STATISTICAL METHODS FOR THE SIMULTANEOUS ANALYSIS OF QUALITY OF LIFE AND SURVIVAL DATA

Thesis submitted for the degree of

Doctor of Philosophy

at the University of Leicester

by

Lucinda Jane Billingham BSc MSc

Department of Epidemiology and Public Health

University of Leicester

2003

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

STATISTICAL METHODS FOR THE SIMULTANEOUS ANALYSIS OF QUALITY OF LIFE AND SURVIVAL DATA

Lucinda Billingham

Abstract

The aim of the thesis is to critically review, apply and where appropriate develop statistical methodology for the analysis of longitudinal quality of life data collected as part of a clinical trial where survival is also a key endpoint on which treatments are being compared.

The thesis focuses on methods that simultaneously analyse quality of life and survival data, partly in order to provide an overall assessment of the treatments in terms of both endpoints and partly as a means to overcome the problem of missing data that results from patients dropping out of the quality of life study due to death. The thesis also extends the methodology to deal with the additional dropout of patients from the quality of life study prior to death. Three key methods are considered: quality-adjusted survival analysis, multistate modelling and joint modelling.

Quality-adjusted survival analysis compares treatments in terms of a composite measure of quality and quantity of life, created by down-weighting survival time according to the reduction in quality of life experienced by patients. Multistate models describe the movement of patients between various health states, defined by levels of quality of life and death, and explore how treatments differ in terms of the transition rates between health states. Joint models describe the change in quality of life over time and the time to death as two interlinked models.

The key pursuit is the practical application of methods to data and the thesis makes use of two real datasets, from the MIC trial in lung cancer and the ESPAC trial in pancreatic cancer, that encompass the typical problems faced by analysts tackling this type of data in the real world. The results from this research provide statisticians analysing quality of life data with a variety of possible methods for the analysis of such data that should yield unbiased results.

Acknowledgements

There are so many people without whom this research would not have been possible.

My heartfelt thanks to my supervisor Keith Abrams for his inspiration, guidance and patience and for making my seven years of research such a pleasure. My thanks also to colleagues at Leicester for their friendship and inspiration, particularly David Jones for his contribution during the early stages of the research.

I am immensely grateful to everyone at The Cancer Research UK Clinical Trials Unit at Birmingham University and especially Janet Dunn who provided me with the opportunity and management support to pursue this research. Special thanks to the Statistics Team at the Trials Unit for their continuous encouragement and overwhelming support, especially during the final consuming stages.

Thanks to the National Health Service Health Technology Assessment Programme for supporting the research in its early stages and Cancer Research UK for supporting my position as Senior Biostatistician during the course of the research, which I hope will benefit the analysis of clinical trials in the future.

I am indebted to Dr Michael Cullen for kindly allowing me to use the quality of life data from the MIC trial, which provided me with the initial impetus to research this field. I am extremely grateful to Professor John Neoptolemos and Deborah Stocken for generously permitting me access to the ESPAC quality of life data.

A huge thank you to my family and friends for their encouragement and interest and especially my parents, who set me off on the right path in life to be able to get to this point. And finally to Jon, who has been there for me every step of the way and who has amazed me with his never ceasing encouragement, enthusiasm and remarkable support.

CONTENTS

CHAPTER 1: INTRODUCTION

1.1 Aims of the Thesis

The aim of the thesis is to critically review, apply and where appropriate develop statistical methodology for the analysis of longitudinal quality of life data collected as part of a clinical trial where survival is also a key endpoint on which treatments are being compared.

The thesis focuses on methods that can be used to simultaneously analyse quality of life and survival data, partly in order to provide an overall assessment of the treatments in terms of both endpoints and partly as a means to overcome the problem of missing data that results from patients dropping out of the quality of life study due to death. The thesis also extends the methodology to deal with the situation when patients may additionally drop out from the quality of life study prior to death.

Three key methods for the simultaneous analysis of quality of life and survival data are considered; quality-adjusted survival analysis, multistate modelling and joint modelling. Quality-adjusted survival analysis has already been established for use with quality of life and survival data but the methods have not been widely applied to longitudinal quality of life data collected in clinical trials. This thesis provides a complete critical review of all methods in terms of their application to longitudinal quality of life data and develops the methods to deal with dropout prior to death. The methods of multistate modelling and Bayesian joint modelling have only received limited attention in relation to quality of life and survival data and this thesis develops their application in this field.

The key pursuit is the practical application of methods to data, using both classical and Bayesian approaches, and the thesis makes use of two real datasets that encompass many of the typical problems faced by analysts dealing with this type of data in cancer clinical trials. Although the focus here is on clinical trial data, specifically in the field of cancer, the methods can be generalised to other contexts.

1.2 Background

The ideal context for evaluating any new treatment for a disease is a randomised phase III clinical trial. In this, the new treatment is compared to the 'standard' treatment using a number of pre-defined clinically relevant outcomes. In the field of cancer, the primary outcome measure is usually length of survival and increasingly quality of life is being included as a secondary outcome. The role of quality of life in cancer clinical trials has become more prominent as improvements in survival due to treatment are either unlikely to be dramatic or likely to be made at the expense of quality of life. In general, the context for the thesis is randomised phase III clinical trials in cancer but many of the issues will be applicable to other scenarios.

A clinical trial will follow patients over time until they are observed to die or until the time of analysis. During this period of study, the patient's quality of life is measured at various time points, usually via a patient-completed questionnaire designed for the purpose. If the study aims to assess quality of life over a fixed period of time then, the death of a patient during this time will result in the early cessation of assessments and patients are said to 'drop out of the quality of life study due to death'. In studies where patients are sufficiently ill for length of survival to be an outcome, subjects will often drop out of the quality of life study prior to death due to illness.

The assessment of treatments in terms of quality of life is usually based on their effect on this outcome over time. Standard longitudinal analysis assumes that any missing data within the time frame of the analysis, which could be a fixed length or defined by the longest individual follow-up time, are missing at random. In studies of quality of life, all measures that are missing after death within the time frame of the analysis will not be missing at random. In addition, measures that are missing due to dropout prior to death are likely to be health-related and thus non-randomly missing. Application of standard methods to longitudinal quality of life data could therefore give biased results.

Issues relating to the assessment and analysis of quality of life data in clinical trials and the problems of dealing with the non-random missing data were originally highlighted during the late 1980's and early 1990's (Fayers and Jones 1983, Olschewski and

Schumacher 1990, Schumacher et al 1991, Pocock 1991, Cox et al 1992, Fletcher et al 1992, Hopwood et al 1994). The field has been developing ever since with workshops dedicated specifically to the subject (Nayfield et al 1992, Bernhard and Gelber 1998, Everitt et al 2002) and research projects on the subject being sponsored by the National Health Service Health Technology Assessment Programme (Billingham et al 1999). More recently, books dealing with the subject have been published (Staquet et al 1998, Fayers and Machin 2000, Fairclough 2002).

1.3 Rationale for Simultaneous Analysis

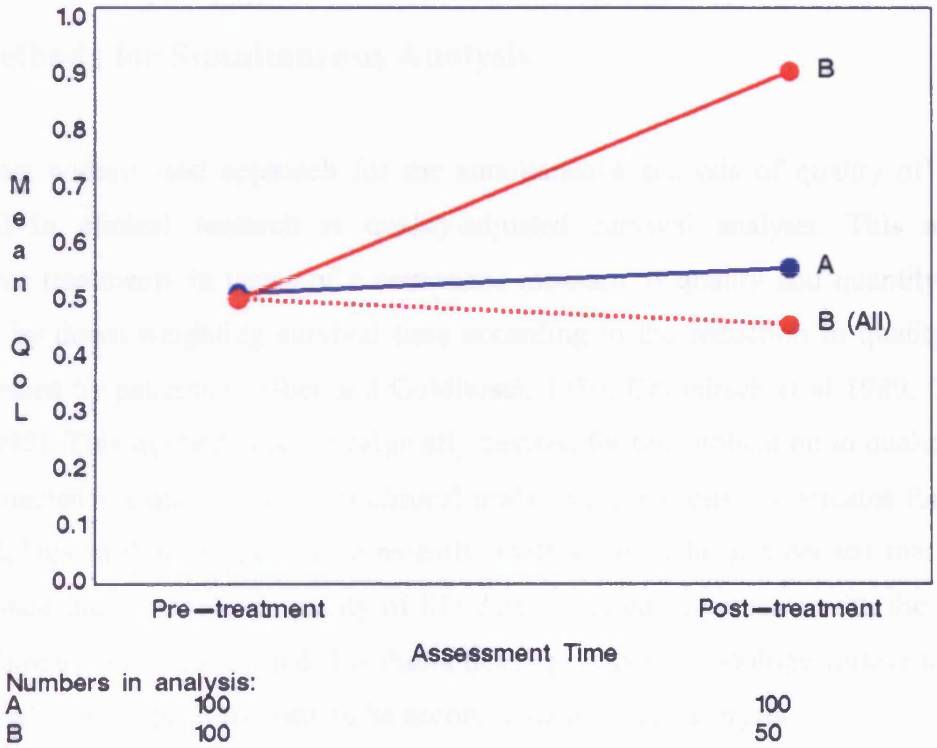
If longitudinal quality of life data are analysed without in any way accounting for any missing data due to dropout or death, then the results at any point in time will represent the quality of life of patients, *conditional* on them being alive and well enough to complete a quality of life assessment. If the aim of a study is to compare changing quality of life for two different treatments and survival or dropout rates differ for the two treatment arms then at any point in time, the comparison of quality of life is not comparing like with like. The only way to make a fair comparison of treatments in terms of quality of life over time is to model the *unconditional* distribution, i.e. that which represents the quality of life of *all* patients, including those that have died. This inclusion of all patients at all time points could be considered equivalent to doing an 'intention-to-treat analysis' which is recognised as the only truly unbiased form of analysis in clinical trials.

In a number of early reviews, simultaneous analysis of quality of life and survival data was suggested as an appropriate approach for the analysis of such data (Fayers and Jones 1983, Olschewski and Schumacher 1990, Schumacher et al 1991, Cox et al 1992). A simultaneous approach allows a more valid assessment of quality of life, since adjusting for the survival data in the longitudinal analysis of quality of life will account for data that are missing within the study time frame as a result of informative dropout due to death. The methods also have the potential to be extended to allow for informative dropout from the quality of life study prior to death, resulting in a further improvement in the validity of the quality of life analysis. As well as the potential to provide a more valid analysis of quality of life, simultaneous analysis may also give

better information to health decision makers. Informed clinical decisions regarding treatment require the quantification and interpretation of differences between treatments in terms of both survival time and change in quality of life measures. Simultaneous analysis enables the assessment of the treatment effect in terms of the two endpoints independently to be supplemented with an assessment in terms of the net effect. This decision may also need to be made in conjunction with other outcomes such as toxicity and cost.

There has been some discussion as to whether the simultaneous analysis of quality of life and survival is appropriate since some argue that it makes no sense to want to estimate the quality of life of patients who are no longer alive (Shih and Quan 1997). The following simple example, illustrated in Figure 1.1, illustrates why simultaneous analysis is necessary for a fair comparison of treatments.

Figure 1.1: Simple example comparing changing quality of life in treatment groups with differential survival rates



Suppose two treatments are being compared, treatment A and treatment B, and treatment A is superior in terms of its effect on survival. Suppose that the aim of the

analysis is to compare treatments in terms of quality of life measured on a 0 to 1 scale at a single fixed time post-treatment and at this time all patients that were originally randomised to treatment A are still alive, including the poorer performance patients, whilst a half of patients on treatment B have died. In general it is the poorer performance patients that will die, leaving the better patients to be assessed for their post-treatment quality of life. For this reason, the post-treatment quality of life of patients on treatment B appears to be better than that on treatment A and, conditional on surviving, it is. However this does not give a true reflection of the comparison of treatments since no account has been made for the differential survival. The only way to truly compare treatments is to include *all* patients in the estimate of post-treatment quality of life, even those that have died. Simultaneous analysis of quality of life and survival will allow this. Simplistically, the dead patients can be included in the analysis by allocating them the worst quality of life score of 0 for their post-treatment assessment (red dotted line in Figure 1.1). The intention to treat analysis, which includes all patients on both treatment arms, then shows that in fact treatment A is superior to treatment B.

1.4 Methods for Simultaneous Analysis

The most widely used approach for the simultaneous analysis of quality of life and survival in clinical research is quality-adjusted survival analysis. This approach compares treatments in terms of a composite measure of quality and quantity of life, created by down-weighting survival time according to the reduction in quality of life experienced by patients (Gelber and Goldhirsch 1986, Goldhirsch et al 1989, Glasziou et al 1990). This method was not originally devised for the application to quality of life data collected via questionnaire in clinical trials and this thesis investigates further the methodology in this context. More recently, methods have been proposed that directly incorporate the longitudinal quality of life data collected on patients with the survival data (Glasziou et al 1998) and this thesis develops this methodology further to enable additional dropout prior to death to be accommodated in the analysis.

Multistate modelling was advocated in a number of the early reviews and discussions on quality of life as a possible means of analysing quality of life and survival

simultaneously (Fayers and Jones 1983, Schumacher et al 1991, Cox et al 1992, Abrams 1992a, Olschewski et al 1992). Multistate models describe the movement of patients between various health states, defined by levels of quality of life and death, and explore how treatments differ in terms of the transition rates between health states. The application of this method to longitudinal quality of life and survival data measured in a clinical trial has not previously been explored in detail and this thesis investigates the validity of the approach in this context and develops the methodology to deal with the problem of additional dropout prior to death.

There is an increasing literature on the joint modelling of longitudinal and event-time data. The application of joint models to quality of life and survival data has been developing within a classical framework (Fairclough et al 1998a, Ribaudo et al 2000, Curran et al 2002, Michiels et al 2002, Fairclough 2002, Pauler et al 2003). The change in quality of life over time and the time to death can be considered as two simultaneous processes occurring in patients, and can be modelled as such. The Bayesian approach to joint modelling has been described by a number of authors (Berzuini 1995, Faucett and Thomas 1996, Carpenter et al 2002, Xu and Zeger 2001, Wang and Taylor 2001) but until recently this approach has not been considered for quality of life and survival data (Wang et al 2002). The application of Bayesian joint models to quality of life and survival data is developed in this thesis and the method is extended to accommodate additional dropout prior to death.

1.5 The Need for a Bayesian Approach to the Analysis

There are two general approaches to statistical analysis, one is known as classical or frequentist and the other as Bayesian (Bland and Altman 1998). Simplistically, statistical analysis is generally interested in estimating an unknown quantity, say θ. Conventional statistical analysis uses a classical approach, which would typically produce an estimate and confidence interval for θ and a probability, known as p-value, associated with comparing the data with a hypothesised value of θ. The alternative Bayesian approach, based on Bayes' theorem, combines prior knowledge regarding the value of θ with the likelihood of the data to produce an estimate of the posterior probability distribution of θ on which all inferences are made. Further details on

Bayesian methodology for clinical trials are given elsewhere (Spiegelhalter et al 2003) and are summarised in Chapter 2.

In estimating an unknown quantity, there are a number of reasons why one may choose to take a Bayesian approach. Some may use it because it allows prior information about θ to be included in the estimation process, with the prior density for θ based on either subjective beliefs or external evidence (from other studies) or a combination of the two. Others use it because, with the results being in the form of a posterior distribution, it allows direct probability statements about the unknown quantity to be made thus giving more clinically relevant conclusions. In many situations it is used because it allows complex analyses to be carried out in a relatively straightforward manner. Computational advances have facilitated the practical application of the Bayesian approach (Gilks et al 1994). In this thesis no attempt is made to include informative prior information since the primary aim for using a Bayesian analysis is to overcome technical difficulties with some of the more complex analyses. Classical approaches to such analyses are generally possible and will be discussed but it is believed that a Bayesian approach provides a more user-friendly and flexible environment with easier implementation via readily available software.

1.6 Background to the Example Datasets

Data from two different trials are used to illustrate and investigate the methodology under consideration. The background to these two trials and the associated quality of life studies are described below. Further details on the measures of quality of life and the timing of the quality of life assessments are given in Chapter 4, whilst the extent of missing data in the quality of life studies is described in Chapter 5.

1.6.1 The MIC Trial

In 1988, two phase III trials were initiated at the Cancer Research UK Clinical Trials Unit in Birmingham to evaluate the role of mitomycin, ifosfamide and cisplatin (MIC) chemotherapy in the treatment of non-small cell lung cancer. The trials were run concurrently and had identical designs and eligibility criteria except for stage of disease,

with the MIC1 trial recruiting patients with localised disease and the MIC2 trial recruiting those with extensive stage disease. The results from both trials are reported elsewhere (Cullen et al 1999) and this thesis uses quality of life and survival data from just the MIC2 trial, henceforth referred to here as the MIC trial.

In the MIC trial, patients were randomly allocated to receive either standard palliative treatment for the relief of symptoms, usually radiotherapy (the PAL arm), or MIC chemotherapy, up to a maximum of 4 courses, followed by standard palliative care (the CT arm). The aim of the trial was to compare treatments in terms of both survival and quality of life. Quality of life was an important endpoint because both treatments were considered largely palliative and MIC chemotherapy was considered by some clinicians to be highly toxic.

The trial closed in March 1996, at which time 351 eligible patients had been randomised. The median survival time for patients in the trial was approximately 6 months. For practical reasons associated with the availability of the research nurse, the quality of life component of the study was carried out only on a subset of trial patients, essentially consisting of patients treated at three main oncology centres. Quality of life data were collected on 109 patients from the trial, 67 on the chemotherapy arm and 42 on the palliative care arm. A questionnaire designed specifically for the trial yielded a global quality of life score derived from the mean of the responses to the 12 questions on the questionnaire. Quality of life was only assessed during the treatment phase of the trial at a maximum of five time points, at approximately 3-weekly time intervals. In total 392 questionnaires were available for analysis.

1.6.2 The ESPAC Trial

The first trial undertaken under the auspices of the European Study Group for Pancreatic Cancer (ESPAC) was a phase III trial investigating the use of both adjuvant (i.e. post-operative) chemotherapy and adjuvant chemoradiotherapy in patients with resectable pancreatic cancer and is referred to here as the ESPAC trial. The trial closed in April 2000 at which point 541 eligible patients with resected pancreatic cancer had been recruited. The primary endpoint for the trial was survival, which was defined as time from surgery to death from any cause. The interim results from the trial, showing a

potential survival benefit for chemotherapy (hazard ratio of 0.66 with 95% confidence interval 0.52 to 0.83), have been published (Neoptolemos et al 2001) and final analysis of the trial is currently being prepared for publication.

Quality of life was considered an important secondary endpoint in the trial and data was collected from trial entry to death on a subgroup of patients who agreed to participate in the study. Some preliminary analysis of the quality of life data was published in the main trial paper (Neoptolemos et al 2001). The data included in this thesis consist of patients from the chemotherapy versus no chemotherapy comparison only and are taken from an early version of the quality of life data. The results included in this thesis will therefore not be properly representative of the trial results and should not be used for clinical evaluation but since the data are from a questionnaire that is one of the most commonly used in cancer clinical trials, they are appropriate for illustrating the methodology in the thesis.

Quality of life data were available for 175 patients, 87 who were randomised to receive chemotherapy (the CT arm) and 88 randomised to no chemotherapy (the NoCT arm). Chemotherapy consisted of 6 cycles of 5-FU plus D-L folinic acid taken monthly. Quality of life was measured using a questionnaire designed by the European Organisation into Research of Treatment for Cancer called the EORTC QLQ-C30 (Aaronson et al 1993). This yields 15 different measures of quality of life including a global health status score, which will be the focus here. Questionnaires were completed at approximately 3-monthly intervals from trial entry to death. The quality of life over time will be affected initially by the patients' recovery from surgery and then by the effect of up to 6 months of chemotherapy for those patients randomised to this arm. In total 710 questionnaires were available for analysis.

1.6.3 Rationale for Inclusion of Specific Examples in the Thesis

Data from the MIC trial and the ESPAC trial are used to illustrate and investigate the methodology under consideration. The design of a quality of life study and the associated analytical issues vary depending on the disease site and treatment under investigation. The trials here relate to the treatment of chemotherapy in two different types of cancer, non-small cell lung cancer and pancreatic cancer. Both trials investigate

the use of a treatment that is likely to have a major impact on patients' quality of life and both involve patients with a disease that, despite treatment, has the potential to progress rapidly resulting in illness and relatively short survival times (median survival time of approximately 6 and 16 months for MIC and ESPAC respectively). This is the typical setting for the inclusion of quality of life as an important endpoint.

The two studies are very different in terms of their design and hence the type of quality of life data for analysis. The MIC study assesses quality of life using a non-standard instrument whose overall mean score provides a reasonably continuous measure of quality of life for analysis. The ESPAC study uses one of the most widely used quality of life questionnaires in cancer clinical trials, the EORTC QLQ-C30 (Aaronson et al 1993), whose global health status score provides a measure of quality of life with a more discrete distribution. The MIC study assesses quality of life for a maximum of five time points during a fixed period of time of 15 weeks, the treatment phase of the trial, whilst in contrast the ESPAC study investigates quality of life every three months from trial entry to death.

1.7 Outline of Thesis

The basic methodology for Bayesian analysis and survival analysis are given in Chapters 2 and 3 respectively, as background information for the reader. Chapters 4 and 5 discuss the nature of quality of life data and the specific problem being addressed here, that of missing data. Chapter 6 considers standard methods for the analysis of longitudinal quality of life data with the focus on the possible models for quality of life data over time. These methods could give biased results in the presence of missing data and this is discussed in the latter part of the chapter. Chapter 7 discusses the possible models that could be used to model the survival data and considers the inclusion of quality of life in such models as a covariate that may explain some of the variation in the survival data. Chapters 8, 9 and 10 are the key chapters in the thesis relating to the three different approaches to simultaneous analysis of quality of life and survival; quality-adjusted survival analysis, multistate modelling and joint modelling. These chapters use the models for quality of life and survival that were introduced in Chapters 6 and 7. The final chapter summarises and discusses the findings from the research and

describes the potential for further work. All key methodology is illustrated using data from the two different studies.

CHAPTER 2: PRINCIPLES OF BAYESIAN ANALYSIS

2.1 Introduction

Some of the models considered for the quality of life and survival data in this thesis are fitted to the data using a Bayesian as well as a classical approach. The aim of this chapter is not to provide a comprehensive review of Bayesian methods but to provide the reader with the fundamental knowledge needed for the understanding of later chapters. There is a wide range of literature that describes Bayesian methods in greater detail (Lee 1989, Bernardo and Smith 1994).

The fundamentals of a Bayesian analysis, as described by others (Bland and Altman 1998, Spiegelhalter et al 1999), are as follows. Given that there is an unknown quantity of primary interest relating to a population, say θ, classical methods regard θ as a fixed unvarying quantity whilst Bayesian methods are based on the idea that since the true value of θ is unknown, it has a probability distribution. A Bayesian analysis combines prior information on θ, in the form of a probability distribution, with the likelihood of the data given θ, to produce a posterior probability distribution for θ. This result is obtained using Bayes theorem. These essential elements of Bayesian analysis are described in Sections 2.2, 2.3 and 2.4. Making inferences about the value of θ requires integration of the posterior distribution, which is often difficult, and in most cases impossible, to evaluate analytically. One option is to use Monte Carlo integration and in particular Markov chain Monte Carlo integration and this is described in Section 2.5

2.2 Likelihood

Suppose that observed data $Y = \{y_1, y_2, ..., y_n\}$ are known to be random, independent observations from a specific type of probability distribution $p(Y)$ but the parameter (or vector of parameters) of the distribution, represented by θ, are unknown. Then, since the observations are independent, the probability of observing the given data is given by:

$$p(Y \mid \theta) = p(y_1, y_2, ..., y_n \mid \theta) = p(y_1 \mid \theta)p(y_2 \mid \theta)...p(y_n \mid \theta) \qquad [2.1]$$

This is the known as the 'likelihood' and since the data values (i.e. $y_1, y_2, ..., y_n$) are known, it is a function of θ and is therefore often represented by $L(\theta)$.

The likelihood function expresses the extent to which different values of θ are supported by the data (Spiegelhalter et al 2003). The *maximum likelihood estimate* of θ is the value of θ that maximises the likelihood, i.e. it is the value of θ that makes the observed data most likely. Ranges of values that are best supported by the data are also desirable and these are given in the form of classical confidence intervals or Bayesian credible intervals.

2.3 Prior Distributions

To be able to estimate the posterior distribution for a parameter, specification of the prior distribution for that parameter is required. If possible, it is convenient to choose the prior distribution to be one that is 'conjugate' to the likelihood function, i.e. one that when combined with the likelihood produces a posterior distribution from the same family as the prior. For example, if a normally distributed prior is used for the unknown mean of a normally distributed likelihood then the posterior distribution will also be normal and if a gamma prior is used for the unknown mean of a Poisson likelihood then the posterior will be gamma.

The prior distribution may be the personal opinion of a single individual, may represent the beliefs of a community, may be a summary of existing evidence, or may be a combination. Here vague priors (also called reference, non-informative or flat priors) are the distribution of choice, that is prior distributions that cover such a wide range of possible values that they provide virtually no information about the unknown parameter. The choice of priors will be discussed in the context of specific models. With a vague prior distribution and a sufficiently large sample, the amount of information contributed by the prior relative to the likelihood is negligible and thus the posterior distribution will be similar to the likelihood function and hence the Bayesian results will be comparable to the classical results (Spiegelhalter et al 2003).

Choosing a vague prior is not necessarily straightforward and different choices could lead to different results, particularly in terms of the variance components of a model and with small amounts of data (Lambert et al 2003a). As recommended by Lambert et al (2003a), priors have been chosen to be vague within a realistic range and sensitivity analysis has been used to assess the influence of the choice of prior on the results.

2.4 Bayesian Inference Derived from Bayes Theorem

Bayes theorem, which is used widely in diagnostic testing, is derived from the basic rules of probability and essentially reverses conditional probabilities. For two events a and b, the theorem states:

$$p(b \mid a) = \frac{p(a \mid b)p(b)}{p(a)} \qquad [2.2]$$

So for parameter (or vector of parameters) given by θ and observed data Y,

$$p(\theta \mid Y) = \frac{p(Y \mid \theta)p(\theta)}{p(Y)} \qquad [2.3]$$

In equation [2.3], $p(\theta \mid Y)$ is the posterior distribution of θ, $p(Y \mid \theta)$ is the likelihood and $p(\theta)$ is the prior distribution. The quantity $p(Y) = \int p(Y \mid \theta)p(\theta)d\theta$ in the denominator of [2.3] is the marginal distribution of the data and is not generally of interest since it does not contain θ, it just ensures that $p(\theta \mid Y)$ is a proper probability distribution in that $\int p(\theta \mid Y)\, d\theta = 1$. The fact that $p(\theta \mid Y) \propto p(Y \mid \theta)p(\theta)$ means that the posterior distribution has the same shape as the likelihood times the prior distribution and this forms the basis for all Bayesian inference.

Bayesian inference essentially involves estimating features, such as mean and variance, of the posterior distribution. These can all be expressed in terms of posterior expectations of functions of θ (Gilks et al 1996). The posterior expectation of a function $g(\theta)$ is given by:

$$E[g(\theta) \mid Y] = \int g(\theta) p(\theta \mid Y) d\theta = \int \frac{g(\theta) p(Y \mid \theta) p(\theta)}{p(Y)} d\theta \qquad [2.4]$$

So for example to estimate the mean of the posterior distribution for θ requires the following calculation:

$$E[\theta \mid Y] = \int \theta \; p(\theta \mid Y) d\theta = \int \frac{\theta \; p(Y \mid \theta) p(\theta)}{p(Y)} d\theta \qquad [2.5]$$

and the variance of the posterior distribution for θ requires:

$$\begin{aligned} Var[\theta \mid Y] &= E[\theta^2 \mid Y] - (E[\theta \mid Y])^2 \\ &= \int \frac{\theta^2 p(Y \mid \theta) p(\theta)}{p(Y)} d\theta - \int \frac{\theta \; p(Y \mid \theta) p(\theta)}{p(Y)} d\theta \end{aligned} \qquad [2.6]$$

The posterior distribution for θ enables direct probability statements to be made about the value of the unknown parameter relative to some clinically relevant value c, i.e.

$$p(\theta > c \mid Y) = \int_c^{\infty} p(\theta \mid Y) d\theta = \int_c^{\infty} \frac{p(Y \mid \theta) p(\theta)}{p(Y)} d\theta \qquad [2.7]$$

Interval estimation should supplement point estimates in Bayesian analysis as confidence intervals do in classical analysis. A 100(1-α)% (usually α=0.05) credible interval (CrI) is an interval in which θ lies with probability (1-α). The lower and upper bounds of the interval (c_L, c_U) are usually estimated from the posterior distribution such that

$$p(\theta < c_L \mid Y) = \alpha/2 \text{ and } p(\theta > c_U \mid Y) = \alpha/2$$

$$\text{i.e. } \int_{-\infty}^{c_L} \frac{p(Y \mid \theta) p(\theta)}{p(Y)} d\theta = \alpha/2 \text{ and } \int_{c_U}^{\infty} \frac{p(Y \mid \theta) p(\theta)}{p(Y)} d\theta = \alpha/2 \qquad [2.8]$$

For skewed posterior distributions it may be preferable to use a highest posterior density (HPD) interval whose values are calculated without assuming equal areas in each tail but such that the values contained within the interval have the highest posterior densities compared to those outside the interval (Spiegelhalter et al 2003). CrI and HPD have a meaning that is often incorrectly ascribed to classical confidence intervals, i.e. they are intervals within which θ lies with 100(1-α)%.

These integrations required for Bayesian inference are often difficult, and in most cases impossible, to evaluate (Gilks et al 1996). Most analyses will involve multi-parameter problems, in which interest is focused on estimating the marginal likelihood for each parameter and therefore the other 'nuisance' parameters need to be integrated out. This integration is usually impossible to evaluate analytically and an alternative approach is required. One option is to use Monte Carlo integration and in particular Markov chain Monte Carlo integration as discussed below.

2.5 Markov Chain Monte Carlo

As described elsewhere (Gilks et al 1996), Monte Carlo integration evaluates $E[g(\theta)]$ by drawing samples $\{\theta_t ; t = 1,...m)$ from the posterior distribution and then the population mean of the function $g(\theta)$ is estimated by the sample mean as follows:

$$E[g(\theta)] = \frac{\sum_{t=1}^{m} g(\theta_t)}{m} \qquad [2.9]$$

In Markov chain Monte Carlo each value is sampled from a distribution that depends only on the previous value, thus θ_t is sampled from the distribution $p(\theta_t \mid \theta_{t-1}, Y)$. The process starts with an arbitrary initial value θ_0 and continues for enough iterations to allow the process to converge to a unique 'stationary' distribution, which should be the posterior distribution for θ. The values sampled during this initial period known as the 'burn-in' are discarded. Once the stationary distribution has been reached, the sampling process continues and enough samples taken to give an accurate estimate of the posterior distribution. Estimates of $E[g(\theta)]$ can be calculated from these retained

sampled values using equation [2.9] which enable features of the posterior distribution to be estimated.

The choice of starting values should not influence the stationary distribution but may influence the speed and ease of convergence to it. The length of the burn-in period required depends on the rate of convergence. Sensitivity analysis should be carried out to assess the influence of starting values and length of the burn-in period. There are a variety of tests to assess convergence but visual inspection of the plot of the sampled values (the trace) is the most commonly used method (Gilks et al 1996). Ideally the trace should move smoothly around a focused range of values without 'sticking' in any one area. Autocorrelations between sampled values at increasing lags should be checked to establish if the sampling process is moving around the stationary distribution adequately. With high autocorrelations a larger number of samples may be needed to fully represent the posterior distribution. Otherwise re-parameterisation of the model to centre covariates may be used to overcome the problem or thinning the chain by sampling values at regular intervals (Gilks et al 1996).

Gibbs sampling is a Markov chain Monte Carlo approach to numerical integration (Gelfand and Smith 1990). It is a technique for sampling from the joint posterior distribution of the unknown quantities in a statistical model (Gilks et al 1994). The Gibbs sampling algorithm is based on the fact that for a given a set of unknown parameters $\theta_1, \theta_2, \theta_3, \ldots, \theta_k$, as the number of iterations approach infinity, the marginal distributions are uniquely determined by the full conditional distributions (Gelfand and Smith 1990). The full conditional distribution of an unknown parameter θ_i (i=1, ...,k) is its conditional distribution conditioning on the values of all other unknown parameters in the model, that is $f(\theta_i|\theta_1, \theta_2, \ldots, \theta_{i-1}, \theta_{i+1}, \ldots, \theta_k)$.

The Gibbs sampler is an algorithm for extracting the marginal distributions from these full conditional distributions. Given an arbitrary starting value $\boldsymbol{\theta}^0 = (\theta_1^0, \theta_2^0, \ldots, \theta_k^0)$, the Gibbs sampler then successively makes random drawings from the full conditional distributions such that for the tth iteration (t=1,2,3,...) the sampled values of the parameters $\boldsymbol{\theta}^t = (\theta_1^t, \theta_2^t, \ldots, \theta_k^t)$ are obtained by sampling:

θ_1^t from $f(\theta_1 \mid \theta_2^{t-1}, \theta_3^{t-1}, \theta_4^{t-1}, ..., \theta_k^{t-1})$

θ_2^t from $f(\theta_2 \mid \theta_1^t, \theta_3^{t-1}, \theta_4^{t-1}, ..., \theta_k^{t-1})$

θ_3^t from $f(\theta_3 \mid \theta_1^t, \theta_2^t, \theta_4^{t-1}, ..., \theta_k^{t-1})$

.....

θ_k^t from $f(\theta_k \mid \theta_1^t, \theta_2^t, \theta_3^t, ..., \theta_{k-1}^t)$

Repeated iterations of this cycle produces a sequence of sampled values $\{\boldsymbol{\theta}^0, \boldsymbol{\theta}^1, \boldsymbol{\theta}^2, ...\}$ which will converge to the true joint posterior distribution. Once in the stationary phase the Gibbs sampler will yield correlated samples from this distribution. It should be noted that when the parameters relate to models with covariates, the covariates should be centred to reduce the posterior correlations between parameters and hence the number of iterations required by the Gibbs sampler (Gilks et al 1996). In this thesis, WinBUGS software (Spiegelhalter et al 2000) is used to estimate posterior distributions of model parameters using Gibbs sampling.

2.6 Summary

Given a model with a set of related unknown parameters, a classical analysis is interested in the estimating the fixed value of each of the parameters. In a Bayesian analysis, the aim is to estimate the posterior distribution for each parameter, given the data and any prior information. The fact that the posterior distribution is proportional to the likelihood times the prior forms the basis for all Bayesian inference. Gibbs sampling can be used to estimate the posterior distribution of each parameter by sampling each from its full conditional distribution. The aim of this thesis is to fit specific models to the data using a Bayesian approach and not necessarily to compare different models in terms of adequacy of fit in the Bayesian framework, although this is possible using Bayes factors, the Deviance Information Criterion and model averaging (Spiegelhalter et al 2003).

In all the analyses included in this thesis, prior distributions for parameters are intended to be vague, an approach referred to as a 'reference Bayes approach' (Spiegelhalter et al 2003) and sensitivity analysis to the choice of specific prior is included. The use of

vague priors will be considered by some purists to not be a proper Bayesian analysis. The main aim however of introducing a Bayesian approach to the analysis here is to allow complex models to be estimated in a relatively straightforward way. Although it is not the aim of the thesis to incorporate prior information into the analysis, the methodology could be extended to do this if so desired.

CHAPTER 3: BASIC SURVIVAL ANALYSIS

3.1 Introduction

The aim of this chapter is to summarise the basic methods for investigating survival data that will be relevant to later chapters. In particular, the definition and estimation of survivor and hazard functions, either parametrically or non-parametrically, is applicable to the models for survival data that are considered both separately and in simultaneous analysis with quality of life.

Measurements of the time between two events, an initial occasion and an endpoint of interest, are generally known as *survival* or *time-to-event data*. In a clinical trial, the initial occasion will usually be defined as the same event for all individuals, such as date of randomisation, date of starting treatment or date of diagnosis. The endpoint of interest will depend on the nature of the disease and treatments under investigation, but is often death or relapse from a period of disease remission. This thesis is specifically interested in the analysis of quality of life and survival data, where by survival we mean time to death, however we are also interested in the time to other events and specifically time to dropout from the quality of life study.

Survival data are different from other types of continuous data because during the period of a study the endpoint of interest is not necessarily observed in all subjects. This may occur because: (i) some patients are lost-to-follow-up, i.e. are not followed to the end of the study and when last seen have not experienced the event of interest; or (ii) the event has not occurred in some patients by the time the study closes for analysis. Such data are referred to as *censored* survival times and are different from missing data in that they provide a lower bound for the actual non-observed survival times. Any analysis carried out on survival data should use statistical methods that do not disregard censored data, and indeed make the fullest possible use of it to avoid loss of information.

Most analytical methods used for survival data with censored observations are only valid if censoring is non-informative. This means that the censoring is not related to any

factors associated with the actual survival time, i.e. the actual survival time, t, of an individual is independent of any mechanism which causes that individual's survival time to be censored at time c, where $c<t$ (Collett 1994). When the censoring mechanism is not independent of survival time, *informative censoring* occurs and standard methods used for survival analysis are invalidated. Informative censoring is a particular problem when analysing individual measures summarising a patient's quality of life data over time and specifically for subject-based approaches to quality-adjusted survival analysis and will be discussed further in Chapters 6 and 8.

There is a large literature describing methods for the analysis of survival data (e.g. Kalbfleisch and Prentice 1980, Collett 1994, Parmar and Machin 1995). The aim of this chapter is to summarise the basic methods for investigating survival data that will be relevant to later chapters, in particular Chapters 7-10. Section 3.2 introduces the survivor and hazard functions, which form the basis for all survival analysis. These functions can be estimated non-parametrically as described in Sections 3.3 and 3.4 or parametrically when the data are believed to follow a known distribution as described in Section 3.5. All of these sections give an initial insight into the survivor and hazard functions for the MIC and ESPAC data. Section 3.6 provides a summary of the chapter and highlights the relevance of findings to later chapters. Models for survival data, which give most flexibility and will be most useful in the simultaneous analysis, are described later in Chapter 7.

3.2 Survivor and Hazard Functions

Survival data is generally described and modelled in terms of two related functions, the *survivor function* and the *hazard function*. The survivor function, $S(t)$, represents the probability that an individual survives from the time origin to some time beyond t. It is given by:

$$S(t) = p(T > t) = 1 - F(t) = 1 - \int_{u=0}^{t} f(u)\, du \qquad [3.1]$$

where T is the random variable representing survival time, and the distribution of survival times is described by $f(t)$, the probability density function, and $F(t)$, the associated cumulative distribution function.

The hazard function, $h(t)$, is defined as:

$$h(t) = \lim_{\Delta\tau \to 0}\left(\frac{P(\tau < t < \tau + \Delta t \mid \tau \leq t)}{\Delta\tau}\right) \qquad [3.2]$$

It represents the instantaneous death rate for an individual surviving to time t and is linked to the probability density function and survivor function by:

$$h(t) = \frac{f(t)}{S(t)} = -\frac{d}{dt}\log S(t) \qquad [3.3]$$

The survivor function and hazard function can be estimated from observed data. If the form of $f(t)$ is not specified then non-parametric procedures can be used, otherwise parametric models can be fitted to the data.

3.3 Non-Parametric Estimation of the Survivor Function

3.3.1 Kaplan-Meier Method

The survivor function can be estimated non-parametrically from observed data, both censored and uncensored, using the Kaplan-Meier method. This method is also called the product-limit method and is based on maximum likelihood estimation (Kaplan and Meier 1958). Suppose deaths occur at distinct times $t_1<t_2<...<t_j<...<t_n$ then the Kaplan-Meier estimate of the survivor function is given by

$$\hat{S}(t) = \prod_{j=1}^{k}\left(1 - \frac{d_j}{n_j}\right) \qquad \text{for all } t_k \leq t \qquad [3.4]$$

where n_j is the number individuals alive just before time t_j and d_j is the number of deaths at time t_j. Survival times censored at time t_j are assumed to occur immediately after the death time when computing the values of the t_j. Confidence intervals for the survivor function can be calculated using a variety of different methods (Collett 1994, Parmar and Machin 1995).

The calculation of Kaplan-Meier estimates is based on the assumption that the deaths of the individuals in the sample occur independently of one another. This allows the probabilities of surviving one interval to the next to be multiplied together to give the survivor function. It should also be noted that the Kaplan-Meier method gives the maximum likelihood estimate of the survivor function only if deaths and censoring are independent (Kaplan and Meier 1958). Thus, for unbiased Kaplan-Meier estimates, it is necessary for the censoring mechanism to be non-informative.

Kaplan-Meier estimates of the survivor function $S(t)$ can be plotted against time t as a survival curve. The survival curve is a stepped plot with the survivor function dropping instantaneously at each time of death and remaining level between successive death times. These provide a useful summary of the data and can be used to determine summary statistics. Median survival time is the value of t for which $S(t)=0.5$ whilst mean survival time is given by the area under the curve. The value for mean survival time is biased if the largest observed survival time is censored. By restricting the area under the curve to a specific time period whose upper limit is less than or equal to the largest observed death time, then the mean within this time period will be unbiased. In SAS, the LIFETEST procedure (SAS Institute 1992) calculates Kaplan-Meier estimates, plots survival curves and gives summary statistics.

3.3.2 Non-Parametric Comparison of Survivor Functions

Plotting Kaplan-Meier survival curves for different groups of patients can be used to compare the groups descriptively, along with the summary statistics such as the median. A more formal comparison of survivor functions can be made using various non-parametric tests. Survival in two or more groups of patients can be compared by using a non-parametric test such as the *log-rank test*, also called the *Mantel-Cox test* (Parmar

and Machin 1995, Collett 1994). This is the most widely used method for comparing survival curves.

The method essentially calculates at each death time, for each group, the expected number of deaths under the null hypothesis of no difference between groups. These are then summed to give the total expected number of deaths in each group, say E_i for group i. The log-rank test compares the observed number of deaths in each group, say O_i for group i, to the expected number by calculating the test statistic

$$\chi^2 = \sum_{i=1}^{g} \frac{(O_i - E_i)^2}{E_i} \qquad [3.5]$$

and comparing it to a chi-square distribution with g-1 degrees of freedom where g is the number of groups.

Alternatively the Mantel-Haenszel version of the log-rank test and weighted Mantel-Haenszel versions may be more suitable than the log-rank test when the number of events is small or the assumption of proportional hazards is not valid for the alternative hypothesis (Parmar and Machin 1995).

3.3.3 Non-Parametric Estimation of Survivor Functions in the MIC Study

Survival time in the MIC trial was defined as time from date of entry to trial to date of death from any cause. Of the 351 patients in the MIC trial, all except 7 patients were dead at the time of analysis. The Kaplan-Meier estimates of the survivor function for all patients (N=351) on each treatment arm in the MIC trial is shown in Figure 3.1. The curves suggest that chemotherapy is beneficial to survival. The median survival time for chemotherapy was 6.7 months (95% confidence interval (CI): 5.4 to 8.0) compared to 4.8 months (95%CI: 4.0 to 5.7) for standard palliative care. A log-rank test shows that this difference in survival is statistically significant at the 5% significance level (p=0.03).

This thesis is specifically interested in the patients in the quality of life study (n=109). The Kaplan-Meier survival curves for these patients are shown in Figure 3.2a. All

patients in the analysis are dead except two with censored survival times greater than 36 months. The medians were 7.9 months versus 4.2 months for CT and PAL respectively and a log-rank test showed the difference in survival to be statistically significant at the 5% level (p=0.003). The subgroup of patients in the quality of life study had a more extreme treatment difference in terms of survival than the overall trial. Quality of life data were only collected during the treatment phase of the trial, that is the first 15 weeks from entry (see Section 4.5), and therefore the survival during this time is particularly important. Figure 3.2b shows the Kaplan-Meier curves for the first 15 weeks of the trial. During this time all patients had full follow-up in terms of survival and 28 patients died, 12 on CT and 16 on PAL. The curves suggest that survival is superior on the CT arm during this time. Medians were not reached but the mean survival time within 15 weeks was 13.9 (standard error = 0.35) for the CT arm and 12.8 (standard error = 0.56) for the PAL arm. If all patients with survival greater than 15 weeks are censored at 15 weeks then a log-rank test shows this difference to be statistically significant at the 5% level (p=0.02).

Figure 3.1 Kaplan-Meier survivor functions for all MIC trial patients (N=351; circles indicate censored survival times)

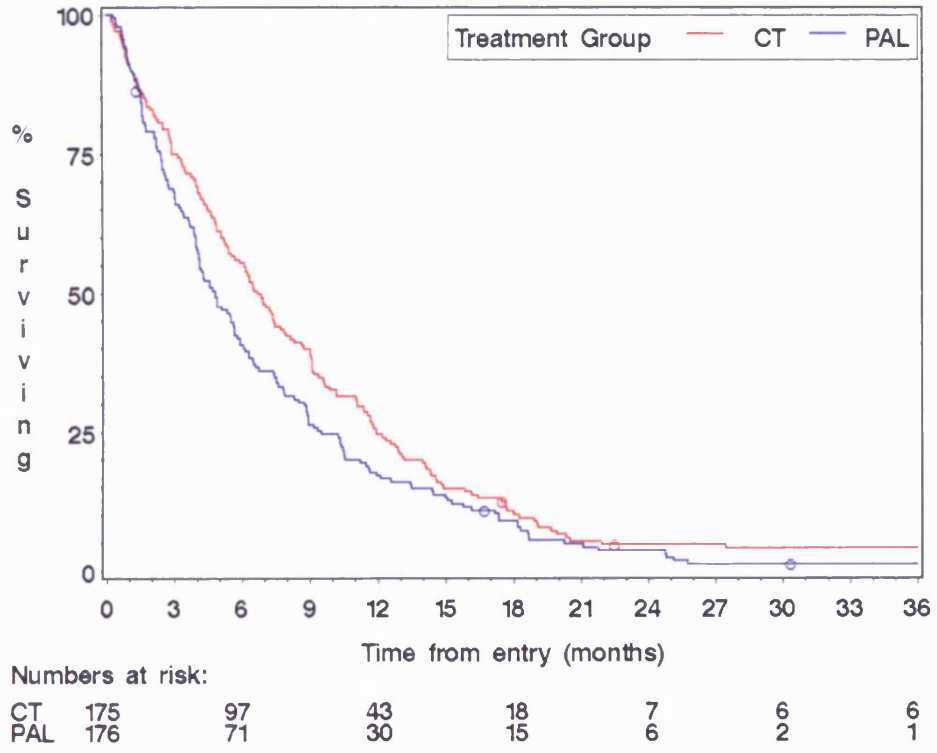
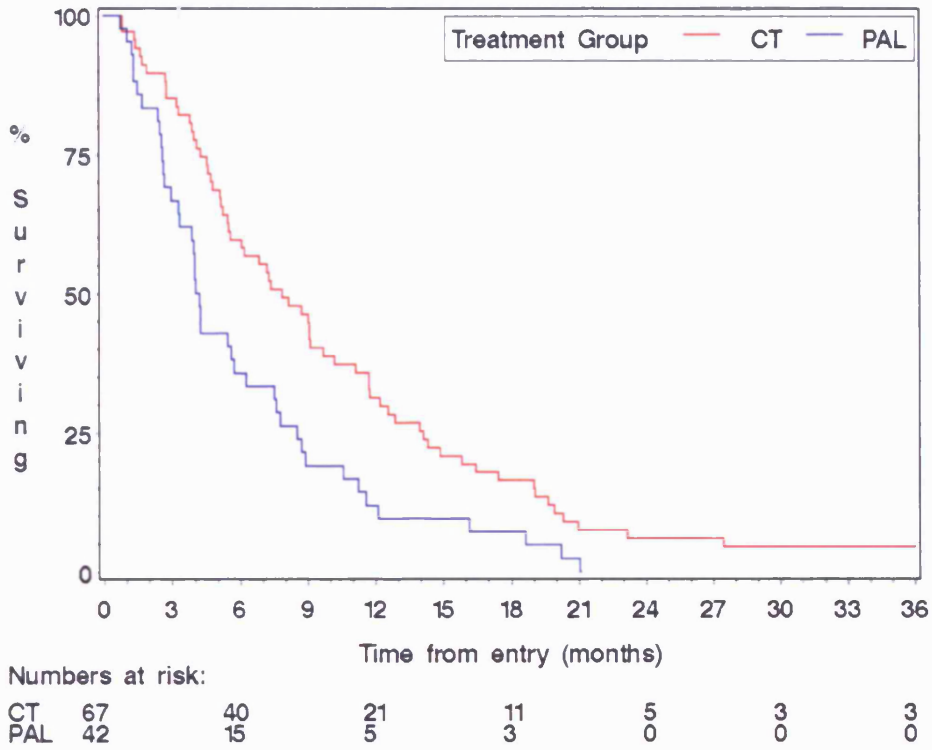
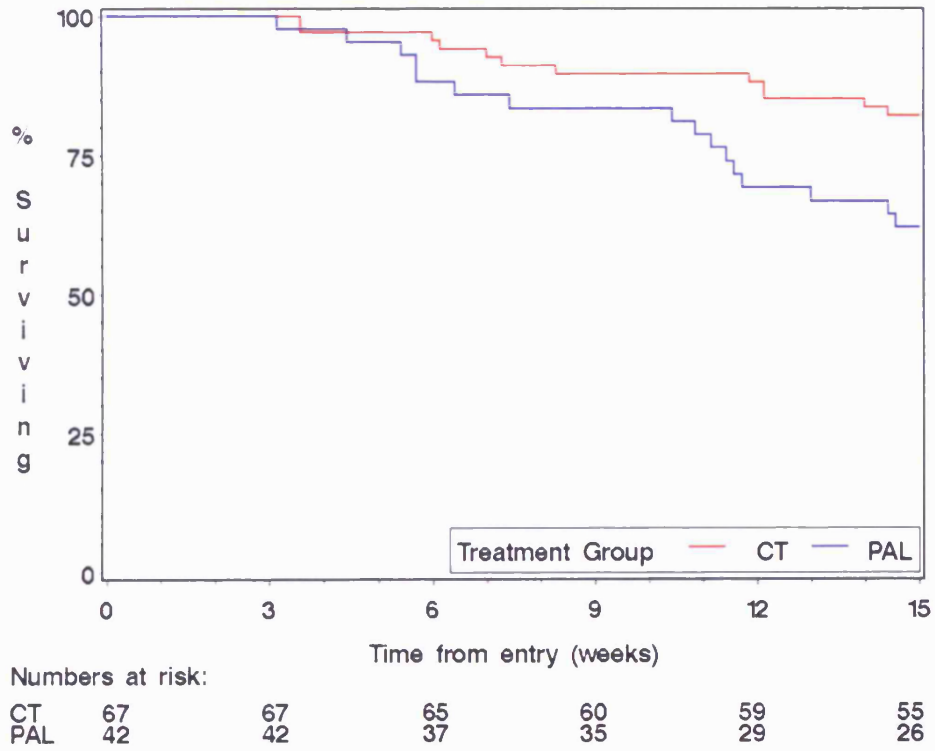


Figure 3.2 Kaplan-Meier survivor functions for patients in MIC quality of life study (N=109)

(a) Over three years



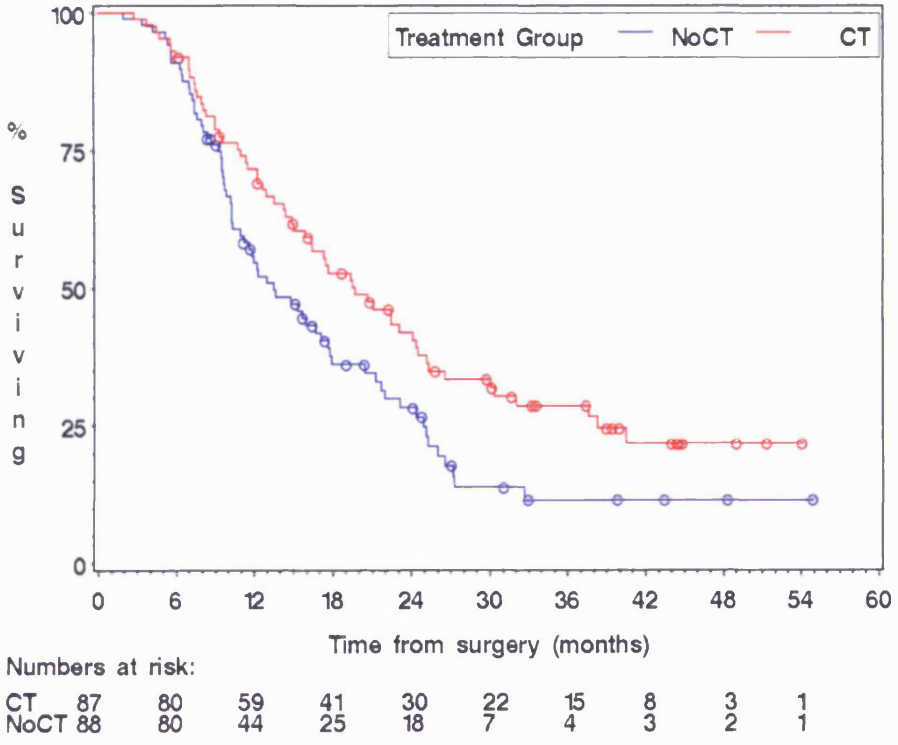
(b) Within 15-week analysis period



3.3.4 Non-Parametric Estimation of Survivor Functions in the ESPAC Study

Survival time in the ESPAC trial was defined as time from date of surgery to date of death from any cause. Of the 175 patients in the ESPAC study, 126 (72%) patients were dead at the time of analysis of which 105 died within 2 years of surgery. The Kaplan-Meier estimates of the survivor function for each treatment arm are shown in Figure 3.3, with censored survival times shown by circles. Some patients are censored relatively early on in the trial, with 9 censored within 1 year (4 on CT and 5 on NoCT) and 22 censored within 2 years (10 on CT and 12 on NoCT). The curves suggest that chemotherapy is beneficial to survival. The median survival time for CT was 19.7 months compared to 13.5 months for NoCT. Mean survival within 24 months was 17.3 (standard error=0.79) for CT compared to 15.0 (standard error=0.78) for NoCT. A log-rank test showed the difference in survival to be statistically significant at the 5% significance level (p=0.02) as it was in the main trial (p=0.0005; Neoptolemos et al 2001).

Figure 3.3 Kaplan-Meier survivor functions for patients in the ESPAC study (circles indicate censored survival times)



3.4 Non-Parametric Estimation of the Hazard Function

3.4.1 Method and Application to the MIC and ESPAC Data

The hazard function can be estimated non-parametrically by dividing time into a series of time intervals and estimating the hazard rate in each interval. The hazard rate for a specific time interval is the risk of death within the interval given survival to the start of the interval. It is estimated by dividing the number of deaths by the total exposure to the risk of death within the interval. The total exposure to the risk of death can be calculated exactly by summing the individual follow-up times within the interval.

In the Kaplan-Meier method for the estimation of the hazard function (Collett 1994), the time intervals are defined by the actual death times. Suppose deaths occur at distinct times $t_1<t_2<...<t_j<...<t_n$ then the Kaplan-Meier estimate of the hazard function at time t where $t_j \leq t < t_{j+1}$ is given by

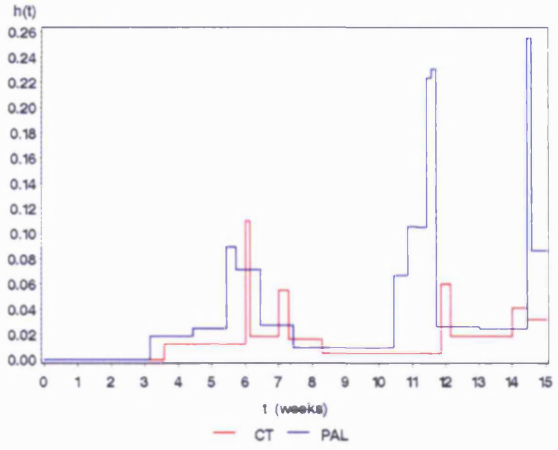
$$\hat{h}(t) = \frac{d_j}{n_j \tau_j} \qquad \text{for } t_j \leq t < t_{j+1} \qquad [3.6]$$

where n_j is the number at risk at time t_j, d_j is the number of deaths in the interval and τ_j is the width of the interval. This assumes the hazard function remains constant between successive death times and that patients who are censored within the interval are actually at risk for the whole interval. These are the same assumptions that are made when calculating the equivalent Kaplan-Meier estimate of the survivor function.

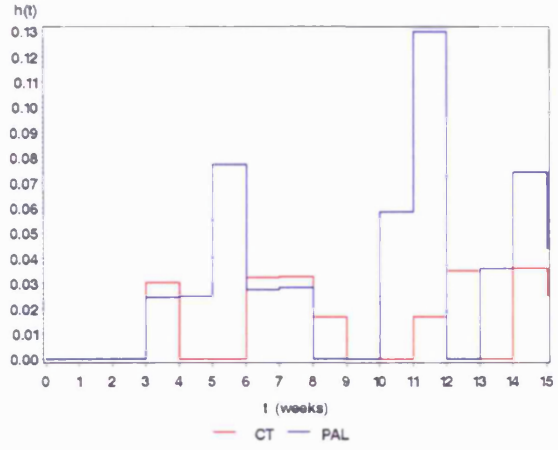
Estimates of the hazard function for the MIC and ESPAC studies are calculated using the Kaplan-Meier method and fixed-width time intervals (see Figures 3.4 and 3.5). The hazard functions for the MIC study are shown for the first 15 weeks from trial entry. During this time the hazard is generally higher for PAL compared to CT and seems to rise over time for PAL and remain relatively constant for CT. For the ESPAC study the hazard functions are shown for 24 months from trial entry. Hazards on both arms rise initially with the NoCT arm rising to a higher level. The hazard then seems to reach a plateau for the CT arm and reduce slightly for the NoCT arm.

Figure 3.4: Estimates of the hazard functions for the MIC study within 15 weeks of entry to trial (see Figure 3.2b for relevant numbers at risk over time)

(a) Kaplan-Meier estimate



(b) Based on weekly hazard rates



(c) Based on 3-weekly hazard rates

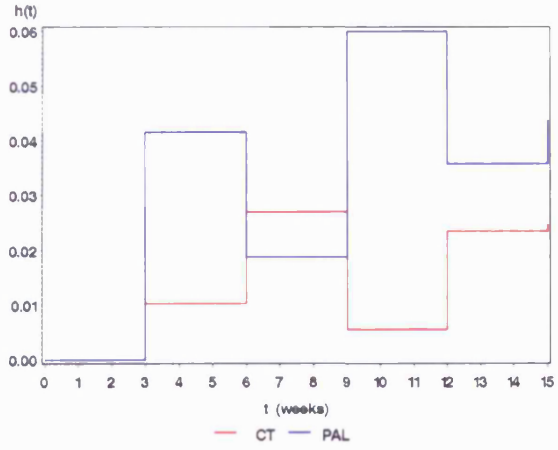
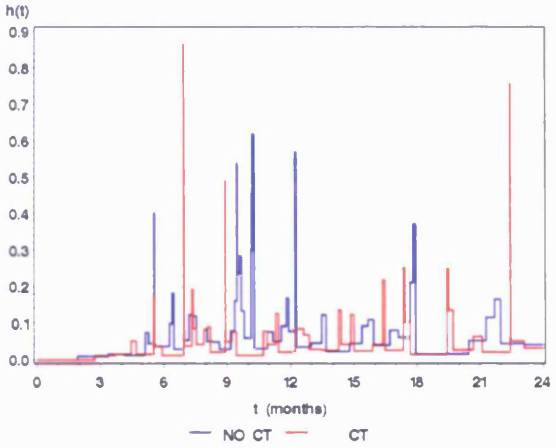
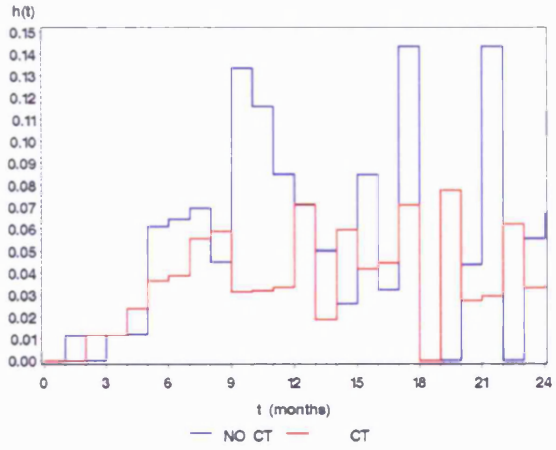


Figure 3.5: Estimates of the hazard functions for the ESPAC study within 24 months from time of surgery (see Figure 3.3 for relevant numbers at risk over time)

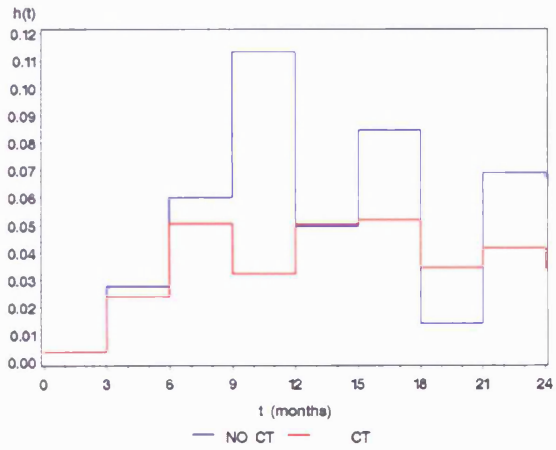
(a) Kaplan-Meier estimate



(b) Based on one-monthly hazard rates



(c) Based on three-monthly hazard rates



3.4.2 Hazard Ratio

The hazard function can be estimated for different groups of patients. At any time *t*, the hazard ratio for two groups of patients is the ratio of the hazard rates at this time and is a measure of the relative survival experience in the two groups. Many survival analysis techniques assume proportional hazards i.e. that this hazard ratio is constant over time. In the situation when the survival for two groups is being compared, the log-rank test is testing the null hypothesis that the ratio of the hazard rates in the two groups is equal to 1.

Using the total number of deaths and the total exposure to the risk of death over the whole follow-up period, the overall hazard rates in two groups can be estimated and these can be divided to give a crude estimate of the hazard ratio. More formally hazard ratios are generally estimated as follows:

$$HR = \frac{O_1 / E_1}{O_2 / E_2} \quad \text{with} \quad SE(\log HR) = \sqrt{\left(\frac{1}{E_1} + \frac{1}{E_2}\right)} \qquad [3.7]$$

Where O_1 and O_2 are the observed numbers of deaths in groups 1 and 2 and E_1 and E_2 are the expected number of deaths in each group of patients calculated using the log-rank method. An alternative Mantel-Haenszel version of the hazard ratio and associated standard error is also available which may be preferable when the number of events is small (Parmar and Machin 1995).

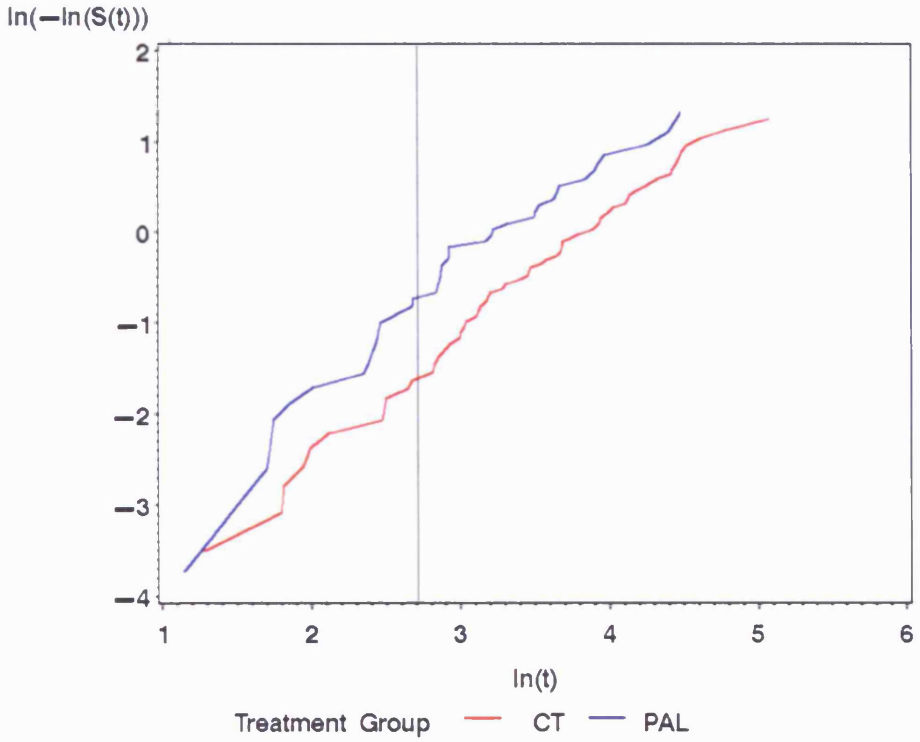
In the MIC study the crude estimates of the hazard rates over the whole study period were 0.0211 on CT and 0.0370 on PAL, giving an estimated hazard ratio of 0.57. This is very similar to that calculated using the log-rank method given in equation [3.7], which gives a hazard ratio of 0.56 (95% CI: 0.37 to 0.86). The crude estimates of the hazard rates over the first 15 weeks were 0.0129 for CT and 0.0298 for PAL giving an estimated hazard ratio of 0.43, comparable to that using the log-rank method of 0.42 (95% CI: 0.19 to 0.90). These are all more extreme than the hazard ratio in the whole trial of 0.79 (95% CI: 0.64 to 0.98).

In the ESPAC study the crude estimates of the hazard rates over the whole study period were 0.0331 on CT compared to 0.0488 on NoCT giving an estimated HR of 0.68. This is very similar to that calculated using the log-rank method of 0.66 (95% CI: 0.47 to 0.95).

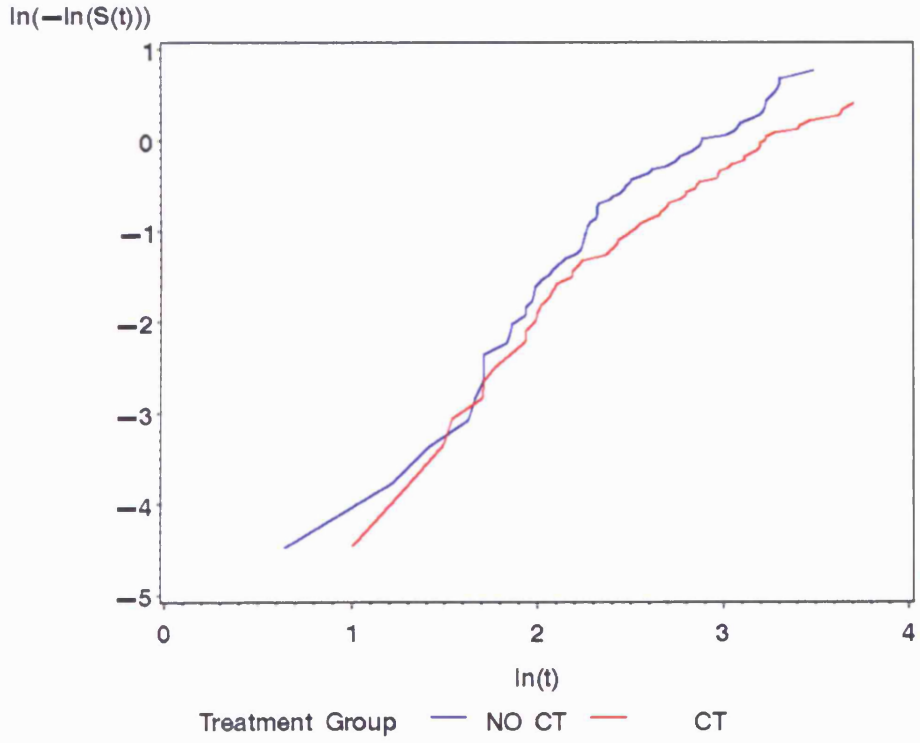
The validity of the assumption of proportional hazards can be assessed using a log cumulative hazard plot. This is a graph of log(-log[$S(t)$]) against log t where $S(t)$ is the survivor function estimated by Kaplan-Meier. If the hazards are proportional then the curves for the two groups should be approximately parallel. In the MIC study, the hazards were reasonable proportional but less so in the early phase (see Figure 3.6a) and similarly for the ESPAC data (see Figure 3.6b).

