liver transplantation. Final report to Department of Health, Brunel University, Uxbridge

Ratcliffe J, Longworth L, Young T. Bryan S. Burroughs A. Buxton M. on behalf of the Cost-Effectiveness of Liver Transplantation Team, *(*2002) Assessing health-related quality of life pre- and post-liver transplantation: a prospective multicenter study. Liver Transplantation; 8(3): 263-270

Ratcliffe J, Young T, Longworth L, Buxton M (2005) An assessment of the impact of informative dropout and nonresponse in measuring health-related quality of life using the euroqol (EQ-5D) descriptive system. Value in Health; 8(1): 53-58

Rebollo P, Ortega F, *et al,* (1998) Health-related quality of life (HRQOL) in end stage renal disease (ESRD) patients over 65 years. Geriatric Nephrology & Urology; 8(2): 85-94

Rebollo P, Ortega F, *et al,* (2000) Health related quality of life (HRQOL) of kidney transplanted patients: variables that influence it. Clinical Transplantation; 14(3): 199-207

Rector TS, Ormaza SM, Kubo SH, (1993) Health status of heart transplant recipients versus patients awaiting heart transplantation: a preliminary evaluation of the SF-36 questionnaire. Journal of Heart and Lung Transplantation; 12(6 Part 1): 983-986

Reddy KS, Stablein D, *et al,* (2003) Long-term survival following simultaneous kidney-pancreas transplantation versus kidney transplantation alone in patients with type 1 diabetes mellitus and renal failure. American Journal of Kidney Disease; 41(2): 464-470

Riether AM, Smith SL, *et al,* (1992) Quality-of-life changes and psychiatric and neurocognitive outcome after heart and liver transplantation. Transplantation; 54(3): 444-450

Roberts SD, Maxwell DR, Gross L, (1980) Cost-effective care of end-stage renal disease: a billion dollar question. Annals of Internal Medicine; 92(1): 243-248

Rodin G, Voshart K, et al, (1985/6) Cadaveric renal transplant failure: the short-term sequelae. International Journal of Psychiatry in Medicine; 15(4): 3573-64

Rovera GM, DiMartini A, *et al, (*1998) Quality of life after intestinal transplantation and on total parenteral nutrition. Transplantation Proceedings; 30(6): 2513-2514

Rovera GM, DiMartini A, *et al, (*1998a) Quality of life of patients after intestinal transplantation. Transplantation; 66(9): 1141-1145

Rubin DB (1987) *Multiple Imputation for Nonresponse in Surveys.* J. Wiley & Sons, New York

Rufat P, Fourquet F, *et al,* (1999) Costs and outcomes of liver transplantation in adults: a prospective, 1-year, follow-up study. Transplantation; 68 (1): 76-83

Russell JD, Beecroft ML, *et al,* (1992) The quality of life in renal transplantation--a prospective study. Transplantation; 54(4): 656-660

Sagmeister MB, Mullhaupt B, *et al,* (2002) Cost-effectiveness of cadaveric and living donor liver transplantation. Transplantation; 73(4): 616-622

Salonen T, Reina T, *et al,* (2003) Cost analysis of renal replacement therapies in Finland. American Journal of Kidney Disease; 42(6): 1228-1238

Sarasin FP, Giostra E, *et al, (*1998) Partial hepatectomy or orthotopic liver transplantation for the treatment of resectable hepatocellular carcinoma: a cost-effectiveness perspective. Hepatology; 28(2): 436-442

Sayag R, Kaplan De-Nour A, *et al,* (1990) Comparison of psychosocial adjustment of male nondiabetic kidney transplant and hospital hemodialysis patients. Nephron; (54)3: 214-218

Schafer JL, (1997) *Analysis of incomplete multivariate data.* Monographs on statistics and applied probability 72. Chapman & Hall/CRC

Schafer, JL. (1999) NORM: Multiple imputation of incomplete multivariate data under a normal model, version 2. Software for Windows 95/98/NT, available from http://www.stat.psu.edu/~jls/misoftwa.html Accessed September 2004

Schafer, JL (2007) The multiple imputation FAQ page. http://www.stat.psu.edu/~jls/mifaq.html Accessed February 2007

Schaubel DE, Desmeules M, *et al,* (1995) Survival experience among elderly end-stage renal disease patients. A controlled comparison of transplantation and dialysis. Transplantation; 60(12): 1389-1394

Schnitzler MA, Smith C, *et al,* (1999) Relative cost of cadaveric versus living donor kidney transplantation. Transplantation; 67(7): S189

Schweitzer EJ, Wiland A, *et al,* (1998) The shrinking renal replacement therapy "break-even" point. Transplantation; 66(12): 1702-1708

Secchi A, Di Carlo V, *et al,* (1991) Effect of pancreas transplantation on life expectancy, kidney function and quality of life in uraemic type 1 (insulin-dependent) diabetic patients. Diabetologia; 34(Supplement 1): S141-S144

Secchi A, Martinenghi S, *et al,* (1998) Effects of pancreas transplantation on quality of life in type I diabetic patients undergoing kidney transplantation. Transplantation Proceedings; 30(2): 339-342

Seedat YK, MacIntosh CG, Subban JV, (1987) Quality of Life for patients in an end-stage renal disease programme. South African Medical Journal; 71(8): 500-504

Sesso R, Eisenberg JM, *et al,* (1990) Cost-effectiveness analysis of the treatment of end-stage renal failure in Brazil. IJTAHC; 6(1): 107-114

Shabahang M, Franceschi D, *et al, (*2002) Comparison of hepatic resection and hepatic transplantation in the treatment of hepatocellular carcinoma among cirrhotic patients. Annals of Surgical Oncology; 9(9): 881-886

Shih FJ, Lee PH, *et al,* (1999) Changes in quality of life and working capacity before and after kidney transplantation. Transplantation Proceedings; 31(5): 1981-1984

Shih FJ, Tsao CI, *et al, (*2002) The context framing the changes in health-related quality of life and working competence before and after lung transplantation: one-year follow-up in Taiwan. Transplantation Proceedings; 34(7): 2801-2806

Shih FJ, Tsao CI, *et al,* (2003) Changes in health related quality of life and working competence before and after heart transplantation: one-year follow-up in Taiwan. Transplantation Proceedings; 35(1): 466-471

Shum-Tim D. Pelletier MP, *et al,* (1999) Transplantation versus coronary artery bypass in patients with severe ventricular dysfunction. Surgical outcome and quality of life. Journal of Cardiovascular Surgery; 40(6): 773-780

Simmons RG, Anderson C, Abress LK, (1990) Quality of life and rehabilitation differences among four end-stage renal disease therapy groups. Scandinavian Journal of Urology & Nephrology Supplementum; 131: 7-22

S-PLUS 6 (2001) S-PLUS 6 for Windows Guide to Statistics, Volume 1 & 2 (2001) Insightful Corporation, Seattle, WA

Spiegelhalter DJ, Abrams KR, Myles JP, (2004) Bayesian approaches to clinical trials and health-care technologies. John Wiley & Sons Ltd, Chichester, UK

Spiegelhalter DJ, Thomas A, Best NG, (1999) *WinBUGS Version 1.2 User Manual*

Starzl TE, Marchioro TL, *et al,* (1963) Homotransplantation of the liver in humans. Surgery Gynaecology and Obstetrics; 117: 659-676

Stavem K, Bjortuft O, *et al, (*2000) Health-related quality of life in lung transplant candidates and recipients. Respiration; 67(2): 159-165

Stavern K, Froland SS, Hellum KB, (2005) Comparison of preference based utilities of the 15D, EQ-5D and SF-6D in patients with HIV/AIDS. Quality of Life Research; 14(4): 971-980

Stratta RJ, Taylor RJ, *et al,* (1993) The analysis of benefit and risk of combined pancreatic and renal transplantation versus renal transplantation alone. Surgery, Gynecology & Obstetrics; 177(2): 163-171

Stratta RJ, Taylor RJ, *et al,* (1993) Combined pancreas-kidney transplantation versus kidney transplantation alone: analysis of benefit and risk. Transplantation Proceedings; 25(1 Part 2): 1298-1301

Strawderman RL, (2000) Estimating the mean of an increasing stochastic process at a censored stopping time. Journal of the American Statistical Association; 95(452): 1192-1208

Sureshkumar KK, Mubin T, *et al,* (2002) Assessment of quality of life after simultaneous pancreas-kidney transplantation. American Journal of Kidney Disease; 39(6): 1300-1306

Sutton AJ, Jones DR, *et al,* (1998) Systematic reviews of randomised trials. *In Health Service Research Methods: A guide to best practice.* Eds. Black N, Brazier J, *et al,* BMJ Books, London

Tarter RE, Erb S, *et al, (*1998) The quality of life following liver transplantation: a preliminary report. Gastroenterology Clinics of North America; 17(1): 207-217

Tarter RE, Switala J, *et al, (*1991) Quality of life before and after orthotopic hepatic transplantation. Archives of Internal Medicine; 151(8): 1521-1526

TenVergert EM, Essink-Bot ML, *et al, (*1998) The effect of lung transplantation on health-related quality of life: a longitudinal study. Chest; 113(2): 358-364

TenVergert EM, Vermeulem KM, *et al, (*2001) Quality of life before and after lung transplantation in patients with emphysema versus other indications. Psychological Reports; 89(3): 707-717

The organ procurement and transplant network (OPTN), (2005) Available at: http://www.optn.org/latestData/rptData.asp. Accessed August 2005

The Royal College of Surgeons, (1999) The Report of the Working Parry to Review Organ Transplantation 1999. Available at: *http://www.rcseng.ac.uk/services/publications/publications/?*pub_id=8. Accessed November 2002

Thomson NM, Scott DF, *et al,* (1989) Morbidity, mortality, and quality of life in long-term survivors of an integrated dialysis/renal transplant programme Transplantation Proceedings; 29(1 Part 2): 2184-2185

Tomasz W, Piotr S, (2003) A trial of objective comparison of quality of life between chronic renal failure patients treated with hemodialysis and renal transplantation. Annals of Transplantation; 8(2): 47-53

Tousignant P, Guttmann RD, Hollomby DJ, (1985) Transplantation and home haemodialysis: their cost-effectiveness. Journal of Chronic Diseases; 38(7): 589-601

Turnbull BW, Brown BW, Hu M, (1974) Survivorship analysis of heart transplant data. Journal of the American Statistical Association; 69(345): 74-80

UK Transplant, (2002) National protocol for the assessment of cardiothoracic transplant patients. http://www.uktransplant.org.uk/ukt/about_transplants/organ_allocation/cardiothoracic/national_protocols_and_standards/protocols_and_standards/patient_assessment.jsp Accessed September 2005

UK Transplant, (2004) More transplants – new lives: Transplant Activity in the UK 2003 – 2004. Available at: http://www.uktransplant.org.uk/ukt/statistics/transplant_activity report/current_activity_reports.jsp/ukt/tx_activity_report_2005_uk_complete-v2.pdf . Accessed May 2006

UK Transplant, (2005) Liver transplant services: Donor organ use – Protocols and guidelines for adults undergoing cadaveric liver transplantation. http://www.uktransplant.org.uk/ukt/about_transplants/organ_allocation/liver/national_protocols_and_guidelines/protocols_and_guidelines/adults.jsp#2 Accessed March 2006

UK Transplant, (2005a) Donor Organ Sharing Scheme: Operating Principles for Liver Transplant Units in the UK and Republic of Ireland. http://www.uktransplant.org.uk/ukt/about_transplants/organ_allocation/liver/liver_organ_sharing_principles/liver_organ_sharing_principles.jsp Accessed March 2006

UK Transplant, www.uktransplant.org.uk Accessed March 2006

UK Transplant, http://www.uktransplant/ukt/about_transplant/organ_allocation/liver/liver.jsp Accessed January 2007

University of Sheffield Library, (2005) STAR – the catalogue of the University of Sheffield Library. Available from: http://library.shef.ac.uk/. Accessed July 2005

Van Enckevort PJ, Koopmanschap MA, Tenvergert EM. (1997) Lifetime costs of lung transplantation: estimation of incremental costs. Health Economics; 6 (5): 479-489

Van Enckevort PJ, TenVergert EM, *et al,* (1998) Technology assessment of the Dutch lung transplantation program. International Journal of Technology Assessment in Health Care; 14(2): 344-356

Van Hout B, Bonsel G J, *et al,* (1993) Heart transplantation in the Netherlands: costs, effects and scenarios. Journal of Health Economics; 12(1): 73-93

Van Hout BA, Al MJ, *et al,* (1994) Costs, effects and C/E ratios alongside a clinical trial. Health Economics; 3(5): 309-19

Venstrom JM, McBride MA, *et al,* (2003) Survival after pancreas transplantation in patients with diabetes and preserved kidney function. JAMA; 290(21): 2817-2823

Vermeulen KM, Ouwens JP, *et al,* (2003) Long-term quality of life in patients surviving at least 55 months after lung transplantation. General Hospital Psychiatry; 25(2): 95-102

Vermeulen KM, van der Bij W, *et al,* (2004) Improved quality of life after lung transplantation in individuals with cystic fibrosis. Pediatric Pulmonology; 37(5): 419-426

Waiser J, Budde K, *et al,* (1998) The quality of life in end stage renal disease care. Transplant International; 11(Supplement 1): S42-S45

Walden J A, Stevenson LW, *et al,* (1989) Heart transplantation may not improve quality of life for patients with stable heart failure. Heart & Lung; 18(5): 497-506

Walden JA, Stevenson LW, *et al,* (1994) Extended comparison of quality of life between stable heart failure patients and heart transplant recipients. Journal of Heart and Lung Transplantation; 13(6): 1109-1118

Wang D, Zhang W, Bakhai A, (2004) Comparison of Bayesian model averaging and stepwise methods for model selection in logistic regression. Statistics in Medicine 23(22); 3451-67

Ware J, Snow KK, *et al,* (1993) *SF-36 Health Survey, manual and interpretation guide.* The Health Institute, New England Medical Centre, Boston, USA

Whiting JF, Zavala EY, *et al,* (1999) The cost-effectiveness of transplantation with expanded donor kidneys. Transplantation Proceedings; 31(1 to 2); 1320-1321

Weisner RH, Porayko MK, *et al,* (1989) Primary sclerosing cholangitis: Natural history, prognostic factors and survival. Hepatology; 10(4): 430-436

Wight J, Edwards L, *et al,* (1998) The SF36 as an outcome measure of services for end stage renal failure. Quality in Health Care; 7(4): 209-221

Willan AR, Chen AB, *et al,* (2003) Incremental net benefit in randomized clinical trials with quality-adjusted survival. Statistics in Medicine; 22(3): 353-362

Willan A. Lin DY, (2001) Incremental net benefit in randomized clinical trials. Statistics in Medicine; 20(11): 1563-1574

Willan A. Lin DY, *et al,* (2002) Using inverse-weighting in cost-effectiveness analysis with censored data. Statistical Methods in Medical Research; 11(6): 539-551

Willan AR, Lin DY, Manca A, (2004) Regression methods for cost-effectiveness analysis with censored data. Statistics in Medicine; 24(1): 131-245

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

Williams JW, Santiogo V, Evans LS, (1987) Socioeconomic aspects of hepatic transplantation. The American Journal of Gastroenterology; 82 (11): 1115-1119

Wolfe RA, Ashby V, *et al,* (1999) Comparison of mortality in all patients on dialysis, patients on dialysis awaiting transplantation, and recipients of a first cadaveric transplant. New England Journal of Medicine; 341(23): 1725-1730

Wu YT, Chien CL, *et al,* (2002) Quality-of-life outcome in cardiac transplantation versus coronary artery bypass patients. Transplantation Proceedings; 34(4): 1269-1270

Yoshimura N, Ohmori Y, *et al,* (1994) Quality of life in renal transplant recipients treated with cyclosporine in comparison with hemodialysis maintenance. Transplantation Proceedings; 26(5): 2542-2543

Young TA, (2005) Estimating mean total costs in the presence of censoring: A comparative assessment of methods. Pharmacoeconomics; 23(12): 1229-1242

Young TA, Longworth L, Ratcliffe J. on behalf of the Cost-Effectiveness in Liver Transplantation (CELT) Team, (2001) Economic evaluation of the liver transplant programme in England and Wales: Survival on the waiting list, post transplant, and estimated survival in the absence of transplantation. Final report to Department of Health, Brunel University, Uxbridge

Young TA, Neuberger J, Longworth L. Ratcliffe J. Buxton M on behalf of the Cost-effectiveness in Liver Transplantation (CELT) Study Team, (2003) Survival gain after liver transplantation for patients with alcoholic liver disease: a comparison across models and centers. Transplantation; 76(10): 1479-1486

Young TA, Thompson S, (2004) The importance of accounting for the uncertainty of published prognostic model estimates. International Journal of Technology Assessment in Health Care; 20(4): 481-487

Younossi ZM, McCormick M, *et al,* (2000) Impact of liver transplantation on health related quality of life. Liver Transplantation; 6 (6): 779-783

Zehrer CL, Gross CR, (1994) Comparison of quality of life between pancreas/kidney and kidney transplant recipients: 1-year follow-up. Transplantation Proceedings; 26(2): 508-509

Zhao H, Tian L, (2001) On estimating medical cost and incremental cost-effectiveness ratios with censored data. Biometrics; 57: 1002-1008

Zhao H, Tsiatis A, (1997) A consistent estimator for the distribution of quality adjusted survival time. Biometrika; 84(2): 339-348

APPENDIX A2.1 ATTRIBUTING CELT ASSESSMENT COSTS FOR NON-TRANSPLANT PATIENTS ACROSS THE TRANSPLANT COHORT

Patients who were assessed but not listed for transplantation, who were listed for transplantation and died on the waiting list, or who were removed from the waiting list prior to transplantation, were excluded from the main CELT analysis. The reason for the exclusion of these patients was that their QALYs and costs would be the same with and without the transplant programme. The exception to this is the assessment costs, which are attributable to the transplant programme. Rather than ignoring these costs, they are attributed across the transplant patients as set out below.

Table A2.1.1 lists the number of patients who were excluded from the main CELT analysis because they were assessed and not listed, removed from the waiting list, or died on the waiting list. Forty seven per cent of ALD patients who were assessed did not receive a transplant during the study period; the majority of these patients were not listed for transplantation. The percentage of assessed patients who did not receive a transplant during the CELT study is much lower for the PBC (34%) and PSC (36%) groups in comparison with the ALD group.

Table A2.1.1 Number of patients assessed in the CELT analysis by disease group

	ALD	PBC	PSC
Total number assessed	**155**	**122**	**70**
Transplanted patients	82	81	45
Patients not listed for transplant	55	28	17
Patients removed from the waiting list	9	5	3
Patients who died on the waiting list	9	8	5

The PSC group will be used to illustrate how the assessment costs for excluded patients were attributed to PSC transplant patients. The total assessment costs to the transplant programme that arose from 25 patients with end-stage PSC who were assessed but did not receive a liver transplant during the CELT study was £206,137 (Table A2.1.2). The total assessment cost for excluded patients was then divided by the number of transplant patients (N = 45) to give a value of £4,581. This value (£4,581) is then added to each transplant patients total study costs, thus attributing the assessment costs for patients who did not receive a transplant during the CELT study period over the transplant patients.

Table A2.1.2 Total assessment costs for PSC patients who did not receive a liver transplant

	N	Total assessment costs
PSC patients not listed for transplantation	17	£122,247
PSC patients removed from the waiting list	3	£15,690
PSC patients who died on the waiting list	5	£68,200
Total	**25**	**£206,137**

APPENDIX A2.2 ESTIMATING COSTS PRIOR TO DEATH

The CELT study team hypothesised that the cost of end-stage liver disease would increase in the month prior to death. To establish whether the hypothesis was true, the cost information from 44 CELT patients with end-stage liver disease who died on the waiting list was examined. Costs were examined over the waiting list period in monthly intervals from month of death, back in time, until point of listing. The median time on the waiting list for the 44 patients prior to death was two months (Range 1 to 14 months).

The cohort of patients who died on the waiting list included patients with liver diseases other than ALD, PBC or PSC, in order to maximise the number of patients included in this exercise. The majority of patients were male and over three quarters of the sample were classified as elective cases (Table A2.2.1).

Table A2.2.1 Demographic characteristics of the 44 CELT patients who died on the liver transplant waiting list

Mean age in years (SD)	51 (11.1)
Males (%)	28 (64%)
Liver disease (%):	
ALD	9 (20%)
PBC	8 (18%)
PSC	5 (11%)
Other diseases	22 (50%)
Elective cases (%)	34 (77%)

Table A2.2.2 details the mean and median costs per month from point of death, back in time, towards point of listing. Both mean and median costs were highest in the month immediately prior to death (30 days immediately prior to transplant). The differences between the mean and median costs per month show that the cost data is highly skewed. Therefore, non-parametric tests were used to test for statistical differences in costs between months one and two, months two and three, and months three and four. Median costs in the month prior death were significantly higher than costs two months prior to death ($p = 0.049$). However, there was no evidence of a significant differences in costs between months two and three, or months three and four prior to death ($p = 0.738$, $p = 0.791$, respectively).

Table A2.2.2 Mean and median monthly costs on the waiting list from time of death

	N	Mean cost	Median cost
In 30 days immediately prior to transplant	44	£3,893	£2,476
30-60 days prior to transplant	22	£2,555	£203
60-90 days prior to transplant	14	£1,549	£50
90-120 days prior to transplant	11	£903	£99

A linear regression model was then fitted to the cost data to predict cost of treatment for end-stage liver disease in the month prior to death. The model was fitted to the 44 patients with end-stage liver diseases who died on the waiting list. Age, gender, disease group[1], type of transplant (emergency or elective) and centre were considered in the model to adjust for case mix. Variables were excluded from the model if they were non-significant predictors of cost at the 5% level (p-value).

Table A2.2.3 presents the results of the regression model. Costs in the month prior to death increased with age, and were higher for emergency cases than for elective cases. Patients with parenchymal liver diseases (this group includes ALD patients) were more likely to have higher costs than other disease groups. Disease group, transplant centre, and the constant term was not statistically significant and was removed from the model.

Table A2.2.3 Results of regression model for predicting costs in the month prior to death

	Regression coefficient	**Standard error**	**t-value**	**P-value**
Age in years	101.6	18.65	5.45	< 0.001
Elective	Base*			
Emergency	13,453.2	2,562.25	5.25	< 0.001
Parenchymal	Base*			
Fulminent	-15,350.7	3,013.17	-5.10	< 0.001
Cholestatic	-1,676.4	1,493.41	-1.12	0.272
Other	-1,940.3	2,292.40	-0.85	0.405

* Base case scenario Adjusted R^2 = 0.744 Root MSE = 3,653.30

The model presented in Table A2.2.3 was used to adjust the non-transplant costs for CELT patients not expected to survive for the full 2.25 year study period.