Figure 3.6: Log cumulative hazard plots for the MIC and ESPAC studies

(a) MIC study (with 15-week cut-off shown)



(b) ESPAC study



3.5 Parametric Approach to Estimation

In some situations it may be appropriate to assume a parametric distribution for $f(t)$, the probability density function of the survival time. The most common distributions used to model survival data are the exponential and Weibull distributions; the exponential being a special form of Weibull distribution. This thesis only considers these two distributions but in general, other distributions such as log-normal, log-logistic and gamma may be more appropriate (Collett 1994).

If survival times follow an exponential distribution then

$$f(t) = \lambda \exp(-\lambda t) \quad \text{for } \lambda > 0 \qquad [3.8]$$

and the survivor function is given by

$$S(t) = \exp(-\lambda t) \qquad [3.9]$$

and the hazard function is given by

$$h(t) = \lambda \qquad [3.10]$$

Thus, an exponential distribution assumes that the hazard rate is constant over time.

If a constant hazard rate is not a valid assumption then a Weibull distribution may be a more appropriate distribution for survival times. In this case

$$f(t) = \lambda \gamma t^{\gamma-1} \exp(-\lambda t^{\gamma}) \quad \text{for } \lambda, \gamma > 0 \qquad [3.11]$$

and the survivor function is given by

$$S(t) = \exp(-\lambda t^{\gamma}) \qquad [3.12]$$

and the hazard function is given by

$$h(t) = \lambda\gamma t^{\gamma-1} \qquad [3.13]$$

The parameters γ and λ determine the shape and scale of the hazard function and are thus called the *shape* and *scale* parameters respectively. In the special case of $\gamma=1$, the distribution of survival times is exponential and the hazard is constant. If $\gamma>1$, the hazard increases with time and if $0<\gamma<1$ then the hazard decreases monotonically.

The suitability of these parametric distributions can be assessed using a log-cumulative hazard plot as described in Section 3.4. If a Weibull distribution is appropriate then the line will be approximately linear, with the slope of the line giving an estimate of the shape parameter for the distribution. If an exponential distribution is appropriate then the slope of the line will be approximately equal to one. The plots for both the MIC and ESPAC studies shown in Figure 3.6 are approximately linear but with slopes greater than 1, indicating that a Weibull may be a more appropriate distribution for the data than an exponential. Examining the plots of the survivor and hazard functions may also give some insight into which parametric distribution may be appropriate for the data. The hazard functions plotted in Figures 3.4 and 3.5 do not appear to be constant over time, indicating that an exponential distribution may not be appropriate for these data.

The parameters for a distribution are estimated from the survival data using the method of maximum likelihood. In SAS, the LIFEREG procedure (SAS Institute Inc 1989) fits parametric models to survival data. Modelling survival data is considered further in Chapter 7 but here parametric distributions are estimated for each treatment group separately in the MIC and ESPAC studies to illustrate the distributions discussed. In the MIC study estimation relates to just the first 15 weeks from trial entry with all survival times greater than this censored at 15 weeks, whilst estimation for the ESPAC study relates to the whole follow-up time. For each study, the exponential and Weibull distributions that best fit the data in each treatment group are shown in Figures 3.7 and 3.8. In particular the Kaplan-Meier survivor function is overlayed with the parametric survivor functions. In both studies, the Weibull appears be more consistent with the data than the exponential. The parameter estimates together with their associated 95%

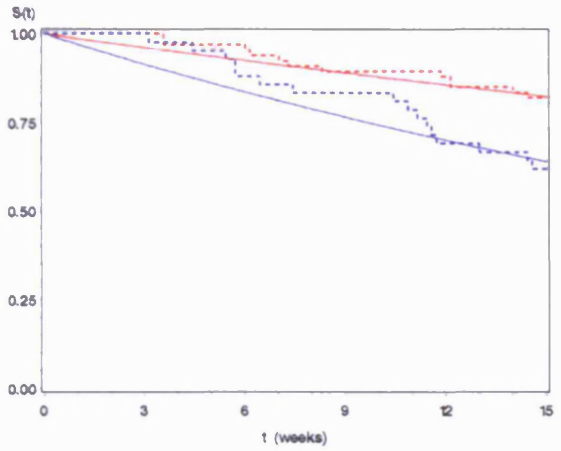
confidence intervals are shown in Table 3.1. The 95% confidence intervals for the Weibull shape parameters for the PAL group in MIC study and both treatment groups in the ESPAC study do not include the value 1 further indicating that the Weibull may be a more appropriate distribution for the data than the exponential.

Table 3.1: Estimates (and 95% confidence intervals) for parameters of exponential and Weibull distributions for MIC and ESPAC survival data by treatment

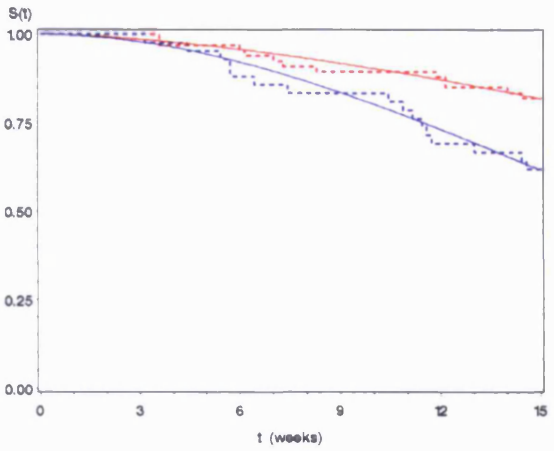
Study	Distribution	Parameter	CT	PAL / NoCT
MIC	**Exponential**	λ	0.0129 (0.0056, 0.0202)	0.0298 (0.0152, 0.0444)
	Weibull	λ (scale)	0.0021 (-0.0032, 0.0075)	0.0025 (-0.0035, 0.0085)
		γ (shape)	1.68 (0.75, 2.60)	1.95 (1.05, 2.84)
ESPAC	**Exponential**	λ	0.0331 (0.0247, 0.0416)	0.0488 (0.0371, 0.0605)
	Weibull	λ (scale)	0.0107 (0.0004, 0.0210)	0.0112 (0.0014, 0.0210)
		γ (shape)	1.35 (1.07, 1.62)	1.49 (1.22, 1.76)

Figure 3.7: Parametric survival curves within 15 weeks of trial entry for each treatment group in the MIC data (CT — ; PAL —)

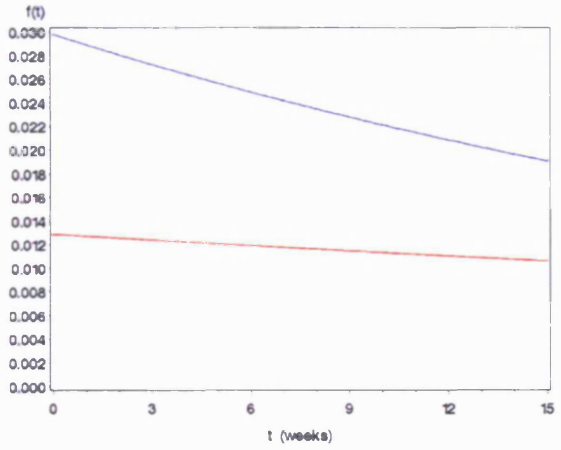
(a) Exponential survivor functions



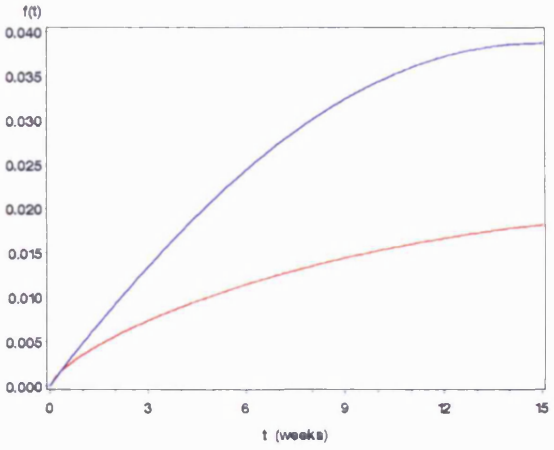
(b) Weibull survivor functions



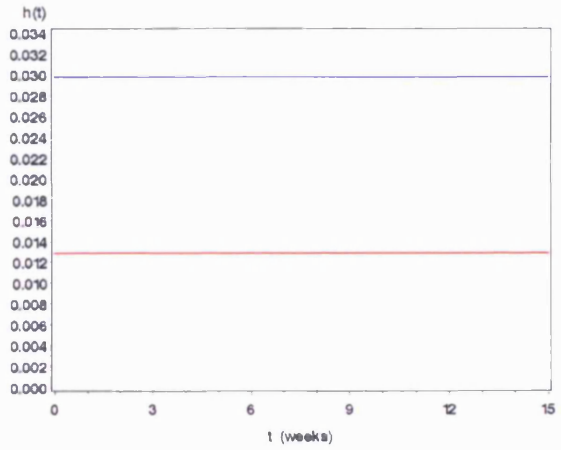
(c) Exponential density functions



(d) Weibull density functions



(e) Exponential hazard functions



(f) Weibull hazard functions

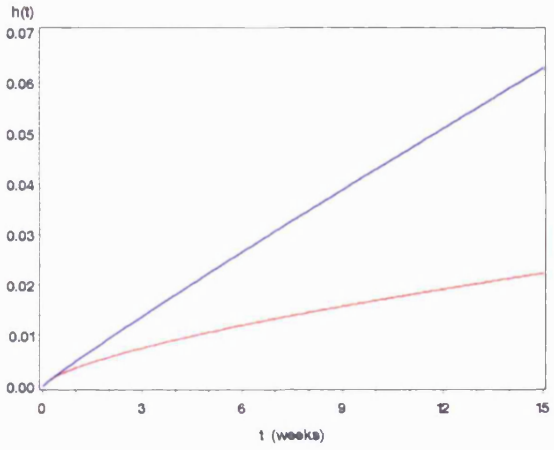
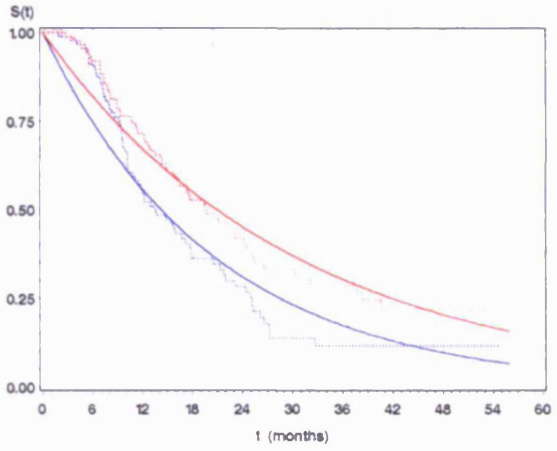
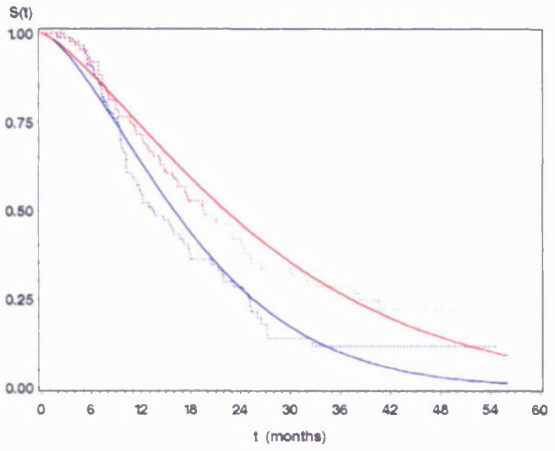


Figure 3.8: Parametric survival curves for each treatment group in the ESPAC study (CT —; NoCT —)

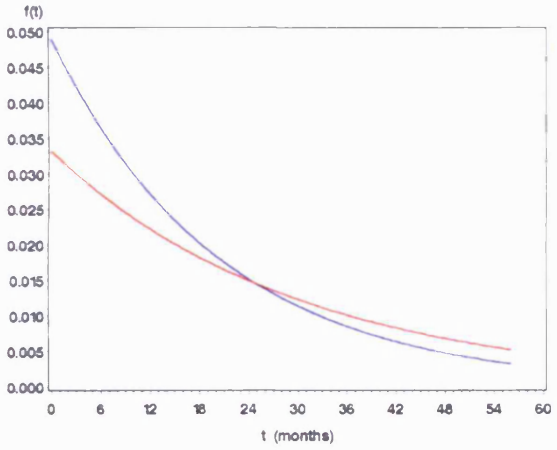
(a) Exponential survivor functions



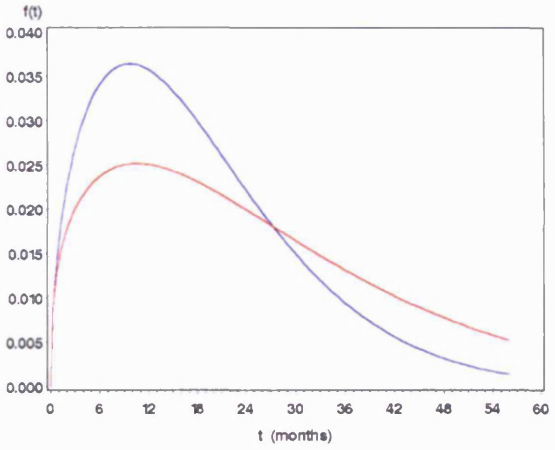
(b) Weibull survivor functions



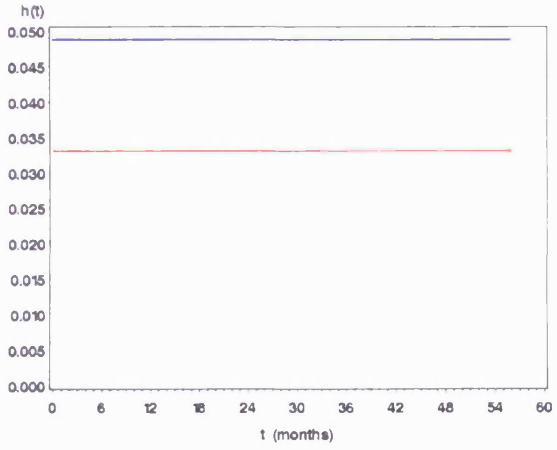
(c) Exponential density functions



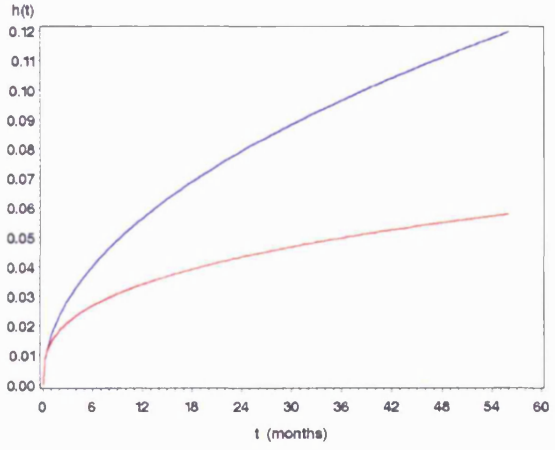
(d) Weibull density functions



(e) Exponential hazard functions



(f) Weibull hazard functions



3.6 Summary and Discussion

This chapter introduced the basic issues and methodology for analysing survival data. Examining the survivor functions gives some indication as to the differential dropout of patients from the quality of life study due to death. In addition, examining hazard functions and cumulative log hazard plots gives some insight into the appropriate models that can be used for the survival data in the simultaneous analysis with quality of life.

Survival analysis of the patients in the MIC quality of life study focused on the 15 weeks from trial entry, the time period during which the quality of life data are being collected. During this time, 28 (26%) patients died and therefore dropped out of the quality of life study, which could cause problems for any standard analysis of quality of life data. The problem is exacerbated by the fact that there appears to be a higher death rate and therefore greater dropout on the PAL arm compared to the CT arm. This indicates that simultaneous analysis of quality of life and survival data are required for an unbiased analysis of the data. It should also be noted from the survival analysis that all patients have full follow-up in terms of survival for the 15-week period from entry to trial and therefore there are no censored survival times within this time period. This will be beneficial in the subject-based approaches to quality-adjusted survival analysis discussed in Chapter 8. In terms of modelling, the hazard does not appear to be constant over the 15-week period, with a Weibull distribution appearing to adequately represent the data, and the assumption of proportional hazards could be considered to be valid.

In the ESPAC study, quality of life is assessed until death and the time frame for the longitudinal quality of life analysis is defined by the longest survival time, which is a censored survival time of 55 months. All deaths that occur prior to this time will cause problems for any standard analysis of the longitudinal quality of life data. Also, with deaths occurring early in the trial, it is not possible to restrict the analysis to a time interval during which no patients die. In addition, as with the MIC study, there is a differential rate of dropout due to death in the two treatment arms. All of these facts indicate that, as with the MIC study, simultaneous analysis of quality of life and survival data are required for an unbiased analysis of the quality of life data. Note that

in this study some patients have survival times censored relatively early and these will cause problems in the subject-based approaches to quality-adjusted survival analysis discussed in Chapter 8. As with the MIC study, in terms of modelling, the hazard does not appear to be constant with a Weibull distribution appearing to adequately represent the data, and the assumption of proportional hazards could be considered to be valid.

The analysis described in this chapter is extended in Chapter 7 to consider models for the survival data. Models that make no assumptions about the distribution of the data are considered in addition to those that assume parametric distributions such as exponential, Weibull and piecewise exponential. Model parameters are estimated using both a classical and Bayesian approach.

CHAPTER 4: ASSESSMENT OF QUALITY OF LIFE

4.1 Introduction

The aim of this chapter is to summarise the issues relating to the assessment of quality of life in a longitudinal study that are relevant to the analysis presented in later chapters. In particular, the type of measures used to reflect quality of life, the multivariate nature of the data and the timings of the assessments, are all issues that will impact on the appropriate analysis of the data.

Quality of life in its most general context is a concept incorporating all factors that might impact on an individual's life. In health service research it is more usual to consider health-related quality of life, which includes only those factors that affect an individual's health. There is no general agreement regarding the identification of such factors (Aaronson 1989, Schumacher et al 1991).

The World Health Organisation defines health as 'a state of complete physical, mental and social well-being, and not merely the absence of disease and infirmity' (World Health Organisation 1947). Quality of life is often referred to in these terms. For example, the dimensions comprising quality of life have been specified as follows (Schumacher et al 1991):

- symptoms of disease and side-effects of treatment (e.g. nausea, pain, anorexia)
- physical and functional status (e.g. mobility, self-care, fatigue)
- emotional status (e.g. anxiety, depression, satisfaction with care)
- social functioning (e.g. family interaction, work/recreation, time with friends)

In this thesis a pragmatic view is taken and quality of life is accepted as any measure that purports to reflect health-related quality of life or some aspect of it.

There are many issues to consider when attempting to measure quality of life, such as what questions should be asked, how should responses be recorded, when should questions be asked, of whom should questions be asked and who should do the asking. These and other aspects of measuring quality of life data in clinical trials have been

extensively reviewed and discussed elsewhere (Fayers and Jones 1983, Aaronson 1989, Fitzpatrick et al 1992, Osoba 1998, Fayers and Machin 2000).

In this chapter, the different methods for measuring quality of life in patients and options for dealing with the multivariate data that are produced are described in Sections 4.2 and 4.3, with specific details of the measures in the MIC and ESPAC studies detailed in Section 4.4. The issues relating to the timing of quality of life assessments, with specific reference to the illustrative studies, are described in Section 4.5. Although some studies may only be interested in the assessment of quality of life at a single time point, in this thesis the interest is on quality of life data that are collected on a number of points over time. Section 4.6 summarises the issues and the relevance to later chapters in the thesis.

4.2 Quality of Life Measures

4.2.1 Instruments for Measuring Quality of Life

The quality of life of a patient is usually measured using an instrument in the form of a questionnaire designed for patient completion. The questionnaire generally comprises sets of questions or *items* relating to the various *dimensions* of quality of life, such as the physical, psychological or social. Responses may be 'yes/no', a series of ordered categories or on a linear analogue scale.

Many instruments are used in the assessment of quality of life (Bowling 1991, Bowling 1996, Fallowfield 1990, Campbell and Gibbard 1998), such as the Rotterdam Symptom Checklist, Nottingham Health Profile, Sickness Impact Profile, Hospital Anxiety and Depression Scale and SF36. Generic quality of life instruments measure general aspects of quality of life and are applicable in a wide range of research settings, whilst non-generic instruments are relevant for a specific disease or treatment. Some questionnaires are dimension-specific in that they only ask questions relating to a particular aspect of quality of life e.g. Hospital Anxiety and Depression Scale. Some questionnaires allow the individual to choose and rate the specific aspects of quality of life that are important

to them. The self-evaluation inventory for quality of life (SEIQOL) is an example of such an instrument (Waldron et al 1999).

Much work has been done to test the adequacy of such instruments. The adequacy is determined by a number of different criteria (Fitzpatrick et al 1998) including for example validity and reliability. An instrument is valid if it is actually measuring what it is designed to measure and it is reliable if, all things being equal, it measures consistently from one occasion to the next. There has been much discussion regarding the requirements of quality of life measures and methods for assessing such requirements, including the use of Cronbach alpha coefficients and factor analysis (Hays et al 1993, Fitzpatrick et al 1998, Hays et al 1998, Fayers and Machin 1998a, Fayers and Machin 2000).

The questionnaires described so far yield *descriptive measures* of quality of life, also called *profile* or *psychometric measures*. Some questionnaires yield a special type of quality of life measure called a *utility*. The responses to the questionnaire enable the patient to be allocated to a health state for which a utility value has been pre-determined (see Section 4.2.3). The relationship between descriptive and valuational approaches has been explored (Bosch et al 1996, Revicki and Kaplan 1993, Kaplan 1998) and measures from each approach are discussed in more detail in Sections 4.2.2, 4.2.3 and 4.2.4 below.

4.2.2 Descriptive Measures

Descriptive measures of quality of life may take various forms (binary, ordinal, discrete or continuous) for which a variety of distributional assumptions are appropriate. Responses to each item on a questionnaire may yield binary data from a yes/no response, ordinal data from a categorical scale, discrete data such as how many times a patient has vomited or how many pills for pain relief a patient has taken or continuous data such as from a linear analogue scale. The aggregation of items, to give a global measurement of a dimension or of overall quality of life, results in data that is usually treated as continuous, despite the fact that the global measure may take only values from a restricted range. The distribution of quality of life variables as measured from a linear analogue scale or as an aggregated global score may or may not be normally

distributed. Sometimes data may be transformed to create normally distributed data for analysis, but on other occasions (e.g. when there is a dominance of a particular response value, perhaps corresponding to 'no problem') this may be difficult.

4.2.3 Utility Measures

The health states experienced by an individual during the course of their disease and treatment are each associated with a quality of life. Each can be assigned a value that measures the preference of the individual for a health state relative to other states. The value reflects the quality of life in that health state. The value generally lies between 0 representing death and 1 representing perfect health, although it can be negative, representing health states judged to be worse than death. Technically, these values are utilities when they are measured under conditions such that an individual is making decisions with uncertain outcomes (von Neumann and Morgenstern 1953), but they may approximate to utilities when measured under other conditions (Torrance 1986).

In a quality of life study, questionnaires determine which health state the patient is in and the corresponding pre-determined utility value can be attached. The utility values associated with health states are measured by questioning a sample of subjects and eliciting their relative preferences for health states. The subjects could be patients, health professionals or the general public. Methods such as *time trade-off* and *standard gamble* (Torrance 1986, 1987) are commonly used methods for evaluating utilities of health states. There are a number of utility-based questionnaires used in quality of life studies but one of the most widely used examples is the EQ-5D (Rabin and de Charro 2001). The EQ-5D consists of 5 questions each with three possible responses and the different combinations of the 5 responses define 243 different health states, each of which has a utility value determined from a previously conducted large-scale valuation study.

The advantage of a utility-based questionnaire is that it generally yields a single value as a measure of quality of life. The fact that the measure is bounded at each end by a value with a clear clinical meaning facilitates the interpretation of the score. In addition the utility has the advantage of being the measure needed in quality-adjusted survival

analysis (see Chapter 8). The utility measure is also necessary when costs of treatment are balanced against quality of life and survival in a cost-utility analysis.

4.2.4 Translating Descriptive Measures Into Utility-Type Measures

Many quality of life studies use instruments that yield descriptive measures rather than utility measures. Some of the methods for simultaneous analysis require utility measures. Ideally, the proposed method of analysis should be determined at the design stage of a study and appropriate measures chosen accordingly. Otherwise it may be desirable to estimate utility values from the descriptive measures that are recorded in the study. For some instruments, it may be possible to convert the descriptive measure into a proper utility measure (Brazier et al 2002, Chancellor et al 1997). If no validated conversion is available, the descriptive measure can be translated onto a 0 to 1 scale to create a *pseudo-utility* measure and may be treated like a utility. It should be noted however that these translated scores are not proper valuations of quality of life, they merely reflect the level of quality of life of the patient on a 0 to 1 scale. Methods of analysis that are based on utility values, such as quality-adjusted survival analysis described in Chapter 8, can be performed with pseudo-utilities but the results require cautious interpretation.

4.3 Handling the Multivariate Nature of the Quality of Life Endpoint

The quality of life data measured in a longitudinal study is generally multivariate in nature. At one extreme, quality of life can be measured by a single global measurement such as the Karnofsky Index (Karnofsky and Burchenal 1949), whilst at the other extreme, assessment is made via a questionnaire containing a multitude of items measuring a variety of conceptual dimensions. The Sickness Impact Profile for example measures 12 dimensions of quality of life via a questionnaire containing 136 items (Bergner et al 1981). The fact that quality of life can be considered in terms of individual items or in terms of separate dimensions means that as an endpoint, it is potentially multivariate in nature. In some situations it may be desirable to consider each item or dimension as a separate quality of life endpoint. In a descriptive analysis this will only cause problems in terms of presentation and overall decisions regarding

the optimal treatment may be difficult if the different items or dimensions give conflicting conclusions. If hypothesis testing is involved, analysis of multiple endpoints will lead to the problem of multiple testing, where the probability of a false positive finding increases as the number of tests carried out increases.

In some studies it may be possible to limit the amount of hypothesis testing by specifying in advance a few key quality of life measures on which hypotheses will be tested, leaving the remaining variables to be analysed purely descriptively (Cox et al 1992, Pocock 1991, Nayfield et al 1992, Fletcher et al 1992). If this approach is not practical or desirable then the analysis will need to account for the multivariate nature of the quality of life endpoint.

There are a variety of ways of handling multiple endpoints in clinical trials (Zhang et al 1997). One approach is to combine multiple endpoints to create global scores before analysis. Another approach is the post-analysis combination of results from the univariate analysis of each separate endpoint. The application of such approaches to quality of life data has been discussed (Schumacher et al 1991, Pocock 1991) and some methods have been applied and compared in the analysis of quality of life data (Tandon 1990). An alternative to these methods is to use a hierarchical approach to the analysis and this has been considered for quality of life data (Beacon 1996, Beacon and Thompson 1996). The essentials of each of these approaches are considered below. In this thesis multiple endpoints are handled using global scores and by restricting analysis to a single global measure of quality of life. The extension of methods to accommodate the multivariate nature of quality of life data is discussed throughout the thesis.

4.3.1 Combining Multiple Endpoints to Create Global Scores

For each individual, the values of the items that make up a quality of life endpoint can be combined in some way to form a global score. In some cases the items within each quality of life dimension may be combined to create dimension-specific global scores, whilst in other cases an overall quality of life global score may be created either by combining all items on a questionnaire or by combining dimension-specific global scores. Use of a single global quality of life score simplifies statistical analysis and should be aimed for when sensible and justifiable (Olschewski and Schumacher 1990).

If treatments are likely to affect dimensions differently then combining dimensions into a global score may not be sensible (Fletcher et al 1992).

Global scores can be calculated using either an unweighted or a weighted sum. It is suggested that unweighted sums should only be used to combine items that are highly positively correlated (Olschewski and Schumacher 1990). On the other hand, a weighted sum may make interpretation difficult and the weights used may be controversial (Fletcher et al 1992). Weights can be determined either from the data using scores from a factor analysis, from decision theory using utility analysis techniques or arbitrarily (Cox et al 1992). Aggregation using data-oriented procedures has been recommended (Olschewski and Schumacher 1990). Most validated questionnaires give specific instructions on how to combine items to form global scores (for example Aaronson et al 1993).

A method for calculating global scores proposed by O'Brien (O'Brien 1984) has been applied to quality of life data (Tandon 1990, Tandon et al 1989). It is a non-parametric approach and creates global scores from the ranks of the item values rather than the actual data values. The data for all treatment groups are pooled and, for each variable in the multivariate quality of life endpoint, the values across all individuals are ranked. A global score is created for each individual by summing the ranks for all variables.

Calculating global scores can be problematic if data are missing on some of the items within a score (Fayers et al 1998). Patients will fail to respond to single items on an otherwise complete questionnaire for a number of different reasons; in some cases the question will have been omitted unintentionally, in other cases it might be that the question was not applicable and in other situations the patient may perceive the question to be too intrusive. If such missing data can be assumed to be missing at random then they will not be a major problem. This assumption would be untenable in situations where consistent non-response to an item suggests that the question is inappropriate or difficult to answer and hence is not missing at random.

A possible solution to this problem is to impute the missing values. This is only feasible if they are limited in number. In a longitudinal study, if a subject has a missing value for an item then the value could reasonably be imputed from:

- values of other items within the dimension for that subject;
- values of other items in the subject's questionnaire;
- values of the item on the subject's questionnaires at time points on either side of the missing value;
- baseline patient characteristics and survival;
- clinical measures that may relate to quality of life taken over time

In general the formula for calculating global scores can incorporate missing data by adjusting for the number of items involved in the calculation. Using a mean rather than a sum allows accommodation of missing values into the global score since the mean can be calculated for a reduced number of items. Alternatively, expressing the sum as the percentage of the maximum achievable score (Fletcher et al 1992) allows for the possibility of a reduced number of items. This approach is indirectly imputing the missing value using the other items within the dimension. If too many of the items within the global score are missing or if the assumption of missing at random is untenable, then the global score should be recorded as missing if any item within it is missing.

4.3.2 Combining Results from Univariate Tests on Multiple Endpoints

The simplest method for testing a global null hypothesis of no treatment effect, using the results from multiple univariate tests, is to use a Bonferroni-type adjustment. The p-values from the multiple univariate tests are adjusted by multiplying each p-value by the number of tests carried out. Each endpoint can be assessed using these adjusted p-values or a global null hypothesis can be assessed using the minimum p-value. This method has been recommended for use with quality of life data (Cox et al 1992) and has been applied and compared to other methods in a quality of life setting (Tandon 1990). The main drawback with the global null hypothesis approach is that it confines attention to the smallest p-value and may be too conservative.

A parametric method for combining results from multiple univariate t-tests, originally proposed by O'Brien (O'Brien 1984) but developed by others (Pocock et al 1987), has

been applied to quality of life data (Tandon 1990; Tandon et al 1989). The following test statistic can be used to assess a global null hypothesis of no treatment effect:

$$J^T S^{-1} t / (J^T S^{-1} J)^{½} \qquad [4.1]$$

where J is a vector of 1's, S is an estimated correlation matrix for the multiple measures and t is a vector of t-statistics from the separate univariate t-tests. The test statistic given in [4.1] has an asymptotic standard normal distribution.

The main drawback of these methods is that they do not give an estimate of the treatment effect, they just provide a test statistic. Such methods will not be pursued further in this thesis.

4.3.3 Hierarchical Approach

Multilevel models have been advocated for the analysis of data that have a hierarchical structure (Goldstein 1995). Longitudinal data can be thought of as hierarchical data with level one of the hierarchy being observations over time within a subject and level two being the subject. Multilevel models have been used to analyse longitudinal quality of life data (Beacon 1996, Beacon and Thompson 1996) and their application is discussed in more detail in Chapter 6. The hierarchical approach provides a means of handling the multivariate nature of the quality of life endpoint (Beacon 1996, Beacon and Thompson 1996). The multiple dimensions that constitute quality of life can simultaneously be analysed in a multilevel model by adding an extra level to the standard longitudinal data model, with the lowest level representing the various quality of life dimensions.

Multilevel models have the advantage over methods discussed previously that they provide estimates of the treatment effect as well as test statistics. Treatment effects are estimated for each dimension separately and, if appropriate, an overall summary estimate may be obtained. The model also allows the correlation between the measures over time for different dimensions to be estimated. Multilevel models are flexible in that they can cope with situations where some of the dimension scores may be missing for some patients. The application of multilevel models to quality of life data in general however is problematic since the method assumes the missing data mechanism is

ignorable, which may not be true for quality of life data. Extending such hierarchical models to include a model for survival data in a joint model as described in Chapter 10 overcomes some of the problems of non-ignorable missing data.

4.4 Quality of Life Measures in the MIC and ESPAC Studies

4.4.1 MIC Study

Quality of life was assessed in a subgroup of 109 of the 351 patients in the trial. Quality of life was assessed using questionnaires completed by the patients with help when needed from a dedicated quality of life research nurse. The questionnaire was designed specifically for the trial but was based on the lung cancer module of the quality-of-life questionnaire designed by the European Organisation for Research and Treatment of Cancer (EORTC), called the EORTC QLQ-LC13 (Aaronson et al 1993).

The questionnaire consisted of 12 questions evaluating various physical and psychological aspects of a patient's quality of life, including a general question on malaise (see Table 1). Responses were given on a four-level ordered categorical scale, 'none', 'a little', 'quite a bit' and 'very much', and for analytical purposes were coded from 0 to 3 respectively.

Table 4.1: Questions from the MIC quality of life questionnaire

ITEM	QUESTION
Cough	Do you have a cough?
Severe Dyspnoea	Do you get breathless on mild activity like dressing?
Moderate Dyspnoea	Do you get breathless when walking on the flat?
Mild Dyspnoea	Do you get breathless on stairs or walking uphill?
Haemoptysis	Have you coughed blood?
Pain	How much pain are you getting?
Appetite	Have you noticed any loss of appetite?
Anxiety	Have you been worrying?
Depression	Have you been depressed?
Dysphagia	Have you any difficulty swallowing?
Nausea	Did you feel sick during or since your last treatment? (CT) Have you been feeling sick? (PAL)
Malaise	Have you been feeling generally ill?

A single measure that summarises all responses from the questionnaire was created by taking the mean of the responses to the 12 individual questions (*MQS*). In the case of questionnaires with missing responses, *MQS* was calculated as the mean of those questions with a response. Of the 392 questionnaires, 349 (89%) were completed fully, 33 (8%) had one response missing, 8 (2%) had two missing responses and the remaining 2 questionnaires had four and five missing.

MQS ranges from the best possible score of 0 (i.e. all symptoms rated as 'none') to the worst possible score of 3 (i.e. all symptoms rated as 'very much'). In order to create a utility-type measure, this mean score was transformed to give a measure on a 0 to 1 scale. The transformed score is calculated such that at one extreme, an *MQS* score of 0 is translated to a value of 1 representing 'best possible health' and at the other extreme an *MQS* value of 3 is translated to a value 0.25, allowing all time after death to be scored as 0. This choice of transformation was subjective and results should be checked for their sensitivity to this definition. The transformed score, henceforth referred to as the global quality of life score (*GQS*), was calculated as follows:

$$GQS = 1 - \frac{MQS}{4} \qquad [4.2]$$

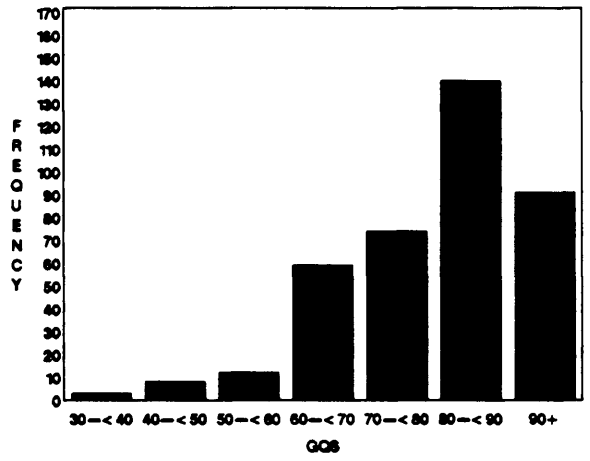
The method for creating a global quality of life score using [4.2] follows the general principles specified by the EORTC for their quality of life questionnaire (Aaronson et al 1993).

The full analysis of the data presented for a clinical audience includes a descriptive analysis of all individual questions in addition to testing hypotheses about *GQS* and the malaise question (*MAL*) (Billingham et al 1997). In this thesis, the analyses focus just on *GQS* and *MAL* as single global measures of quality of life. Despite the fact that *GQS* is made up of ordered categorical responses, the large number of questions contributing to the score results in a variable that is continuous in nature with a slightly skewed distribution over all questionnaires (see Figure 4.1a). The distribution of the ordinal variable *MAL* over all questionnaires also has a skewed distribution (See Figure 4.1b). There is an association between the values of *GQS* and *MAL* with mean *GQS* reducing

for increasing values of *MAL* (mean *GQS* is 89, 79, 66, 53 for categories 0, 1, 2, 3 respectively).

Figure 4.1 Distribution of global measures of quality of life in the MIC study across all assessments

(a) *GQS*



(b) *MAL*

4.4.2 ESPAC Study

The quality of life instrument used in the ESPAC study was the core module of the cancer-specific questionnaire designed and validated by the European Organisation for Research into Treatment for Cancer called the EORTC QLQ-C30 (Aaronson et al 1993). This contains 30 different questions measuring 15 different dimensions of quality of life. An additional questionnaire designed, specifically for pancreatic cancer, was also used. The main analysis of the study analysed all quality of life dimensions assessed by these questionnaires (Neoptolemos et al 2001). In this thesis, however, a single measure of quality of life known as the 'global health status score' is selected as the variable for analysis and, although simplistic, it nonetheless illustrates the essential elements of the various methods for simultaneous analysis and allows insight into the associated problems. The global health status score (*GHSS*) is constructed from 2 questions each with a 7-level ordinal response (see Figure 4.2). If one of the questions is missing the score is calculated from the single non-missing response (Aaronson et al 1993). Of the 710 questionnaires returned in the ESPAC study, 7 had a missing response for one of the two questions.

Figure 4.2: Global health status score (*GHSS*) from the EORTC QLQ-C30

	Very poor						Excellent
*Q*29: How would you rate your overall physical condition during the past week?	1	2	3	4	5	6	7
*Q*30: How would you rate your overall quality of life during the past week?	1	2	3	4	5	6	7

$$GHSS = \frac{Mean - 1}{6} \times 100$$

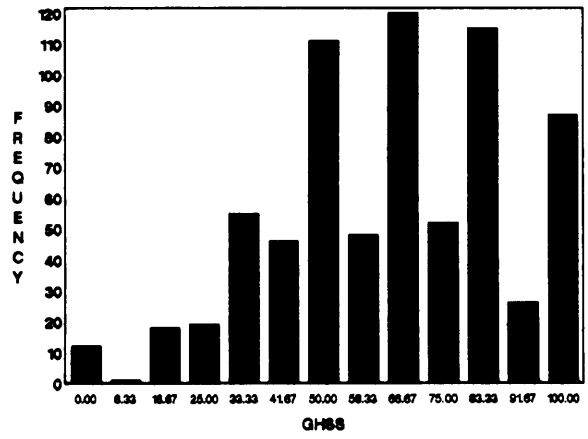
where $Mean = (Q29 + Q30)/2$

0 ⟶ 100

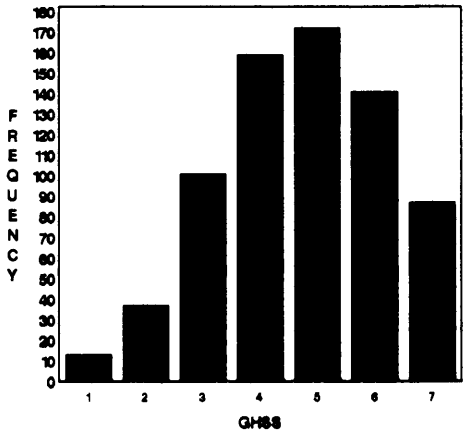
Values for *GHSS* range from 0 to 100 and take 13 equally-spaced possible values in this range. The distribution of the values in the ESPAC study (see Figure 4.3a) is affected by the fact that patients are more likely to give the same response to both questions (i.e. scores of 0, 16.67, 33.33, 50, 66.67, 83.33, 100) rather than different responses. By combining scores into a 7-level ordinal variable, the distribution of GHSS becomes more normal (see Figure 4.3b).

Figure 4.3 Distribution of *GHSS* in the ESPAC study

(a) In terms of actual values



(b) As a 7-level ordinal variable



4.5 Timing of Quality of Life Assessments

Quality of life data is generally longitudinal in nature. Some studies assess quality of life at one time point only or take a baseline measure and a follow-up measure, but generally quality of life is recorded at more than two time points during the course of a study. There can be any number of time points and these will not necessarily be evenly spaced or consistent across individuals.

4.5.1 Planned Versus Actual Timing

In most studies, the timing of the quality of life assessments will be planned for specific time points; sometimes assessments are planned for fixed times from date of entry to the trial e.g. 3 months, 6 months, 9 months etc, implying a pre-defined fixed time between assessments but often they are planned to relate to the timing of treatment and follow-up visits. In the first case, for administrative reasons or maybe due to delays by the patient in responding to the questionnaire, the actual timings may vary considerably from what was planned. In the latter situation, although theoretically patients should all receive treatment and be followed-up at approximately the same time points, the reality is that patients may differ greatly in these timings and consequently in the timings of the quality of life assessments. Delays in chemotherapy mean that timing in relation to date of entry to trial may be variable, but in relation to timing of treatment then the assessments may conform exactly to what was planned e.g. day 1 of each cycle.

One option is to allocate assessments to specific planned time points according to whether the actual timing falls within the pre-defined 'window' for that time point. This process is not straightforward. For example, more than one assessment may fall within a certain window and since only one assessment is required per time point, a subjective choice needs to be made as to which one to discard and which one to keep, or alternatively an average value could be used. The choice of window needs to be wide enough to minimise exclusions but narrow enough to give a true reflection of quality of life at that time point.

Examining plots that show actual timing of each assessment for each patient in terms of time since entry to trial together with the planned times, will show how much variability there is in the actual timings around the planned ones. Examples are given for the MIC and ESPAC studies (see Figures 4.4 and 4.6 respectively) and for other studies (Fairclough 2002). This type of plot can also be produced to show timings in relation to timing of treatment received (see Figure 4.5).

4.5.2 Time Frame for the Quality of Life Study

The assessment of treatments in terms of quality of life is usually based on their effect on this outcome over time. Some studies, may be interested in the effects during a fixed time frame of interest, such as the treatment period or the first year from entry to trial for example. This can be defined at the design stage of the study such that quality of life data is only collected during this time or can be defined at the analysis stage in order to minimise the amount of missing data in the analysis. In other studies the focus may be on the quality of life during a patient-related time frame, such as from entry to trial until disease progression or death. In this case, the length of study will vary for each patient but the overall time frame of the longitudinal analysis will be defined by the longest individual study time. For both the fixed and overall patient-related time frame, even though patients may have a complete set of quality of life measures until death, a standard longitudinal analysis of the quality of life outcome will treat the non-existent data at all times within the time frame after death as missing and will assume that they are missing at random. Clearly because they are missing due to death this assumption is invalid and any longitudinal analysis should account for this. Patients who die within the fixed or overall patient-related time frame are said to have 'dropped out' of the quality of life study and because the resulting missing data are not missing at random, the dropout is said to be 'informative'.

For a fixed time frame of interest, variability in the actual timings of questionnaires may result in some occurring after the fixed time frame of the study. Although any analysis will generally focus purely on making inferences about the quality of life within the fixed time frame, it is possible to use data from beyond the end of the study period to make inferences regarding quality of life within the study period. This raises the issue of how to deal appropriately with these post-study assessments, and specifically whether

they should contribute to any analysis of the data. By including these assessments, potential biases could be introduced. The responses post-study are conditional on patients surviving long enough to give additional 'late' information and even for the longer survivors, there is a possibility that the patients who answered later questionnaires were different in some way to those that did not and hence the inclusion of these data could bias the results. It may be that the values recorded at post-study time points are reasonably consistent with the values recorded during the study period but in some cases the values may be highly influential on the resulting models. The option therefore that causes least bias is to ignore all assessments that occurred after the fixed time frame of the study. However, if treatment delays and administrative errors cause some assessments to fall just a short way outside the time frame then one may still wish to include them in the analysis, especially if they are the only measure of say post-treatment quality of life. It will generally be desirable to exclude those that fall a considerable time after the end of the study period. The MIC study had assessments that fell outside the fixed time frame of the study and these are discussed in Section 4.5.3.

4.5.3 Timing of Assessments in the MIC Study

The MIC study was designed to assess quality of life during the treatment phase of the trial. For patients on the CT arm, assessments were planned to be taken on the first day of each cycle of chemotherapy. In addition one post-treatment questionnaire was to be taken. The first cycle of chemotherapy was expected to occur soon after randomisation and cycles were 3 weekly, thus the expected timings of the questionnaires were 0, 3, 6, 9 and 12 weeks after entry to trial. Assessments on the palliative arm were planned to be 0, 3, 6 and 9 weeks from entry to trial, in expectation that these times would match those for patients on the chemotherapy arm. For some reason five questionnaires were planned on the chemotherapy arm and only four on the palliative care arm.

The quality of life study period was defined as 15 weeks from entry to trial. The choice of 15 weeks from entry to trial as a cut-off to the quality of life study is subjective. It was necessary to choose a time that was early enough to minimise the length of time from the scheduled last assessment (12 weeks for CT patients and 9 weeks for PAL patients) to the end of the study period but needed to be late enough to minimise the number of assessments that fell outside the study period. In total 6 assessments out of

392 fell outside the study period. Delays to chemotherapy resulted in 6 patients having their first post-treatment assessment a short time after the 15-week cut-off (actual times ranged from 15.1 to 17.3 weeks from date of entry). It is not desirable to ignore this information, especially when they occurred so close to the 15-week cut-off and therefore these assessments were included in the subsequent analysis.

The actual timings of the questionnaires varied considerably around the planned times. Summary measures for the distribution of actual timings in relation to trial entry, previous assessment and chemotherapy cycles are given in Table 4.2. The actual timings of the questionnaires in relation to time of randomisation are plotted for each patient in each treatment group in Figure 4.4. In addition the timings in relation to each cycle of chemotherapy are plotted for the CT arm in Figure 4.5. On average the patients conformed reasonably well to the scheduled time points (see Table 4.2) especially in terms of timing in relation to previous assessment time and timing in relation to treatment time but there was some variability. The pattern in Figure 4.4 shows that the rationale for timing in the early part of the study was to time questionnaires in relation to trial entry whilst Figure 4.5 shows that the latter patients were scheduled according to treatment times and hence their questionnaires were more variable in terms of the time from trial entry. Notice the 6 patients in Figure 4.4 with questionnaires outside the 15-week study period.

Table 4.2: Timing of assessments (A1 to A5) in the MIC study in weeks in relation to trial entry, previous assessment and treatment - median (minimum, maximum)

(a) CT arm (N=67)

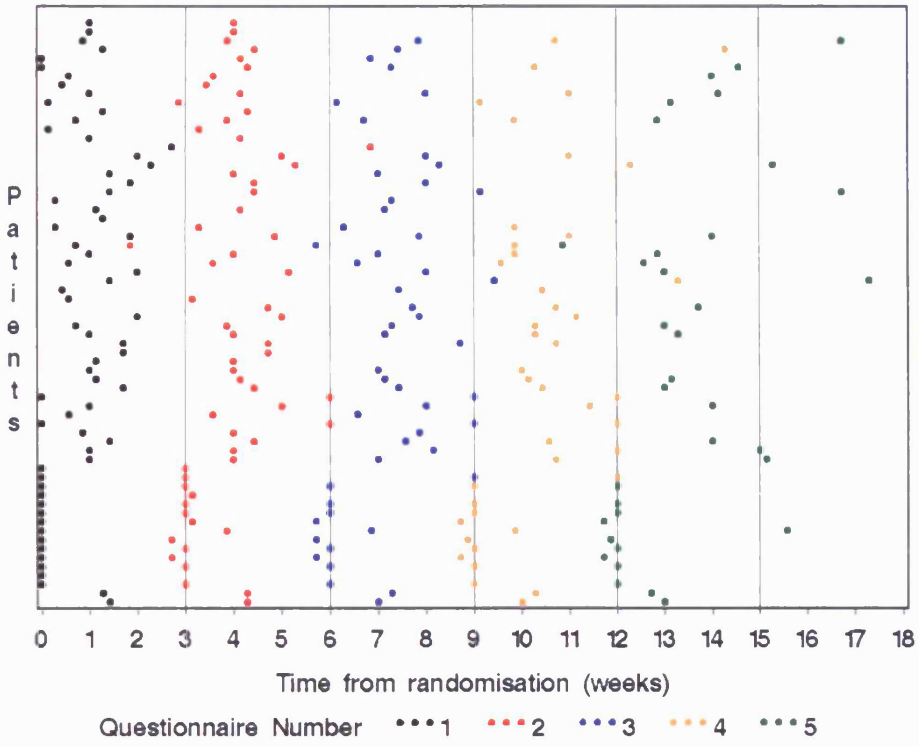
	A1	A2	A3	A4	A5
Scheduled time in relation to trial entry	0	3	6	9	12
Actual time in relation to trial entry	0.9 (0, 2.7)	4.0 (1.9, 6.9)	7.2 (5.7, 9.4)	10.3 (8.7, 14.3)	13.0 (10.9, 17.3)
Actual time in relation to previous assessment	-	3.0 (1.1, 6.0)	3.0 (2.6, 6.0)	3.0 (2.0, 6.9)	3.0 (1.0, 6.0)
Scheduled time in relation to treatment	Prior to cycle 1	Prior to cycle 2	Prior to cycle 3	Prior to cycle 4	Post cycle 4
Actual time in relation to day 1 of that cycle	0 (-4.1, 0.4)	0 (-2.3, 0.7)	0 (-2.1, 0.4)	-0.1 (-2.3, 0.4)	3 (0.7, 6.0)

(b) PAL arm (N=42)

	A1	A2	A3	A4
Scheduled time in relation to trial entry	0	3	6	9
Actual time in relation to trial entry	0.0 (0, 2.9)	3.6 (3.0, 5.6)	7.1 (5.4, 9.0)	9.9 (8.7, 13.3)
Actual time in relation to previous assessment	-	3.0 (2.0, 4.0)	3.0 (2.4, 4.4)	3.0 (2.0, 5.1)

Figure 4.4: Timing of each assessment in the MIC study in terms of entry to trial

(a) CT arm (N=67)



(b) PAL arm (N=42)

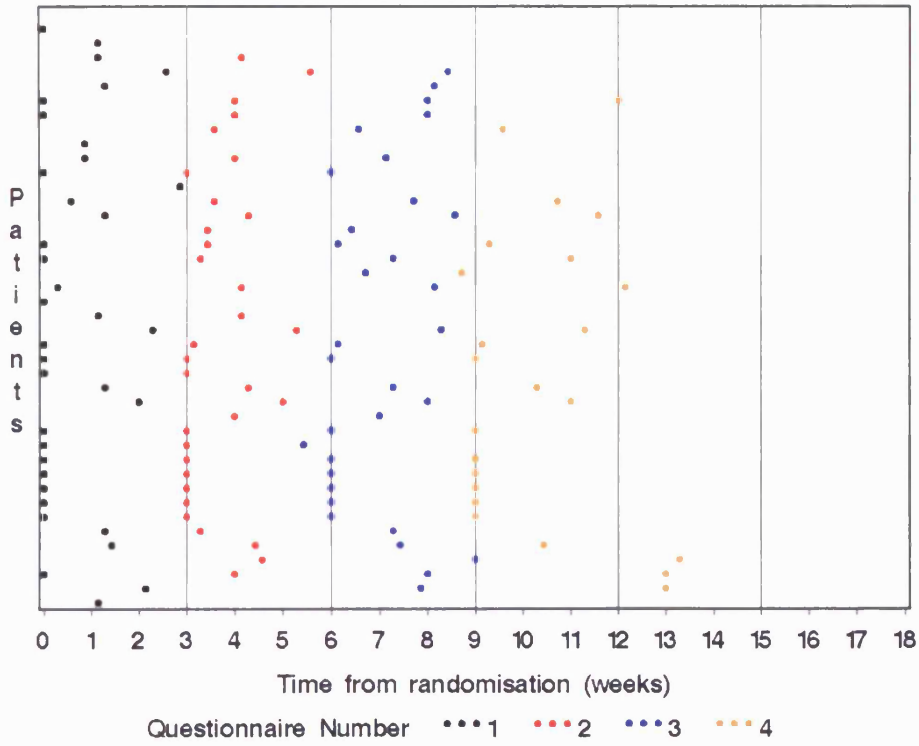
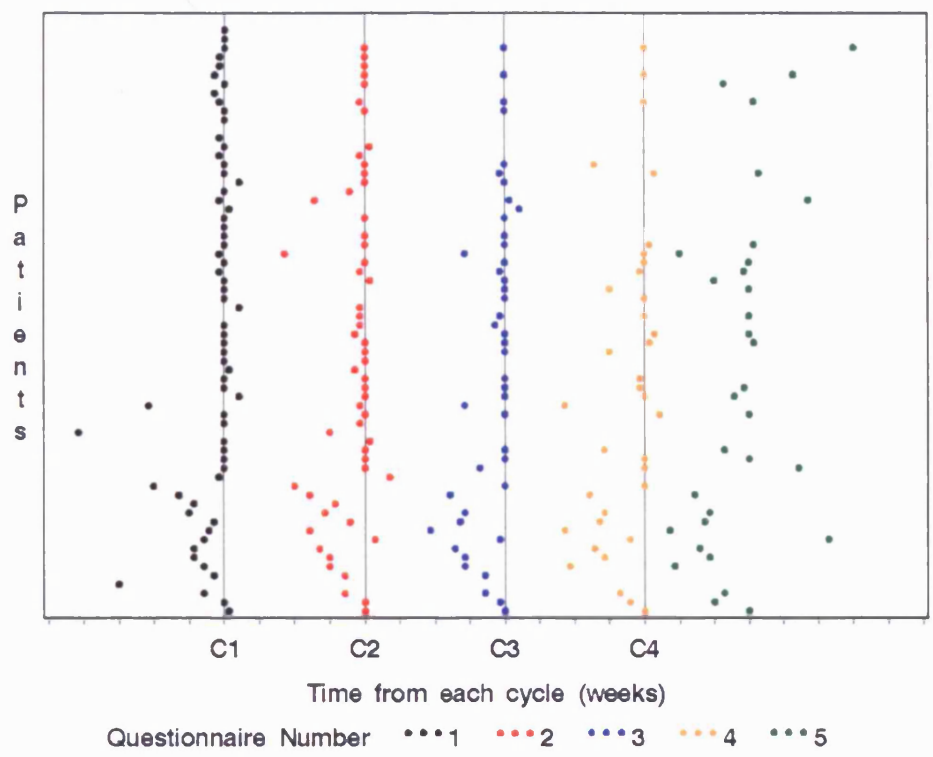


Figure 4.5: Timing of each questionnaire on the CT arm of the MIC study in relation to the time of its associated cycle of chemotherapy



4.5.4 Timing of Assessments in the ESPAC Study

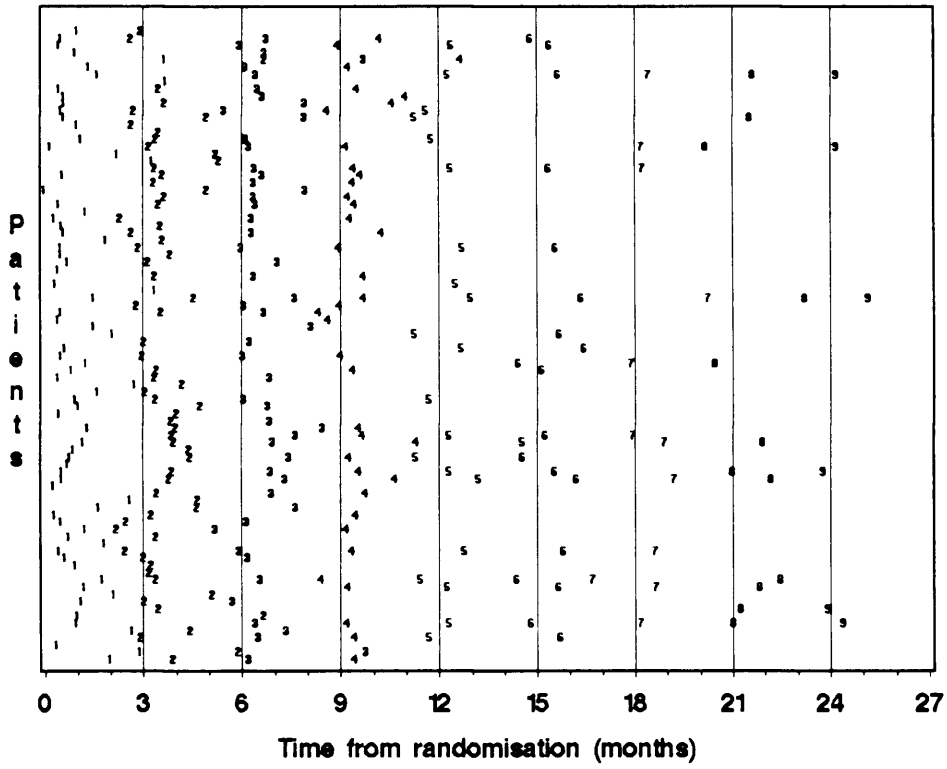
In the ESPAC study, the baseline quality of life assessment was planned to be taken after surgery and before the start of adjuvant treatment. Further assessments were planned every 3 months from baseline until death. Summary measures for the distribution of actual timings in relation to time of surgery are given in Table 4.3. The actual timings of the questionnaires 1 to 9 in relation to time of surgery are plotted for each patient in each treatment group in Figure 4.6. These show that on average the questionnaires conform approximately to their expected schedule but for each assessment there is considerable variation around the scheduled time. The variability was similar for both treatment arms.

Table 4.3: Summary measures for timings of assessments (in months) in the ESPAC study in relation to time of surgery

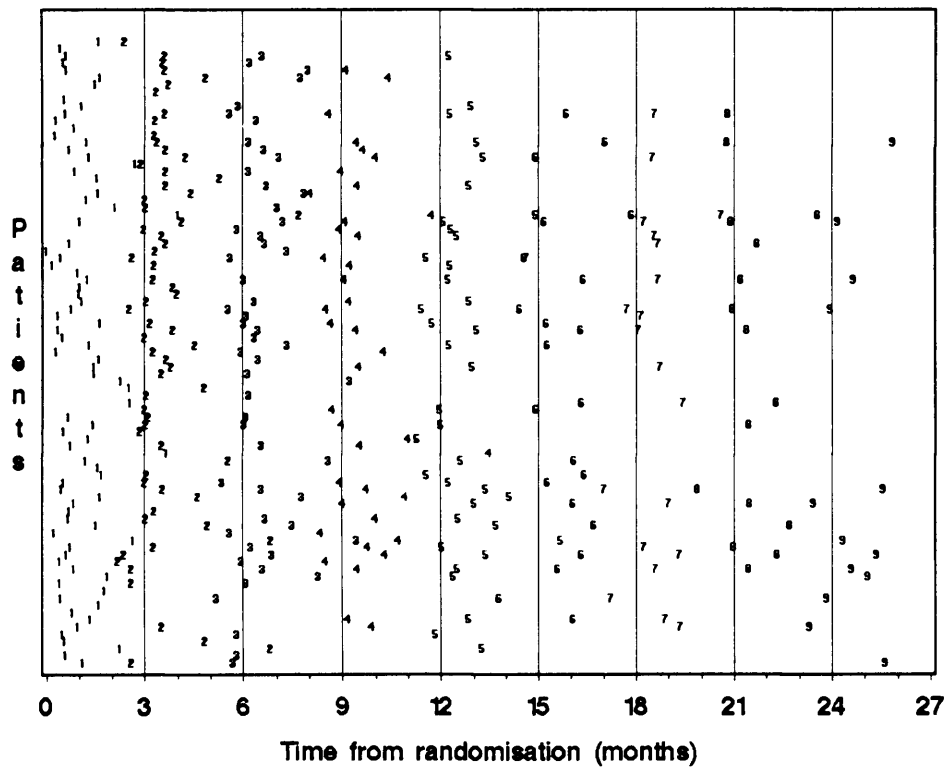
Assessment Number	N	Scheduled Time	Actual Time		
			Median	Minimum	Maximum
No CT					
1	75	0	0.9	0	3.7
2	67	3	3.5	2.2	6.7
3	53	6	6.5	5.2	9.8
4	39	9	9.4	8.3	12.7
5	22	12	12.3	11.3	14.5
6	19	15	15.5	14.4	16.4
7	12	18	18.3	16.7	20.3
8	12	21	21.6	20.2	23.2
9	7	24	24.2	23.8	28.0
10	4	27	27.6	26.9	28.5
11	3	30	30.2	30.0	30.2
12	3	33	33.5	33.0	34.2
13	1	36	35.7	35.7	35.7
14	2	39	39.4	38.9	40.0
15	1	42	41.9	41.9	41.9
16	0	45	-	-	-
17	1	48	48.4	48.4	48.4
CT					
1	81	0	1.0	0.0	4.0
2	66	3	3.5	2.2	7.7
3	54	6	6.4	5.2	9.4
4	40	9	9.5	8.1	13.5
5	38	12	12.5	11.3	15.7
6	22	15	16.0	13.8	17.9
7	21	18	18.6	14.6	20.6
8	16	21	21.4	19.9	23.6
9	13	24	24.6	23.3	25.9
10	11	27	27.6	26.3	28.9
11	8	30	30.4	30.0	33.9
12	6	33	33.7	33.0	36.6
13	5	36	37.2	33.4	40.4
14	3	39	40.0	39.5	40.6
15	3	42	43.3	43.3	44.5
16	1	45	46.5	46.5	46.5
17	1	48	49.6	49.6	49.6

Figure 4.6 Timing of first 9 assessments in the ESPAC study in relation to time of surgery

(a) NoCT group (N=88)



(b) CT group (N=87)



4.6 Summary

Quality of life of patients is generally measured via questionnaires. These questionnaires could yield descriptive or utility measures. In some types of analysis such as quality-adjusted survival analysis (see Chapter 8), utility measures are required and it may be necessary to transform descriptive measures to utility-type measures in order to be able to perform such an analysis. The measures of quality of life chosen for analysis for the MIC and ESPAC studies are typical of the types of measure collected in cancer clinical trials.

A major issue with quality of life data is the fact that instruments can yield multiple measures of quality of life. Although in this chapter an overview of possible approaches for handling the multivariate nature of quality of life data has been given, the issue of multiple endpoints in this thesis is not directly addressed. Full consideration of the issue however is given in the discussion section for each method of analysis in terms of extending the method to deal with multiple endpoints. Throughout the thesis the methods of analysis assume a single measure of quality of life. This could be a single item from a questionnaire, a score created from a number of items or a utility measure yielded by a questionnaire. This assumes that either the analysis would be restricted to one key endpoint, probably a global measure of quality of life or the analysis would need to be repeated for each of a number of key endpoints.

Quality of life assessments will generally be scheduled for a number of fixed points over time either in relation to trial entry or treatment but there could be a considerable amount of variation in the actual timing of assessments around these scheduled assessment times. For some types of analysis that are based on the scheduled time points, the level of variability may be important but if the analysis uses the actual timings of the assessments rather than the scheduled times then this variability may not be so relevant.

Studies of quality of life will relate either to a fixed time frame of interest or a patient-related time frame where the overall time frame for the study is defined by the longest individual study time. In both cases, even though all patients may have complete quality

of life data until death, the dropout of patients within the study time period creates non-random missing data and this needs to be accounted for in any analysis. Simultaneous analysis of quality of life and survival is the best approach to deal with the problem of informative dropout due to death and these methods are developed in Chapters 8, 9 and 10.

CHAPTER 5: PROBLEM OF MISSING QUALITY OF LIFE DATA

5.1 Introduction

The aim of this chapter is to discuss the types of missing data that occur in quality of life studies and the mechanisms that create them. The problems caused by missing data will be highlighted and the possible approaches for dealing with them will be discussed. The extent of missing data in the MIC and ESPAC studies is reported.

The occurrence of missing data is one of the main problems in analysing longitudinal quality of life data. Missing data do not only reduce the amount of information for inference and cause imbalance in the data structure, in terms of the number of measures per patient, but may cause the results from an analysis to be biased. It is therefore important to assess and report the extent and possible causes of missing data prior to any analysis. This enables the results from any analysis that ignores the missing data to be interpreted with the appropriate amount of caution and provides some indication for the need to use methods that make some account for the missing data.

The issues relating to missing data in studies of quality of life are discussed routinely in any publication on quality of life but have been the specific focus for some (Fairclough et al 1998b, Simes et al 1998, Curran et al 1998a and c, Fayers et al 1998, Troxel et al 1998). Some of the concepts relating to missing data are introduced in Section 5.2 and the potential bias caused by missing data is described in Section 5.3. The extent of missing data in the MIC and ESPAC studies is reported in Section 5.4. There are a number of different approaches for dealing with missing data and these are outlined in Section 5.5. Missing data can occur within a returned quality of life questionnaire. This can cause problems in calculating global scores or could cause problems if the missing data relate to an item on the questionnaire that has been chosen as one of the variables to be analysed. The problems relating to missing items has been discussed elsewhere (Fayers et al 1998) and was discussed earlier in Section 4.3.1 in the context of global scores. The focus here is on missing whole questionnaires. There are three different types of missing questionnaires that are considered; those resulting from late entry of patients to the trial, those that occur intermittently and those that result from a patient

'dropping out' of the quality of life study prior to the end of study. These are each considered in turn in Sections 5.6-5.8 with specific reference to the MIC and ESPAC studies. Section 5.9 provides a summary and discussion of the issues and the relevance for later chapters in the thesis.

5.2 Concepts of Missing Data

5.2.1 Definition of Missing Data

In the context of quality of life studies, missing data can mean two different things and it is important to clarify what we mean by 'missing data'. Within the time frame of the quality of life study (see Section 4.5.2), values that are not observed at scheduled assessment times are deemed as missing and will be treated as such in any standard analysis. There is a distinction however between missing data that the study never planned to collect and data that one would not expect to be missing. When a patient dies, the unrecorded responses to quality of life questionnaires that occur after death but within the time frame of the quality of life study are expected. The design of the study might also be that quality of life assessments cease at time of disease progression or at end of treatment for example and thus unrecorded data after these times but within the time frame of the study are also expected.