[1] Disease group 1 = parenchymal (includes ALD), 2 = fulminant, 3 = cholestatic (includes PBC, PSC), 4 = other

APPENDIX A3.1 LITERATURE REVIEW SEARCH STRATERGY

A list of search terms are presented in sections A3.1, A3.2 and A3.3. Section A3.1 presents a list of terms relating to organ transplantation and Section A3.2 a list of search terms relating to the seven organs included in this review. Section A3.3 lists the search terms applied in order to identify different types of study. Potentially relevant studies included at least one term from each of the three lists (A3.1, A3.2 and A3.3) in the abstract, title or keywords.

e.g. "transplant" AND "kidney" AND "cost-utility"
e.g. "graft" AND "lung" AND "mortality"

A3.1 Transplantation

Graft
Grafting
Transplant
Transplantation
Transplants

A3.2 Transplant organs

Bowel
Cardiac
Heart
Heart-Lung
Hepatic
Intestinal
Intestine
Kidney
Liver
Lung
Pancreas
Pancreatic
Renal

A3.3 Type of study

Benefits

Benefits and costs

Charges

Cost

Cost analysis

Costs and benefits

Costs and cost analysis

Cost-benefit analysis

Cost-benefit data

Cost-effective

Cost-effectiveness

Cost measures

Cost-utility

Cost-utilities

Costing

Death rate

Expenses

Mortality

Outcome

Pricing

Quality of life

Survival

Surviving

Utilities

Utility

APPENDIX A3.2 SOLID ORGAN TRANSPLANTATION LITERATURE REVIEW

This appendix lists the references for 202 articles included in the literature review detailed in Chapter 3. Articles are listed by organ.

Heart Transplantation

Bocchi EA, Bellotti G, *et al,* (1996) Mid-term results of heart transplantation, cardiomyoplasty, and medical treatment of refractory heart failure caused by idiopathic dilated cardiomyopathy. Journal of Heart and Lung Transplantation; 15(7): 736-745

Bortman G, Delgardo D, *et al,* (1999) Analysis of quality of life before and after heart transplantation. Transplantation Proceedings; 31(6): 2555

Buxton MJ, Acheson R, *et al,* (1985) Costs and benefits of the heart transplant programme at Harefield and Papworth hospitals. Department of Health and Social Security; Report No. 12: HMSO

Caine N, Sharples LD, *et al,* (1990) Prospective study comparing quality of life before and after heart transplantation. Transplantation Proceedings; 22(4): 1437-1439

Caine N, Sharples LD, *et al,* (1996) Measurement of health-related quality of life before and after heart-lung transplantation. Journal of Heart and Lung Transplantation; 15 (10): 1047-1058

Clark DA, Stinson EB, *et al,* (1971) Cardiac Transplantation in Man: VI. Prognosis of patients selected for cardiac transplantation. Annals of Internal Medicine; 75(1); 15-21

Cope JT, Kaza AK, *et al,* (2001) A cost comparison of heart transplantation versus alternative operations for cardiomyopathy. Annals of Thoracic Surgery; 72(4): 1298-1305

Cotrufo M, Romano G, *et al,* (2005) Treatment of extensive ischemic cardiomyopathy: quality of life following two different surgical strategies. European Journal of Cardio-Thoracic Surgery; 27(3): 481-487

Evangalista LS, Moser D, *et al,* (2004) Functional status and perceived control influence quality of life in female heart transplant recipients. Journal of Heart and Lung Transplantation; 23(3): 360-367

Fisher DC, Lake KD, *et al,* (1995) Changes in health-related quality of life and depression in heart transplant recipients. Journal of Heart and Lung Transplantation; 14(2): 373-381

Gajarski RJ, Towbin JA, Garson A, (1996) Fontan palliation versus heart transplantation: a comparison of charges. American Heart Journal; 131(6): 1169-1174

Grady KL, Jalowiec A, White-Williams C, (1996) Improvement in quality of life in patients with heart failure who undergo transplantation. Journal of Heart and Lung Transplantation; 15(8): 749-757

Haberman S, (1980) Heart transplants: putting a price on life. Health and Social Service Journal; 90(4700): 877-879

Hellinger FJ. (1982) An analysis of a public programme for heart transplantation; Journal of Human Resources; 17(2): 307-313

Hummel *et al,* (2000) How expensive is heart transplantationn (HTX) when compared with treatment of end stage heart failure (ESHF) before HTx? A single centre analysis of 350 patients. Journal of Heart and Lung Transplantation; 19(1): 44

Jones BM, Taylor F, *et al,* (1992) Longitudinal study of quality of life and psychological adjustment after cardiac transplantation. Medical Journal of Australia; 157(1): 24-26

Lietz K, John R, *et al,* (2004) Outcomes in cardiac transplant recipients using allografts from older donors versus mortality on the transplant waiting list: implications for donor selection criteria. Journal of the American College of Cardiology; 43(9): 1553-1561

Mai FM, McKenzie FM, Kostuk WJ, (1990) Psychosocial adjustment and quality of life following heart transplantation. Canadian Journal of Psychiatry - Revue Canadienne de Psychiatrie; 35(3): 223-227

O'Brien B, Banner NR, *et al,* (1988) The Nottingham Health Profile as a measure of quality of life following combined heart and lung transplantation. Journal of Epidemiology and Community Health; 42(3): 232-234

O'Brien B, Buxton MJ, Ferguson B, (1987) Measuring the effectiveness of heart transplant programmes: quality of life data and their relationship to survival analysis. Journal of Chronic Diseases; 40(Supplement 1): 137S-158S

Packa DR. (1989) Quality of life of adults after a heart transplant. Journal of Cardiovascular Nursing; 3(2):12-22

Rector TS, Ormaza SM, Kubo SH, (1993) Health status of heart transplant recipients versus patients awaiting heart transplantation: a preliminary evaluation of the SF-36 questionnaire. Journal of Heart and Lung Transplantation; 12(6 Part 1): 983-986

Riether AM, Smith SL, *et al,* (1992) Quality-of-life changes and psychiatric and neurocognitive outcome after heart and liver transplantation. Transplantation; 54(3): 444-450

Shih FJ, Tsao CI, *et al,* (2003) Changes in health related quality of life and working competence before and after heart transplantation: one-year follow-up in Taiwan. Transplantation Proceedings; 35(1): 466-471

Shum-Tim D. Pelletier MP, *et al,* (1999) Transplantation versus coronary artery bypass in patients with severe ventricular dysfunction. Surgical outcome and quality of life. Journal of Cardiovascular Surgery; 40(6): 773-780

Turnbull BW, Brown BW, Hu M, (1974) Survivorship analysis of heart transplant data. Journal of the American Statistical Association; 69(345): 74-80

Van Hout B, Bonsel G J, *et al,* (1993) Heart transplantation in the Netherlands: costs, effects and scenarios. Journal of Health Economics; 12(1): 73-93

Walden J A, Stevenson LW, *et al,* (1989) Heart transplantation may not improve quality of life for patients with stable heart failure. Heart & Lung; 18(5): 497-506

Walden JA, Stevenson LW, *et al,* (1994) Extended comparison of quality of life between stable heart failure patients and heart transplant recipients. Journal of Heart and Lung Transplantation; 13(6): 1109-1118

Wu YT, Chien CL, *et al,* (2002) Quality-of-life outcome in cardiac transplantation versus coronary artery bypass patients. Transplantation Proceedings; 34(4): 1269-1270

Intestinal Transplantation

DiMartini A, Rovera GM *et al,* (1998) Quality of life after small intestinal transplantation and among home parenteral nutrition patients. Journal of Parenteral and Enteral Nutrition; 22(6): 357-362

Longworth L, Young T, *et al,* (2003a) An economic evaluation of small bowel transplantation for paediatric patients. Final report to the Department of Health, Brunel University, Uxbridge

Rovera GM, DiMartini A, *et al,* (1998) Quality of life after intestinal transplantation and on total parenteral nutrition. Transplantation Proceedings; 30(6): 2513-2514

Rovera GM, DiMartini A, *et al,* (1998a) Quality of life of patients after intestinal transplantation. Transplantation; 66(9): 1141-1145

Kidney Transplantation

Aranzabal J, Perdigo L, *et al,* (1991) Renal transplantation costs: an economic analysis and comparison with dialysis costs. Transplantation Proceedings; 23(5): 2574

Baltzan MA, Ahmed S, *et al,* (1997) Variations in living donor graft rates by dialysis clinic: effect on outcome and cost of chronic renal failure therapy. Clinical Nephrology; 47(6): 351-355

Baltzan MA, Shoker AS, *et al,* (1996) HLA-identity-long-term renal graft survival, acute vascular, chronic vascular, and acute interstitial rejection. Transplantation; 61(6): 881-885

Blommers TJ, Schanbacher B, Corry RJ, (1984) Transplant and dialysis: the cost/benefit question. Iowa Medicine; 74(1): 15-17

Bonal J, Cleries M, Vela E, (1997) Transplantation versus haemodialysis in elderly patients. Renal Registry Committee. Nephrology Dialysis Transplantation; 12(2): 261-264

Chantler C, Carter JE, *et al,* (1980) 10 years' experience with regular haemodialysis and renal transplantation. Archives of Disease in Childhood; 55(6): 435-445

Churchill DN, Torrance GW, *et al,* (1987) Measurement of quality of life in end-stage renal disease: the time-trade off approach. Clinical and Investigative Medicine - Medicine Clinique et Experimentale; 10(1): 14-20

Croxson BE, Ashton T. (1990) A cost-effectiveness analysis of the treatment of end-stage renal failure. New Zealand Medical Journal; 103(888): 171-174

De By TM, D'Amaro J, *et al,* (1982) A retrospective analysis of hospital haemodialysis and kidney transplantation costs in The Netherlands. Netherlands Journal of Medicine; 25(3): 83-87

de Wit AG, Ramsteijn PG, de Charro F, (1998) Economic evaluation of end stage renal disease treatment. Health Policy; 44(3): 215-232

Devlins GM, Mandin H, *et al,* (1990) Illness intrusiveness and quality of life in end-stage renal disease: comparison and stability across treatment modalities. Health Psycology; 9(2): 117-142

Djamali A, Kendziorski C, *et al,* (2003) Disease progression and outcomes in chronic kidney disease and renal transplantation. Kidney International; 64(5): 1800-7

Eggers P, (1992) Comparison of treatment costs between dialysis and transplantation. Seminars in Nephrology; 12(3): 284-289

Evans RW, Manninen DL, *et al,* (1985) The quality of life of patients with end-stage renal disease. New England Journal of Medicine; 312(9): 553-559

Fujisawa M, Ichikama Y, *et al,* (2000) Assessment of health-related quality of life in renal transplant and hemodialysis patients using the SF-36 health survey. Urology; 56(2): 201-206

Garner TI, Dardis R, (1987) Cost-effectiveness analysis of end-stage renal disease treatments. Medical Care; 25(1): 25-34

Gudex CM. (1995) Health-related quality of life in endstage renal failure. Quality of Life Research; 4(4): 359-366

Hathaway DK, Winsett RP, *et al,* (1998) Post kidney transplant quality of life prediction models. Clinical Transplantation; 12(3): 168-74

Haycox A, Jones D, (1996) The cost effectiveness of renal provision in the UK. Journal of management in medicine; 10(1): 6-15

Jacobson SH, Fryd D, *et al,* (1988) Transplantation, hemodialysis, and continuous ambulatory peritoneal dialysis for end-stage renal disease in diabetic patients. Journal of Diabetic Complications; 2(3): 150-157

Jassal SV, Krahn MD, *et al,* (2003) Kidney transplantation in the elderly: a decision analysis. Journal of the American Society of Nephrology; 14(1): 187-196

Jofre R, Lopez-Gomez JM, *et al,* (1998) Changes in quality of life after renal transplantation. American Journal of Kidney Disease; 32(1): 93-100

Johnson JP, McCauley CR, Copley JB, (1982) The quality of life of haemodialysis and transplant patients. Kidney International; 22(3): 286-291

Kalo Z, Jaray J, Nagy J, (2001) Economic evaluation of kidney transplantation versus hemodialysis in patients with end-stage renal disease in Hungary. Progress in Transplantation; 11(3): 188-193

Kaminota M, (2001) Cost-effectiveness analysis of dialysis and kidney transplants in Japan. Keio Journal of Medicine; 50(2): 100-108

Karlberg I, (1992) Cost analysis of alternative treatments in end-stage renal disease. Transplantation Proceedings; 24(1): 335

Karlberg I, Nyberg G, (1995) Cost-effectiveness studies of renal transplantation. International Journal of Technology Assessment in Health Care; 11(3): 611-622

Koch U, Muthny FA, (1991) Quality of life in patients with end-stage renal disease in relation to the method of treatment. Psychotherapy & Psychosomatics; 54(2 to 3): 161-171

Krakauer H, (1986) Assessment of alternative technologies for the treatment of end-stage renal disease. Israel Journal of Medical Sciences; 22(3 to 4): 245-259

Kutner NG, Brogan D, Kutner MH, (1986) End-stage renal disease treatment modality and patients' quality of life. Longitudinal assessment. American Journal of Nephrology; 6(5): 396-402

Laupacis A, Keown P, *et al,* (1996) A study of the quality of life and cost-utility of renal transplantation. Kidney International; 50(1): 235-242

Laupacis A, Pus N, *et al,* (1993) Disease-specific questionnaire for patients with a renal transplant. Nephron; 64(2): 226-231

Lopez-Neblina F, Alvarez Jimenez H, Finkelstein London I, (1999) High efficiency kidney transplantation: concept, technique, results, and cost analysis. Transplantation Proceedings; 31(7): 3025

Loubeau PR, Loubeau JM, *et al,* (2001) The economics of kidney transplantation versus hemodialysis Progress in Transplantation; 11(4): 291-7

Ludbrook A, (1981) A cost-effectiveness analysis of the treatment of chronic renal failure. Applied Economics; 13: 337-50

Madrigal G, (1994) Cost estimate of kidney transplants in Costa Rica: comparison to chronic dialysis. Transplantation Proceedings; 26(1): 121

Matas AJ, Schnitzler MA, (2003) Payment for living donor (vendor) kidneys: a cost-effectiveness analysis. American Journal of Transplantation; 4(8): 216-221

McDonald SP, Craig JC, (2004) Long-term survival of children with end-stage renal disease. New England Journal of Medicine; 350(26): 2654-2662

McDonald SP, Russ GR, (2002) Survival of recipients of cadaveric kidney transplants compared with those receiving dialysis treatment in Australia and New Zealand, 1991-2001. Nephrology Dialysis Transplantation; 17(12): 2212-2219

Mendez R, Asward S, *et al,* (1992) Cost and financing of kidney transplantation in the United States. Transplantation Proceedings; 24(5):2127-2128

Morris PL, Jones B, (1988) Transplantation versus dialysis: a study of quality of life. Transplantation Proceedings; 20(1): 23-26

Morris PL, Jones B, (1989) Life satisfaction across treatment methods for patients with end-stage renal failure. Medical Journal of Australia; 150(8): 428-432
Muthny FA, Koch U, (1991) Quality of life of patients with end-stage renal failure. A comparison of hemodialysis, CAPD, and transplantation. Contributions to Nephrology; 89: 265-273

Nishimura R, Dorman JS, *et al,* (2003) Incidence of ESRD and survival after renal replacement therapy in patients with type 1 diabetes: a report from the Allegheny County Registry. American Journal of Kidney Disease; 42(1): 117-124

Ohi G, Hasegawa T, *et al,* (1986) Why are cadaveric renal transplants so hard to find in Japan? An analysis of economic and attitudinal aspects. Health Policy; 6: 269-278

Ost L, Groth CB, *et al,* (1980) Cadaveric renal transplantation in patients of 60 years and above. Transplantation; 30(5): 339-340

Parfrey PS, Vavasour HM, *et al,* (1987) Symptoms in end-stage renal disease: dialysis v transplantation. Transplantation Proceedings; 19(4): 3407-3409

Parfrey PS, Vavasour HM, *et al,* (1988) Clinical features and severity of nonspecific symptoms in dialysis patients. Nephron; 50(2): 121-128

Parfrey PS, Vavasour HM, Gault MH, (1988a) A prospective study of health status in dialysis and transplant patients. Transplantation Proceedings; 20(6): 1231-1232

Parfrey PS, Vavasour HM, *et al,* (1989) Development of a health questionnaire specific for end-stage renal disease. Nephron; 52(1): 20-28

Park H, Bang WR, *et al,* (1992) Journal Quality of life of ESRD patients: development of a tool and comparison between transplant and dialysis patients. Transplantation Proceedings; 24(4): 1435-1437

Park IH, Yoo HJ, *et al,* (1996) Changes in the quality of life before and after renal transplantation and comparison of the quality of life between kidney transplant recipients, dialysis patients, and normal controls. Transplantation Proceedings; 28(3): 1937-1938

Pietrabissa A, Ciaramella A, *et al,* (1992) Effect of kidney transplantation on quality of life measures. Transplant International; 5(Supplement 1): S708-10

Rebollo P, Ortega F, *et al,* (1998) Health-related quality of life (HRQOL) in end stage renal disease (ESRD) patients over 65 years. Geriatric Nephrology & Urology; 8(2): 85-94

Rebollo P, Ortega F, *et al,* (2000) Health related quality of life (HRQOL) of kidney transplanted patients: variables that influence it. Clinical Transplantation; 14(3): 199-207

Roberts SD, Maxwell DR, Gross L, (1980) Cost-effective care of end-stage renal disease: a billion dollar question. Annals of Internal Medicine; 92(1): 243-248

Rodin G, Voshart K, et al, (1985/6) Cadaveric renal transplant failure: the short-term sequelae. International Journal of Psychiatry in Medicine; 15(4): 357-364

Russell JD, Beecroft ML, *et al,* (1992) The quality of life in renal transplantation--a prospective study. Transplantation; 54(4): 656-660

Salonen T, Reina T, *et al,* (2003) Cost analysis of renal replacement therapies in Finland. American Journal of Kidney Disease; 42(6): 1228-1238

Sayag R, Kaplan De-Nour A, et al, (1990) Comparison of psychosocial adjustment of male nondiabetic kidney transplant and hospital hemodialysis patients. Nephron; (54)3: 214-218

Schaubel DE, Desmeules M, *et al,* (1995) Survival experience among elderly end-stage renal disease patients. A controlled comparison of transplantation and dialysis. Transplantation; 60(12): 1389-1394

Schnitzler MA, Smith C, *et al,* (1999) Relative cost of cadaveric versus living donor kidney transplantation. Transplantation; 67(7): S189

Schweitzer EJ, Wiland A, *et al,* (1998) The shrinking renal replacement therapy "break-even" point. Transplantation; 66(12): 1702-1708

Seedat YK, MacIntosh CG, Subban JV, (1987) Quality of Life for patients in an end-stage renal disease programme. South African Medical Journal; 71(8): 500-504

Sesso R, Eisenberg JM, *et al,* (1990) Cost-effectiveness analysis of the treatment of end-stage renal failure in Brazil. IJTAHC; 6(1): 107-114

Shih FJ, Lee PH, *et al,* (1999) Changes in quality of life and working capacity before and after kidney transplantation. Transplantation Proceedings; 31(5): 1981-1984

Simmons RG, Anderson C, Abress LK, (1990) Quality of life and rehabilitation differences among four end-stage renal disease therapy groups. Scandinavian Journal of Urology & Nephrology Supplementum; 131: 7-22

Thomson NM, Scott DF, *et al,* (1989) Morbidity, mortality, and quality of life in long-term survivors of an integrated dialysis/renal transplant programme Transplantation Proceedings; 29(1 Part 2): 2184-2185

Tomasz W, Piotr S, (2003) A trial of objective comparison of quality of life between chronic renal failure patients treated with hemodialysis and renal transplantation. Annals of Transplantation; 8(2): 47-53

Tousignant P, Guttmann RD, Hollomby DJ, (1985) Transplantation and home haemodialysis: their cost-effectiveness. Journal of Chronic Diseases; 38(7): 589-601

Waiser J, Budde K, *et al,* (1998) The quality of life in end stage renal disease care. Transplant International; 11(Supplement 1): S42-5

Whiting JF, Zavala EY, *et al,* (1999) The cost-effectiveness of transplantation with expanded donor kidneys. Transplantation Proceedings; 31(1 to 2); 1320-1321

Wight J, Edwards L, *et al,* (1998) The SF36 as an outcome measure of services for end stage renal failure. Quality in Health Care; 7(4): 209-221

Wolfe RA, Ashby V, *et al,* (1999) Comparison of mortality in all patients on dialysis, patients on dialysis awaiting transplantation, and recipients of a first cadaveric transplant. New England Journal of Medicine; 341(23): 1725-1730

Yoshimura N, Ohmori Y, *et al,* (1994) Quality of life in renal transplant recipients treated with cyclosporine in comparison with hemodialysis maintenance. Transplantation Proceedings; 26(5): 2542-2543

Liver Transplantation

Batra N, (2001) Hepatitis C screening and treatment versus liver transplantation: a financial option appraisal and commissioning model for purchasers. Disease Management & Health Outcomes; 9(7): 371-384

Majno PE, Sarasin FP, *et al,* (2000) Primary liver resection and salvage transplantation or primary liver transplantation in patients with single, small hepatocellular carcinoma and preserved liver function: an outcome-oriented decision analysis. Hepatology; 31(4): 899-906

O'Carroll RE, Smith K, *et al,* (2000),A comparison of the WHOQOL-100 and the WHOQOL-BREF in detecting change in quality of life following liver transplantation. Quality of Life Research; 9(1): 121-124

Payne JL, McCarty KR, *et al,* (1996) Outcomes analysis for 50 liver transplant recipients: the Vanderbilt experience. American Surgeon; 62(4): 320-325

Price CE, Lowe D, *et al,* (1995) Prospective study of the quality of life in patients assessed for liver transplantation: outcome in transplanted and not transplanted groups. Journal of the Royal Society of Medicine; 88(3): 130-135

Ratcliffe J, Eldabi T, *et al,* (2001) A simulation modelling approach to evaluating alternative policies for the management of the waiting list for liver transplantation. Health Care Management Science; 4(2): 117-124

Ratcliffe J, Longworth L, *et al,* (2002) Assessing health-related quality of life pre- and post-liver transplantation: a prospective multicenter study. Liver Transplantation; 8(3): 263-270

Riether AM, Smith SL, *et al,* (1992) Quality-of-life changes and psychiatric and neurocognitive outcome after heart and liver transplantation. Transplantation; 54(3): 4444-50

Rufat P, Fourquet F, *et al,* (1999) Costs and outcomes of liver transplantation in adults: a prospective, 1-year, follow-up study. Transplantation; 68(1): 76-83

Sagmeister MB, Mullhaupt B, *et al,* (2002) Cost-effectiveness of cadevaric and living donor liver transplantation. Transplantation; 73(4): 616-622

Sarasin FP, Giostra E, *et al,* (1998) Partial hepatectomy or orthotopic liver transplantation for the treatment of resectable hepatocellular carcinoma: a cost-effectiveness perspective. Hepatology; 28(2): 436-442

Shabahang M, Franceschi D, *et al,* (2002) Comparison of hepatic resection and hepatic transplantation in the treatment of hepatocellular carcinoma among cirrhotic patients. Annals of Surgical Oncology; 9(9): 881-886

Tarter RE, Erb S, *et al,* (1998) The quality of life following liver transplantation: a preliminary report. Gastroenterology Clinics of North America; 17(1): 207-217

Tarter RE, Switala J, *et al,* (1991) Quality of life before and after orthotopic hepatic transplantation. Archives of Internal Medicine; 151(8): 1521-1526

Williams JW, Santiogo V, Evans LS, (1987) Socioeconomic aspects of hepatic transplantation. The American Journal of Gastroenterology; 82(11): 1115-1119

Young TA, Neuberger J, *et al,* (2003) Survival gain after liver transplantation for patients with alcoholic liver disease: a comparison across models and centers. Tranplantation; 76(10): 1479-1486

Younossi ZM, McCormick M, *et al,* (2000) Impact of liver transplantation on health-related quality of life. Liver Transplantation; 6(6): 779-783

Lung Transplantation

Al MJ, Koopmanschap M, *et al,* (1998) Cost-effectiveness of lung transplantation in the Netherlands: a scenario analysis. Chest; 113(1): 124-130

Anyanwu AC, McGuire A, *et al,* (2001) Assessment of quality of life in lung transplantation using a simple generic tool. Thorax; 56(3): 218-222

Anyanwu A C, McGuire A, *et al,* (2002) An economic evaluation of lung transplantation. Journal of Thoracic & Cardiovascular Surgery; 123(3): 411-420

Busschbach JJ, Horikx PE, *et al,* (1994) Measuring the quality of life before and after bilateral lung transplantation in patients with cystic fibrosis. Chest; 105(3): 911-917

Cohen L, Littlefield C, *et al,* (1998) Predictors of quality of life and adjustment after lung transplantation. Chest; 113(3): 633-644

Groen H, van der Bij W, *et al,* (2004) Cost-effectiveness of lung transplantation in relation to type of end-stage pulmonary disease. American Journal of Transplantation; 4(7) 1155-1162

Gross CR, Savik K, *et al,* (1995) Long-term health status and quality of life outcomes of lung transplant recipients. Chest; 108(6): 1587-1593

Hosenpud JD, Bennett LE, *et al,* (1998) Effect of diagnosis on survival benefit of lung transplantation for end-stage lung disease. The Lancet; 351(9095): 24-27

Kugler C, Strueber M, *et al,* (2004) Quality of life 1-year after lung transplantation. Progress in Transplantation; 14(4): 331-336

Lanuza DM, Lefaiver C, *et al,* (2000) Prospective study of functional status and quality of life before and after lung transplantation Chest; 118(1): 115-122

Limbos MM, Chan CK, Kesten S, (1997) Quality of life in female lung transplant candidates and recipients. Chest; 112(5): 1165-1174

Limbos MM, Joyce DP, *et al,* (2000) Psychological functioning and quality of life in lung transplant candidates and recipients. Chest; 118(2): 408-416

Liou TG, Adler FR, *et al,* (2001) Survival effect of lung transplantation among patients with cystic fibrosis. JAMA; 286(21): 2683-2689

MacNaughton KL, Rodrigue JR, *et al,* (1998) Health-related quality of life and symptom frequency before and after lung transplantation. Clinical Transplantation; 12(4): 320-323

Ramsey S, Larson, (1998) Lung Transplantation. The Lancet; 351(9111)1285

Ramsey S, Patrick DL, *et al,* (1995a) The cost-effectiveness of lung transplantation. A pilot study. Chest; 108(6): 1594-1601

Ramsey S, Patrick DL, *et al,* (1995b) Improvement in quality of life after lung transplantation: a preliminary study. The University of Washington Medical Center Lung Transplant Study Group. Journal of heart and lung transplantation; 14(5): 870-877

Shih FJ, Tsao CI, *et al,* (2002) The context framing the changes in health-related quality of life and working competence before and after lung transplantation: one-year follow-up in Taiwan. Transplantation Proceedings; 34(7): 2801-2806

Stavem K, Bjortuft O, *et al,* (2000) Health-related quality of life in lung transplant candidates and recipients. Respiration; 67(2): 159-165

TenVergert EM, Essink-Bot ML, *et al,* (1998) The effect of lung transplantation on health-related quality of life: a longitudinal study. Chest; 113(2): 358-364

TenVergert EM, Vermeulem KM, *et al,* (2001) Quality of life before and after lung transplantation in patients with emphysema versus other indications. Psychological Reports; 89(3): 707-717

Van Enckevort PJ, Koopmanschap MA, *et al,* (1997) Lifetime costs of lung transplantation: estimation of incremental costs. Health Economics; 6(5): 479-489

van Enckevort PJ, TenVergert EM, *et al,* (1998) Technology assessment of the Dutch lung transplantation program. International Journal of Technology Assessment in Health Care; 14(2): 344-356

Vermeulen KM, Ouwens JP, *et al,* (2003) Long-term quality of life in patients surviving at least 55 months after lung transplantation. General Hospital Psychatry; 25(2): 95-102

Vermeulen KM, van der Bij W, *et al,* (2004) Improved quality of life after lung transplantation in individuals with cystic fibrosis. Pediatric Pulmonology; 37(5): 419-426

Pancreas Transplantation

Johnson JL, Schellberg J, *et al,* (1990) Does pancreas transplantation really improve the patient's quality of life? Transplantation Proceedings; 22(2): 575-576

Kiberd BA, Larson T, (2000) Estimating the benefits of solitary pancreas transplantation in nonuremic patients with type I diabeties mellitus. Transplantation; 70(7): 1121-1127

Piehlmeier W, Bullinger M, *et al,* (1992) Quality of life in diabetic patients prior to or after pancreas transplantation in relation to organ function. Transplantation Proceedings; 24(3): 871-873

Simultaneous Pancreas-Kidney Transplantation

Adang EM, Engel GL, *et al,* (1996) Comparison before and after transplantation of pancreas-kidney and pancreas-kidney with loss of pancreas--a prospective controlled quality of life study. Transplantation; 62(6): 754-759

Cheung AH, Sutherland DE, *et al,* (1992) Simultaneous pancreas-kidney transplant versus kidney transplant alone in diabetic patients. Kidney International; 41(4): 924-929

Cosimi AB, Auchincloss H Jr, *et al,* (1998) Combined kidney and pancreas transplantation in diabetics. Archives of Surgery; 123(5): 621-625

Douzdjian V, Ferrara D, Silvestri G, (1998) Treatment strategies for insulin-dependent diabetics with ESRD: a cost-effectiveness decision analysis model. American Journal of Kidney Disease; 31(5): 794-802

Douzdjian V, Ferrara D, Silvestri G, (1998) Cost-utility analysis of pancreas transplantation compared to other treatment options for type I diabetics with end-stage renal disease. Transplantation Proceedings; 30: 278

Esmatjes E, Ricart MJ, *et al,* (1994) Quality of life after successful pancreas-kidney transplantation. Clinical Transplantation; 8(2 Part 1): 75-78

Frank A, Deng S, *et al,* (2004) Transplantation for type I diabetes: comparison of vascularized whole-organ pancreas with isolated pancreatic islets. Annals of Surgery; 240(4): 631-643

Gaber AO, Hathaway DK, *et al,* (1994) Improved autonomic and gastric function in pancreas-kidney vs kidney-alone transplantation contributes to quality of life. Transplantation Proceedings; 26(2): 515-516

Gross CR, Kangas JR, *et al,* (1995) One-year change in quality-of-life profiles in patients receiving pancreas and kidney transplants. Transplantation Proceedings; 27(6): 3067-3068

Gross CR, Zehrer CL, (1993) Impact of the addition of a pancreas to quality of life in uremic diabetic recipients of kidney transplants. Transplantation Proceedings; 25(1 Part 2): 1293-1295

Hathaway DK, Hartwig MS, *et al,* (1994) A prospective study of changes in quality of life reported by diabetic recipients of kidney-only and pancreas-kidney allografts. Journal of Transplant Coordination; 4(1): 12-17

Hathaway DK, Hartwig MS, *et al,* (1994a) Improvement in quality of life reported by diabetic recipients of kidney-only and pancreas-kidney allografts. Transplantation Proceedings; 26(2): 512-514

Holohan TV, (1996) Cost-effectiveness modelling of simultanious pancreas-kidney transplantation. International Journal of Technology Assessment in Health Care; 12(3): 416-424

Kiebert GM, van Oosterhout EC, *et al,* (1994) Quality of life after combined kidney-pancreas or kidney transplantation in diabetic patients with end-stage renal disease. Clinical Transplantation; 8(3 Part 1): 239-245

Matas AJ, McHugh L, *et al,* (1998) Long-term quality of life after kidney and simultaneous pancreas-kidney transplantation. Clinical Transplantation; 12(3): 233-242

Milde FK, Hart LK, Zehr PS, (1995) Pancreatic transplantation. Impact on the quality of life of diabetic renal transplant recipients. Diabetes Care; 18(1): 93-95

Nakache R, Tyden G, Groth CG, (1989) Quality of life in diabetic patients after combined pancreas-kidney or kidney transplantation. Diabetes; 38(Supplement 1): 40-42

Nakache R, Tyden G, Groth CG, (1994) Long-term quality of life in diabetic patients after combined pancreas-kidney transplantation or kidney transplantation. Transplantation Proceedings; 26(2): 510-511

Nathan DM, Fogel H, *et al,* (1991) Long-term metabolic and quality of life results with pancreatic/renal transplantation in insulin-dependent diabetes mellitus. Transplantation; 52(1): 85-91

Piehlmeier W, Bullinger M, *et al,* (1991) Quality of life in type 1 (insulin-dependent) diabetic patients prior to and after pancreas and kidney transplantation in relation to organ function. Diabetologia; 34(Supplement 1): S150-S157

Piehlmeier W, Bullinger M, *et al,* (1996) Evaluation of the quality of life of patients with insulin-dependent diabetes mellitus before and after organ transplantation with the SF 36 health survey. European Journal of Surgery; 162(12): 933-940

Reddy KS, Stablein D, *et al,* (2003) Long-term survival following simultaneous kidney-pancreas transplantation versus kidney transplantation alone in patients with type 1 diabetes mellitus and renal failure. American Journal of Kidney Disease; 41(2): 464-470

Secchi A, Di Carlo V, *et al,* (1991) Effect of pancreas transplantation on life expectancy, kidney function and quality of life in uraemic type 1 (insulin-dependent) diabetic patients. Diabetologia; 34(Supplement 1): S141-S144

Secchi A, Martinenghi S, *et al,* (1998) Effects of pancreas transplantation on quality of life in type I diabetic patients undergoing kidney transplantation. Transplantation Proceedings; 30(2): 339-342

Stratta RJ, Taylor RJ, *et al,* (1993) The analysis of benefit and risk of combined pancreatic and renal transplantation versus renal transplantation alone. Surgery, Gynecology & Obstetrics; 177(2): 163-171

Stratta RJ, Taylor RJ, *et al,* (1993a) Combined pancreas-kidney transplantation versus kidney transplantation alone: analysis of benefit and risk. Transplantation Proceedings; 25(1 Part 2): 1298-1301

Sureshkumar KK, Mubin T, *et al,* (2002) Assessment of quality of life after simultaneous pancreas-kidney transplantation. American Journal of Kidney Disease; 39(6): 1300-1306

Venstrom JM, McBride MA, *et al,* (2003) Survival after pancreas transplantation in patients with diabetes and preserved kidney function. JAMA; 290(21): 2817-2823

Zehrer CL, Gross CR, (1994) Comparison of quality of life between pancreas/kidney and kidney transplant recipients: 1-year follow-up. Transplantation Proceedings; 26(2): 508-509

APPENDIX A3.3 FURTHER DETAILS ON THE 158 STUDIES INCLUDED IN THE LITERTURE REVIEW; BY ORGAN

Sections A3.3.1 to A3.3.7 summarise the results of the literature review presented in Chapter 3 in further detail. Results are presented by organ: heart (A3.3.1), intestine (A3.3.2), kidney (A3.3.3), liver (A3.3.4), lung (A3.3.5), pancreas (A3.3.6) and pancreas-kidney (A3.3.7). Each section presents a general overview of the studies reviewed and gives further comments in detail by type of study: survival, HRQL, cost or cost-effectiveness. A summary of the studies included in this review is also summarised in tabular fashion by organ.

A3.3.1 Heart Transplantation

A total of 26 studies compared heart transplantation with a non-transplant comparator group[1] (Table A3.1a to A3.1c). Seven (27%) of studies used a quasi-experimental control group and the remainder of the studies compared transplantation with outcomes on the waiting list prior to transplantation. Quasi-experimental control groups were: Coronary Artery Bypass Grafting (CABG), mitral valve repair, left ventricular reconstruction, cardiomyoplasty, Fontan palliation, ventricular resection, and medical treatment of end-stage heart failure. Only one study matched quasi-experimental and transplant patients in terms of disease characteristics [Shum-Tim, 1999], though a further study did acknowledge that the control cohort might bias the results, as some of the patients in the control group did not have end-stage heart failure [Cope *et al,* 2001].

Survival (Table A3.1a)

Five studies focused on comparing survival with and without transplantation. Four studies compared transplantation with survival on the waiting list whilst the fifth study compared transplant survival with a quasi-experimental comparator group. The study presented by Clark and colleagues is likely to exaggerate the effect of survival after transplantation as two of the 14 patients in the waiting list group were removed from the waiting list due to stabilisation or improvement; in order to maintain an "intention to treat" approach these two patients should have remained in the comparator group for the analysis [Clark *et al,* 1971]. The most thorough survival study compared transplant survival with survival on the waiting list for heart transplantation using a series of statistical models [Turnbull *et al,* 1974]. Turnbull and colleagues proposed a series of alternative statistical techniques to allow for biases in using an intervention delay cohort of waiting list patients, however their approach assumes patients are not given transplants selectively, which is unlikely to be the case.

Two of the four studies that addressed both survival and HRQL had a sample sizes of less than 30 patients. The results from these two studies should be interpreted with caution as they may not be powered to detect a statistical difference between groups. The other study of note in this series of papers evaluated the costs and benefits of the UK heart transplant programme [Buxton *et al,* 1985]. Buxton and colleagues used the methods suggested by Turnbull *et al,* to allow for

[1] A simultaneous heart-lung transplant study is included within this group of studies [Caine *et al,* 1996]

potential biases when comparing a cohort of heart transplant patients with a cohort of non-transplant patients awaiting transplantation [Turnbull *et al,* 1974]. In addition to this the authors also consider a series of functions for extrapolating the survival data beyond the observed study period (linear, Weibull & exponential) [O'Brien *et al,* 1987]. Buxton's study presents costs for the transplant programme but does not compare these costs with a non-transplant cohort of patients.

HRQL (Table A3.1b)

Over half of the heart transplant studies reporting HRQL (62%) compared patients with and without transplantation, with one study comparing the HRQL of heart-lung transplantation [Caine *et al,* 1996; O'Brien *et al,* 1988]. A total of 12 studies compared HRQL with and without transplantation. Two studies compared HRQL after transplantation with a quasi-experimental control group. The remaining 10 HRQL studies compared transplantation pre and post transplant, three of whom used an unpaired comparisons design, measuring HRQL at one time point only, and seven used a paired comparison, evaluating HRQL at a series of time points pre and post transplant.

The studies by Bortman *et al* and Packa *et al* will be subject to recall bias as transplanted patients were asked to value their pre-transplant HRQL several months after transplantation [Packa *et al,* 1989; Bortman *et al,* 1999]. One HRQL study asked experts to value patient's HRQL before and after transplantation using an unvalidated HRQL instrument [Mai *et al,* 1990].

Costs & Cost-effectiveness (Table A3.1c)

Five studies evaluated the costs or cost-effectiveness of heart transplantation in comparison with either quasi-experimental or intervention delay groups. Of note amongst these studies is one that consisted of a sample of one patient, a 25 year old male [Haberman 1980]. A series of hypothesised scenarios were considered for this patient, including transplantation, survival without transplantation for two years and immediate death. The cost-effectiveness of heart transplantation was then estimated for this patient. The results are unlikely to be generalisable to the population, where the average heart transplant patient is unlikely to be a 25-year old male.

Table A3.1a Summary of the literature review results for heart transplantation (N = 9 survival or survival and HRQL studies)

Study	Type of study	Comparator group	Statistical modelling (Y/N)	If modelling used was uncertainty allowed for?	Was censoring allowed for, if C or CE study?	Type of censoring method used	Sensitivity analysis (Y/N)	Type of sensitivity analysis	Small sample size for either treatment or comparator group (N < 30)
Bocchi *et al*	Survival	Quasi	N	N/a	N/a		N		Greater than 30
Clark *et al*	Survival	Intervention	N	N/a	N/a		N		Less than 30
Hellinger *et al*	Survival	Intervention	N	N/a	N/a		N		Greater than 30
Leitz *et al*	Survival	Intervention	N	N/a	N/a		N		Greater than 30
Turnbull *et al*	Survival	Intervention	Y	N	N/a		Y	One-way	Greater than 30
Buxton *et al*[1]	Survival & HRQL	Intervention	Extrapolate	N	N/a		Y	One-way	Greater than 30
Cotrufo *et al*	Survival & HRQL	Quasi	N	N/a	N/a		N		Greater than 30
Fisher *et al*	Survival & HRQL	Before & After	N	N/a	N/a		N		Less than 30
Shum-Tim *et al*	Survival & HRQL	Quasi	N	N/a	N/a		N		Less than 30

[1] Three papers were published for this study: Buxton *et al*, 1985; Caine *et al*, 1990; O'Brien *et al*, 1987

HRQL – Quality of life Intervention – Intervention delay waiting list cohort

Quasi – Quasi-experimental N/a – Not applicable

Table A3.1b Summary of the literature review results for heart transplantation (N = 12 HRQL studies)

Study	Type of study	Comparator group	Statistical modelling (Y/N)	If modelling used was uncertainty allowed for?	Was censoring allowed for, if C or CE study?	Type of censoring method used	Sensitivity analysis (Y/N)	Type of sensitivity analysis	Small sample size for either treatment or comparator group (N < 30)
Bortman *et al*	HRQL	Before & After	N	N/a	N/a		N		Greater than 30
Caine *et al*[1]	HRQL	Before & After	Last value carried forward	N	N/a		N		Greater than 30
Evangalista *et al*	HRQL	Cross section	N	N/a	N/a		N		Greater than 30
Grady *et al*	HRQL	Before & After	N	N/a	N/a		N		Greater than 30
Jones *et al*	HRQL	Before & After	N	N/a	N/a		N		Greater than 30
Mai *et al*	HRQL	Expert opinion	N	N/a	N/a		N		Less than 30
Packa *et al*	HRQL	Before & After	N	N/a	N/a		N		Greater than 30
Rector *et al*	HRQL	Cross section	N	N/a	N/a		N		Greater than 30
Riether *et al*	HRQL	Before & After	N	N/a	N/a		N		Greater than 30
Shih *et al*	HRQL	Before & After	N	N/a	N/a		N		Less than 30
Walden *et al*[2]	HRQL	Quasi	N	N/a	N/a		N		Less than 30
Wu *et al*	HRQL	Quasi	N	N/a	N/a		N		Less than 30

[1] Two papers were published for this study: Caine *et al*, 1996; O'Brien *et al*, 1988 [3] [2] Two papers were published for this study: Walden *et al*, 1989; Walden *et al*, 1994

HRQL – Health related quality of life Quasi – Quasi-experimental N/a – Not applicable

Table A3.1c Summary of the literature review results for heart transplantation (N = 5 cost or cost-effectiveness studies)

Study	Type of study	Comparator group	Statistical modelling (Y/N)	If modelling used was uncertainty allowed for?	Was censoring allowed for, if C or CE study?	Type of censoring method used	Sensitivity analysis (Y/N)	Type of sensitivity analysis	Small sample size for either treatment or comparator group (N < 30)
Hummel *et al*	Cost	Before & After	N	N/a	N		N		Greater than 30
Cope *et al*	Cost-effectiveness	Quasi	N	N/a	N		N		Less than 30
Gajarski *et al*	Cost-effectiveness	Quasi	N	N/a	N		N		Greater than 30
Haberman *et al*	Cost-effectiveness	Intervention (P)	Y	N	N/a		Y	One way	Less than 30
Van-Hout *et al*	Cost-effectiveness	Intervention	Simulation model	N	N/a		N		Greater than 30

Intervention – Intervention delay waiting list cohort P – paired comparison with and without transplantation N/a – Not applicable

A3.3.2 Intestine Transplantation

Intestinal transplantation is an alternative treatment to parenteral nutrition (PN) for patients with chronic intestinal failure. Only two studies were identified that compared transplantation with a non-transplant alternative for intestinal transplantation (Table A3.2). The first study, comprising three papers looked at HRQL using two approaches: a) HRQL before and after transplantation, b) compared the HRQL of listed patients on PN (an alternative treatment to transplantation) with HRQL after transplantation. In both cases the number of patients per group was ten or less.

The second study explored the cost-effectiveness of intestinal transplantation and again used two approaches a) a combination of historical data from a prognostic model and an intervention delay group b) intervention delay waiting list data only [Longworth *et al*, 2003a]. The two approaches produce very different results and the authors acknowledge that there is likely to be bias using the second approach as sicker patients are given priority of healthier patients for transplanted organs. However, the first approach does rely on the validity of the prognostic model, which the authors acknowledge has not been validated, and does not incorporate model parameter uncertainty.