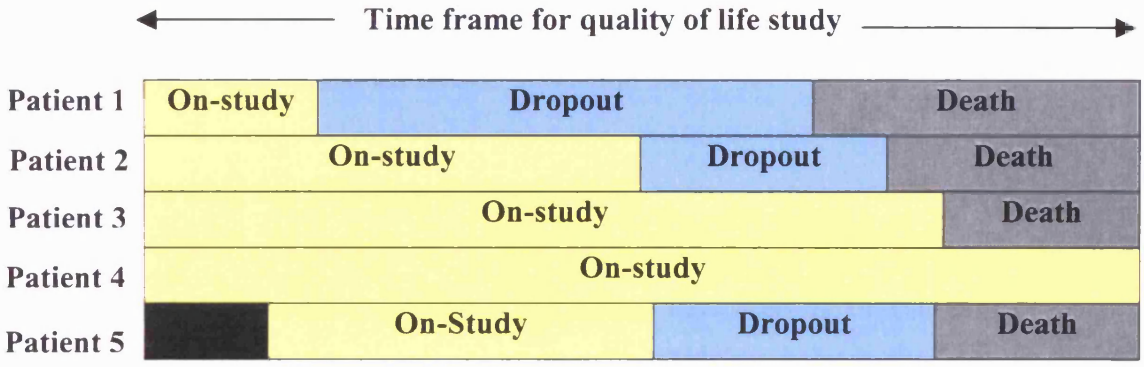
Whether the missing data are expected or unexpected, standard analysis does not distinguish between the two. The advantage of simultaneous analysis of quality of life and survival data is that the unobserved quality of life data after death will not be treated as missing data. These methods can be extended so that other expected unobserved values are not treated as missing. If the analysis is considering quality of life over time on a continuous rather than a discrete time scale then one could argue that the data at all time points between the actual assessment times are missing.

5.2.2 Types of Missing Data

The study period for each patient within the time frame of the quality of life study can be split into in three phases (see Figure 5.1). The *on-study* phase is the time during

which the patient participates in the quality of life study and is defined by the first and last questionnaire that the patient completes. If the patient enters the quality of life study late this may result in missing questionnaires prior to the on-study period (e.g. patient 5 in Figure 5.1). During the on-study phase the patient may have missing questionnaires at scheduled time points between two completed questionnaires and these are called *intermittent* missing questionnaires. If a patient ceases participation in the quality of life study prior to all the scheduled questionnaires being completed then they are said to *dropout* of the quality of life study (e.g. patients 1, 2 and 5 in Figure 5.1). Time of dropout can be defined in a number of ways (see Section 5.8). If the time of dropout is defined as the time of the first missed scheduled assessment for example and death occurs prior to this time then it can be said that the dropout occurs directly due to death and the patient does not actually enter the dropout phase (e.g. patient 3 in Figure 5.1). If however death occurs some time after the time of dropout then the patient can be said to enter a *dropout* phase prior to death. The problems associated with each form of missing data and the ways of handling them are discussed in Sections 5.6 to 5.8.

Figure 5.1: Illustration of the three stages in the study of quality of life for 5 different patients (black shaded area indicates phase when patient not on study)



5.2.3 Missing Data Mechanisms

The validity of the analysis of data with missing values is dependent on the mechanism associated with the missing data. A quality of life assessment missing at a particular time point may be categorised in three ways (Little and Rubin 1987, Laird 1988): *missing completely at random* (MCAR), when the probability of non-response is independent of all observed and unobserved quality of life values; *missing at random*

(MAR), when the probability of non-response depends on observed quality of life values but not on any unobserved values; *missing not at random* (MNAR), when the probability of response depends on unobserved quality of life values and possibly on observed values as well. Note that if the probability of non-response depends on certain known covariates such as patient characteristics or treatment group then the missing data can still be MCAR.

The extent to which the missing data can be ignored depends on the inferential framework (Rubin 1976, Laird 1988). With classical inference, only MCAR is ignorable for least squares but both MCAR and MAR are ignorable under maximum likelihood. With Bayesian inference both MCAR and MAR are ignorable. In all cases MNAR is non-ignorable in the sense that if it is ignored it will result in biases.

Determining if the mechanism is not MCAR is relatively straightforward since one can use the observed data to investigate any dependencies. Distinguishing between MAR and MNAR is much more difficult and requires assumptions which are at best only partially verifiable (Curran et al 1998a). With quality of life data, testing the mechanism is usually redundant since dropout due to death is known to be MNAR and in many cases dropout prior to death can often be assumed to be MNAR. Investigating the dropout mechanism is discussed further in Section 5.8.

5.3 Bias Caused by Missing Data

The bias caused by missing quality of life data is discussed by Curran et al (1998c). They show that the bias for all patients at any scheduled time point is given by:

$$Bias = (1 - P)(\mu_{nr} - \mu_r) \qquad [5.1]$$

where P is the proportion of responders, μ_{nr} is the mean response for the non-responders and μ_r is the mean response for the responders. Clearly the value of μ_{nr} will not be known but one can conclude that the amount of bias is dependent on the proportion of non-responders and the size of difference in means between the responders and non-responders. Thus, the missing data will only cause bias if the non-responders are

expected to have the different mean response to the responders and the greater the difference the greater the bias.

Another important observation that Curran et al (1998c) make is regarding the difference between two treatment arms. They show that, at a specific time point, the bias in the mean difference in response between treatments is given by the difference in the bias observed on each treatment arm. Thus, a treatment comparison is unbiased if the bias is the same in both treatment arms. Therefore it could be argued that if the non-response rate and reasons for non-response are equivalent on both treatment arms then the bias is likely to be similar and thus the treatment comparison will be unbiased. One can never be sure however that this is really the case and accounting for the missing data rather than ignoring them will always be a more appropriate approach. Also with the more common and complex scenario of longitudinal data with missing data at each successive time point then the assumption of equivalent bias on both treatment arms is less likely to be valid.

5.4 Reporting the Extent of Missing Data

The extent of missing data should be reported in conjunction with any analysis of the data to enable the readers to judge for themselves the effect the missing data may have on the results (Hopwood et al 1994). For each treatment arm, the amount of missing data that occurs over time should be reported. The levels of missing data resulting from late entry, intermittent missing and dropout should be distinguished. In some trials the reasons for non-response, when it occurs, are collected as part of the data and this information should also be reported if available. Other clinical data collected in the trial, such as recurrence or toxicity data, may give some indication as to the reason for the missing data.

The missing data are often reported in terms of the actual number of questionnaires that are returned at each scheduled time point in the study compared to the expected number (Hopwood et al 1994, Machin and Weeden 1998). The expected number at every scheduled time point could be the taken as the total number of patients in the study and non-response due to death and other reasons reported as a percentage of this. Some

authors adjust the expected number of questionnaires over time to reflect the reduction in patients due to death as this is an allowable reason for non-completion. If other reasons for non-completion have been defined in the study protocol such as disease progression or end of treatment then the expected number of questionnaires may be adjusted for this. Calculation of expected numbers in this way is not always straightforward because of the variation in the actual timing compared to the scheduled time. For example, a patient may complete the required questionnaire earlier than scheduled and then die prior to the scheduled time so that in terms of the expected number of assessments at the scheduled time they would not be included but they would be included in the actual number.

5.4.1 Extent of Missing Data in the MIC Study

In the MIC study a fixed number of assessments were scheduled, 5 on the CT arm and 4 on the PAL arm and the numbers of returned questionnaires diminishes over time (Table 5.1). By investigating the patterns of returned questionnaires the extent of the missing data can be determined (see Table 5.2).

Table 5.1: Number of returned questionnaires at each scheduled assessment time (A1 to A5) in the MIC study

	A1	A2	A3	A4	A5	Total
CT (N=67)	66	62	52	43	36	263
PAL (N=42)	37	33	33	24	-	129

Table 5.2 Patterns of returned questionnaires in the MIC study

	CT (N=67)	PAL (N=42)
Complete	31	20
Complete but with intermittent missing	4	1
Complete but with late entry	1	3
Dropout with no intermittent missing	29	15
Dropout with intermittent missing	2	1
Dropout with late entry	0	2

There were 6 patients in total with late entry (1 on CT and 5 on PAL) and 8 patients with intermittent missing (6 on CT and 2 on PAL). On the CT arm, 31 (46%) patients returned all planned questionnaires with a further 5 (7%) patients completing the full set apart from some intermittent missing or late entry. The remaining 31 (46%) patients dropped out of the study early, 2 of which also had intermittent missing. On the PAL arm, 20 (48%) patients returned all planned questionnaires with a further 4 (9%) patients completing the full set apart from some intermittent missing or late entry. The remaining 18 (43%) patients dropped out of the study early, 3 of which also had intermittent missing or were late entries. Further investigation of the patients who dropped out of the study showed that 6 (9%) patients on the CT arm and 4 (10%) on the PAL arm dropped out as a direct result of death, i.e. they died within 3 weeks of their last recorded assessment.

In terms of questionnaires (see Table 5.3), 386 (77%) of the 503 scheduled assessments were returned. The missing rate was similar on both treatment arms. Of the 117 missing questionnaires, the majority (86%) were missing due to dropout. The problem of late entry and intermittent missing is negligible in this study. The dropout rate of 20% of all scheduled questionnaires was equivalent on both treatment arms. By assuming that the timings of the missing questionnaires would have been at 3-weekly intervals from the last recorded assessment, it was possible to determine which questionnaires were missing as a direct consequence of death. If the timings of the missing questionnaires occurred after death then these were categorised as missing as a result of death. In total 35 (7%) questionnaires were missing as a result of death and the rate was similar on both treatment arms.

Table 5.3: Missing questionnaires in the MIC study

	CT (N=67x5=335)	PAL (N=42x4=168)	Total (N=503)
Scheduled assessments:			
Returned	259 (77%)	127 (76%)	386 (77%)
Missing	76 (23%)	41 (24%)	117 (23%)
Missing due to late entry	1	6	7
Missing due to intermittent	7	2	9
Missing due to dropout:			
Total	68 (20%)	33 (20%)	101 (20%)
Due to Death	21 (6%)	14 (8%)	35 (7%)
Other	47	19	66

5.4.2 Extent of Missing Data in the ESPAC Study

In the ESPAC study patients were required to complete questionnaires every 3 months from trial entry to death. In total the 175 patients in the study completed 710 questionnaires with patients completing assessments up to the 17th scheduled time point at 48 months, with later assessments being more frequent for the CT arm (see Table 5.4).

Table 5.4: Number of returned questionnaires at each scheduled time point in the ESPAC study

Assessment	CT	NoCT	Total
1	81	75	156
2	66	67	133
3	54	53	107
4	40	39	79
5	38	22	60
6	22	19	41
7	21	12	33
8	16	12	28
9	13	7	20
10	11	4	15
11	8	3	11
12	6	3	9
13	5	1	6
14	3	2	5
15	3	1	4
16	1	0	1
17	1	1	2
Total	**389**	**321**	**710**

On both treatment arms more than 50% of patients completed all assessments prior to dropout. A high proportion of patients had intermittent missing assessments and some had missing assessments due to late entry to the study. These patterns are summarised in Table 5.5 and discussed in further detail in Sections 5.6-5.8.

Table 5.5: Patterns of response for patients in the ESPAC study

(a) CT arm (N=87)

	Censored follow-up	Dropout due to death	Dropout prior to death	Dropout prior to censoring	Total
Complete	3	19	21	5	48 (55%)
With Intermittent Missing	8	7	10	8	33 (38%)
With Late Entry	1	0	2	1	4 (5%)
With Intermittent Missing and Late Entry	1	0	0	1	2 (2%)
Total	13 (15%)	26 (30%)	33 (38%)	15 (17%)	87

(b) NoCT arm (N=88)

	Censored follow-up	Dropout due to death	Dropout prior to death	Dropout prior to censoring	Total
Complete	1	26	21	3	51 (59%)
With Intermittent Missing	11	4	5	4	24 (28%)
With Late Entry	1	6	4	1	12 (14%)
With Intermittent Missing and Late Entry	0	0	1	0	1 (1%)
Total	13 (15%)	36 (41%)	31 (35%)	8 (9%)	88

5.5 General Approaches for Dealing with Missing Data

5.5.1 Complete Case Analysis

This is the simplest approach to missing data. The analysis is restricted only to those patients who have a complete set of values over time. This enables standard methods for complete data to be used but there are a number of problems with this. The sample size for the study will be reduced and if, for example, there are a large number of patients with just the odd intermittent missing questionnaire then the sample size could be dramatically reduced despite only a small proportion of the total data being missing. Not only does this reduce the power to detect differences in treatment but also it is wasteful of the data that has been collected on the incomplete cases. Further to this,

patients who complete all assessments must be survivors and compliers and hence probably are a 'healthy' subgroup of all patients in the quality of life study, not a random sample, and the analysis will result in over-estimates of quality of life. The longer the quality of life study period, the greater the level of attrition and the more reduced, and possibly more biased, the sample for complete case analysis will be. This may be the case even if the treatment groups are comparable in terms of the amount of missing data and the causes. Complete case analysis is not recommended unless the proportion of patients with missing data is very small, say less than 5% (Fayers and Machin 2000).

5.5.2 Available Case Analysis

Many standard methods of analysis do not require complete data and can be used to analyse all available data on all patients. This is preferable to complete case analysis but in general these methods assume that the missing data are missing at random and in some cases missing completely at random. Much of the missing data in quality of life studies are likely to be missing not at random and therefore the standard methods for available case analysis will give biased results. If the reasons for missing data are incorporated into the analysis then using the available quality of life data should give valid results and indeed this is the main approach that is adopted in this thesis.

Standard methods of analysis make use of the Expectation-Maximisation (EM) algorithm (Dempster et al 1977). The EM algorithm is a widely used iterative procedure for maximum likelihood estimation in the presence of missing data. The process iterates between an expectation step and a maximisation step. In the expectation step the conditional expectation of the missing data given the observed data and current parameter estimates are estimated. The maximisation step updates the parameter estimates by maximising the complete data log-likelihood as though the missing data had been filled in. It is not the purpose of the EM algorithm to fill in the missing values but to fill in the functions relating to the missing data in the complete data likelihood (Little and Rubin 1987). The algorithm proceeds until convergence.

5.5.3 Imputation

An alternative approach to complete or available case analysis is to fill-in the missing data with plausible values, a process called imputation (Little and Rubin 1987). The use of imputation for quality of life data has been discussed (Curran et al 1998b, Fayers and Machin 2000). Plausible values are estimated using available quality of life data and other information that may be accessible about the patient. The artificial complete data set can then be analysed in a standard way.

The information that can be used for imputation is extensive. Quality of life data are generally longitudinal and multidimensional and so for a missing item on a given patient at a given time point, imputation could be based on values for other items on the same questionnaire, values for the same item at the same time point but for other individuals or values for the same item and the same individual but at other time points. In addition there are baseline patient characteristics, survival data and other clinical data collected in the trial that may be informative for imputation. Clinical trials will often collect clinician-assessed performance status and weight regularly throughout the trial and also details of treatment-induced toxic events. These longitudinal clinical data may all be related to the quality of life of the patient at any time.

There are a number of methods that replace missing data values with a single imputed value. Simple mean imputation replaces the missing value with the mean for patients with observed values. Regression imputation replaces the missing value with a predicted value from a regression model that quantifies the relationship between the quality of life measure and a number of covariates in those with observed data. Last value carried forward replaces the missing value by the previously observed value or similarly first value carried back replaces the missing value with the next observed value. By assuming a linear change between observed values on either side of a missing value, linear interpolation can be used to impute the missing value. If the quality of life values are a limited set of ordinal values then Markov chain imputation can be used. In this approach, using the probabilities of transition between pairs of values estimated from the observed data and the values observed on either side of the missing value, a randomly generated value is imputed.

In some situations it may be preferable to impute values less formally than any of the above approaches. If information is known about why values for a patient are missing then sensible values could be imputed for those data. The most extreme example of this is death. If a patient has missing data because they have died then it may be sensible to give them the worst quality of life value for all the assessments after death. Indeed if the quality of life measure is a utility then a value of 0 is valid. This approach is often used without necessarily being referred to as imputation. Similarly, if it is known that a patient has not responded to a questionnaire because of poor health then an appropriate poor quality of life value can be imputed for that missing data. Sensitivity analysis can be used to investigate the effect of allocating extreme values to all missing data.

In all cases, when missing values are replaced by single imputed values, the variability of data will be underestimated in any analysis since the imputed values will be treated in exactly the same way as the observed values and no account is taken of the uncertainty in the imputation. Multiple imputation is an alternative approach to single value imputation that attempts to overcome these problems (Rubin 1987, Schafer 1997). In multiple imputation, each missing value is replaced by a number of plausible values. A small number, say 5, is usually sufficient to represent the uncertainty adequately. Each of the artificial complete datasets is analysed using standard methods and estimates are combined using Rubin's rules (Rubin 1987) such that the variability *between* the estimates from the different data sets is accounted for as well as the variability *within* each of the datasets. For this reason multiple imputation is generally a more valid approach than single value imputation and although standard software is available to perform multiple imputation such as the MI procedure in SAS, the approach may not be feasible for quality of life data. The longitudinal nature of the data also needs to be accounted for in the multiple imputation as well as other variables that may be predictive of the missing data creating a complex model for multiple imputation (Schafer 2001). If the data are missing not at random then it may not be possible to predict the missing values using observed data. In addition when complex methods of analysis are being used to analyse the data, combining estimates from the multiple datasets is not necessarily straightforward.

5.6 Dealing with Late Entry of Patients into the Quality of Life Study

If a patient enters the quality of life study late, it means that at a minimum they have not completed a baseline questionnaire. In addition their entry may have been so late as to miss other early questionnaires. Probably the most common reason for late entry into the study and certainly for missing baselines will be administrative. Eligibility criteria for entry to trial will usually include a requirement for the patient to be fit enough to participate in the trial and willing and able to participate in the quality of life study. Thus, in general, providing the patient conformed to eligibility criteria for entering the study, late entry of patients to the quality of life study should not be due to the health of the patient. For this reason it is likely that the data from the missing questionnaires will be MAR and possibly even MCAR.

Missing baseline data will not be a problem for methods of analysis that are based on modelling the quality of life data over time (see Sections 6.4, 8.7 and Chapter 10) as long as they are MAR. For other methods of analysis such as summary measures analysis (see Section 6.3), subject-based approaches to quality-adjusted survival analysis (See Sections 8.4 and 8.5) and multistate modelling of quality of life and survival (see Chapter 9), the baseline values of quality of life for all subjects are required. Omitting patients with missing baseline data will provide unbiased results if the data are MCAR, but is wasteful of information if the patients have completed the remainder of their questionnaires. It may therefore be preferable to impute values for the missing baseline data. Having imputed missing baseline values, any other subsequent missing questionnaires due to late entry will become intermittent missing values and should be dealt with accordingly (see Section 5.7).

Missing values could be replaced by the mean baseline value for all patients in the trial. This approach is fairly crude and does not account for the fact that baseline quality of life may be associated with survival time. It may therefore be preferable to impute a value based on the known survival time of the patients. Baseline values may also be associated with other baseline patient characteristics such as age or stage of disease and these could also be used for imputation. These approaches do not account for the fact that baseline values will also be related to later values in the patients' series and maybe

also to how the values change over time. Replacing missing values with the patients' first recorded value, i.e. first value carried backwards, is a crude way of imputation but at least makes some account of the value in the series which is likely to be most highly correlated with the baseline value. A more sophisticated approach would be to use all the known information on a patient to impute a baseline value. Imputing just a single value however will not account for the uncertainty in the imputation and hence multiple imputation based on all known information would give the most valid results.

Decisions as to which approach to use in dealing with missing baselines will be based on the extent of the problem. If the number of missing baselines is relatively small then it may not be worthwhile employing the more sophisticated approaches, as the imputed data should not have a large influence on the overall results. The effect of the decision on the overall results could be examined in a sensitivity analysis. Missing baseline values are not the focus of this thesis and hence relatively simple approaches are adopted to overcome the problem.

The methods of analysis mentioned above that require baseline values of quality of life, require not only a baseline value but also more specifically a value of quality of life at time of entry into the study, i.e. time 0. In some studies, the administration of questionnaires may be such that the baseline questionnaire is taken prior to treatment but not necessarily at time of entry to study, for example baseline questionnaires may be administered on the first day of the first cycle of chemotherapy, which may be a week or so after trial entry. Thus, even though a patient may have a baseline questionnaire, they may still cause problems in such an analysis if the timing of the questionnaire is not at entry to study. Imputation of the quality of life values in such patients at time 0 is required. The simplest approach and the one that is generally used here is to use first value carried backwards. This approach is generally adequate when the time delay between entry and baseline assessment is short as it is unlikely the health status of the patient will change much in this short time span *before* treatment, but the imputation is less valid for greater delays. The extent of this problem in the MIC and ESPAC studies is discussed further below.

5.6.1 Late Entry of Patients in the MIC Study

Of the 109 patients in the MIC study, 6 patients (1 CT and 5 PAL) entered the study late and therefore had missing baseline questionnaires. It is believed that these are likely to be missing purely for administrative reasons and with such a small number of patients it would be feasible just to omit these patients. The size of the study however is relatively small and the omission would have the greatest impact on the PAL arm, which already has a reduced number of patients compared to the CT arm and therefore it was decided that these patients should be included. Missing baseline values for *GQS* were replaced by a single imputed value based on a regression model of baseline *GQS* with age, sex, performance status, survival time and *GQS* at second assessment. This does not take account of the full series of quality of life values and will underestimate the variability but with so few missing values it provides an adequate means to overcoming the problem.

In the remaining 103 patients who completed a baseline questionnaire, 37 (18 on CT and 19 on PAL) were completed on day of entry to study. The timings of questionnaires in the MIC study were described previously in Section 4.5.3. On the CT arm 44 (67%) patients completed baselines within 1 week from trial entry with maximum time 2.7 weeks. On the PAL arm 23 (62%) patients completed baselines within 1 week with maximum 2.9 weeks. Thus in the majority of cases the use of first value carried backwards to time of trial entry will be adequate as the period of imputation is generally short.

5.6.2 Late Entry of Patients in the ESPAC Study

Of the 175 patients in the ESPAC study there were 19 (11%) patients with missing baseline questionnaires (see Table 5.5). The incidence was greater on the NoCT arm compared to the CT arm (15% versus 7%). As with the MIC study, it is believed that these are likely to be missing purely for administrative reasons and as such these will be treated as MCAR. Since the timings (see Section 4.5.4) of some of the baseline questionnaires are not dissimilar to the timings of the second questionnaires for these patients with missing baselines it was decided to use the same method of imputation for both missing baselines and delayed baselines. Thus, when the analysis requires values

of *GHSS* at time of entry to study, then a first value carried back will be used in all cases.

The timings of questionnaires in the ESPAC study were described previously in Section 4.5.4. On both treatment arms, baseline questionnaires on average were completed within approximately 1 month of trial entry, with a maximum delay of 4 months. Thus, in the majority of cases, the use of first value carried backwards to time of trial entry will be adequate, as the period of imputation is relatively short and is prior to treatment. The imputation may not be so appropriate for those with a long delay between trial entry and baseline assessment nor for those whose first questionnaire occurs at the second scheduled assessment after treatment has started.

5.7 Dealing with Intermittent Missing Quality of Life Questionnaires

An intermittent missing questionnaire occurs when a patients does not complete a questionnaire at a scheduled time point but has completed questionnaires at scheduled time points before and after the missing form. The occurrence of intermittent missing questionnaires will often be due to administrative reasons and in such cases it may be possible to assume that these data are missing completely at random. An intermittent missing questionnaire however could occur because the patient felt too ill on that occasion to participate in the study, especially if they are undergoing active treatment, or conversely because the treatment worked so well that the patient felt that quality of life was no longer relevant. In such cases the probability of response may depend on covariates, such as treatment, or may depend on the quality of life experienced at that time, in which case it would be invalid to assume that missing data mechanism was ignorable. Methods for dealing with such missing data are complex (Diggle et al 1994). Intermittent missing values are not the focus of this thesis and hence in all cases the missing data mechanism is assumed to be ignorable.

In the MIC study, of the 503 scheduled assessments, there were only 9 intermittent missing questionnaires that occurred in 8 patients (6 on CT of which 1 has two consecutive intermittent values and 2 on PAL). These are likely to have occurred for administrative reasons. Intermittent missing questionnaires therefore are not a problem

in the MIC study and assuming such data to be MAR should not have a great influence on the conclusions.

In the ESPAC study, 35% of patients had at least one intermittent missing questionnaire (see Table 5.5). The incidence of intermittent missing questionnaires is to be expected because the study was over a longer period of time and the longer the series of questionnaires, the greater the chance of an intermittent missing questionnaire occurring. The incidence was greater on the CT arm compared to the NoCT arm (40% versus 29%) and this could have been because patients on the CT arm generally had longer series of questionnaires. No details are available as to the reason for these questionnaires being missing. The majority of them are likely to be missing for administrative reasons and all intermittent missing questionnaires will be assumed MAR.

5.8 Dealing with Dropout from the Quality of Life Study

If a quality of life study aims to collect data at scheduled points over a fixed time frame, such as 1 year from trial entry, then a patient who ceases to participate in the study before completing all scheduled assessments is defined as a *dropout*. Patients may drop out prior to the fixed end-of-study time for a number of reasons including death, disease progression, illness, early cessation of treatment, lack of treatment effect, lost to follow-up or patient choice. In addition, if a patient is still on-study at the time of analysis then they will appear as a dropout in the analysis, as their quality of life assessments will cease prior to the fixed end-of-study time. If the quality of life study aims to collect data at scheduled points over a patient-related time frame, such as until death or disease progression, then even if a patient completes all scheduled assessments within their time frame, in terms of analysis they are still deemed to have dropped out at the time of the event such as death that defines their end-of-study time (see Section 4.5.2).

Once a patient has dropped out of a study, no more information on quality of life is available from that point onwards. Quality of life information could be considered as *censored* at the date of dropout. This term can cause confusion as the censoring of the survival times may be at a different time to the censoring of the quality of life data. In

terms of analysis, dropouts cause problems because they create missing data. In studies where survival, as well as quality of life, is an endpoint, patients are generally severely ill and therefore dropout due to death or illness will be a common occurrence. In such situations, the dropout process may depend on the unobserved measurements (i.e. those measurements that would have been observed had the patient not dropped out) and the incomplete follow-up of subjects is called *informative* (or *non-ignorable*) *dropout* (Diggle and Kenward 1994). The missing quality of life data after dropout will be MNAR and needs to be accounted for in any analysis of the data.

The dependence of the probability of dropout on previously observed values can be assessed by splitting the patients into different groups according to their dropout time and then plotting the mean quality of life scores over time for each of these groups (Hopwood et al 1994, Curran et al 1998c). Similarly, means over time can be plotted according to whether patients complete their next scheduled assessment or not (Carpenter et al 2002). If a relationship is observed, then the missing data cannot be MCAR and must be either MAR or MNAR. See Figures 6.6 and 6.7 for examples of these types of plots given for the MIC and ESPAC data.

If the reasons for dropout have been collected as part of the study or if clinical data are available that may indicate the reasons for dropout then this information can be used in any analysis. Dropouts that occur for different reasons may be treated differently. At a minimum it should be possible to categorise the reason for dropout as death or non-death, although determining if a dropout is directly attributable to death is not necessarily straightforward. If a patient dies then they drop out of the study at their time of death and since assessments occur at discrete points in time, their last assessment will invariably be some time prior to death. It will be necessary to define the maximum time that can have elapsed between last assessment and death for the reason for dropout to be given as death. If the reason for dropout is known to be unrelated to the health of the patient then the missing data for these individuals can be treated as missing at random.

One way of dealing with missing data from dropouts is to impute values to replace the missing data, from data that already exists. There is a variety of methods for doing this (see Section 5.5.3) but it may be difficult if there is a large amount of missing data and only valid if data are missing at random. In general therefore imputation will not be

appropriate and the methods of analysis must allow for informative dropout. Methods that simultaneously analyse quality of life and survival data overcome the problems of missing data associated with dropout due to death, but additional informative dropout prior to death may also need to be considered in any analysis.

One way to deal with dropout prior to death is to simultaneously analyse time to dropout with quality of life instead of time to death and this is considered in Chapters 8, 9 and 10. In all methods, the time to dropout needs to be defined. The exact dropout time is not usually known or recorded and therefore needs to be estimated. The only information that is available is the first missed scheduled assessment. The time of this missed scheduled assessment could be taken as the time of dropout or, if assessments are scheduled at regular time intervals, then the time of dropout for an individual could be this amount of time after their last recorded assessment. If the reason for dropout is known to be unrelated to the health of the patient then the time of dropout could be treated as a censored observation of their true dropout time (Henderson et al 2000, Hogan and Laird 1997a). More complex approaches to estimating dropout time have been considered where each dropout time is imputed from a uniform distribution over the appropriate time interval between last observed and first missed assessment time (Henderson et al 2000). Sensitivity analysis can be used to determine if the definition of time to dropout has any effect on the conclusions.

Time to dropout can be compared for each treatment group in a clinical trial using standard survival analysis techniques such as Kaplan-Meier survival curves and log-rank tests (see Chapter 3). If there is no difference between the groups in terms of time to dropout then it could be argued that dropout will not bias the treatment comparison (see Section 5.3). The survivor function for the dropout event, that includes both dropout prior to death in addition to dropout directly due to death, can be thought of as *dropout-free survival*. This is a comparable concept to outcomes such as disease-free survival and progression-free survival, which are commonly used in cancer clinical trials. In these endpoints the event is some clinical indication that the cancer has returned or is progressing, which in some cases will be death. Here dropout is taken to be an event that indicates worsening health, which at the extreme is indicated by death.

5.8.1 Dropout in the MIC Study

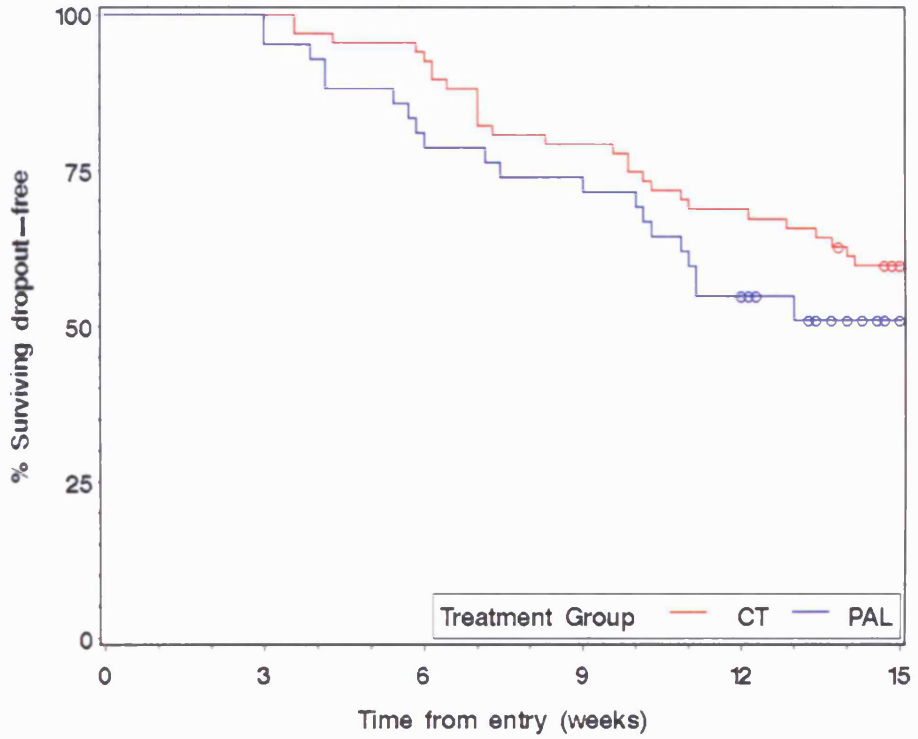
In the MIC study, patients who did not complete the final scheduled questionnaire (5^{th} on CT and 4^{th} on PAL) were defined as dropouts. As discussed earlier, there were 31 dropouts on the CT arm and 18 on the PAL arm. If a patient died within 3 weeks of the last recorded assessment then the time of dropout was defined as the time of death and the dropout was attributed directly to death. There were 6 dropouts due to death on the CT arm and 4 on the PAL arm. For the remaining patients who dropped out, the time of dropout was defined as 3 weeks after the last recorded assessment. The time of dropout for 4 patients on the CT arm was after the 15-week study period and thus the dropout time was censored at 15 weeks. For those patients who did not drop out (36 on CT and 24 on PAL), their time of dropout was censored 3 weeks after their final scheduled assessment time or at 15 weeks, whichever came first. Patients were censored prior to the end-of-study even though they continued on follow-up during this time because had they been obliged to complete further assessments then they may have refused to participate because of illness. On the CT arm only 4 patients were censored for dropout prior to 15 weeks with the earliest censored at 13.9 weeks, whilst on the PAL arm, because only 4 assessments were required, 17 patients were censored prior to 15 weeks with the earliest at 12 weeks. If a non-dropout died within 3 weeks of their last assessment then the time of dropout was recorded as an event at their time of death (2 patients on CT arm). This choice was made because had further assessments been required then it is known that dropout would have occurred at this time.

The details of dropouts and survival are given in Table 5.6 and dropout-free survival function on each treatment arm is shown as Kaplan-Meier survival curves in Figure 5.2. The curves suggest that time to dropout or death tended to be earlier on the PAL arm compared to the CT arm but a log-rank test showed that this difference was not statistically significant at the 5% level ($p=0.22$). Nevertheless, the observed difference between treatment arms in terms of dropout may have an impact on the comparison of treatments in terms of quality of life.

Table 5.6: Details of dropout and survival within 15-week study period in the MIC study

		CT (N=67)	PAL (N=42)
Non-dropouts	Died within 15 weeks	0	3
	Survived 15 weeks	36	21
Dropouts	Censored dropout time within 15 weeks	4	0
	Direct result of death	6	4
	Died after dropout within 15 weeks	6	9
	Survived after dropout until 15 weeks	15	5

Figure 5.2 Kaplan-Meier curves for dropout-free survival within 15 weeks in the MIC study (censored values shown by circles)



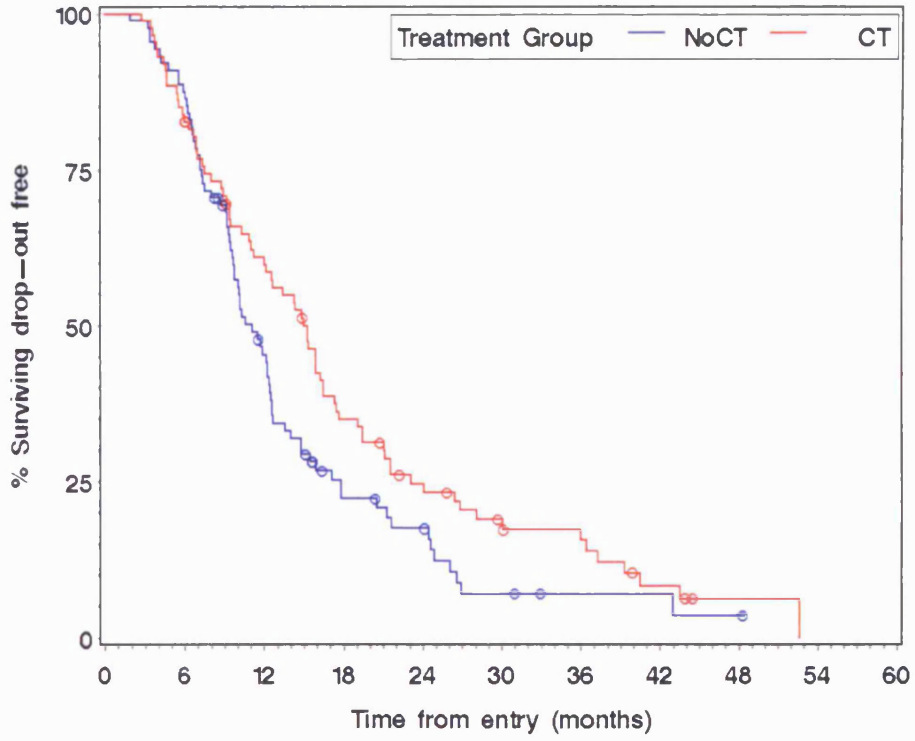
5.8.2 Dropout in the ESPAC Study

In the ESPAC study patients are required to complete assessments at three-monthly intervals until death and the definition of time to dropout therefore is more straightforward than for the MIC study. For those patients who die within 3 months of their last recorded assessment, the time of dropout is recorded as an event at their time of death and the reason for dropout is attributed directly to death. For those patients

whose last recorded assessment was more than 3 months prior to their last survival follow-up (whether alive or dead), the time of dropout was recorded as an event 3 months after their last recorded assessment. For those patients whose survival time is censored within 3 months of their last recorded assessment, the time of dropout is censored at this time too.

From Section 5.4.2 it can be seen that on the NoCT arm 36 patients dropped out of the study directly due to death, 31 prior to death and 8 prior to censoring with the remaining 13 patients censored for dropout at their last follow-up time. On the CT arm 26 patients dropped out of the study directly due to death, 33 prior to death and 15 prior to censoring with the remaining 13 patients censored for dropout at their last follow-up time. Kaplan-Meier curves for dropout-free survival are shown in Figure 5.3 with censored dropout times shown by circles. Dropout on the NoCT arm tended be earlier than the CT arm and a log-rank test showed this to be statistically significant at the 5% level (p=0.05). This differential dropout may have an impact on the comparison of treatments in terms of quality of life and should be accounted for in any analysis.

Figure 5.3 Kaplan-Meier curves for dropout-free survival in ESPAC study (censored values shown by circles)



5.9 Summary and Discussion

Missing data are a common occurrence in longitudinal studies of quality of life. In general, missing data cause problems in terms of unbalanced data and reduced precision in estimation but, in quality of life studies, the additional issue that the some of the missing data are likely to be missing not at random causes particular problems for analysis. If missing data are not missing at random then standard analyses will produce biased results.

There are a number of different types of missing data that occur in quality of life studies. Missing data that occur due to patients entering a study later than planned will often be missing for administrative reasons and it is therefore feasible to either impute missing values or ignore the missing data in the analysis. Similarly, it will often be valid to assume that non-compliance to an assessment in a series of completed assessments will produce intermittent missing data that are missing at random. The key problem with quality of life studies and the focus of this thesis is missing data that occur due to dropout.

The definition of dropout will be specific to a quality of life study and relates to the time frame of the study. In terms of analysis, whatever the design of the study, deaths that occur during the analysis period cause non-ignorable missing data for all time from death onwards. The missing data caused by dropout due to death should therefore be accounted for in any analysis of the quality of life data to ensure unbiased results. The methods for simultaneous analysis of quality of life and survival discussed in Chapters 8, 9 and 10 all deal with informative missing data resulting from dropout due to death.

In addition, patients may dropout prior to death. The time period between last assessment and last survival follow-up that defines when a patient drops out of the study needs to be specified. In general, a patient will be deemed as a dropout prior to death at the first missed scheduled assessment time after their last recorded assessment. If dropout is believed to be due to poor health, which may often be the case in studies of quality of life, then the dropout is informative and the missing data are non-ignorable and this should be accounted for in any analysis of the data. Methods of simultaneous

analysis of quality of life and survival are extended in Chapters 8, 9 and 10 to account not only for death but also for additional dropout prior to death. Methods of simultaneous analysis that are based on health states, such as some forms of quality-adjusted survival analysis and multistate modelling, enable death and dropout prior to death to be considered as two separate states. Other forms of quality-adjusted survival analysis and also joint modelling treat death and dropout prior to death as equivalent events, which may not always be appropriate. Methods may need to account for different reasons for dropout when these are available to ensure that those patients who drop out of the quality of life study for non-health related reasons are treated differently from those who drop out due to illness for example.

Dropout rates in both the MIC and ESPAC studies appeared to differ between the treatment arms, with dropout occurring earlier on the control arms compared to the chemotherapy arms. This differential dropout may have an impact on the comparison of treatments in terms of quality of life and should be accounted for in any analysis. In all subsequent analysis, because the reasons for dropout are not known, dropout is assumed to occur for reasons of ill health. This is likely to be a valid assumption for the majority of patients and provides conclusions in terms of a 'worst-case scenario'.

In summary, the extent of missing data in studies of quality of life should always be reported in conjunction with any analysis of the data to enable the reader to assess the validity of the analysis and associated results. The key problem for analysis will be the informative dropout of patients from the quality of life study due to death and possibly the additional dropout of patients prior to death for reasons related to health. Standard methods of longitudinal analysis will provide biased results for such data. Methods for the simultaneous analysis of quality of life and either survival or dropout-free survival, that are described and developed in Chapters 8, 9 and 10, will provide an unbiased comparison of treatments in terms of quality of life.

CHAPTER 6: ANALYSIS OF QUALITY OF LIFE DATA USING STANDARD LONGITUDINAL METHODS

6.1 Introduction

The aim of this chapter is to describe and illustrate some of the standard methods of analysis for longitudinal data and discuss the problems of their application to quality of life data. In general, the aim of any analysis of longitudinal quality of life data is to evaluate the change in quality of life over time and the effect of covariates, especially treatment, on this change. All methods discussed in this chapter are standard approaches to analysing longitudinal data and all assume that the mechanism that gives rise to any missing data is ignorable. Application of these methods to longitudinal quality of life data, which typically include non-ignorable missing data, will therefore generally give biased results. The methods however provide an initial starting point in the analysis of quality of life data, provided that the potentially biased results are highlighted and interpreted with caution. In the context of this thesis, the methods give some initial insight into the quality of life data for the MIC and ESPAC studies introduced in Chapter 4 and provide background material to later chapters.

Measures of quality of life may be continuous, ordinal or binary (see Section 4.2). For continuous measures, such as those derived from a linear analogue scale or an aggregated global score, the appropriateness of parametric or non-parametric methods depends on whether the data can be considered to be normally distributed. It may be possible to normalise a distribution of a continuous measure using a transformation, such as log or square root, and then compare treatments using a parametric approach on the transformed variable (Armitage and Berry 1987). It could be argued that since the continuous measures have truncated distributions, parametric methods will always be inappropriate (O'Brien et al 1987) but this view is not adopted here. In addition, since aggregated global scores are usually made up of a number of ordinal responses, the variable will often only take a restricted number of values on the continuous scale, again compromising the validity of treating the measure as continuous.

The analysis of longitudinal quality of life data should begin descriptively. Section 6.2 discusses exploratory analysis, which provides an initial impression of the data using graphical methods. It investigates the distribution of the data, highlights outliers and examines trends over time and hence possible appropriate models. The simplest method of analysis is to summarise the longitudinal data for each patient in the form of a summary measure, which can then be analysed using standard non-longitudinal methods. These methods are discussed in Section 6.3 but since this approach does not make full use of the data, modelling approaches are generally preferable and form the main focus of this chapter. Models for the data are discussed in Section 6.4. Models are applied to the MIC and ESPAC data in Section 6.5 using a classical approach to parameter estimation, whilst a Bayesian approach is discussed in Section 6.6. All analyses include only observed quality of life data and the implications of the missing data on the analysis are discussed throughout the chapter and specifically in Section 6.7, which summarises the chapter.

6.2 Exploratory Data Analysis

The interpretation of longitudinal quality of life data can be difficult and an initial exploratory analysis often gives an insight into the data before any formal testing or modelling is carried out. Descriptive methods of analysis do not have the problems associated with multiple testing and hence the quality of life data can be explored as extensively as desired. This section discusses the variety of ways to graphically explore the data both at the patient level and the treatment group level. Descriptive methods for analysing quality of life data have been reviewed in detail elsewhere (Fayers and Jones 1983, Beacon 1996, Machin and Weeden 1998, Fayers and Machin 1998b).

6.2.1 Patient Profiles

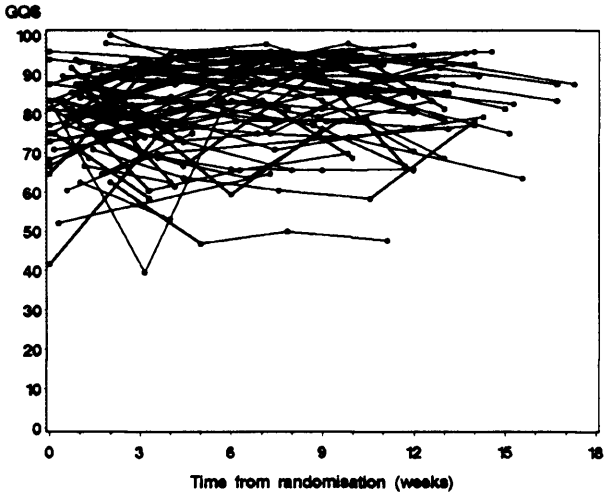
Patient profiles can be examined by plotting individual patient scores over time. They may reveal a consistent pattern across patients, give some indication to the distribution of the data and will highlight errors, outliers and patterns of missing data. The values of quality of life can be plotted as single points at the appropriate times. Often the quality of life study is designed to assess patients at scheduled time points but there is usually

some variation in the timings of the actual assessments around these scheduled times and a choice needs to be made whether the scheduled or actual assessment times are most appropriate. Usually these points are joined so that a patient's quality of life, measured at a number of discrete time points, is described by a 'curve' over continuous time. By connecting measures at discrete time points in this way, all quality of life values between the actual assessment times are effectively being imputed.

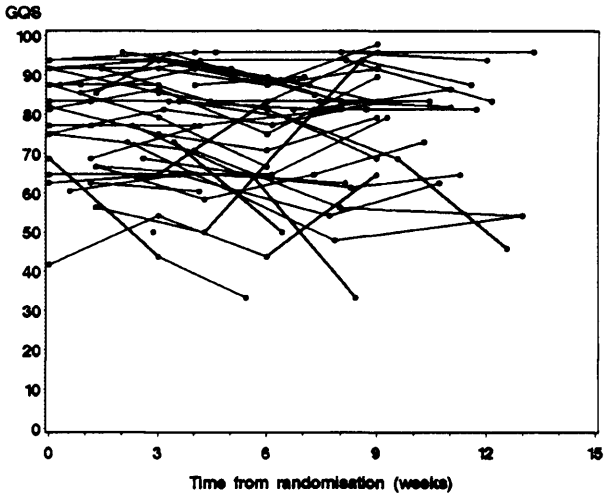
Each option for joining the points is associated with an inherent underlying assumption and choices should be made that are appropriate to the situation. A step function can be used if it is thought that the quality of life remains constant until the next measure. In some cases it may be appropriate to assume that the change happens midway between two time points. Another option is to assume that there is a linear change from one time point to the next. If there are intermittent missing assessments, a decision needs to be made as to whether it is reasonable to assume that such missing data are missing at random. If so, then the missing data can be ignored and the quality of life values on either side can be connected as normal. If not, then gaps should be left in the curve at these times. A profile ending early will signify a patient dropout.

Figure 6.1: Patient profiles of *GQS* over time in the MIC study

(a) CT Group (N=67)



(b) PAL Group (N=42)

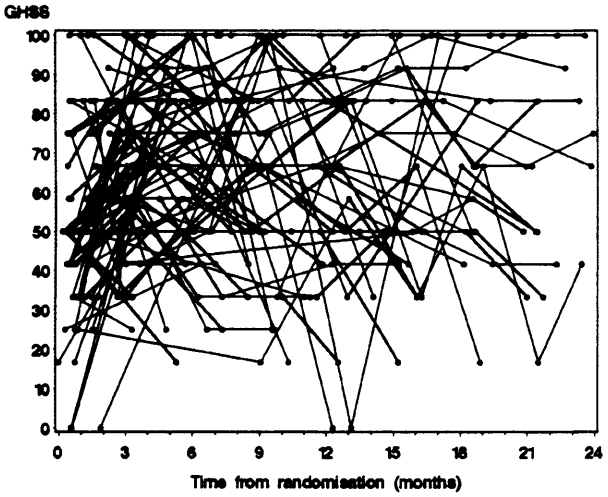


In the MIC and ESPAC studies, quality of life profiles were created by plotting the *GQS* and *GHSS* scores against the actual assessment times, assuming a linear change between assessments and ignoring intermittent missing values (see Figures 6.1 and 6.2). The

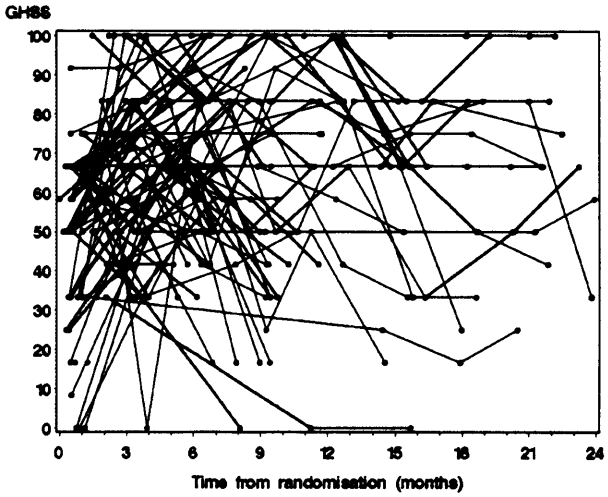
patient profiles for the ESPAC study are shown for 24 months from date of entry to trial, as the data are very sparse after this time. In each study, individual profiles were overlayed on one graph for each treatment group separately to enable patterns in each treatment group to be compared. An alternative method of display is to plot the profiles for each patient as a set of mini-graphs (Beacon and Thompson 1996). It is often impractical to display data from large numbers of patients using overlayed graphs or mini-graphs and one alternative is to plot a simple random sample of the patients in a study (Beacon 1996).

Figure 6.2: Patient profiles of *GHSS* over 24 months from randomisation in the ESPAC study

(a) CT Group (N=88)



(b) NoCT Group (N=87)



The graphs for the MIC study do not show any obvious pattern in change over time or any clear differences between the two treatment groups. At any point in time, the scores are slightly skewed towards the upper end of the range, particularly in the CT group. In the ESPAC study the patient profiles are extremely noisy with no clear trend over time in either of the treatment groups and no apparent differences between the groups.

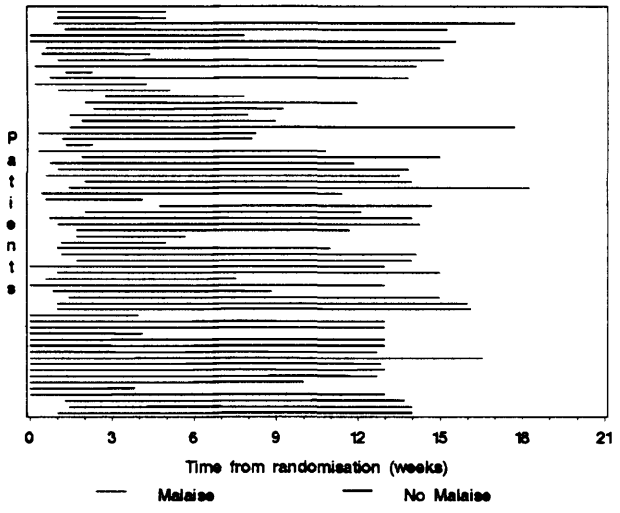
For binary data or ordinal data with a small number of categories, it may be preferable to plot each patient's profile in the form of a line over time that varies in style depending on the value of a quality of life variable at that time. The patient profiles are usually ordered by date of entry to study. For ordinal data, several different types of line can be used to represent the different levels of the variable, but the diagram becomes

difficult to interpret as the number of levels increases. Continuous and ordinal level data can be adapted for this type of presentation by grouping the values into a small number of levels, ideally two.

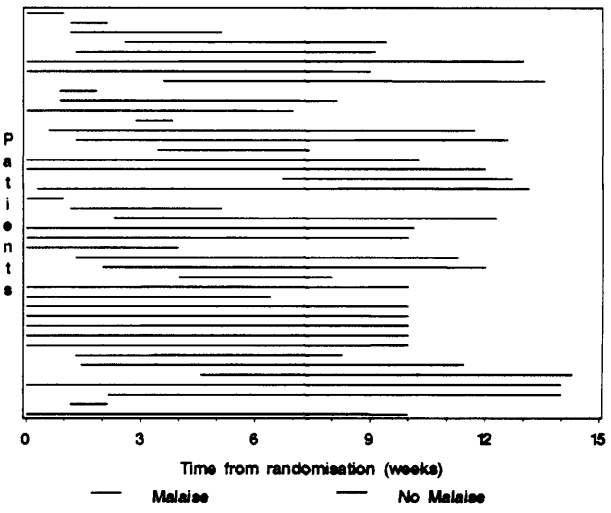
Patient profiles showing time spent with malaise (*MAL*=1, 2 or 3) or without malaise (*MAL*=0) were plotted for the MIC Study (see Figure 6.3). The value of malaise measured at each assessment is assumed to carry over to the next recorded assessment and the last recorded value is carried over for one week post-assessment (to enable this last value to be displayed). As with *GQS*, the graphs do not show any obvious pattern in change over time or any clear differences between the two treatment groups.

Figure 6.3: Patient profiles of *MAL* over time in the MIC study

(a) CT Group (N=67)



(b) PAL Group (N=42)



Patient profiles can be grouped and overlayed according to differing lengths of follow-up and also, possibly, differing reasons for shortened follow-up, giving possible insight into the association of the dropout process with previous quality of life (see Section 6.2.2 for further details on this approach relating to group profiles).

6.2.2 Group Profiles

After examining individual patient profiles of quality of life, it is often necessary to summarise the experience of patients within each treatment group to enable the treatments to be compared more clearly in relation to their effect on quality of life over

time. Group summary measures can only be calculated and plotted if the data relate to scheduled assessment times. Plots of group profiles over time, overlayed on the same graph, enable a clear visual comparison of groups.

There are two main types of group summary statistic that can be plotted over time, the mean or median value of quality of life or a proportion with a certain level of quality of life. The choice depends on the type of quality of life measure being summarised. The mean or median quality of life scores in each group can be plotted over time. Bars at each time point representing 95% confidence intervals for the mean or median should be included. This is the most useful way to represent continuous data. For ordinal data, the use of means is not theoretically correct, but will often be more informative than the median especially with floor or ceiling effects and may be considered if the ordinal scale is long. If the data are binary then the proportion with a symptom or side-effect can be plotted over time. Bars at each time point representing 95% confidence intervals for the proportion should be included. If the data are ordinal or continuous then the proportion reaching or exceeding a certain level of quality of life over time can be plotted but detail on the severity of the symptom or side effect will be lost.

Summarising the quality of life of patients in each treatment group over time is complicated by the problem of missing data. Group summary measures at each time point can be calculated and plotted either for the subgroup of complete cases or for all available data. As discussed in Section 5.5, the patients included in a complete case analysis are unlikely to be a representative subset of the overall sample and the group summary values will probably be over-estimates of the true values. Also, due to the reduced sample size, the confidence intervals will be wide. If all available data are used then the reduction in the sample size over time should be made clear in any graph or table by specifying the number contributing to the descriptive measure at each time point. The confidence intervals should also reflect the decreasing sample size by their increasing width. Interpretation of such data needs caution since a comparison of results at different time points compares different groups of patients. Subjects contributing to summary measures at later time points are likely to be the 'healthier' members of the original group. Available data analysis is not incorrect but should be interpreted as a conditional analysis. Plotting all available data shows the mean quality of life at each

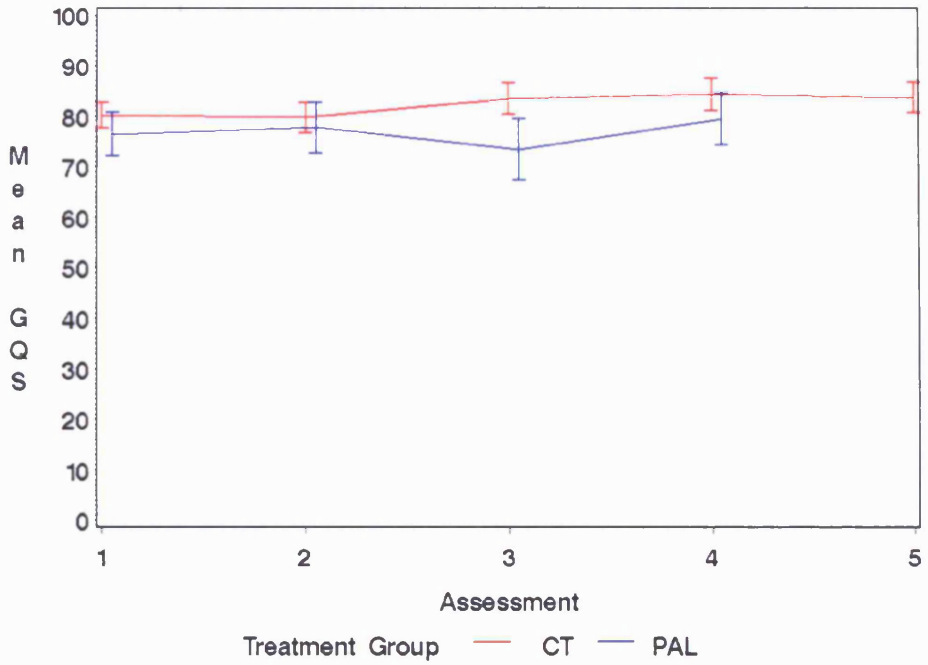
time point conditional on the patient being alive and well enough to complete an assessment.

For each treatment group in the MIC study, the mean *GQS* score at each scheduled assessment time together with the 95% confidence interval are plotted and group profiles created by assuming a linear change from one assessment to the next (Figure 6.4a). The reduction in the numbers of patients over time is shown to aid judicious interpretation. Conditional on patients being alive and well enough to complete questionnaires, there does not appear to be a change over time in the two treatment groups although a slight initial improvement in the CT group and a slight deterioration in the PAL group results in a difference at the third assessment.

The proportions of patients with malaise (*MAL*=1,2,3) together with the associated 95% confidence intervals are also plotted over time for each treatment group (Figure 6.4b). This shows the extent of malaise in the two groups is relatively high at around 60% and the profiles are reasonably comparable over time. In both groups the extent of malaise reduces initially but then increases.

Figure 6.4: Group profiles for the MIC study: summary statistics together with 95% confidence intervals at each scheduled assessment time

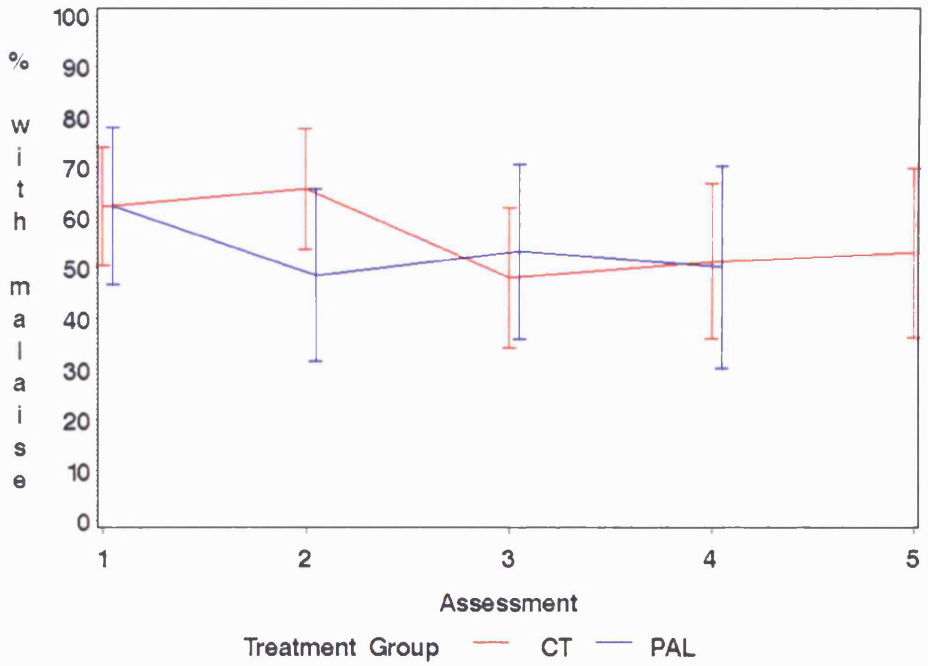
(a) Mean *GQS*



Numbers at risk:

MIC+ PC	66	62	52	43	36
PC	37	33	33	24	

(b) Proportion with malaise

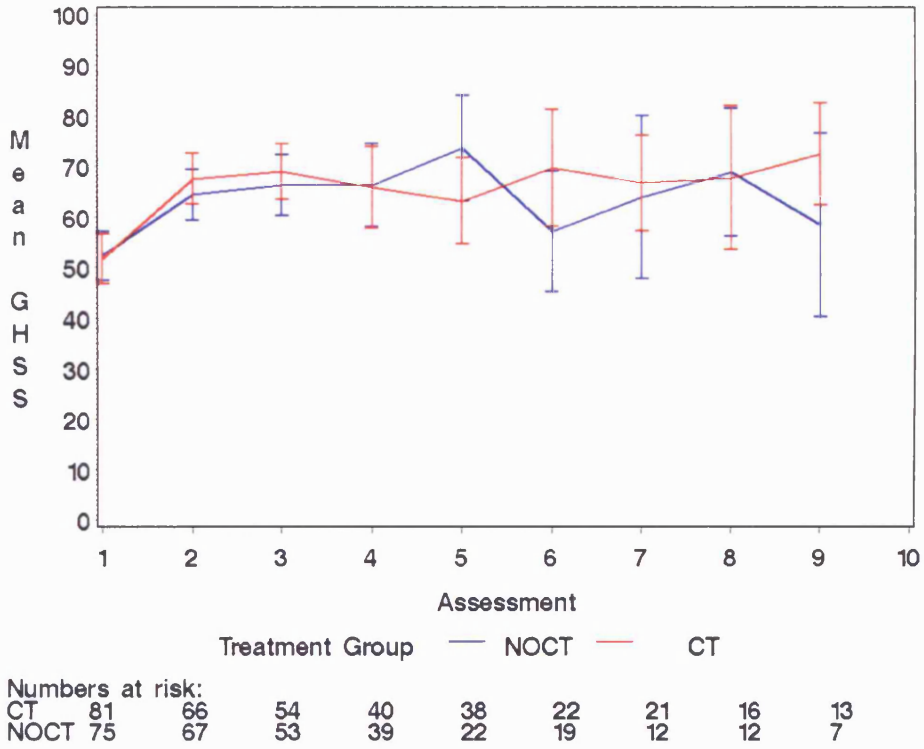


Numbers at risk:

MIC+ PC	66	61	50	41	34
PC	37	33	32	24	

For each treatment group in the ESPAC study, the mean *GHSS* score at each scheduled assessment time together with the 95% confidence interval are plotted and group profiles created by assuming a linear change from one assessment to the next (Figure 6.5). The reduction in the numbers of patients over time is shown to aid judicious interpretation. Although some patients were assessed at scheduled time points after the ninth scheduled assessment, the numbers of patients were too small for a reliable estimation of the mean *GHSS* at those times and therefore have been excluded. Conditional on patients being alive and well enough to complete questionnaires, there does not appear to be any difference between the treatment groups, with both groups improving in quality of life initially and then fluctuating over time with no obvious trend.

Figure 6.5: Group profiles of mean *GHSS* and 95% confidence intervals over time in the ESPAC study



Another useful exploratory plot of group summary measures over time is to split patients into subgroups according to differing lengths of follow-up. Subgroups can be formed according to number of completed assessments (Hopwood et al 1994) or a more detailed breakdown based on different reasons for dropout, such as censoring, death or

lost to follow-up could be used (Cox et al 1992). Quality of life over time can be compared across subgroups to establish its association with the dropout process. If there are no obvious differences between the subgroups then it may be valid to combine the data for analysis (Hopwood et al 1994). The quality of life study needs a large number of participants for subgroups to contain adequate numbers of subjects for this type of analysis. Pattern mixture modelling described in Chapter 10 is based on this type of approach of dividing patients into subgroups according to their dropout pattern and modelling the longitudinal values in each group separately. The method provides a means for analysing longitudinal quality of life whilst accounting for informative dropout.

For the MIC and ESPAC studies, the mean *GQS* and *GHSS* are plotted over time for dropout groups defined by the number of their last scheduled assessment (see Figures 6.6 and 6.7). In the MIC study, there were 6 dropout groups: those who dropped out after 1 assessment (N=8), 2 assessments (N=15) or 3 assessments (N=16), those on the CT group who dropped out after 4 assessments (N=10), those on the PAL group who completed all 4 scheduled assessments (N=24) and those on the CT group who completed all 5 scheduled assessments (N=36). In general, observed quality of life appeared to deteriorate prior to dropout. In the ESPAC study there were 10 dropout groups defined by the last assessment time between 1 and 9 (N=16, 31, 26, 31, 15, 14, 9, 7, 9 for times 1 to 9 respectively) and those who completed an assessment at a later time point than the 9^{th} assessment and therefore did not dropout within this 24 month time period (N=17). All profiles increase to a peak and then decrease over time prior to dropout.

Frequency distribution plots of the individual quality of life values over all time points (see Chapter 4, Figures 4.1 and 4.3) give an initial indication of the distribution of the data, which will guide what analysis may be appropriate. Distributions can also be plotted by scheduled assessment time (see Figure 6.8). Box and whisker plots of data at each scheduled time point gives an indication on the distribution of data at each time point. The decrease in sample size over time can be reflected by making the width of the box proportional to the number of subjects at that time point (Cnaan et al 1997).

Figure 6.6 Mean *GQS* over time by dropout group in the MIC study (red lines indicate groups of patients who have dropped out; black lines represent completers; dotted lines represent CT group only; solid black line represents PAL only)

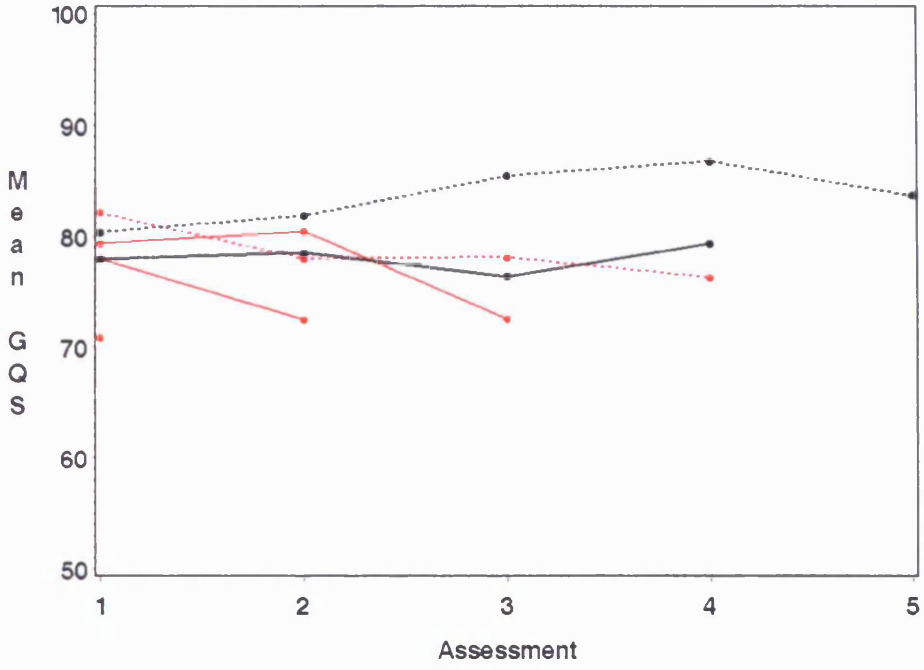
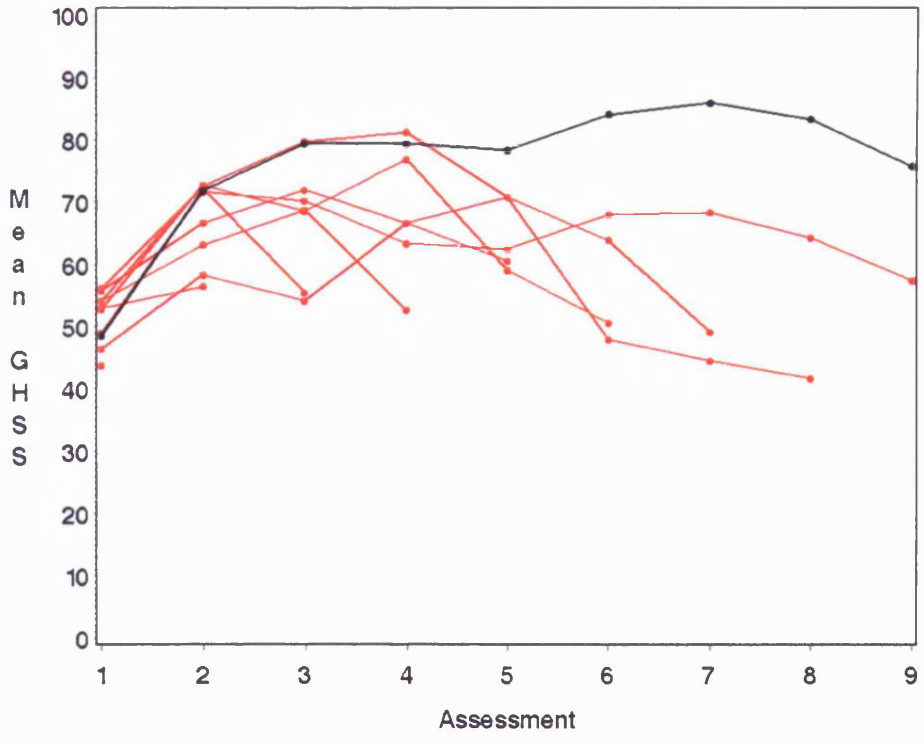


Figure 6.7 Mean *GHSS* over time by dropout group in the ESPAC study (red lines indicate groups who have dropped out prior to or at the 9th assessment; black line indicates group who went on to complete assessments after the 9th)



6.2.3 Autocorrelations and the Covariance Matrix

Another important aspect of the quality of life data to explore is the correlations between measures taken at different time points. These are known as *autocorrelations* or *serial correlations*. This can only be done when the data relate to scheduled assessment times and the variation around the scheduled time is not too large. If there are T scheduled time points for quality of life assessment and ρ_{kl} (k=1,2,...,T and l=1,2,...,T) represents the correlation between a specific quality of life measure at time point k with those at time point l then these autocorrelations can be expressed in the form of a $T \times T$ symmetric matrix where the diagonal elements equal 1 and $\rho_{kl} = \rho_{lk}$ for all values of k and l ($k \neq l$). Since $\rho_{kl} = \sigma_{kl} / \sigma_k \sigma_l$, where σ_{kl} is the covariance and σ_k^2 and σ_l^2 are the variances of measurements taken at time points k and l, the autocorrelation matrix has a direct relationship with the covariance matrix. The scatter plots of values taken at each pair of time points can be plotted in a matrix. The types of structure that the covariance matrix takes is important for modelling the quality of life data and this is discussed further in Section 6.4.2.