Table A3.2 Summary of the literature review results for intestinal transplantation (N = 2 studies)

Study	Type of study	Comparator group	Statistical modelling (Y/N)	If modelling used was uncertainty allowed for?	Was censoring allowed for, if C or CE study?	Type of censoring method used	Sensitivity analysis (Y/N)	Type of sensitivity analysis	Small sample size for either treatment or comparator group (N < 30)
DiMartini *et al*[1]	HRQL	1. Before & After 2. Quasi	N	N/a	N/a		N		Less than 30
Longworth *et al*	Cost-effectiveness	1. Intervention & historical 2. Intervention	Y N	N N/a	Y Y	Lin Lin	Y	One-way	Less than 30 Less than 30

[1] Three papers were published for this study: DiMartini *et al*, 1998; Rovera *et al*, (1998); Rovera *et al*, (1998a) HRQL – Health related quality of life Intervention – Intervention delay waiting list cohort Quasi – Quasi-experimental N/a – Not applicable

A3.3.3 Kidney Transplantation

Dialysis is an alternative therapy for patients with end-stage renal failure and many of the studies identified in the review of kidney transplant studies compare transplantation with dialysis. In total 64 studies were identified that compared kidney transplantation with alternative treatments for kidney failure, 53 (83%) of which compared transplantation with dialysis, the remaining 11 compared kidney transplantation with an intervention delay cohort of waiting list patients. Of the 64 studies identified ten (16%) focused on survival, 25 (39%) on HRQL, nine (14%) costs only and 20 (31%) were cost-effectiveness studies (Table A3.3a to A3.3d).

Survival (Table A3.3a)

Four of the ten kidney transplant studies that address survival do not adjust for differences in case mix, using Cox PH models, between the non-transplant control group and the kidney transplant cohort. All six of the studies that do adjust for case mix differences between groups use Cox proportional hazard models to do so. However, only one study matched transplant patients with a quasi-experimental control group of listed patients undergoing dialysis, by time spent on the waiting list, prior to analysis [Schaubel *et al,* 1995]. In two of the studies the effect of kidney transplantation is likely to be exaggerated, in favour of transplantation, because transplant patients must survive for either one year [Djamali *et al,* 2003] or five years [Thomson *et al,* 1989] before being included in the study.

HRQL (Table A3.3b)

Nine (35%) of the 26 HRQL studies performed a paired analysis of HRQL pre and post transplantation (before and after study), and the remaining studies compared HRQL post transplant with a quasi-experimental dialysis group. Only one study adjusted for difference in case mix between the quasi-experimental control group and the transplant group before comparing HRQL [Evans *et al,* 1985]. Two studies selected a quasi-experimental dialysis group of patients on dialysis who were listed for transplantation [Jofre *et al,* 1998; Fujisawa *et al,* 2000]. These two studies have more comparable control groups than other studies using quasi-experimental control groups as they considered only patients who had developed end-stage kidney failure (patients with end-stage failure become eligible for transplantation), whereas other studies included patients who were not considered for transplant.

One before and after study will be subject to recall bias as patients were asked to recall their pre-transplant HRQL after transplantation [Koch & Muthny, 1991; Muthny & Koch, 1991].

Costs (Table A3.3c)

All nine cost studies compare transplantation with a quasi-experimental group of dialysis patients. In all but one study the dialysis cohort consists of a group of observed patients, Loupez-Neblina *et al* compare transplantation with simulated cohort of 50 dialysis patients [Loupez-Neblina *et al,* 1999].

Table A3.3a Summary of the literature review results for kidney transplantation (N = 10 survival studies)

Study	Type of study	Comparator group	Statistical modelling (Y/N)	If modelling used was uncertainty allowed for?	Was censoring allowed for, if C or CE study?	Type of censoring method used	Sensitivity analysis (Y/N)	Type of sensitivity analysis	Small sample size for either treatment or comparator group (N < 30)
Bonal *et al*	Survival	Quasi	Cox	N	N/a		N		Greater than 30
Chantler *et al*	Survival	Quasi	N	N/a	N/a		N		Greater than 30
Djamali *et al*	Survival	Quasi	Cox	N	N/a		N		Greater than 30
Jacobson *et al*	Survival	Quasi	N	N/a	N/a		N		Greater than 30
McDonald *et al*[1]	Survival	Intervention	Cox	N	N/a		N		Greater than 30
Nishimura *et al*	Survival	Quasi	Cox	N	N/a		N		Greater than 30
Öst *et al*	Survival	Quasi	N	N/a	N/a		N		Greater than 30
Schaubel *et al*	Survival	Quasi	Cox	N	N/a		N		Greater than 30
Thomson *et al*	Survival	Quasi	N	N/a	N/a		N		Greater than 30
Wolfe *et al*	Survival	Intervention	Cox	N	N/a		N		Greater than 30

[1] Two papers were published for this study: McDonald & Ross, 2002; McDonald & Craig, 2004 Intervention – Intervention delay waiting list cohort Quasi – Quasi-experimental

N/a – Not applicable

Table A3.3b Summary of the literature review results for kidney transplantation (N = 25 HRQL studies)

Study	Type of study	Comparator group	Statistical modelling (Y/N)	If modelling used was uncertainty allowed for?	Was censoring allowed for, if C or CE study?	Type of censoring method used	Sensitivity analysis (Y/N)	Type of sensitivity analysis	Small sample size for either treatment or comparator group (N < 30)
Churchill *et al*	HRQL	Quasi	N	N/a	N/a		N		Greater than 30
Devlins *et al*	HRQL	Quasi	N	N/a	N/a		N		Greater than 30
Evans *et al*	HRQL	Quasi	Regression	N	N/a		N		Greater than 30
Fujisawa *et al*	HRQL	Quasi	N	N/a	N/a		N		Greater than 30
Gudex	HRQL	Quasi	N	N/a	N/a		N		Greater than 30
Hathaway *et al*	HRQL	Before & After	N	N/a	N/a		N		Greater than 30
Jofré *et al*	HRQL	Before & After	N	N/a	N/a		N		Greater than 30
Johnson *et al*	HRQL	Quasi	N	N/a	N/a		N		Less than 30
Koch & Muthny[1]	HRQL	1. Before & After 2. Quasi	N N	N/a	N/a		N		Greater than 30
Kutner *et al*	HRQL	Before & After	N	N/a	N/a		N		Greater than 30
Morris & Jones[2]	HRQL	Quasi	N	N/a	N/a		N		Less than 30
Parfrey *et al*[3]	HRQL	Quasi	N	N/a	N/a		N		Greater than 30
Park *et al*[4]	HRQL	1. Before & After 2. Quasi	N N	N/a N/a	N/a N/a		N N		Less than 30 Greater than 30
Pietrabissa *et al*	HRQL	Quasi	N	N/a	N/a		N		Greater than 30
Rebello *et al*[5]	HRQL	Quasi	N	N/a	N/a		N		Greater than 30
Rodin *et al*	HRQL	Before & After	N	N/a	N/a		N		Greater than 30
Russell *et al*	HRQL	Before & After	N	N/a	N/a		N		Greater than 30

Table A3.3b Summary of the literature review results for kidney transplantation (N = 25 HRQL studies)

Study	Type of study	Comparator group	Statistical modelling (Y/N)	If modelling used was uncertainty allowed for?	Was censoring allowed for, if C or CE study?	Type of censoring method used	Sensitivity analysis (Y/N)	Type of sensitivity analysis	Small sample size for either treatment or comparator group (N < 30)
Churchill *et al*	HRQL	Quasi	N	N/a	N/a		N		Greater than 30
Devlins *et al*	HRQL	Quasi	N	N/a	N/a		N		Greater than 30
Evans *et al*	HRQL	Quasi	Regression	N	N/a		N		Greater than 30
Fujisawa *et al*	HRQL	Quasi	N	N/a	N/a		N		Greater than 30
Gudex	HRQL	Quasi	N	N/a	N/a		N		Greater than 30
Hathaway *et al*	HRQL	Before & After	N	N/a	N/a		N		Greater than 30
Jofré *et al*	HRQL	Before & After	N	N/a	N/a		N		Greater than 30
Johnson *et al*	HRQL	Quasi	N	N/a	N/a		N		Less than 30
Koch & Muthny[1]	HRQL	1. Before & After 2. Quasi	N N	N/a	N/a		N		Greater than 30
Kutner *et al*	HRQL	Before & After	N	N/a	N/a		N		Greater than 30
Morris & Jones[2]	HRQL	Quasi	N	N/a	N/a		N		Less than 30
Parfrey *et al*[3]	HRQL	Quasi	N	N/a	N/a		N		Greater than 30
Park *et al*[4]	HRQL	1. Before & After 2. Quasi	N N	N/a N/a	N/a N/a		N N		Less than 30 Greater than 30
Pietrabissa *et al*	HRQL	Quasi	N	N/a	N/a		N		Greater than 30
Rebello *et al*[5]	HRQL	Quasi	N	N/a	N/a		N		Greater than 30
Rodin *et al*	HRQL	Before & After	N	N/a	N/a		N		Greater than 30
Russell *et al*	HRQL	Before & After	N	N/a	N/a		N		Greater than 30

Study	Type of study	Comparator group	Statistical modelling (Y/N)	If modelling used was uncertainty allowed for?	Was censoring allowed for, if C or CE study?	Type of censoring method used	Sensitivity analysis (Y/N)	Type of sensitivity analysis	Small sample size for either treatment or comparator group (N < 30)
Sayag *et al*	HRQL	Quasi	N	N/a	N/a		N		Greater than 30
Seedat *et al*	HRQL	Quasi	N	N/a	N/a		N		Greater than 30
Shih *et al*	HRQL	Before & After	N	N/a	N/a		N		Greater than 30
Simmons *et al*	HRQL	Quasi	N	N/a	N/a		N		Greater than 30
Tomosz & Piotr	HRQL	Quasi	N	N/a	N/a		N		Greater than 30
Waiser *et al*	HRQL	Quasi	N	N/a	N/a		N		Greater than 30
Wight *et al*	HRQL	Quasi	N	N/a	N/a		N		Greater than 30
Yoshimura *et al*	HRQL	Before & After	N	N/a	N/a		N		Greater than 30

[1] Two papers were included for this study: Koch & Muthny, 1991; Muthny & Koch 1991 [2] Two papers were included for this study: Morris & Jones, 1988; Morris & Jones, 1989

[3] Four papers were included for this study: Parfrey *et al*, 1987; Parfrey *et al*, 1988; Parfrey *et al*, 1988a Parfrey *et al*, 1989

[4] Two papers were included for this study: Park *et al*, 1992; Park *et al*, 1996 [5] Two papers were included for this study: Rebello *et al*, 1998; Rebello *et al*, 2000

HRQL – Health related quality of life Intervention – Intervention delay waiting list cohort Quasi – Quasi-experimental N/a – Not applicable

Table A3.3c Summary of the literature review results for kidney transplantation (N = 9 cost studies)

Study	Type of study	Comparator group	Statistical modelling (Y/N)	If modelling used was uncertainty allowed for?	Was censoring allowed for, if C or CE study?	Type of censoring method used	Sensitivity analysis (Y/N)	Type of sensitivity analysis	Small sample size for either treatment or comparator group (N < 30)
Baltzan *et al*[†]	Costs	Quasi	N	N/a	N		N		Greater than 30
Haycox & Jones	Costs	Quasi	N	N/a	N		N		Greater than 30
Lopez-Neblina *et al*	Costs	Quasi	N	N/a	N		N		Greater than 30
Loubeau *et al*	Costs	Quasi	N	N/a	N		N		Greater than 30
Madrigal	Costs	Quasi	N	N/a	N		N		Greater than 30
Mendez *et al*	Costs	Quasi	N	N/a	N		N		Greater than 30
Salonen *et al*	Costs	Quasi	N	N/a	N		N		Greater than 30
Schnitzler *et al*	Costs	Quasi	N	N/a	N		N		Greater than 30
Schweitzler *et al*	Costs	Quasi	N	N/a	N		N		Greater than 30

[†] Two papers were published for this study: Baltzan *et al*, 1996; Baltzan *et al*, 1997 Intervention – Intervention delay waiting list cohort Quasi – Quasi-experimental

N/a – Not applicable

Table A3.3d Summary of the literature review results for kidney transplantation (N = 20 cost-effectiveness studies)

Study	Type of study	Comparator group	Statistical modelling (Y/N)	If modelling used was uncertainty allowed for?	Was censoring allowed for, if C or CE study?	Type of censoring method used	Sensitivity analysis (Y/N)	Type of sensitivity analysis	Small sample size for either treatment or comparator group (N < 30)
Aranzabal *et al*	Cost-effectiveness	Quasi	N	N/a	N		N		Greater than 30
Bloomers *et al*	Cost-effectiveness	Quasi	N	N/a	N		N		Greater than 30
Croxson & Ashton	Cost-effectiveness	Quasi	Markov	N	N/a		Y	One-way	Greater than 30
De By *et al*	Cost-effectiveness	Quasi	N	N/a	N		N		Greater than 30
De Wit *et al*	Cost-effectiveness	Quasi	Markov	N	N/a		Y	One-way	Greater than 30
Eggars	Cost-effectiveness	Quasi	N	N/a	N		Y	One-way	Greater than 30
Garner & Dardis	Cost-effectiveness	Quasi	Y	N	Y	Lin	Y	One-way	Greater than 30
Jassal *et al*	Cost-effectiveness	Quasi	Markov	N	N/a (Lifetime)		Y	One-way	Greater than 30
Kaló *et al*	Cost-effectiveness	Intervention	Extrapolate	N	N/a		Y	Discount rate	Greater than 30
Kaminota *et al*	Cost-effectiveness	Quasi	Y	N	N/a (Lifetime)		Y	One-way	Greater than 30
Karlberg[1]	Cost-effectiveness	Quasi	N	N/a	N		Y	One-way	Greater than 30
Krakauer	Cost-effectiveness	Quasi	Cox	N	N		N		Greater than 30
Laupacis *et al*[2]	Cost-effectiveness	Intervention	Extrapolate	N	N		N		Greater than 30
Ludbrook	Cost-effectiveness	Quasi	Markov	N	N		Y	One-way	Greater than 30
Matas & Schnitzler	Cost-effectiveness	Quasi	Markov	N	N		N		Greater than 30
Ohi *et al*	Cost-effectiveness	Quasi	N	N/a	Y	Lin	N		Greater than 30
Roberts *et al*	Cost-effectiveness	Quasi	Modelling	N	N/a (Lifetime)		N		Greater than 30
Sesso *et al*	Cost-effectiveness	Quasi	N	N/a	N		N		Greater than 30

Study	Type of study	Comparator group	Statistical modelling (Y/N)	If modelling used was uncertainty allowed for?	Was censoring allowed for, if C or CE study?	Type of censoring method used	Sensitivity analysis (Y/N)	Type of sensitivity analysis	Small sample size for either treatment or comparator group (N < 30)
Tousignant *et al*	Cost-effectiveness	Quasi	N	N/a	N		N		Less than 30
Whiting *et al*	Cost-effectiveness	Quasi	Markov	N	N/a		Y	One-way	Greater than 30

[1] Two papers were included for this study: Karlberg, 1992; Karlberg & Nyberg, 1995 [2] Laupacis *et al*, 1993; Laupacis *et al*, 1996 Intervention – Intervention delay waiting list cohort

Quasi – Quasi-experimental N/a – Not applicable

A3.3.4 Liver Transplantation

A total of 27 studies were identified as comparing survival, HRQL, costs or cost-effectiveness of non-transplant versus transplant cohorts: three survival, 14 compared HRQL, two costs and eight reported cost-effectiveness analysis (Table A3.4a to A2.4d).

Survival (Table A3.4a)

Of the three survival studies only one used a prognostic model to estimate survival in the absence of transplantation. The prognostic model was based on historical patients with end stage liver disease and was applied to the transplant patients, using information from their experience on the waiting list, to estimate survival in the absence of transplantation [Christensen *et al*, 1999; Liemann Garcia *et al*, 2001]. The other two survival studies used a quasi-experimental comparator group consisting of patients receiving an alternative treatment for end-stage organ failure [Llovet *et al*, 1999; Shabahang *et al*, 2002].

HRQL (Table A3.4b)

Over half (ten) of the HRQL studies used a before and after approach for the comparison of HRQL with and without transplantation. Seven of these studies excluded patients who did not respond at all time points during a study or those who died before the study was completed, making these studies biased towards healthier patients. However, Payne *et al* and Younossi *et al*, who both used a time series approach to measuring HRQL, included patient information up to the time point a patient dropped out of the study [Payne *et al*, 1996; Younossi *et al*, 2000]. The remaining studies used an unpaired comparator approach and all included survivors.

Costs (Table A3.4c)

Rufat and colleagues cost study uses cost incurred on the waiting list for liver transplantation to estimate costs in the absence of transplantation, whereas Batra uses a quasi-experimental alternative treatment group [Rufat *et al*, 1999; Batra, 2001]. Of the two studies only Rufat et al conduct one-way sensitivity analysis on the assumptions they make in the cost analysis.

Cost-effectiveness (Table A3.4c)

Over half of the cost-effectiveness studies used modelling techniques to estimate costs and effects with and without transplantation, with all but one study conducting one-way sensitivity analysis.

The remaining three cost-effectiveness studies used a quasi-experimental comparator group. Burroughs *et al* used a cohort of patients with end-stage cirrhosis who were been treated for complications at the same centre as the transplant patients [Burroughs *et al*, 1992]. The comparator cohort is likely to be of a different case mix to the transplant cohort and is likely to contain patients who are not eligible for transplantation. It is not clear how these potential differences between groups could affect the results as patients clinical details are not supplied in the paper. Caution should also be applied to the interpretation of the paper by Manjo and

colleagues where it would be difficult to generalise the results as the transplant cohort consists of only four patients [Manjo *et al,* 2000]. Finally, Williams *et al* study is likely to exaggerate the effects of transplantation [Williams *et al,* 1987]. The transplant cohort is selective and contains patients who survive for at least six months post transplant, whilst although their non-transplant cohort met listing criteria for transplantation they included patients who died in the six month period preceding transplantation.

Table A3.4a Summary of the literature review results for liver transplantation (N = 3 survival studies)

Study	Type of study	Comparator group	Statistical modelling (Y/N)	If modelling used was uncertainty allowed for?	Was censoring allowed for, if C or CE study?	Type of censoring method used	Sensitivity analysis (Y/N)	Type of sensitivity analysis	Small sample size for either treatment or comparator group (N < 30)
Christensen *et al*[1]	Survival	Combined historical & Intervention	PM	N	N/a		N		Greater than 30
Llovet *et al*	Survival	Quasi	Cox	N	N/a		N		Greater than 30
Shabahang *et al*	Survival	Quasi	N	N/a	N/a		N		Greater than 30

[1] Two papers were published for this study: Christensen *et al*, 1999; Liemann Garcia *et al*, 2001 Quasi – Quasi-experimental PM – Prognostic model N/a – Not applicable

Table A3.4b Summary of the literature review results for liver transplantation (N = 14 HRQL studies)

Study	Type of study	Comparator group	Statistical modelling (Y/N)	If modelling used was uncertainty allowed for?	Was censoring allowed for, if C or CE study?	Type of censoring method used	Sensitivity analysis (Y/N)	Type of sensitivity analysis	Small sample size for either treatment or comparator group (N < 30)
Belle *et al*	HRQL	Before & After	N	N/a	N/a		N		Greater than 30
Cole *et al*	HRQL	Before & After	N	N/a	N/a		N		Greater than 30
De Bona *et al*	HRQL	Cross section	N	N/a	N/a		N		Greater than 30
Gross *et al*	HRQL	Before & After	N	N/a	N/a		N		Greater than 30
Karam *et al*	HRQL	Before & After	N	N/a	N/a		N		Greater than 30
Kober *et al*	HRQL	Cross section	N	N/a	N/a		N		Greater than 30
Levy *et al*	HRQL	Cross section	N	N/a	N/a		N		Greater than 30
O'Carroll *et al*	HRQL	1. Before & After 2. Cross section	N	N/a	N/a		N		Greater than 30 Less than 30
Payne *et al*	HRQL	Before & After	N	N/a	N/a		N		Greater than 30
Price *et al*	HRQL	Cross section	N	N/a	N/a		N		Greater than 30
Riether *et al*	HRQL	Before & After	N	N/a	N/a		N		Greater than 30
Tarter *et al*, 1988	HRQL	Before & After	N	N/a	N/a		N		Less than 30
Tarter *et al*, 2001	HRQL	Before & After	N	N/a	N/a		N		Greater than 30
Younossi *et al*	HRQL	Before & After	N	N/a	N/a		N		Greater than 30

HRQL – Health related quality of life Quasi – Quasi-experimental N/a – Not applicable

Table A3.4c Summary of the literature review results for liver transplantation (N = 10 cost or cost-effectiveness studies)

Study	Type of study	Comparator group	Statistical modelling (Y/N)	If modelling used was uncertainty allowed for?	Was censoring allowed for, if C or CE study?	Type of censoring method used	Sensitivity analysis (Y/N)	Type of sensitivity analysis	Small sample size for either treatment or comparator group (N < 30)
Batra	Costs	Quasi	N	N/a	N		N		Greater than 30
Rufat *et al*	Costs	Intervention	N	N/a	N		Y	One way	Greater than 30
Bonsel *et al*[1]	Cost-effectiveness	Combined historical & Intervention	PM	N	N		N		Greater than 30
Burroughs *et al*	Cost-effectiveness	Quasi	N	N/a	N		N		Less than 30
Farinati *et al*	Cost-effectiveness	Quasi	N	N/a	N		N		Less than 30
Longworth *et al*[2]	Cost-effectiveness	Combined historical & Intervention	PM	N	N/a		Y	One way	Greater than 30
Majno *et al*	Cost-effectiveness	Quasi	Markov	N	N/a		Y	One way	Greater than 30
Sagmeister *et al*	Cost-effectiveness	Intervention	Markov	N	N/a		Y	One-way	Greater than 30
Sarasin *et al*	Cost-effectiveness	Quasi	Decision	N	N/a		Y	One-way	Greater than 30
Williams *et al*	Cost-effectiveness	Historical	N	N/a	N		N		Less than 30

[1] Four papers were published for this study: Bonsel *et al*, 1900; Bonsel *et al*, 1900a; Bonsel *et al*, 1900b; Bonsel *et al*, 1992

[2] Five papers were published for this study: Longworth *et al*, 2003; Longworth & Bryan, 2003; Ratcliffe *et al*, 2001; Ratcliffe *et al*, 2002; Young *et al*, 2003

Intervention – Intervention delay waiting list cohort Quasi – Quasi-experimental PM – Prognostic model N/a – Not applicable

A3.3.5 Lung Transplantation

Lung transplantation is accepted as a treatment of choice for some patients with end-stage lung failure, where either single or double lung transplantation can occur. The literature review identified 16 studies that compared lung transplantation with a non-transplant cohort; two studies looked at survival alone, 11 were HRQL studies and three were cost-effectiveness studies of lung transplantation (Table A3.5).