Although there is some variation in the timing around the scheduled assessment times in the MIC Study (see Section 4.5.3), it is possible to examine the association between measures of *GQS* taken at different time points. The associations are examined for both treatment groups combined. Figure 6.8 shows a matrix of scatter plots, with frequency distributions on the diagonal to indicate variability and Table 6.1 shows the associated correlation coefficients, variances and covariances. The correlation between successive measures is high and generally decreases slightly as the time between measures increases, as might be expected.

Scatter plots of the *GHSS* scores at different time points in the ESPAC study is not informative as, with a limited set of 13 possible discrete values for GHSS, many of the points are superimposed. Also with the larger number of scheduled assessment times the matrix of scatter plots become less manageable. The correlations, variances and covariances for the first 9 scheduled assessment times are given in Table 6.2. The pattern of the covariance matrix is less clearly defined than that for the MIC study.

Figure 6.8 Matrix of scatter plots showing association of *GQS* at different assessment times (A1-A5) in the MIC Study

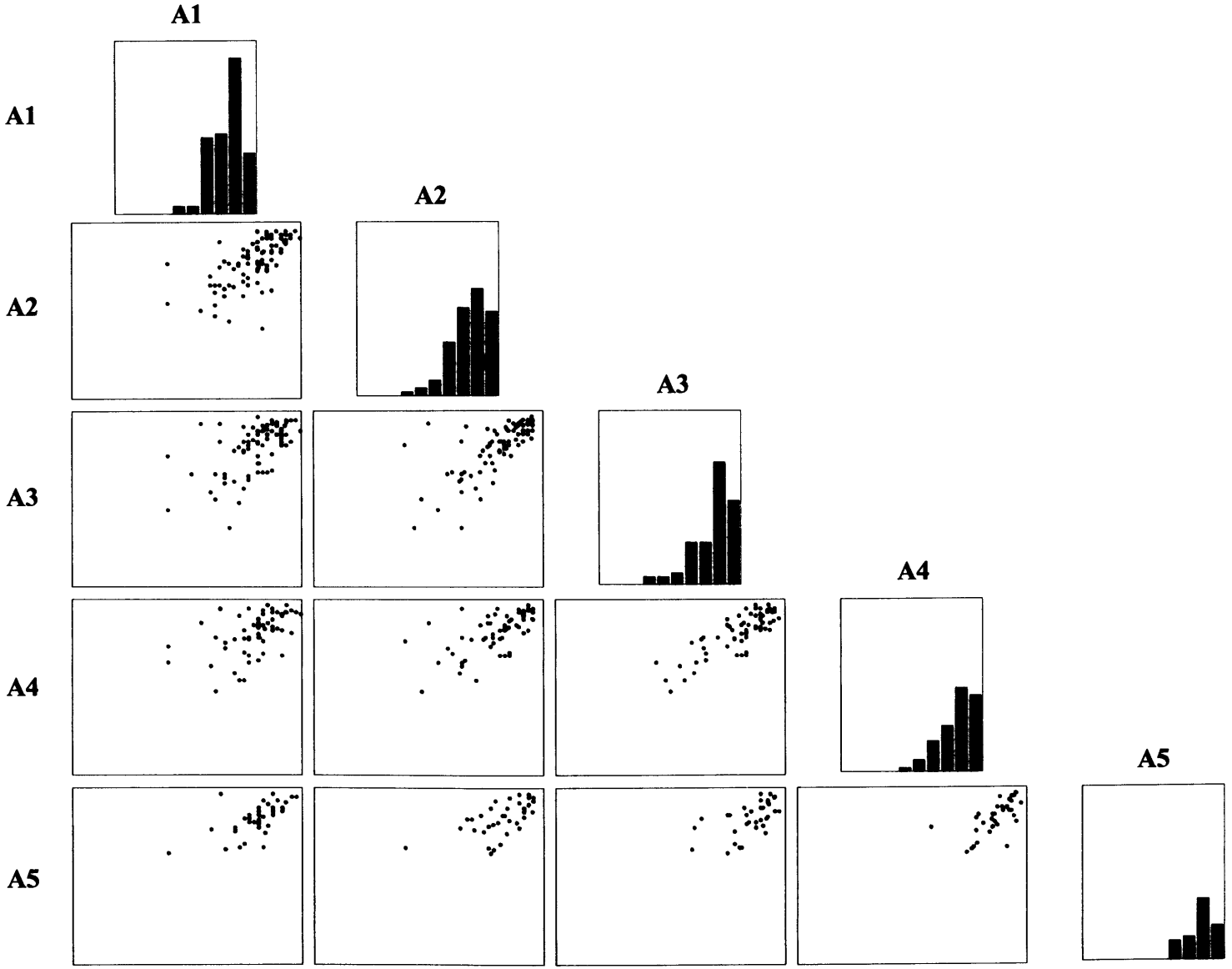


Table 6.1 Matrix of Pearson correlation coefficients (below diagonal) for association of *GQS* at different assessment times (A1-A5) in the MIC study and associated variance and covariance terms (on diagonal and above)

	A1	A2	A3	A4	A5
A1	142.44	100.52	102.09	68.22	70.02
A2	0.64	174.33	140.83	99.31	74.31
A3	0.57	0.71	223.48	146.86	85.30
A4	0.48	0.64	0.83	139.04	71.42
A5	0.62	0.59	0.60	0.64	89.72

Table 6.2: Matrix of Pearson correlation coefficients (below diagonal) for association of *GHSS* at different assessment times (A1-A9) in the ESPAC study and associated variance and covariance terms (on diagonal and above) for the first 9 assessments

	A1	A2	A3	A4	A5	A6	A7	A8	A9
A1	489.35	100.79	64.05	105.32	-12.00	135.96	6.79	42.95	-157.79
A2	0.22	437.14	207.61	166.88	149.68	80.69	13.31	231.60	-93.18
A3	0.13	0.46	466.89	241.39	165.91	125.59	69.01	277.07	220.42
A4	0.18	0.31	0.43	672.67	314.72	446.84	111.02	441.25	316.73
A5	-0.02	0.27	0.29	0.46	696.72	316.90	102.12	75.79	13.82
A6	0.22	0.14	0.21	0.63	0.44	759.06	413.01	623.08	355.83
A7	0.01	0.03	0.13	0.18	0.16	0.62	589.23	541.61	319.65
A8	0.08	0.43	0.50	0.66	0.11	0.87	0.86	669.00	320.66
A9	-0.34	-0.21	0.48	0.57	0.02	0.61	0.62	0.58	452.49

6.3 Summary Measures Analysis

Using summary measures has been promoted as the simplest method for analysing longitudinal data (Matthews et al 1990, Matthews 1993). It reduces the repeated measures over time for an individual to a *single* summary measure, which can then be analysed using standard non-longitudinal statistical methods such as for treatment group comparisons. The choice of summary measure needs to be clinically meaningful and will depend on the nature of the measure together with the disease and treatment under investigation.

There is a wide range of summary measures that could be chosen to represent longitudinal quality of life data. For example, the mean, median, minimum or maximum

value measured over time or last score may be appropriate. Alternatively, the change in quality of life between two time points or the slope representing change over time for each individual or the area under the curve could be computed (Lydick et al 1995, Hollen et al 1997). Time to the occurrence of a quality-of-life-related event could be the summary measure of choice. Some of the more widely used summary measures for quality of life data are discussed in more detail below. If the data are highly imbalanced with respect to the number of repeated observations per patient, then for some patients the summary measure will be based on a large number of observations and for others it may be based on very few. In a simple summary measures analysis, no account is taken of how many measures contributed to the summary outcome, the uncertainty in each summary outcome and the length of time over which the summary is taken. In addition if the calculation of the summary measure involves missing data or it is not possible to calculate the summary measure because of missing data and the missing data are non-ignorable then the results from the summary measures analysis may be biased.

6.3.1 Palliative Indicators

Simple binary indicator summary measures can be used to define palliation (Stephens and Hopwood 1995, MRC Lung Cancer Working Party 1991 and 1992). For example, an indicator may be set to show whether an individual experienced a decrease in the level of severity of a particular symptom or in overall quality of life at any time compared to baseline, or alternatively whether an individual experienced total disappearance of a symptom at any time. Other summary measures include duration of palliation and percentage of patient survival time during which there was palliation (MRC Lung Cancer Working Party 1991, 1992). These summary measures will only be valid if all patients have died, otherwise they should be restricted to a set follow-up time for which all patients in the study have been followed. The problems in defining palliation have been discussed (Stephens et al 1999) and problems with using these types of summary measures are discussed further elsewhere (Billingham and Cullen 2003; see Appendix I).

6.3.2 Time to the Occurrence of a Quality-of-Life-Related Event

Longitudinal quality of life data can be summarised by a single value representing the time to the occurrence of a quality-of-life-related event. The most widely used clinical endpoint, which in some situations will be quality-of-life-related, is relapse-free survival, i.e. time from study entry to disease relapse. Time to first occurrence of an important clinical adverse event or disease progression (Nabholtz et al 1996) and time until first occurrence of Karnofsky index of less than 60 (Rosenman and Choi 1982) have both been considered as quality-of-life-oriented endpoints and used as summary measures on which to compare treatments. Time to first improvement or time to first worsening of quality of life from baseline have also been suggested as summary measures, with times for patients not achieving such targets being censored (Hopwood et al 1994). Time to palliation of various symptoms, in those patients with the symptom present pre-treatment, has also been used as a summary measure for treatment comparison (MRC Lung Cancer Working Party 1996).

Once a quality-of-life-oriented endpoint has been defined, standard survival analysis techniques can be used to analyse the data. One advantage of this method is that patients who do not achieve the endpoint because of either death or dropout prior to death would still be included in the analysis as a censored data point, thus dealing with the problem of informative dropout. However, because the censoring mechanism may be related to the time-to-event, standard survival analysis techniques may be invalidated by informative censoring. Another advantage of this summary measure is that it can be used in situations where quality of life has been assessed at varying time points. However, if quality of life is assessed at only a few widely spaced time points, then the summary measure will be very crude.

An additional issue with the approach is that the potential for change, whether in terms of improvement or worsening, depends on the baseline value. For example, the worse a patient is at baseline, the greater the potential for improvement, so patients who do not experience a symptom at baseline are not able to improve and are therefore excluded from the analysis. This will only be a problem in the comparison of treatments if the treatment groups differ with respect to baseline symptoms. The fact that a treatment

may prevent a symptom from starting may be important in itself and would not be assessed by this type of endpoint.

6.3.3 Area Under the Curve

A standard summary measure used for longitudinal data is area under the curve (AUC). For each individual, the quality of life measures over time can be plotted as a curve as discussed in section 6.2.1. The AUC for each individual can be calculated and used as a summary measure (Lydick et al 1995, Hollen et al 1997, Fairclough 1997). If the measure of quality of life is a utility then the AUC is the quality-adjusted survival time for that individual for the period of assessment. This will be discussed in more detail in Chapter 8.

If the length of time over which quality of life is assessed differs across individuals then this needs to be adjusted for in the calculation of the summary measure. Some authors divide the area by the length of the observation time from trial entry to last assessment and compare treatments using a standardised area under the curve (SAUC) (Qian et al 2000, Fayers and Machin 2000, Bailey et al 1998). This is effectively an analysis of the distribution of quality of life during the time that they participate in the study. The problem with this approach is that individuals with short follow-up time will have equal weighting to those with long follow-up time (Ganiats et al 1995). In particular, if death is not accounted for then, for example, an individual with a quality of life value of 0.7 for 3 months who then dies will be treated with equal weight as someone who survives say for 3 years with a quality of life value of 0.7. An alternative approach is to choose a time period for which all individuals have completed all assessments and calculate the AUC for this fixed time period. This may not always be possible and may require imputation to complete curves for individuals within the fixed time period. This is discussed further in relation to the subject-based approach to quality-adjusted survival analysis in Sections 8.4 and 8.5.

6.3.4 Summary Measures Analysis for the MIC and ESPAC Studies

For both the MIC and ESPAC studies, two different summary measures were chosen for analysis; the change in quality of life between the third and first questionnaire and the

standardised area under the curve (SAUC). The third questionnaire was chosen for analysis in both studies as it was a clinically relevant time point and the choice of an earlier rather than later questionnaire minimises the number of dropouts. For the MIC study the third questionnaire represents the time when patients on the CT arm are half way through treatment and patients on the PAL arm will be approximately 6 weeks from entering the study. In the ESPAC study the third questionnaire is taken approximately 6 months from surgery and patients on the CT arm should have completed their chemotherapy by this time. The AUC summary measure needed to be standardised as there was considerable variation between patients in terms of their length of follow-up due to death, censoring and dropout. The results are given in Tables 6.3 and 6.4 for the MIC and ESPAC studies respectively.

For the change in quality of life summary measure, patients who do not have a third questionnaire or a baseline questionnaire are omitted from the analysis. This is a considerable number for the ESPAC study (see Table 6.4). In both studies, the majority of patients with a missing third questionnaire are due to dropout prior to the third questionnaire and this is likely to bias the results. However, for those patients who are alive and well enough to complete quality of life assessments, there is a statistically significant difference between CT and PAL in the MIC study in terms of 0- to 6-week change in *GQS* with patients improving on average on the CT arm and deteriorating on the PAL arm over this time (see Table 6.3). There was no evidence of difference between treatments in terms of 0- to 6-month change in *GHSS* in the ESPAC study (see Table 6.4).

For the SAUC, only those patients with a single assessment are excluded from the analysis. In the MIC study, there was evidence that the treatment groups differed in terms of this summary measure with CT patients having greater values and thus better quality of life over their follow-up time than PAL patients. In the ESPAC study there was no evidence of any difference between the treatment groups in terms of SAUC. Again these results are conditional on patients being alive and well enough to complete quality of life assessments.

Table 6.3: Summary measures analysis of *GQS* in the MIC study: number with summary measure, means, standard errors (SE), medians and interquartile (IQ) ranges for each summary measure

Summary Measure	Statistic	CT (N=67)	PAL (N=42)	P-value
0- to 6-week change	Number	52	33	
	Mean (SE)	2.02 (1.46)	-4.88 (2.43)	0.01 (Students t-test)
	Median (IQ range)	2.60 (-4.17, 7.67)	-2.75 (-10.98, 2.08)	0.005 (Wilcoxon)
SAUC	Number	66	36	
	Mean (SE)	81.23 (1.26)	76.14 (2.19)	0.03 (Students t-test)
	Median (IQ range)	82.81 (74.48, 89.89)	78.13 (65.09, 87.87)	0.07 (Wilcoxon)

Table 6.4: Summary measures analysis of *GHSS* in the ESPAC study: number with summary measure, means, standard errors (SE), medians and interquartile (IQ) ranges for each summary measure

Summary Measure	Statistic	CT (N=88)	NoCT (N=87)	P-value
0- to 6-month change	Number	50	43	
	Mean (SE)	14.83 (4.22)	14.15 (4.27)	0.91 (Students t-test)
	Median (IQ range)	16.67 (0, 33.33)	16.67 (0, 33.33)	0.90 (Wilcoxon)
SAUC	Number	78	78	
	Mean (SE)	63.45 (1.93)	62.12 (1.96)	0.63 (Students t-test)
	Median (IQ range)	63.36 (50, 78.44)	62.86 (51.47, 73.71)	0.77 (Wilcoxon)

6.4 Models For Longitudinal Data

6.4.1 Introduction

Exploratory data analysis (as described in 6.2) gives some insight into how quality of life changes over time and whether there are any salient differences between treatments with regards to these changes. In modelling the longitudinal quality of life data, the aim is to quantify these observed patterns. Modelling must account for the possible

correlation between the successive measures taken on each patient. There is a considerable literature on modelling longitudinal data (Diggle et al 1994, Hand and Crowder 1996, Lindsey 1994) and such methods are considered here in terms of their application to quality of life data.

Longitudinal quality of life studies usually give rise to data that are unbalanced in terms of the number and timing of assessments per subject, which in part will be due to missing data. The impact of different types of missing data mechanisms on longitudinal data analysis has been discussed (Laird 1988) and is discussed further in Section 6.7 but for the moment we only consider the problem it causes in terms of unbalanced data. Analysis of longitudinal quality of life data requires modelling techniques that capture the dynamic nature of the data and can cope with the unbalanced structure. Models for normally distributed continuous quality of life measures will be the focus in the following sections but consideration will be given to other types of measures in Section 6.7.

6.4.2 Covariance Matrix

In modelling the quality of life data it is necessary to specify a structure for the covariance matrix that is being assumed. Although the covariance matrix is not necessarily of direct interest, the structure chosen will affect the estimation of the regression coefficients in the model. The exploratory analysis of the observed autocorrelations in a dataset (see Section 6.2.3) may give some insight into the likely covariance structure that should be used in modelling that data.

Given a vector of data $\boldsymbol{Q} = (Q_{11}, Q_{12}, \ldots, Q_{1t}, Q_{21}, Q_{22}, \ldots, Q_{2t}, \ldots, Q_{n1}, Q_{n2}, \ldots, Q_{nt})$ of length nt where Q_{ij} represents the observed quality of life of patient i ($i = 1,\ldots,n$) at time point j ($j = 1,\ldots,t$). The covariance matrix for $\boldsymbol{Q}$ will be of size $nt \times nt$. Observations on different subjects are assumed to be independent and therefore the covariance matrix has a block diagonal structure with all off-diagonal blocks as zero matrices. The $t \times t$ blocks on the diagonal represent the individual covariance matrices each reflecting variation between the times within the individual. All individuals are assumed to have the same covariance matrix. In specifying the structure for the covariance matrix of $\boldsymbol{Q}$ therefore it is sufficient to specify the structure of the individual covariance matrix. The

possible structures are detailed elsewhere (Burton et al 1998) but some of the possible structures are summarised in Table 6.5. The simple structure assumes that all assessment times have the same variance (σ^2) and that there is no correlation between values at different assessment times. This is unlikely to be valid for repeated measures data. A compound symmetric structure assumes that there is a correlation between the values at different assessment times (ρ) but the correlation is the same whatever the distance between the assessments. The autoregressive structure allows the correlation to diminish as the distance between the assessments increases. The unstructured covariance matrix assumes different variability of measurements at different assessment times and does not assume any relationship between the correlations.

Table 6.5 Possible covariance structures for modelling longitudinal quality of life data

	$Var(Q_{ij})$ for patient i and time point j	$Cov(Q_{ij}, Q_{ik})$ for patient i and time points j, k ($j \neq k$)
Simple	σ^2	0
Compound Symmetric	σ^2	$\sigma^2\rho$
Autoregressive (order 1)	σ^2	$\sigma^2\rho^{\lvert j-k\rvert}$
Unstructured	σ_j^2	σ_{jk}

6.4.3 Repeated Measures Analysis of Variance Model

When the assessments relate to a small number (say ≤3) of fixed scheduled time points that are common to all patients and with rarely missing values then it may be appropriate to model the data using a repeated measures analysis of variance model (Diggle et al 1994). With this approach, time is treated as a categorical variable rather than as a continuous measure such that the quality of life for a patient is given by the mean for the combination of levels of treatment and time. The quality of life for patient i in treatment group g at scheduled time point j is given as follows:

$$Q_{ijg} = \mu + \alpha_g + \beta_j + \varepsilon_{ijg} \qquad [6.1]$$

where μ is the overall mean and α_g represents the effect of treatment group g, β_j represents the effect of time point j and ε_{ijg} represents remaining variation around this mean. The model assumes that ε_{ijg} are normally distributed with mean **0** and covariance matrix $\boldsymbol{\Sigma}$. The fact that the data are repeated measures can be accommodated by specifying an appropriate structure for the covariance matrix $\boldsymbol{\Sigma}$.

Missing data for any of the scheduled assessment times causes the data to be unbalanced. One strategy to enable the use of repeated measures analysis of variance in this situation is to undertake a complete case analysis (see Section 5.5.1), where only patients with complete data are used to compare treatments. Problems of reduced numbers for analysis and potential bias make this approach problematic and so an alternative strategy that allows all available data to be included is preferable for quality of life data (Zwinderman 1992). The GLM procedure in SAS will perform this type of analysis as long as the data consist of one record per patient per repeated measure (Zwinderman 1992). This model is also called a cell means model and has been applied to quality of life data using the MIXED procedure in SAS with a REPEATED statement included to specify the structure of the covariance matrix (Fairclough 2002). This approach however is still based on the assumption that missing data are missing at random and this will generally not be a valid assumption for quality of life data.

6.4.4 Random Effects Models

Many studies assess quality of life at four or more scheduled time points, as do the MIC and ESPAC studies considered in this thesis. In many cases assessments will be irregularly spaced and it may be more useful to model the data using the actual timing of the assessment rather than allocating the assessment to a scheduled time point. In general therefore it will be more useful to consider time as continuous and model quality of life over continuous time.

Measures of quality of life over time can be described by a polynomial function of time as follows:

$$Q_{ij} = \alpha_1 + \alpha_2 t_{ij} + \alpha_3 t_{ij}^2 + \ldots + \varepsilon_{ij} \qquad [6.2]$$

Q_{ij} is the measure of quality of life for the ith patient taken at the jth assessment and t_{ij} is the time that the jth assessment was taken for the ith patient in relation to some fixed time origin, usually date of entry to trial. In some cases, when assessments relate to scheduled time points the t_{ij} for a given j will be the same for all patients i.e. the t_{ij} can be replaced by t_j in the model. The regression parameter α_1 represents the mean quality of life at time 0 and the parameters α_2, α_3, ... represent the mean change over time for the population. The variation of the observation for the ith patient taken at the jth assessment around the mean is represented by ε_{ij}.

More complex polynomial functions are not always easy to interpret and hence where possible it is generally preferable to model the data using a linear function. If the change over time is not linear then it may be possible to transform either the quality of life measure or time scale to create a linear relationship (Carpenter et al 2002). Alternatively a piecewise linear model may be preferable to fitting a complex polynomial function (Fairclough 2002). Given cut points at increasing values of time given by $C_0=0>C_1>C_2>\ldots > C_m$ the piecewise linear model is given as:

$$Q_{ij} = \alpha_1 + \alpha_2 t_{ij}^{[1]} + \alpha_3 t_{ij}^{[2]} + \ldots + \alpha_{m+2} t_{ij}^{[m+1]} + \varepsilon_{ij} \qquad [6.3]$$

where for $x=1,\ldots,m$

$$t_{ij}^{[x]} = \max[(\min(t_{ij}, C_x) - C_{x-1}), 0] \text{ and } t_{ij}^{[m+1]} = \max(t_{ij}, C_m) - C_m$$

In this model, the data between each pair of cut points C_p and C_{p+1} are modelled as a linear function with slope given by α_{p+1}. This type of model has been applied to quality of life data using a slightly different but equivalent parameterisation (Fairclough 2002). The form of the model presented here provides direct estimates of the slopes for each piece of the model.

With repeated measures data, the random terms ε_{ij} are unlikely to be independent. The dependence can be modelled either by specifying an appropriate covariance structure or by extending the model to include random effects as well as fixed effects in a *mixed model*. Mixed models allow population average effects to be measured whilst allowing

for within-patient correlation. The model is made up of a fixed part and a random part, with the fixed part estimating the population change over time and the random part estimating the departure of individuals from the population average. The simplest model to fit is a linear random-effects model over time as below (Laird and Ware 1982).

$$Q_{ij} = \alpha_1 + \alpha_2 t_{ij} + \alpha_{1i} + \alpha_{2i} t_{ij} + \varepsilon_{ij} \qquad [6.4]$$

The subject-specific intercept and slope terms are assumed to have a joint normal distribution as follows:

$$\begin{pmatrix} \alpha_{1i} \\ \alpha_{2i} \end{pmatrix} \sim MVN\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix} \right) \qquad [6.5]$$

The covariance structure of the data for each individual is now defined by the covariance matrix for the random effects part of the model and the covariance matrix for the error term. Although other structures can be considered, the covariance matrix for the random effects will usually be of the unstructured form as specified in [6.5] where the variability between the individual intercepts is assumed different to the variability between the individual slopes and some correlation between the intercepts and slopes is expected. The unstructured option is the least restrictive in terms of assumptions about the structure of the data and generally is feasible to estimate with only 3 unknown parameters. With the autocorrelations now incorporated into the random effect part of the model, it will often be adequate to assume the errors ε_{ij} to be normally distributed with mean **0** and covariance matrix $\sigma_\varepsilon^2\mathbf{I}$. Alternative structures for the covariance matrix as specified in Table 6.5 can also be considered.

The errors not only represent measurement error but also the departure of the true curve from linearity. If a linear model does not seem appropriate, higher order terms for time, such as quadratic or cubic can be introduced but these terms should also include random effect terms. Alternatively a piecewise linear model as specified in [6.3] can be estimated with random effects for the time intervals.

The model can be extended to include other covariates either as random or fixed effects. Here the model in [6.4] is extended to include a covariate for treatment group G_i as a fixed effect. Within-subject variances and covariances are assumed to be the same for both treatment groups. Assuming mean quality of life follows a linear function of time, the effect of treatment can be investigated by including treatment as a covariate in the model together with a treatment by time interaction as follows:

$$Q_{ij} = \alpha_1 + \alpha_2 t_{ij} + \alpha_{1i} + \alpha_{2i} t_{ij} + \delta_1 G_i + \delta_2 (G_i \times t_{ij}) + \varepsilon_{ij} \qquad [6.6]$$

With treatment included as a fixed effect, the δ_1 regression coefficient measures the effect of treatment on the overall intercept of the linear function and δ_2 measures the effect of treatment on the overall slope. If this model is estimated for the quality of life data alone, any missing data are assumed to be missing at random and hence no account is made of informative dropout due to death.

Estimates of the parameters in a random effects model can be obtained using maximum likelihood methods. Maximum likelihood estimates can be computed using the EM algorithm (Laird and Ware 1982; see also Section 5.5.2) or using numerical techniques such as the Newton-Raphson Algorithm (Lindstrom and Bates 1988; see also Section 7.3). Restricted maximum likelihood estimation (REML) may give better parameter estimates than maximum likelihood when sample sizes are small and similar results if samples are large (Diggle et al 1994) and hence REML is used here. Most standard software is now able to estimate such models. The adequacy of different models can be assessed using formal model fitting techniques that compare nested models by assessing the change in minus twice log likelihood and compare non-nested models using model-fit criteria such as Akaike's Information Criterion (AIC). Parameter estimation can also be carried out using a Bayesian approach as described in Section 6.6.

6.5 Models for the MIC and ESPAC Data

6.5.1 Linear Random Effects Model for *GQS* in the MIC Study

As discussed in Section 4.4.1, although *GQS* is not strictly speaking a continuous measure the range of values it can take means that it could be considered as such. The distribution of *GQS* in general is not normally distributed (Figure 4.1a and Figure 6.8) but since the distributions are only slightly skewed and no transformations of the outcome appear to improve the normality of the distribution, *GQS* will be assumed to be normally distributed.

Examination of the individual profiles over time (Figure 6.1) indicates that although the responses for some individuals are far from linear and others only have one or two measures on which to base a model, on the whole a linear model is a reasonable model to consider for the change in quality of life over time and hence the values for *GQS* over time were modelled using a linear random effects model as specified in [6.6], i.e.

$$Q_{ij} = \alpha_1 + \alpha_2 t_{ij} + \alpha_{1i} + \alpha_{2i} t_{ij} + \delta_1 G_i + \delta_2 (G_i \times t_{ij}) + \varepsilon_{ij} \qquad [6.7]$$

where

$$\begin{pmatrix} \alpha_{1i} \\ \alpha_{2i} \end{pmatrix} \sim MVN\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix} \right) \text{ and } \varepsilon_{ij} \sim \mathrm{N}(0,\ \sigma_e^2)$$

The time and treatment covariates, t_{ij} and G_i, used in this model are centred round 0 to enable a direct comparison of results with those from the Bayesian approach in the next section. So G_i=0.5 represents the CT group and G_i=-0.5 represents the PAL group and the time origin is 6 weeks from trial entry.

The PROC MIXED procedure in SAS (Littell et al 2000, Singer 1997) is used to fit linear random effects models to the longitudinal quality of life data. SAS calculates restricted maximum likelihood estimates of the model parameters using the Newton-Raphson procedure. Wald tests are used to assess the statistical significance of the parameters.

Table 6.6 Parameter estimates for the linear random effects model for quality of life over time in the MIC study including treatment and treatment by time interaction

	Parameter estimates using REML	Standard error	p-value
α_1	78.12	1.14	<0.0001
α_2	-0.06	0.13	0.63
δ_1	5.71	2.28	0.01
δ_2	0.45	0.26	0.08
σ_1^2	111.88	18.40	<0.0001
σ_2^2	0.41	0.27	0.06
σ_{12}	1.93	1.53	0.21
σ_e^2	51.11	5.70	<0.0001

The intercept estimates, α_1 and δ_1, are not of direct interest but show that the mean *GQS* score at 6 weeks was 80.98 on CT and 75.26 on PAL and this difference was statistically significant. The estimate for α_2 suggests that the slope representing overall change in quality of life over time was not significantly different from zero but the estimate for δ_2 indicates that there was some evidence of borderline statistical significance that the change over time differed for the two treatment groups. Combining the estimates gives an estimated slope for the CT group of 0.17 compared to that of –0.29 for the PAL group. This suggests that on average the quality of life of the CT group improves over time whilst that for the PAL group deteriorates. Estimates of the variance and covariance show that there is a significant variability between the individual intercepts (i.e. value at 6 weeks) and some evidence of variability between the individual slopes. There is no evidence of a correlation between the individual intercepts and slopes.

6.5.2 Piecewise Linear Random Effects Model for *GHSS* in the ESPAC Study

As discussed in Section 4.4.2, *GHSS* is essentially a discrete variable with 13 distinct values. This number of values is deemed sufficient to consider modelling the variable as a continuous measure. The alternative options of treating it as an ordinal variable would ideally require reducing the measures to a manageable number of categories in which some of the detail in the data would be lost. The distribution of *GHSS* is not normally

distributed (Figure 4.3) but since no transformation of the outcome appears to improve the normality of the distribution, *GHSS* for the purposes of this section will be assumed to be normal.

Examination of the individual profiles over time (Figure 6.2) indicates that the responses are very variable and some individuals only have one or two measures on which to base a model. The group profiles indicate that a linear model is probably not appropriate. A quadratic model and a piecewise linear model with a change-point at 6 months were considered as alternatives and the piecewise linear model was found to provide the best fit to the data (AIC had smallest value). The values for *GHSS* over time were therefore modelled using a piecewise linear random effects model as follows:

$$Q_{ij} = \alpha_1 + \alpha_2 t_{ij}^{[1]} + \alpha_3 t_{ij}^{[2]} + \alpha_{1i} + \alpha_{2i} t_{ij}^{[1]} + \alpha_{3i} t_{ij}^{[2]} + \delta_1 G_i + \delta_2 (G_i \times t_{ij}^{[1]}) + \delta_3 (G_i \times t_{ij}^{[2]}) + \varepsilon_{ij} \quad [6.8]$$

where

$$t_{ij}^{[1]} = \min(t_{ij}, 6) \text{ and } t_{ij}^{[2]} = \max(t_{ij}, 6) - 6$$

and

$$\begin{pmatrix} \alpha_{1i} \\ \alpha_{2i} \\ \alpha_{3i} \end{pmatrix} \sim MVN\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_2^2 & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_3^2 \end{pmatrix} \right) \text{ and } \varepsilon_{ij} \sim \mathrm{N}(0, \sigma_e^2)$$

The treatment covariate G_i, used in this model is centred round 0 to enable a direct comparison of results with those from the Bayesian approach in the next section. So G_i=0.5 represents the CT group and G_i=-0.5 represents the NoCT group.

The PROC MIXED procedure in SAS (Littell et al 2000, Singer 1997) is used to fit the piecewise linear random effects models to the longitudinal values of *GHSS*. The Newton-Raphson procedure is used to calculate REML estimates for the parameters. Wald tests are used to assess significance of parameters.

Table 6.7 Parameter estimates for the piecewise linear random effects model for *GHSS* over time in the ESPAC study with change-point at 6 months and including treatment and treatment by time interaction

	Parameter estimates using REML	Standard error	p-value
α_1	51.05	1.98	<0.001
α_2	2.81	0.48	<0.001
α_3	-0.87	0.21	<0.001
δ_1	-1.01	3.97	0.800
δ_2	0.32	0.95	0.740
δ_3	0.50	0.42	0.238
σ_1^2	225.27	78.84	0.002
σ_2^2	11.18	4.38	0.005
σ_3^2	1.18	0.66	0.037
σ_{12}	-27.82	16.49	0.092
σ_{13}	-12.42	4.97	0.013
σ_{23}	1.71	1.07	0.109
σ_e^2	283.15	23.25	<0.001

The estimates of α_1 and δ_1 show that the mean value of *GHSS* at baseline (time=0) was approximately 51 and the intercepts for both treatment arms were equivalent (p=0.80). The estimates of α_2 and α_3 together with δ_1 and δ_2 show that on average *GHSS* significantly increases over the first 6 months indicating an initial improvement in quality of life but significantly decreases from 6 months onwards and there is no significant difference between the treatment groups in terms of this pattern over time. Estimates of the variances and covariances show that there is a significant variability between the individual intercepts and between the individual slopes in both the early and late phase. There is some evidence of a correlation between the individual intercepts and slopes in the post 6-month period.

6.6 Bayesian Analysis of Longitudinal Quality of Life Data

In the previous section, the parameters in the models for quality of life over time for the MIC and ESPAC studies were estimated using maximum likelihood methods. The same model parameters can be estimated using a Bayesian approach in WinBUGS. This requires specification of the likelihood and prior distributions for the parameters. For

the purposes of this thesis, vague prior distributions will be used, though in theory external prior evidence could be included.

6.6.1 Linear Random Effects Model for *GQS* in the MIC Study

To estimate the parameters in the linear random effects model specified in equation [6.7] for the MIC data from a Bayesian approach, we specify the likelihood as follows:

$$Q_{ij} \sim N(\mu_{ij},\ \sigma_e^2) \qquad [6.9]$$
$$\mu_{ij} = \theta_{1i} + \theta_{2i}t_{ij} + \delta_1 G_i + \delta_2 (G_i \times t_{ij})$$

The time and treatment covariates were centred to aid convergence. The individual intercepts and slopes were assumed to have a bivariate normal distribution with mean $\boldsymbol{\alpha}$ and covariance matrix $\boldsymbol{\Sigma}$ as follows:

$$\begin{pmatrix}\theta_{1i}\\ \theta_{2i}\end{pmatrix} \sim MVN\left(\begin{pmatrix}\alpha_1\\ \alpha_2\end{pmatrix}, \begin{pmatrix}\sigma_1^2 & \sigma_{12}\\ \sigma_{12} & \sigma_2^2\end{pmatrix}\right)$$

Prior distributions were specified for each of the unknown parameters. It was intended that vague prior distributions be used in all cases. For the fixed regression parameters, normal distributions with relatively large variances were used as follows:

$$\alpha_1,\ \alpha_2,\ \delta_1,\ \delta_2 \sim N(0,\ 10000)$$

The prior distribution for σ_e^2 was specified in terms of its inverse, representing the precision, which was allocated a Gamma prior distribution as follows:

$$\frac{1}{\sigma_e^2} \sim Gamma(0.001,\ 0.001)$$

The most difficult part is specifying a prior distribution for the covariance matrix $\mathbf{\Sigma}$. One option is to use a Wishart distribution (Spiegelhalter et al 1995), a multivariate extension of a chi-square distribution, as a prior for the precision matrix $\mathbf{\Sigma}^{-1}$:

$$\mathbf{\Sigma}^{-1} \sim \text{Wishart}(\mathbf{R}, \rho)$$

where $\mathbf{R}$ is an initial prior estimate at the magnitude of the covariance matrix and ρ is the number of degrees of freedom. The smallest possible value that ρ can take is the rank of $\mathbf{\Sigma}$, which in this case is 2, and this will represent vague prior knowledge and mean that the choice for $\mathbf{R}$ is not crucial. We chose $\mathbf{R}$ with 100 and 0.5 on the main diagonal and 0 on the off diagonal. The elements of the main diagonal were chosen so that the relative scale of the variance of the intercepts compared to slopes that was observed in the classical analysis was maintained here.

Because of the difficulties of specifying the parameters of a Wishart prior distribution, an alternative approach to specifying a prior distribution for $\mathbf{\Sigma}$ is to use a product normal formulation (Spiegelhalter et al 1995). In this approach the bivariate normal model is reformulated as two inter-related univariate normal models as follows:

$$\begin{aligned} \theta_{1i} &\sim N(\mu_1, \omega_1^2) \\ \theta_{2i} &\sim N(\mu_{2i}, \omega_2^2) \\ \text{where } \mu_{2i} &= \lambda_1 + \lambda_2 \theta_{1i} \end{aligned} \qquad [6.10]$$

Here the individual intercepts are assumed to have a univariate normal distribution and the individual slopes have a univariate normal distribution conditional on the value of the individual's intercept.

Vague prior distributions can be specified for the λ's and inverse ω^2's as follows:

$$\lambda_1, \lambda_2 \sim N(0, 10000)$$

$$\frac{1}{\omega_1^2}, \frac{1}{\omega_2^2} \sim Gamma(0.001, 0.001)$$

Gamma prior distributions of this form have been shown to work well as vague priors for precision parameters (Spiegelhalter et al 1995) but other options were also considered for the precision parameters as part of a sensitivity analysis. A uniform distribution over a relatively large interval for the standard deviations was considered i.e.

$$\omega_1, \omega_2, \sigma_e \sim Uniform(0, 10000)$$

A half normal distribution (i.e. truncated at zero) with a relatively large variance for the standard deviations was also considered i.e.

$$\omega_1, \omega_2, \sigma_e \sim N(0, 10000)I(0,)$$

The original parameters in the bivariate normal distribution can be obtained by back-transforming as follows:

$$\begin{aligned} \alpha_1 &= \mu_1 \\ \alpha_2 &= \lambda_1 + \lambda_2\mu_1 \\ \sigma_1^2 &= \omega_1^2 \\ \sigma_2^2 &= \omega_2^2 + \lambda_2^2\omega_1^2 \\ \sigma_{12} &= \lambda_2\omega_1^2 \end{aligned} \qquad [6.11]$$

Starting values for the parameters were based on results from the classical parameter estimates. The model allowed a 'burn-in' of 50,000 iterations and posterior estimates were based on 50,000 sampled values. WinBUGS code can be found in Appendix II as part of the joint model discussed in Chapter 10 and the results of fitting the different models are given in Table 6.8 and are compared to the classical estimates obtained in Section 6.5.1.

Table 6.8: Bayesian estimates of parameters (and standard errors) in linear random effects model for *GQS* over time in the MIC study with different vague prior distributions

	Classical (REML)	Bayesian	Bayesian – product normal formulation		
		Wishart prior on precision matrix	Gamma priors on precisions	Half normal priors on standard deviations	Uniform priors on standard deviations
α_1	78.12 (1.14)	78.10 (1.14)	78.11 (1.13)	78.11 (1.17)	78.12 (1.16)
α_2	-0.0635 (0.1312)	-0.0563 (0.1311)	-0.0521 (0.124)	-0.0681 (0.1404)	-0.0602 (0.1258)
δ_1	5.71 (2.28)	5.84 (2.27)	5.86 (2.22)	5.78 (2.36)	5.71 (2.31)
δ_2	0.4542 (0.2623)	0.46 (0.2573)	0.4587 (0.2504)	0.465 (0.2686)	0.4625 (0.2604)
σ_1^2	111.88 (18.40)	112.4 (18.80)	111.9 (19.14)	115.3 (19.19)	114.2 (19.02)
σ_2^2	0.4091 (0.2683)	0.3399 (0.2039)	0.2605 (0.2362)	0.4552 (0.2760)	0.3459 (0.2924)
σ_{12}	1.93 (1.53)	1.71 (1.48)	1.80 (1.55)	2.11 (1.58)	1.85 (1.55)
σ_e^2	51.11 (5.70)	53.35 (5.55)	54.88 (5.90)	52.45 (5.76)	54.01 (6.15)

All forms of prior distribution give broadly similar answers, which are comparable to the classical estimates. The conclusions from the Bayesian model are therefore the same as those given for the classical estimates in Section 6.5.1. Although the different forms of prior distribution gave comparable results, the model performance varied. The model with the Wishart prior distribution was the most efficient in terms of speed. There were problems with the Gamma and Uniform priors in that the estimate for the slope parameter and hence the variance and covariance parameters on a few occasions sampled repeatedly at around zero for a number of iterations. This could potentially lead to underestimates of the variability of these parameters.

6.6.2 Piecewise Linear Random Effects Model for *GHSS* in the ESPAC Study

To estimate the parameters in the piecewise linear random effects model specified in equation [6.8] for the ESPAC data from a Bayesian approach, we specify the likelihood as follows:

$$Q_{ij} \sim N(\mu_{ij}, \sigma_e^2)$$

$$\mu_{ij} = \theta_{1i} + \theta_{2i} t_{ij}^{[1]} + \theta_{3i} t_{ij}^{[2]} + \delta_1 G_i + \delta_2 (G_i \times t_{ij}^{[1]}) + \delta_3 (G_i \times t_{ij}^{[2]})$$

The treatment covariate G_i, used in this model is centred round 0 to aid convergence with $G_i = 0.5$ representing the CT group and $G_i = -0.5$ representing the NoCT group. The individual intercepts and slopes were assumed to have a bivariate normal distribution with mean $\boldsymbol{\alpha}$ and covariance matrix $\boldsymbol{\Sigma}$ as follows:

$$\begin{pmatrix} \theta_{1i} \\ \theta_{2i} \\ \theta_{3i} \end{pmatrix} \sim MVN\left(\begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_2^2 & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_3^2 \end{pmatrix} \right)$$

The prior distributions specified for each of the parameters in the model were very similar to those chosen for the linear random effects model in the MIC study. Vague prior distributions were used in all cases. For the fixed regression parameters, normal distributions with large variances were used as follows:

$$\alpha_1, \alpha_2, \alpha_3, \delta_1, \delta_2, \delta_3 \sim N(0, 10000)$$

The prior distribution for σ_e^2 was specified in terms of its inverse, representing the precision and the precision was allocated a Gamma prior as follows:

$$\frac{1}{\sigma_e^2} \sim Gamma(0.001, 0.001)$$

A Wishart distribution was specified as a prior distribution for the precision matrix $\boldsymbol{\Sigma}^{-1}$:

$$\boldsymbol{\Sigma}^{-1} \sim Wishart(\mathbf{R}, \rho)$$

where $\mathbf{R}$ was chosen with 200, 10 and 1 on the main diagonal and 0 on the off diagonal to maintain the relative scale of the variances that was observed in the classical analysis. The smallest possible value of 3 was chosen for ρ to represent vague prior knowledge

and means that the choice for **R** is not crucial. The product normal formulation could be considered as an alternative to the Wishart distribution although it would be more unmanageable with 3 random effects, but a sensitivity analysis on the Wishart prior distribution was preferred, as described below.

Starting values for the parameters were based on results from the classical parameter estimates. The model allowed a burn-in of 50,000 iterations and posterior estimates were based on 50,000 sampled values. WinBUGS code can be found in Appendix II as part of the joint model presented in Chapter 10 and the results of fitting the model are given in Table 6.9. Alternative versions of the model were also run in a sensitivity analysis. The base model (as specified above) was altered as follows: (i) **R** was defined with values of 10,10 and 10 on the main diagonal and 0 on the off-diagonal (ii) 200,000 values were sampled (iii) 200,000 iterations with every 10^{th} value sampled. These latter two options were included to overcome the problem of high autocorrelations observed in the base model.

Table 6.9: Bayesian estimates of parameters (and standard errors) in piecewise linear random effects model for *GHSS* over time in the ESPAC study with different vague prior distributions

	Classical (REML)	Bayesian			
		Base model	R with (10, 10, 10) on diagonal	200,000 sampled values	Sample every 10th value in 200,000
α_1	51.05 (1.98)	51.05 (1.20)	50.86 (1.91)	51.03 (1.97)	51.05 (1.97)
α_2	2.81 (0.48)	2.82 (0.49)	2.94 (0.48)	2.82 (0.48)	2.82 (0.48)
α_3	-0.87 (0.21)	-0.87 (0.28)	-1.04 (0.31)	-0.86 (0.27)	-0.87 (0.27)
δ_1	-1.01 (3.97)	-0.97 (3.94)	-0.90 (3.79)	-0.95 (3.95)	-0.96 (3.95)
δ_2	0.32 (0.95)	0.32 (0.94)	0.27 (0.94)	0.31 (0.94)	0.31 (0.94)
δ_3	0.50 (0.42)	0.47 (0.44)	0.53 (0.52)	0.48 (0.43)	0.48 (0.43)
σ_1^2	225.27 (78.84)	203 20 (69.86)	159.1 (86.37)	203.1 (69.03)	203.1 (68.99)
σ_2^2	11.18 (4.38)	9.75 (3.48)	9.01 (3.88)	9.71 (3.49)	9.70 (3.46)
σ_3^2	1.18 (0.66)	1.20 (0.57)	2.29 (0.84)	1.20 (0.57)	1.20 (0.57)
σ_{12}	-27.82 (16.49)	-22.83 (13.55)	-16.18 (16.03)	-22.7 (13.56)	-22.68 (13.5)
σ_{13}	-12.42 (4.97)	-10.14 (4.60)	-8.62 (5.40)	-10.36 (4.44)	-10.37 (4.45)
σ_{23}	1.71 (1.07)	1.52 (0.92)	0.82 (1.16)	1.55 (0.89)	1.55 (0.89)
σ_e^2	283.15 (23.25)	293.10 (22.84)	292.1 (25.97)	292.9 (22.47)	293.0 (22.43)

The Bayesian parameter estimates from the base model are comparable to the classical estimates, especially the fixed parameters. The conclusions therefore that were made in Section 6.5.2 apply here. The autocorrelations were high for all parameters in the base model. Sampling for 200,000 iterations rather than 50,000 did not greatly affect the results, nor did thinning the chain and therefore the sample of 50,000 values appeared to be sufficient to ensure that posterior distributions were adequately estimated. The model using the different estimate for **R** also gave reasonably comparable results to the base model, indicating that the results were not sensitive to this choice in relation to the prior distribution.

6.7 Discussion

In general, the aim of a quality of life study is to examine the change in quality of life over time and determine any differences between treatments. To achieve this, the observed quality of life data could be analysed using standard methods for longitudinal data analysis. This includes plotting patient and group profiles to give an initial impression of the nature of the data and highlight the possible trends over time and any differences between treatments. A summary measures analysis and more formal modelling when appropriate may follow this descriptive analysis.

The application of standard methods to longitudinal quality of life data however will be problematic because of missing data. This is due partly to the unbalanced data that results when missing data are present but more importantly because in general the missing data will not be missing at random. Standard methods of analysis all assume that any missing data are ignorable and thus application to quality of life data with informative dropout could give biased results. The standard methods actually provide a conditional analysis, that is they provide inferences on the quality of life of over time conditional on patients being alive and well enough to complete quality of life assessments. As discussed in Chapter 1, for an unbiased comparison of treatments, all patients at all time points should be included in any longitudinal analysis in an intention-to-treat type of approach.

The method of summary measures is attractive due to its simplicity and interpretability. The analysis, however, can be inefficient in highly unbalanced data when there is considerable variation between patients in terms of the number of repeated measures (Albert 1999). Summary measures may be difficult to calculate when informative dropout is present in the data, and the analysis in such situations may give biased results. It therefore may not be ideal for quality of life data. It may be possible to accommodate dropouts into the analysis by replacing all missing data with appropriate imputed values prior to analysis, but full allowance of corresponding uncertainty is ideally required. In particular, the worst value of quality of life could be assigned to all time points after death (for example Hollen et al 1997). One of the most widely used summary measures is area under the curve and the application of this summary measure

to the quality of life data over time in conjunction with the survival data is one of the simplest approaches to quality-adjusted survival analysis. For this type of area-under-the-curve analysis, quality of life must be measured on a 0 to 1 scale, where 0 represents quality of life equivalent to death and patients who dropout of the quality of life study due to death are included in the analysis by allocating them a quality of life score of zero for all time after death. This type of analysis is discussed further in Chapter 8.

Modelling the quality of life data over time will provide a more accurate analysis than the method of summary measures, as long as there is some underlying measurable trend to the data. Random effects models in the form of multi-level models have been used before to model longitudinal quality of life data (Beacon 1996, Beacon and Thompson 1996). They have the advantage that they can be extended to a further level to enable more than one quality of life outcome measure to be modelled at the same time (see Section 4.3.3). The application of these models however to observed quality of life data will generally give biased results as any missing data are assumed missing at random. As with the summary measures, one approach to dealing the problem of informative dropout may be to impute appropriate values of quality of life for all time points after dropout prior to analysis. An alternative and potentially more appealing approach would be to combine the model for quality of life over time with the probability of survival in a quality-adjusted survival analysis, as demonstrated in Chapter 8. Alternatively by jointly modelling the quality of life and survival data, the informative dropout due to death is accounted for in the estimation of the parameters in the random effects model for quality of life over time. This approach is discussed further in Chapter 10.

The models considered here are simplistic in the sense that they only include a treatment covariate. Other covariates such as age and performance status for example could influence quality of life over time and the models considered here could be extended to include these additional baseline covariates. From a classical viewpoint, the statistical significance of different covariates could be assessed using formal model fitting techniques that compare nested models by assessing the change in minus twice log likelihood. From a Bayesian viewpoint, different models could be compared using Bayes factors or model averaging (Spiegelhalter et al 2003). Bayesian estimation of parameters in the models for longitudinal quality of life data may be of interest in its

own right, but the main reason for including a Bayesian approach to parameter estimation here is in preparation for the joint modelling carried out in Chapter 10.

The models considered in this thesis assume the quality of life outcome is normally distributed. If the distribution of a continuous measure is not normal then it may be possible to normalise it using an appropriate transformation such as log or square root and then model the transformed variable instead of the original. If the measure is an ordinal variable with a large number of levels then the assumption of normality may not be unreasonable, however as the number of categories decreases so the assumption becomes more untenable.

Approaches for non-normally distributed measures are based on the theory of generalised linear models (Dobson 1990) and allow the analysis of binary, ordinal or continuous response variables. As such, the expected response is related to the explanatory variables in the model via a 'link' function. With normally distributed outcomes the model directly predicts the expectation of the response variable and this is a generalised linear model with an identity link. For non-normally distributed outcomes, the model can be defined to predict a *function* of the expectation of the response where the function, known as the link function, depends on the type of response variable. For example, a logit or probit link may be appropriate for binary responses, cumulative logit link for ordinal responses and log link for counts. This type of model could be considered for quality of life outcomes, for which the assumption of normality is not tenable, by specifying an appropriate link function and modelling a function of the expectation of the variable rather than the expectation itself. These types of models have been applied to quality of life outcomes (Beacon 1996).

The standard methods of analysis applied to the MIC quality of life data showed that in terms of *GQS* there may be a small benefit for chemotherapy compared to standard palliative care. The summary measures analysis showed that on average in terms of 0- to 6-week change, patients on CT improve whilst those on PAL deteriorate, conditional on patients being alive and well enough to complete an assessment at the 6-week time point. This difference was statistically significant (p=0.005). In addition, the standardised area under the curve analysis demonstrated that the mean quality of life score for those alive and well enough to complete quality of life assessments was

greater on CT compared to PAL and this was of borderline statistical significance (p=0.07). Although there was a large variation in individual patient profiles, the overall trend in quality of life over the 15-week study period was approximately linear with patients on CT improving on average from baseline and those on PAL deteriorating. The difference in linear change over time between treatments was of borderline statistical significance (p=0.08). Again this is conditional on patients being alive and well enough to complete assessments.

The standard methods of analysis applied to the ESPAC quality of life data showed that in terms of *GHSS* there did not appear to be any difference between CT and NoCT in terms of quality of life over time. The summary measures analysis showed that on average, in terms of 0- to 6-month change, patients on both treatment arms improve conditional on patients being alive and well enough to complete an assessment at the 6-month time point. There was no difference between the treatment arms in terms of the level of improvement. In addition, the area under the curve analysis demonstrated that the mean quality of life score for those alive and well enough to complete quality of life assessments was equivalent on both treatment arms. The individual patient profiles were very noisy and the data are not ideal for modelling. The group profiles indicated that a piecewise linear model with one cut-point at 6 months from surgery may be an appropriate model. This model showed that, conditional on patients being alive and well enough to complete assessments, quality of life on average initially improves in both treatment groups up to 6-months from surgery and then deteriorates from that point. There was no statistically significant difference between treatments in terms of these trends over time (p=0.74 for interaction of treatment with initial slope and p=0.24 for interaction of treatment with second slope). The non-normality of the *GHSS* outcome could be accommodated by grouping values into a limited number of categories and using a proportional odds model to model the change in scores over time.

In summary, the application of standard longitudinal methods of analysis to quality of life data provides an initial insight into the data and provides inferences on quality of life over time conditional on patients being alive and well enough to participate in the study. For an unbiased comparison of treatments, the analysis of quality of life should be based on an intention-to-treat principle which means including in the analysis all patients at all points in time of the quality of life study period, but informative dropout

often precludes this in practice. The simultaneous analysis of quality of life and survival will enable this and therefore is the recommended approach for the analysis of longitudinal quality of life data with informative dropout.

CHAPTER 7: MODELLING SURVIVAL DATA

7.1 Introduction

The aim of this chapter is describe the models that can be considered for survival data. The models will first include treatment as a fixed covariate and then be extended to include quality of life as a time dependent covariate. The models considered in this chapter are relevant for all three approaches that simultaneously analyse quality of life and survival.

The variation in survival data can be explained by means of a regression model. Variables that could potentially affect the survival time of a patient can be included as covariates in the model. The key covariate of interest will usually be treatment. In addition to treatment, covariates may include demographic variables such as age and sex, physiological variables such as white blood cell count or disease-related variables such as type of tumour or stage of disease at entry to study. Covariates that keep the same value for the duration of the study are called fixed, whilst those with possibly changing values over time are called time-dependent (Collett 1994). In this chapter, as well as being interested in the effect of treatment on survival as a fixed covariate, the inclusion of quality of life as a time-dependent covariate in the model enables the association between survival and the changing values of quality of life over time to be examined.

There are two types of regression model that can be considered. The first, which is generally referred to as a Cox regression model or more specifically a Cox proportional hazards regression model, models the hazard function (Cox 1972) whilst the second, generally called an accelerated failure time model, models the survivor function (Collett 1994). In general, Cox regression models are used when no distributional form is assumed for the survival data and accelerated failure time models are used when the data take on a specific distributional form, although either can be used in both situations. For the exponential and Weibull distributions, there is a direct link between the parameters of the two types of model. For the purposes of fitting the models to the data, it will sometimes be necessary to consider the models in the form of an accelerated

failure time model but in all cases, results will be reported in terms of the Cox regression model parameters.

There is much literature describing methods for modelling survival data (Collett 1994, Parmar and Machin 1995, Hosmer and Lemeshow 1999). This chapter summarises those methods that will be relevant in analysing quality of life and survival data simultaneously and thus provides a background to later chapters in this thesis. Modelling the survival data gives insight into the survival process in the MIC and ESPAC studies in preparation for analysing it simultaneously with quality of life. For this reason, the analysis is carried out only on the patients included in the quality of life study and only for the quality of life study period. Within these restrictions, the survival analysis will estimate the effect of treatment on survival and will also explore the relationship between quality of life and survival. Consideration is given in the joint modelling presented in Chapter 10 to extending the method to allow the inclusion of survival data from all trial patients. Estimation is considered from both a classical and Bayesian point of view. There is some literature discussing the Bayesian approach to survival analysis (Abrams 1992b, Abrams 1998, Ibrahim et al 2001) but these have not been widely applied to clinical trial data.

The chapter starts in Section 7.2 by describing the Cox regression model and considering the different forms that the baseline hazard can take. Classical estimation of the model parameters is described in Section 7.3 and these methods are applied to a simple model for the survival data in the MIC and ESPAC studies with just treatment as a fixed covariate in Section 7.4. The model is extended in Section 7.5 to include quality of life as time-dependent covariate. The alternative Bayesian approach to the estimation of the parameters in this model is given in Section 7.6. The problems related to including quality of life as a time-dependent covariate are discussed in Section 7.7, highlighting the need for a joint modelling approach, which is considered in Chapter 10.

7.2 Cox Regression Model

7.2.1 Basic Form of the Cox Regression Model

The Cox regression model (Cox 1972) allows the variation in survival, as expressed by the hazard function, to be explained by certain covariates. The hazard function for patient i, $h_i(t)$ is modelled as:

$$h_i(t) = h_0(t)\exp(\underline{\beta}'\underline{x}_i) \qquad [7.1]$$

where $h_0(t)$ is the underlying baseline hazard function, $\underline{x}_i$ is a vector of fixed covariates for patient i and $\underline{\beta}$ is a vector of regression coefficients. In this thesis, a Cox regression model always refers to the form of the model given by Cox (1972) as specified in [7.1] where the baseline hazard can be either arbitrary or parametric.

This model is based on the assumption that the hazard for different patient groups defined by the covariate values are proportional and the ratio of the hazards is approximately constant over time and hence this model is often called a proportional hazards regression model. The Cox regression model can be extended to incorporate time-dependent covariates (see Section 7.5).

7.2.2 Alternative Formulation of Cox Regression Model

An alternative formulation of the Cox model involves the consideration of the data in terms of a counting process (Andersen and Borgan 1985, Andersen et al 1993, Abrams 1998). Each patient i has a counting process $N_i(t)$ counting the number of events that occur up to time t and an at-risk process $Y_i(t)$ which is 1 if patient i is at risk of the event at time t and 0 otherwise. The rate of change of the counting process is known as the intensity process $\alpha_i(t)$ and this can modelled using a multiplicative intensity model which has the following form:

$$\alpha_i(t) = Y_i(t)\alpha_0(t)\exp(\underline{\beta}'\underline{x}_i)$$

where $\alpha_0(t)$ is the baseline intensity process and $\underline{\beta}$ is a vector of the regression parameters measuring the effect of the covariates $\underline{x}_i$ on the baseline intensity. The inclusion of the indicator variable $Y_i(t)$ in the model ensures that patients only contribute to the likelihood when they are at risk of the event.

If the event is death, then the intensity process is equivalent to the hazard function and the multiplicative intensity model is equivalent to a Cox proportional hazards model. So in counting process notation we can re-write the model specified in [7.1] as:

$$h_i(t) = Y_i(t)h_0(t)\exp(\underline{\beta}'\underline{x}_i) \qquad [7.2]$$

Considering the Cox model in this counting process formulation enables more complex scenarios such as multiple events per patient and patients moving in and out of different health states to be modelled using a reasonably straightforward and coherent approach.

7.2.3 Form for the Baseline Hazard

In general, the model does not assume any particular form of probability distribution for the survival times and thus the underlying baseline hazard is allowed to be *arbitrary*. For this reason, the Cox model is often referred to as a semi-parametric model. In some circumstances, however, if survival times follow a particular distribution then it may be more efficient to assume a parametric form for the underlying baseline hazard. Assuming a parametric form for the baseline hazard has the advantage of being able to report an estimate of the hazard of death as well as the effect of covariates on this hazard and enables prediction when required. In this thesis, as well as allowing the baseline hazard in the Cox regression model to be completely arbitrary, three other forms are considered; piecewise exponential, exponential and Weibull. Although the piecewise exponential model is a parametric form, it has close links with the non-parametric form for the baseline hazard, as described in 7.2.5.

7.2.4 Exponential and Weibull Models

If survival times are assumed to follow an exponential distribution then the baseline hazard function is constant over time (see Section 3.5) and the hazard for patient *i* at any point in time given by the Cox regression model is:

$$h_i = \lambda \exp(\underline{\beta}' \underline{x}_i) \qquad [7.3]$$

If survival times are assumed to follow a Weibull distribution then the baseline hazard function has a known form (see Section 3.5) and the Cox regression model becomes:

$$h_i(t) = \lambda \gamma t^{\gamma-1} \exp(\underline{\beta}' \underline{x}_i) \qquad [7.4]$$

In the parametric modelling of survival times T, the accelerated failure time model for patient *i* has the general form:

$$\log(T_i) = \mu + \underline{\alpha}' \underline{x}_i + \sigma \varepsilon_i \qquad [7.5]$$

where $\underline{x}_i$ is a vector of covariates, $\underline{\alpha}$ is a vector of unknown regression coefficients, σ is an unknown scale parameter, μ is an unknown intercept parameter and ε_i is an error term where errors are assumed to follow a particular distribution. If survival times are assumed to follow a Weibull distribution then there is a direct correspondence between the parameters under an accelerated failure time model (μ, σ and $\underline{\alpha}$ in [7.5]) and those under a proportional hazards model (λ, γ and $\underline{\beta}$ in [7.4]). The parameter estimates can be transformed using the following formulae (Collett 1994):

$$\hat{\lambda} = \exp\left(\frac{-\hat{\mu}}{\hat{\sigma}}\right), \qquad \hat{\gamma} = \frac{1}{\hat{\sigma}}, \qquad \underline{\hat{\beta}} = \left(\frac{-1}{\hat{\sigma}}\right)\underline{\hat{\alpha}} \qquad [7.6]$$

Using the Taylor series approximation to the variance of a function of random variables (Collett 1994), the estimates of the variance of $\hat{\lambda}$, $\hat{\gamma}$ and the *k*th element of the vector of regression coefficients $\hat{\beta}_k$, are given by:

$$Var(\hat{\lambda}) \approx \frac{\hat{\lambda}^2}{\hat{\sigma}^4}\left(\hat{\sigma}^2 Var(\hat{\mu}) + \hat{\mu}^2 Var(\hat{\sigma}) - 2\hat{\mu}\hat{\sigma}Cov(\hat{\mu},\hat{\sigma})\right)$$

$$Var(\hat{\gamma}) \approx \frac{1}{\hat{\sigma}^4}\left(Var(\hat{\sigma})\right) \qquad [7.7]$$

$$Var(\hat{\beta}_k) \approx \frac{1}{\hat{\sigma}^4}\left(\hat{\sigma}^2 Var(\hat{\alpha}_k) + \hat{\alpha}_k^2 Var(\hat{\sigma}) - 2\hat{\alpha}_k\hat{\sigma}Cov(\hat{\alpha}_k,\hat{\sigma})\right)$$

For an exponential distribution, all of the above follows with $\sigma = 1$ and thus the conversion is as follows:

$$\hat{\lambda} = \exp(-\hat{\mu}), \quad Var(\hat{\lambda}) = \hat{\lambda}^2 Var(\hat{\mu}), \quad \hat{\beta}_k = -\hat{\alpha}_k, \quad Var(\hat{\beta}_k) = Var(\hat{\alpha}_k) \qquad [7.8]$$

7.2.5 Piecewise Exponential Model

Instead of allowing the baseline hazard in the Cox regression model to be completely arbitrary, it may be preferable to split time into a number of intervals and assume that the baseline hazard varies for different time intervals but is constant within each time interval, i.e. survival times follow different exponential distributions within each time interval.

Suppose the time scale is divided into J intervals $(0,a_1]$, $(a_1, a_2]$, ..., (a_{J-1},∞) where the jth interval is given by $(a_{j-1}, a_j]$. These time intervals can be defined in numerous ways and whilst they do not have to be the same length, they should be chosen so that the assumption of constant hazard of death within each interval is not unreasonable. The choice should also be such that ideally there are approximately equal numbers of deaths in each interval and at least one death in each time interval is needed for adequate modelling of the baseline hazard in that time interval. The choice of time intervals is arbitrary and sensitivity analysis should be used to assess the effect of the choice of divisions on the results. As the time intervals become narrower, so the results from the piecewise exponential model will approach the results from the fitting the Cox model with an arbitrary baseline hazard. In the limit, where the a_1, a_2, ..., a_{J-1} are the distinct death times of the individuals, the piecewise exponential model is equivalent to the Cox model with arbitrary baseline hazard.

Given the J time intervals, the hazard of death for patient i within the jth time interval is given by:

$$h_{ij} = \lambda_j \exp(\underline{\beta}^T \underline{x}_i) \qquad [7.9]$$

If the model is estimated within an accelerated failure time model framework then:

$$\log(T_i) = \mu + \mu_j + \underline{\alpha}' \underline{x}_i + \varepsilon_i \qquad [7.10]$$

then the parameters in the accelerated failure time model (μ, μ_j, $\underline{\alpha}$ in [7.10]) can be transformed to the parameters in the Cox model (λ_j, $\underline{\beta}$ in [7.9]) as follows:

$$\begin{aligned} &\hat{\lambda}_j = \exp(-\hat{\mu} - \hat{\mu}_j), \\ &Var(\hat{\lambda}_j) \approx \hat{\lambda}_j^2 \left(Var(\hat{\mu}) + Var(\hat{\mu}_j) + 2Cov(\hat{\mu}, \hat{\mu}_j)\right) \qquad [7.11] \\ &\hat{\beta}_k = -\hat{\alpha}_k, \\ &Var(\hat{\beta}_k) = Var(\hat{\alpha}_k) \end{aligned}$$

7.3 Classical Estimation of Parameters in Survival Models

The regression parameters, $\underline{\beta}$, and the underlying baseline hazard function are estimated using maximum likelihood estimation. Maximum likelihood estimators are consistent, asymptotically efficient and asymptotically normal. The likelihood function is given by the probability of the observed data under the model (see Section 2.2). Let t_i represent the survival time for patient i and d_i represent whether that survival time is uncensored (d_i =1) or censored (d_i =0). Let $f(t)$ represent the probability density function for the survival times and $S(t)$ represent the associated survivor function. For uncensored data, the probability of an observed survival time t_i is given by $f(t_i)$ and for censored data, the probability of an observed survival time t_i is given by $S(t_i)$, the probability of surviving to time t_i. The full likelihood function for a sample of n independent observations (t_i, d_i) with i=1 to n, is therefore given by

$$L = \prod_{i=1}^{n} \left([f(t_i)]^{d_i} \times [S(t_i)]^{1-d_i} \right) \qquad [7.12]$$

Using the relationship between the hazard function and the probability density and survivor function (see [3.2]), [7.12] can be written as

$$L = \prod_{i=1}^{n} \left([h(t_i)S(t_i)]^{d_i} \times [S(t_i)]^{1-d_i} \right)$$

i.e. $$L = \prod_{i=1}^{n} \left([h(t_i)]^{d_i} \times S(t_i) \right)$$

From the relationship given in [3.2],

$$S(t) = \exp\left(- \int_{u=0}^{t} h(u)du \right)$$

and hence under the assumption of a Cox model

$$S(t) = \left(S_0(t)\right)^{\exp(\beta x)}$$

Thus, under the assumption of a Cox model the likelihood function is

$$L = \prod_{i=1}^{n} \left([h_0(t_i)\exp(\beta x_i)]^{d_i} \times [S_0(t_i)]^{\exp(\beta x_i)} \right)$$

and the log-likelihood is given by

$$\ln(L) = \sum_{i=1}^{n} \left(d_i \ln[h_0(t_i)] + d_i \beta x_i + \exp(\beta x_i) \ln[S_0(t_i)] \right) \qquad [7.13]$$

If the survival times follow a particular distribution such as exponential or Weibull then the hazard and survivor functions for these distributions (see Section 3.5) can be substituted into the log likelihood in [7.13] and this can then be maximised with respect

to all the parameters. If no distributional form is being assumed for the survival data then it is not possible to maximise the full likelihood.

As an alternative to the full likelihood, Cox (1972) proposed using a function called the *partial likelihood*. The partial likelihood is the likelihood of just the observed events. The likelihood of the death that occurs to patient i at time t_i is given by:

$$L_i = \frac{h_i(t_i)}{\sum_{j \in R(t_i)} h_j(t_i)}$$

where $R(t_i)$ represents the set of patients who are at risk of death at time t_i and includes patient i.

Under the Cox regression model and with the arbitrary baseline hazard cancelling out in both the numerator and denominator, this likelihood becomes:

$$L_i = \frac{\exp(\underline{\beta}'\underline{x}_i)}{\sum_{j \in R(t_i)} \exp(\underline{\beta}'\underline{x}_j)}$$

The partial likelihood function is the product of the likelihoods for the events and is given by:

$$PL(\underline{\beta}) = \prod_{i=1}^{n} \left(\frac{\exp(\underline{\beta}'\underline{x}_i)}{\sum_{j \in R(t_i)} \exp(\underline{\beta}'\underline{x}_j)} \right)^{d_i} \qquad [7.14]$$

Since this partial likelihood function contains only the $\underline{\beta}$ regression parameters, the distribution of the survival times can be ignored and the function can be maximised with respect to $\underline{\beta}$. Values of $\underline{\beta}$ that maximise the partial likelihood would also maximise the full likelihood. These maximum partial likelihood estimates are consistent and asymptotically normal but not fully efficient; the standard errors are slightly larger than they would be under the full likelihood function (Allison 1995, Cantor 1997).

The partial likelihood given in [7.14] assumes that there are no tied values amongst the observed survival times, which in general will not be the case. An exact expression for the partial likelihood with tied data has been suggested (Kalbfleisch and Prentice 1980) and although some standard software may provide an option of using this, it is not easily computed and approximations may be preferable (Breslow 1974, Efron 1977). The approximation advocated by Breslow is used here.

Determining the values of the parameters that maximise the full or partial likelihood function usually requires an iterative numerical method. Here we use the Newton-Raphson algorithm as follows (Collett 1994). If $\boldsymbol{U}(\beta)$ is the $p\times 1$ vector of first derivatives of the log likelihood with respect to β (the vector of efficient scores) and $\boldsymbol{I}(\beta)$ is the $p\times p$ matrix of negative second derivatives of the log likelihood evaluated at β (the Hessian matrix or observed information matrix) with inverse $\boldsymbol{I}^{-1}(\beta)$ then at the $(j+1)$th iteration:

$$\beta_{j+1} = \beta_j + \boldsymbol{I}^{-1}(\beta_j)\, \boldsymbol{U}(\beta_j)$$

Starting values for β are taken as least squares estimates with censored treated as uncensored. Iterations are repeated until estimates for β converge. Once values for β have been determined, the square root of the diagonal elements of the information matrix will give estimates of the standard errors for the parameter estimates.

The Wald statistic is calculated as $\left(\hat{\beta}/SE(\hat{\beta})\right)^2$ and by comparing this value to a chi-squared distribution on 1 degree of freedom it can be used to assess whether the estimate is statistically significantly different from zero.

7.4 Survival Models with Treatment as Fixed Covariate

7.4.1 Method

The simple Cox regression model that assesses the effect of the single fixed covariate treatment group on survival is as follows:

$$h_i(t) = h_0(t)\exp(\beta G_i) \qquad [7.15]$$

where G_i is the treatment group for patient i.

This model is applied to both the MIC and ESPAC survival data under the assumption that the baseline hazard is either arbitrary, exponential, Weibull or piecewise exponential. A classical approach to parameter estimation is used in this section.

The PHREG procedure in SAS (SAS Institute Inc 1992) was used to fit the model with an arbitrary baseline hazard to the data. The data for analysis consists of one line per patient with the survival time, an indicator variable to denote if the survival time is censored or not and the value of the treatment group covariate. SAS uses a Newton-Raphson iterative algorithm to estimate the value of β that maximises the partial likelihood as described in Section 7.3.

The LIFEREG procedure in SAS (SAS Institute Inc 1989) was used to fit the models with exponential and Weibull baseline hazards to the data using the DIST=EXPONENTIAL and DIST=WEIBULL option in the MODEL statement. The format for the data is the same as for the PHREG procedure. SAS models the survival time as an accelerated failure time model rather than modelling the hazard as in a Cox model and therefore the regression coefficients given by SAS were transformed so that they correspond to the regression coefficients for a Cox model by using the formulae given in [7.6], [7.7] and [7.8]. SAS uses the Newton-Raphson algorithm to estimate the values of the parameters that maximise the full likelihood given in [7.13].

The piecewise exponential model required some data manipulation before it could be applied to the data. The first step was to decide on the time intervals for analysis. Then for each subject one record was created for every time interval during which the subject was at risk of death. A variable representing the time interval to which each record related was included in the data. Time was reset to 0 at the beginning of each time interval. If a subject did not die in the interval then survival time variable is assigned the full length of the time interval and censoring indicator is set to 0. If a subject died within the interval then the survival time variable was assigned the length of time from

the start of the interval to death and the censoring indicator was set to 1. If a subject was last observed alive within the time interval then the survival time variable was assigned the length of time from the start of the interval to the last observation time and the censoring indicator was set to 0. The value of the fixed treatment covariate was replicated for each interval. The LIFEREG procedure in SAS with the DIST=EXPONENTIAL option in the MODEL statement was then used to fit the piecewise exponential model to the data. The variable representing the time interval was declared as a categorical variable and included in the model as a covariate. SAS sets up an indicator variable for each interval, the value for the last interval is set to zero and all other intervals are compared to this. The hypothesis of no difference between the hazards for the intervals was assessed using the Wald test for the overall interval variable. The regression coefficients for the accelerated failure time model given by SAS were translated into the regression coefficients for a Cox model using [7.11].

7.4.2 Application to the MIC Study

The analysis focuses on the difference in survival between the two treatment arms within the first 15 weeks from entry to trial, the time during which quality of life data is collected. There were 28 deaths (12 on CT and 16 on PAL) occurring at 24 distinct times during this period and the remainder of patients had full follow-up and are therefore censored at 15 weeks in the analysis. In Chapter 3, the Kaplan-Meier survival curves (see Figure 3.2b) suggested that the CT arm had superior survival and a log-rank test showed the difference to be statistically significant (p=0.02). The aim here is to estimate the effect of treatment on survival within the first 15 weeks from entry using the Cox regression model specified in [7.15] where $G_i = 0.5$ if patient i is in the CT arm or $G_i = -0.5$ for the PAL arm. These values were chosen for the covariate in order to correspond to the Bayesian analysis in which the covariate is centred round zero to aid convergence.

In fitting a piecewise exponential model, time intervals had to be chosen in accordance with the requirements outlined in 7.2.5. One option was to use the most extreme form of the model with intervals defined by the 24 distinct death times, which should give equivalent results to the model with arbitrary baseline hazard. Alternatively it seemed clinically sensible to choose 3-weekly time intervals as this corresponds to the timing of

cycles of chemotherapy and of quality of life assessments. Estimates of the hazard function are shown in Figure 3.4 and it does not seem unreasonable to assume constant hazard within 3-weekly intervals from study entry. Since the first death did not occur until just after 3 weeks, the first time interval needed to be greater than 3 weeks. Thus time intervals of (0,6], (6,9], (9,12] and (12,15] were used for the piecewise exponential model. The maximum likelihood estimates (and associated standard errors) for the parameters in each of the models are given in Table 7.1.

Table 7.1: Maximum likelihood estimates (and standard errors) for survival models with treatment as fixed covariate in the MIC study

Form of baseline hazard	Parameter	MLE (SE)
Arbitrary	β	-0.8796 (0.3822)
Arbitrary (alternative*)	β	-0.8799 (0.3822)
Exponential	λ	0.0196 (0.0037)
	β	-0.8376 (0.3819)
Weibull	λ	0.0022 (0.0020)
	γ	1.8249 (0.3294)
	β	-0.8792 (0.3821)
Piecewise Exponential	$\lambda_{0\text{-}6}$	0.0126 (0.0045)
	$\lambda_{6\text{-}9}$	0.0211 (0.0086)
	$\lambda_{9\text{-}12}$	0.0258 (0.0098)
	$\lambda_{12\text{-}15}$	0.0292 (0.0111)
	β	-0.8640 (0.3822)

*** fit as a piecewise exponential with 24 time intervals defined by death times**

The estimate of β under an arbitrary baseline hazard shows that chemotherapy significantly reduces the risk of death (p=0.02). The hazard ratio is estimated as 0.41 (95% CI: 0.20 to 0.88). The results from the piecewise exponential version of this model are virtually identical. The estimates of β under an exponential, Weibull and piecewise exponential baseline hazard are similar to this with the exponential being the most different. The exponential model estimates the constant hazard over time as 0.0126 in the CT group and 0.0304 in the PAL group, which are comparable to the estimates of the hazard calculated in Section 3.4.2. The estimates of the constant hazard in each time interval of the piecewise exponential model shows the baseline hazard increasing over time, although the Wald test for the time interval variable showed that there was no statistically significant difference between the hazards for the four intervals (p=0.37). The estimate of the shape parameter γ for the Weibull model is

significantly greater than 1 indicating that the hazard increases over time, and therefore an exponential distribution appears inappropriate.

7.4.3 Application to the ESPAC Study

Simple survival analysis of the ESPAC study in Chapter 3 suggested that the CT arm had superior survival to the NoCT arm (Figure 3.3; log-rank test p=0.02). The aim here is to estimate the effect of treatment on survival using the Cox regression model specified in [7.15] where, similar to the MIC study, $G_i = 0.5$ if patient i is in the CT arm or $G_i = -0.5$ for the NoCT arm. The maximum follow-up time is approximately 54 months and 126 deaths had occurred at the time of analysis. Since the numbers included in the analysis reduce to 30 and 18 in the CT and NoCT groups respectively by 24 months, the analysis is repeated for survival within 24 months as well as for the full follow-up period. There were 105 deaths within 24 months from surgery and the 48 patients still alive and on follow-up at 24 months were censored at this time. Three-monthly time intervals were chosen for the piecewise exponential model i.e. (0,3], (3,6],...,(21,24], (24,27], (27,∞) as this tied in with the timing of the quality of life questionnaires and were such that at least one death happens in each interval. Estimates of the hazard rates on the two treatment arms are shown in Figure 3.5c and the assumption of constant hazard within 3-monthly time intervals does not seem unreasonable. The number of deaths occurring after 27 months (6 in CT group and 3 in NoCT group) was too few for any sensible time intervals after this time. The parameter estimates for each of the models are given in Table 7.2. Note that the alternative form of the model with arbitrary baseline hazard that was used for the MIC study was not feasible here as the number of distinct death times was too large for a manageable form of the piecewise exponential model.

Table 7.2: Maximum likelihood estimates (and standard errors) for survival models with treatment as fixed covariate in the ESPAC study

Form of baseline hazard	Parameter	Over Whole FU	Within 24 Montbs
Arbitrary	β	-0.4180 (0.1807)	-0.3925 (0.1971)
Exponential	λ	0.0402 (0.0036)	0.0393 (0.0039)
	β	-0.3873 (0.1785)	-0.3502 (0.1963)
Weibull	λ	0.0107 (0.0036)	0.0058 (0.0024)
	γ	1.4221 (0.1000)	1.6640 (0.1392)
	β	-0.5074 (0.1807)	-0.4363 (0.1969)
Piecewise Exponential	$\lambda_{0\text{-}3}$	0.0037 (0.0026)	0.0037 (0.0026)
	$\lambda_{3\text{-}6}$	0.0252 (0.0070)	0.0252 (0.0070)
	$\lambda_{6\text{-}9}$	0.0541 (0.0111)	0.0542 (0.0111)
	$\lambda_{9\text{-}12}$	0.0691 (0.0141)	0.0692 (0.0142)
	$\lambda_{12\text{-}15}$	0.0536 (0.0138)	0.0536 (0.0138)
	$\lambda_{15\text{-}18}$	0.0619 (0.0165)	0.0619 (0.0165)
	$\lambda_{18\text{-}21}$	0.0274 (0.0122)	0.0274 (0.0122)
	$\lambda_{21\text{-}24}$	0.0534 (0.0189)	0.0534 (0.0189)
	$\lambda_{24\text{-}27}$	0.1081 (0.0312)	-
	λ_{27+}	0.0262 (0.0088)	-
	β	-0.4222 (0.1806)	-0.3963 (0.1970)

The estimate of β under an arbitrary baseline hazard shows that chemotherapy significantly reduces the risk of death (p=0.02). For the full follow-up period, the hazard ratio is estimated as 0.66 (95% CI: 0.46 to 0.94). The estimates of β under an exponential, Weibull and piecewise exponential baseline hazard are reasonably similar to this with the piecewise exponential being most comparable. For the full follow-up period, the exponential model estimates the constant hazard over time as 0.0331 in the CT group and 0.0488 in the NoCT group, which are the same as the estimates of the hazard calculated in Section 3.4.2. The estimates of the constant hazard in each time interval of the piecewise exponential model shows the baseline hazard fluctuating over time and the Wald test for the time interval variable showed that these differences were statistically significant ($p<0.0001$ for full follow-up and $p=0.002$ for within 24 months). The estimate of the shape parameter γ for the Weibull model is significantly greater than 1 indicating that the hazard increases over time and therefore an exponential distribution would appear not to be appropriate. In all models the reduction in hazard of death by chemotherapy is slightly less when the analysis is restricted to within 24 months from surgery, although the results are reasonably comparable.

7.5 Extending Model to Include Quality of Life as a Time-Dependent Covariate

7.5.1 Cox Regression Model with Time-Dependent Covariates

The Cox regression model specified in [7.1] can be extended to incorporate time-dependent covariates. The model for the hazard function can be written as:

$$h_i(t) = h_0(t)\exp(\underline{\beta}'\underline{x}_i + \underline{\omega}'\underline{z}_i(t)) \qquad [7.16]$$

where $\underline{z}_i(t)$ is a vector of time-dependent covariates for patient i with associated regression coefficients $\underline{\omega}$. The variables in $\underline{z}_i(t)$ represent successive measures of a binary, ordinal or continuous covariate. When time-dependent variables are included in the model the relative hazard becomes time-dependent and so the model ceases to be a proportional hazards model.

In a Cox regression model with fixed covariates it is possible to consider a parametric form for the baseline hazard and in the previous section exponential and Weibull models were considered. Such parametric models however cannot easily accommodate time-dependent covariates (Petersen 1986). In fitting the Cox regression model with quality of life as a time-dependent covariate the assumption of an exponential or Weibull distribution for the survival times is no longer considered. The only alternative that we consider to the Cox regression model with arbitrary baseline hazard is the piecewise exponential model which, although it is based on an underlying parametric assumption, can incorporate time-dependent covariates.

Suppose the death times of the n individuals in a study are $t_1<t_2<...<t_i< ....<t_n$ and suppose $R(t_i)$ is the set of individuals at time t_i that are at risk of death. The partial likelihood specified in [7.14] is still applicable but the covariates become time-dependent and thus to maximise the partial likelihood with a time-dependent covariate $z(t)$, the *true* value of the covariate $z(t_i)$ at each death time t_i is needed for all individuals in the risk set $R(t_i)$, including the individual who dies at that time. In practice, the covariate is measured at discrete time points that in general do not coincide with event

times and the covariate is measured with error, which may be substantial (see Section 7.7 for further discussion). The usual method is to impute the values of the covariate at the event times (Collett 1994). The most common approach is to impute the value of the covariate at time t to be the last recorded value prior to t. In situations where there are recorded values either side of the time of interest, then it may be preferable to use either the value at the closest time or, for continuous variables, a linearly interpolated value. Alternative more sophisticated options have been suggested such as using a 'window approach', where by defining what is a proximate value to a failure time, only these are used as covariate values (Gail 1981).

During a study, a patient may experience changes in quality of life as time passes. This could be described either by a changing quality of life score or by the movement in and out of various quality of life health states. The change in score or pattern of movement between states may help to explain survival differences and can be considered for inclusion in any survival model as a time-dependent covariate. The inclusion of such a covariate however is not necessarily straightforward. As described above, a quality of life value is required for all patients still alive and on follow-up at each death time. The difficulties of estimating values at each death time when few or no patients have values actually measured at that time are exacerbated by the additional problem of patients dropping out of the quality of life study despite having continued follow-up for survival, resulting in non-ignorable missing covariate values for all time after dropout.

7.5.2 Model with Quality of Life as a Time-Dependent Covariate

The Cox regression model with treatment as a fixed covariate as specified in [7.14] is extended to include quality of life as time-dependent covariate as follows:

$$h_i(t) = h_0(t)\exp(\beta G_i + \omega Q_i(t)) \qquad [7.17]$$

where $Q_i(t)$ is the quality of life of patient i at time t.

This model is fitted to the MIC and ESPAC data with an arbitrary and a piecewise exponential baseline hazard function.

The aim here is to illustrate the application of standard methods to the quality of life and survival data and thus standard methods for imputation of quality of life values at death times, as described below, are used. These methods will generally be inadequate as they take no account of measurement error and do not account for the fact that the missing data after dropout will not be missing at random but alternative methods that overcome these problems are discussed in Section 7.7 and developed in Chapter 10. For the arbitrary baseline hazard, the values of quality of life at each death time that are required to maximise the partial likelihood are estimated using the last observed value carried forward. This assumes that patients remain in steady state from one assessment to the next and from the last assessment to death. For a piecewise exponential model, time-dependent covariates are assumed to be constant for each interval and the model requires the value of the covariate for each interval for all patients at risk at the start of the interval. The last observed value prior to the start of the interval is used as the covariate value for the interval. As an alternative for a continuous covariate, the mean value for the interval is estimated as the mean of the values at the start and end of the interval, calculated by linear interpolation of the observed values. If a patient dies or is censored within the interval then the last observed quality of life value within the interval is used. For all analyses, dropout is accommodated by carrying the last observed value forward to either death, last survival follow-up or end of study time. All time points prior to the first assessment are imputed by carrying the first value backwards to time 0.

The PHREG procedure in SAS (SAS Institute Inc 1992) was used to fit the model with an arbitrary baseline hazard to the data. As before, the data for analysis consists of one line per patient but in addition to the survival data and treatment covariate, the quality of life data per patient is attached as a series of variables representing the successive assessment times and the corresponding quality of life values. Data statements within the PHREG procedure allow the values included at each death time to be that previously observed.

7.5.3 Application to the MIC Study

In modelling the survival within 15 weeks for patients in the MIC study, quality of life was included as a time-dependent covariate as both *GQS* and *MAL*. Patients' successive

values for these variables were included as time-dependent covariates in the Cox regression model. *GQS* was included in terms of the changing value over time whilst *MAL* was included in terms of the movement between two different health states: well (*MAL*=0) and ill (*MAL*=1,2,3). Time spent in the well health state was allocated the value 0 and the ill state was allocated the value 1. The values of *GQS* and *MAL* at each death time for the arbitrary baseline hazard or for each time interval in the piecewise exponential model are calculated using last value carried forward. Also the values of *GQS* for each time interval in the piecewise exponential model are included as an interval mean, calculated as specified in Section 7.5.2. In all cases, the time from study entry to the first assessment is allocated the value of quality of life at the first assessment, i.e. first value carried backwards. The maximum likelihood estimates of the model parameters are given in Table 7.3.

In the model with an arbitrary baseline hazard, the regression coefficients for the quality of life covariates suggest that there is a significant relationship between the changing values of both *GQS* and *MAL* and survival within 15 weeks (Wald chi-square $p<0.0001$ and $p=0.01$ respectively). The hazard ratio for *GQS* is estimated as 0.94 (95% CI: 0.92 to 0.97) suggesting that increasing values of *GQS* (i.e. improving quality of life) are associated with a reduced hazard of death. The hazard ratio for *MAL* is estimated as 3.60 (95% CI: 1.37 to 9.47) suggesting that moving into the ill state is associated with an increased hazard of death. The regression coefficients for treatment show that after adjusting for the effect of changing values of *GQS* over time, survival within 15 weeks is no longer significantly different between treatments ($p=0.13$), whereas treatment remains significant after adjusting for the effect of *MAL* ($p=0.02$). The alternative version of the model with arbitrary baseline gives very similar results.

In the piecewise exponential model, the values for the baseline hazard are comparable across the three different models and all show the hazard to increase with time but in all cases the Wald chi-square statistic for the interval variable showed no significant difference between the intervals ($p>0.25$). The estimates of the regression parameters are similar to those from the model with arbitrary baseline hazard. The relationship between quality of life and survival is significant in all models, particularly when the interval mean value for *GQS* is used ($P<0.0001$) and the effect of treatment is reduced with the inclusion of *GQS* but remains significant with the inclusion of *MAL*.

Table 7.3: Maximum likelihood estimates (and standard errors) for survival models with treatment and quality of life as covariates in the MIC study

Form of baseline hazard	Values for QoL covariate	Parameter	MLE (SE)
Arbitrary	*GQS* LVCF	β	-0.5853 (0.3906)
		ω	-0.0577 (0.0127)
Arbitrary (alternative*)	*GQS* LVCF	β	-0.6030 (0.3898)
		ω	-0.0562 (0.0127)
Piecewise Exponential	*GQS* LVCF	$\lambda_{0\text{-}6}$	0.0104 (0.0039)
		$\lambda_{6\text{-}9}$	0.0173 (0.0073)
		$\lambda_{9\text{-}12}$	0.0212 (0.0085)
		$\lambda_{12\text{-}15}$	0.0264 (0.0102)
		β	-0.6518 (0.3908)
		ω	-0.0415 (0.0128)
	GQS Interval mean	$\lambda_{0\text{-}6}$	0.0091 (0.0035)
		$\lambda_{6\text{-}9}$	0.0153 (0.0066)
		$\lambda_{9\text{-}12}$	0.0188 (0.0078)
		$\lambda_{12\text{-}15}$	0.0237 (0.0094)
		β	-0.4964 (0.3999)
		ω	-0.0549 (0.0133)
Arbitrary	*MAL* LVCF	β	-0.9306 (0.3827)
		ω	1.2802 (0.4939)
Arbitrary (alternative*)	*MAL* LVCF	β	-0.9306 (0.3827)
		ω	1.2903 (0.4939)
Piecewise Exponential	*MAL* LVCF	$\lambda_{0\text{-}6}$	0.0083 (0.0034)
		$\lambda_{6\text{-}9}$	0.0151 (0.0068)
		$\lambda_{9\text{-}12}$	0.0186 (0.0079)
		$\lambda_{12\text{-}15}$	0.0209 (0.0088)
		β	-0.9137 (0.3827)
		ω	1.5569 (0.5407)

*** fit as a piecewise exponential with 24 time intervals defined by death times**

7.5.4 Application to the ESPAC Study

Successive values of *GHSS* were included as a time-dependent covariate in the Cox regression model for survival in the ESPAC study. With so few subjects completing quality of life assessments after 24 months, the survival analysis was restricted to this 24-month period from surgery. The values of *GHSS* at each death time for the arbitrary baseline hazard or for each time interval in the piecewise exponential model are calculated using last value carried forward. Also the values of *GHSS* for each time interval in the piecewise exponential model are included as an interval mean, calculated

as specified in Section 7.5.2. In all cases, the time from study entry to the first assessment is allocated the value of quality of life at the first assessment, i.e. first value carried backwards. The maximum likelihood estimates of the model parameters are given in Table 7.4.

Table 7.4: Maximum likelihood estimates (and standard errors) for survival models with treatment and quality of life as covariates in the ESPAC study

Form of baseline hazard	Values for QoL covariate	Parameter	Within 24 Montbs
Arbitrary	***GHSS*** **LVCF**	β	-0.4766 (0.1983)
		ω	-0.0325 (0.0042)
Piecewise Exponential	***GHSS*** **LVCF**	$\lambda_{0\text{-}3}$	0.0027 (0.0019)
		$\lambda_{3\text{-}6}$	0.0188 (0.0055)
		$\lambda_{6\text{-}9}$	0.0511 (0.0106)
		$\lambda_{9\text{-}12}$	0.0688 (0.0142)
		$\lambda_{12\text{-}15}$	0.0533 (0.0139)
		$\lambda_{15\text{-}18}$	0.0633 (0.0170)
		$\lambda_{18\text{-}21}$	0.0270 (0.0121)
		$\lambda_{21\text{-}24}$	0.0583 (0.0206)
		β	-0.4645 (0.1977)
		ω	-0.0237 (0.0043)
Piecewise Exponential	***GHSS*** **Interval mean**	$\lambda_{0\text{-}3}$	0.0026 (0.0019)
		$\lambda_{3\text{-}6}$	0.0212 (0.0060)
		$\lambda_{6\text{-}9}$	0.0454 (0.0097)
		$\lambda_{9\text{-}12}$	0.0572 (0.0123)
		$\lambda_{12\text{-}15}$	0.0437 (0.0117)
		$\lambda_{15\text{-}18}$	0.0547 (0.0149)
		$\lambda_{18\text{-}21}$	0.0246 (0.0111)
		$\lambda_{21\text{-}24}$	0.0517 (0.0184)
		β	-0.4112 (0.1975)
		ω	-0.0335 (0.0045)

All models give similar estimates for the treatment and quality of life regression coefficients and the estimates of the constant baseline hazard under both piecewise exponential models are similar to those calculated previously in the model with just the treatment covariate. Regression coefficients for treatment show that after adjusting for the effect of changing values of *GHSS* over time, survival within 24 months is still significantly different between treatments (p=0.02). From the model with arbitrary baseline hazard the hazard ratio is estimated as 0.62 (95% CI: 0.42 to 0.92) compared to 0.66 without the quality of life covariate. The regression coefficients for the quality of

life covariate suggest that there is a very strong relationship between the changing values of *GHSS* and survival within 24 months ($p<0.0001$). From the model with arbitrary baseline hazard the hazard ratio is estimated as 0.97 (95% CI: 0.96 to 0.98), which suggests that increasing values of *GHSS* (i.e. improvement in quality of life) are associated with a reduction in the hazard of death.

7.6 Bayesian Approach to Parameter Estimation in Survival Models

The parameters in a Cox regression model can be estimated using Bayesian inference (Abrams 1998). The aim is to determine a posterior distribution for each of the unknown parameters given the data and prior distributions for the parameters. Vague prior distributions are used here, which should give approximately equivalent results to maximum likelihood estimation. As with the classical approach, different forms for the baseline hazard can be considered. By necessity, all models fitted using Bayesian inference are parametric. It is however still possible to fit the equivalent of a model with an arbitrary baseline hazard by considering it as an extreme form of a piecewise exponential model (as discussed in 7.2.5). The aim of this section is to estimate the parameters in the Cox regression model with treatment and quality of life as covariates as specified in [7.17]. Therefore only an arbitrary and piecewise exponential baseline hazard will be considered, with the arbitrary form just being an extension of the piecewise exponential model.

Any random variable that equals the number of times an event occurs in a given interval of time can be modelled as a Poisson distribution. Thus, given any time interval $(a_{j-1},a_j]$, the number of times a death occurs in this interval can be modelled as a Poisson distribution. The number of deaths for patient i in time interval $(a_{j-1}, a_j]$ has a Poisson distribution with mean μ_{ij} where, given a constant hazard of death h_{ij} within the interval,

$$\mu_{ij} = r_{ij} h_{ij}$$

where r_{ij} is the length of time patient i is at risk of death within the interval. Using the Cox model for the hazard of death as specified in [7.17] with piecewise exponential baseline hazard gives:

$$\mu_{ij} = r_{ij}\lambda_j \exp(\beta G_i + \omega Q_{ij}) \qquad [7.18]$$

By setting λ_j=exp(β_{0j}) the model can be written as:

$$\log(\mu_{ij}) = \log(r_{ij}) + \beta_{0j} + \beta G_i + \omega Q_{ij} \qquad [7.19]$$

Thus to estimate the parameters in [7.17] from a Bayesian approach the likelihood is specified as $d_{ij} \sim Poisson(\mu_{ij})$, where d_{ij} is the number of deaths per patient per time interval which can only be 0 or 1 and the model for μ_{ij} is given in [7.19]. The data must also include for each patient and each time interval the actual time that the patient is at risk during the interval, r_{ij}, and the values of the covariates G_i and Q_{ij}, ideally centred to aid convergence.

Prior distributions for β_{0j} needed to be specified as well as for β and ω. Vague priors were used for all parameters to produce equivalent results to the maximum likelihood estimates from the classical piecewise exponential model. Normal distributions with large variances were used as follows:

$$\beta_{0j}, \beta, \omega \sim N(0, 10000)$$

If the time scale for the data is divided into intervals $(0, a_1]$, $(a_1, a_2]$, ... ,$(a_{J-1}, a_J]$ such that $a_1, a_2, \ldots, a_J$ are the distinct observed death times then the piecewise exponential model will be equivalent to the Cox model with an arbitrary baseline hazard. Since the deaths that occur in an interval all occur at the end, the length of time at risk within each interval is the same for all subjects, that is all patients are at risk for the width of the time interval w_j (j=1,2,..,J) and the model given in [7.18] can be simplified as:

$$\mu_{ij} = w_j\lambda_j \exp(\beta G_i + \omega Q_{ij})$$

By setting $w_j\lambda_j$=exp(β_{0j}) the model can be written as:

$$\log(\mu_{ij}) = \beta_{0j} + \beta G_i + \omega Q_{ij} \qquad [7.20]$$

In this model, patients who are censored within the time interval are assumed to survive the whole time interval, which is not an unreasonable assumption to make when the time intervals are short and one that is also indirectly made in any classical analysis. The results for β and ω from this model with vague priors are equivalent to the maximum likelihood estimates for the Cox model with arbitrary baseline hazard.

If the data being analysed has a large number of deaths, then the number of observed distinct death times is large and therefore the number of time intervals and corresponding β_{0j} are also large and this model may not be computationally feasible. In this case a piecewise exponential model with wider time intervals but in which the actual time at which deaths occur during the interval is used to estimate the model parameters may be a more preferable option.

WinBUGS software (Spiegelhalter et al 2000) was used to fit the piecewise exponential models described here to the MIC and ESPAC data. The form of the piecewise exponential model equivalent to fitting an arbitrary baseline hazard was only feasible for the MIC study. In all cases a burn-in of 50000 iterations was used and estimates of the posterior distributions for the parameters were based on samples of 50000. The models were all checked for convergence by examining the trace plots and autocorrelations and all appeared to be adequate. In both the MIC and the ESPAC studies, the Bayesian estimates of the regression coefficients and underlying baseline hazard were very similar to those obtained using maximum likelihood estimation (see Tables 7.5 and 7.6).

Table 7.5: Comparison of Bayesian versus classical parameter estimates (and standard errors) for survival models in the MIC study

Form of baseline hazard	Values for QoL covariate	Parameter	Classical	Bayesian
Arbitrary	*GQS* LVCF	β	-0.5853 (0.3906)	-
		ω	-0.0577 (0.0127)	-
Arbitrary (alternative*)	*GQS* LVCF	β	-0.6030 (0.3898)	-0.6161 (0.3967)
		ω	-0.0562 (0.0127)	-0.0560 (0.0128)
Piecewise Exponential	*GQS* LVCF	$\lambda_{0\text{-}6}$	0.0104 (0.0039)	0.0105 (0.0038)
		$\lambda_{6\text{-}9}$	0.0173 (0.0073)	0.0168 (0.0072)
		$\lambda_{9\text{-}12}$	0.0212 (0.0085)	0.0206 (0.0083)
		$\lambda_{12\text{-}15}$	0.0264 (0.0102)	0.0257 (0.0099)
		β	-0.6518 (0.3908)	-0.6559 (0.3973)
		ω	-0.0415 (0.0128)	-0.0414 (0.0129)
	GQS Interval mean	$\lambda_{0\text{-}6}$	0.0091 (0.0035)	0.0089 (0.0034)
		$\lambda_{6\text{-}9}$	0.0153 (0.0066)	0.0149 (0.0065)
		$\lambda_{9\text{-}12}$	0.0188 (0.0078)	0.0183 (0.0076)
		$\lambda_{12\text{-}15}$	0.0237 (0.0094)	0.0230 (0.0091)
		β	-0.4964 (0.3999)	-0.5001 (0.4070)
		ω	-0.0549 (0.0133)	-0.0549 (0.0134)
Arbitrary	*MAL* LVCF	β	-0.9306 (0.3827)	-
		ω	1.2802 (0.4939)	-
Arbitrary (alternative*)	*MAL* LVCF	β	-0.9306 (0.3827)	-0.9378 (0.3891)
		ω	1.2903 (0.4939)	1.3620 (0.5162)
Piecewise Exponential	*MAL* LVCF	$\lambda_{0\text{-}6}$	0.0083 (0.0034)	0.0080 (0.0033)
		$\lambda_{6\text{-}9}$	0.0151 (0.0068)	0.0144 (0.0066)
		$\lambda_{9\text{-}12}$	0.0186 (0.0079)	0.0176 (0.0076)
		$\lambda_{12\text{-}15}$	0.0209 (0.0088)	0.0198 (0.0086)
		β	-0.9137 (0.3827)	-0.9226 (0.3885)
		ω	1.5569 (0.5407)	1.6630 (0.5739)

*** fit as a piecewise exponential with 24 time intervals defined by death times**

Table 7.6: Comparison of Bayesian versus classical parameter estimates (and standard errors) for survival models in the ESPAC study

Form of baseline hazard	Values for QoL covariate	Parameter	Classical	Bayesian
Arbitrary	*GHSS* LVCF	β	-0.4766 (0.1983)	-
		ω	-0.0325 (0.0042)	-
Piecewise Exponential	*GHSS* LVCF	$\lambda_{0\text{-}3}$	0.0027 (0.0019)	0.0027 (0.0019)
		$\lambda_{3\text{-}6}$	0.0188 (0.0055)	0.0188 (0.0054)
		$\lambda_{6\text{-}9}$	0.0511 (0.0106)	0.0507 (0.0106)
		$\lambda_{9\text{-}12}$	0.0688 (0.0142)	0.0683 (0.0141)
		$\lambda_{12\text{-}15}$	0.0533 (0.0139)	0.0529 (0.0137)
		$\lambda_{15\text{-}18}$	0.0633 (0.0170)	0.0628 (0.0169)
		$\lambda_{18\text{-}21}$	0.0270 (0.0121)	0.0267 (0.0120)
		$\lambda_{21\text{-}24}$	0.0583 (0.0206)	0.0580 (0.0205)
		β	-0.4645 (0.1977)	-0.4661 (0.1989)
		ω	-0.0237 (0.0043)	-0.0236 (0.0043)
Piecewise Exponential	*GHSS* Interval mean	$\lambda_{0\text{-}3}$	0.0026 (0.0019)	0.0026 (0.0019)
		$\lambda_{3\text{-}6}$	0.0212 (0.0060)	0.0210 (0.0060)
		$\lambda_{6\text{-}9}$	0.0454 (0.0097)	0.0452 (0.0096)
		$\lambda_{9\text{-}12}$	0.0572 (0.0123)	0.0568 (0.0123)
		$\lambda_{12\text{-}15}$	0.0437 (0.0117)	0.0434 (0.0117)
		$\lambda_{15\text{-}18}$	0.0547 (0.0149)	0.0543 (0.0148)
		$\lambda_{18\text{-}21}$	0.0246 (0.0111)	0.0244 (0.0110)
		$\lambda_{21\text{-}24}$	0.0517 (0.0184)	0.0515 (0.0183)
		β	-0.4112 (0.1975)	-0.4114 (0.1979)
		ω	-0.0335 (0.0045)	-0.0335 (0.0045)

7.7 Summary and Discussion

The focus of this chapter has been to consider the possible model options for survival data. These models will be relevant for all three approaches that simultaneously analyse survival and quality of life data. Models with just treatment as a fixed covariate are particularly relevant for the multistate modelling in Chapter 9. Survival models are the simplest form of a multistate model, describing the movement of patients from an alive state to a dead state. In Chapter 9 this simple two-state model is extended to incorporate several different health states and the models described here with just treatment as a covariate illustrate the forms that are considered for modelling the transition rates between health states. The models that incorporate quality of life as a time-dependent covariate as well as treatment are particularly relevant for the joint modelling of quality

of life and survival in Chapter 10. In this later chapter, joint modelling is achieved by allowing the survival model illustrated here to include modelled values of quality of life rather than observed values. Since a Bayesian approach to joint modelling is taken, the models that include quality of life as a time-dependent covariate are considered here from a Bayesian as well as a classical viewpoint.

The model with piecewise exponential baseline hazard turns out to be the only practical option for modelling survival data with quality of life as a time-dependent covariate in both a Bayesian and classical framework. A Cox regression model with a piecewise exponential baseline hazard using a reasonable number of time intervals should generally be a good approximation to that with an arbitrary baseline hazard. Problems may occur when there are no early deaths since the first interval may have to be particularly wide and changes in quality of life that occur over this wide time interval which may be related to survival will not be incorporated into the model.

The analysis presented here also gives further insight into the underlying hazard of death and the effect of treatment on this hazard in the two illustrative studies. More importantly it also provides an initial look at the relationship between changing quality of life and survival. In both studies the models with just the treatment covariate showed that the underlying hazard was changing over time and chemotherapy significantly reduced the hazard of death. The inclusion of the quality of life covariate in all formats showed strong evidence of an association between quality of life and survival, with improving quality of life associated with a decreased hazard of death.

Although treatment and quality of life are the only covariates that are included in the models here, the models could easily be extended to include other covariates. These simplistic models however enable the key covariates to be investigated and are adequate for illustrating the methodology, with both fixed and time-dependent covariates being represented. Since the aim of the modelling was not to find the set of covariates that best explained the variation in the survival or the form of the underlying baseline hazard that best fit the data, consideration has not been given to model adequacy but all the standard methods apply. From a classical perspective, changes between nested models in minus twice log likelihood can be used to assess the significant contribution of covariates to the model. For non-nested models the Akaike Information Criterion (AIC)

can be used, with smaller values of AIC indicating better models. Discrimination between an arbitrary baseline hazard and a parametric form is not always easy but the model that yields the smaller standard errors for the regression parameters indicates a more efficient model. Model adequacy can be checked using plots of residuals (Collett 1994). From Bayesian perspective Bayes factors can be used to compare models although this is not straightforward as it requires an informative prior for the baseline hazard (Abrams 1998).

One of the key aims in this chapter has been to include quality of life as a time-dependent covariate in the Cox model, enabling the relationship between quality of life and survival to be investigated. Including quality of life as a time-dependent covariate however is problematic as values are required for all patients in the risk set at each death time. Since covariate values are generally not available at these specific times, simple methods of imputation are normally used to estimate the required values. Last value carried forward is probably the simplest and most widely used option and is used here. Linear interpolation can also be used to estimate covariate values at death times but cannot be applied for the period of follow-up time after the last assessment. For the piecewise exponential model the value of the covariate is fixed for each interval and using the last observed value prior to the interval as the value for that interval may be very unrepresentative, especially if the interval is long, as it ignores any changes in quality of life during the interval. Using an estimate of the mean quality of life for the interval as the covariate as used here may therefore be preferable but again this cannot be applied to the intervals after the last assessment and last value carried forward is still used for all time on follow-up after the last assessment. Clearly these methods of imputation are inadequate for quality of life data where missing data after dropout will generally not be missing at random and where the dropout period may be of considerable length. In addition, maximising the partial likelihood requires the value of quality of life of patients at their own actual death time and this will be estimated as their last observed value, which will clearly be wrong. The value of quality of life at death could be imputed as the worst value, which may be zero if the measure is a utility or utility-type measure but this induces a strong relationship between quality of life and survival, which defeats one of the objectives of the analysis, which is to investigate such a relationship.

The imputed covariate values of quality of life will be treated in the model as true values and no account is made for measurement error. It has been shown that the presence of measurement error in the covariate values will cause the estimated parameters to be biased towards the null (Prentice 1982). One option to reduce this bias is to use a two-stage model (Tsiatis et al 1995) where the covariate values included at each death time are updated estimates of the true values based on a random effects model for all information prior to the death time. With new random effects models being fitted for every death time, this is a computationally complex method. In addition this two-stage model does not use any survival information in the modelling of the covariate process and thus a joint modelling approach would be preferable (Wulfsohn and Tsiatis 1997, Faucett et al 1998). Approaches to joint modelling are discussed in Chapter 10.

CHAPTER 8: QUALITY-ADJUSTED SURVIVAL ANALYSIS

8.1 Introduction

The aim of this chapter is to review in detail the different methods for quality-adjusted survival analysis in terms of their applicability to longitudinal quality of life data collected on patients. The methodology is developed to deal with the additional informative dropout of patients from quality of life studies prior to death.

Quality-adjusted survival analysis is the most widely used approach in clinical research for the simultaneous analysis of quality of life and survival data. The approach combines the amount of time patients spend in a number of different health states with weights reflecting the quality of life of those health states to create a composite measure of quality and quantity of life. The method was originally devised for health states defined using clinical data (Gelber and Goldhirsch 1986) and has not been widely used for the analysis of quality of life data collected on patients in a clinical trial. More recently methods have been proposed that directly incorporate the longitudinal quality of life collected on patients with the survival data (Glasziou et al 1998) and this alternative to the health-state based method is far more disposed to the analysis of longitudinal data collected in clinical trials.

There are two main approaches to quality-adjusted survival analysis depending on the level of aggregation of the quality of life and survival data. The subject-based approach combines quality of life and survival at the subject level, thus creating a single endpoint for each individual on which to compare treatments. The group-based approach aggregates quality of life and survival at a group level, such as treatment group. For both levels of aggregation there are two different approaches to using the longitudinal quality of life data; either the quality of life data can be used to determine the time that patients spend in different health states over time and then quality weights are attached to these times or the values of quality of life measured over time, as long as they are utility-type measures, can be used directly to 'down-weight' the survival time.

In this thesis all methods for quality-adjusted survival analysis will be reviewed in terms of their applicability to longitudinal quality of life data collected on patients. The application of all the different approaches to longitudinal quality of life data has not been demonstrated previously and this thesis allows the methods to be reviewed in detail. The methods are developed to deal with the problem of informative dropout (see Chapter 5). The thesis introduces the concept of *survival-adjusted quality-of-life*, which reflects the fact that the method adjusts for the dropout of patients due to death in the analysis of quality of life. By interpreting quality-adjusted survival analysis in this way, the extension of the methodology to deal with additional dropout prior to death can be meaningfully interpreted as *dropout-adjusted quality-of-life.*

Quality-adjusted survival analysis is based on the concept of quality-adjusted life years (QALYs) and QALY models, particularly TWiST and QTWiST, are described in Section 8.2. The different approaches to combining quality of life are introduced in Section 8.3 and then the following Sections 8.4 to 8.7, illustrate in detail the four different approaches: subject- and group-based aggregation based on either health-states or direct inclusion of values. The MIC and ESPAC studies are used in all cases to illustrate the methodology and the extension of the methods to deal with dropout prior to death is considered in each section. The final section 8.8 provides a critical review of the methods for the application to longitudinally collected quality of life data.

8.2 QALY Models

8.2.1 The General QALY Model

Quality-adjusted survival analysis is based on the concept of quality-adjusted life years (QALYs) (Glasziou et al 1990) where quality and quantity of survival are combined into a single composite measure. To calculate QALYs, years of life are multiplied by a fraction, the quality-adjustment fraction, which expresses the impairment in quality of life experienced during this time. The quality-adjustment fraction ranges from 0 to 1, with 0 representing quality of life equivalent to death and 1 representing perfect health. Negative values can be used if the quality of life is thought to be worse than death. These weights are intended to reflect the relative desirability of the state and are usually

referred to as 'health state utilities' (Torrance 1987), as described earlier in Section 4.2.3.

If the patient experiences or is expected to experience a series of health states H_j (j=1 to J), with different levels of quality of life as measured by utilities u_j (j=1 to J) and the time spent in each state H_j is given by t_j (j=1 to J), then the conventional approach to calculating QALYs is to sum the weighted times spent in the different states. This gives the following standard form for the QALY model:

$$QALY = \sum_{j=1}^{J} u_j t_j \qquad [8.1]$$

Glasziou et al (1990) discuss the assumptions which the QALY model is based on, which are as follows:

- *utility independence* - the utility value for a health state does not depend on the time spent in that state;
- *context independence* - the utility value assigned to a health state is independent of previous or future quality of life or the amount of remaining life;
- *risk neutrality* - all life years are valued equivalently.

The risk neutrality assumption means that time is included in the model as a linear term, t, rather than as a non-linear function of time, i.e. $f(t)$. More general models, that include different forms of discounting or risk-adjustment, have been suggested (Glasziou et al 1990).

8.2.2 The TWiST Model

A special QALY endpoint for comparing therapies, that incorporates both length and quality of survival into a single measure, was developed in a subject-based approach to quality-adjusted survival analysis (Gelber et al 1989). The endpoint that they devised, TWiST (Time spent Without Symptoms of disease and Toxicity of treatment), was developed as a measure of the 'good' quality time experienced by the patient. It was originally developed to assess treatments for breast cancer (Gelber and Goldhirsch 1986,

Gelber et al 1987, Gelber et al 1989) and has also been used in the assessment of treatments for ovarian cancer (Willemse et al 1990, Willemse et al 1992).

TWiST is calculated for each patient by subtracting from overall survival those periods of time during which treatment or disease reduce their quality of life. This is equivalent to calculating QALYs for a patient with utility values of 0 for times with symptoms and toxicity, and utility values of 1 otherwise. The definitions of time with symptoms of disease and time with toxicity of treatment can be adjusted for different clinical situations depending on the disease and treatment under study. Defining the untoward events that can occur and determining the importance attached to each one, in terms of the amount of time subtracted from total survival, are of paramount importance in creating a meaningful TWiST measure (Feldstein 1991).

The definitions of TWiST in the literature have in general been based on clinical criteria rather than patient-based measures of quality of life. Quality of life data collected via questionnaires at repeated assessments over time could however be used to define TWiST. For example in one study, researchers gave a quality weighting of 1 to the survival time during which a patient had either an unchanged high quality of life and no signs of symptomatically progressive disease or improvement in quality of life estimates without being hospitalised; all other survival time was weighted as zero (Glimelius et al 1995). Further examples use survival time spent with 'normal' quality of life scores, as defined by the quality of life instrument, as the definition of TWiST (Beacon 1996, Allen-Mersh et al 1994). In the MIC study, patients are never strictly speaking in a TWiST state as they are never without disease and in many cases are undergoing treatment throughout the quality of life study period. In Section 8.4.2 we give an example of a TWiST-type analysis using the MIC data by interpreting all time without malaise (MAL=0) as a TWiST-type state and allocating it a utility of 1, whilst allocating all other time a value of 0.

The TWiST model is a simplistic way of incorporating quality of life into a survival-type endpoint. The model has been criticised for many reasons including the fact that the amounts of time deducted from overall survival to create the TWiST endpoint are arbitrary and that the model does not account for the quality-of-life experienced during

these times (Brunner 1989). Is was recognised that time with toxicity and symptoms could be added with appropriate weights to TWiST to avoid equating these periods of time to death (Gelber et al 1989) and this forms the basis for the Q-TWiST model described below in Section 8.2.3.

8.2.3 The Q-TWiST Model

The Q-TWiST (Quality-adjusted Time Without Symptoms of disease and Toxicity of treatment) endpoint is a natural extension of the quality-of-life oriented endpoint TWiST and an adaptation of the general QALY model. The TWiST model is extended so that periods of time spent with toxicity or relapse are included in the analysis but are weighted to represent their quality value relative to TWiST. Thus, overall survival is scaled downwards by arbitrarily giving survival during treatment or symptoms a reduced value.

The Q-TWiST model was originally developed and used to assess the effects of adjuvant therapy in women with breast cancer (Glasziou et al 1990, Goldhirsch et al 1989, Gelber and Goldhirsch 1989, Gelber et al 1991). In the original breast cancer application (Goldhirsch et al 1989) the following health states were defined:

TOX time having subjective toxic side-effects
TWiST time without symptoms or toxicity
REL time following systemic relapse (this includes time spent recovering from treatment for local recurrence)

The clinical criteria which defined the sections of a patient's follow-up time that would fall into these health states were fully specified.

The model used for Q-TWiST is a QALY model with weights for *TOX* and *REL* representing the quality values of each health state relative to *TWiST*, which has a value of 1. The original form of the Q-TWiST model is given as (Goldhirsch et al 1989):

$$Q\text{-}TWiST = u_t\, TOX + TWiST + u_r\, REL \qquad [8.2]$$

where *TOX* is the time spent with toxicity from treatment, *REL* is the time spent with symptoms of disease and u_t and u_r are the utilities associated with these periods of survival time. Note that values of (u_t, u_r) equal to (1,1), (1,0) and (0,0) gives overall survival, disease-free survival and TWiST respectively.

Applications of the Q-TWiST model tend to use either the same health states as the original breast cancer example with slightly different definitions or a slightly modified version. For example the original breast cancer states have been modified to include an additional 'recovery' state (Feldstein 1991). It has been suggested that in some trials, it may be necessary, in addition to *TOX*, to define a second period of toxicity to represent the late toxic effects of treatments on a patient's quality of life (Gelber et al 1991). The model has been applied to other cancers such as lung (Rosenthal et al 1992), lymphoma (Cole et al 1995) and rectal (Gelber et al 1996) and in other disease settings such as AIDS (Gelber et al 1992). In this AIDS application, patients could move from *TWiST* to a state where first adverse events were experienced and then on to a state of disease progression. Further work is being undertaken to develop Q-TWiST so that it can be applied to neurological diseases such as multiple sclerosis (Schwartz et al 1995a) and epilepsy (Schwartz et al 1995b). Applications of Q-TWiST models have almost entirely been such that the health states are defined using clinical criteria, but patient-assessed quality of life data has been used (Beacon 1996). In this model, three progressive states were defined relating to the quality of life data collected: (i) treatment-related abnormal quality of life, (ii) normal quality of life, and (iii) quality of life deterioration.

Essentially Q-TWiST is a particular type of QALY model. When the model was originally introduced from both a clinical viewpoint (Goldhirsch et al 1989) and a mathematical one (Glasziou et al 1990) and in a more recent overview (Gelber et al 1995), the method of *partitioned survival analysis* (see Section 8.6) was proposed as the appropriate methodology for calculating Q-TWiST. This methodology has therefore become synonymous with the Q-TWiST model and is often referred to as the 'Q-TWiST method' when in fact it can be applied to any QALY model. This group-based approach to quality-adjusted survival analysis is therefore not discussed here in terms of the Q-TWiST model but is discussed later in Section 8.6 in terms of general QALY models. It should be noted here that the method of partitioned survival analysis requires

progressive health states and therefore Q-TWiST models are generally defined in terms of progressive health states, which are not always feasible (for example Schwartz et al 1995a and b). In this thesis, the application of partitioned survival analysis to quality of life data, which is demonstrated in Section 8.6, provides an illustration of the 'Q-TWiST method' in this context.

8.3 Possible Approaches to Combining Quality of Life and Survival

8.3.1 Defining the Four Possible Approaches

The different approaches to quality-adjusted survival analysis can be summarised into four distinct categories, depending on whether the quality of life and survival data are combined at the subject or the group level and depending on whether the actual values of quality of life data collected on patients over time are used directly in the analysis or indirectly by using them to allocate patients to different health states for periods of their follow-up time. Table 8.1 provides a summary of these four approaches and further details are given below.

Quality of life and survival data can be combined in a QALY model either at the subject level or at the group level. The subject-based approach calculates a QALY value for each subject from their quality of life and survival data and uses standard methods of analysis to compare treatment groups in terms of this composite endpoint. This is a particular type of summary measures analysis that was introduced in Section 6.3. The group-based approach estimates group values for quality of life and survival and aggregates them into a QALY at the group level. Although the subject-based approach is generally more straightforward, it can be problematic when there are censored survival times, as discussed below in Section 8.3.2, in which case the group-based approach may be preferable.

In both the subject-based approach and the group-based approach, there are two different ways to use the longitudinal quality of life data in the analysis. The first uses longitudinal quality of life data collected on patients to determine the amount of time spent in a number of different health states and then, given utility values for those states

derived externally, the QALYs can be calculated. These health states can be defined in any way that is appropriate to the study and clinically relevant, and in some cases may conform to the health states defined in the Q-TWiST model. In the group-based approach there is the added proviso that the heath states must be progressive. The second method uses the actual values of quality of life collected over time on patients to determine the QALYs. This method ideally requires the quality of life measures to be utility values but if utilities are not available then utility-type measures can be used to give QALY-type results.

Table 8.1: Summary of the four different approaches to quality-adjusted survival analysis

	Subject-based approaches	Group-based approaches
Using health states	▪ Survival time for a subject partitioned into a series of health states ▪ Times in states for subject combined with associated external utility values ▪ Requires full follow-up of patients in terms of survival for analysis period ▪ Section 8.4	▪ Partitioned survival analysis ▪ Requires progressive health states ▪ Mean times in states for group combined with external utility values ▪ Deals with censored survival times ▪ Section 8.6
Using actual values	▪ Area under subject quality-of-life curve ▪ Ideally requires utility measure ▪ Requires full follow-up of patients in terms of survival for analysis period ▪ Section 8.5	▪ Integrated quality-survival product (IQSP) ▪ Quality of life function for group combined with survivor function for group ▪ Deals with censored survival times ▪ Ideally requires utility measure ▪ Section 8.7

8.3.2 Issues Relating to Subject-Based Approaches

If all patients in the study are followed up until death and quality of life is assessed for the full duration of this time, then a QALY can be calculated for each patient. Treatments can be compared using standard univariate analytical methods such as t-tests, Wilcoxon tests, analysis of variance or regression analysis depending on the

distribution of the outcome and the research question being investigated. In general, however, all subjects will not be followed up until death and for those patients with censored survival times, their QALY will also be censored. It would seem sensible to analyse such censored QALY data using standard survival analysis techniques, but the use of quality weightings in calculating the QALY endpoint creates an informative censoring and thus renders such methods invalid (Glasziou et al 1990). Survival time for patients with poor quality of life will receive a lower weighting than that for patients with good quality of life. Patients with poor quality of life will therefore accumulate QALYs at a slower rate and will therefore be censored earlier on the QALY timescale than those with good quality of life, thus resulting in informative censoring (Glasziou et al 1990).

One option is to restrict the period of study, say to time L, such that all subjects either die within this time or are known to survive this time and hence, in all cases, survival times *within* the restricted period will be uncensored. The QALYs gained within this study time, $QALY(L)$, can be calculated and as long as quality of life values are available for the full duration of each patient's survival time, there will be no censored values of $QALY(L)$ and the comparison of treatments in terms of this measure using standard analytical methods will be valid (Ganiats et al 1995). If, for some patients, quality of life values are not available for the full duration of the survival time because they have dropped out of the study prior to death, then to avoid censored values of $QALY(L)$ it will be necessary to impute quality of life values for the periods of survival with no measures. Further details of dealing with dropout are given for each specific method in Sections 8.4 and 8.5.

If there are a number of survival times censored at an early time point then choosing a fixed time period with no censored survival times may not be a practical approach and using survival analysis techniques even with just a small number of censored values could give biased results (Gelber et al 1989). An alternative option is therefore to replace the censored QALYs with imputed values and analyse using standard statistical methods for non-censored data. Censored data can be replaced with a range of possible values (Gelber et al 1989). The minimum possible QALY would occur if the patient died immediately after the last follow-up. The maximum possible QALY would occur if

the patient survived the rest of the study period with perfect quality of life. A mean value is obtained by assuming the patient survived the rest of the study period with quality of life valued at 0.5 (Gelber et al 1989). This is illustrated and discussed further in Section 8.5.3. A more appropriate solution to overcome the problem of censored QALYs is to use a group-based approach.

8.3.3 Issues Relating to Group-Based Approaches

In the group-based approach, rather than combining quality of life and survival into a composite QALY measure for each subject and averaging them across each treatment group, average quality of life and survival are combined at the group level. In the QALY model, the mean amount of time a group spends in each health state can be calculated using partitioned survival analysis (Glasziou et al 1990). This method requires the health states to be progressive. The QALY is then calculated by combining this time with utilities reflecting the average quality of life of the group in each health state. The analysis generally uses utilities obtained from external valuational studies or considers the full range of possible values in a threshold utility analysis (see Section 8.6). The application of partitioned survival analysis in a QALY model using quality of life data to define the progressive health states is demonstrated in Section 8.6. This method, however, is based on the ability to define a clinically meaningful set of *progressive* health states and this is often not possible with quality of life data. A simple methodology that directly combines longitudinal quality of life data with survival data at a group level has been proposed by a number of authors (Ganiats et al 1995, Hwang et al 1996, Glasziou et al 1998) and has been referred to as the *integrated quality-survival product* (Beacon 1996). It is comparable to that proposed for the analysis of censored cost data (Lin et al 1997). The method is described in Section 8.7. The problem of dealing with additional dropout prior to death in these group-based approaches is discussed in Sections 8.6 and 8.7 in relation to the specific methodologies.

8.3.4 Approaches for the MIC and ESPAC Studies

The MIC study lends itself well to subject-based quality-adjusted survival analysis. The quality of life study is restricted to 15 weeks from trial entry for reasons of data

availability, and during this time all patients have full survival follow-up. There are therefore no censored survival times within 15 weeks and consequently no censored quality-adjusted survival times within this period. Thus having calculated the number of quality-adjusted life weeks within 15 weeks ($QALW(15)$) for each individual, the analysis of this outcome is straightforward.

Methods for calculating $QALW(15)$ for each individual are not so straightforward and two different approaches are considered. The malaise variable is ideal for defining two health states: *well* ($MAL=0$) and *ill* ($MAL=1$, 2 or 3) and the time spent in each of these states can be calculated for each subject and combined with appropriate utilities for these health states to form $QALW(15)$ (see Section 8.4.2). Alternatively for each individual the values of GQS over time can be considered as a utility-type outcome and the area under the curve defined by the values of GQS gives the $QALW(15)$ for the individual (see Section 8.5.2).

With no censored values for $QALW(15)$ there is no real need for a group-based approach to the analysis but it is included for comparative purposes. Firstly, using the health states defined by the malaise variable, the method of partitioned survival analysis is used to calculate $QALW(15)$ in each treatment group (see Section 8.6.3). This method requires progressive health states and as the example will show, it will not generally be appropriate for quality of life data collected over time and the alternative method of the integrated quality-survival product is far more amenable to the MIC data (see Section 8.7.2).

In the ESPAC study, the subject-based approach is problematic because of the censored survival times and therefore the group-based approach is the ideal method for this data. The subject-based approach (see Section 8.5.2) however is carried out in order to illustrate methods for dealing with the censored data and in order to compare the results with those from the group-based approach (see Section 8.7.2). The analysis is restricted to 24 months from trial entry to reduce the amount of censored data in the subject-based approach and to ensure that the estimation in the group-based approach is based on a reasonable number of patients. In both approaches the quality-adjusted life months within 24 months from trial entry ($QALM(24)$) are estimated by considering the actual

values of *GHSS* measured over time as a utility. The health-state methods would require health states to be defined based on the values of *GHSS* and although this is possible, there is no obvious choice for the health states and therefore the subject and group-based approaches that use the values of *GHSS* over time are the preferred options.

8.4 Subject-Based Approach Using Health States

8.4.1 Method

The values of quality of life collected on a patient over time can be used to partition the subject's survival time into a series of health states (see Figure 8.1). The patient can move in and out of J different health states H_1, H_2, ..., H_J in any order before finally moving to a death state (D). The definition of the health states should be based on the quality of life data being collected. The choice of questions and cut-off values that define the health states will be subjective, but the health states should be clinically meaningful and ideally should relate to states with known utility values, although this is generally not possible. The total time spent in each health state can be calculated and for patient i is given as t_{1i}, t_{2i},..., t_{Ji}. If the utility values for these health states u_1, u_2, ..., u_J have been established from an external valuational study then for each patient they can be combined with the times in the usual QALY model as specified in [8.1] to give a value for the individual as follows:

$$QALY_i = \sum_{j=1}^{J} u_j t_{ji} \qquad [8.3]$$

If utility values are not known then $QALY_i$ can be calculated in terms of a variety of possible utility values in a sensitivity analysis. Having calculated $QALY_i$ for each patient, the outcome can be compared using standard statistical methods as described in Section 8.3.2. If there are censored survival times within the period of analysis then these should be analysed using methods described in Section 8.3.2.

Figure 8.1: Example of partitioning survival time for patient *i* into a series of health states and calculating $QALY_i$

← Patient's survival time from trial entry →

Health states	H_2	H_1	H_2	H_3	D
Utility values for health states	u_2	u_1	u_2	u_3	0
Time spent in health state*	$t_{2i}^{[1]}$	t_{1i}	$t_{2i}^{[2]}$	t_{3i}	

* [1], [2] indicates first and second visit to the same state

$$QALY_i = \sum_{j=1}^{3} u_j t_{ji} = (u_1 \times t_{1i}) + (u_2 \times (t_{2i}^{[1]} + t_{2i}^{[2]})) + (u_3 \times t_{3i})$$

Establishing the amount of time that each subject spends in each health state from the longitudinal quality of life data is not necessarily straightforward. Various assumptions are necessary in order to estimate at what point in time a subject moves from one state to another. If subject i is in health state H_j at the assessment taken at time T_{qi} and is in health state H_k ($\neq H_j$) at a subsequent assessment taken at time T_{ri}, then there are a number of options to determine the time of transition. The subject could move to H_k either (a) at time T_{qi}; or (b) at time T_{ri}; or (c) somewhere between T_{qi} and T_{ri}, say mid-way at time $T_{qi} + [(T_{ri} - T_{qi}) / 2]$. This choice will be subjective and may depend on the time-frame to which the quality of life questionnaire relates.

In general it is assumed that subjects remain in a given health state until the subsequent assessments indicate otherwise. If there are large time gaps between assessments, either due to the design of the study or due to intermittent missing data then this assumption may be questionable. The assumption means that patients remain in the health state recorded at the last assessment until death. Again this assumption becomes more tenuous the larger the time gap between the last assessment and death and especially if the last recorded health state was a 'good' one. In particular if the patient drops out of the study then it will generally be invalid to assume that the last value carries forward. If a patient drops out of the quality of life study then the QALY could be censored at the time of the last assessment or time of dropout (in whatever way that is defined). This approach is generally not recommended since, as discussed previously in Section 8.3.2,

censored individual QALYs are problematic to analyse and their occurrence should be minimised. It is therefore preferable to consider imputing the health states from the last assessment to death, ideally using a method other than last value carried forward.

If the health states that the patient experiences during the dropout time can be ascertained from other clinical information such as clinically-assessed performance status then this may enable the QALY to be estimated. Alternatively, 'dropout' could be considered as a health state that occurs prior to death and as such allocated an appropriate utility value. The patient moves into this health state at the time of dropout (in whatever way that is defined). Different utility values could be considered in a sensitivity analysis. For example in Figure 8.1, if patient i dropped out of the study at time $t_{2i} + t_{1i}$, then the health state H_3 could represent the dropout state with utility value u_3. If dropout occurs for different reasons and these reasons are recorded then the analysis could include several dropout states, each with a different utility value. This approach is illustrated in Section 8.4.2 in relation to the MIC study.

8.4.2 Application to the MIC Study

For each individual in the MIC study, the responses to the malaise question at assessments over the quality of life study period are used to determine the health states experienced. The QALY model is defined with two possible health states, *well* (*MAL*=0) and *ill* (*MAL*=1,2,3). The analysis is restricted to 15 weeks from study entry for reasons of quality of life data availability and the number of quality-adjusted life weeks for patient i within the 15-week study period is given as follows:

$$QALW_i(15) = u_W t_{Wi} + u_I t_{Ii} \qquad [8.4]$$

where u_W and u_I are the utility values for the *well* and *ill* states respectively and t_{Wi} and t_{Ii} represent the length of time that patient i spends in the *well* and *ill* states respectively within the 15 week analysis period.

For each patient, the time spent in each health state, t_{Wi} and t_{Ii}, can be determined from the longitudinal values of the malaise variable. Changes in health state are assumed to

occur on the day of the assessment at which the change in state is recorded and patients are assumed to remain in steady state until the next observed change. As discussed in Section 8.3.4, all patients have full follow-up in terms of survival over the 15-week analysis period and therefore, using the assumption of steady state from one assessment to the next, the values for t_{Wi} and t_{Ii} within 15 weeks can be completely determined. Initially, dropout prior to death is ignored and last value carried forward is used for time from dropout to death or 15-week cut-off. The model is also extended to include a dropout state as follows:

$$QALW_i(15) = u_W t_{Wi} + u_I t_{Ii} + u_D t_{Di} \qquad [8.5]$$

where u_D is the utility value for the *dropout* state and t_{Di} represents the length of time that patient i spends in the *dropout* state within the 15-week analysis period. A patient can only move into the *dropout* state if they have not completed the full schedule of questionnaires and the time that a patient moves into the *dropout* state is calculated as 3 weeks after the last assessment (see Section 5.8.1 for further details on dropout in the MIC study).

Since known utility values for the health states in models [8.4] and [8.5] are not available, $QALW_i(15)$ is calculated for a variety of different utility values in a sensitivity analysis. The utility values chosen are purely subjective and represent a range of possible values. For the two-state model, the following pairs of values for (u_W, u_I) are considered: (1,1) representing standard survival analysis within 15 weeks; (1,0) representing a TWiST-type analysis; (1, 0.8) and (1, 0.5) representing perfect health in the well state and reduced quality of life in the ill state; (0.8, 0.5) representing reduced quality of life in both states. When the additional dropout state is included the following values are considered for (u_W, u_I, u_D): (1,0,0), (1, 0.8, 0.4), (1, 0.8, 0.2), (1, 0.5, 0.2), (0.8, 0.5, 0.2). In all these choices, the dropout state is assumed to have worse quality of life than the ill state.

Since there are no censored survival times within the 15-week analysis period, the $QALW(15)$ outcome can be analysed using standard statistical methods. The distribution of the outcome is not normally distributed and therefore the treatments were compared

using a Wilcoxon two-sample test. The means and standard errors (SE) for each treatment group together with the median and inter-quartile (IQ) range are shown in Table 8.2.