Survival (Table A3.5)

Of the two survival studies, one used a time dependent non-parametric survival model to compare survival with and in the absence of transplantation [Hosenpud *et al,* 1998] The second study used a modelling approach to estimate non-transplant and transplant survival up-to five years, adjusting for demographic and clinical characteristics [Liou *et al,* 2001]. The latter study did not consider allowing for model parameter uncertainties.

HRQL (Table A3.5)

In four studies the authors compared a non-transplant cohort with a transplant cohort at a single time point (cross sectional study) and eight studies used a before and after comparison of HRQL with and without transplantation (one study used both a cross sectional and before and after approach [Cohen *et al,* 1998]). All of the studies that used a before and after approach ignored patients with incomplete HRQL profiles over time and patients who died during the study.

Cost-effectiveness (Table A3.5)

All three cost-effectiveness studies compare lung transplantation with an intervention delay group of waiting list patients over the lifetime of the study cohort, in all three studies data are extrapolated beyond the observed study period. The authors use various statistical modelling techniques including the declining exponential function (DEALE) method and the Weibull model to extrapolate survival beyond the study period.

Ramsey and colleagues extrapolate costs based on the monthly cost on the waiting list, which varies by time post-listing [Ramsey *et al,* 1995]. Van Enckevort *et al* use a slightly more sophisticated approach; they extrapolate transplant and non-transplant costs based on the average per patient weekly cost at the end of the observed post-transplant follow-up period or time spent on the waiting list [Van Enckevort *et al,* 1997]. The Dutch study of Van Enckevort and colleagues also make an adjustment to both transplant and non-transplant costs in the three months prior to death. Anyanwu *et al,* state that they use the same approach as Van Enckevort to estimate the lifetime costs of lung transplantation and a non-transplant cohort of patients in the UK [Anyanwu *et al,* 2002].

Table A3.5 Summary of the literature review results for lung transplantation (N = 16 studies)

Study	Type of study	Comparator group	Statistical modelling (Y/N)	If modelling used was uncertainty allowed for?	Was censoring allowed for if C or CE study?	Type of censoring method used	Sensitivity analysis (Y/N)	Type of sensitivity analysis	Small sample size for either treatment or comparator group (N < 30)
Hosenpud *et al*	Survival	Intervention	Non-parametric	N	N/a		N		Greater than 30
Liou *et al*	Survival	Quasi	Cox	N	N/a		N		Greater than 30
Busschbach *et al*	HRQL	Before & After	N	N/a	N/a		N		Less than 30
Cohen *et al*	HRQL	Before & After Cross section	N N	N/a N/a	N/a N/a		N N		Greater than 30 Less than 30
Gross *et al*	HRQL	Before & After	N	N/a	N/a		N		Less than 30
Kugler *et al*	HRQL	Before & After	N	N/a	N/a		N		Greater than 30
Lanuza *et al*	HRQL	Before & After	N	N/a	N/a		N		Less than 30
Limbos *et al*	HRQL	Cross section	N	N/a	N/a		N		Less than 30
Limbos *et al*	HRQL	Cross section	N	N/a	N/a		N		Greater than 30
MacNaughton *et al*	HRQL	Before & After	N	N/a	N/a		N		Less than 30
Shih *et al*	HRQL	Before & After	N	N/a	N/a		N		Less than 30
Staverm *et al*	HRQL	Cross section	N	N/a	N/a		N		Less than 30
Vermeulen *et al*[1]	HRQL	Before & After	N	N/a	N/a		N		Less than 30
Anyanwu *et al*[2]	Cost-effectiveness	Intervention	Extrapolate	N	N/a		N		Greater than 30
Ramsey *et al*[3]	Cost-effectiveness	Intervention	Extrapolate	N	N (Lifetime)		N		Less than 30
Van Enkevort *et al*[4]	Cost-effectiveness	Intervention	Extrapolate	N	N (Lifetime)		N		Greater than 30

[1] Two papers were published for this study: Vermeulen *et al*, 2003; Vermeulen *et al*, 2004 [2] Two papers were published for this study: Anyanwu *et al*, 2001; Anyanwu *et al*, 2002

[3] Three papers were published for this study: Ramsey *et al*, 1998; Ramsey *et al*, 1995; Ramsey *et al*, 1995a

[4] Six papers were published for this study: Al *et al*, 1998; Groen *et al*, 2004; TenVergert *et al*, 1998; TenVergert *et al*, 2001; van Enckervort *et al*, 1997; van Enckervort *et al*, 1998

HRQL – Health related quality of life Intervention – Intervention delay waiting list cohort control group Quasi – Quasi-experimental N/a – Not applicable

A3.3.6 Pancreas Transplantation

A total of three studies compared pancreas transplantation with a non-transplant control group (Table A3.6). One HRQL study compared transplantation with dialysis on the waiting list [Piehlmeier *et al,* 1992], one HRQL study compared transplant patients' treatment on the waiting list with their post-transplant experience [Johnson *et al,* 1990] and a cost-effectiveness study compared transplantation with treatment for diabetes mellitus [Kiberd & Larson, 2000].

Table A3.6 Summary of the literature review results for pancreas transplantation (N = 3 studies)

Study	Type of study	Comparator group	Statistical modelling (Y/N)	If modelling used was uncertainty allowed for?	Was censoring allowed for, if C or CE study?	Type of censoring method used	Sensitivity analysis (Y/N)	Type of sensitivity analysis	Small sample size for either treatment or comparator group (N < 30)
Johnson *et al*	HRQL	Before & After	N	N/a	N/a		N		Less than 30
Piehlmeier *et al*	HRQL	Quasi	N	N/a	N/a		N		Greater than 30
Kiberd & Larson	Cost-effectiveness	Quasi	Markov	N	N (Lifetime)		Y	One-way	Greater than 30

HRQL – Health related quality of life Intervention – Intervention delay waiting list cohort Quasi – Quasi-experimental N/a – Not applicable

A3.3.7 Simultaneous Pancreas Kidney Transplant (SPK)

A total of 20 studies were identified that compared SPK with either kidney transplant alone, haemodialysis, or an intervention delay (waiting list) control group, with five studies comparing SPK with more than one control group. These studies are detailed in Tables A3.7a to A3.7c below.

Survival (Table A3.7a)

All three survival studies made adjustments for case mix factors. However, none of the three studies looking at survival and HRQL adjusted for case mix factors. These three studies all compared simultaneous pancreas-kidney transplantation with kidney transplant alone. In addition to this Secchi and colleagues also compared the survival and HRQL of SPK patients with a group of haemodialysis patients [Secchi *et al*, 1991] and Stratta *et al,* compared HRQL pre and post transplant [Stratta *et al,* 1993].

HRQL (Table A3.7b)

A total of ten studies compared HRQL with and without transplantation: five studies compared HRQL pre and post transplant, two compared SPK with a quasi-experimental (haemodialysis) control group, and six studies compared SPK with kidney transplant alone. Only one paper matched SPK recipients by gender, age and year of transplant [Sureshkumar *et al,* 2002]. Milde and colleagues comparison of HRQL pre and post SPK is subject to recall bias as patients are asked to recall what their HRQL was before they had a transplant [Milde *et al,* 1995]. The study does not give details of how long after transplantation patients are asked about their HRQL. Two studies used more than one comparator group.

Cost and Cost-Effectiveness (Table A3.7c)

The cost study of Cosimi *et al,* gives scant details of their methodology and present only the average charges of transplantation in comparison with kidney transplant alone [Cosimi *et al,* 1988]. Of the three cost-effectiveness studies, two use a modelling approach, one over a patient's lifetime [Holohan, 1996] and the other over a five year period [Douzdjian *et al,* 1998; Douzdjian *et al,* 1999]. The remaining cost-effectiveness study uses observed data to compare transplantation with a quasi-experimental control group [Frank *et al,* 2004].

Table A3.7a Summary of the literature review results for simultaneous pancreas-kidney transplantation (N = 6 survival or HRQL and survival studies)

Study	Type of study	Comparator group	Statistical modelling (Y/N)	If modelling used was uncertainty allowed for?	Was censoring allowed for, if C or CE study?	Type of censoring method used	Sensitivity analysis (Y/N)	Type of sensitivity analysis	Small sample size for either treatment or comparator group (N < 30)
Cheung *et al*	Survival	Kidney alone	Y (for age)	N	N/a		N		Greater than 30
Reddy *et al*	Survival	Kidney alone	Cox	N	N/a		N		Greater than 30
Venstrom *et al*	Survival	Intervention	Cox	N	N/a		N		Greater than 30
Nathan *et al*	Survival & HRQL	Kidney alone	N	N/a	N/a		N		Less than 30
Secchi *et al*[1]	Survival & HRQL	1 Kidney alone 2 Quasi	N N	N/a N/a	N/a		N N		Less than 30 Greater than 30
Stratta *et al*[2]	Survival & HRQL	1 Kidney alone 2 Before & After	N N	N/a N/a	N/a		N N		Greater than 30 Greater than 30

[1] Two papers were published for this study: Secchi *et al*, 1991; Secchi *et al*, 1998 [2] Two papers were published for this study: Stratta *et al*, 1993; Stratta *et al*, 1993a

HRQL – Health related quality of life Intervention – Intervention delay waiting list cohort Quasi – Quasi-experimental N/a – Not applicable

Table A3.7b Summary of the literature review results for simultaneous pancreas-kidney transplantation (N = 10 HRQL studies)

Study	Type of study	Comparator group	Statistical modelling (Y/N)	If modelling used was uncertainty allowed for?	Was censoring allowed for, if C or CE study?	Type of censoring method used	Sensitivity analysis (Y/N)	Type of sensitivity analysis	Small sample size for either treatment or comparator group (N < 30)
Adang*et al*	HRQL	Before & After	N	N/a	N/a		N		Less than 30
Esmatjes *et al*	HRQL	Quasi	N	N/a	N/a		N		Greater than 30
Gaber *et al*[1]	HRQL	Before & After	N	N/a	N/a		N		Less than 30
Gross *et al*[2]	HRQL	1 Kidney alone 2 Quasi 3 Intervention	N N N	N/a N/a N/a	N/a N/a N/a		N N N		Greater than 30 Greater than 30 Less than 30
Kiebert *et al*	HRQL	Kidney alone	N	N/a	N/a		N		Greater than 30
Matas *et al*	HRQL	Kidney alone	N	N/a	N/a		N		Greater than 30
Milde *et al*	HRQL	1 Kidney alone 2 Before & After	N N	N/a N/a	N/a N/a		N N		Greater than 30 Greater than 30
Nakache *et al*[3]	HRQL	Kidney alone	N	N/a	N/a		N		Less than 30
Piehlmeir *et al*[4]	HRQL	Before & After Cross section	N	N/a	N/a		N		Greater than 30
Sureshkumar *et al*	HRQL	Kidney alone	N	N/a	N/a		N		Less than 30

[1] Three papers were published for this study: Gaber *et al*, 1994; Hathaway *et al*, 1994a; Hathaway *et al*, 1994b

[2] Three papers were published for this study: Gross & Zehrer, 1993; Gross *et al*, 1995; Zehrer & Gross, 1994 [3] Two papers were published for this study: Nakache *et al*, 1989; Nakache *et al*, 1994

[4] Two papers were published for this study: Piehlmeier *et al*, 1991; Piehlmeier *et al*, 1996 HRQL – Health related quality of life Intervention – Intervention delay waiting list cohort

Quasi – Quasi-experimental N/a – Not applicable

Table A3.7c Summary of the literature review results for simultaneous pancreas-kidney transplantation (N = 4 cost or cost-effectiveness studies)

Study	Type of study	Comparator group	Statistical modelling (Y/N)	If modelling used was uncertainty allowed for?	Was censoring allowed for, if C or CE study?	Type of censoring method used	Sensitivity analysis (Y/N)	Type of sensitivity analysis	Small sample size for either treatment or comparator group (N < 30)
Cosimi *et al*	Costs	Kidney alone	N	N/a	N		N		Less than 30
Douzdjian *et al*[1]	Cost-effectiveness	1 Kidney alone 2 Quasi	Markov Markov	N N	N N		Y Y	One-way One-way	Greater than 30 Greater than 30
Frank *et al*	Cost-effectiveness	Quasi	N	N/a	N		Y	One-way	Less than 30
Holohan	Cost-effectiveness	Kidney alone	Modelling	N	N (Lifetime)		Y	One-way	Greater than 30

[1] Two papers were published for this study: Douzjian *et al*, 1998; Douzjian *et al*, 1999 Intervention – Intervention delay waiting list cohort Quasi – Quasi-experimental N/a – Not applicable

APPENDIX A4.1 LITERATURE REVIEW SEARCH STRATERGY FOR IDENTIFYING STUDIES CONSIDERING CENSORED COST METHODOLOGY

A literature review was conducted in order to identify existing methods for estimating mean total study costs in the presence of censored data. The following databases were searched: Ovid Medline [Ovid, 2005], BIDS Social Science database [BIDS, 2005], NHS EED [Centre for Review and Dissemination, 2005] and OHE HEED [OHE HEED, 2005]. A list of search terms is presented in Table A4.1.1; any term listed in the first column of the table was combined with the search terms listed in the second column of the table.

Table A4.1.1 Search terms used to identify censored cost methods

The following terms:	Were combined with:
Censor Censors Censored Censoring	Cost Costs Costing
Incomplete	Cost Costs Costing
Missing	Cost Costs Costing

The titles and abstracts of articles that were identified in the literature review were scanned and a total of 30 articles were obtained and read to establish whether they contained methodology for estimating mean total study costs in the presence of censoring. A further three articles were added to the review from the reference list of reviewed articles, giving a total of 33 articles.

Of the 33 articles that were obtained and read:

14 (42%) were rejected as they did not contain original methodology
Four (12%) were rejected as they treated censoring as a missing data problem
Two (6%) were rejected as methods dealt with censoring of QALY data
Six (18%) were rejected as methods dealt with censoring of cost-effectiveness data
Seven (21%) were accepted as they contained original methods for estimating mean study costs in the presence of censoring.

The seven articles that were accepted listed a possible 12 methods for estimating mean total costs in the presence of censoring. Three articles were identified by Ovid, one by both Ovid and NHS EED, and the remaining three articles from the reference lists of other articles.

References

Bang H, Tsaitis AA, (2000) Estimating medical costs with censored data. Biometrika; 87(2): 329-343

Carides GW, Heyse JF, Iglewicz B (2000) A regression based method for estimating mean treatment cost in the presence of right-censoring. Biostatistics; 1(3): 299-313

Fenn P, McGuire A, *et al,* (1995) The analysis of censored treatment cost data in economic evaluation. Medical Care; 33(8): 851-63

Fenn P, McGuire A, *et al,* (1996) Modelling programme costs in economic evaluation. Journal of Health Econimics; 15: 115-125

Lin DY, Feuer EJ, *et al,* (1997) Estimating medical costs from incomplete follow-up data. Biometrics: 53(2): 419-434

Lin DY (2000) Regression analysis of incomplete medical cost data. Statistics in Medicine; 22(7): 1181-1200

Lipscomb J, Ancukiewicz M, *et al,* (1998) Predicting the cost of illness: A comparison of alternative models applied to stroke. Medical Decision Making; 18(Supplement 2): S39-S56

APPENDIX A4.2 CENSORED COST RESULTS

Tables A4.2.1 to A4.2.15 present mean total costs and standard errors for each of the 12 censoring methods under different levels of censoring and different censoring mechanisms. In each table the difference between the estimated mean and observed mean, and the estimated standard error and the observed standard error, are presented. Tables A4.2.1 to A4.2.9 also present the sampling standard error (SSE) – the variation in the mean results across 5,000 simulations. The smaller the SSE the more consistent the method is at predicting the mean total cost estimate.

Table A4.2.1 Random censoring (10% censored data)

	Mean total cost	Difference in Means (Estimated – Observed)	Mean Ranking	Standard Error (SE)	Difference in SE (Estimated – Observed)	SE Ranking	SSE
CELT results	**£36,045**			**1,517**			
Cost Histories Known							
WK: Weighted costs	£35,311	-734	7	1,510	-7	1	407
RK: Lin's regression	£37,184	1,139	8	1,047	-470	9	298
LK: Lin's method	£33,785	-2,260	9	1,327	-190	5	214
PC: Partitioned Cox	£20,321	-15,724	10	345	-1,172	10	120
Cost Histories Unknown							
C: Carides	£36,145	100	4	1,295	-222	7	550
IC: Ignoring Censoring	£35,359	-686	6	1,500	-17	2	233
LU: Lin's method	£36,071	26	1	1,557	40	3	545
RU: Lin's regression	£35,863	-182	5	1,134	-383	8	530
WU: Weighted costs	£35,968	-77	3	1,723	206	6	616
CC: Complete cases	£35,981	-64	2	1,671	154	4	413
Cox: Cox regression	£69,266	33,221	11	2,708	1,191	11	1,241
KM: Kaplan-Meier	£133,752	97,707	12	11,651	10,134	12	2,323

Table A4.2.2 Random censoring (30% censored data)

	Mean total cost	Difference in Means (Estimated – Observed)	Mean Ranking	Standard Error (SE)	Difference in SE (Estimated – Observed)	SE Ranking	SSE
CELT results	**£36,045**			**1,517**			
Cost Histories Known							
WK: Weighted costs	£35,219	-826	7	1,505	-12	1	485
RK: Lin's regression	£35,951	-94	3	997	-520	9	462
LK: Lin's method	£33,597	-2,448	9	1,335	-182	4	362
PC: Partitioned Cox	£20,795	-15,250	10	405	-1,112	10	213
Cost Histories Unknown							
C: Carides	£36,179	134	4	1,538	21	2	1,044
IC: Ignoring Censoring	£34,263	-1,782	8	1,469	-48	3	361
LU: Lin's method	£36,052	7	1	1,780	263	6	1,027
RU: Lin's regression	£35,809	-236	6	1,200	-317	7	981
WU: Weighted costs	£35,972	-73	2	1,742	225	5	650
CC: Complete cases	£36,256	211	5	1,892	375	8	967
Cox: Cox regression	£68,594	32,549	11	2,915	1,398	11	1,838
KM: Kaplan-Meier	£136,791	100,746	12	12,347	10,830	12	3,881

Table A4.2.3 Random censoring (50% censored data)

	Mean total cost	Difference in Means (Estimated – Observed)	Mean Ranking	Standard Error (SE)	Difference in SE (Estimated – Observed)	SE Ranking	SSE
CELT results	**£36,045**			**1,517**			
Cost Histories Known							
WK: Weighted costs	£33,333	-2,712	8	1,529	12	1	1,729
RK: Lin's regression	£34,730	-1,316	6	949	-568	7	554
LK: Lin's method	£33,373	-2,672	7	1,307	-210	5	479
PC: Partitioned Cox	£21,385	-14,660	10	505	-1,012	10	303
Cost Histories Unknown							
C: Carides	£36,173	128	2	1,988	471	6	1,692
IC: Ignoring Censoring	£33,072	-2,973	9	1,437	-80	3	430
LU: Lin's method	£35,952	-93	1	2,198	681	8	1,633
RU: Lin's regression	£35,534	-511	4	1,464	-53	2	1,472
WU: Weighted costs	£35,619	-416	3	1,725	208	4	2,412
CC: Complete cases	£36,571	526	5	2,336	819	9	1,397
Cox: Cox regression	£67,836	31,791	11	3,123	1,606	11	2,194
KM: Kaplan-Meier	£140,485	104,440	12	13,210	11,693	12	5,134

Table A4.2.4 End-of-Study censoring (10% censored data, censored from 1.75 years)

	Mean total cost	Difference in Means (Estimated – Observed)	Mean Ranking	Standard Error (SE)	Difference in SE (Estimated – Observed)	SE Ranking	SSE
CELT results	**£36,045**			**1,517**			
Cost Histories Known							
WK: Weighted costs	£35,078	-967	7	1,505	-12	2	1,267
RK: Lin's regression	£37,866	1,821	8	1,076	-441	7	50
LK: Lin's method	£33,882	-2,163	9	1,804	-433	6	35
PC: Partitioned Cox	£21,385	-14,656	10	506	-1,011	10	146
Cost Histories Unknown							
C: Carides	£36,408	363	5	1,284	-233	5	496
IC: Ignoring Censoring	£35,983	-62	2	1,514	-3	1	45
LU: Lin's method	£36,391	346	4	1,552	35	3	480
RU: Lin's regression	£36,059	14	1	1,030	-487	8	454
WU: Weighted costs	£35,898	-147	3	2,016	499	9	1,060
CC: Complete cases	£36,794	749	6	1,694	177	4	50
Cox: Cox regression	£69,583	33,538	11	2,573	1,056	11	100
KM: Kaplan-Meier	£133,339	97,294	12	11,440	9,923	12	1,855

Table A4.2.5 End-of-Study censoring (30% censored data, censored from 1.75 years)

	Mean total cost	Difference in Means (Estimated – Observed)	Mean Ranking	Standard Error (SE)	Difference in SE (Estimated – Observed)	SE Ranking	SSE
CELT results	**£36,045**			**1,517**			
Cost Histories Known							
WK: Weighted costs	£35,838	-207	4	1,549	32	3	1,809
RK: Lin's regression	£37,785	1,740	8	1,072	-445	8	64
LK: Lin's method	£33,248	-2,797	9	1,471	-46	4	3,512
PC: Partitioned Cox	£21,389	-14,656	10	506	-1,011	10	302
Cost Histories Unknown							
C: Carides	£36,414	369	6	1,493	-24	2	946
IC: Ignoring Censoring	£35,935	-110	2	1,511	-6	1	39
LU: Lin's method	£36,368	323	5	1,736	219	5	938
RU: Lin's regression	£36,074	29	1	1,167	-350	6	894
WU: Weighted costs	£35,886	-159	3	1,993	476	9	2,005
CC: Complete cases	£36,718	673	7	1,930	413	7	29
Cox: Cox regression	£69,512	33,467	11	2,563	1,046	11	109
KM: Kaplan-Meier	£135,678	99,633	12	11,417	9,900	12	321

Table A4.2.6 End-of-Study censoring (50% censored data, censored from 1.75 years)

	Mean total cost	Difference in Means (Estimated – Observed)	Mean Ranking	Standard Error (SE)	Difference in SE (Estimated – Observed)	SE Ranking	SSE
CELT results	**£36,045**			**1,517**			
Cost Histories Known							
WK: Weighted costs	£35,861	-184	5	1,506	-11	2	4
RK: Lin's regression	£36,122	77	1	1,033	-484	6	1,084
LK: Lin's method	£34,305	-1,740	9	1,387	-130	3	2,077
PC: Partitioned Cox	£22,345	-13,700	10	2,178	661	8	673
Cost Histories Unknown							
C: Carides	£36,495	424	7	1,864	347	5	1,407
IC: Ignoring Censoring	£35,866	-179	4	1,507	-10	1	42
LU: Lin's method	£36,421	375	6	2,071	554	7	1,451
RU: Lin's regression	£36,170	-125	2	1,379	-138	4	1,272
WU: Weighted costs	£35,895	-150	3	2,478	961	10	1,450
CC: Complete cases	£35,037	-1,008	8	2,340	823	9	73
Cox: Cox regression	£69,409	33,364	11	2,548	1,031	11	123
KM: Kaplan-Meier	£160,124	124,079	12	14,291	12,774	12	482

Table A4.2.7 End-of-Study censoring (10% censored data, censored from 1.25 years)

	Mean total cost	Difference in Means (Estimated – Observed)	Mean Ranking	Standard Error (SE)	Difference in SE (Estimated – Observed)	SE Ranking	SSE
CELT results	**£36,045**			**1,517**			
Cost Histories Known							
WK: Weighted costs	£35,689	-356	7	1,630	113	3	3,102
RK: Lin's regression	£36,253	208	5	1,031	-486	9	1,090
LK: Lin's method	£33,798	-2,247	9	1,352	-165	4	303
PC: Partitioned Cox	£21,501	-14,544	10	2,138	621	10	210
Cost Histories Unknown							
C: Carides	£36,129	84	4	1,292	-225	5	493
IC: Ignoring Censoring	£35,975	-70	3	1,514	-3	1	39
LU: Lin's method	£36,042	-3	1	1,545	28	2	491
RU: Lin's regression	£36,063	18	2	1,035	-482	8	482
WU: Weighted costs	£35,742	-303	6	1,770	253	6	3,144
CC: Complete cases	£36,739	694	8	1,931	414	7	35
Cox: Cox regression	£69,559	33,514	11	2,570	1,053	11	164
KM: Kaplan-Meier	£135,861	99,816	12	11,506	9,989	12	345

Table A4.2.8 End-of-Study censoring (30% censored data, censored from 1.25 years)

	Mean total cost	Difference in Means (Estimated – Observed)	Mean Ranking	Standard Error (SE)	Difference in SE (Estimated – Observed)	SE Ranking	SSE
CELT results	**£36,045**			**1,517**			
Cost Histories Known							
WK: Weighted costs	£35,865	-180	6	1,507	-10	3	69
RK: Lin's regression	£36,183	138	5	1,095	-422	7	2,576
LK: Lin's method	£33,844	-2,201	9	1,077	-440	8	7
PC: Partitioned Cox	£21,611	-14,434	10	2,129	612	10	317
Cost Histories Unknown							
C: Carides	£36,163	118	4	1,517	0	1	939
IC: Ignoring Censoring	£35,859	-186	7	1,508	-9	2	56
LU: Lin's method	£36,094	49	2	1,750	233	5	942
RU: Lin's regression	£36,048	3	1	1,182	-335	6	904
WU: Weighted costs	£35,986	-59	3	1,975	458	9	919
CC: Complete cases	£36,661	616	8	1,687	170	4	64
Cox: Cox regression	£69,368	33,323	11	2,540	1,023	11	239
KM: Kaplan-Meier	£132,631	96,586	12	11,302	9,785	12	485

Table A4.2.10 Informative censoring – Too ill (10th percentile of EQ5D distribution, 13% censored)

	Mean total cost	Difference in Means (Estimated – Observed)	Mean Ranking	Standard Error (SE)	Difference in SE (Estimated – Observed)	SE Ranking
CELT results	**£36,045**			**1,517**		
Cost Histories Known						
WK: Weighted costs	£32,696	-3,349	7	1,549	32	2
RK: Lin's regression	£34,317	-1,728	3	990	-527	9
LK: Lin's method	£31,589	-4,456	9	1,278	-239	7
PC: Partitioned Cox	£19,662	-16,383	10	288	-1,229	10
Cost Histories Unknown						
C: Carides	£35,120	-925	1	1,345	-172	5
IC: Ignoring Censoring	£31,897	-4,148	8	1,439	-78	3
LU: Lin's method	£34,866	-1,179	2	1,740	223	6
RU: Lin's regression	£22,575	-2,470	4	1,078	-439	8
WU: Weighted costs	£33,492	-2,553	6	1,647	130	4
CC: Complete cases	£33,560	-2,485	5	1,495	-22	1
Cox: Cox regression	£72,596	36,551	11	4,245	2,728	11
KM: Kaplan-Meier	£136,841	100,796	12	13,720	12,203	12

Table A4.2.9 End-of-Study censoring (50% censored data, censored from 1.25 years)

	Mean total cost	Difference in Means (Estimated – Observed)	Mean Ranking	Standard Error (SE)	Difference in SE (Estimated – Observed)	SE Ranking	SSE
CELT results	**£36,045**			**1,517**			
Cost Histories Known							
WK: Weighted costs	£35,713	-332	7	1,493	-24	2	280
RK: Lin's regression	£36,092	47	3	1,083	-434	6	2,642
LK: Lin's method	£33,405	-2,640	9	1,404	-113	4	2,485
PC: Partitioned Cox	£21,933	-14,112	10	557	-960	9	2,589
Cost Histories Unknown							
C: Carides	£36,081	36	2	1,926	409	5	1,512
IC: Ignoring Censoring	£35,744	-301	6	1,502	-15	1	62
LU: Lin's method	£36,096	51	4	2,135	618	7	1,576
RU: Lin's regression	£36,076	31	1	1,421	-96	3	1,424
WU: Weighted costs	£35,931	-114	5	2,525	1,008	11	1,476
CC: Complete cases	£34,931	-1,114	8	2,333	816	8	98
Cox: Cox regression	£69,174	33,129	11	2,512	995	10	280
KM: Kaplan-Meier	£159,605	123,560	12	14,381	12,864	12	625

Table A4.2.10 Informative censoring – Too ill (10th percentile of EQ5D distribution, 13% censored)

	Mean total cost	Difference in Means (Estimated – Observed)	Mean Ranking	Standard Error (SE)	Difference in SE (Estimated – Observed)	SE Ranking
CELT results	**£36,045**			**1,517**		
Cost Histories Known						
WK: Weighted costs	£32,696	-3,349	7	1,549	32	2
RK: Lin's regression	£34,317	-1,728	3	990	-527	9
LK: Lin's method	£31,589	-4,456	9	1,278	-239	7
PC: Partitioned Cox	£19,662	-16,383	10	288	-1,229	10
Cost Histories Unknown						
C: Carides	£35,120	-925	1	1,345	-172	5
IC: Ignoring Censoring	£31,897	-4,148	8	1,439	-78	3
LU: Lin's method	£34,866	-1,179	2	1,740	223	6
RU: Lin's regression	£22,575	-2,470	4	1,078	-439	8
WU: Weighted costs	£33,492	-2,553	6	1,647	130	4
CC: Complete cases	£33,560	-2,485	5	1,495	-22	1
Cox: Cox regression	£72,596	36,551	11	4,245	2,728	11
KM: Kaplan-Meier	£136,841	100,796	12	13,720	12,203	12

Table A4.2.11 Informative censoring – Too ill (20th percentile of EQ5D distribution, 31% censored)

	Mean total cost	Difference in Means (Estimated – Observed)	Mean Ranking	Standard Error (SE)	Difference in SE (Estimated – Observed)	SE Ranking
CELT results	**£36,045**			**1,517**		
Cost Histories Known						
WK: Weighted costs	£28,526	-7,519	8	1,482	-35	2
RK: Lin's regression	£30,475	-5,570	6	849	-668	9
LK: Lin's method	£29,032	-7,013	7	1,197	-320	6
PC: Partitioned Cox	£20,075	-15,970	10	280	-1,237	10
Cost Histories Unknown						
C: Carides	£33,896	-2,149	1	1,459	-58	3
IC: Ignoring Censoring	£26,794	-9,251	9	1,348	-169	4
LU: Lin's method	£33,560	-2,485	2	2,041	524	8
RU: Lin's regression	£31,031	-5,014	3	1,237	-280	5
WU: Weighted costs	£30,759	-5,286	5	1,847	330	7
CC: Complete cases	£30,977	-5,068	4	1,494	-23	1
Cox: Cox regression	£74,468	38,423	11	5,295	3,778	11
KM: Kaplan-Meier	£145,445	109,400	12	16,644	15,127	12

Table A4.2.12 Informative censoring – Too well (80th percentile of EQ5D distribution, 21% censored)

	Mean total cost	Difference in Means (Estimated – Observed)	Mean Ranking	Standard Error (SE)	Difference in SE (Estimated – Observed)	SE Ranking
CELT results	**£36,045**			**1,517**		
Cost Histories Known						
WK: Weighted costs	£34,409	-1,636	4	1,524	7	1
RK: Lin's regression	£36,315	270	1	1,013	-507	10
LK: Lin's method	£33,008	-3,037	9	1,284	-233	6
PC: Partitioned Cox	£20,244	-15,801	10	3,215	1698	11
Cost Histories Unknown						
C: Carides	£35,131	-914	2	1,538	-21	2
IC: Ignoring Censoring	£33,671	-2,374	8	1,493	-24	3
LU: Lin's method	£34,753	-1,292	3	1,747	230	5
RU: Lin's regression	£34,300	-1,745	6	1,230	-287	7
WU: Weighted costs	£34,160	-1,885	7	1,847	330	8
CC: Complete cases	£34,401	-1,644	5	1,624	107	4
Cox: Cox regression	£64,998	28,953	11	1,985	468	9
KM: Kaplan-Meier	£130,421	94,376	12	11,244	9,727	12

Table A4.2.13 Informative censoring – Too well (90th percentile of EQ5D distribution, 14% censored)

	Mean total cost	Difference in Means (Estimated – Observed)	Mean Ranking	Standard Error (SE)	Difference in SE (Estimated – Observed)	SE Ranking
CELT results	**£36,045**			**1,517**		
Cost Histories Known						
WK: Weighted costs	£35,627	-418	3	1,482	-35	2
RK: Lin's regression	£37,630	1,585	8	1,065	-452	9
LK: Lin's method	£33,850	-2,193	9	1,397	-120	5
PC: Partitioned Cox	£20,284	-15,761	10	321	-1,196	11
Cost Histories Unknown						
C: Carides	£36,405	360	2	1,432	-85	4
IC: Ignoring Censoring	£35,182	-863	7	1,520	3	1
LU: Lin's method	£35,885	-160	1	1,670	153	6
RU: Lin's regression	£35,259	-786	5	1,150	-367	8
WU: Weighted costs	£35,201	-844	6	1,780	263	7
CC: Complete cases	£35,328	-717	4	1,593	76	3
Cox: Cox regression	£70,863	34,818	11	2,516	999	10
KM: Kaplan-Meier	£132,764	96,719	12	11,321	9,804	12

Table A4.2.14 Partial censoring – One time resource collection (collected 2.25 years after the CELT study first began, 80% censored)

	Mean total cost	Difference in Means (Estimated – Observed)	Mean Ranking	Standard Error (SE)	Difference in SE (Estimated – Observed)	SE Ranking
CELT results	**£36,045**			**1,517**		
Cost Histories Known						
WK: Weighted costs	£34,359	-1,676	2	1,543	26	1
RK: Lin's regression	£35,615	-430	1	999	-518	6
LK: Lin's method	£33,503	-2,542	4	1,232	-285	3
PC: Partitioned Cox	£22,765	-13,280	7	758	-759	9
Cost Histories Unknown						
C: Carides	£10,117	-25,928	8	365	-1,152	10
IC: Ignoring Censoring	£34,120	-1,925	3	1,460	-57	2
LU: Lin's method	£9,361	-26,684	9	871	-646	7
RU: Lin's regression	£42,119	6,704	6	2,917	1,400	11
WU: Weighted costs	£8,369	-27,676	10	1,104	-413	4
CC: Complete cases	£42,195	6,150	5	2,027	510	5
Cox: Cox regression	£64,272	28,227	11	2,171	654	8
KM: Kaplan-Meier	£141,108	105,063	12	12,824	11,307	12

Table A4.2.15 Partial censoring – Fixed resource collection (collected 2.25 years after the end of recruitment at each centre, 15% censored)

	Mean total cost	Difference in Means (Estimated – Observed)	Mean Ranking	Standard Error (SE)	Difference in SE (Estimated – Observed)	SE Ranking
CELT results	**£36,045**			**1,517**		
Cost Histories Known						
WK: Weighted costs	£35,991	-54	2	1,502	-15	3
RK: Lin's regression	£37,912	1,867	7	1,079	-438	8
LK: Lin's method	£33,924	-2,121	8	1,290	-227	6
PC: Partitioned Cox	£20,199	-15,846	10	84	-1,433	11
Cost Histories Unknown						
C: Carides	£37,699	1,654	6	1,233	-284	7
IC: Ignoring Censoring	£36,035	-10	1	1,517	0	1
LU: Lin's method	£37,585	1,540	5	1,506	-11	2
RU: Lin's regression	£37,194	1,149	=3	1,060	-457	9
WU: Weighted costs	£31,456	-4,589	9	1,463	-54	4
CC: Complete cases	£37,194	1,149	=3	1,655	138	5
Cox: Cox regression	£69,374	33,629	11	2,588	1,071	10
KM: Kaplan-Meier	£132,074	96,029	12	11,290	9,773	12

APPENDIX A4.3 RESULTS FROM KENDALL'S CONCORDANCE STATISTIC

Kendall's concordance statistic is a measure of the agreement between several sets of rankings. Given that the test is measuring agreement a significant p-value ($p \leq 0.05$) indicated that the sets of rankings agree with each other, whereas a p-value > 0.05 indicates that the rankings differ. For the CELT data a significant p-value indicates that the ordering of the accuracy of the mean (or standard error) in the presence of censoring does not differ, and a non-significant p-value indicates that the accuracy of methods does differ.

Table A4.3.1 presents the results from comparing the rankings of the mean (or standard error) estimates across different censoring mechanisms, or censoring levels, using Kendall's concordance statistic. The following comparisons were made firstly across the rankings for the mean estimates and secondly across the rankings for the standard error estimate:

- **random censoring:** comparing rankings for mean estimates at 10%, 30% and 50%
- **end of study censoring (1.75 years):** comparing rankings for estimates at 10%, 30% and 50%
- **end of study censoring (1.25 years):** comparing rankings for estimates at 10%, 30% and 50%
- **informative censoring:** too ill (10^{th} and 20^{th} percentiles) and too well (80^{th} and 90^{th} percentiles)
- **informative censoring: too ill** (10^{th} and 20^{th} percentiles)
- **informative censoring: too well** (80^{th} and 90^{th} percentiles)
- **partial censoring:** one time (80%), varied by centre (15%)
- **10% censoring:** random, end of study (1.75 years), end of study (1.25 years), informative too ill (10^{th} percentile)[1], informative too well (90^{th} percentile)[1], partial (varied by centre)[1]
- **30% censoring:** random, end of study (1.75 years), end of study (1.25 years), informative too ill (20^{th} percentile)[2], informative too well (80^{th} percentile)[2]
- **50% censoring:** random, end of study (1.75 years), end of study (1.25 years)

[1] Although informative censoring levels and partial censoring levels are slightly greater than 10% they have been included in this comparison

[2] Although informative censoring levels are slightly greater than 30% they have been included in this comparison

Table A4.3.1 Results from applying Kendall's Concordance across censoring mechanisms or censoring levels to compare agreement in a) mean rankings and b) standard error rankings

	Mean		Standard error	
	Kendall's concordance statistic	p-value	Kendall's concordance statistic	p-value
Random censoring	0.93	0.001	0.84	0.003
End of study censoring (1.75 years)	0.89	0.002	0.89	0.002
End of study censoring (1.25 years)	0.94	0.001	0.82	0.004
Informative censoring (all levels)	0.89	<0.001	0.92	<0.001
Informative censoring: too ill	0.97	0.030	0.95	0.034
Informative censoring: too well	0.90	0.049	0.97	0.030
Partial censoring	0.83	0.074	0.89	0.052
10% censoring	0.67	<0.001	0.85	<0.001
30% censoring	0.79	<0.001	0.88	<0.001
50% censoring	0.82	0.004	0.90	0.002

Table A4.3.2 presents the results after comparing the rankings for the accuracy of the mean estimates with the rankings for the accuracy of the standard error estimates for each individual censoring mechanism by level.

Table A4.3.2 Results from applying Kendall's Concordance for individual censoring mechanisms by censoring levels to compare agreement between mean rankings and standard error rankings

	Kendall's concordance statistic	p-value
Random censoring 10%	0.81	0.082
Random censoring 30%	0.69	0.196
Random censoring 50%	0.69	0.172
End of study censoring 1.75 years 10%	0.78	0.103
End of study censoring 1.75 years 30%	0.82	0.082
End of study censoring 1.75 years 50%	0.75	0.122
End of study censoring 1.25 years 10%	0.83	0.078
End of study censoring 1.25 years 30%	0.77	0.113
End of study censoring 1.25 years 50%	0.75	0.125
Informative censoring: too ill 10^{th} percentile (13%)	0.72	0.144
Informative censoring: too ill 20^{th} percentile (31%)	0.78	0.106
Informative censoring: too well 80^{th} percentile (21%)	0.76	0.115
Informative censoring: too well 90^{th} percentile (14%)	0.83	0.075
Partial censoring: one time (80%)	0.81	0.087
Partial censoring: varied by centre (15%)	0.87	0.060

DIVIDING THE STUDY PERIOD INTO ALTERNATIVE INTERVAL LENGTHS WHEN ESTIMATING MEAN STUDY COSTS (AND STANDARD ERRORS) IN THE PRESENCE OF CENSORING

The weighted cost method (KCH), Lin's regression method (KCH), Lin's method (KCH), partitioned Cox cost method, Carides' method and Lin's method (UCH) required the study time interval to be divided into smaller intervals, in order to estimate mean study costs in the presence of censoring. The accuracy of mean (and standard error) estimates was compared across different choices of interval length. For the weighted cost method (KCH), Lin's regression method (KCH) and the partitioned Cox method mean cost estimates were compared for interval lengths of 2, 3, 6, and 12 months; for Lin's method (KCH), Carides method and Lin's method (UCH) mean cost estimates were compared for interval lengths of 1, 2, 3, 6 and 12 months. The accuracy of methods was examined for five alternative censoring mechanisms:

- 10% random censoring
- 10% end of study censoring (1.75 years)
- 13% informative censoring too ill (10^{th} percentile)
- 14% informative censoring too well (90^{th} percentile)
- 15% partial censoring (fixed time resource collection).

The results are presented in Tables A4.4.1 to A4.4.6 below. In each table the interval length that gives the most accurate estimate of mean total costs are printed in blue, for each censoring mechanism. Thus, for the weighted cost method (Table A4.4.1) under 10% random censoring, dividing the study period into six monthly interval lengths, the most accurate estimates of mean total costs. Interval lengths that gave the least accurate mean cost estimate are printed in red for each censoring mechanism.

interval lengths of 2, 3, 6, and 12 months

	Mean cost	Difference in Means*	Standard error	Difference in SE*
CELT results	**£36,045**		**1,517**	
2 monthly intervals				
10% random	£46,238	10,193	1,877	360
10% end of study (1.75 years)	£48,270	12,225	1,918	401
13% informative: too ill	£42,946	6,901	1,860	343
14% informative too well	£47,136	11,091	1,997	480
15% partial: fixed time	£48,278	12,233	1,982	465
3 monthly intervals				
10% random	£35,311	-734	1,510	-7
10% end of study (1.75 years)	£35,078	-967	1,505	-12
13% informative: too ill	£32,695	-3,349	1,549	32
14% informative too well	£35,627	-418	1,482	-35
15% partial: fixed time	£35,991	-54	1,502	-15
6 monthly intervals				
10% random	£36,133	88	1,691	174
10% end of study (1.75 years)	£36,518	473	1,665	148
13% informative: too ill	£30,994	-5,051	1,650	133
14% informative too well	£34,556	-1,489	1,721	204
15% partial: fixed time	£36,497	452	1,680	163
12 monthly intervals				
10% random	£35,723	-322	1,592	75
10% end of study (1.75 years)	£35,553	-492	1,621	104
13% informative: too ill	£29,474	-6,571	1,533	16
14% informative too well	£32,769	-3,276	1,583	66
15% partial: fixed time	£35,953	-92	1,507	-10

* Estimated – Observed

Table A4.4.2 Lin's regression method (KCH) mean and standard error estimates for interval lengths of 2, 3, 6, and 12 months

	Mean cost	Difference in Means*	Standard error	Difference in SE*
CELT results	**£36,045**		**1,517**	
2 monthly intervals				
10% random	£45,980	9,935	1,116	-401
10% end of study (1.75 years)	£46,722	10,677	1,142	-375
13% informative: too ill	£40,238	4,193	926	-591
14% informative too well	£42,835	6,790	969	-548
15% partial: fixed time	£46,775	10,730	1,145	-372
3 monthly intervals				
10% random	£37,184	1,139	1,047	-470
10% end of study (1.75 years)	£37,866	1,821	1,076	-441
13% informative: too ill	£34,462	-1,583	989	-528
14% informative too well	£37,640	1,595	1,067	-450
15% partial: fixed time	£37,912	1,867	1,079	-438
6 monthly intervals				
10% random	£36,707	662	1,012	-505
10% end of study (1.75 years)	£37,203	1,158	1,033	-484
13% informative: too ill	£34,897	-1,148	962	-555
14% informative too well	£37,433	1,388	1,039	-478
15% partial: fixed time	£37,257	1,212	1,035	-482
12 monthly intervals				
10% random	£39,013	2,968	1,140	-377
10% end of study (1.75 years)	£39,255	3,210	1,152	-365
13% informative: too ill	£39,008	2,963	1,119	-398
14% informative too well	£39,847	3,802	1,173	-344
15% partial: fixed time	£39,398	3,263	1,155	-362

* Estimated – Observed

Table A4.4.3 Lin's method (KCH) mean and standard error estimates for interval lengths of 1, 2, 3, 6, and 12 months