Table 8.2 Summary statistics for *QALW*(15) for each treatment group in the MIC study for various utility values in the two- and three-state models

	CT		PAL		p-value from Wilcoxon test
	Mean (SE)	Median (IQ range)	Mean (SE)	Median (IQ range)	
Two-State Model					
(1, 1)	13.90 (0.35)	15.00 (15.00,15.00)	12.80 (0.56)	15.00 (11.43, 15.00)	0.01
(1, 0)	5.67 (0.64)	4.43 (0, 10.00)	5.75 (0.87)	5.86 (0,11.00)	0.95
(1, 0.8)	12.25 (0.34)	12.89 (12.00, 14.00)	11.39 (0.53)	12.00 (10.49, 13.80)	0.16
(1, 0.5)	9.78 (0.41)	9.71 (7.50, 12.50)	9.27 (0.59)	9.14 (7.50, 12.79)	0.42
(0.8, 0.5)	8.65 (0.30)	8.83 (7.50, 10.50)	8.12 (0.44)	7.99 (7.50, 10.56)	0.30
Three-State Model					
(1, 0, 0)	5.37 (0.62)	4.00 (0, 9.86)	5.54 (0.86)	4.71 (0, 9.57)	>0.99
(1, 0.8, 0.4)	11.58 (0.36)	12.00 (10.66, 13.46)	10.81 (0.57)	12.00 (8.91, 13.80)	0.42
(1, 0.8, 0.2)	11.27 (0.39)	12.00 (9.69, 13.46)	10.55 (0.60)	12.00 (7.29, 13.80)	0.46
(1, 0.5, 0.2)	9.17 (0.43)	9.36 (7.24, 11.50)	8.77 (0.63)	8.46 (5.57, 12.00)	0.59
(0.8, 0.5, 0.2)	8.10 (0.32)	8.40 (6.86, 9.90)	7.66 (0.48)	7.50 (5.57, 10.20)	0.47

The first row in Table 8.2 illustrates the overall survival comparison as was seen in Chapter 3, with the CT group having longer survival within 15 weeks compared to the PAL group, which was statistically significant at the 5% level. The TWiST-type models with utility values of (1,0) and (1,0,0), which only count time in the *well* state, show no difference between the CT and PAL group. In the remaining models, the trend is for the CT group to have a greater *QALW*(15) on average than the PAL group but the difference is relatively small and in all cases not statistically significant at the 5% level. This analysis indicates that although there is a difference between the treatment groups in

terms of survival within 15 weeks, the difference diminishes after adjusting for the quality of life experienced during this time and the analysis suggests that there is no evidence of a difference between treatments in terms of quality-adjusted survival within 15 weeks.

8.5 Subject-Based Approach Using Actual Values

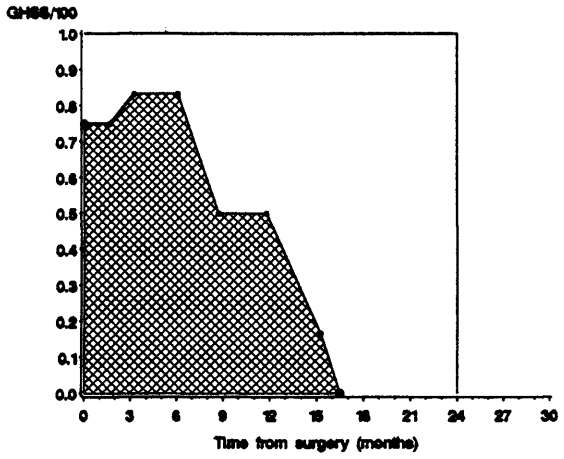
8.5.1 Method

The simplest approach to quality-adjusted survival analysis is to use the actual values of quality of life collected over time to calculate the number of QALYs for each subject (Ganiats et al 1995, Korn 1993). The measures of quality of life taken at discrete points over time for an individual can be used to create a 'curve' that describes the quality of life from entry to trial to either death or to a fixed end-of-study time. By connecting measures at discrete time points in this way, all quality of life values between the actual assessment times are effectively being imputed. When the quality of life measure is a utility, the area under this curve represents the QALYs for that subject. If the quality of life measure is not a utility then it may be possible to transform it to a 0 to 1 scale, to create a 'pseudo utility' (see Section 4.2.4), and the resulting area under the curve may be interpreted like a QALY.

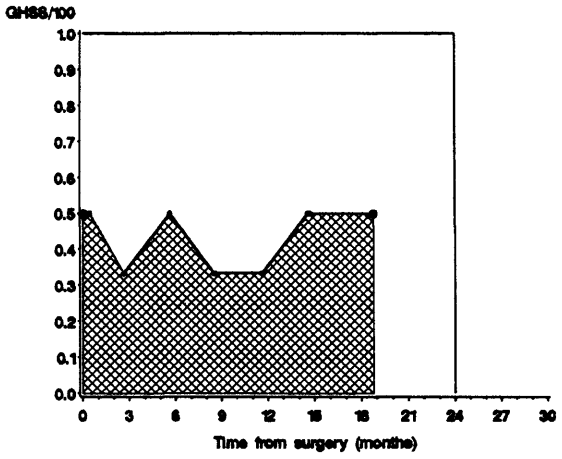
There are a number of practical issues to consider when describing a patient's quality of life, measured at a number of discrete time points, by a curve in continuous time. Some of the issues, such as using actual or scheduled assessment times and using a step function or assuming a linear change between assessments, were discussed in Section 6.2.1. In addition, for an unbiased analysis individual curves need to all relate to the same length of follow-up (Ganiats et al 1995) and therefore to be able to calculate QALYs for each individual, the curve needs to be complete from trial entry to either death or fixed end-of-study time. Various assumptions need to be made to achieve this and these are described below and illustrated using specific examples of patients' longitudinal values of *GHSS* (divided by 100 for a 0 to 1 scale) in the ESPAC study (see Figure 8.2).

Figure 8.2: Examples of quality of life curves for the ESPAC study (red dots indicate imputed values)

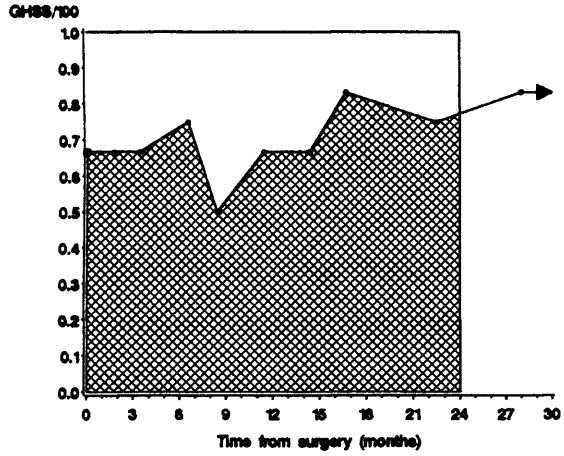
(a) Death within the 2-year study period



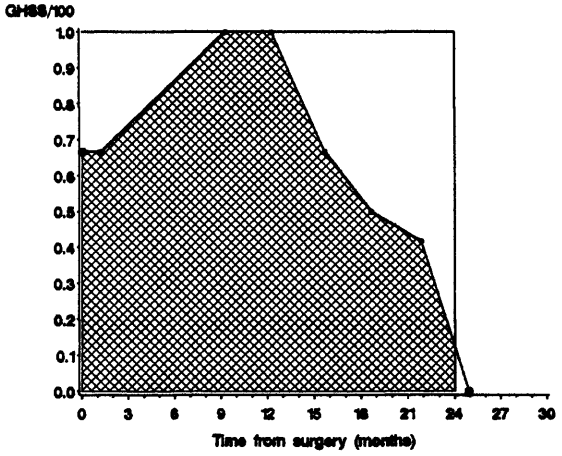
(b) Censored within the 2-year study period



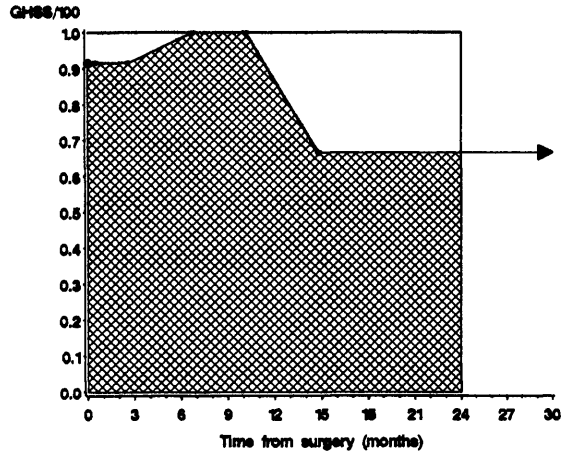
(c) Post 2-year assessment available



(d) Death after 2-year study period



(e) Censored survival greater than 2 years



Value at Trial Entry

If actual assessment times are being used and the first assessment is some time after the trial entry time, then the value of quality of life at entry needs to be imputed and connected to the value at the first assessment. The simplest approach is to carry the first value back to the trial entry time. All examples in Figure 8.2 have imputed values at trial entry.

Intermittent Missing Values

If the assessments are made at scheduled times and there are intermittent missing assessments, a decision needs to be made as to whether it is reasonable to assume that such missing data are missing at random. If so, then the missing data can be ignored and the quality of life values on either side can be connected as normal. If not, then an approach that accounts for informative intermittent missing data needs to be used, such methodology however is not well developed (Diggle et al 1994). In Figure 8.2(d), values between baseline and 9 months are indirectly imputed by assuming a linear change between the observed values.

Values Between Last Assessment and Death

If a patient dies then their measure on the utility-type scale becomes zero from that point onwards. The curve between the last measure before death and the zero measure can be completed either as a step function or as a linear decline (see Figure 8.2(a)).

Values Between Last Assessment and Censored Survival Time

If a patient's survival time is censored then the quality of life curve will stop at the point when the patient was last known to be alive and the QALY will be censored. This can cause problems for the analysis as discussed in Section 8.3.2. The curve between last assessment and censored survival time can be completed using last value carried forward (see Figure 8.2(b)).

Completing Curves Within Fixed Study Periods

The quality of life analysis will often be restricted to a fixed study time, either because that is the period of data collection or in order to minimise the number of patients with censored survival times (as discussed in Section 8.3.2). If a patient is known to survive

the full duration of this quality of life study period, then the curve must be completed within this time by connecting the last measure taken within the study period to the fixed end-of-study time. This can be done in a number of ways depending on the nature of the post-study information. In some situations, it may be necessary or desirable to use purely the information within the study period for the analysis but if post-study information is available then it may enable the values imputed by the quality of life curve at the end of the study period to be more accurate in the following way. If a post-study quality of life assessment is available then the last measure before the fixed cut-off can be connected to the first measure after the cut-off using either a step function or linear change (see Figure 8.2(c)). Otherwise if there is no post-study quality of life data available but the date of death is known, the last measure can be connected to the value zero at the time of death in the most appropriate way (see Figure 8.2(d)). In both cases the part of the curve within the study period completes the quality of life curve, with the value at the cut-off time calculated using interpolation. If a linear decline is assumed between last assessment and death and death occurs a long time after the end-of-study time then decline will be shallow and the interpolated value at the fixed end-of study time will be similar to a value using last value carried forward. If neither of these events occur then the curve between the last measure and the cut-off time can be completed by carrying the last value forward (see Figure 8.2(e)).

Dealing With Dropout Prior to Death

If a patient drops out of the quality of life study but has continued follow-up in terms of survival then either the curve could stop at the last recorded assessment giving a censored QALY, which as discussed in 8.3.2 can cause problems in the analysis, or the curve could continue to the last survival follow-up point (whether it be dead or alive) by imputing the missing values. There are a number of options available for imputation (Fayers and Machin 2000, Fairclough 2002 and see Section 5.5.3), the validity of which may be compromised if the time from last quality of life assessment to last survival follow-up is very long. The simplest approach is to carry the last value forward. Alternatively, the worst or best value could be carried forward from some appropriate point after the last assessment. If the approximate health state of the patient during the dropout time can be ascertained from other clinical information such as clinically-assessed performance status then, at some appropriate point after the last assessment, the

curve could take on a value that reflects the estimated health state of the patient. If the time of death of the patient is known then the curve could linearly decrease to zero over the dropout time. More sophisticated methods such as multiple imputation could also be considered for imputing values at the scheduled times after dropout (whenever that is defined to be) for the formation of the quality of life curves.

The approach taken here is to assume a linear decline in quality of life from the value at the last assessment to the value of zero at death for patients who have a known death time. This assumes that a patient who drops out has progressively deteriorating quality of life over the dropout period, which in most cases will be an appropriate assumption. In both studies most patients have a known death time, but in the few patients with censored survival, dropout is dealt with by using last value carried forward. This may be a valid assumption, especially in those with a long survival time, as the fact that they have not died may be some indication that the dropout may not have occurred due to ill health. An alternative more extreme approach for dealing with dropout prior to death is actually to treat dropout prior to death as an event equivalent to death and to assume they have a value of zero at the point when they dropout. In this way, the quality of life curves can be completed by assuming a linear decrease from the value at the last assessment to the value of zero at the dropout time. This is an equivalent approach to that used for the group-based method using actual values, when the dropout-free survivor function is used instead of the survivor function in the integrated quality-survival product (see Section 8.7.4).

8.5.2 Application to the MIC and ESPAC Studies

In both studies, the quality of life curves for each subject were created by plotting the values of *GQS* for MIC and *GHSS* for ESPAC (both divided by 100 to put the measure on a 0 to 1 scale) against actual assessment times, assuming a linear change between assessments and ignoring intermittent missing values. Any values between the actual assessment times are effectively imputed by the formation of the quality of life curve. Although *GQS* and *GHSS* are not proper utility measures, they are scaled like a utility and therefore the area under the curves that they trace over time will give QALY-type measures. For reasons of data availability, the analysis for the MIC study was restricted

to the 15-week period from entry to trial. For the ESPAC study, the analysis was restricted to 24 months from trial entry in order to reduce the censoring. Quality of life curves need to be defined completely for this analysis period.

The value of quality of life at time of entry to study was imputed using first value carried backwards (see Section 5.6). For those that die within the analysis period (28 in MIC and 105 in ESPAC), a linear decrease in quality of life was assumed from the value at the last assessment to the value of zero at the time of death (regardless of the time span). For patients with censored survival times within the analysis period (0 for MIC and 22 for ESPAC), values of quality of life at the last assessment were carried forward until the time last seen alive (regardless of the time span). For those patients that have survival follow-up greater than the analysis period, post-study information in terms of quality of life where available and survival otherwise was used to complete the curves within the analysis period. For these patients, the 15-week value for MIC or 24-month value for ESPAC is interpolated from the linear change between the two values that straddle the end-of-study cut-off time. When post-study quality of life values were available (6 in MIC and 22 in ESPAC) then the value at the last assessment within the analysis period was joined to the first value outside the period. Otherwise, if time of death was known (73 in MIC and 17 in ESPAC) then the value at the last assessment within the analysis period was joined to zero at the time of death. If the survival time was censored after the analysis period (2 in MIC and 9 in ESPAC) then the value at the last assessment was carried forward to the end-of-study cut-off time.

For each subject, the area under the quality of life curve within the analysis period was calculated to give the quality-adjusted life weeks achieved within 15 weeks for MIC ($QALW(15)$) and quality-adjusted life months achieved within 24 months for ESPAC ($QALM(24)$). For the MIC study there were no censored $QALW(15)$ within the analysis period and thus the outcome can be analysed using standard statistical methods. The means together with standard errors and medians with inter-quartile ranges are given for $QALW(15)$ on each treatment arm in Table 8.3. The difference between treatment arms in terms of this outcome was assessed using Student t-tests and Wilcoxon tests. In the ESPAC study, there were 22 (10 on CT and 12 on NoCT) censored $QALM(24)$ values which were dealt with by (i) retaining them as censored values and using Kaplan-Meier

estimates, and (ii) imputing extreme and average values for them. These individual *QALM*(24) were compared across the two treatment arms and statistically tested using log-rank for (i) and Wilcoxon two-sample tests for (ii) (Table 8.4).

Table 8.3: Quality-adjusted life weeks within 15 weeks, *QALW*(15), by treatment group in the MIC study calculated using subject-based approach

	CT (N=67)	PAL (N=42)	P-values
Mean (standard error)	10.71 (0.40)	8.81 (0.57)	0.006 (t-test)
Median (IQ range)	11.97 (9.83, 13.38)	9.33 (6.55, 11.98)	0.002 (Wilcoxon test)

In the MIC study, patients gain on average approximately 2 extra quality-adjusted life weeks within the 15 weeks from trial entry on CT compared to PAL. This difference between treatment arms was shown to be statistically significant at the 1% level.

Table 8.4: Quality-adjusted life months within 24 months, *QALM*(24), by treatment group in the ESPAC study calculated using subject-based approach

	CT (N=87)	NoCT (N=88)	P-values
Using Kaplan-Meier estimates:			
Mean (standard error)	10.19 (0.70)	8.97 (0.69)	0.29
Median (95% CI)	9.80 (7.11-11.84)	6.38 (5.59-9.62)	(log-rank)
Using best values for censored patients:			
Mean (standard error)	10.67 (0.73)	9.48 (0.72)	0.22
Median (IQ range)	9.86 (4.81-16.51)	6.46 (3.93-15.66)	(Wilcoxon)
Using worst values for censored patients:			
Mean (standard error)	9.43 (0.64)	8.12 (0.60)	0.13
Median (IQ range)	8.44 (4.81-13.42)	5.87 (3.80-11.81)	(Wilcoxon)
Using average values for censored patients:			
Mean (standard error)	10.05 (0.65)	8.80 (0.64)	0.18
Median (IQ range)	9.86 (4.81-14.79)	6.38 (3.93-14.52)	(Wilcoxon)

In all cases the chemotherapy arm achieved, in terms of means, between 1 and 1.5 extra quality-adjusted life month within 24 months, whilst the medians showed a difference of between 2.5 and 3.5 months between the treatment arms. None of these differences were statistically significant at the 5% level. None of these approaches deal with the censored

data adequately and the alternative group-based approach, shown in Section 8.7, is preferred.

8.6 Group-Based Approach Using Health States: Partitioned Survival Analysis

8.6.1 Method

The subject-based approach to calculating QALYs by dividing a patients follow-up into the time spent into different health states was described in Section 8.4. This approach becomes problematic when patients have censored survival times within the analysis period as this causes the individual QALY values to be censored. Application of survival analysis techniques to the QALY endpoint is not valid as the censoring is informative (see Section 8.3.2) and therefore in such situations a group-based approach, which overcomes such difficulties, is preferable.

Partitioned survival analysis was proposed as a group-based approach to quality-adjusted survival analysis (Glasziou et al 1990) and provides a means of handling the problem of informative censoring. The method requires a set of *progressive* health states H_j (j=1 to J) to be defined that completely describe the series of health states experienced by patients from trial entry to death. Overall survival is partitioned into the time spent in each health state and the mean duration in each state for each group are combined as a weighted sum according to the QALY model as given in [8.1]. Thus the QALY model for group G is given by:

$$QALY_G = \sum_{j=1}^{J} u_j t_{jG} \qquad [8.6]$$

where t_{jG} represents the mean time that group G spend in health state j and u_j is the utility for that state which is assumed to be the same for all groups. Weighting the time spent in each health state by the utility at the group level rather than the subject level

avoids the need to weight individual censored survival times and thus overcomes the problem of informative censoring.

The method is ideal for the Q-TWiST model for which the defined health states are progressive and it has been widely used in this context (Goldhirsch et al 1989, Gelber et al 1991, Rosenthal et al 1992, Gelber et al 1995, Gelber et al 1996). Defining progressive health states using longitudinal quality of life data however is problematic and therefore examples of the use of partitioned survival analysis in this context are limited (Beacon 1996). The application of partitioned survival analysis to the MIC data in Section 8.6.3 provides an illustration of the problems.

To estimate $QALY_G$ for each group the mean time spent in each state needs to be estimated from the data. The date of exiting each successive health state is regarded as an endpoint and Kaplan-Meier estimates are calculated for the survival function from a fixed origin, such as date of entry to trial, to each endpoint. If the exit time from one health state is censored at time t for a patient, then all subsequent exit times will be censored at time t. If a state H_j is skipped, then the exit time for H_j will be the same as that for H_{j-1}. Kaplan-Meier survival curves corresponding to each transition time can be overlayed on one graph to show the partitioning of overall survival. These are called partitioned survival plots and separate graphs should be produced for each treatment group (see Figure 8.3 relating to the MIC data).

Estimating the time of exiting each state from the longitudinal quality of life data requires a number of assumptions to be made. These were discussed in relation to the subject-based approach using health states (Section 8.4.1) and are explicitly stated for the analysis of the MIC data in Section 8.6.3. In addition, the problem of dealing with dropout needs to be considered. If a subject is assumed to remain in the health state that they occupied at the last assessment until death, censoring or end-of study time then this may not be appropriate if this time span is long. In such circumstances, it may be preferable to assume that patients move to a 'dropout' state prior to death and include this as the final state in the series of progressive health states. This method of dealing with dropout is applied to the MIC study in Section 8.6.4.

In general, for survival time T, the area under a survival curve defined by the survivor function $S(t)$ provides an estimate of the mean survival time and is given by:

$$E(T) = \int_{t=0}^{\infty} S(t)\,dt \qquad [8.7]$$

The areas under the survival curves for successive endpoints can therefore be estimated and used to compute the areas between the curves giving estimates of the mean duration of each health state. If the last observed survival time is censored, then the entire survival curve cannot be estimated and so these areas can only be calculated if a specified time from randomisation is chosen as the upper time limit for the analysis, i.e. the upper limit of the integral will be this time limit rather than infinity. This may be the upper limit of observation or could be based on the follow-up time of the study cohort. Alternatively if a parametric version of the survivor function is assumed then no upper limit is required.

Mean times from randomisation to exiting each health state, restricted to the upper time limit, are calculated from the area beneath each estimated survivor function from 0 to the chosen finite limit. In practice, the area under a survivor function is estimated by summing the rectangular areas under the Kaplan-Meier curve using the following formula:

$$Area = \sum_{i=0}^{L} \hat{S}(t_i)(t_{i+1} - t_i) \qquad [8.8]$$

where each t_i (i=1 to L-1) is a death time, with t_0 defined to be zero and t_L defined to be the chosen upper time limit. The mean survival time given by SAS in their LIFETEST procedure (SAS Institute Inc 1989) gives the area under the survival curve and can be calculated for the restricted time period defined by upper time limit L by setting all survival times greater than L to be equal to this time and uncensored.

Differences between successive restricted means for time from randomisation to exiting each health state give the restricted mean duration in each state. The restricted mean

quality-adjusted survival is estimated by combining the restricted mean durations as a weighted sum according to the QALY model. Restricted means based on the product limit method are asymptotically unbiased and normally distributed (Breslow and Crowley 1974). Consequently statistical inferences for quality-adjusted survival can be based on the asymptotic normality of the estimates and require the calculation of standard errors of the estimates. The variance for quality-adjusted survival can be estimated from the vector of utility weights and the variance-covariance matrix for the mean times in each state (Glasziou et al 1990) as follows:

$$Var(QALY_G) = \underline{u}' \, \underline{W_G} \, \underline{u} \qquad [8.9]$$

where $\underline{u}$ is the Jx1 vector of utilities defined in model [8.6] and $\underline{W_G}$ is the JxJ variance-covariance matrix for $\hat{t}_{jG}$ (j=1 to J), the estimates of the mean time that group G spend in each health state.

There is no simple expression for the covariance terms when dealing with restricted means and hence the variance-covariance matrix is estimated using a bootstrap method (Glasziou et al 1990). The bootstrap method is carried out as follows (Hinkley 1988, Efron and Tibshirani 1993). A new sample of patients of size N is created by repeatedly sampling with replacement from the N individuals in the study. This process is repeated thousands of times to obtain a whole series of new data sets. Restricted means for times spent in each state are calculated for each data set to produce an empirical sampling distribution, called a bootstrap sampling distribution, for the statistic. The variances and covariances computed from these values are used as the variance-covariance estimates. Variance-covariance estimates have previously been based on a series of 1000 new data sets (Gelber et al 1991).

8.6.2 Comparing Treatments in Terms of QALYs

In some situations it may be possible to compare the QALYs for treatment groups using specific utility values in the QALY model. Values may be chosen arbitrarily if no patient-derived information is available or they could be based on a time trade-off or

standard gamble study, for example (Torrance 1986, 1987). For example, in a study of patients with small cell lung cancer, the researchers justified their arbitrary choice of utility coefficients in a Q-TWiST model for their final conclusions (u_t = 0.75 and u_r = 0.25) as those which they perceived to most closely resemble the clinical experience of the patients (Rosenthal et al 1992). They compared their Q-TWiST results to those from a comparable study, who obtained their utility coefficients (u_t = 0.57 and u_r = 0.15) from a proxy group of patients and health professionals (Goodwin et al 1988).

In most cases, utility weights will be unknown. Treatments can be compared in terms of QALYs using a spectrum of utility values. For the Q-TWiST model specified in [8.2] with two unknown utility values u_t and u_r, a commonly used approach is to carry out a threshold utility analysis, a form of sensitivity analysis, in which the trial data is used to determine the utility values which would give no difference between treatments, i.e. where the restricted mean quality-adjusted survival times in the two treatments is equal. When there are two unknown utility coefficients, the set of values that give equal quality-adjusted survival is described by a straight line on a two dimensional plot. Confidence limits can be calculated for this 'threshold line' by finding the pairs of utility coefficient values for which the confidence interval for the treatment effect captures zero (Glasziou et al 1990, Gelber et al 1995). Such sensitivity analysis becomes more difficult when there are more than two unknown utilities, which may occur when an additional dropout state is included in the model.

In general, threshold utility analysis determines the values of u_j such that for two groups A and B:

$$\hat{\Delta}(QALY) = \hat{QALY}_A - \hat{QALY}_B = \sum_{j=1}^{J} u_j(\hat{t}_{jA} - \hat{t}_{jB}) = 0 \qquad [8.10]$$

If $\underline{\hat{d}}$ represents the Jx1 vector of estimates of differences in mean times between the two groups A and B, i.e. $\hat{d}_j = \hat{t}_{jA} - \hat{t}_{jB}$ for j=1 to J, and $\underline{u}$ represents the Jx1 vector of utilities and the JxJ matrix $\underline{W}$ represents the sum of the variance-covariance matrices for

the two groups $\underline{W_A}$ and $\underline{W_B}$ then the α-level confidence limits for the threshold utility values are given by:

$$\underline{u}'(t^2_{\alpha/2}\underline{W} - \underline{\hat{d}\hat{d}'})\,\underline{u} = 0 \qquad [8.11]$$

where $t_{\alpha/2}$ is the value of the t distribution relating to the $\alpha/2$ percentage.

8.6.3 Partitioned Survival Analysis of the MIC Data

The main problem with carrying out a partitioned survival analysis on the MIC data concerns defining progressive health states. Progressive health states had to be considered that could be derived from the data and were also clinically meaningful. The malaise variable expressed as a two level variable, no malaise (*MAL*=0) and malaise (*MAL*=1,2 or 3) was used to define *well* and *ill* quality of life states respectively. Various sequences of *well* (*W*) and *ill* (*I*) quality of life periods were experienced by patients within the 15 weeks from study entry and these were explored to help define the progressive sequence of health states (see Table 8.5).

Table 8.5 Sequences of *well* (*W*) and *ill* (*I*) quality of life states observed within 15 weeks of trial entry in the MIC study

Sequence	Frequency
W	18
WI	19
WIW	3
WIWI	2
I	38
IW	17
IWI	10
IWIW	2

To incorporate the sequences experienced by *all* patients (i.e. to include both *IWIW* and *WIWI*), the definition of progressive health states would need to include 5 progressive health states (i.e. *IWIWI* or *WIWIW*). This was considered to be beyond both the limit of intelligibility and the limited amount of data. Thus, although a few patients would need

to be excluded from the analysis, a definition consisting of 4 progressive health states was considered to be preferable.

Having decided on a definition consisting of 4 progressive health states, there were two possible sequences to consider; *WIWI* or *IWIW*. The first option would exclude the 2 patients with an *IWIW* sequence and the second option would exclude the 2 patients with a *WIWI* pattern. In general, given several options for a definition, the final choice should be based on what is most clinically meaningful. In this case, neither option seemed clinically preferable to the other and there was no clear clinical explanation for the potential continued fluctuation between the two health states. The decision, therefore, to use the second option described above (i.e. *IWIW*), was arbitrary but was based on the assumption that patients were more likely be an *ill* rather than a *well* state on entry to trial. Sensitivity analysis should be performed to determine if using a different definition affects the conclusions of the analysis. The following analysis therefore excludes 2 patients from the chemotherapy arm.

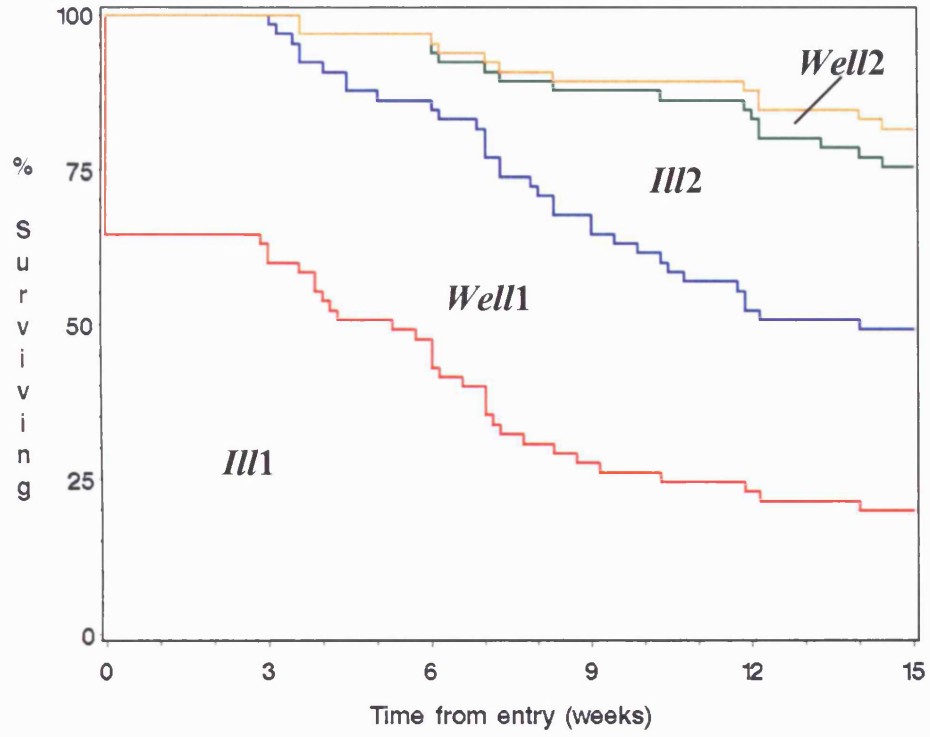
The progressive health states were defined as *Ill1*, *Well1*, *Ill2* and *Well2* where the numbering of the states indicates the first and second visits to the 'same' health state. With this definition of progressive health states, where a patient returns to a health state previously visited, the degree to which the first and second visit to the state are similar needs to be considered. For example, the *ill* quality of life state visited after the patient has been in an *well* state (i.e. *Ill2*) might describe a very different experience to the *ill* quality of life state visited first (i.e. *Ill1*).

Having defined the 4 progressive health states, the date of exit from each health state formed successive endpoints for analysis. The time from entry to study to each endpoint was calculated for each patient, with exit times from a state set equal to the exit time from the previous state if the state was skipped. For example, if a patient was in a *well* state on entry to the study, then they were assumed to have skipped the first state, and their exit time from *Ill1* was set to be zero. Also, for example, if a patient exited *Well1* at 15 weeks and therefore did not experience the other two states within the 15-week analysis period, then exit times from *Ill2* and *Well2* were both set at 15 weeks. Kaplan-

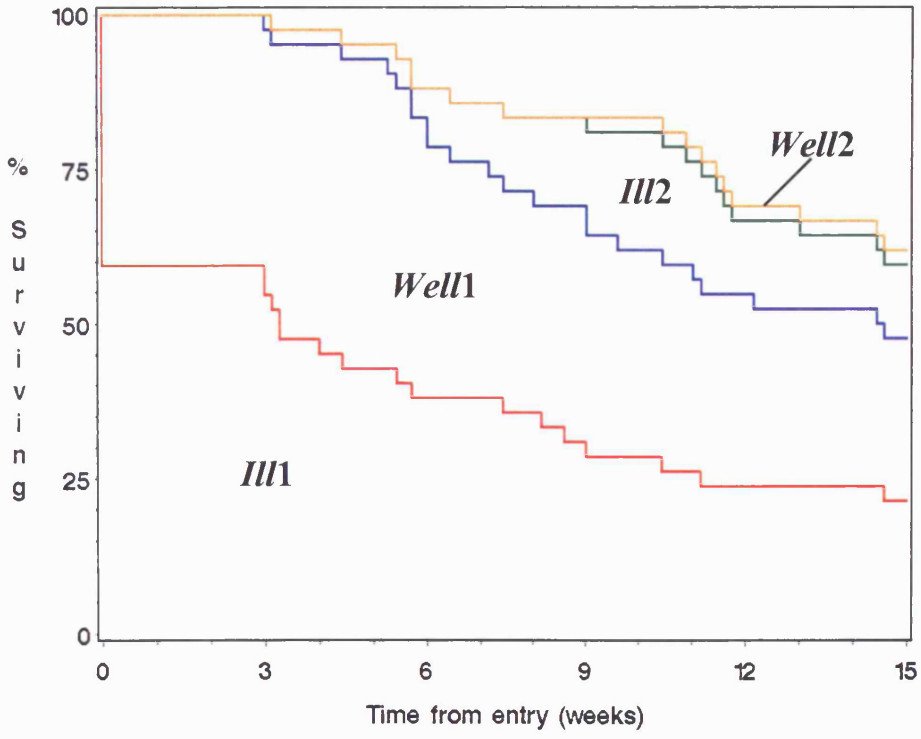
Meier survival curves for successive endpoints were calculated and overlayed to give a partitioned survival plot for each treatment group (see Figure 8.3).

Figure 8.3 Partitioned survival analysis of the MIC data

(a) CT arm



(b) PAL arm



The area under each curve, with an upper time limit of 15 weeks, was obtained from SAS (corresponding to [8.8]) and gave the restricted mean survival times from date of entry to trial to each endpoint. The differences between successive means gave the mean time spent in each health state (see Table 8.6). Standard errors for these times were obtained from 1000 bootstrap samples.

Table 8.6 Restricted mean survival times (and standard errors) in weeks spent in each health state in the MIC study

Health State	CT arm (N=65)	PAL arm (N=42)
Ill1	5.95 (0.72)	5.70 (0.95)
Well1	5.31 (0.61)	5.61 (0.90)
Ill2	2.33 (0.45)	1.34 (0.41)
Well2	0.28 (0.16)	0.14 (0.14)

The QALY model for each treatment group in the MIC data was defined as follows:

$$QALW(15) = u_{I1}\, t_{Ill1} + t_{Well1} + u_{I2}\, t_{Ill2} + t_{Well2} \qquad [8.12]$$

where the utility values u_{I1} and u_{I2} ($0 \leq u_{I1} \leq 1$, $0 \leq u_{I2} \leq 1$) are unknown and reflect the reduction in quality of life during the first and second visits respectively to the *ill* quality of life state. This model assumes that the quality of life in the *well* state, whether at first or second visit, was equivalent to 'perfect health'. It also assumes that the quality of life experienced in the *ill* state differed depending on whether it was the first or second visit.

Utility values for the *ill* states are not known and a sensitivity analysis to specific chosen values of utilities was chosen in preference to a threshold utility analysis. The values considered for (u_{I1}, u_{I2}) were: (0.8, 0.8) to reflect equal utility in both *ill* states; (0.8, 0.5) to reflect the second *ill* state being of worse quality of life than the first; and (0.5, 0.8) to reflect the first *ill* state being worse than the second. The mean $QALW(15)$ for each treatment group was calculated using [8.12] with these utility values and the mean times spent in states given in Table 8.6. The results are given in Table 8.7, with standard errors and 95% confidence intervals obtained from 1000 bootstrap samples.

Table 8.7: Mean *QALW*(15) on each treatment arm of the MIC study with standard errors and 95% confidence intervals for different utility values

	CT (N=65)			PAL (N=42)		
Utility values	**Mean *QALW*(15)**	**Standard Error**	**95% Confidence Interval**	**Mean *QALW*(15)**	**Standard Error**	**95% Confidence Interval**
(0.8, 0.8)	12.21	0.34	11.55-12.88	11.39	0.53	10.35-12.41
(0.8, 0.5)	11.51	0.34	10.85-12.19	10.98	0.52	9.97-11.99
(0.5, 0.8)	10.42	0.43	9.57-11.27	9.68	0.64	8.42-10.92

The CT arm consistently has greater quality-adjusted survival time within 15 weeks than the PAL arm but 95% confidence intervals for the two treatment arms overlap and thus these observed differences are not statistically significant at the 5% level.

8.6.4 Dealing with Dropout in the MIC Analysis

As previously discussed in Section 8.6.1, in the above analysis a subject is assumed to remain in the health state that they occupied at the last assessment until death, censoring or end-of study time. This may not be appropriate if the patient is a dropout, especially the time span is long. The analysis is therefore re-run with patients who drop out of the quality of life study being moved to a *dropout* state prior to death three weeks after their last assessment. The definition of dropout and time to dropout for the MIC study is given in Section 5.8.1. This dropout state is included as the final state in the series of progressive health states and the QALY model given in [8.12] is extended as follows:

$$QALW(15) = u_{I1}\, t_{Ill1} + t_{Well1} + u_{I2}\, t_{Ill2} + t_{Well2} + u_D\, t_{Dropout} \qquad [8.13]$$

where the $t_{Dropout}$ is the mean time spent in the *dropout* state estimated from the partitioned survival analysis and u_D is the unknown utility value associated with this state. Mean times spent in each health state with the *dropout* state included were estimated using partitioned survival analysis and results for each treatment arm given in Table 8.8. These means were combined with the same utility values for (u_{I1}, u_{I2}) that were considered previously i.e. (0.8,0.8), (0.8, 0.5) and (0.5,0.8) and a utility value of 0.2 was assumed for the dropout state. The mean *QALW*(15) for each treatment group

were estimated from these utility values and the means given in Table 8.8 using [8.13] and the results are given in Table 8.9. Standard errors and 95% confidence intervals were obtained from 1000 bootstrap samples.

Table 8.8 Restricted mean survival times (and standard errors) in weeks spent in each health state in the MIC study

Health State	CT arm (N=65)	PAL arm (N=42)
Ill1	5.35 (0.66)	4.91 (0.84)
Well1	5.00 (0.60)	5.39 (0.88)
Ill2	1.64 (0.33)	1.02 (0.30)
Well2	0.28 (0.16)	0.14 (0.14)
Dropout	1.60 (0.35)	1.33 (0.42)

Table 8.9: Mean *QALW*(15) on each treatment arm of the MIC study with standard errors and 95% confidence intervals, for different utility values with dropout state included (u_D=0.2)

	CT (N=65)			PAL (N=42)		
Utility values for (u_{I1}, u_{I2})	Mean *QALW*(15)	Standard Error	95% Confidence Interval	Mean *QALW*(15)	Standard Error	95% Confidence Interval
(0.8, 0.8)	11.19	0.40	10.42-11.97	10.55	0.60	9.36-11.73
(0.8, 0.5)	10.70	0.39	9.94-11.46	10.24	0.59	9.09-11.40
(0.5, 0.8)	9.59	0.44	8.71-10.45	9.07	0.67	7.76-10.39

In comparison to the results given in Table 8.7, the inclusion of the dropout state reduces the mean quality-adjusted survival times gained by patients within 15 weeks from trial entry on both treatment arms. The comparison between treatment arms is similar to that without the dropout state with CT on average having greater quality-adjusted time than PAL. The difference between treatment groups appears to be reduced with the inclusion of the dropout state and thus remains statistically non-significant at the 5% level.

8.7 Group-Based Approach Using Actual Values: Integrated Quality-Survival Product

8.7.1 Method for Quality of Life and Survival Data

An alternative method that directly combines longitudinal quality of life data with survival data at a group level has been proposed by a number of authors, each advocating the same model but suggesting alternative approaches to estimation (Glasziou et al 1998, Ganiats et al 1995, Hwang et al 1996). The method has been referred to as the *integrated quality-survival product* (IQSP) and applied to data from cancer clinical trials (Beacon 1996). It is comparable to that proposed for the analysis of censored cost data (Lin et al 1997).

The method multiplies the survivor function $S(t)$ by the quality of life function $Q(t)$ for the group, where $S(t)$ is the proportion of subjects that survive to time t and $Q(t)$ is a summary measure of the quality of life of those survivors. In this way a quality-adjusted survival curve is created for the group. The area under this curve, usually calculated for a restricted time period, say up to time L, gives the mean $QALY$ for the group for this period, thus

$$QALY(L) = \int_0^L Q(t)S(t)dt \qquad [8.14]$$

The key decision is what estimators to use for the quality of life and survivor functions, $Q(t)$ and $S(t)$ respectively in [8.14]. The survivor function can be estimated from the sample of survival data using standard methods such as the Kaplan-Meier product-limit estimator, the life-table method or by fitting a parametric model. There are a number of options for the quality of life function. One option is to estimate it using a model for quality of life over time fitted to the observed data. For example, a simple linear regression model fitted to all the available data could be used as an estimate of the quality of life function (Beacon 1996). Alternatively a lowess or kernel-type smoother could be applied to the sample to estimate mean quality of life for the group over time

(Hwang et al 1996, Beacon 1996). A similar, but simpler approach, is to calculate the mean quality of life of survivors at the scheduled assessment times and the quality of life function over continuous time can be created by connecting these estimates using either a step function or by assuming a linear change (Glasziou et al 1998). If quality of life assessments are not taken at scheduled times, or the use of actual rather than scheduled assessment times is preferable, then the quality of life curves for each individual can be used to determine the quality of life at any time point t_j and the mean of these values across all individuals that are alive and uncensored at this time gives an estimate of the quality of life function at time t_j, that is

$$\hat{Q}(t_j) = 1/n_j \sum_{i=1}^{nj} q_i(t_j) \qquad [8.15]$$

where $q_i(t_j)$ is the estimate of quality of life at time t_j for surviving and uncensored individual i (i=1 to n_j). Having used individual values to calculate $\hat{Q}(t_j)$ at a number of discrete time points, it is still necessary to join these values using say a step or linear function so that $\hat{Q}(t)$ is a continuous function over time.

Having estimated the survivor and quality of life functions, the quality-adjusted survival curve is created by calculating their product and the area beneath this curve gives an estimate of the mean $QALY(L)$. Clearly if the estimators for both survival and quality of life are simple functions of time then they can be multiplied and the integral of the quality-survival product can be calculated for time from 0 to L to give an estimate of the mean $QALY(L)$. However, the most obvious estimator for the survival function is the Kaplan-Meier estimate, which is non-parametric and cannot be expressed as a simple function of time. The Kaplan-Meier method estimates the survival function $S(t_j)$ at each death time t_j (j=1 to k) and connects these estimates using a step function, i.e. it assumes the survival function remains constant until the next death time. With this as a survival function, the estimator for $QALY(L)$ becomes:

$$\hat{QALY}(L) = \sum_{j=0}^{k}\left[\hat{S}(t_j)\int_{t_j}^{t_{j+1}}\hat{Q}(t)dt\right] \qquad [8.16]$$

where t_0=0 and t_{k+1}=L.

If the estimator for the quality of life function is also piecewise constant between death times t_j, then the estimator for $QALY(L)$ becomes

$$\hat{QALY}(L) = \sum_{j=0}^{k}\left[\hat{S}(t_j)\hat{Q}(t_j)(t_{j+1}-t_j)\right] \qquad [8.17]$$

Thus having established the death times t_j (j=1 to k) of the patients, individual quality of life curves can be used to estimate the quality of life at each death time $q_i(t_j)$ (j=1 to k), with values set to missing for all patients who died at this time or previously. $\hat{Q}(t_j)$ can then be estimated as the mean of the values for the survivors at these times and by assuming a step function, the quality of life estimator will be piecewise constant between death times. A valid alternative to using step functions would be to assume a linear change between time points for both of the functions (Hwang et al 1996).

The standard error for the mean $\hat{QALY}(L)$ is mathematically complicated and bootstrapping is recommended as the method for estimation (Glasziou et al 1998, Hwang et al 1996), enabling confidence intervals to be calculated and hypothesis tests to be carried out. The method of bootstrapping (Hinckley 1988, Efron and Tibshirani 1993) was described earlier in Section 8.6.1.

8.7.2 Application to the MIC and ESPAC Studies

In the MIC study, the aim of the analysis was to estimate the mean quality-adjusted life weeks within 15 weeks of entry to the trial, $QALW(15)$, for each treatment arm using the IQSP. In the ESPAC study, quality-adjusted life months could theoretically be estimated for the whole follow-up period but with the small number of patients contributing to the survivor and hazard functions in the later stages it was decided to restrict the analysis to

24 months from date of entry and estimate *QALM*(24) on each treatment arm. At this point the number of patients still included in estimation of the survivor and quality of life functions was 48 (30 CT and 18 NoCT) and 20 (13 CT and 7 NoCT) respectively.

For both studies, a non-parametric survivor function was chosen to be combined with the quality of life function in the IQSP. Kaplan-Meier estimates of the survival function were calculated for each death time and, in the standard way, these estimates were assumed to remain constant until the next death time creating a stepped function. The survivor functions for the two treatment arms in the MIC study are shown in Figure 3.2b and those for the ESPAC study are shown in Figure 3.3. In the MIC study, within 15 weeks from trial entry, there were 12 deaths occurring at 10 distinct times on the CT arm and 16 deaths occurring at 15 distinct times on the PAL arm. In the ESPAC study, there were 47 deaths occurring at 45 distinct times within the 24-month analysis period on the CT arm and 58 deaths occurring at 55 distinct death times on the NoCT arm.

The non-parametric Kaplan-Meier survivor function was combined in the IQSP with quality of life functions estimated from the observed values of *GQS* and *GHSS* over time for the MIC and ESPAC studies respectively. Although these measures are not utilities, they can be transformed onto a 0 to 1 scale by dividing the measures by 100 and treated as utilities in the analysis (see Section 4.2.4). The quality of life function was estimated from a series of mean quality of life values determined from either (a) the mean quality of life at each scheduled time point, or (b) the mean quality of life at each death time determined from individual patient quality of life curves.

In option (a), the means at scheduled time points were simply calculated from the observed values available at each of those times. The quality of life function for each treatment arm is estimated by assuming linear changes between the means. These were shown earlier in Figure 6.4a for the MIC study and Figure 6.5 for the ESPAC study. The quality of life function needs to be a continuous function over the full analysis period and as can be seen in Figure 6.4a, this is not available for the MIC study. The last scheduled time point was 12 weeks for the CT arm and 9 weeks for the PAL arm. To complete the functions during the 15-week study period, the means at these last scheduled time points were carried forward to the 15-week cut-off time. This

assumption has greater impact on the PAL arm than the CT arm.

In option (b), the mean quality of life at each death time within each treatment group was estimated from the individual values obtained from the patient quality of life curves as shown in Figures 6.1 and 6.2 for the MIC and ESPAC studies respectively. The individual curves include only observed values of quality of life over time with linear changes assumed between values. The individual curves are curtailed at the patients' final assessment. The means at each death time are joined using a step function, such that the steps coincide with those on the Kaplan-Meier survivor function.

In both options (a) and (b), although patients remain in the analysis during the time between their last assessment and death, they do not contribute to the quality of life function during this time. This method essentially assumes that the missing data during this period are missing at random. If the dropout of patients from the quality of life study prior to death is believed to be informative, then this approach will give biased results. Methods for dealing with dropout prior to death, in addition to dropout directly due to death, are discussed in Section 8.7.4 and may be more appropriate than the approach taken here.

The quality-adjusted survival functions for each treatment group were created by multiplying the survivor and quality of life functions together and calculating the areas under these curves to give for each group the estimated mean $QALW(15)$ in the MIC study and $QALM(24)$ in the ESPAC study. For option (a), the area was calculated using [8.16], whilst for option (b) calculations were based on [8.17]. To estimate the standard error of the mean quality-adjusted survival time, 1000 bootstrap samples were taken from the observed data on each treatment arm and standard errors were estimated from the distribution of bootstrap sample means. There is no standard software available for performing quality-adjusted survival analysis using IQSP and data manipulation and analysis was programmed in SAS (SAS Institute Inc 1989) with bootstrapping carried out using a SAS macro. The results are given in Tables 8.10 and 8.11 for MIC and ESPAC studies respectively.

Table 8.10: Mean *QALW*(15) (and standard errors) for each treatment arm of the MIC study using Kaplan-Meier survivor function, unadjusted for quality of life and adjusted for two different quality of life functions

Survivor Function	Quality of Life Function	CT Mean *QALW*(15) (standard error)	PAL Mean *QALW*(15) (standard error)
Kaplan-Meier estimates stepped	No adjustment for quality of life	13.90 (0.35)	12.80 (0.56)
	Mean *GQS* at scheduled time points - linear changes	11.47 (0.34)	9.90 (0.54)
	Mean *GQS* at death times from individual curves – stepped	11.09 (0.67)	9.66 (0.74)

Table 8.11: Mean *QALM*(24) (and standard errors) for each treatment arm of the ESPAC study using Kaplan-Meier survivor function, unadjusted for quality of life and adjusted for two different quality of life functions

Survivor Function	Quality of Life Function	CT Mean *QALM*(24) (standard error)	NoCT Mean *QALM*(24) (standard error)
Kaplan-Meier estimates stepped	No adjustment for quality of life	17.29 (0.79)	15.02 (0.78)
	Mean *GHSS* at scheduled time points - linear changes	11.41 (0.71)	9.64 (0.68)
	Mean *GHSS* at death times from individual curves – stepped	11.01 (0.65)	9.52 (0.68)

In Table 8.10, the results for the MIC study include the area under the unadjusted Kaplan-Meier survivor function giving the mean survival time within 15 weeks, which was previously presented in Section 3.3.3 and shows that patients on the CT arm have greater mean survival time than those on the PAL arm. After adjusting these estimates for the quality of life experienced during this time, the CT arm still has superior quality-adjusted survival compared to the PAL arm but 95% confidence intervals overlap indicating that this difference is statistically non-significant at the 5% level. The estimates of mean *QALW*(15) are slightly higher with this group-based approach in comparison to the subject-based approach (see Table 8.3) and the difference between

treatment groups slightly less (1.57 and 1.43 compared to 1.90).

In Table 8.11, the results for the ESPAC study show that on both treatment arms there is a reduction of approximately 6 months when adjusting the survival time over 24 months for the quality of life experienced. The results for the two different forms of quality of life function were similar. The mean *QALM*(24) was greater for the CT group compared to the NoCT group, with chemotherapy giving patients on average between 1.5 and 2 extra quality-adjusted life months within the 24 months from entry to trial but 95% confidence intervals overlap indicating that this difference is not statistically significant at the 5% level. In this example, the results obtained with the group approach give slightly higher estimates for mean *QALM*(24) than those obtained using the subject-based approach (see Table 8.4) and with slightly greater differential between treatments.

8.7.3 Alternative Interpretation of Quality-Adjusted Survival Analysis

The IQSP has been interpreted as a method for calculating quality-adjusted survival time, in the sense that the quality of life function in [8.14] is considered as a weight for the survivor function. This enables the mean survival time for a group of patients to be down-weighted for the quality of life experienced. The survivor function however in the IQSP, which is on a 0 to 1 scale, could be considered as a weight for the quality of life function. In this way, the IQSP could be considered as calculating the *survival-adjusted quality-of-life*. This interpretation reflects the fact that the method adjusts for the dropout of patients due to death in the analysis of quality of life, which is often the key aim of simultaneous analysis of the two endpoints. By interpreting the IQSP in this way, the extension of the methodology to deal with dropout prior to death can be meaningfully interpreted as *dropout-adjusted quality-of-life* (see Section 8.7.4).

Only one simple adjustment is required to the methodology to enable this alternative interpretation. The IQSP calculates the mean survival-adjusted quality-of-life score over the whole follow-up period. The survival-adjusted quality-of-life score would be more easily interpreted if it were expressed as a proportion of the maximum obtainable score during this time, which is 1×L, i.e. survived the whole study period with perfect quality of life. Thus the mean survival-adjusted quality-of-life (*SAQL*) within a restricted time

period defined by L is given by:

$$SAQL(L) = \frac{1}{L}\int_0^L Q(t)S(t)dt \qquad [8.18]$$

The mean survival-adjusted quality-of-life and standard errors, restricted to the appropriate time period, were estimated for each treatment arm in the MIC and ESPAC studies using the two different forms of the quality of life function that were considered in Section 8.7.2. The area under the quality of life function, unadjusted for survival was also calculated for comparison purposes. This unadjusted analysis is comparable to the standardised area under the curve (SAUC) summary measures analysis presented in Section 6.3.4. The results for the two studies are given in Tables 8.12 and 8.13

Table 8.12: Mean *SAQL*(15) (and standard errors) for each arm of the MIC study using two different quality of life functions, unadjusted for survival and also adjusted for the Kaplan-Meier survivor function

Quality of Life Function	Survivor Function	CT Mean *SAQL*(15) (standard error)	PAL Mean *SAQL*(15) (standard error)
Mean *GQS* at scheduled time points - linear changes	Unadjusted for survival	0.83 (0.0120)	0.78 (0.0215)
	Kaplan-Meier estimates - stepped	0.76 (0.0227)	0.70 (0.0357)
Mean *GQS* at death times from individual curves - stepped	Unadjusted for survival	0.80 (0.0482)	0.67 (0.0453)
	Kaplan-Meier estimates - stepped	0.74 (0.0444)	0.64 (0.0490)

Using the quality of life function defined by the means at scheduled time points and without adjusting for dropout due to death, patients on the CT arm of the MIC study on average achieved 83% (95% CI: 81-85) of their maximum attainable quality of life during 15 weeks from entry whilst those on the PAL arm on average achieved 78% (95% CI: 74-82) of their maximum during this time. This comparison is biased by the informative dropout of patients from the quality of life study due to death. Adjusting these results for dropout due to death by weighting the quality of life function by the survivor function reduces these values and increases the standard error to give 76% (95% CI: 72-80) on the CT arm and 70% (95%CI: 63-77) on the CT arm. The mean

SAQL(15) are further reduced and the standard errors are further increased using the alternative form of the quality of life function based on the individual curves. For both forms of the quality of life function, patients on the CT arm achieve a greater proportion of their maximum attainable quality of life than those on the PAL arm, although 95% confidence intervals overlapped each other.

Table 8.13: Mean *SAQL*(24) (and standard errors) for each arm of the ESPAC study using two different quality of life functions, unadjusted for survival and also adjusted for the Kaplan-Meier survivor function

Quality of Life Function	Survivor Function	CT Mean *SAQL*(24) (standard error)	NoCT Mean *SAQL*(24) (standard error)
Mean *GHSS* at scheduled time points - linear changes	Unadjusted for survival	0.66 (0.0265)	0.64 (0.0290)
	Kaplan-Meier estimates - stepped	0.48 (0.0295)	0.40 (0.0282)
Mean *GHSS* at death times from individual curves - stepped	Unadjusted for survival	0.64 (0.0280)	0.64 (0.0355)
	Kaplan-Meier estimates - stepped	0.46 (0.0270)	0.40 (0.0282)

In the ESPAC study, using the quality of life function defined by the means at scheduled time points and without adjusting for dropout due to death, patients on the two treatment arms achieved very similar proportions of their maximum attainable quality of life during 24 months from entry. On average, patients on the CT arm achieved 66% (95% CI: 61-71) whilst those on the PAL arm achieved 64% (95% CI: 58-70). This comparison is biased and adjusting for dropout due to death reduces the mean on both arms to 48% (95% CI: 42-54) and 40% (95% CI: 34-46) on the CT and NoCT arms respectively. The results using the alternative form of the quality of life function based on the individual curves are comparable to these results. Unlike the MIC study, the standard errors are similar for adjusted compared to unadjusted and for the two different forms of quality of life function.

8.7.4 Dealing with Dropout Prior to Death

Estimating the quality of life function over time for a group of patients using only the observed quality of life data assumes that all missing data are missing at random. In Section 8.7.3 the missing data that occur due to death is accounted for by weighting the quality of life function by the survivor function. Patients however that drop out of the quality of life study prior to death will be included in the analysis until the point that they die but will not contribute to the quality of life function during the time between their last assessment and death. The estimation of the quality of life function therefore assumes that the missing data between dropout and death are missing at random. If this is not a valid assumption and it is assumed that patients generally drop out due to poor health then the results that adjust only for dropout due to death will be biased and the dropout prior to death should be accounted for in the analysis.

Choosing how to deal with the missing data between last assessment and death or end of study is not a trivial issue and may have implications on the results. One approach may be to impute values of quality of life for the dropout period. For example, assuming a linear change between the value at the last assessment and the value of zero at death may provide adequate imputed values for the dropout period. Imputing single values for missing data has the advantage of being a simple approach but underestimates the variability in the data and although methods such as multiple imputation would help to address this, they may overcomplicate the analysis.

One option, which has not previously been proposed, is to calculate the IQSP using dropout or death, whichever occurs first, as the event rather than death alone. Thus the dropout-free survivor function (as described in Section 5.8) is used in place of the survivor function in the IQSP given in [8.18]. The mean *dropout-adjusted quality-of-life (DAQL)* within the restricted time period defined by L is given by:

$$DAQL(L) = \frac{1}{L}\int_0^L Q(t)DFS(t)dt \qquad [8.19]$$

where DFS is the dropout-free survivor function.

The dropout-free survivor function for the MIC study was defined in Section 5.8.1 and Kaplan-Meier estimates of the function for each treatment group are shown in Figure 5.2. In summary, if a patient does not complete their final scheduled assessment (5^{th} assessment on CT arm and 4^{th} on PAL arm) then they are defined as having an event three weeks after their last recorded assessment, providing they survive this time. Any patient who dies within 3 weeks of the last recorded assessment is defined as having an event at their time of death. All other patients have censored dropout times either 3 weeks after their last completed assessment or at 15 weeks from trial entry, whichever is first. Given this definition of a dropout event, there were 27 events on the CT arm occurring at 22 distinct times and 20 events on the PAL arm occurring at 17 distinct times. The dropout-free survivor function is combined with the two different forms of quality of life function using [8.19] and the results are presented in Table 8.14. For comparison purposes, the previous results from Section 8.7.3 for the unadjusted quality-of-life and the survival-adjusted quality-of-life are also included.

Table 8.14: Mean *DAQL*(15) (and standard errors) for each arm of the MIC study using two different quality of life functions, adjusted for Kaplan-Meier dropout-free survivor function in comparison to unadjusted and adjusted for survivor function

Quality of Life Function	Adjustment Function	CT Mean *DAQL*(15) (standard error)	PAL Mean *DAQL*(15) (standard error)
Mean *GQS* at scheduled time points - linear changes	Unadjusted	0.83 (0.0120)	0.78 (0.0215)
	Kaplan-Meier estimates for survival	0.76 (0.0227)	0.70 (0.0357)
	Kaplan-Meier estimates for dropout-free survival	0.69 (0.0273)	0.59 (0.0400)
Mean *GQS* at death times from individual curves - stepped	Unadjusted	0.80 (0.0482)	0.67 (0.0453)
	Kaplan-Meier estimates for survival	0.74 (0.0444)	0.64 (0.0490)
	Kaplan-Meier estimates for dropout-free survival	0.66 (0.0264)	0.53 (0.0401)

Adjusting for dropout in addition to death in the MIC study, further reduces the mean quality of life experienced over 15 weeks. The results from the two different quality of life functions are comparable. In both cases the CT arm has an additional 10% of their maximum attainable quality of life compared to the PAL arm. The difference between treatments becomes larger as dropout due to death and then additional dropout prior to death are adjusted for in the analysis.

The dropout-free survival function for the ESPAC study was defined in Section 5.8.2 and Kaplan-Meier estimates of the function for each treatment group are shown in Figure 5.3. In summary, if the last recorded assessment for a patient is more than 3 months prior to their last survival follow-up then the patient is recorded as having a dropout event 3 months after their last recorded assessment. For those patients who die within 3 months of their last recorded assessment, the event is recorded at their time of death. The remaining patients are recorded as having censored dropout times at their last survival follow-up. Since the analysis is restricted to 24 months from trial entry, all events that occur after this time are censored at 24 months. Given this definition of a dropout event, there were 62 events on the CT arm occurring at 61 distinct times and 68 events on the NoCT arm occurring at 67 distinct times. The dropout-free survivor function is combined with the two different forms of quality of life function using [8.19] and the results are presented in Table 8.15. For comparison purposes, the previous results from Section 8.7.3 for the unadjusted quality-of-life and the survival-adjusted quality-of-life are also included.

Adjusting for dropout in addition to death in the ESPAC study, further reduces the mean quality of life experienced over 24 months. The results from the two different quality of life functions are comparable. In both cases the difference between treatments is small with the CT arm having an additional 6% of their maximum attainable quality of life compared to the NoCT arm.

Table 8.15: Mean *DAQL*(24) (and standard errors) for each arm of the ESPAC study using two different quality of life functions, adjusted for Kaplan-Meier dropout-free survivor function in comparison to unadjusted and adjusted for survivor function

Quality of Life Function	Adjustment Function	CT Mean *DAQL*(24) (standard error)	NoCT Mean *DAQL*(24) (standard error)
Mean *GHSS* at scheduled time points - linear changes	Unadjusted	0.66 (0.0265)	0.64 (0.0290)
	Kaplan-Meier estimates for survival	0.48 (0.0295)	0.40 (0.0282)
	Kaplan-Meier estimates for dropout and death	0.40 (0.0274)	0.34 (0.0243)
Mean *GHSS* at death times from individual curves - stepped	Unadjusted	0.64 (0.0280)	0.64 (0.0355)
	Kaplan-Meier estimates for death	0.46 (0.0270)	0.40 (0.0282)
	Kaplan-Meier estimates for dropout and death	0.39 (0.0253)	0.33 (0.0230)

8.8 Discussion and Critical Review of Quality-Adjusted Survival Analysis

Quality-adjusted survival analysis provides a relatively straightforward approach for the simultaneous analysis of quality of life and survival data. Although this type of analysis has been widely used in clinical research, its application has almost entirely related to health states defined using clinical criteria and the application to quality of life data collected on patients in a clinical trial is limited (Allen-Mersh et al 1994, Beacon 1996). This thesis provides the most comprehensive review to date of the different approaches to quality-adjusted survival analysis in their specific application to longitudinal quality of life data collected on patients in a clinical trial.

The different approaches to quality-adjusted survival analysis can be summarised into four distinct categories, depending on whether the quality of life and survival data are combined at the subject or the group level and depending on whether the actual values

of quality of life data collected on patients over time are used directly in the analysis or indirectly by using them to allocate patients to different health states for periods of their follow-up time. These four different approaches each have their own advantages and disadvantages and each are more appropriate for some situations than others, as discussed in detail in this section.

All approaches are based on the concept of combining quality of life and survival data in a QALY model, with the well-known TWiST and Q-TWiST models being special forms of the general model. The general form of the QALY model that has been used in this thesis is based on a number of assumptions: the independence of the utility value for a health state to the time spent in the health state; the independence of the utility value for a health state to previous or future health states and to remaining amount of life; and the equivalent valuation of all follow-up time. These assumptions have been discussed in relation to a Q-TWiST model (Till and de Haes 1991, Gelber and Goldhirsch 1991). A more general form for the model combining quality of life and survival data has been discussed (Cole et al 1994) and may overcome some of the assumptions of the QALY model. One of the advantages of the approaches discussed in this thesis is that no distributional assumptions are required for the quality of life or survival data although the IQSP can be partly or fully parametric if desired.

The approaches that directly incorporate the actual values of quality of life with survival either at a subject or group level ideally require the quality of life measures over time to be utilities. Questionnaires that yield utility values for patients over time are not widely used in clinical trials but the approaches can be applied to any measure from a quality of life instrument provided that it is transformed onto a 0 to 1 utility-type scale. The choice of transformation will generally be linear but in some cases it may be more appropriate to use one that is non-linear (Beacon 1996). The interpretation of the resultant measure as a QALY however may be questionable if the quality of life measures are not true utilities. As the methods described in this chapter become more widely used, so clinical trials will more routinely use questionnaires that can allocate utility values to patients over time as well as or instead of questionnaires that yield descriptive measures, and the application of the actual value approaches will be more valid.

If the longitudinal quality of life outcome for analysis is not a utility and cannot easily be translated into a utility then the approaches using actual values will not be applicable. In such situations the health state approaches, in which the values of quality of life over time are used purely to allocate patients to certain health states during their follow-up time, will be preferable. This approach lends itself well to categorical outcomes and to situations where multiple measures of quality of life are relevant, since the health states can be defined in terms of any values for any number of measures. For valid QALY outcomes the utilities for the quality-of-life-defined health states need to be known. This may be possible if an external valuation study is carried out. If however the utilities are not known then the range of utilities can be investigated in a sensitivity analysis. The greater the number of different utilities in the QALY model, the more unmanageable the sensitivity analysis becomes.

The subject-based approach of combining quality and quantity of life into individual QALYs either using actual values or health states is the simplest form of quality-adjusted survival analysis. The method is straightforward to understand and implement and is therefore accessible for clinicians. It is a type of summary measures analysis and as such has the advantage that having calculated the QALY for each subject, standard univariate methods of analysis can potentially be applied to the QALY endpoint to determine differences between treatment groups for example. If the analysis contains censored survival times then the QALY endpoint will also be censored and, since standard survival techniques would be invalidated for the QALY endpoint due to informative censoring, the analysis becomes problematic. In some situations it may be possible to restrict the period of analysis to one in which all patients have full survival follow-up thus eliminating the censored survival times but if this is not desirable or possible then a subject-based approach may not be a valid option. The choice of an upper time limit is subjective, and the possible inclusion of available quality of life and survival information after this time in the calculation of the QALY within the analysis period needs to be considered carefully to ensure that no bias is introduced. In many studies restricting the analysis to an upper time limit will merely minimise, rather than eliminate, the number of censored QALYs and in this case, although imputing values for the censored QALYs in a sensitivity analysis is an option, in general a subject-based approach is not advisable if censored data are present. The alternative to a subject-based

approach is a group-based approach and since, in general, censored survival times are likely to be present in any analysis, this approach will be the most applicable. One problem with the group-based approach is the difficulty in calculating standard errors, although software to perform bootstrapping is becoming more readily available.

A number of assumptions need to be made to create subject-based utility-type curves over time from which QALYs can be calculated. Different assumptions will produce different curves and give different results (Ganiats et al 1996) and the assumptions may be questionable if the time spans between assessments are large. Although the method deals with the problem of missing quality of life data due to death by allocating a value of zero to all time beyond death, there are a number of options for dealing with missing data resulting from dropout from the quality of life study prior to death. Values for these missing data can be imputed, in particular if the reasons for dropout are known or clinical data are available an appropriate value can be allocated to the time spent as a dropout to reflect the quality of life during that time. Other simple methods of imputation such as last value carried forward may not be adequate and more sophisticated methods such as multiple imputation need to be considered and their use within this context needs to be assessed. By assuming a linear decrease from the value at the last assessment to the value of zero at death, as in this thesis, the method is indirectly accounting for dropout, since it is assuming that patients gradually deteriorate during their course of their dropout time. Some authors who use an area under the curve approach to analyse longitudinal quality of life data, deal with dropout, including that due to death, by dividing the area by the length of the observation time from trial entry to last assessment and compare treatments using a standardised area under the curve (Qian et al 2000) This is effectively an analysis of the distribution of quality of life during the time that they participate in the study. The problem with this approach is that individuals with a short follow-up time will have equal weighting to those with a long follow-up time (Ganiats et al 1995). In particular, if death is not accounted for then, for example, an individual with a quality of life value of 0.7 for 6 months and then dies will be treated with equal weight as someone who survives for the whole 2 year analysis period with a quality of life value of 0.7.

Partitioned survival analysis is generally difficult to apply to longitudinal quality of life data because of the need for progressive health states. It may be possible to overcome this problem by specifying different phases of the same state, as was done in the MIC study, but this can become clumsy and may lose clinical meaning. There is no unique way to divide the survival time of patients into periods of differing quality of life and the accuracy depends on the frequency of quality of life assessments. Different divisions should be considered as part of a sensitivity analysis. Additional dropout prior to death can easily be accounted for by incorporating the time spent as a dropout into the model with an appropriate weighting to reflect quality of life.