	Mean cost	Difference in Means*	Standard error	Difference in SE*
CELT results	**£36,045**		**1,517**	
1 monthly intervals				
10% random	£36,017	-28	1,535	18
10% end of study (1.75 years)	£36,009	-36	1,519	2
13% informative: too ill	£33,977	-2068	1,563	46
14% informative too well	£36,382	337	1,539	22
15% partial: fixed time	£36,062	-17	1,430	-87
2 monthly intervals				
10% random	£34,602	-1,443	1,318	-199
10% end of study (1.75 years)	£34,644	-1,401	1,329	-188
13% informative: too ill	£32,296	-3,749	1,272	-245
14% informative too well	£34,952	-1,093	1,313	-204
15% partial: fixed time	£34,813	-1,232	1,380	-137
3 monthly intervals				
10% random	£33,785	-2,260	1,327	-190
10% end of study (1.75 years)	£33,882	-2,163	1,084	-433
13% informative: too ill	£31,589	-4,456	1,278	-239
14% informative too well	£33,852	-2,193	1,397	-120
15% partial: fixed time	£33,924	-2,121	1,290	-227
6 monthly intervals				
10% random	£31,688	-4,357	1,274	-243
10% end of study (1.75 years)	£31,905	-4,140	1,283	-234
13% informative: too ill	£28,903	-7,142	1,210	-307
14% informative too well	£31,726	-4,319	1,298	-219
15% partial: fixed time	£31,946	-4,099	1,272	-245
12 monthly intervals				
10% random	£29,212	-6,833	1,212	-305
10% end of study (1.75 years)	£29,578	-6,467	1,224	-293
13% informative: too ill	£26,377	-9,668	1,143	-374
14% informative too well	£29,130	-6,915	1,230	-287
15% partial: fixed time	£29,619	-6,426	1,181	-336

* Estimated – Observed

Table A4.4.4 Partitioned Cox cost method mean and standard error estimates for interval lengths of 2, 3, 6, and 12 months

	Mean cost	Difference in Means*	Standard error	Difference in SE*
CELT results	**£36,045**		**1,517**	
2 monthly intervals				
10% random	£19,240	-16,805	112	-1,405
10% end of study (1.75 years)	£19,115	-16,930	100	-1,417
13% informative: too ill	£18,697	-17,348	106	-1,411
14% informative too well	£19,031	-17,014	101	-1,416
15% partial: fixed time	£19,194	-16,854	103	-1,414
3 monthly intervals				
10% random	£20,321	-15,724	345	-1,172
10% end of study (1.75 years)	£21,385	-14,660	506	-1,011
13% informative: too ill	£19,662	-16,383	288	-1,229
14% informative too well	£20,284	-15,761	321	-1,196
15% partial: fixed time	£20,199	-15,846	84	-1,433
6 monthly intervals				
10% random	£21,969	-14,076	79	-1,436
10% end of study (1.75 years)	£21,651	-14,394	76	-1,441
13% informative: too ill	£21,421	-14,624	81	-1,436
14% informative too well	£22,113	-13,932	81	-1,436
15% partial: fixed time	£21,729	-14,316	79	-1,438
12 monthly intervals				
10% random	£24,198	-11,847	192	-1,325
10% end of study (1.75 years)	£23,439	-12,606	180	-1,337
13% informative: too ill	£23,403	-12,642	88	-1,429
14% informative too well	£24,236	-11,809	84	-1,433
15% partial: fixed time	£23,533	-12,512	81	-1,436

* Estimated – Observed

Table A4.4.5 Carides' method mean and standard error estimates for interval lengths of 1, 2, 3, 6, and 12 months

	Mean cost	Difference in Means*	Standard error	Difference in SE*
CELT results	**£36,045**		**1,517**	
1 monthly intervals				
10% random	£36,122	77	1,294	-223
10% end of study (1.75 years)	£36,091	46	1,283	-234
13% informative: too ill	£34,451	-1,594	1,319	-198
14% informative too well	£35,450	-595	1,319	-198
15% partial: fixed time	£36,874	829	1,267	-250
2 monthly intervals				
10% random	£36,907	862	1,313	-204
10% end of study (1.75 years)	£36,854	809	1,300	-217
13% informative: too ill	£35,029	-1,016	1,374	-143
14% informative too well	£36,381	336	1,423	-94
15% partial: fixed time	£37,670	1,625	1,335	-182
3 monthly intervals				
10% random	£36,145	100	1,295	-222
10% end of study (1.75 years)	£36,408	363	1,284	-233
13% informative: too ill	£35,120	-925	1,345	-172
14% informative too well	£36,405	360	1,432	-85
15% partial: fixed time	£37,699	1,654	1,233	-284
6 monthly intervals				
10% random	£36,091	46	1,283	-234
10% end of study (1.75 years)	£36,659	813	1,306	-211
13% informative: too ill	£35,029	-1,016	1,346	-171
14% informative too well	£36,380	335	1,360	-157
15% partial: fixed time	£37,682	1,637	1,271	-246
12 monthly intervals				
10% random	£36,823	778	1,351	-166
10% end of study (1.75 years)	£36,869	824	1,343	-174
13% informative: too ill	£34,553	1,492	1,272	-245
14% informative too well	£36,379	334	1,386	-131
15% partial: fixed time	£37,689	1,644	1,309	-208

* Estimated – Observed

Table A4.4.6 Lin's method (UCH) mean and standard error estimates for interval lengths of 1, 2, 3, 6, and 12 months

	Mean cost	Difference in Means*	Standard error	Difference in SE*
CELT results	**£36,045**		**1,517**	
1 monthly intervals				
10% random	£36,096	51	1,553	36
10% end of study (1.75 years)	£36,107	62	1,541	24
13% informative: too ill	£34,638	-1,407	1,591	74
14% informative too well	£35,341	-704	1,672	155
15% partial: fixed time	£36,877	832	1,610	93
2 monthly intervals				
10% random	£36,539	494	1,561	44
10% end of study (1.75 years)	£36,576	531	1,549	32
13% informative: too ill	£34,620	-1,425	1,612	95
14% informative too well	£34,682	-363	1,658	141
15% partial: fixed time	£37,419	1,374	1,553	36
3 monthly intervals				
10% random	£36,071	26	1,557	40
10% end of study (1.75 years)	£36,391	346	1,552	35
13% informative: too ill	£34,866	-1,179	1,740	223
14% informative too well	£35,885	-160	1,670	153
15% partial: fixed time	£37,985	1,940	1,493	-24
6 monthly intervals				
10% random	£36,640	595	1,579	62
10% end of study (1.75 years)	£36,659	614	1,559	42
13% informative: too ill	£34,698	-1,347	1,642	125
14% informative too well	£35,759	-286	1,598	81
15% partial: fixed time	£37,518	1,473	1,545	28
12 monthly intervals				
10% random	£36,597	552	1,613	96
10% end of study (1.75 years)	£36,699	654	1,603	85
13% informative: too ill	£34,179	-1,866	1,687	170
14% informative too well	£35,806	-239	1,758	241
15% partial: fixed time	£37,541	1,496	1,596	79

* Estimated – Observed

APPENDIX A5.1 RESULTS FOR A SERIES OF COX PH MODELS FITTED TO THE PBC MAYO COHORT

This appendix presents further results of a series of Cox PH models fitted to the PBC Mayo data.

A5.1.1 Cox PH Model Fitted to the Combined PBC Mayo (N = 312) and PBC CELT (N = 81) data sets

Table A5.1.1 presents the Cox PH results for the model presented in Section 5.3.4.

Table A5.1.1 Regression coefficients and standard errors for a combined PBC Mayo cohort and PBC CELT cohort Cox PH model

	Regression coefficients	Standard Error	Z	p-value
Age	0.03	0.01	3.25	0.001
Gender: female	-0.51	0.24	-2.15	0.032
Ascities present	0.26	0.21	1.22	0.220
Log_e(bilirubin)	0.77	0.10	7.47	< 0.001
Log_e(prothrombin time)	0.70	0.56	1.24	0.210
Oedema score 0.5	0.39	0.22	1.78	0.075
Oedema score 1	0.45	0.24	1.90	0.057
Albumin	-0.65	0.18	-3.57	< 0.001
Study: CELT cohort	3.04	0.41	7.36	< 0.001

Likelihood ratio test = 270 on 9 Degrees of Freedom (DF), $p < 0.001$

The results of the proportionality test are detailed in Table A5.1.2, variables should have a non-significant p-value ($p > 0.05$) should the proportionality assumption hold.

Table A5.1.2 PH results for the for a combined PBC Mayo cohort and PBC CELT cohort Cox model

	χ_1^2	p-value
Age	3.77	0.052
Gender: female	0.09	0.760
Ascities present	0.29	0.594
Log_e(bilirubin)	0.95	0.329
Log_e(prothrombin time)	1.06	0.304
Oedema score 0.5	0.89	0.345
Oedema score 1	3.67	0.056
Albumin	0.16	0.689
Study: CELT cohort	0.85	0.355

A5.1.2 Cox PH Model Fitted to the Combined PBC Mayo (N = 312) data sets

The regression coefficients for the Cox PH model are presented in the main text in Tables 5.2 and 5.3. Table A5.1.3 presents the proportionality tests for the Cox PH results for the model presented in Section 5.4.

Table A5.1.3 PH results for the for the PBC Mayo cohort Cox model

	χ_1^2	p-value
Age	1.05	0.305
Gender: female	0.04	0.837
Ascities present	0.55	0.458
Log_e(bilirubin)	1.68	0.195
Log_e(prothrombin time)	0.00	0.979
Oedema score 0.5	1.46	0.228
Oedema score 1	3.66	0.056
Albumin	0.06	0.811

A5.1.3 Cox PH Model Fitted to the PBC Mayo (N = 312) data sets – excluding prothrombin time and ascities

Table A5.1.4 presents the Cox PH results for the model presented in Section 5.7.3; excluding non-significant variables from the PBC Mayo model (prothombin time and ascities).

Table A5.1.4 Regression coefficients and standard errors for the PBC Mayo cohort Cox PH model (excluding prothrombin time and ascities)

	Regression coefficients	Standard Error	Z	p-value
Age	0.03	0.01	3.08	0.002
Gender: female	-0.60	0.25	-2.40	0.016
Log_e(bilirubin)	0.91	0.11	8.07	< 0.001
Oedema score 0.5	0.44	0.23	1.89	0.059
Oedema score 1	0.67	0.24	-2.82	0.049
Albumin	-0.97	0.21	-4.67	< 0.001

Likelihood ratio test = 245 on 6 DF, $p < 0.001$

The results of the proportionality test are detailed in Table A5.1.5.

Table A5.1.5 PH results for the for the PBC Mayo cohort Cox model (excluding prothrombin time and ascities)

	χ_1^2	p-value
Age	0.59	0.442
Gender: female	0.14	0.708
Log_e(bilirubin)	0.96	0.327
Oedema score 0.5	1.63	0.201
Oedema score 1	3.53	0.060
Albumin	0.00	0.999

A5.1.4 Cox PH Model Fitted to the PBC Mayo (N = 312) data sets – including bilirubin, albumin, age, and ascities

Table A5.1.6 presents the Cox PH results for the model presented in Section 5.7.3; including variables that were common to the PBC Mayo model, Royal Free model, and European model (bilirubin levels, albumin levels, age, and ascities).

Table A5.1.6 Regression coefficients and standard errors for the PBC Mayo cohort Cox PH model (including bilirubin, albumin, age, and ascities)

	Regression coefficients	Standard Error	Z	p-value
Age	0.03	0.01	3.80	< 0.001
Ascities present	0.37	0.20	1.86	0.064
Log_e(bilirubin)	0.94	0.11	8.43	< 0.001
Albumin	-0.96	0.21	-4.67	< 0.001

Likelihood ratio test = 237 on 4 DF, p < 0.001

The results of the proportionality test are detailed in Table A5.1.7.

Table A5.1.7 PH results for the for the PBC Mayo cohort Cox model (including bilirubin, albumin, age and ascities)

	χ_1^2	p-value
Age	1,19	0.276
Ascities present	0.02	0.875
Log_e(bilirubin)	1.74	0.187
Albumin	0.03	0.860

APPENDIX A5.2 S-PLUS COMPUTER CODE FOR ADJUSTING FOR MODEL PARAMETER UNCERTAINY

A5.2.1 S-PLUS Code for adjusting for model parameter uncertainty – using standard errors

This appendix presents the relevant S-PLUS code for obtaining 3,000 Monte Carlo simulations in order to adjust for model uncertainty using published information from Cox PH models on regression coefficients and standard errors [S-PLUS 6, 2001].

The regression coefficients for the log transformation of serum bilirubin, serum albumin, oedema scores 0.5 (oedema present without diuretics or oedema resolved by diuretics), oedema score 1 (oedema present despite diuretics), presence of ascities, female gender, patient age and the log transformation of prothrombin time were: 0.87, -0.93, 0.45, 0.52, 0.19, -0.64, 0.027, and 1.10 and their standard errors were: 0.112, 0.215, 0.234, 0.252, 0.221, 0.250, 0.009, and 0.660, respectively. `rmvnorm` is a function in S-PLUS for simulating randomly generated multivariate normal distributions. In the case of measuring model parameter uncertainty the mean is set to be the value of the regression coefficients from the prognostic model and the standard deviation is set at the standard error of the regression coefficients. The number of simulations run was 3,000.

```
Mayo <- rmvnorm(3000, mean = c(0.87, -0.93, 0.45, 0.52, 0.19, -0.64,
0.027, 1.10), sd = c(0.112, 0.215, 0.234, 0.252, 0.221, 0.25, 0.009,
                                0.66))
```

The risk score for the CELT population is derived by fitting each of the 3,000 Monte Carlo simulated data sets to the PBC CELT data, using the following S-PLUS command:

```
RS <- PBCCELT %*% t(Mayo)
```

where `PBCCELT` is a vector of the serum bilirubin score, serum albumin, oedema, ascities, gender, age, and prothrombin time for the 81 CELT patients with end-stage PBC, and `Mayo` is a vector of the 3,000 simulated regression coefficients after applying the command `rmvnorm`. The adjusted risk score is then obtained:

```
RSadj <- RS - 2.41

ExpSurv <- exp(RSadj)
```

The next step is to obtain estimates for the baseline survival probabilities of surviving to time point t for each individual in the CELT cohort for each of the 3,000 Monte Carlo simulations. The baseline survival probabilities were presented in Table 5.3 and are applied to the CELT cohort at three monthly intervals from 3 months to 60 months.

```
ProbSurv3m <- 0.917^ExpSurv
ProbSurv6m <- 0.835^ ExpSurv
ProbSurv9m <- 0.782^ ExpSurv
ProbSurv12m <- 0.734^ ExpSurv
ProbSurv15m <- 0.686^ ExpSurv
ProbSurv18m <- 0.644^ ExpSurv
ProbSurv21m <- 0.644^ ExpSurv
ProbSurv24m <- 0.632^ ExpSurv
ProbSurv27m <- 0.547^ ExpSurv
ProbSurv30m <- 0.547^ ExpSurv
ProbSurv33m <- 0.547^ ExpSurv
ProbSurv36m <- 0.489^ ExpSurv
ProbSurv39m <- 0.489^ ExpSurv
ProbSurv42m <- 0.489^ ExpSurv
ProbSurv45m <- 0.489^ ExpSurv
ProbSurv48m <- 0.489^ ExpSurv
ProbSurv51m <- 0.462^ ExpSurv
ProbSurv54m <- 0.462^ ExpSurv
ProbSurv57m <- 0.435^ ExpSurv
ProbSurv60m <- 0.379^ ExpSurv
```

The mean probability of survival to time point t is then estimated for the CELT cohort for each of the 3,000 Monte Carlo simulated data sets.

```
CohortMeanSP0m <- rep(1, 3000)
CohortMeanSP3m<- round(apply(ProbSurv3m, 2, mean), 6)
CohortMeanSP6m<- round(apply(ProbSurv6m, 2, mean), 6)
CohortMeanSP9m<- round(apply(ProbSurv9m, 2, mean), 6)
CohortMeanSP12m <- round(apply(ProbSurv12m, 2, mean), 6)
CohortMeanSP15m <- round(apply(ProbSurv15m, 2, mean), 6)
CohortMeanSP18m <- round(apply(ProbSurv18m, 2, mean), 6)
CohortMeanSP21m <- round(apply(ProbSurv21m, 2, mean), 6)
CohortMeanSP24m <- round(apply(ProbSurv24m, 2, mean), 6)
CohortMeanSP27m <- round(apply(ProbSurv27m, 2, mean), 6)
CohortMeanSP30m <- round(apply(ProbSurv30m, 2, mean), 6)
CohortMeanSP33m <- round(apply(ProbSurv33m, 2, mean), 6)
CohortMeanSP36m <- round(apply(ProbSurv36m, 2, mean), 6)
CohortMeanSP39m <- round(apply(ProbSurv39m, 2, mean), 6)
CohortMeanSP42m <- round(apply(ProbSurv42m, 2, mean), 6)
CohortMeanSP45m <- round(apply(ProbSurv45m, 2, mean), 6)
CohortMeanSP48m <- round(apply(ProbSurv48m, 2, mean), 6)
CohortMeanSP51m <- round(apply(ProbSurv51m, 2, mean), 6)
```

```
CohortMeanSP54m <- round(apply(ProbSurv54m, 2, mean), 6)
CohortMeanSP57m <- round(apply(ProbSurv57m, 2, mean), 6)
CohortMeanSP60m <- round(apply(ProbSurv60m, 2, mean), 6)
```

Finally the estimated mean CELT PBC survival in the absence of transplantation is obtained for each of the 3,000 Monte Carlo simulations from the area under the survival curve [Collett, 1994].

```
ExpNonTxSurv <- (3*CohortMeanSP3m)+(3*CohortMeanSP6m)+
(3*CohortMeanSP9m)+(3*CohortMeanSP12m)+(3*CohortMeanSP15m)+
(3*CohortMeanSP18m)+(3*CohortMeanSP21m)+(3*CohortMeanSP24m)+
(3*CohortMeanSP27m)+(3*CohortMeanSP30m)+(3*CohortMeanSP33m)+
(3*CohortMeanSP36m)+(3*CohortMeanSP39m)+(3*CohortMeanSP42m)+
(3*CohortMeanSP45m)+(3*CohortMeanSP48m)+(3*CohortMeanSP51m)+
(3*CohortMeanSP54m)+(3*CohortMeanSP57m)+(3*CohortMeanSP60m)+
        ((CohortMeanSP0m - CohortMeanSP3m)*1.5)+
        ((CohortMeanSP3m - CohortMeanSP6m)*1.5)+
        ((CohortMeanSP6m - CohortMeanSP9m)*1.5)+
       ((CohortMeanSP9m - CohortMeanSP12m)*1.5)+
       ((CohortMeanSP12m - CohortMeanSP15m)*1.5)+
        ((CohortMeanSP15m - CohortMeanSP21m)*3)+
       ((CohortMeanSP21m - CohortMeanSP24m)*1.5)+
       ((CohortMeanSP24m - CohortMeanSP33m)*4.5)+
       ((CohortMeanSP33m - CohortMeanSP48m)*7.5)+
        ((CohortMeanSP48m - CohortMeanSP54m)*3)+
       ((CohortMeanSP54m - CohortMeanSP57m)*1.5)+
        ((CohortMeanSP57m - CohortMeanSP60m)*1.5)
```

The mean area under the curve is equivalent to the 50th percentile of the distribution ExpNonTxSurv, i.e. the mean of the 1,500 and 1,501th observations when the 3,000 Monte Carlo simulations are ordered smallest to largest.

```
(Sort(ExpNonTxSurv))[1500]
```

And the 95% percentile CI is equivalent to the 2.5th and the 97.5th percentiles

```
(Sort(ExpNonTxSurv))[75]
(Sort(ExpNonTxSurv))[2925]
```

Finally the number of life years gained can be obtained by subtracting the mean expected survival length in the absence of transplantation (ExpNonTxSurv) from the mean CELT survival to five-years post-liver transplantation (4.4 years or 53 months).

```
ExpSurvGain <- 53 - ExpNonTxSurv
    (Sort(ExpSurvGain))[1500]
     (Sort(ExpSurvGain))[75]
    (Sort(ExpSurvGain))[2925]
```

A5.2.2 S-PLUS code for adjusting for model parameter uncertainty – using standard errors and the correlation matrix

There is no additional programming required in order to adjust for the correlation between regression coefficients and the standard errors of the regression coefficients. All that is needed is some extra information on the correlation matrix of the regression coefficients, when running the `rmvnorm` function in S-PLUS.

The correlation matrix for the Mayo model is presented in Table 5.5 in Chapter 5, lets call this matrix "`Mayo.cor`"

$$Mayo.cor = \begin{bmatrix} 1.00 & -0.19 & -0.03 & 0.05 & 0.09 & 0.19 & 0.39 & -0.79 \\ -0.19 & 1.00 & -0.07 & -0.09 & 0.10 & -0.19 & 0.01 & 0.33 \\ -0.03 & -0.07 & 1.00 & 0.85 & -0.38 & -0.45 & -0.36 & -0.24 \\ 0.05 & -0.09 & 0.85 & 1.00 & -0.51 & -0.52 & -0.31 & -0.30 \\ 0.09 & 0.10 & -0.38 & -0.51 & 1.00 & -0.00 & -0.38 & -0.30 \\ 0.19 & -0.19 & -0.45 & -0.52 & -0.00 & 1.00 & 0.82 & -0.04 \\ 0.39 & 0.01 & -0.36 & -0.31 & -0.38 & 0.82 & 1.00 & -0.01 \\ -0.79 & 0.33 & -0.24 & -0.30 & -0.30 & -0.04 & -0.01 & 1.00 \end{bmatrix}$$

This extra information was added to `rmvnorm` and, as before, 3,000 simulations were run.

```
Mayo <- rmvnorm(3000, mean = c(0.87, -0.93, 0.45, 0.52, 0.19, -0.64,
  0.027, 1.10), cov = Mayo.cor, sd = c(0.112, 0.215, 0.234, 0.252,
                      0.221, 0.25, 0.009, 0.66))
```

The process described in section A5.2.1 is then repeated in order to obtain the survival gain after transplantation and the probabilities of surviving to time point t, with 95% CIs for model parameter uncertainty.

APPENDIX A6.1 FURTHER DETAILS OF THREE METHODS FOR ESTIMATING INDIVIDUAL PATIENT OUTCOMES

This Appendix describes how Methods 6.4.1 to 6.4.3 are applied to a cohort of patients in order to estimate individual patient outcomes (Chapter 6, Section 6.4). The PBC Mayo prognostic model will be applied to 81 PBC CELT patients to estimate, what would have been, their survival and survival outcome in the absence of liver transplantation over a five-year period. Methods are illustrated for one of the simulation runs.

A.6.1.1 Method 6.4.1: Probability of Survival Equivalent for all Patients, Random Selection of Survivors (p = 0.315)

Table A6.1 presents non-transplant survival estimates for 81 PBC CELT patients, using the PBC Mayo model to predict survival over five years. Survival estimates are first presented prior to allowing for patients survival. Table A6.1 also presents the expected survival outcomes for the 81 patients for one simulation run, where each patient has a 31.5% probability of being selected as surviving to five years. Finally, the table presents each patient's individual survival time, after adjusting the survival times for patients expected to survive for five years.

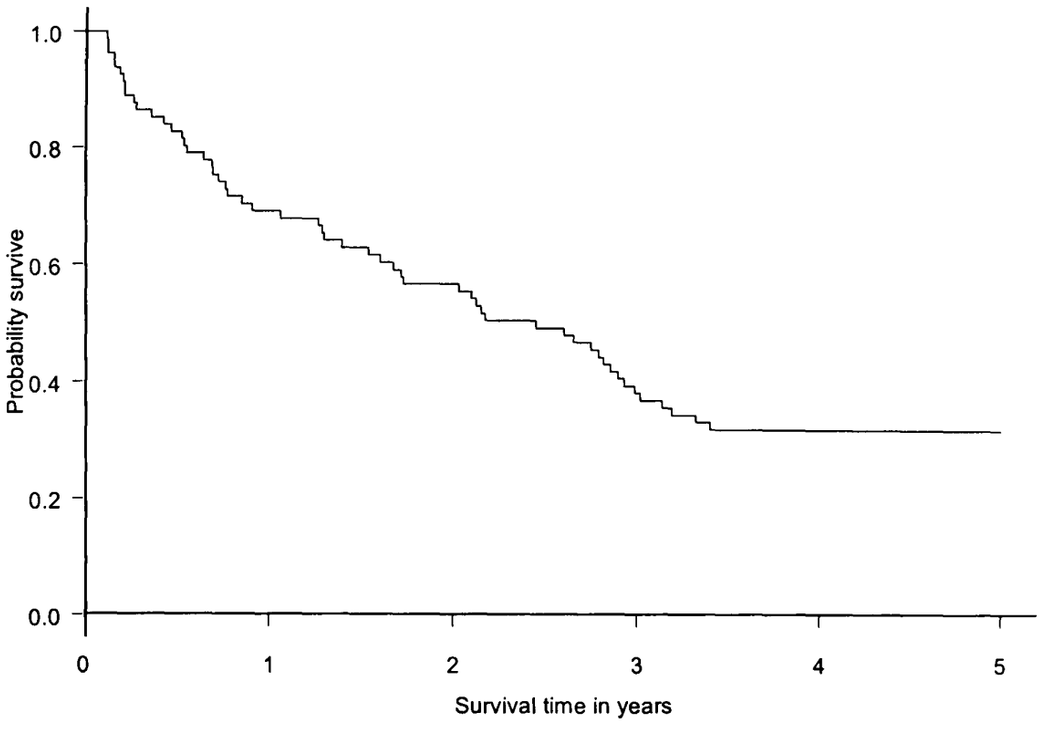
The Kaplan-Meier survival curve for this simulation run is presented in Chapter 6 (Figure 6.1) and the mean survival time for the cohort is 3.0 years.

A6.1.2 Method 6.4.2: Selecting the n Patients with the Longest Survival Times as Survivors (n = 26)

The mean expected number of survivors at five years for the PBC CELT cohort is 26. The 26 patients with the longest survival times are assumed to survive the full five-year study period. Table A6.2 shows which patients would be expected to survive to five years and presents their adjusted survival times, where survivors have an expected survival length of five years and deaths a survival length equivalent to their predicted survival from the PBC Mayo prognostic model.

Figure A6.1 presents the Kaplan-Meier survival curve after selecting the 26 patients with the longest survival times as survivors. The mean survival time for the cohort after applying Method 6.4.2 is 2.59 years.

Figure A6.1 **Kaplan-Meier survival curve for 81 CELT PBC patients where expected individual patient survival is predicted assuming the 26 patients with the longest survival times survive (Method 6.4.2)**

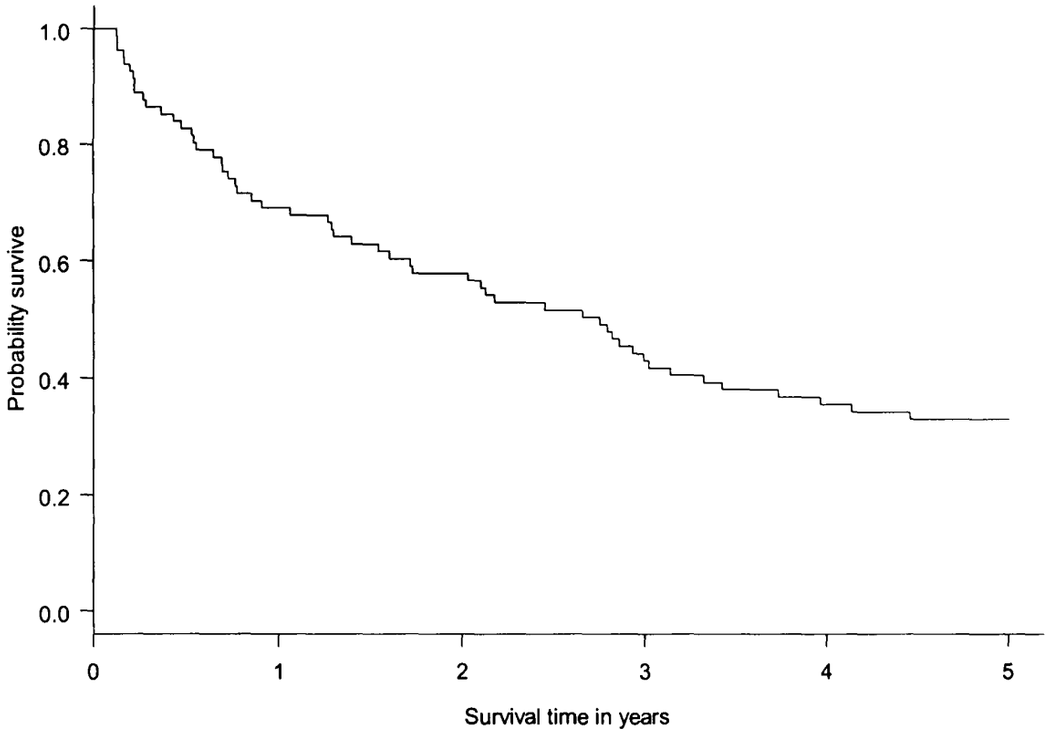


A6.1.3 Method 6.4.3: Probabilistic Sensitivity Analysis (PSA)

Table A6.3 presents the results from one PSA simulation run. Each patient's predicted outcome is estimated using information from the expected non-transplant survival probability at five years. As with the previous two methods, survivors have an expected survival length of five years and deaths a survival length equivalent to their predicted survival from the PBC Mayo prognostic model.

Figure A6.2 presents the Kaplan-Meier survival curve after applying PSA for one simulation run. The mean survival time for the cohort after applying Method 6.4.3 is 2.70 years.

Figure A6.2 Kaplan-Meier survival curve for 81 CELT PBC patients where expected individual patient survival is predicted using PSA to predict individual patient outcomes[1] (Method 6.4.3)



[1] Results from one simulation run

Table A6.1 Estimating individual survival outcomes for 81 CELT PBC patients assuming that the probability of survival equivalent for all patients, random selection of survivors (p = 0.315)

Patient ID	PBC Mayo estimated survival length	Simulated patient outcome	Estimated survival length	Patient ID	PBC Mayo estimated survival length	Simulated patient outcome	Estimated survival length
1	3.42	D	3.42	41	1.73	D	1.73
2	1.72	S	5.00	42	2.45	D	2.45
3	3.19	D	3.19	43	0.45	S	5.00
4	2.99	D	2.99	44	0.56	D	0.56
5	2.17	D	2.17	45	2.89	D	2.89
6	4.32	D	4.32	46	1.30	S	5.00
7	4.92	S	5.00	47	0.27	S	5.00
8	0.47	D	0.47	48	0.85	D	0.85
9	0.12	D	0.12	49	0.78	D	0.78
10	0.19	D	0.19	50	1.60	S	5.00
11	3.01	D	3.01	51	4.53	D	4.53
12	2.13	D	2.13	52	0.22	S	5.00
13	4.61	D	4.61	53	4.27	D	4.27
14	3.80	S	5.00	54	1.40	S	5.00
15	4.62	S	5.00	55	0.54	S	5.00
16	1.27	D	1.27	56	2.82	S	5.00
17	4.24	D	4.24	57	0.70	D	0.70
18	3.52	S	5.00	58	0.16	D	0.16
19	0.91	D	0.91	59	0.28	D	0.28
20	2.85	D	2.85	60	2.10	D	2.10
21	0.13	D	0.13	61	4.67	D	4.67

Table A6.1 (continued)Estimating individual survival outcomes for 81 CELT PBC patients assuming that the probability of survival equivalent for all patients, random selection of survivors (p = 0.315)

Patient ID	PBC Mayo estimated survival length	Simulated patient outcome	Estimated survival length	Patient ID	PBC Mayo estimated survival length	Simulated patient outcome	Estimated survival length
22	4.23	D	4.23	62	2.75	D	2.75
23	0.21	S	5.00	63	3.96	D	3.96
24	2.93	D	2.93	64	3.47	S	5.00
25	0.70	D	0.70	65	3.32	D	3.32
26	1.54	D	1.54	66	4.06	S	5.00
27	0.13	D	0.13	67	0.44	D	0.44
28	0.53	D	0.53	68	4.49	D	4.49
29	3.95	D	3.95	69	1.68	D	1.68
30	4.84	D	4.84	70	4.63	S	5.00
31	3.40	S	5.00	71	4.45	S	5.00
32	4.38	D	4.38	72	0.37	D	0.37
33	2.03	S	5.00	73	2.79	D	2.79
34	4.13	D	4.13	74	2.15	D	2.15
35	3.73	D	3.73	75	4.06	S	5.00
36	1.29	D	1.29	76	0.73	D	0.73
37	2.60	D	2.60	77	0.22	S	5.00
38	0.77	D	0.77	78	1.06	D	1.06
39	0.16	D	0.16	79	2.66	D	2.66
40	3.42	D	3.42	80	3.14	D	3.14
				81	3.54	D	3.54

Table A6.2 Estimating individual survival outcomes for 81 CELT PBC patients assuming that the 26 patients with the longest survival times survive (PBC Mayo prognostic model)

Patient ID	PBC Mayo estimated survival length	Simulated patient outcome	Estimated survival length	Patient ID	PBC Mayo estimated survival length	Simulated patient outcome	Estimated survival length
1	3.42	S	5.00	41	1.73	D	1.73
2	1.72	D	1.72	42	2.45	D	2.45
3	3.19	D	3.19	43	0.45	D	0.45
4	2.99	D	2.99	44	0.56	D	0.56
5	2.17	D	2.17	45	2.89	D	2.89
6	4.32	S	5.00	46	1.30	D	1.30
7	4.92	S	5.00	47	0.27	D	0.27
8	0.47	D	0.47	48	0.85	D	0.85
9	0.12	D	0.12	49	0.78	D	0.78
10	0.19	D	0.19	50	1.60	D	1.60
11	3.01	D	3.01	51	4.53	S	5.00
12	2.13	D	2.13	52	0.22	D	0.22
13	4.61	S	5.00	53	4.27	S	5.00
14	3.80	S	5.00	54	1.40	D	1.40
15	4.62	S	5.00	55	0.54	D	0.54
16	1.27	D	1.27	56	2.82	D	2.82
17	4.24	S	5.00	57	0.70	D	0.70
18	3.52	S	5.00	58	0.16	D	0.16
19	0.91	D	0.91	59	0.28	D	0.28
20	2.85	D	2.85	60	2.10	D	2.10
21	0.13	D	0.13	61	4.67	S	5.00

Table A6.2 (continued)Estimating individual survival outcomes for 81 CELT PBC patients assuming that the 26 patients with the longest survival times survive (PBC Mayo prognostic model)

Patient ID	PBC Mayo estimated survival length	Simulated patient outcome	Estimated survival length	Patient ID	PBC Mayo estimated survival length	Simulated patient outcome	Estimated survival length
22	4.23	S	5.00	62	2.75	D	2.75
23	0.21	D	0.21	63	3.96	S	5.00
24	2.93	D	2.93	64	3.47	S	5.00
25	0.70	D	0.70	65	3.32	D	3.32
26	1.54	D	1.54	66	4.06	S	5.00
27	0.13	D	0.13	67	0.44	D	0.44
28	0.53	D	0.53	68	4.49	S	5.00
29	3.95	S	5.00	69	1.68	D	1.68
30	4.84	S	5.00	70	4.63	S	5.00
31	3.40	D	3.40	71	4.45	S	5.00
32	4.38	S	5.00	72	0.37	D	0.37
33	2.03	D	2.03	73	2.79	D	2.79
34	4.13	S	5.00	74	2.15	D	2.15
35	3.73	S	5.00	75	4.06	S	5.00
36	1.29	D	1.29	76	0.73	D	0.73
37	2.60	D	2.60	77	0.22	D	0.22
38	0.77	D	0.77	78	1.06	D	1.06
39	0.16	D	0.16	79	2.66	D	2.66
40	3.42	S	5.00	80	3.14	D	3.14
				81	3.54	S	5.00

Table A6.3 Estimating individual survival outcomes for 81 CELT PBC patients using information from the individual patient survival probabilities from the PBC Mayo prognostic model to estimate survival (Method 6.4.3: PSA)

Patient ID	PBC Mayo survival probability at five years	PBC Mayo estimate survival length	Simulated patient outcome	Estimated survival length	Patient ID	PBC Mayo survival probability at five years	PBC Mayo estimate survival length	Simulated patient outcome	Estimated survival length
1	0.48	3.42	S	5.00	41	0.11	1.73	D	1.73
2	0.11	1.72	D	1.72	42	0.24	2.45	D	2.45
3	0.42	3.19	S	5.00	43	0.01	0.45	D	0.45
4	0.37	2.99	D	2.99	44	0.002	0.56	D	0.56
5	0.19	2.17	D	2.17	45	0.34	2.89	S	5.00
6	0.76	4.32	S	5.00	46	0.05	1.30	D	1.30
7	0.97	4.92	S	5.00	47	0.000	0.27	D	0.27
8	0.001	0.47	D	0.47	48	0.01	0.85	D	0.85
9	0.000	0.12	D	0.12	49	0.01	0.78	D	0.78
10	0.000	0.19	D	0.19	50	0.09	1.60	D	1.60
11	0.37	3.01	D	3.01	51	0.83	4.53	S	5.00
12	0.18	2.13	D	2.13	52	0.000	0.22	D	0.22
13	0.86	4.61	S	5.00	53	0.74	4.27	S	5.00
14	0.59	3.80	S	5.00	54	0.06	1.40	D	1.40
15	0.86	4.62	S	5.00	55	0.002	0.54	D	0.54
16	0.05	1.27	D	1.27	56	0.32	2.82	D	2.82
17	0.73	4.24	S	5.00	57	0.01	0.70	D	0.70
18	0.51	3.52	S	5.00	58	0.000	0.16	D	0.16
19	0.02	0.91	D	0.91	59	0.000	0.28	D	0.28
20	0.33	2.85	D	2.85	60	0.17	2.10	D	2.10

Table A6.3 (continued)Estimating individual survival outcomes for 81 CELT PBC patients using information from the individual patient survival probabilities from the PBC Mayo prognostic model to estimate survival (Method 6.4.3: PSA)

Patient ID	PBC Mayo survival probability at five years	PBC Mayo estimate survival length	Simulated patient outcome	Estimated survival length	Patient ID	PBC Mayo survival probability at five years	PBC Mayo estimate survival length	Simulated patient outcome	Estimated survival length
21	0.000	0.13	D	0.13	61	0.88	4.67	S	5.00
22	0.73	4.23	S	5.00	62	0.31	2.75	D	2.75
23	0.000	0.21	D	0.21	63	0.64	3.96	D	3.96
24	0.35	2.93	D	2.93	64	0.50	3.47	S	5.00
25	0.01	0.70	D	0.70	65	0.46	3.32	D	3.32
26	0.08	1.54	D	1.54	66	0.68	4.06	S	5.00
27	0.000	0.13	D	0.13	67	0.001	0.44	D	0.44
28	0.002	0.53	D	0.53	68	0.82	4.49	S	5.00
29	0.64	3.95	S	5.00	69	0.10	1.68	S	5.00
30	0.94	4.84	S	5.00	70	0.87	4.63	S	5.00
31	0.48	3.40	S	5.00	71	0.81	4.45	D	4.45
32	0.78	4.38	S	5.00	72	0.000	0.37	D	0.37
33	0.16	2.03	D	2.03	73	0.32	2.79	D	2.79
34	0.70	4.13	D	4.13	74	0.18	2.15	S	5.00
35	0.57	3.73	D	3.73	75	0.68	4.06	S	5.00
36	0.05	1.29	D	1.29	76	0.01	0.73	D	0.73
37	0.28	2.60	S	5.00	77	0.000	0.22	D	0.22
38	0.01	0.77	D	0.77	78	0.03	1.06	D	1.06
39	0.00	0.16	D	0.16	79	0.29	2.66	D	2.66
40	0.48	3.42	D	3.42	80	0.41	3.14	D	3.14
					81	0.52	3.54	S	5.00

the PBC Mayo prognostic model is used as a starting point. The mean probability of survival at five years is 0.315 for the 81 PBC CELT patients.

The mean survival probability can be expressed statistically by assuming that the probability of survival follows a binomial distribution, where the probability of an individual patient surviving to five years is 0.315. Formally, this probability is written as $P_{Si} = 0.315$, where i = 1 to 81 refers to the patient identifier. A random number generator is used to simulate the expected outcome for each individual patient in a cohort. This random number generator produces a number, either zero or one, for each patient. A survival prediction can be generated for each of the 81 cases in the PBC CELT cohort by assuming that the occurrence of a one denotes that the individual will survive the full study period without transplantation and a zero denotes a patient who will die within the five-year study period.

The following computer syntax is used in S-PLUS [S-PLUS 6, 2001] to simulate this process:

```
rbinom(81, 1, 0.315)
```

For one simulation run, a series of estimated events for each of the 81 PBC patients are obtained, as illustrated below:

0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 1, 1, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 1, 1, 0, 0, 1, 0, 1, 0, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 1, 0, 1, 0, 0, 0, 1, 1, 0, 0, 0, 1, 0, 1, 0, 0, 0, 0

The number and proportion of survivors in the cohort can then be calculated. For the simulation run presented above, the number of survivors equals 22, giving an estimated five-year survival of 27% (for this simulation run the percentage of survivors is lower than the "true" percentage of 31.5%).

It is assumed that all estimated survivors survive the full study period and have a survival length of five years. The survival length for patients predicted to die is estimated by applying the PBC Mayo prognostic model to the PBC CELT cohort. The mean non-transplant survival length for the PBC CELT cohort can then be estimated from the area under the Kaplan-Meier survival curve.

Belle SH, Porayko MK, *et al,* (1997) Changes in quality of life after liver transplantation among adults. National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) Liver Transplantation Database (LTD). Liver Transplantation and Surgery; 3(2): 93-104

Bonsel GJ, Essink-Bot M, et al, (1990) Orthotopic liver transplantation in the Netherlands: the results and impact of a medical technology assessment. Health Policy; 16: 147-161

Bonsel GJ, Klompmaker IJ, *et al,* (1990a) Cost-effectiveness analysis of the Dutch liver transplantation programme. Transplantation Proceedings; 22(4): 1481-1484

Bonsel GJ, Klompmaker IJ, *et al,* (1990b) Use of prognostic models for assessment of value of liver transplantation in primary biliary cirrhosis. The Lancet; 335(8688): 493-497

Bonsel GJ, Essink-Bot ML, *et al,* (1992) Assessment of the quality of life before and following liver transplantation. First results. Transplantation; 53(4): 796-780

Burroughs A, *et al,* (1992) Comparative hospital costs of liver transplantation and the treatment of complications of cirrhosis. European Journal of Gastroenterology and Hepatology; 4: 123-128

Christensen E, Gunson B, Neuberger J, (1999) Optimal timing of liver transplantation for patients with primary biliary cirrhosis: use of prognostic modelling. Journal of Hepatology; 30(2): 285-292

Cole CR, Bucuvalas JC, *et al,* (2004) Impact of liver transplantation on HRQOL in children less than 5 years old. Pediatric Transplantation; 8(3): 222-227

De Bona M, Ponton P, *et al,* (2000) The impact of liver disease and medical complications on quality of life and psychological distress before and after liver transplantation. Journal of Hepatology; 33(4): 609-615

Farinati F, Gianni S, *et al,* (2001) Does the choice of treatment influence survival of patients with small hepatocellular carcinoma in compensated cirrhosis? European Journal of Gastroenterology and Hepatology; 13(10): 1217-1224

Gross CR, Malinchoc M, *et al,* (1999) Quality of life before and after liver transplantation for cholestatic liver disease. Hepatology; 29(2): 356-364

Karam V, Castaing D, *et al,* (2003) Longitudinal prospective evaluation of quality of life in adult patients before and one year after liver transplantation. Liver Transplantation; 9(7): 703-711

Kober B, Kuchler T, *et al,* (1991) A psychological support concept and quality of life research in a liver transplantation program: an interdisciplinary multicenter study. Psychotherapy & Psychosomatics; 54(2 to 3): 117-131

Levy MF, Jennings L, *et al,* (1995) Quality of life improvements at one, two, and five years after liver transplantation. Transplantation; 59(4): 515-518

Liermann Garcia RF, Evangelista Garcia C, *et al,* (2001) Transplantation for primary biliary cirrhosis: retrospective analysis of 400 patients in a single center. Hepatology; 33(1): 22-27

Llovet JM, Fuster J, Bruix J, (1999) Intention to treat analysis of surgical treatment for early hepatocellular carcinoma: resection versus transplantation. Hepatology; 30(6): 1434-1440

Longworth L, Bryan S, (2003) An empirical comparison of EQ-5D and SF-6D in liver transplant patients. Health Economics; 12(12): 1061-1067

Longworth L, Young T, *et al,* (2003) Midterm cost-effectiveness of the liver transplantation program of England and Wales for three disease groups. Liver Transplantation; 9(12): 1295-1307