In general, methods for quality-adjusted survival analysis require the analysis to be restricted to an upper time limit, as methods are based on the calculation of areas under curves. In some situations, as with the MIC study, the quality of life data may only be collected for a limited amount of time, and the period of analysis will automatically be restricted to an upper time limit. Otherwise, the effect of imposing an upper time limit should be investigated using a sensitivity analysis. Such methods have been proposed specifically in relation to the Q-TWiST model (Glasziou et al 1990, Gelber et al 1995). A parametric approach to quality-adjusted survival analysis as suggested by Cole et al (1994) may overcome this limitation. The group-based approaches both allow parametric forms for the functions under which areas need to be calculated and this may enable the analysis to be unrestricted in terms of time.

The IQSP provides a simple method for combining quality of life and survival data at the group level whilst dealing with censored survival times. It is preferable to the method of partitioned survival analysis since it is not based on progressive health states, which generally are difficult to define in relation to quality of life data. Instead the method directly incorporates longitudinal quality of life data collected on patients in a quality-adjusted survival analysis. The inclusion of covariates in the analysis has not been considered and further research is required to extend the method to incorporate covariates.

There are number of choices to be made when implementing the IQSP, in particular estimates for the quality of life and survival functions need to be chosen and the analysis

in general will need to be restricted to an upper time limit. Sensitivity analysis is recommended to assess the robustness of the results to these choices. The method is flexible in terms of the choice of functions with either or both being non-parametric or parametric. The use of parametric functions for quality of life and survival may enable quality-adjusted survival for the whole follow-up period to be calculated rather than just for a fixed period of time. The most obvious choice for the survival function is the Kaplan-Meier survival curve. For the quality of life function, using individual's functions of quality of life over time to estimate the average quality of life of survivors at a number of given time points is the most flexible approach in that it allows for differing assessment times. For both the quality of life and survival functions, assumptions are required to realistically map the values estimated at discrete time points to their estimated course in continuous time. The method does not fully account for the uncertainty in the estimation of the two functions and in particular the increase in uncertainty in the estimation of the functions over time that is related to the reduction in patient numbers.

By interpreting the results from the IQSP as mean survival-adjusted quality-of-life rather than mean quality-adjusted survival time, the results can be expressed as the proportion of the maximum obtainable quality of life during the period which may be more clinically intuitive than quality-adjusted survival time. In addition, the method can easily be extended to deal with dropout by including the dropout-free survivor function rather than the survivor function. Unlike the subject-based approach, the individual utility-type curves only need to include the observed data and this reduces the need for imputation. The method of quality-adjusted survival analysis using IQSP is a relatively new methodology and research in this thesis has shown that it is a useful approach for the simultaneous analysis of quality of life and survival. This thesis has developed the method to deal with additional dropout prior to death, which further indicates the potential of the method for the analysis of quality of life data with informative dropout.

CHAPTER 9: MULTISTATE MODELS FOR QUALITY OF LIFE AND SURVIVAL DATA

9.1 Introduction

The aim of this chapter is to describe the method of multistate modelling and develop the methodology for application to the simultaneous analysis of quality of life and survival data. Multistate models were first proposed for use in a medical context in 1951 (Fix and Neyman 1951) and have more recently been advocated in a number of reviews and discussions regarding the analysis of quality of life data, as a possible means of analysing quality of life and survival data simultaneously (Fayers and Jones 1983, Schumacher et al 1991, Cox et al 1992, Abrams 1992a, Olschewski et al 1992). Some authors have discussed in more detail the use of this approach for the analysis of quality of life data in cancer clinical trials (Olschewski and Schumacher 1990) but there is little evidence of its application to such data.

The statistical background to multistate survival analysis is derived from the analysis of event-history data (Clayton 1988, Andersen and Keiding 2002) and stochastic processes (Cox and Miller 1965, Chiang 1980). Many of the theoretical aspects of multistate models fall within a counting process framework (Andersen and Borgan 1985, Fleming and Harrington 1991, Andersen et al 1993). The study of events occurring in individuals over time generates event-history data. In studies such as these, individuals can be thought of as occupying one of a finite number of states at any point in time and the movement between states can be described by conditional probabilities or transition rates. This dynamic process is known as a stochastic process. Quality of life assessment in clinical trials generates event-history type data, with events being defined as entry and exit from pre-defined health states. The movement of individuals between quality of life states can then be considered as a stochastic process and modelled accordingly.

Multistate modelling requires the definition of a finite number of health states, including death, that patients experience during the study. Defining these health states and the possible transitions between them describes the multistate model. The transition rates, which describe the movement between health states, can then be modelled, possibly

using covariates. In this way the time-dependent structure and dynamic nature of quality of life data can be incorporated into an analysis comparing treatments and the effect of explanatory variables on transition rates from one state to another can be investigated. Multistate models do not require any distributional assumptions about the quality of life data and may therefore provide a convenient approach for non-normally distributed or categorical quality of life measures, although defining states is not necessarily straightforward.

Multistate models have been applied in a variety of clinical settings such as diabetes (Andersen 1988, Marshall and Jones 1995), rheumatoid arthritis (Young et al 1999), liver transplantation (Hansen et al 1994), liver cirrhosis (Andersen et al 2000), bone marrow transplantation (Klein et al 1993, Klein and Shu 2002), heart transplantation (Kay 1982, Wu 1982), breast cancer (Kay 1984, Perez-Ocon et al 2001a and b), prostate cancer (Myers et al 1980), leukaemia (Chevret et al 2000) and HIV (Gentleman et al 1994, Sypsa et al 2001). The applications, however, all use clinical criteria rather than quality of life data to define health states. There are limited examples of applications of multistate models to quality of life data (Olschweski 1984, Abrams 1992b, Charlesworth and Skene 1997).

This thesis applies multistate models to the quality of life and survival data collected in the MIC and ESPAC studies, allowing the feasibility of the methodology for such data to be investigated. In addition, the development of such models to include a 'dropout' state may provide a means for overcoming the problem of informative dropout from the quality of life study. This approach for dealing with dropout in a longitudinal study has not previously been considered and is investigated here for the MIC and ESPAC studies. The multistate models that are considered for the MIC and ESPAC data are described in Section 9.2 and Section 9.3. These models are applied to quality of life and survival data in the MIC and ESPAC studies in Section 9.4. The models are extended in Section 9.5 to include a dropout state with application to the MIC and ESPAC data. A discussion and critical review of the use of multistate modelling for the simultaneous analysis of quality of life and survival is given in Section 9.6, with areas of potential future research highlighted.

9.2 Defining the Model

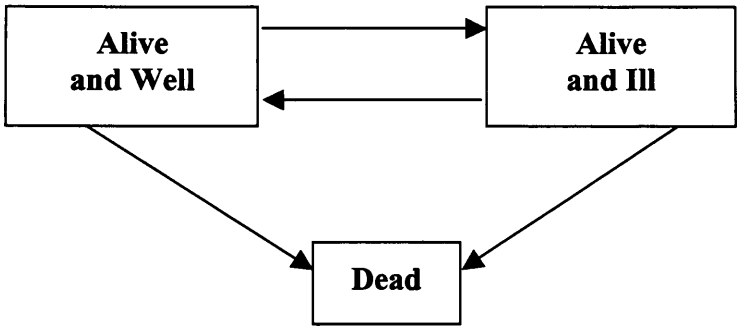
9.2.1 Health States

The set of health states chosen for the model should be clinically meaningful and fully describe the experiences of the patients. They should be mutually exclusive and exhaustive. The number of states should be restricted so that the model does not become overcomplicated and to ensure that the number of patients passing from one state to another is sufficient for adequate estimation of parameters in the model.

There are two main types of health states. A *transient* state is one that a patient can pass through during the course of their follow-up and an *absorbing* state is one that a patient cannot leave once it has been entered. The standard model for survival analysis corresponds to the simplest multistate model, where the patient can be in one of two possible states, a transient 'alive' state or an absorbing 'death' state. The competing risks model (David and Moeschberger 1978) is an extension of this two-state survival model and forms a multistate model with one transient alive state and several absorbing death states corresponding to different causes of death. In terms of modelling quality of life data it is more relevant to extend the simple two-state survival model so that there are several transient alive states and a single absorbing death state.

The simplest version of a multistate model for quality of life is the three-state illness-death model (also called the 'three-state disability model'), where there are two transient alive states: *well* and *ill*, and one absorbing death state. This model was used for the MIC and ESPAC data (see Figure 9.1).

Figure 9.1 Three-state illness-death model for the MIC and ESPAC studies



In the MIC study, two different models were considered, one where the patients' health states were determined from the global quality of life score (*GQS*) and one where they were determined from the malaise question (*MAL*). The *GQS* outcome was continuous in nature and to define health states, appropriate cut-off values needed to be determined. Patients were categorised as being in the *well* state if the patient had a value of *GQS* greater than 83, the median value over all questionnaires at all time points, and were categorised as being in the *ill* state if the patient had a value of *GQS* less than or equal to 83. For the second model, patients were defined as *well* if they had no malaise (*MAL*=0) and were defined as *ill* if they had malaise at any level (*MAL*=1, 2 or 3). For the ESPAC data the global health status score (*GHSS*) was used to define the health states with patients categorised as *well* if the patient had a value greater than 50 and *ill* if they had a value less than or equal to 50. The value of 50 was chosen to represent a clinically-determined cut-off rather than a data-driven one as was used in the MIC study. A score of 50 is obtained when a patient's overall physical condition and overall quality of life on average are rated half way between very poor and excellent (see Section 4.4.2).

For both studies, the health states used in the model could have been defined using different cut-offs or could have been based on other questions asked in the study, possibly multiple questions. More complex models with more than two alive states defined by quality of life values were considered, but in both studies such a formulation would have meant that there were too few transitions for adequate modelling of the data. The addition of a *dropout* state to the above models is considered in Section 9.5.

9.2.2 Health State Transitions

The movement between states can be described using either a *transition probability* or a *transition rate*. A transition probability is the likelihood of an individual moving from one state to another within a specified time period; a transition rate is the instantaneous potential of transition at any point in time (Miller and Homan 1994). Whilst the transition probability can take values between 0 and 1, the transition rate, sometimes called *transition intensity*, has no upper bound. In the simple two-state survival model, the transition rate from a transient alive state to an absorbing death state is the standard hazard rate function for the survival time distribution (see Section 3.2).

The two measures are related (Miller and Homan 1994) and, when the instantaneous transition rate r remains constant during a period of time, say t_1 to t_2, a transition probability p can be estimated using:

$$p = 1 - \exp\left[-r\left(t_2 - t_1\right)\right] \qquad [9.1]$$

In some situations, patients in the trial may experience 'recovery' during the follow-up time and in these circumstances the model may need to include reverse transitions, allowing the ability to return to a state previously occupied. For example, in a three-state illness-death model for the data from a liver cirrhosis trial (Andersen et al 1991), the authors allow patients with low prothrombin index to recover and return to the 'alive with normal prothrombin index' state and thus incorporate reverse transitions in their model. If reverse transitions are possible then repeat transitions may occur, that is a patient may experience a particular transition more than once during their follow-up. In some circumstances, such as when it is thought that the transition rate would be different for a repeat transition compared to an initial transition, it may be preferable to model repeat and initial transitions as separate events (Islam 1994). This would only be possible if the number of subjects experiencing repeat transitions was large enough for adequate modelling.

In the MIC and ESPAC studies, patients could move either way between the transient health states, sometime several times, until finally moving to the absorbing state of

death. The model therefore included reverse and hence repeat transitions as shown in Figure 9.1 by arrows in both directions between the alive states.

9.2.3 Assumptions of Underlying Stochastic Process

In the most general model, the transition rates would depend on the whole history of the patient. However, it is often reasonable to assume that at any time point the state currently occupied by a patient contains all the information relevant to a patient's future course. Under this assumption the model represents a Markov process. If the transition rates from each state are conditional on the duration of time spent in the state, often called the *sojourn time*, then the model represents a semi-Markov process (Clayton 1988).

It may be necessary to define the health states of the model so that the Markov assumption is valid. For example, in a three-state illness-death model, if the transition of patients between the well and death states depended on whether the patient had originally been in the ill state, then it would be necessary to create two well states: an 'always been well' state and a 'was ill but now well' state. In this way the validity of the Markov assumption could be retained.

If the transition rates remain constant over time, i.e. are independent of time, then the process is time-homogeneous, otherwise if they vary over time, i.e. are functions of time, then the process is time-nonhomogeneous. In some situations the transition rates may be piecewise constant such that they are constant over defined subdivisions of the follow-up time.

9.3 Modelling the Transition Rates

9.3.1 Cox Regression Model for Transition Rates

In a standard survival model, the transition rate from the alive state to the dead state is commonly represented by a Cox regression model (Cox 1972) as described in Chapter 7. The application of Cox regression models to the more general multistate framework

which allows several transient disease states between entry to study and death has been discussed (Kay 1982, Hsieh et al 1983, Cox 1984, Prentice et al 1981, Islam and Singh 1992). The exact times of transition from one state to another, or estimates of them, are needed for this type of analysis. Cox regression models have been used to model the transition rates in various multistate survival analysis applications (Wu 1982, Kay 1984, Andersen 1988, Klein et al 1993, Hansen et al 1994, Islam 1994, Young et al 1999, Andersen et al 2000, Chevret et al 2000, Perez-Ocon et al 2001a and b).

The transition rate from state k to state l for patient i, $h_{i[kl]}(t)$ can be modelled using a Cox regression model as follows:

$$h_{i[kl]}(t) = h_{0[kl]}(t)\exp(\underline{\beta}'_{[kl]}\underline{x}_{i[kl]}) \qquad [9.2]$$

where $h_{0[kl]}(t)$ is a baseline transition rate for the transition from k to l, $\underline{x}_{i[kl]}$ is a vector of covariates specific to that transition and $\underline{\beta}_{[kl]}$ is a vector of unknown regression coefficients specific to that transition. The model could be generalised to include time-dependent covariates by replacing $\underline{x}_{i[kl]}$ with $\underline{x}_{i[kl]}(t)$.

9.3.2 Markov or Semi-Markov

When modelling the transition rates, consideration needs to be given to whether the process is Markov or semi-Markov. Semi-Markov processes have been discussed in relation to multistate models (Lagakos et al 1978, Prentice et al 1981, Wu 1982, Hsieh et al 1983, Cox 1984, Chevret et al 2000, Andersen et al 2000). In a Markov process, as discussed in Section 9.2.3, the transition rate to another state depends only on the present state occupied, whilst in a semi-Markov process the transition rate is also dependent upon the duration of time in the present state (Andersen et al 1993).

The type of process is related to the time scale on which time in the model for a transition rate is measured, i.e. it relates to the 'time at which the clock starts'. If, in modelling the transition rate from state k to state l, time t is measured from the time of entry to the study (i.e. t=0), then the model represents a Markov process, since duration of time in state k is not included. Otherwise, if time is measured in the model from time

of entry to state k, say w in relation to study entry time, then time is included in the model as t-w rather than t and, with sojourn time now included, the model represents a semi-Markov process (Kay 1984). In a semi-Markov process, the clock is effectively 'reset to zero' every time a state is entered (Clayton 1988). An alternative way to fit a semi-Markov model is to model t but include sojourn time as a time-dependent covariate in the model (Kay 1984, Andersen et al 2000).

Information regarding the history of the process prior to entering state k may be included as covariates. In particular, when individuals can experience the same transition more than once, the transition rate from state k to state l may depend on aspects such as whether state k has been occupied before, the number of times state k has been visited before and the total time previously spent in state k. If the covariates in the model do not include information regarding states prior to the current one, then the model implicitly assumes that the state changes form a Markov process (Kay 1982). The time from study entry to entry to state k could also be included as a covariate (Kay 1984). When the time origin is taken as date of entry to study, then the covariates containing information on history prior to entering state k will be time-dependent, whilst if the origin is taken as date of entry to state k, then the history is already determined at the origin and the information will be included as fixed covariates.

9.3.3 Form of Underlying Baseline Transition Rate

In modelling the transition rates the underlying baseline transition rate may be considered to be arbitrary (Kay 1982). In some circumstances, however, if transition times follow a particular distribution then it may be more efficient to assume a parametric form for the underlying baseline transition rate. The most commonly used distributions for survival data are the exponential and Weibull distribution, the exponential being just a special form of Weibull distribution (see Section 3.5). These distributions have been assumed for the baseline transition rates in multistate models (Beck 1979, Lagakos 1976, Young et al 1999), as well as piecewise versions of them (Wu 1982, Perez-Ocon et al 2001b).

If an exponential distribution is assumed for the transition times from state k to state l then the underlying baseline transition rate is constant:

$$h_{0[kl]}(t) = \lambda_{[kl]} \quad [9.3]$$

and the model is thus *time-homogeneous*, whilst for a Weibull distribution the underlying baseline transition rate takes the form:

$$h_{0[kl]}(t) = \lambda_{[kl]}\, \gamma_{[kl]}\, t^{\gamma[kl]-1} \quad [9.4]$$

and the model is *time-nonhomogenous*. If an exponential distribution is assumed for the transition times then, since the underlying transition rate is not dependent on time, the models under the assumption of a either a Markov or a semi-Markov process will be equivalent. This, however, is not the case for other distributions such as Weibull, since the baseline transition rate changes over time.

9.3.4 Estimating Parameters

Estimation of the $\underline{\beta}_{[kl]}$ regression parameters in model [9.2] is achieved using the ideas of partial likelihood as described in Section 7.3 and introduced by Kay (1982). If $t_1 < t_2 < t_3 < \ldots < t_T$ represents the T ordered transition times and transition time t_j represents the time when the subject with covariate vector x_j moves from state k to state l, then the likelihood for this event is given by:

$$L_j(\underline{\beta}_{[kl]}) = \frac{\exp(\underline{\beta}'_{[kl]}\underline{x}_{j[kl]})}{\sum_{s \in R(tj,k)} \exp(\underline{\beta}'_{[kl]}\underline{x}_{s[kl]})} \quad [9.5]$$

where $R(t_j, k)$ represents the subjects who are in state k at time t_j and therefore at risk of transition to state l.

The partial likelihood is the likelihood of the observed events, which is the product of the likelihoods for the T events. Since each element that makes up the partial likelihood (as specified in [9.5]) only contains the regression parameters for a specific transition, the partial likelihood can be factorised into a partial likelihood function for each transition. The partial likelihood function for the transition from k to l is given by:

$$PL(\underline{\beta}_{[kl]}) = \prod_{j \in T[kl]} \left(\frac{\exp(\underline{\beta}'_{[kl]} \underline{x}_{j[kl]})}{\sum_{s \in R(tj,k)} \exp(\underline{\beta}'_{[kl]} \underline{x}_{s[kl]})} \right) \qquad [9.6]$$

where $T[kl]$ represents the set of times at which transitions from state k to state l occurred.

Thus the partial likelihood for a transition from state k to state l is identical to the partial likelihood for the standard Cox regression model, with transitions from k to states other than l treated as censored data (Kay 1982). Thus, for each transition in the model, estimates of the β regression coefficients are obtained by maximising the relevant partial likelihood. In situations where individuals may experience the same transition more than once, the partial likelihood is still valid (Kay 1982) but the likelihood function assumes that the events are independent. A time-dependent covariate can be included in the model to account for the number of times previously in the state.

Significance testing and confidence intervals for the β parameters can be based either on the asymptotic normality of the distribution of the estimators or on the large sample likelihood ratio test for nested models (Kay 1982), where changes in minus twice the log likelihood are compared to a chi-square distribution. Akaike's information criterion (AIC) can be used to compare non-nested models (Collett 1994). AIC is calculated as:

$$AIC = -2 \log \hat{L} + \alpha q \qquad [9.7]$$

where $\hat{L}$ is the maximised likelihood, q is the number of unknown β regression parameters in the model and α is a pre-determined constant with a value of 3 being equivalent to a 5% significance level. The smaller the value of AIC, the better the model.

If a parametric form is assumed for the underlying transition rate then the parameters can be estimated by maximising the full likelihood (see Section 7.3). As with the partial likelihood, the full likelihood factorises into the terms for each transition. The

parameters for each transition can therefore be estimated separately by maximising the relevant part of the full likelihood with transitions to states other than the one of interest being censored for that transition (Andersen et al 1993).

9.4 Modelling Transition Rates in the MIC and ESPAC Studies

9.4.1 Translating the Longitudinal Quality of Life Data into Health State Transitions

Three-state illness-death models were used to describe the quality of life and survival data in the MIC and ESPAC studies (see Figure 9.1 and Section 9.2.1). In the MIC study, quality of life data were only collected during the treatment period of the trial and hence the analysis was restricted to the 15-week period from study entry (see Section 4.5.3) with all patients still alive at 15 weeks censored at this time. The ESPAC study collected quality of life from trial entry to death and therefore in theory the model could be applied to the whole follow-up period. Quality of life data however become sparser over time and were available on only a few patients after 24 months and hence the analysis was restricted to 24 months from study entry. All patients with survival follow-up longer than 24 months were censored at this time and all quality of life assessments after 24 months were excluded from the analysis.

Quality of life was assessed at distinct time points and assumptions were necessary to infer values over continuous time in order to estimate the exact transition date from one health state to another. Measures of quality of life were assumed to remain constant from one assessment to the next. If the value at an assessment reflected a different health state to the previous one then the time of the transition to that new health state was taken as the time of that assessment. This enabled the exact dates of transition between the two alive states to be estimated. Dates of death for patients in both studies were known and so exact dates for transitions to death were available. In general, sensitivity analysis should be used to assess the impact of the assumptions on the conclusions of the analysis. Other options considered included (i) assuming the change happened mid-way between assessments and (ii) for continuous measures of quality of

life assuming a linear change from one assessment to the next and estimating the time of transition using linear interpolation.

Another issue was how to deal with missing assessments due to dropout. Patients who dropped out of the quality of life study because of death were not a problem since their transition to death was included in the model. However, patients who dropped out prior to death were a problem, since the health states that they occupied between their last assessment and death or end of study time were not known. The easiest option was to assume that patients remained in the health state that they occupied at their last assessment until their death or end of study time and this is the approach we adopt here. This is equivalent to imputing values after dropout using a 'last value carried forward' approach. Other methods of imputation (see Section 5.5.3) could also be considered. An alternative strategy to imputation for patients who dropped out for reasons other than death is to censor the transition times at the date of dropout (see Section 5.8 for definitions of date of dropout). With the dropout generally informative, this is likely to result in informative censoring which would invalidate the models and hence this was not considered further. Another alternative strategy is to include *dropout* as a state in the model and this is considered in Section 9.5.

9.4.2 Models Fitted to the Data

In the MIC study, during the 15-week period from study entry, 85 transitions in total were experienced using the *GQS* variable and 102 using the *MAL* variable (see Table 9.1). In the ESPAC study, during the 24 months from study entry, 265 transitions in total were experienced based on *GHSS* (see Table 9.2). These are the data used to model the four transition rates. For each transition from a state, the number at risk is the number of occasions when patients are in that health state and at risk of transition to another. For example, in the ESPAC study, there were 163 at risk in the well state, which included 121 patients who occupied the state only once, 18 who occupied it on two separate occasions and 2 who occupied it three times (see footnotes to Tables 9.1 and 9.2). The number of transitions from well to ill and from ill to well, include a small number of patients who experience that transition more than once (see footnotes to Tables 9.1 and Table 9.2).

Table 9.1 Frequency of transitions in the 3-state model for the MIC data

Transition	Number at Risk			Number of Transitions		
	Total	CT	PAL	Total	CT	PAL
***GQS*-Defined Health States**						
Well→Ill	78[a]	53	25	28[c]	21	7
Ill→Well	88[b]	57	31	29[d]	22	7
Well→Dead	78[a]	53	25	4	0	4
Ill→Dead	88[b]	57	31	24	12	12
***MAL*-Defined Health States**						
Well→Ill	78[e]	52	26	38[g]	27	11
Ill→Well	105[f]	69	36	36[h]	27	9
Well→Dead	78[e]	52	26	5	2	3
Ill→Dead	105[f]	69	36	23	10	13

[a] 66 patients occupying state once and 6 occupying twice

[b] 64 patients occupying once and 12 occupying twice (1 occasion had 0 duration)

[c] 26 patients with one transition and 1 with two

[d] 25 patients with one transition and 2 with two

[e] 64 patients occupying once and 7 occupying twice

[f] 77 patients occupying once and 14 occupying twice

[g] 34 patients with one transition and 2 with two

[h] 32 patients with one transition and 2 with two

Table 9.2 Frequency of transitions in the 3-state model for the ESPAC data

Transition	Number at Risk			Number of Transitions		
	Total	CT	NoCT	Total	CT	NoCT
***GHSS*-Defined Health States**						
Well→Ill	163[i]	81	82	80[k]	42	38
Ill→Well	172[j]	94	78	80[l]	46	34
Well→Dead	163[i]	81	82	37	15	22
Ill→Dead	172[j]	94	78	68	32	36

[i] 121 patients occupying once,18 twice and 2 three times (1 occasion had 0 duration)

[j] 110 patients occupying once, 20 twice, 6 three times and 1 four times (3 occasions had 0 duration)

[k] 57 patients with one transition, 10 with two and 1 with three

[l] 61 patients with one transition, 8 with two and 1 with three

For both studies, the transition rates were modelled using Cox regression models as specified in [9.2] with only a covariate for treatment included as follows:

$$h_{i[kl]}(t) = h_{0[kl]}(t)\exp(\beta_{[kl]}G_i) \qquad [9.8]$$

where G_i represents the treatment group for patient i with G_i=1 representing the chemotherapy arm (CT in MIC and ESPAC) and G_i=0 representing the arm with no chemotherapy (PAL in MIC and NoCT in ESPAC).

For illustrative purposes a variety of models were fitted, all adaptations of the basic Cox regression model according to whether the process was assumed to be Markov or semi-Markov and whether a parametric form for the underlying baseline transition rate was specified (see Table 9.3).

The transition rates were analysed one at a time. When modelling the transition rate from state k to state l, only individuals who occupied state k at some point during the analysis period contributed to the model, whilst individuals who passed out of state k at some point and then returned back to it later contributed twice to the model. If an individual was in state k, then they were 'at risk' for the state k to state l transition. If they moved to state m rather than state l, then the time for the k to l transition was censored at the time of passing to state m. If they did not move from state k before the end-of-study time, then the k to l transition time was censored at that time (15 weeks for the MIC study and 24 months for the ESPAC study).

Table 9.3 Models fitted to transition rates in the MIC and ESPAC studies

Model	Distribution for transition times	Process	Model for transition rate from state k to state l
1	**Arbitrary**	**Markov**	$h_{[kl]}(t\|G)=h_{0[kl]}(t)\exp(\beta_{[kl]}G)Y_{[kl]}(t)$ $h_{0[kl]}(t)$ is baseline transition rate; $\beta_{[kl]}$ is the regression parameter for treatment; $Y_{[kl]}(t)$ is the 'at risk' process – if an individual is in state k at time t and therefore at risk for transition to state l then $Y_{[kl]}(t)=1$, otherwise $Y_{[kl]}(t)=0$; t represents time from study entry.
2	**Arbitrary**	**Semi-Markov**	$h_{[kl]}(t\|G, t_k)=h_{0[kl]}(t-t_k)\exp(\beta_{[kl]}G)$ $h_{0[kl]}(t-t_k)$ is the baseline transition rate; $\beta_{[kl]}$ is the regression parameter for treatment; t_k represents time from study entry to entry to state k.
3	**Exponential** Time-homogeneous i.e. transition rates assumed to be constant over time	**Not-applicable** (Markov equivalent to semi-Markov)	$h_{[kl]}(t\|G)=\lambda_{[kl]}\exp(\beta_{[kl]}G)$ $\lambda_{[kl]}$ is the constant baseline transition rate; $\beta_{[kl]}$ is the regression parameter for treatment; t can be either the time from study entry or the time from entry to state k.
4	**Weibull** Time non-homogeneous i.e. transition rates assumed to vary over time	**Markov**	$h_{[kl]}(t\|G)=\lambda_{[kl]}\gamma_{[kl]}t^{\gamma[kl]-1}\exp(\beta_{[kl]}G)Y_{[kl]}(t)$ $\lambda_{[kl]}$ and $\gamma_{[kl]}$ are the scale and shape parameters for the Weibull distribution; $\beta_{[kl]}$ is the regression parameter for treatment; $Y_{[kl]}(t)$ is the 'at-risk' process as described in Model 1; t represents time from study entry.
5	**Weibull** Time non-homogeneous i.e. transition rates assumed to vary over time	**Semi-Markov**	$h_{[kl]}(t\|G, t_k)=\lambda_{[kl]}\gamma_{[kl]}(t-t_k)^{\gamma[kl]-1}\exp(\beta_{[kl]}G)$ $\lambda_{[kl]}$ and $\gamma_{[kl]}$ are the scale and shape parameters for the Weibull distribution; $\beta_{[kl]}$ is the regression parameter for treatment; t_k represents time from study entry to state k.

The PHREG procedure in SAS (SAS Institute Inc. 1992) was used to fit the Cox regression models with arbitrary baseline transition rate and the LIFEREG procedure (SAS Institute Inc 1989) was used for models with exponential and Weibull baseline transition rates. The LIFEREG procedure models the transition rates as accelerated failure time models but if transition times are assumed to follow an exponential or Weibull distribution then there is a direct correspondence between the parameters under an accelerated failure time model and those under a Cox regression model (see Sections 7.2.4 and 7.2.5). The results here are reported as parameters from a Cox regression model.

The PHREG procedure was able to fit Markov and semi-Markov models but the LIFEREG procedure was only able to fit semi-Markov models as it does not allow subjects to move in and out of the risk set. STATA software (StatCorp 2002) was therefore used to fit Weibull models under the Markov assumption. The set-up of the data for analysis was different depending on whether the model being fitted was Markov or semi-Markov. In all cases the data for the transition from state k to state l consisted of one line per patient per visit to state k. Each line also consisted of an indicator variable that was 1 for an exit to state l, representing an actual event, and 0 for either an exit to state m or an exit time censored at end-of-study time or last survival follow-up. For the Markov model, each line consisted of the time of entry to and exit from state k in relation to trial time, which determined the periods of time in their follow-up that they were in the risk set for the transition. For the semi-Markov model, each line consisted of the duration of time spent in state k on that visit. The data set up for the time-homogeneous model (Model 3) was the same as that for a semi-Markov model.

9.4.3 Results for Treatment Comparison

The models with arbitrary baseline transition rates (Models 1 and 2) give estimates only for the treatment regression parameter, whilst the parametric models (Models 3, 4 and 5) also give estimates for the parameters of the underlying distributions for the transition times. The effect of treatment on each transition rate is the main interest and thus the estimates of the treatment regression parameter from each model are given here (see Tables 9.4 and 9.5). Regression parameters represent the log hazard ratio and

negative values indicate that the hazard of transition for the chemotherapy arm is less than that for the non-chemotherapy arm. Parameter estimates for the exponential and Weibull distributions have not been given but a comparison of models is discussed in the text.

Table 9.4: Treatment regression parameters (with standard errors) in the 3-state model for the MIC study

	Model 1: Arbitrary Markov	**Model 2:** Arbitrary Semi-Markov	**Model 3:** Exponential Markov / Semi-Markov	**Model 4:** Weibull Markov	**Model 5:** Weibull Semi-Markov
***GQS*-Defined Health States**					
Well →Ill	0.41 (0.44)	0.40 (0.44)	0.42 (0.44)	0.41 (0.44)	0.43 (0.44)
Ill →Well	0.73 (0.43)$^{+}$	0.72 (0.43)$^{+}$	0.72 (0.43)$^{+}$	0.71 (0.43)$^{+}$	0.75 (0.43)$^{+}$
Well →Dead	-	-	-	-	-
Ill →Dead	-0.45 (0.41)	-0.39 (0.41)	-0.43 (0.41)	-0.44 (0.41)	-0.38 (0.41)
***MAL*-Defined Health States**					
Well →Ill	0.43 (0.36)	0.51 (0.36)	0.45 (0.36)	0.42 (0.36)	0.51 (0.36)
Ill →Well	0.49 (0.39)	0.43 (0.39)	0.48 (0.38)	0.47 (0.39)	0.50 (0.39)
Well →Dead	-0.76 (0.92)	-0.79 (0.91)	-0.86 (0.91)	-0.88 (0.91)	-0.79 (0.91)
Ill →Dead	-0.94 (0.42)*	-0.87 (0.42)*	-0.89 (0.42)*	-0.93 (0.42)*	-0.83 (0.42)$^{+}$

+ statistically significant at the 10% level
* statistically significant at the 5% level

In the MIC study, although the numbers of subjects is relatively small there were adequate subjects and events to model all transitions except for the *well* to *dead* transition defined by *GQS*, but the small study size is reflected in the relatively large standard errors. The treatment parameter estimates for the five models are comparable for both the *GQS* and *MAL* models. The results suggest that, compared to PAL, CT increases the hazard of transition between the two different quality of life states but decreases the hazard of transition to death from both the *well* and *ill* states. The increase in the relative hazard of transition from *ill* to *well* is greater for the *GQS* model than the *MAL* model with regression parameters statistically significant at the 10% level. There

is a greater reduction in relative hazard of death from the *ill* state for the *MAL* model compared to the *GQS* model, with regression parameters statistically significant at the 5% level.

The Markov models give very similar results in general to the semi-Markov models and comparison of AIC for the two different forms with arbitrary baseline suggest Markov generally fit better than semi-Markov. This implies that taking account of duration of time in a state in this study has relatively little impact on the effect of treatment on transition rates out of the state. Also the models with parametric baseline hazards gave very similar results to those with an arbitrary baseline hazard indicating that assuming such underlying distributions for the transition times between states does not greatly influence the treatment effect on the transition rates. Examining changes in minus twice log likelihood suggest that the exponential is a satisfactory model for all transitions except *ill* to *dead* for which the underlying transition rate may not be homogeneous over time.

Table 9.5: Treatment regression parameters (with standard errors) in the 3-state model for the ESPAC study

	Model 1: Arbitrary Markov	**Model 2:** Arbitrary Semi-Markov	**Model 3:** Exponential Markov / Semi-Markov	**Model 4:** Weibull Markov	**Model 5:** Weibull Semi-Markov
Well →Ill	0.05 (0.22)	0.11 (0.22)	0.08 (0.22)	0.03 (0.22)	0.06 (0.22)
Ill →Well	0.10 (0.23)	-0.01 (0.23)	-0.02 (0.23)	-0.05 (0.23)	-0.07 (0.23)
Well →Dead	-0.47 (0.34)	-0.45 (0.34)	-0.41 (0.33)	-0.50 (0.34)	-0.46 (0.34)
Ill →Dead	-0.48 (0.25)$^{+}$	-0.43 (0.24)$^{+}$	-0.44 (0.24)$^{+}$	-0.59 (0.24)*	-0.47 (0.24)$^{+}$

\+ statistically significant at the 10% level
* statistically significant at the 5% level

In the ESPAC study, the results from all five models were reasonably comparable and all gave similar inferences. The results show that chemotherapy appears to have very little effect on the relative hazard of transition between quality of life health states defined by *GHSS* but does reduce the relative hazard of transition to death from both quality of life states but particularly the ill state for which the regression parameters are

statistically significant at the 10% level. This treatment effect was greatest in the Weibull Markov model. The arbitrary Markov model suggested an opposite treatment effect to the other models for the *ill* to *well* transition but this is probably due to random fluctuation around a zero effect. Comparison of AIC for the arbitrary Markov and semi-Markov models suggest that for all transitions the Markov is a better fit for the data than the semi-Markov. This implies that the duration of time spent in a health state does not have an effect on the relative hazard of transition. Comparison of exponential and Weibull using change in minus twice log likelihood suggests that the Weibull is a better fit for the data than the exponential and thus the hazard of transition is non-homogeneous over time.

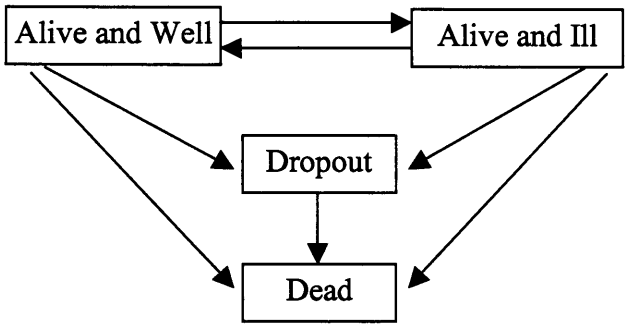
9.5 Extending Models to Include Dropout State

Patients who drop out of the longitudinal quality of life study due to death i.e. who die prior to the next scheduled assessment are assumed to move from the health state they occupied at the last assessment to the dead state at the time of death. In general, this will be a reasonable assumption to make. Patients who drop out prior to death need to be dealt with adequately in any analysis. In the previous section we have used a last value carried forward approach to impute the time between the last assessment and death or end-of-study time or last survival follow-up. This may be a reasonable assumption when the time gap is small, but for those who drop out early this may lead to unrealistic estimation of transition times.

An alternative, and perhaps more appropriate, option might be to consider *dropout* as a type of health state and include this as an extra transient alive state in the multistate model. Patients will only move to death from this state. If dropout is believed to be due to illness then this could be regarded as a state worse than the *ill* state i.e. their *ill* state was so bad as to cease participation in the quality of life study. If dropout is believed to be due to different reasons then it may be preferable to include a number of different *dropout* states. This is especially important if some patients drop out for reasons of poor health and others drop out for reasons of good health. By including a *dropout* state, patients are effectively being taken out of the risk sets for the transitions in the three-state model at the point when they drop out of the quality of life study. In the MIC and

ESPAC studies, it is assumed that all patients dropout due to poor health (see section 5.8.1 and 5.8.2) and therefore one *dropout* state was added to illness-death model, as shown in Figure 9.2.

Figure 9.2 Illness-death model with dropout state



In the MIC study patients were defined as dropouts if they did not return their final scheduled assessment (5^{th} assessment for CT arm and 4^{th} assessment for PAL) and in the ESPAC study they were defined as dropouts if their last assessment was more than 3 months prior to their last survival follow-up, whether that be dead or alive (see Section 5.8 for more details on dropouts). Patients who drop out were assumed to move to the *dropout* state at the time when the next planned assessment was due i.e. 3 weeks after the last recorded assessment in the MIC study and 3 months after the last recorded assessment in the ESPAC study, provided this was within their survival follow-up and within the analysis period (i.e. within 15 weeks for MIC and 24 months for ESPAC). Note that in the MIC study, patients who completed all scheduled assessments remain at risk of transition to dropout for all follow-up time after their last assessment despite the fact that the definition of dropout in this study precludes them from entering this state. An alternative option that may address this problem is to censor the transition to dropout three weeks after the last assessment, if the patient is still on follow-up at this time.

The amount of data contributing to the model for each transition rate is shown in Tables 9.6 and 9.7. With the increased complexity of the model, the amount of data for each transition is reduced. As with the three-state model, there are not sufficient data in the MIC study to adequately model the *well* to *dead* transition when these states are defined by *GQS*.

Table 9.6 Frequency of transitions in the dropout model for the MIC data

Transition	Number at Risk			Number of Transitions		
	Total	CT	PAL	Total	CT	PAL
***GQS*-Defined Health States**						
Well→Ill	78	53	25	28	21	7
Ill→Well	88	57	31	29	22	7
Well→Dropout	78	53	25	14	8	6
Ill→Dropout	88	57	31	21	13	8
Well→Dead	78	53	25	1	0	1
Ill→Dead	88	57	31	12	6	6
Dropout→Dead	35	21	14	15	6	9
***MAL*-Defined Health States**						
Well→Ill	78	52	26	38	27	11
Ill→Well	105	69	36	36	27	9
Well→Dropout	78	52	26	8	5	3
Ill→Dropout	105	69	36	27	16	11
Well→Dead	78	52	26	2	1	1
Ill→Dead	105	69	36	11	5	6
Dropout→Dead	35	21	14	15	6	9

Table 9.7 Frequency of transitions in the dropout model for the ESPAC data

Transition	Number at Risk			Number of Transitions		
	Total	CT	PAL	Total	CT	PAL
***GHSS*-Defined Health States**						
Well→Ill	163	81	82	80	42	38
Ill→Well	172	94	78	80	46	34
Well→Dropout	163	81	82	40	18	22
Ill→Dropout	172	94	78	34	21	13
Well→Dead	163	81	82	13	5	8
Ill→Dead	172	94	78	43	18	25
Dropout→Dead	74	39	35	49	24	25

The 7 different transition rates can be modelled as before using Cox regression models. Models with arbitrary, exponential and Weibull baseline hazards under both a Markov and semi-Markov assumption were estimated for each transition and the results for the treatment regression parameter are shown in Tables 9.8 and 9.9. As before, regression parameters represent the log hazard ratios and negative values indicate that the hazard of transition for the chemotherapy arm is less than that for the non-chemotherapy arm.

Table 9.8: Treatment regression parameters (with standard errors) in the dropout model for the MIC study

	Model 1: Arbitrary Markov	**Model 2:** Arbitrary Semi-Markov	**Model 3:** Exponential	**Model 4:** Weibull Markov	**Model 5:** Weibull Semi-Markov
***GQS*-Defined Health States**					
Well →Ill	0.34 (0.44)	0.35 (0.44)	0.36 (0.44)	0.33 (0.44)	0.35 (0.44)
Ill →Well	0.78 (0.44)$^{+}$	0.81 (0.43)$^{+}$	0.78 (0.43)$^{+}$	0.80 (0.43)$^{+}$	0.85 (0.43)$^{+}$
Well →Dead	-	-	-	-	-
Ill →Dead	-0.31 (0.58)	-0.33 (0.58)	-0.36 (0.58)	-0.33 (0.58)	-0.32 (0.58)
Well →Dropout	-0.44 (0.54)	-0.43 (0.54)	-0.45 (0.54)	-0.53 (0.54)	-0.48 (0.54)
Ill →Dropout	0.19 (0.45)	0.20 (0.45)	0.12 (0.45)	0.15 (0.45)	0.21 (0.45)
Dropout →Dead	-0.63 (0.57)	-0.72 (0.54)	-0.93 (0.53)	-0.75 (0.53)	-0.92 (0.53)
***MAL*-Defined Health States**					
Well →Ill	0.43 (0.36)	0.52 (0.36)	0.45 (0.36)	0.44 (0.36)	0.53 (0.36)
Ill →Well	0.45 (0.39)	0.44 (0.39)	0.45 (0.38)	0.44 (0.39)	0.33 (1.42)
Well →Dead	-0.32 (1.42)	-0.38 (1.41)	-0.44 (1.41)	-0.47 (1.41)	-0.48 (0.39)
Ill →Dead	-0.87 (0.61)	-0.83 (0.61)	-0.83 (0.61)	-0.87 (0.61)	-0.79 (0.61)
Well →Dropout	0.14 (0.73)	0.09 (0.73)	0.07 (0.73)	0.04 (0.73)	0.14 (0.73)
Ill →Dropout	-0.29 (0.39)	-0.27 (0.39)	-0.27 (0.39)	-0.32 (0.39)	-0.24 (0.39)
Dropout →Dead	-0.63 (0.57)	-0.72 (0.54)	-0.93 (0.53)	-0.75 (0.53)	-0.92 (0.53)

+ statistically significant at the 10% level
* statistically significant at the 5% level

In the MIC study, as with the three-state model, the results for the five different models are comparable. The addition of the dropout state has not greatly changed the conclusions from the previous 3-state model, with chemotherapy still increasing the relative hazard of transition between well and ill states but reducing the relative hazard of transition to death. The hazard of death from dropout is reduced with chemotherapy. In the *GQS*-defined dropout model, chemotherapy can now be seen to reduce the

relative hazard of transition to dropout from the well state but increase it from the ill state. Treatment has the opposite effect on these transitions in the *MAL*-defined dropout model, indicating that the conclusions are not robust to the definition of the health states.

As with the 3-state model, comparison of AIC for the Markov and semi-Markov models under an arbitrary baseline indicates that the Markov model fits the data better, which suggests that in this dropout model, duration of time in a state does not affect the relative hazard of transition to other states. Examination of changes in minus twice log likelihood between exponential and Weibull models suggested that Weibull was a better fit for the *GQS*-defined *well* to *dropout* and *ill* to *dropout* transitions and for the *MAL*-defined *well* to *ill*, *ill* to *well* and *ill* to *dropout* transitions. For all transitions to death the exponential was an appropriate model suggesting that the underlying hazard of transition to death from any state is constant over time.

Table 9.9: Treatment regression parameters (with standard errors) in the dropout model for the ESPAC study

	Model 1: Arbitrary Markov	**Model 2:** Arbitrary Semi-Markov	**Model 3:** Exponential Markov / Semi-Markov	**Model 4:** Weibull Markov	**Model 5:** Weibull Semi-Markov
Well →Ill	-0.02 (0.23)	0.05 (0.23)	0.04 (0.22)	-0.05 (0.23)	0.00 (0.22)
Ill →Well	0.13 (0.23)	0.06 (0.23)	0.03 (0.23)	-0.00 (0.23)	0.01 (0.23)
Well →Dead	-0.58 (0.57)	-0.54 (0.57)	-0.53 (0.57)	-0.63 (0.57)	-0.57 (0.57)
Ill →Dead	-0.58 (0.31)$^{+}$	-0.58 (0.31)$^{+}$	-0.60 (0.31)$^{+}$	-0.72 (0.31)*	-0.60 (0.31)
Well →Dropout	-0.35 (0.32)	-0.27 (0.32)	-0.26 (0.32)	-0.38 (0.32)	-0.32 (0.32)
Ill →Dropout	0.07 (0.36)	0.17 (0.35)	0.21 (0.35)	0.07 (0.36)	0.18 (0.35)
Dropout →Dead	-0.19 (0.29)	-0.26 (0.29)	-0.17 (0.29)	-0.17 (0.29)	-0.17 (0.29)

+ statistically significant at the 10% level
* statistically significant at the 5% level

In the ESPAC study, as with the three-state model, the results from all five models are reasonably comparable and as before the results imply that chemotherapy has very little effect on the hazard of transition between *well* and *ill* states but reduces the relative hazard to death from both of these states, significantly so from the *ill* state. The inclusion of a dropout state in the model has therefore not affected the conclusions here regarding the effect of treatment on these transitions. The estimates suggest that chemotherapy reduces the relative hazard of transition to *dropout* from a *well* state but possibly increases it from an *ill* state. The relative hazard of transition to death from *dropout* is reduced with chemotherapy,

Comparison of AIC for the arbitrary Markov and semi-Markov models suggested that, as in the three-state model, for all transitions the Markov model is a better fit for the data than the semi-Markov. This implies that the duration of time spent in a health state does not have an effect on the relative hazard of transition. Comparison of exponential and Weibull using change in minus twice log likelihood suggested that the Weibull is a better fit for the data than the exponential for the transitions between alive states but exponential is better for the transitions to death. This implies that, after adjusting for dropout, the hazard of transition to death is constant whilst the hazard of transition from one alive state to another is non-homogeneous over time.

9.6 Discussion and Critical Review of Multistate Modelling

Although multistate modelling has been recognised as a potential approach for the simultaneous analysis of quality of life and survival data (Fayers and Jones 1983, Olschewski and Schumacher 1990, Schumacher et al 1991, Cox et al 1992, Abrams 1992a, Olschewski et al 1992), applications have been very limited (Olschweski 1984, Abrams 1992b, Charlesworth and Skene 1997). This thesis develops the method of multistate modelling for quality of life and survival data, applying the methodology to two different datasets and thus allowing the application of such models in this field to be investigated. In addition, the methodology is extended to deal with informative dropout by including a *dropout* state in the model. This approach for dealing with informative dropout in a longitudinal study has not previously been considered.

Multistate modelling provides a flexible approach for the simultaneous analysis of quality of life and survival data. One of the major advantages of multistate modelling is that no distributional assumptions regarding the quality of life outcome are needed and the method can be applied to any type of quality of life outcome whether continuous, ordinal or binary. The models considered here define health states in terms of a single quality of life outcome. In practice there may be more than one quality of life outcome of interest. The definition of health states can be based on responses to more than one measure thus providing a simple option for accommodating multiple quality of life outcomes into the analysis. Multivariate models for the transition rates could be considered as a means for dealing with multiple outcomes and this has been attempted for a model with health states based on two clinical outcomes using a bivariate exponential model (Young et al 1999).

Defining health states for a multistate model however is not necessarily straightforward. The definitions based on quality of life data are subjective and different definitions need to be considered as part of a sensitivity analysis. The investigator has to make decisions on the quality of life variable to be used, the number of health states to be included and the cut-off values used to discriminate between health states. At one extreme, the model needs to be complex enough to be clinically meaningful and to ensure information from the data is utilised to a maximum. At the other extreme a simple model is needed to allow an adequate number of transitions between health states, enabling transition rates to be estimated with sufficient precision and also to ease the interpretation of the analysis.

Definition of the health states may be such that clinically important information is lost. The most clinically important transition may not be able to be included in the model because of the small number of participants in the study experiencing it. For example, in the MIC study the moderately ill (MAL=2) to severely ill (MAL=3) transition could not be modelled because of a lack of numbers but this could be a very important transition from a clinical viewpoint. Also the transition between health states may be as a result of very small changes in quality of life whilst large changes in quality of life are not reflected in the model. For example, in the MIC study when a GQS value of 83 is used to discriminate between the *well* and *ill* states, then a small change in quality of life from 82 to 84 would result in a transition whilst a large change from say 30 to 80 would

not be reflected in the model. This could be addressed by defining health states in terms percentage reduction in quality of life score. In addition, the measurement error in the quality of life outcome is not accounted for when determining transitions between states. Smoothing techniques and random effects models have been considered for the longitudinal data that determines the transitions (Sypsa et al 2001) and 'adhoc smoothing' which required subjects to be in a new state for two consecutive measurements for a proper transition was found to be the best approach.

The data requirements for the approaches to multistate modelling described in this chapter are strict, with not only dates of entry, death and censoring needed, but also 'exact' dates of transition between health states. In prospective studies of quality of life and survival, although the exact time of transition to death will generally be known, it is not possible to observe the actual transitions of patients from one quality-of-life health state to another, all that is known for a patient is the health state occupied at each scheduled assessment time. Thus, if the health state of a patient changes from one assessment to the next then the actual time of transition is not known, just that it occurred somewhere between the two assessments. The transition times in such a situation are *interval-censored* (Collett 1994). The approach that has been used in this thesis is to estimate the exact transition times from the data. The most widely used approximations to exact transition dates are the actual follow-up dates, as used here, or the mid-point between follow-up dates. The accuracy of estimation is determined by the frequency of the quality of life assessments. If the assessments are widely spaced then these estimates may not be very accurate and it has been shown that estimates of transition rates obtained from approximated data will not always be correct (Andersen et al 1991). Thus, when quality of life assessment times are limited and widely spaced, alternative approaches to modelling transition rates that do not require exact transition times may be more appropriate (Kay 1986, Hillis 1986, Longini et al 1989, Andersen 1991, Gentleman et al 1994, Lu and Stitt 1994, Marshall and Jones 1995, Gottschau and Hogh 1995) but these methods are not readily accessible to researchers since specialised software is required.

Multistate modelling requires a relatively large amount of data to ensure that transitions rates have sufficient patients for adequate modelling and to ensure that there is sufficient power to detect real differences between treatments in terms of the transition

rates between states. The amount of data in the MIC study was not really sufficient for multistate modelling and some transitions may not be adequately modelled. The results were not robust to the choice of quality of life variable chosen to define the health states and this may have been a result of small study size. With the increase in importance of quality of life in clinical trials, the collection of data will be given greater priority and the problem of lack of data should be less of an issue (Abrams 1992a). Further, if multistate models are proposed at the design stage of a clinical trial, then the collection of data can be planned so it yields adequate and appropriate data.

The models illustrated in this chapter have been in the form of a Cox regression model but other forms such as accelerated failure time models could be considered (Cole et al 1994). Parametric and semi-parametric models have been illustrated under both Markov and semi-Markov assumptions. In general, the conclusions in these studies have been reasonably robust to these choices. The models have only included treatment as a covariate but with sufficient data additional covariates could easily be included with, if appropriate, different covariates included for different transitions. The usual methods for model fitting can be applied with changes in minus twice the log likelihood used to assess the statistical significance of different covariates within each transition. Repeat transitions have been treated as independent events in the analysis here but this may not always be deemed appropriate and if the number is not inconsiderable then the violation of the assumption may invalidate the results. Covariates could be included in the model to account for the non-independence of repeat events by, for example including a time-dependent covariate to represent the number of times a patient has previously visited the current health state (Kay 1982).

The analysis here was restricted to fixed time periods. In the case of the MIC study this was because the data collection was restricted to a fixed time period. In the case of the ESPAC study the analysis was restricted because of sparse data but theoretically if the quality of life study collects data over the full follow-up period then the models could be unrestricted in terms of the period of analysis.

The inclusion of a dropout state in the model removes the need to impute values of quality of life from the last assessment to death, censor date or fixed end-of-study time. This is particularly relevant when the length of period for imputation is long. By

including a dropout state in the model, subjects are removed from the risk sets for transitions to other health states at the time of dropout. This may provide more valid estimates of the transition rates between the quality of life states and death. In the illustrative studies presented here, the inclusion of a dropout state in the model did not greatly affect the conclusions regarding the relative hazards of transition between well, ill and dead states for the two treatment groups but it allowed the dropout process to be investigated. In the ESPAC study the results suggested that if a patient is in a well state then chemotherapy is better at keeping them there. The method does not make any assumption about what type of health the patients have in the dropout state, just that all patients in the dropout state have the same level of health. If patients are dropping out for different reasons, some related to poor health and others for administrative reasons for example, then different dropout states can be included for different reasons.

Methods discussed in this chapter have been based on a classical approach and, although there has been some work on Bayesian approaches to multistate modelling (Abrams 1992b, Charlesworth and Skene 1997), further development in this field would be beneficial. Although complex models such as multivariate multistate models can be fit using a classical approach, a Bayesian analysis may provide a more flexible framework. A specific example of a complex model involving multistate modelling that has been proposed using a classical approach but may be better implemented in a Bayesian framework is the parametric method for quality-adjusted survival analysis proposed by Cole et al (1994). The method uses the results from a parametric multistate model and simulation to estimate the time spent in different health states for incorporation into a QALY-type model. Parameter estimation in such a complex model may be more easily achieved in a Bayesian analysis.

Multistate models with clinically-defined health states have been advocated as being preferable to an overall survival model since a greater biological insight may be gained by analysing the steps in a disease/treatment process (Andersen 1988). Similarly multistate models with quality-of-life-defined health states provide insight into the quality of life process with the ability to investigate the effects of different covariates on different transitions. The inclusion of a death state in the model overcomes the problem of dropout due to death in the analysis of quality of life data and the further inclusion of a dropout state overcomes the problem of dropout prior to death. The inclusion of the

dropout state enables the dropout process to be explored and more than one dropout state can be included if appropriate. Given sufficient data, multistate modelling, despite being previously neglected, would appear to provide a practical and clinically intuitive option for the simultaneous analysis of quality of life and survival data.

CHAPTER 10: JOINT MODELLING OF QUALITY OF LIFE AND SURVIVAL DATA

10.1 Introduction

The aim of this chapter is to describe the types of model that have been considered for the joint modelling of repeated measures and time-to-event data, highlighting those that have been applied to quality of life and survival or dropout data, and to develop a Bayesian approach to joint modelling for application to such data. Although the classical approach to joint modelling has been emerging as a useful method for the analysis of quality of life and survival or dropout data, the application of the Bayesian approach for this type of data has until recently not previously been undertaken and this is therefore the focus for this chapter.

There is an increasing literature on the joint modelling of repeated measures and time-to-event data. The majority of examples have used classical approaches (Wu and Bailey 1988 and 1989, Wu and Carroll 1988, Diggle and Kenward 1994, Little 1995, Schluchter 1992, Pawitan and Self 1993, Wulfsohn and Tsiatis 1997, Hogan and Laird 1997a and b, Henderson et al 2000) and there are a number of recent examples of the application of such methods to longitudinal quality of life data (Fairclough et al 1998a, Ribaudo et al 2000, Curran et al 2002, Michiels et al 2002, Fairclough 2002, Pauler et al 2003). The Bayesian approach to joint modelling has been described by a number of authors (Berzuini 1995, Faucett and Thomas 1996, Carpenter et al 2000, Xu and Zeger 2001, Wang and Taylor 2001) but there is only one recent example of the application of such an approach in the field of quality of life (Wang et al 2002).

In some studies, the aim of joint modelling may be to deal with the problem of non-ignorable missing data in the analysis of repeated measures over time. By jointly modelling the event that causes dropout from the longitudinal study with the repeated measures data, the estimation of parameters in the repeated measures model are 'adjusted' to allow for the informative dropout. In other studies the aim may be to provide more valid estimates of the relationship between the repeated measures and the

time to an event by modelling the repeated measures covariate that is included in the time-to-event model. In some situations, both questions will be of interest.

The change in quality of life over time and the time to death can be considered as two simultaneous processes occurring in patients, and can be modelled as such. Models for analysing quality of life data and survival data as two separate processes were discussed in Chapters 6 and 7 respectively and these are developed within a joint model here. Joint modelling of quality of life and survival data will account for the dropout due to death in the model for changing quality of life over time and will provide a more appropriate estimate of the relationship between the two processes. By jointly modelling quality of life and time to dropout or death, the model can also account for other reasons for dropout as well as death and can provide some insight into the dropout process.

The classical approaches to joint modelling are reviewed in Section 10.2 with particular reference to applications to quality of life data. The Bayesian approach is described in Section 10.3 and its specific application to the quality of life and survival data in the MIC and ESPAC studies are described in Section 10.4. Modelling quality of life and survival data only deals with the problem of missing data that results from death and the problem of informative dropout prior to death is discussed in Section 10.5 and dealt with by modelling time to dropout or death rather than time to death. A critical review of the methodology is given in Section 10.6, in which the limitations and issues for further research are discussed.

10.2 Classical Approaches to Joint Modelling

10.2.1 General Review

In the classical approach to joint modelling, the joint density can be factorised in two ways. The two different forms of factorisation give rise to two different types of models usually referred to as *mixture models* and *selection models*. There is an extensive literature discussing these types of models for longitudinal data in the presence of informative dropout (Little 1995, Hogan and Laird 1997a and 1997b, Kenward and Molenberghs 1999). If Y represents the longitudinal data, D represents the dropout

process that gives rise to the missing data, X represent covariates of interest such as treatment, θ and ϕ represent the unknown parameters relating X with Y and D respectively, then the joint density $f(Y, D \mid X, \theta, \phi)$ can be factorised either as

$$f(Y, D \mid X, \theta, \phi) = f(Y \mid D, X, \theta) f(D \mid X, \phi) \qquad [10.1]$$

or as

$$f(Y, D \mid X, \theta, \phi) = f(D \mid Y, X, \phi) f(Y \mid X, \theta) \qquad [10.2]$$

The first factorisation in [10.1] is the form of a mixture model and the second factorisation in [10.2] is the form of a selection model. In the mixture model, the model for the longitudinal data is conditional on the dropout process whilst in the selection model, the model for the dropout process is conditional on the longitudinal data. In both cases the majority of the literature focus on the repeated measures being a normally distributed outcome. As such, the most commonly used model for the repeated measures over time is a random effects model (Laird and Ware 1982), usually in a linear form. The dropout process can be expressed in a number of different ways. If dropout specifically due to death is of interest then death rather than dropout will be the event of interest. In either case the occurrence of the event or the time to the event may be modelled.

Early literature on dealing with the problem of informative dropout in the analysis of repeated measures uses a form of mixture model (Wu and Bailey 1988, 1989). A conditional linear random effects model is assumed for the repeated measures over time such that the individual slopes are linear functions of the total follow-up time. Two methods for estimating the population slope based on a weighted least squares approach were suggested: the linear minimum variance unbiased estimator and the linear minimum mean squared errors estimator. This methodology was developed further using empirical Bayes inference to enable not only the population slope parameter to be estimated but also individual slope parameters (Mori et al 1992). Further extensions of the approach included the individual intercepts and slopes in a linear random effects model being functions of the time to a censoring event, with time-to-event modelled as a Weibull distribution (Pawitan and Self 1993).

The problem with the conditional linear model approach is that at least two measurements per subject are needed and the method assumes that the potential follow-up time is the same for all subjects and thus methods were developed to overcome this (Schluchter 1992). Schluchter's model assumes a linear random effects model for the repeated measures over time and assumes that the individual intercept, slope and a transformation of the dropout time, generally a log transformation, follow a trivariate normal distribution. Maximum likelihood estimates of the model parameters are obtained using the EM algorithm, which enables censored dropout times to be accommodated. Applications of the method have been demonstrated (Schluchter et al 2001, Schluchter et al 2002) but the approach is computationally complex, may require large amounts of data and is based on an entirely parametric model.

The properties of a multivariate normal distribution mean that Schluchter's model can be considered as both a mixture model and a selection model. On the one hand, the conditional expected values of the parameters from the random effects model given the transformed dropout times can be expressed as a linear function of the transformed dropout times, whilst on the other the mean of the transformed dropout times can be expressed as linear function of the random intercept and slope (Schluchter et al 2001). In this way both the question of examining the change in the repeated measure over time whilst accounting for informative dropout and the question of examining the relationship between the repeated measure and the risk of dropout can be addressed simultaneously. In that sense the approach by Schluchter is similar to the Bayesian approach described and developed in Section 10.3.

A more general form of the model has been described (De Gruttola and Tu 1994) and the model has also been adapted to use residuals from an accelerated failure time model for the survival times rather then the log survival times themselves in the trivariate normal model (Touloumi et al 1999). In this adaptation of the model, the method of restricted iterative generalised least squares is used to estimate parameters within the maximisation step of the EM algorithm. A mixture model that does not assume any parametric form for the event-time distribution has also been described (Hogan and Laird 1997a). Further details on pattern-mixture models are discussed in Section 10.2.2 in relation to their application to quality of life data.

Selection models have also been proposed as a method for modelling repeated measures over time with missing data caused by informative dropout. In these joint models the model for the dropout process, however defined, is conditional on the change over time in the repeated measures. In one of the earliest applications, a probit model is assumed for the probability of dropout with the probability conditional on the individual intercept and slope from the linear random effects model for the repeated measures over time (Wu and Carroll 1988). A more general approach to the problem of informative dropout using selection models proposed a linear logistic regression model for the probability of dropout conditional on the observed measures of the repeated outcome prior to the dropout time and the value of the repeated outcome that would have been observed at the dropout time had the subject not dropped out (Diggle and Kenward 1994). A simplex algorithm was used to maximise the log-likelihood but the EM algorithm may also be applicable. This model has been criticised for its sensitivity to the model assumptions, which are difficult to assess (Kenward 1998).

Rather than modelling the probability of dropout, the time to dropout can be modelled in a selection model. Wulfsohn and Tsiatis (1997) use a linear random effects model for the longitudinal data and a Cox regression model for survival data and include the modelled value of the longitudinal variable as a time-dependent covariate in the survival model. The EM algorithm is used to estimate the parameters. The selection model proposed by Henderson et al (2000) is given in a generic form such that it incorporates many of the specific models proposed by other researchers and specifies that the joint distribution of repeated measurements and time to dropout in each subject is modelled via a 'latent zero-mean bivariate Gaussian process' defined as follows. The model for repeated measures for subject i contains a latent variable $W_{1i}(t)$ and the model for the time to dropout contains a latent variable $W_{2i}(t)$ such that there is a link between these latent variables. For example setting $W_{1i}(t) = U_{1i}+U_{2i}\,t$ where U_{1i} and U_{2i} are subject-specific intercept and slope in a linear random effects models and $W_{2i}(t)=\gamma W_{1i}(t)$ gives a model used by a number of other researchers (Faucett and Thomas 1996, Wulfsohn and Tsiatis 1997). This general model formulation by Henderson et al (2000) and more specifically that proposed by Wulfsohn and Tsiatis (1997) most closely resembles the models developed and discussed in the Bayesian approach in Section 10.3.

10.2.2 Application to Quality of Life Studies

The joint modelling of quality of life and dropout from a classical point of view has been emerging as a means of overcoming the problems of missing quality of life data resulting from informative dropout (Fairclough et al 1998a, Ribaudo et al 2000, Curran et al 2002, Michiels et al 2002, Fairclough 2002, Pauler et al 2003). The two key types of joint model that have been applied to longitudinal quality of life data are the selection model and the pattern-mixture model.

Various forms of mixed model have been used for the quality of life data. In a pattern-mixture model, this model for quality of life over time is conditional on the dropout pattern whereas the selection model is interested in the marginal density for quality of life. Models for dropout have either been in terms of probability of dropout (Curran et al 2002, Michiels et al 2002, Pauler et al 2003) or the time of dropout (Fairclough et al 1998a, Ribaudo et al 2000)

Ribaudo et al (2000) simultaneously analyses the quality of life and survival data using the model originally proposed by Schluchter (1992) in which the intercept and slope from a linear random effects model of quality of life over time together with the log survival time are assumed to follow a trivariate normal distribution. Estimation of the parameters of this joint distribution is problematic when the data includes censored survival times, but these difficulties are overcome by using the EM algorithm. Fairclough et al (1998a) also illustrate the use of Schluchter's model for quality of life and dropout. This method could be extended to encompass a more complex random effects model, but convergence problems for parameter estimation could occur unless there are a large number of subjects (Fairclough et al 1998a). Also the performance of the model may deteriorate as the degree of censoring in the data increases (Ribaudo et al 2000).

Joint modelling of quality of life and dropout data has also taken the form of a selection model in which the quality of life data is modelled using a mixed model and the probability of dropout is modelled using a logistic regression model (Curran et al 2002, Michiels et al 2002). Pattern-mixture models have been shown to be a useful approach

as they are clinically intuitive (Fairclough et al 1998a, Curran et al 2002, Michiels et al 2002, Pauler et al 2003). The use of pattern-mixture models for quality of life data has been described in detail elsewhere (Fairclough 2002) and the methodology is summarised as follows. If the quality of life data are subject to informative dropout then the patients can be stratified according to p different dropout patterns. The regression parameters and covariance matrix in the model for quality of life over time are estimated separately for each missing data pattern. The population estimates for the model parameters are given by the weighted average of the estimates from the p dropout patterns where the weights are determined by the proportion of patients with that missing data pattern. So given a parameter β from a model for the longitudinal data, the population estimate of this parameter is given by:

$$\hat{\beta} = \sum_{\forall p} \hat{\pi}_p \hat{\beta}_p \qquad [10.3]$$

where $\hat{\pi}_p$ is the proportion of patients that are in the pth dropout pattern and $\hat{\beta}_p$ is the estimate of the β parameter in the pth dropout pattern.

One of the key advantages of the method is that a model for the dropout process does not need to be specified. There are however a number of practical problems in applying the methodology. Specification of the dropout patterns is subjective and patterns need to be defined in order to ensure that the number of patients in each stratum is sufficient for adequate modelling of the data. The most simplistic categorisation is to create two strata, one consisting of 'completers' and a second of dropouts at any time. This may too crude and ideally, depending on the number of subjects available, additional strata can be added according to time of, and reason for, dropout. Patterns may have small numbers of patients if the total number of patients in the study is small or there are a large number of time points at which a patient can drop out. For some patterns it may not be possible to estimate the model parameters and further assumptions may be required in order to be able to fit the model to all strata. For example, if a linear model is assumed for the quality of life data over time then at least two observations per patient are required to be able to estimate the intercept and slope. Thus, for patients who drop out after their baseline assessment, it may be necessary to assume that their slope is the

same as that for the patients who drop out after the second assessment. An alternative approach is to place assumptions on the model parameters using the time of dropout as a covariate (Fairclough 2002).

The 'mixing' of the parameters from the separate dropout patterns could be achieved using multilevel models rather than just using a weighted summation as described above. De Stavola and Christensen (1996) proposed a method for dealing with the problem of informative dropout in longitudinal studies using multilevel models. Overlapping risk sets are created such that for a series of time points $T_1,..,T_K$ the risk set at time T_k (k=1,...,K) consists of all subjects still alive at time T_k and multilevel models can be fit to the data with level 1 as the repeated measures over time, level 2 as the subjects, level 3 as the last risk set to which the subject belongs. This method has not been considered for quality of life data but may be a good approach as multilevel models have been shown to be good way of dealing with multidimensionality of quality of life data (Beacon and Thompson 1996).

10.3 Bayesian Approach to Joint Modelling

A Bayesian approach to joint modelling of repeated measures and time-to-event data was first described in detail by Faucett and Thomas (1996) and is summarised as follows. As with the classical approaches, models are specified for the repeated measures data (called the 'covariate tracking model' by Faucett and Thomas) and the time-to-event data (called the 'disease risk model' by Faucett and Thomas). Joint modelling is achieved through linking the models by including the *modelled* value of the repeated measure as a covariate in the time-to-event model. This model formulation specified by Faucett and Thomas (1996) is analogous to a selection model and is described in more detail below.

A measurement error model is assumed for the repeated measures data, with a linear random effects model assumed for the true values over time. If z_{ij} represents the value of the repeated measure on the jth occasion for subject i taken at time t_{ij} then

$$z_{ij} = x_i(t_{ij}) + \varepsilon_{ij} \qquad [10.4]$$

where $x_i(t_{ij})$ is the value of the true unobserved repeated measure at time t_{ij} and ε_{ij} are independent normally distributed errors with zero mean and constant variance and

$$x_i(t) = \alpha_i + \beta_i t \qquad [10.5]$$

where the random effects α_i and β_i have a bivariate normal distribution with unknown mean vector μ and unknown covariance matrix Σ.

A Cox proportional hazards model is specified for the time-to-event data such that the hazard of death for subject i at time t, $\lambda_i(t)$ is modelled using:

$$\lambda_i(t) = \lambda_0(t)\exp[\gamma x_i(t)] \qquad [10.6]$$

where $\lambda_0(t)$ represents the underlying baseline hazard function and γ is the regression coefficient for the true value of the repeated measures covariate. The baseline hazard is modelled as a piecewise exponential model.

Faucett and Thomas (1996) use Gibbs sampling (see Chapter 2) to fit the models in a single analysis. This gives the joint posterior distribution for all unknown parameters in both models. The advantages of Gibbs sampling for estimating complex models in a variety of medical contexts was highlighted in the early 1990s and was promoted as "a way of reducing the problem of dealing with a large number of related unknown parameters and missing data into a simpler problem dealing with one unknown quantity at a time, sampling each from its full conditional distribution" (Gilks et al 1993). In fact one of the medical contexts that was highlighted in this paper was the joint modelling of repeated measures over time and time-to-event data in the context of modelling serial CD4 counts and the onset of AIDS.

The method proposed by Faucett and Thomas (1996) is flexible in that it allows for unequally spaced and missing repeated measures data with varying numbers of

observations per subject and it allows for censored event times. Simulation to compare results from modelling each process separately with results from the combined model showed the separate models underestimated parameters whilst the combined model virtually eliminated the bias (Faucett and Thomas 1996).

The model proposed by Faucett and Thomas (1996) has been extended and adapted for application to various scenarios in which all have focused on using the joint model as a method for examining the effect of a time-dependent covariate on survival. (Faucett et al 1998, Xu and Zeger 2001, Wang and Taylor 2001, Taylor and Wang 2002, Pauler and Finkelstein 2002). The approach has been modified for a binary time-dependent covariate in order to examine the effect of post-operative smoking status over time on lung cancer recurrence and survival (Faucett et al 1998). In an application investigating the use of PANSS (positive and negative symptoms scale) as a marker for disease status in schizophrenia, the simple measurement error model that was originally proposed has been extended to a generalised linear model for the repeated measures data and this underlying true value or latent variable is modelled as a random effects model with covariates and serial autocorrelation (Xu and Zeger 2001). Wang and Taylor (2001) extend the model to give a more general form of the random effects model that includes a random effect called the integrated Ornstein-Uhlenbeck stochastic process, which accounts for the random fluctuation of the variable around the population average. They also include additional covariates in each part of the model and point out that by including a covariate in both parts of the model the joint model allows the effects of a covariate on the repeated measures process to be distinguished from its effect on the hazard of death or dropout. They use MCMC to fit the model and for some parameters use Metropolis-Hastings sampling because sampling directly from the full conditional distribution was not feasible. The potential of their model for the investigation of longitudinal biomarkers as surrogate endpoints has also been examined (Taylor and Wang 2002). Joint Bayesian models have been used to predict prostate cancer recurrence from serial measures of a tumour marker (Pauler and Finkelstein 2002). Piecewise linear random effects model are used with an individually-defined single change-point for the longitudinal measures of a tumour marker and a Cox regression model with a piecewise exponential baseline hazard for the time to prostate cancer recurrence. The models are linked by including functions of the longitudinal parameters

as covariates in the Cox model, in particular whether or not a change-point has occurred.

A comparable approach to that proposed by Faucett and Thomas (1996) for the joint modelling of serial measurements of disease markers and failure time data has been proposed in which the time scale is divided into a series of intervals or 'slots' (Berzuini 1995). The measurements of disease markers over time are modelled in terms of the time slots rather than continuous time using a linear random effects model. Simultaneously, the hazard of failure within a slot is modelled using a Cox regression model with constant baseline hazard assumed for each slot. A discrete-time approximation to the baseline hazard function of Cox is used which, unlike the piecewise exponential model, does not include the time of the event within the interval in the model.

There are a few examples where the aim of the Bayesian joint modelling is to adjust for the missing data, resulting from dropout, in the model for the repeated measure over time (Best et al 1996, Carpenter et al 2002, Wang et al 2002) with one specifically relating to quality of life data (Wang et al 2002). In the early example, joint modelling was used to estimate parameters in a selection model for repeated measures data taken on two occasions with dropout (Best at al 1996). The response at a second time point was modelled using a simple regression model including first response as a covariate and a logistic regression model was used for the probability of non-response conditional on the measured (not modelled) value of the response at the second time point. Informative priors were used to estimate the posterior distributions in a fully Bayesian analysis. Carpenter et al (2002) in a similar manner used a selection model formulation similar to that proposed by Diggle and Kenward (1994) in order to jointly model serial measures of forced expiratory volume (FEV) with the probability of dropout in a clinical trial of treatments for asthma. The logistic regression model for the probability of dropout includes a covariate for the measured (not modelled) value for FEV and rather than estimating the regression parameter for this covariate, the model is estimated using specifically chosen values for the regression parameter.

There is one recent example of joint modelling of quality of life data and survival data using a Bayesian approach applied to data collected in a heart failure trial (Wang et al

2002). The proposed methodology recognises that quality of life data usually consists of multiple ordinal responses designed to measure an underlying latent trait of quality of life. Rather than analysing the single score that is derived from summing these items, the model analyses the multiple item responses. A proportional odds model is used to jointly model the response frequencies from the multiple quality of life items making up the latent variable of quality of life. A Cox regression model is used to model the survival data and includes a covariate representing the latent variable for quality of life. Parameter estimation uses the Metropolis-Hastings algorithm rather than Gibbs sampling.

The method described by Faucett and Thomas (1996) was recognised for its potential for the analysis of quality of life and survival data (Billingham et al 1999, Billingham and Abrams 2002 – see Appendix I). Joint modelling should improve the parameter estimates in the quality of life model because the estimation will account for informative dropout due to death. At the same time, the estimates in the survival model will also be improved because the quality of life covariate values are estimated from the model for quality of life data over time fitted to all subjects and are adjusted for covariate measurement error. The method has been developed for application to quality of life and survival data and investigated through its application to the MIC and ESPAC data. To apply the methodology to quality of life and survival data a mixed model that adequately describes the pattern of quality of life over time should be considered for the quality of life process, as discussed in Section 6.4. For the survival process, Cox regression models or accelerated failure time models, either semi-parametric or parametric and possibly piecewise, as described in Chapter 7 could all be considered.

10.4 Joint Model for the MIC and ESPAC Studies

10.4.1 Model Specification

Suppose $Q(t_{ij})$ represents the value of quality of life for patient i at their jth time point t_{ij}, then the likelihood for the data is specified as follows:

$$Q(t_{ij}) \sim N(\mu(t_{ij}),\ \sigma_e^2)$$

where $\mu(t_{ij})$ represents the true unobserved value of quality of life at time t_{ij} and an appropriate model is chosen for this latent variable. Appropriate models for longitudinal quality of life data were discussed in Section 6.4.

For the MIC study a linear random effects model with a fixed treatment covariate was chosen as the most appropriate model to represent the true underlying value of *GQS* over time (see Section 6.5.1 and 6.6.1) and the model is specified as follows:

$$\mu(t_{ij}) = \theta_{1i} + \theta_{2i}t_{ij} + \delta_1 G_i + \delta_2 (G_i \times t_{ij}) \qquad [10.7]$$

where G_i represents the treatment group for patient i (0.5 for CT, -0.5 for PAL) and δ_1, δ_2 are fixed unknown regression parameters for treatment and treatment by time interaction. The individual intercepts and slopes are assumed to have a bivariate normal distribution with unknown mean $\boldsymbol{\alpha}$ and covariance matrix $\boldsymbol{\Sigma}$ as follows:

$$\begin{pmatrix}\theta_{1i}\\ \theta_{2i}\end{pmatrix} \sim MVN\left(\begin{pmatrix}\alpha_1\\ \alpha_2\end{pmatrix}, \begin{pmatrix}\sigma_1^2 & \sigma_{12}\\ \sigma_{12} & \sigma_2^2\end{pmatrix}\right)$$

For the ESPAC study a piecewise linear random effects model with a fixed change-point at 6 months was chosen as the most appropriate model to represent the true underlying value of *GHSS* over time (see Section 6.5.2 and 6.6.2) and the model is specified as follows:

$$\mu(t_{ij}) = \theta_{1i} + \theta_{2i}t_{ij}^{[1]} + \theta_{3i}t_{ij}^{[2]} + \delta_1 G_i + \delta_2 (G_i \times t_{ij}^{[1]}) + \delta_3 (G_i \times t_{ij}^{[2]}) \qquad [10.8]$$

where $t_{ij}^{[1]} = \min(t_{ij}, 6)$, $t_{ij}^{[2]} = \max(t_{ij}, 6) - 6$, G_i represents the treatment group for patient i (0.5 for CT, -0.5 for NoCT) and δ_1, δ_2, δ_3 are fixed unknown regression parameters for treatment and treatment by time interactions. The individual intercepts and slopes are assumed to have a trivariate normal distribution with unknown mean $\boldsymbol{\alpha}$

and covariance matrix Σ as follows:

$$\begin{pmatrix}\theta_{1i}\\ \theta_{2i}\\ \theta_{3i}\end{pmatrix} \sim MVN\left(\begin{pmatrix}\alpha_1\\ \alpha_2\\ \alpha_3\end{pmatrix},\begin{pmatrix}\sigma_1^2 & \sigma_{12} & \sigma_{13}\\ \sigma_{12} & \sigma_2^2 & \sigma_{23}\\ \sigma_{13} & \sigma_{23} & \sigma_3^2\end{pmatrix}\right)$$

At the same time as modelling the quality of life data, the hazard of death for subject i is modelled using a piecewise exponential model as described in Section 7.6. In the MIC study an initial six-week time interval and 3 further three-week time intervals were used i.e. (0,6], (6,9], (9,12], (12,15] with all patients surviving after 15 weeks censored at this time. In the ESPAC study three-monthly time intervals up to 24 months were chosen (0,3], (3,6], (6,9], ..., (18,21], (21,24] with all patients that are known to survive longer 24 months censored at this time. The likelihood for the data d_{ik}, the number of deaths for patient i in time interval k (which can only be 0 or 1) is specified as follows:

$$d_{ik} \sim Poisson(\eta_{ik})$$

where η_{ik}, the mean number of deaths for patient i in time interval k, is modelled as:

$$\log(\eta_{ik}) = \log(r_{ik}) + \beta_{0k} + \beta G_i + \omega Q_{ik} \qquad [10.9]$$

In this model, r_{ik} is the length of time that patient i is at risk during interval k; G_i is the treatment covariate taking the same values as specified for the quality of life model with unknown regression parameter β; β_{0k} is the unknown underlying constant hazard in time interval k; Q_{ik} is the time-dependent covariate representing the quality of life of patient i in time interval k with associated unknown regression parameter ω. This model was fitted to the survival data in Chapter 7 and values for Q_{ik} were estimated from the observed data using imputation. Here the values included in the survival model are a function of the true unobserved value of quality of life estimated from the model for quality of life over time. Note that the inclusion of the treatment covariate in both parts of the joint model enables treatment to have not only a direct effect on the hazard of death but also an indirect effect through its effect on quality of life.

With the kth time interval defined as $(a_{k-1}, a_k]$, the options considered for Q_{ik} for both the MIC and ESPAC studies are as follows:

- The underlying value at the *start* of the interval: $Q_{ik} = \mu(a_{k-1})$
- The underlying *mean* of the values at each end of the interval: $Q_{ik} = \frac{\mu(a_{k-1}) + \mu(a_k)}{2}$

In addition for the MIC study, a third option was considered:

- The underlying *slope* over time: $Q_{ik} = \theta_{2i} + \delta_2 G_i$

Thus by replacing Q_{ik} in [10.9] by either the start, mean or slope value as defined above, this provides a link between the model for quality of life and the model for survival enabling the parameters in both models to be jointly estimated.

Prior distributions need to be specified for all unknown parameters in both models. Vague prior distributions were used for the separate modelling of quality of life and survival in Chapters 6 and 7 and the same prior distributions are considered here in the joint model.

In the quality of life model, normal distributions with relatively large variances are used for the fixed regression parameters as follows:

$$\alpha_p, \delta_p \sim N(0, 10000) \quad p=1,2 \text{ for MIC and } p=1,2,3 \text{ for ESPAC}$$

For both studies a Wishart distribution was used as a prior for the precision matrix Σ^{-1} as specified Section 6.6 with a Gamma distribution used as the prior for the inverse of σ^2_e as follows:

$$\frac{1}{\sigma_e^2} \sim Gamma(0.001, 0.001)$$

For the MIC study, the product normal formulation was considered as an alternative approach to specifying a Wishart prior distribution for $\mathbf{\Sigma}^{-1}$ as part of a sensitivity analysis. As described in Section 6.6.1, three different forms were considered for the variance terms; gamma prior distributions on the inverse variances (as specified for σ_e^2 above); half normal prior distributions with large variances on the standard deviations or uniform distributions over a wide range for the standard deviations. For the ESPAC study, the sensitivity analysis carried out in Section 6.6.2 showed that the model with the Wishart prior referred to as the base model (**R** with diagonal terms 200, 10, 1 and off-diagonal terms 0 and $\rho = 3$) was robust and this is used here in the joint model.

For the parameters in the survival model, normal distributions with large variances were used as follows:

$$\beta_{0j}, \beta, \omega \sim N(0, 10000) \qquad j=1,2,3,4 \text{ in MIC and } j=1,2,..,8 \text{ in ESPAC}$$

The marginal posterior distribution of all parameters in the joint model, given the data and the vague prior distributions, were obtained using Gibbs sampling. As described in Chapter 2, this process samples values from the full conditional distributions of each parameter given the data and the current values of all other parameters. WinBUGS was used to perform the analysis and the process was run for 50,000 iterations before retaining the next 50,000 sampled values as estimates from the posterior distributions. The WinBUGS code for the models for the MIC and ESPAC data is given in Appendix II.

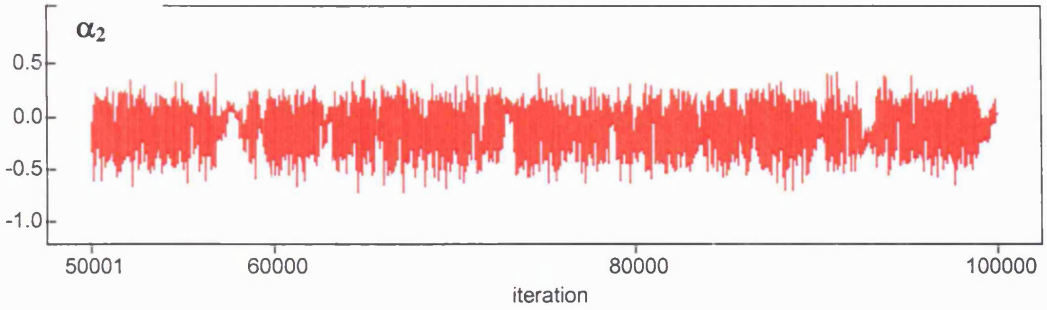
10.4.2 Results for the MIC Data

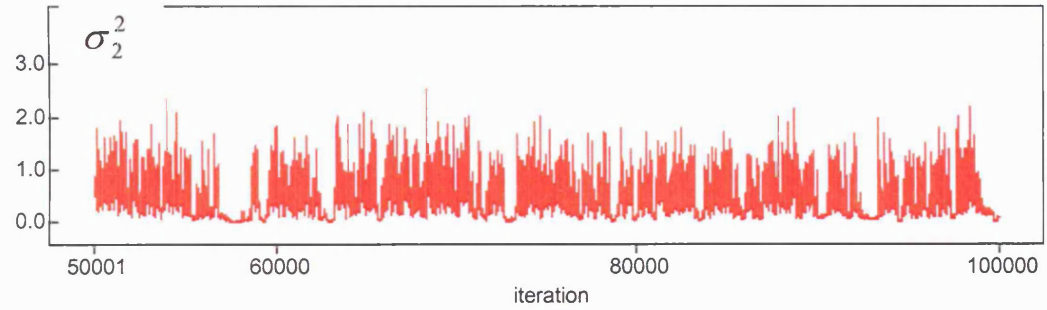
The results from the separate modelling of quality of life and survival in the MIC study using a Bayesian approach are given in Sections 6.6.1 and 7.6 respectively. In the separate modelling, the Bayesian approach with vague priors gave comparable estimates of model parameters to those obtained with a classical approach. Table 10.1 therefore includes the parameter estimates from the separate modelling obtained using a Bayesian

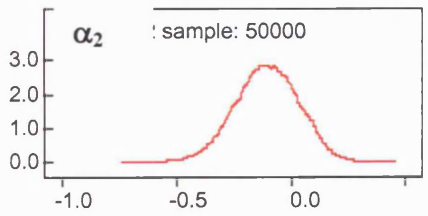
approach for comparison with those from the Bayesian joint model.

The performance of the models with the start, mean and slope values as the covariate in the survival model was similar. Examination of the trace plots from running each model showed that for the product normal formulation the process on occasions appeared to become trapped in a certain part of the density i.e. successive iterations had the same value, which occasionally produced non-standard shaped posterior densities (see example in Figure 10.1). This problem did not occur using the Wishart prior distribution, although the autocorrelations were generally high for some parameters with this form of the model, especially that with the slope value as the covariate in the survival model (see example in Figure 10.2 and output in Appendix IIIC). Taking a large number of sampled values i.e. 50,000 should overcome the problem of high autocorrelations. Despite the differences in performance, all the different forms of prior distributions gave comparable estimates of the parameters. To reduce the quantity of results presented, the parameter estimates obtained using a Wishart prior distribution are given here. The means and standard errors from the posterior marginal distributions for each parameter for all versions of the joint model (i.e. that with either the start, mean or slope value as covariate in the survival model) are given in Table 10.1 and with output from WinBUGS given in Appendix IIIA-D.

Figure 10.1: Trace plot and posterior density for α_2 and σ_2^2 in the joint model for the MIC data with start value as the covariate in the survival model and product normal formulation with gamma prior distributions on the precision parameters







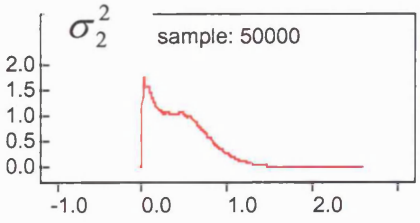


Figure 10.2: Trace plot and autocorrelations for δ_1, β and ω in the joint model for the MIC data with start value as covariate in survival model and Wishart prior distribution on the precision matrix

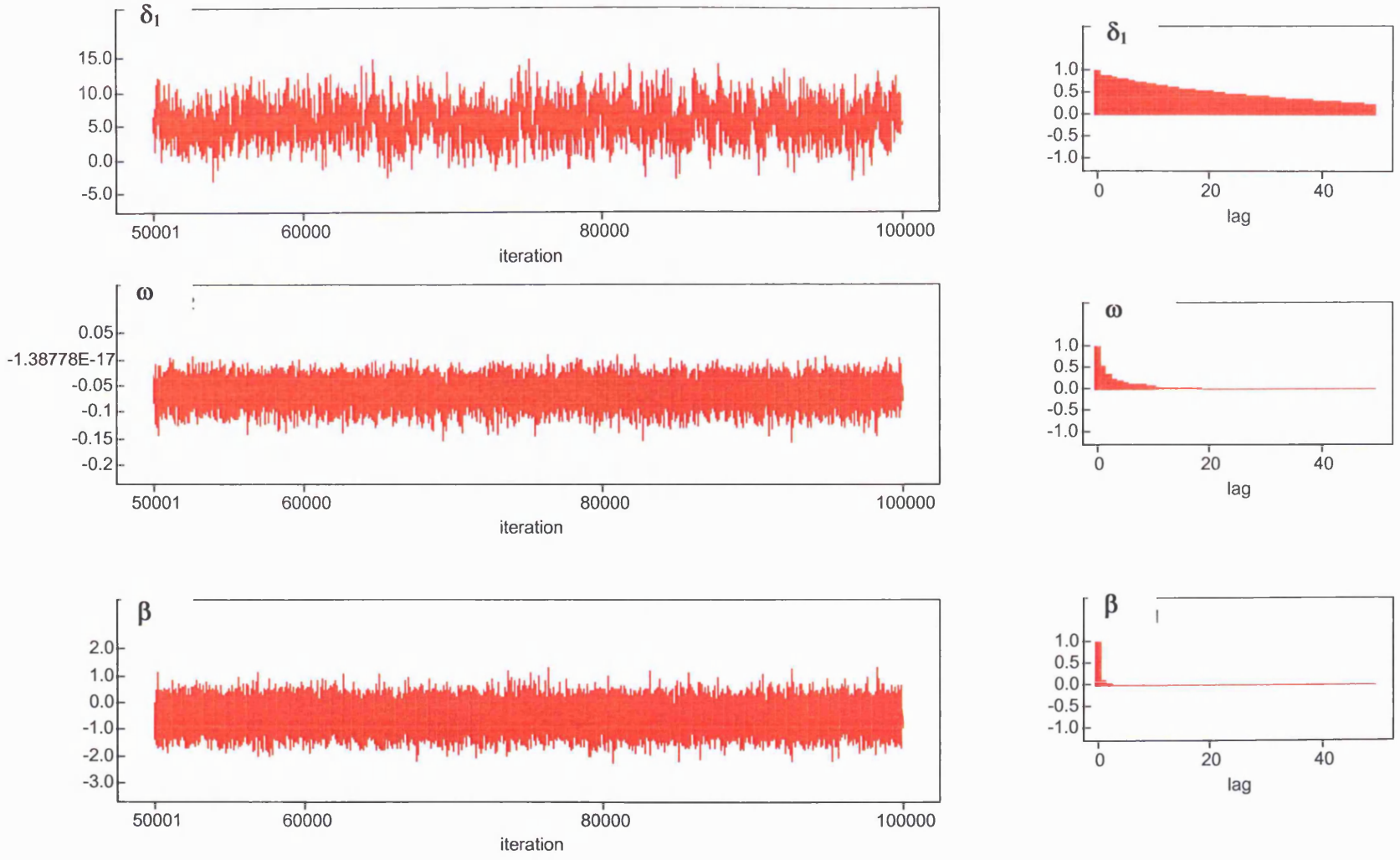


Table 10.1: Parameter estimates (and standard errors) from separate and joint modelling of quality of life (*GQS*) and survival data in the MIC study using a Bayesian approach with Wishart prior distribution; quality of life covariate values in survival model taken as value at start of interval, mean for interval and slope for change in *GQS* over time

		Model for *GQS*	Model for Survival		Joint Model for *GQS* and Survival			
			Start value as covariate	Mean value as covariate	Start value as covariate	Mean value as covariate	Slope value as covariate	Slope with interaction
Intercept	α_1	78.10 (1.14)			77.87 (1.17)	77.82 (1.17)	77.44 (1.20)	77.45 (1.21)
Time	α_2	-0.06 (0.13)			-0.11 (0.14)	-0.13 (0.14)	-0.20 (0.14)	-0.21 (0.13)
Treatment (on *GQS*)	δ_1	5.84 (2.27)			5.88 (2.33)	5.80 (2.28)	6.15 (2.27)	5.91 (2.37)
Treatment by time	δ_2	0.46 (0.26)			0.48 (0.26)	0.48 (0.27)	0.51 (0.26)	0.50 (0.26)
Variance Intercept (I)	σ_1^2	112.40 (18.80)			115.70 (19.22)	116.30 (19.38)	122.10 (20.42)	123.60 (20.59)
Variance Slope (S)	σ_2^2	0.34 (0.20)			0.43 (0.23)	0.48 (0.26)	0.35 (0.21)	0.37 (0.22)
Covariance I and S	σ_{12}	1.71 (1.48)			2.19 (1.56)	2.28 (1.65)	3.39 (1.45)	3.63 (1.51)
Error	σ_e^2	53.35 (5.55)			52.02 (5.49)	51.33 (5.51)	53.59 (5.22)	53.21 (5.23)
Treatment (on survival)	β		-0.66 (0.40)	-0.50 (0.41)	-0.49 (0.42)	-0.42 (0.43)	0.82 (1.33)	-0.78 (1.86)
GQS	ω		-0.04 (0.0129)	-0.05 (0.0134)	-0.06 (0.0199)	-0.06 (0.0201)	-3.87 (2.03)	-4.36 (2.09)
Treatment by *GQS*	ϕ							-2.73 (2.40)

The key purpose for this analysis is to adjust for the informative dropout due to death in the model for quality of life. The comparison of the separate and joint model parameter estimates for the quality of life model is therefore the key focus (see Table 10.1). The separate model for *GQS* showed that the on average *GQS* decreased slightly over time. In the joint model, as a result of accounting for the informative dropout due to death, the decrease over time becomes appreciably larger, particularly when the slope value is used as the covariate in the survival model. The standard errors show that the 95% credible intervals for these parameter estimates measuring change in *GQS* over time all contain zero. Despite having an effect on the overall change in *GQS* over time, the joint model does not appear to have an impact on the estimate for the treatment by time interaction. Even after adjusting for dropout due to death, the effect of treatment on the change in *GQS* over time remains the same, with CT on average improving quality of life over time (slopes for change in *GQS* range from 0.13 to 0.04 depending on the model) and PAL leading to deteriorating quality of life over time (slopes for change in *GQS* range from –0.35 to –0.46 depending on model). Standard errors show that zero falls just inside the 95% credible intervals for these parameters. The intercept, treatment effect, variance and covariance estimates for the joint model were in broad agreement with those from the separate model.

In terms of survival, the joint models with start value and mean value of modelled *GQS* used as the quality of life covariate are directly comparable to the separate models that use these values of *GQS* as a time-dependent covariate (see Table 10.1). Comparison shows that the effect of treatment on survival is reduced slightly with joint modelling. This may be because some of the effect of treatment on survival is now taken up through its effect on quality of life. The relationship between quality of life and survival in the joint model is comparable to that in the separate models but the standard errors are larger, as measurement error in *GQS* is now included in the model.

The joint model that includes the modelled slope value of *GQS* over time as a covariate in the survival model is more complex to interpret. It is not directly comparable to the models estimated for survival alone because the separate models measure the relationship between quality of life value and survival not *change* in quality of life and survival. When the model specified in [10.9] was applied to the data the treatment effect

changed in direction suggesting that after adjusting for the effects of change in quality of life on survival, chemotherapy *increased* the relative hazard of death. This unexpected result suggested that there might be an interaction between treatment and change in quality of life on survival and this term was added to the model. The results (final column in Table 10.1) showed that after including an interaction term, the treatment effect on survival returned to being negative. The negative regression coefficients for all terms in the model suggest that chemotherapy reduces the hazard of death and improving quality of life also reduces the hazard of death but the combination of improving quality of life on chemotherapy further contributes to the reduction in the hazard of death. The standard error for the effect of *GQS* slope value on survival was much greater than that for the effect of start and mean value of *GQS* on survival. This may be because the individual slopes for *GQS* over time were more variable across individuals than the predicted actual values, and this uncertainty is propagated through both models.

10.4.3 Results for the ESPAC Data

The results from the separate modelling of quality of life and survival in the ESPAC study from a Bayesian approach are given in Sections 6.6.2 and 7.6 respectively. In the separate modelling, the Bayesian approach with vague prior distributions gave comparable estimates of model parameters to those obtained with a classical approach. Table 10.2 therefore includes the parameter estimates from the separate modelling obtained using a Bayesian approach for comparison with those from the Bayesian joint model.

The performance of the two models with the start and mean values as the covariate in the survival model was similar. Examination of the trace plots and the autocorrelations showed that for all of the parameters except the regression parameter for treatment on survival (β) and the underlying constant baseline hazards, there were large autocorrelations at every lag (see examples in Figure 10.3 and output from WinBUGS in Appendix IV). The parameter estimates from 20,000 sampled values after a burn-in of 20,000 were compared to those from 50,000 sampled values after a burn-in of 50,000 and the values were very comparable suggesting that, despite the high autocorrelations,

the parameter estimates are reasonably robust. The means and standard errors from the posterior marginal distributions for each parameter in both versions of the joint model (i.e. that with either the start or mean value as covariate in the survival model) are given in Table 10.2 with output from WinBUGS given in Appendix IV.

Figure 10.3: Trace plot and autocorrelations for α_3, δ_1 and β in the joint model for the ESPAC data with mean value as covariate in survival model

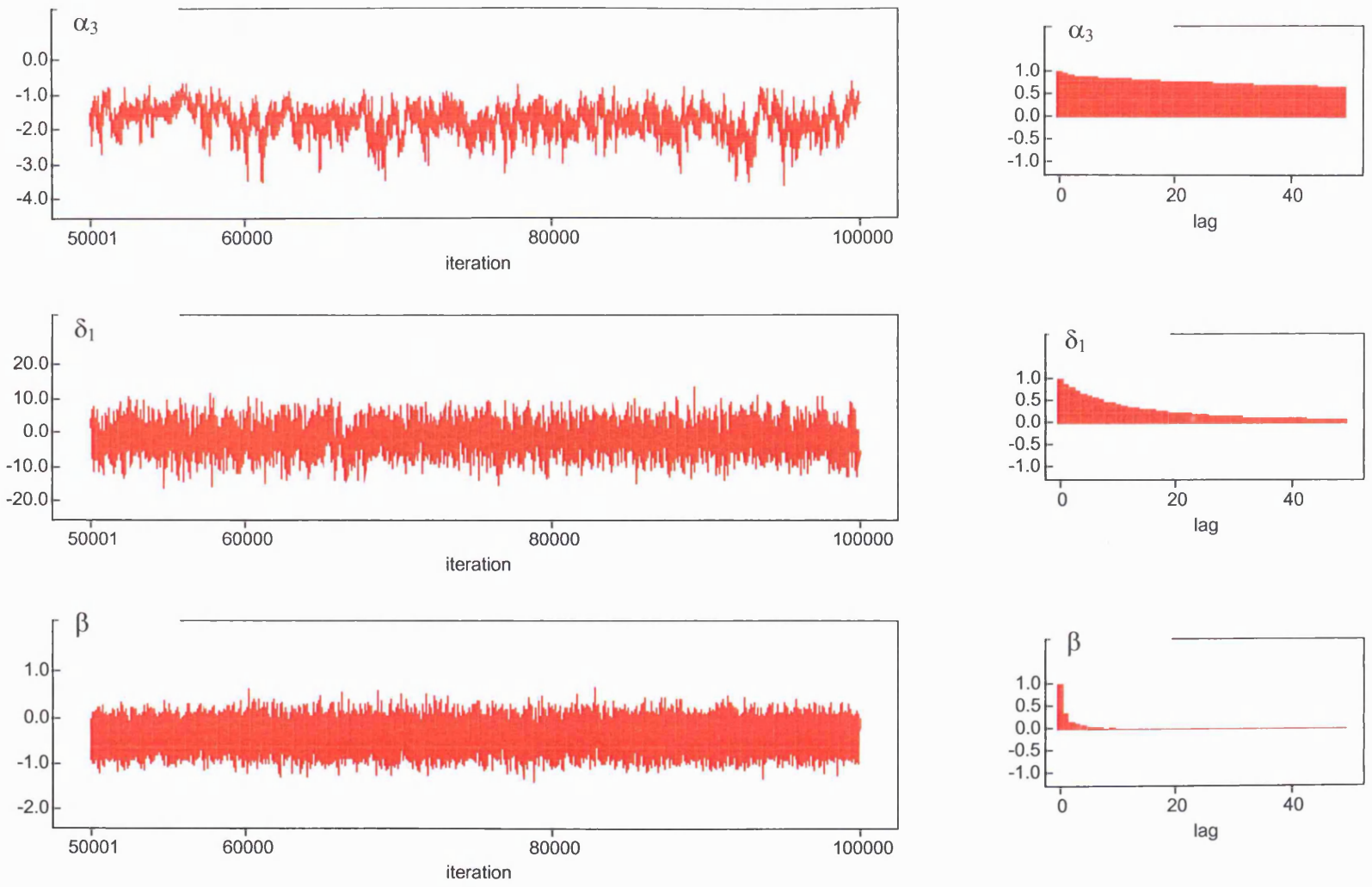


Table 10.2: Parameter estimates (and standard errors) from separate and joint modelling of quality of life (*GHSS*) and survival data in the ESPAC study using a Bayesian approach; quality of life covariate values in survival model taken as value at start of interval and mean for interval

		Model for *GHSS*	Model for Survival		Joint Model for *GHSS* and Survival	
			Start value as covariate	Mean value as covariate	Start value as covariate	Mean value as covariate
Intercept	α_1	51.05 (1.20)			50.99 (2.05)	51.28 (2.03)
Time0-6	α_2	2.82 (0.49)			2.78 (0.51)	2.68 (0.52)
Time6+	α_3	-0.87 (0.28)			-1.77 (0.39)	-1.96 (0.44)
Treatment (on *GHSS*)	δ_1	-0.97 (3.94)			-1.33 (3.82)	-1.37 (3.80)
Treatment by Time0-6	δ_2	0.32 (0.94)			0.30 (0.91)	0.31 (0.93)
Treatment by Time6+	δ_3	0.47 (0.44)			0.83 (0.50)	0.96 (0.53)
Variance Intercepts (I)	σ_1^2	203 20 (69.86)			210.4 (66.19)	223.9 (72.19)
Variance Slope(S)0-6	σ_2^2	9.75 (3.48)			10.22 (3.50)	11.31 (4.07)
Variance Slope(S)6+	σ_3^2	1.20 (0.57)			2.66 (1.14)	3.20 (1.39)
Covariance I and S0-6	σ_{12}	-22.83 (13.55)			-25.11 (13.16)	-27.60 (14.88)
Covariance I and S6+	σ_{13}	-10.14 (4.60)			-12.54 (5.81)	-14.52 (5.81)
Covariance S0-6 and S6+	σ_{23}	1.52 (0.92)			2.89 (1.30)	3.07 (1.35)
Error	σ_e^2	293.10 (22.84)			283.20 (22.47)	277.6 (22.00)
Treatment (on survival)	β		-0.47 (0.20)	-0.41 (0.20)	-0.39 (0.24)	-0.31 (0.26)
GHSS	ω		-0.02 (0.0043)	-0.03 (0.0045)	-0.06 (0.0136)	-0.07 (0.0134)

In the separate model for quality of life, *GHSS* on average increases over the first 6 months and then decreases over time from 6 months onwards. After adjusting for dropout due to death in the joint model (see Table 10.2), the initial slope is comparable but the decrease from 6 months onwards is more severe for both versions of the model.

In the separate model, the treatment by time interaction for the second period is such that the CT arm has less severe deterioration in quality of life than the NoCT arm and in both versions of the joint model this treatment effect is larger. The variance and covariance parameters are reasonably comparable across models.

In the survival model, the treatment effect on survival, which shows that CT reduces the relative hazard of death, is reduced in both versions of the joint model, possibly because some of the treatment effect is now taken up indirectly through the quality of life covariate (see Table 10.2). In fact the 95% credible interval for the treatment regression parameter does not include zero in the separate model but does in the joint model. The relationship between *GHSS* and survival is slightly increased in both versions of the joint model and has an increased standard error reflecting the fact that the model now incorporates measurement error.

10.5 Dealing with Dropout Prior to Death

In section 10.4, quality of life data was jointly modelled with survival data. In this joint model, the model for quality of life over time accounts for the dropout of patients from the quality of life study due to death, such that all missing data after death is no longer treated as missing at random in the quality of life model. The problem of dropout prior to death still remains an issue, as all missing data between the last recorded assessment and death will be treated as missing at random. For some patients this may be a valid assumption but in many cases the dropout will occur due to poor health.

One option is to impute the missing data between last assessment and death. This may be possible using other clinical data such as recorded toxicity, performance status or evidence of disease relapse. Alternatively an appropriate value of quality of life could be allocated to time points after dropout that reflect the assumed quality of life of the patients during this time. If the reasons for dropout are available then this could contribute to the imputation. The problem with these approaches is that the imputed values will be treated as *observed* values in the model. Multiple imputation provides an alternative approach that accounts for the uncertainty of the imputation but methods are complex and assume the missing data are missing at random (Schafer 2001).

An alternative approach is to model the time to dropout or death, whichever occurs first, instead of time to death and this will account for all missing data due to dropout or death. The approach of including other reasons for dropout as well as death in the joint model was advocated by Lindsey (1997) who modelled time to dropout with a survival model with patients who did not drop out being uninformatively censored. The model will treat all dropouts in the same way and thus assume the reason for dropout is the same for all patients. In particular it will treat dropout due to death equivalently to dropout prior to death. If the reason for dropout was known for some patients to be random such as administrative errors then these patients could be excluded from the dropout process until the point where they drop out due to death. In this thesis, since the reasons for dropout are not recorded, all dropouts are treated identically. The definition of time to dropout is subjective and was discussed in Section 5.8. The definitions, which differ slightly for the two studies are reiterated here for each of the studies.

10.5.1 Joint Modelling of Quality of Life and Time to Dropout for the MIC Study

The time to dropout was examined for the MIC study in Section 5.8.1. In the MIC study patients were required to complete a fixed number of quality of life assessments (5 on CT arm and 4 on PAL arm) within the 15-week study period, which makes the definition of time of dropout slightly more complex. For those patients who complete all scheduled assessments, the time of dropout is either (i) censored three weeks after their last completed assessment if they survive this time, or (ii) censored at 15 weeks if their last assessment is less than or equal to 3 weeks prior to the end of study, or (iii) recorded as an event at their time of death if they die within 3 weeks of their last assessment. For those patients who do not complete the final scheduled assessment (i.e. 5^{th} on CT and 4^{th} on PAL) the time of dropout is either (i) recorded as an event three weeks after their last completed assessment if they survive this time, or (ii) censored at 15 weeks if the last recorded assessment is less than 3 weeks prior to the end of study, or (iii) recorded as an event at their time of death if they die within 3 weeks of their last assessment.

Section 5.8.1 presents an analysis of time to dropout. Of the 109 patients, 47 dropped out within the 15-week analysis period (27 on CT and 20 on PAL). These are a combination of dropouts prior to death and dropouts directly due to death. Kaplan-Meier

curves showed a trend for patients on the PAL arm to dropout earlier than those on the CT arm but a log-rank analysis showed that this was not significant at the 5% level (p=0.22).

The same joint models that were specified in Section 10.4 are re-run with time to dropout (as defined above) used instead of time to death. As with the joint quality of life and survival model various forms of prior distribution were used in a sensitivity analysis. All forms gave comparable estimates for parameters. The results using a Wishart prior distribution are shown in Table 10.3. Three versions of the joint model are given for the different forms of quality of life covariate included in the dropout model: (i) *GQS* at start of each time interval, (ii) mean *GQS* for interval (iii) slope for *GQS* over time. Each version of the model is compared to the equivalent joint model of quality of life and survival that was presented in Section 10.4, which only accounts for dropout due to death.

Examination of the trace plots and autocorrelations showed that for the start and mean value versions of the model, the performance was similar to that with the survival data with the key issue being high autocorrelations. The trace plots for the slope version of the model indicated that the model occasionally sampled extreme values for a parameter but this was a rare occurrence and examining means and medians showed that these extreme values appear to have little influence on the final results.

Table 10.3: Parameter estimates (and standard errors) from joint modelling of quality of life (*GQS*) and time to dropout compared to the joint model with time to death in the MIC study using a Bayesian approach with a Wishart prior distribution

		Start value as covariate		Mean value as covariate		Slope value as covariate	
		Death	**Dropout**	**Death**	**Dropout**	**Death**	**Dropout**
Intercept	α_1	77.87 (1.17)	77.91 (1.15)	77.82 (1.17)	77.82 (1.17)	77.45 (1.21)	77.15 (1.23)
Time	α_2	-0.11 (0.14)	-0.10 (0.14)	-0.13 (0.14)	-0.13 (0.14)	-0.21 (0.13)	-0.30 (0.16)
Treatment (on *GQS*)	δ_1	5.88 (2.33)	5.72 (2.31)	5.80 (2.28)	5.83 (2.26)	5.91 (2.37)	5.95 (2.28)
Treatment by time	δ_2	0.48 (0.26)	0.43 (0.26)	0.48 (0.27)	0.44 (0.27)	0.50 (0.26)	0.49 (0.27)
Intercept variance	σ_1^2	115.70 (19.22)	114.30 (18.96)	116.30 (19.38)	115.80 (19.32)	123.60 (20.59)	123.90 (21.01)
Slope variance	σ_2^2	0.43 (0.23)	0.43 (0.26)	0.48 (0.26)	0.50 (0.26)	0.37 (0.22)	0.48 (0.28)
Covariance	σ_{12}	2.19 (1.56)	1.97 (1.52)	2.28 (1.65)	2.27 (1.68)	3.63 (1.51)	3.51 (1.68)
Error	σ_e^2	52.02 (5.49)	52.08 (5.53)	51.33 (5.51)	51.02 (5.49)	53.21 (5.23)	51.45 (5.35)
Treatment (on event)	β	-0.49 (0.42)	-0.11 (0.32)	-0.42 (0.43)	-0.05 (0.32)	-0.78 (1.86)	0.59 (1.16)
GQS	ω	-0.06 (0.02)	-0.04 (0.01)	-0.06 (0.02)	-0.04 (0.01)	-4.36 (2.09)	-2.51 (1.22)
Treatment by *GQS*	ϕ					-2.73 (2.40)	-0.47 (1.19)

For the start and mean value versions of the joint model, modelling time to dropout rather than time to death has very little effect on the parameter estimates in the quality of life part of the model. When slope is included as a covariate in the dropout model, the reduction in *GQS* over time is further reduced after accounting for dropout, although the treatment by time interaction remains stable. In the dropout part of the model with start and mean values as covariates, the parameter estimates for the treatment covariate indicate that chemotherapy reduces the relative hazard of dropout, although in all cases the 95% credible interval contains zero. Also the hazard of dropout is reduced for increased values of *GQS* (95% credible interval –0.07 to –0.01) confirming that dropout is informative. The estimates for the dropout part of the model with the slope as the covariate are more unexpected and very different to the results from the survival model. In this model, treatment appears to have an opposite effect on dropout with chemotherapy increasing the hazard of dropout. The model was re-run excluding the interaction term but the regression parameter for the treatment effect on time to dropout remained positive. The parameter estimates are relatively small with large standard errors so this discrepancy between models for survival and dropout could be merely an effect of the estimates being close to zero.

10.5.2 Joint Modelling of Quality of Life and Time to Dropout for the ESPAC Study

The time to dropout was examined for the ESPAC study in Section 5.8.2. Since patients are required to complete assessments at three-monthly intervals until death, the definition of time to dropout is more straightforward than the MIC study. For those patients whose last recorded assessment was more than 3 months prior to their last survival follow-up, the time of dropout was recorded as an event 3 months after their last recorded assessment. For those patients who die within 3 months of their last recorded assessment, the time of dropout is recorded as an event at their time of death. For those patients whose survival time is censored within 3 months of their last recorded assessment, the time of dropout is censored at this time too. Section 5.8.2 shows an analysis of time to dropout. In addition to the 62 patients who dropped out due to death, a further 87 dropped out prior to death, with the remaining 26 having censored dropout times at their last follow-up time. Kaplan-Meier curves showed a trend for patients on

the NoCT arm to dropout earlier than those on the CT arm and a log-rank analysis showed that this was just statistically significant at the 5% level (p=0.05). The analysis here is restricted to 24 months from trial entry and if a patient's dropout time is greater than 24 months then their time of dropout is censored at 24 months.

As with the MIC study, the same joint model that was specified in Section 10.4 is re-run with time to dropout used instead of time to death, with start and mean values of *GHSS* included as covariates in the dropout model. The results are shown in Table 10.4 with each version of the model compared to the equivalent joint model of quality of life and survival that was presented in Section 10.4, which only accounts for dropout due to death. Examination of the trace plots and autocorrelations showed that the performance of the joint model with dropout was similar to that with the survival function.

Table 10.4: Parameter estimates (and standard errors) from joint modelling of quality of life (*GHSS*) and time to dropout compared to the joint model with time to death in the ESPAC study using a Bayesian approach

		Start value as covariate		Mean value as covariate	
		Death	**Dropout**	**Death**	**Dropout**
Intercept	α_1	50.99 (2.05)	50.82 (2.04)	51.28 (2.03)	51.13 (1.93)
Time0-6	α_2	2.78 (0.51)	2.97 (0.51)	2.68 (0.52)	2.82 (0.49)
Time6+	α_3	-1.77 (0.39)	-2.39 (0.50)	-1.96 (0.44)	-2.77 (0.52)
Treatment (on *GHSS*)	δ_1	-1.33 (3.82)	-3.05 (3.88)	-1.37 (3.80)	-2.08 (3.90)
Treatment by Time0-6	δ_2	0.30 (0.91)	0.88 (0.92)	0.31 (0.93)	0.54 (0.93)
Treatment by Time6+	δ_3	0.83 (0.50)	1.02 (0.60)	0.96 (0.53)	1.31 (0.63)
Variance Intercepts (I)	σ_1^2	210.4 (66.19)	205.4 (61.09)	223.9 (72.19)	212.4 (63.52)
Variance Slope(S)0-6	σ_2^2	10.22 (3.50)	10.18 (3.46)	11.31 (4.07)	10.69 (3.36)
Variance Slope(S)6+	σ_3^2	2.66 (1.14)	4.03 (1.64)	3.20 (1.39)	4.88 (1.86)
Covariance I and S0-6	σ_{12}	-25.11 (13.16)	-25.52 (12.66)	-27.60 (14.88)	-25.43 (12.45)
Covariance I and S6+	σ_{13}	-12.54 (5.81)	-19.31 (6.99)	-14.52 (5.81)	-22.22 (7.47)
Covariance S0-6 and S6+	σ_{23}	2.89 (1.30)	4.17 (1.55)	3.07 (1.35)	4.63 (1.68)
Error	σ_e^2	283.20 (22.47)	287.7 (22.34)	277.6 (22.00)	284.1 (22.06)
Treatment (on event)	β	-0.39 (0.24)	-0.26 (0.22)	-0.31 (0.26)	-0.16 (0.24)
GHSS	ω	-0.06 (0.01)	-0.07 (0.01)	-0.07 (0.01)	-0.07 (0.01)

When time to dropout is modelled in the ESPAC data, rather than time to death, the reduction in *GHSS* over the second time period from 6 months onwards time is further reduced after accounting for dropout. The interaction of treatment with both the slope prior to 6 months and the slope after 6 months is also increased slightly, indicating an increased effect of treatment on change in quality of life after adjusting for dropout. In the dropout part of the model, the parameter estimates for the treatment covariate indicate that chemotherapy reduces the relative hazard of dropout, although in all cases

the 95% credible interval contains zero. Also the hazard of dropout is reduced for increased values of *GHSS* confirming that dropout is informative.

10.6 Discussion and Critical Review of Joint Modelling

Joint modelling of quality of life and survival data enables the estimates in the model for quality of life to account for dropout due to death and thus any missing data that occurs due to death will not be treated as missing at random. At the same time the modelling of survival data with quality of life as a time-dependent covariate is enhanced since covariate values are estimated from the model for quality of life data over time fitted to all subjects. By modelling time to dropout rather than survival time the approach can overcome the problem of informative dropout prior to death in the quality of life model and examine the relationship between the true underlying value of quality of life and time to dropout.

Classical approaches to joint modelling have been successfully applied to quality of life data (Fairclough et al 1998a, Ribaudo et al 2000, Curran et al 2002, Michiels et al 2002, Pauler et al 2003) but the methods can be problematic. Selection models can be computationally intensive and require specialised software and large amounts of data (Hogan and Laird 1997a). Pattern mixture models require a subjective stratification of patients into subgroups according to their dropout pattern. The numbers of patients in each subgroup may not be sufficient for adequate modelling if the study is relatively small and assumptions may be required to fit the model in subgroups with sparse data.

In general, the Bayesian approach to joint modelling allows a greater degree of flexibility. The models for each process can be as complex as required, though a balance has to be maintained between complexity and ease of interpretation. Gibbs sampling simplifies the estimation of parameters in complex models (Gilks et al 1993). Availability of WinBUGS software has made such methods more accessible to researchers although caution is required to ensure that WinBUGS is not used as a 'black box tool for statistical analysis' (Best et al 1996). Convergence may be a problem, especially with small numbers. Examination of trace plots, autocorrelations and density plots are essential to assess model convergence and sensitivity analysis should be carried

out to ensure that choice of starting values, length of burn-in, length and thinning of sampled chain does not affect the results. Analysis should also be carried out to assess the sensitivity of the results to choice of prior distribution including the type of distribution and the level of vagueness. Such analysis can be very time-consuming especially with more complex models that may take some time to run. Although the analysis here used vague prior distributions, the method also provides a means for including external data sources as prior information in the estimation of parameters (Spiegelhalter et al 2003). Bayesian joint models have been compared indirectly to the classical approach by comparing the classical and Bayesian parameter estimates for each part of the model but further work is required to compare the Bayesian joint modelling approach directly to the classical approach.

Unlike quality-adjusted survival analysis where quality of life needs to be a utility-type measure, the joint modelling approach does not restrict the quality of life variable to any particular scale, however, the distribution of the variable is relevant. In the joint models considered here the focus has been on models that assume the repeated measures of quality of life follow a normal distribution. This assumption will often be plausible, especially if the quality of life measure for analysis is a global measure. In some circumstances, an appropriate transformation of the quality of life outcome may improve the normality of the data and the transformed variable can be modelled instead of the original quality of life variable. In some cases the assumption of normality will be untenable, especially if the outcome is an ordinal variable with few categories. The Bayesian approach however could easily be adapted to be applicable to any type of distribution. By using general linear mixed models, non-normal quality of life data can also be accommodated (Xu and Zeger 2001).

The repeated measures data, in a joint model with time-to-event data, is generally modelled using a mixed effects model and in particular a linear random effects model (for example Wu and Bailey 1988, Schluchter 1992, Berzuini 1995, Faucett and Thomas 1996, Wulfsohn and Tsiatis 1997, Pauler et al 2003). The assumption of a linear trend in quality of life over time will often be plausible or it may be possible to transform either the quality of life variable or the time variable to ensure that it is (Carpenter et al 2000). Modelling the change in quality of life over time as a linear function provides the most

easily interpreted results from a clinical viewpoint. In circumstances where a linear trend is inappropriate a higher order polynomial function could be considered for the model or, as implemented here for the ESPAC data, a piecewise linear random effects model may be applicable (De Stavola and Christensen 1996, Ribaudo et al 2000, Pauler and Finkelstein 2002). The latter will generally be more easily interpreted from a clinical point of view than a complex polynomial function. The piecewise linear random effects model implemented here assumed a single fixed pre-defined change-point. In general, any number of change-points can be used that are appropriate for the clinical situation and the given data. The change-points can be included in the model as fixed effects or as random effects such that the values are allowed to vary between individuals (Pauler and Finkelstein 2002). The random effects models considered in most papers on joint modelling, and in this thesis, assume that the random effects are normally distributed. Joint models that do not put any restrictions on the distribution the random effects have been developed (Tsiatis and Davidian 2001). An alternative form of the random effects model, which has been shown to be more appropriate for modelling CD4 counts over time and has been used in a joint modelling framework using Bayesian estimation, includes an integrated Ornstein-Uhlenbeck stochastic process (Wang and Taylor 2001). Instead of including a random effect for the slope, the model includes a fixed effect for the slope representing the population average rate of change in the dependent variable and a subject-specific function of time that allows for the random fluctuation around the population average.

The joint modelling considered in this chapter assumes that there is a single measure of quality of life for which there are repeated measures over time. Often there are several measures of quality of life, such as different dimensions measured by an instrument, that may require analysis. One option is to repeat the joint modelling for each outcome independently and make separate inferences relating to each one. This however does not account for the correlations between the measures. The random effects models considered in this thesis are simple univariate forms of a hierarchical model with repeated observations at the lowest level nested within patients at a second level. The advantage of these models in the quality of life setting is their ability to be extended to a third level to enable two or more quality of measures to be considered in a single model and allowing the relationship between the measures to be investigated (Beacon 1996,

Beacon and Thompson 1996). This methodology has recently been extended to jointly model measures of quality of life with survival using Schluchter's model in a classical approach (Ribaudo et al 2000). The pattern-mixture type of approach using multi-level models suggested by De Stavola and Christensen (1996) provides a potentially useful approach for quality of life data as these models can be extended to incorporate more than one quality of life outcome.

The event in the time-to-event data is often death, disease progression or dropout. For the time-to-event component of the joint model, a Cox regression model has been widely used both in the classical setting (Wulfsohn and Tsiatis 1997, Henderson 2000) and particularly the Bayesian setting (Faucett and Thomas 1996, Faucett et al 1998, Xu and Zeger 2001, Wang and Taylor 2001, Wang et al 2002, Pauler and Finkelstein 2002). In the classical approach to joint modelling an arbitrary baseline hazard is usually assumed but with the Bayesian approach to joint modelling it is not possible to assume an arbitrary baseline hazard since a fully parametric approach is required. The piecewise exponential baseline hazard is the closest alternative and the most flexible form for the baseline hazard. This is the form of the model that is generally used in the Bayesian setting and is the model that is used in this thesis. The model requires at least one death in each interval and thus if there are no early deaths then the first interval in the piecewise exponential model will be wide and the details of the changes in quality of life over this interval will be lost.

Other parametric models could be used for the survival data although this would be more difficult in a selection model formulation, as the incorporation of time-dependent covariates in parametric survival models is not well developed (Petersen 1986). If the slope from a linear random effects model representing a subject's change in quality of life over time is used as the covariate in the survival model then this ceases to be time-dependent and a fully parametric form for the model rather than a piecewise exponential can be considered.

In the joint models used in this thesis, some measure of the underlying value of quality of life is included as a covariate in the survival model. The covariate value at a given point in time is determined from the model for quality of life to which values previous

to and after the given time point have contributed. In this way the survival model is conditioning on future events. This may be a problem if the model was to be used for predictive purposes.

The joint models considered here have only included treatment as a covariate in each part of the model. Other covariates that are thought to explain the variation in quality of life over time and survival could be included in either or both parts of the joint model. With a Bayesian approach, model averaging over several different models with different covariates could be considered (Draper 1995) or Bayes factors could be used to compare models (Spiegelhalter et al 2003). One of the advantages of the joint model illustrated in this thesis is that the inclusion of the same covariate in both parts of the model enables the direct effect of the covariate on the hazard of death to be estimated in addition to the indirect effect via the effect of quality of life on the hazard (Wang and Taylor 2001).

In this thesis the time to dropout has been modelled jointly with quality of life to adjust for missing data resulting from informative dropout. Lindsey (1997) believes that the dropout process should be modelled together with the longitudinal measurements process since it is an integral part of the phenomenon under study. In this analysis, dropout prior to death and dropout due to death are treated as equivalent events, such that once a patient drops out prior to death, their time of death after dropout is not accounted for in the model. It may be appropriate to consider the two forms of dropout as separate types of event, and which both need to be accounted for in the joint model. It may be possible to incorporate a model for the dropout process to account for data missing for reasons other than death (Lambert et al 1999). Further to this, it may be appropriate to account for more than one type of dropout prior to death if the different reasons for dropout have been recorded. The need to distinguish between dropout due to death and dropout prior to death has been addressed (Pauler et al 2003). In a pattern-mixture model approach, the parameter estimates in the model for quality of life over time are combined across the dropout patterns but remain conditional on survival. The authors concede that it results in a non-randomised comparison of treatment arms in terms of quality of life outcome, which is exactly what the joint modelling is trying to address. Shih and Quan (1997) raise concerns about modelling the hypothetical complete data marginal distribution for quality of life suggesting that it does not make

clinical sense to make inferences about quality of life for those who have died. As Pauler et al (2003) highlight, this is the only way to provide an unbiased comparison of treatment arms.

Unlike some forms of quality-adjusted survival analysis, joint modelling does not necessarily require the analysis to be restricted to a specific time period. If the quality of life data is collected during a fixed time period then it may be wise to restrict the survival data also to this period so that the model for quality of life fitted to the data is not extrapolated outside the period of data collection. It may also be preferable to fix the time period for the analysis as the data becomes sparse as the time from entry increases. In restricting the analysis to a fixed time period, values of quality of life that fall outside that time period could still contribute to the estimation of the model for quality of life over time even though only the values of quality of life within the fixed time period will be included as covariates in the survival model. For example in the MIC study there were 7 values of quality of life that fell just outside the 15-week cut-off period and as with other analyses in this thesis, they were included in the estimation of the quality of life model as they were the only estimates of post-treatment quality of life available for those patients. In other studies it may be more appropriate to exclude the values measured outside the fixed analysis period.

The application of joint modelling of quality of life and survival data to the MIC and ESPAC studies showed that by adjusting for informative dropout due to death, the deterioration in quality of life over time that was observed in the unadjusted model becomes more severe. In these examples adjusting for dropout prior to death as well as directly due to death further increased the severity of the deterioration. The treatment effects in these examples were not greatly affected by the joint modelling. The inclusion of the slope as the covariate in the survival model produced more extreme results than just including the actual value of quality of life over time. This indicates that modelling the relationship between the change in quality of life and time to death or dropout in the joint model has a greater impact on the conclusions. With the patient profiles for quality of life over time being relatively noisy, using the individual slopes as the covariate in the survival model may not provide stable results. Joint modelling also had an effect on the relationship between quality of life and survival measured in the survival model with

standard errors for the regression parameters being increased due to accounting for measurement error. The impact of joint modelling on the data is similar to that has been previously observed in simulations (Faucett and Thomas 1996).

In summary, the Bayesian approach to joint modelling has been shown to be a practical methodology for the simultaneous analysis of quality of life and survival data. Sensitivity analysis is crucial to assess the effect of model assumptions on the results. The validity of inferences from the joint model relies on appropriateness of models used and methods for assessing the fit of joint models requires further research (Faucett et al 1998).

CHAPTER 11: CONCLUSIONS AND DISCUSSION

11.1 Summary of Research

In the assessment of treatments in clinical trials, particularly those in cancer, quality of life is often an important endpoint alongside length of survival. Quality of life data are usually collected via patient-completed questionnaires in a longitudinal study. The key problem in the analysis of the longitudinal quality of life data is that patients drop out of the quality of life study due to death and all missing data after death will not be missing at random. In addition, in studies where patients are sufficiently ill for length of survival to be an outcome, subjects will often drop out of the quality of life study prior to death due to illness. In such circumstances the missing data resulting from dropout prior to death will also not be missing at random. Standard methods for longitudinal analysis that assume missing data are missing at random will give biased results and methods for the simultaneous analysis of quality of life and survival or dropout that provide an unbiased analysis were investigated.

Three different approaches have been investigated for the simultaneous analysis of quality of life and survival. Quality-adjusted survival analysis compares treatments in terms of a composite measure of quality and quantity of life. Multistate modelling explores how treatments differ in terms of the transition rates between various health states defined by levels of quality of life and death. Joint modelling describes the quality of life and survival data in terms of two interlinked models enabling treatments to be compared in terms of each endpoint but adjusted for the influence of the other. Methods were applied to data from two clinical trials, one in patients with non-small cell lung cancer and one in patients with pancreatic cancer. These are typical of the quality of life studies generally encountered in cancer clinical trials. Application of the methods to the data from these trials provides insight into the practical issues associated with such approaches. The direct comparison of these three methods for the simultaneous analysis of quality of life and survival data has not previously been undertaken, neither has the extension of these methods to deal with additional dropout prior to death

There a four key methods for quality-adjusted survival analysis defined by whether the aggregation of quality of life and survival is at the subject or the group level and whether the longitudinal values of quality of life data are used directly to down-weight survival time or used to determine the time spent in specific health states used in the model. The application of all four methods to longitudinal quality of life data that is carried out in this thesis has not been demonstrated previously and enables the methods to be reviewed and compared in detail. In addition, the methodology is extended to deal with the problem of informative dropout prior to death. In particular, in relation to the group-based approach using actual values of quality of life known as the integrated quality-survival product, the thesis recognises that quality-adjusted survival can be interpreted as survival-adjusted quality-of-life. This new concept reflects the fact that the method adjusts for the dropout of patients due to death in the analysis of quality of life. By interpreting the quality-adjusted survival analysis in this way, the extension of the methodology to deal with informative dropout prior to death can be meaningfully interpreted as dropout-adjusted quality-of-life.

Multistate modelling has been advocated as a possible means for analysing quality of life and survival data simultaneously and although this approach has been applied to data with health states defined by clinical criteria, there are limited examples of applications of multistate models to quality of life data. In this thesis, multistate models under a variety of different modelling assumptions were applied to the illustrative examples, allowing the feasibility of such methodology to be investigated. In addition the development of the method to include a specific dropout state provided a means of overcoming the problem of informative dropout from the quality of life study prior to death. This approach for dealing with dropout in a longitudinal study has not previously been considered.

Joint modelling of repeated measures and time-to-event data has been an expanding area of development. The majority of the research is based on classical approaches and some of these classical approaches have been applied to longitudinal quality of life data together with survival or dropout data in order to overcome the problem of informative dropout. The Bayesian approach to joint modelling has been described by a number of authors but the application of such an approach in the field of quality of life has, until recently, not previously been undertaken and was therefore chosen as the focus for

research. This thesis develops Bayesian joint modelling for application to quality of life and survival data and extends the method to account also for dropout prior to death.

11.2 Discussion of Findings

11.2.1 Separate Versus Simultaneous Analysis

Simultaneous analysis of quality of life and survival provides an overall assessment of the treatments in a clinical trial in terms of both endpoints. Simultaneous analysis could be said to supplement the separate analysis of each endpoint but in fact separate analysis of quality of life in general should not be considered. The analysis of survival as a single endpoint is not a problem and will often be considered as the primary endpoint in cancer clinical trials for example but the analysis of quality of life as a separate endpoint is problematic. Analysing longitudinal quality of life data that is subject to informative dropout due to death will provide inferences on the quality of life over time conditional on patients surviving but this does not provide an unbiased comparison of treatments since at any point in time it is not comparing like with like. The only truly unbiased comparison is an intention-to-treat type analysis in which all patients at all time points are included in the analysis. This unbiased comparison can only be achieved by the simultaneous analysis of both endpoints. Some have questioned the meaning of such simultaneous analysis especially in the context of joint modelling, suggesting that it does not make clinical sense to make inferences about the quality of life for those who have died (Shih and Quan 1997). Simultaneous analysis of quality of life and survival however is necessary to provide an unbiased comparison of treatments.

11.2.2 Overall Comparison of the Three Simultaneous Methods

The health state methods for quality-adjusted survival analysis are based on a similar methodology to multistate modelling and hence both have similar advantages and disadvantages and both deal with the problem of dropout prior to death in a similar way. In both approaches, health states are defined by the quality of life outcomes being measured and the longitudinal quality of life data are used to estimate the time of transition from one state to the next. In the quality-adjusted survival analysis these

transition times are used to estimate the time spent in each state for combining with appropriate utility values whereas in multistate modelling the transition times are used to model the hazards of transition from one health state to the next. Methodology has been proposed that combines the methods in a single analysis, using multistate modelling to estimate the time spent in different health states for incorporation into a quality-adjusted survival analysis (Cole et al 1994). Further research into this approach, possibly within a Bayesian framework is required.

The advantage of these health state methods is the fact that no distributional assumptions are required about the quality of life outcome. The method can be applied to any type of outcome and can incorporate multiple outcomes into the definition of the health states. The definition of health states however is not straightforward. The choice is subjective and choosing states that are clinically meaningful and at the same time include sufficient data for adequate estimation may not always be possible. Sensitivity analysis should be used to assess the robustness of the results to the definition. The health state approach may be problematic when there are censored survival times within the analysis period as the group-based approach of partitioned survival analysis that overcomes the problem of censoring requires progressive health states which will often not be easy to define in terms of quality of life. The main advantage of multistate modelling over the group-based approach to quality-adjusted survival analysis, is that the health states in the model do not need to be progressive. The other potential problem with quality-adjusted survival analysis is that utility values are required for the different health states in order to be able to combine the results into a composite measure.

The health state methods both deal with dropout prior to death in a similar way. Patients are assumed to move to a dropout state at an appropriately defined time, such as the time when the next assessment was due. In the quality-adjusted survival analysis, the time spent in the dropout state can be allocated a utility value that reflects the expected quality of life in that state. In the multistate modelling, the transition rates to the dropout state and from the dropout state to death can all be explicitly modelled. The health state methods allow dropout prior to death to be treated differently to dropout due to death by treating them as two separate states in the model. Given sufficient patients in a study it would also be possible to extend this to have different dropout states relating to

different reasons. These would need to be progressive for partitioned survival analysis but not for multistate modelling.

Quality-adjusted survival analysis is a more straightforward method than multistate modelling. The results are expressed in terms of a composite measure of quality and quantity of life and are therefore likely to be more clinically intuitive. The methods involve the familiar simple concept of area under a curve and are therefore accessible to statisticians and clinicians. Multistate modelling requires more assumptions than quality-adjusted survival analysis in order to be able to model the transition rates. Multistate modelling decomposes the quality of life and survival processes, which may be more difficult to interpret but may provide a better insight into clinical pathway of patients. In addition the models may be used to provide insight into the missing data mechanism by comparing the risk of death from the dropout state with that from other health states in the model.

The group-based method for quality-adjusted survival analysis that directly uses the longitudinal values of quality of life in an integrated quality-survival product (IQSP) is comparable to the joint modelling. In the IQSP the function for quality of life over time is multiplied by the survivor function to give a quality-adjusted survivor function or survival-adjusted quality-of-life function whereas in the joint modelling the parameters in the models for quality of life over time and survival are estimated simultaneously.

One advantage of the IQSP is that the method does not require parametric forms for the quality of life and survivor functions whereas the joint modelling is a fully parametric approach. The IQSP approach however does require the quality of life measure ideally to be a utility measure although transformation of a descriptive measure to a utility-type measure enables the method to be applied to other measures. The interpretability of the results when the quality of life outcome is not a true utility, just a measure on a 0 to 1 scale, is debatable. As the method of IQSP becomes more widely known, quality of life studies will be designed to collect utility measures either alongside or instead of descriptive measures of quality of life enabling the results from IQSP to be more valid.

The parametric nature of joint modelling more readily allows the inclusion of covariates in the analysis and has greater scope for development to more complex scenarios,

especially within a Bayesian framework. Possible extensions include the incorporation of cost data in a cost-utility analysis and multiple measures of quality of life and the feasibility to deal with the issue of quality of life only collected in a subgroup of trial patients (see Section 11.3). The Bayesian rather than the classical approach to joint modelling was pursued as it allows complex models to be fit in a relatively straightforward way and the methodology had not previously been considered in the area of quality of life research. Vague prior distributions were used as it was not the aim of the thesis to include prior information into the analysis, but the methodology could be extended to incorporate external evidence, such as results from other studies on quality of life in similar patient populations, if so desired. Bayesian joint models have been compared indirectly to the classical approach by comparing the classical and Bayesian parameter estimates for each part of the model but further work is required to compare the Bayesian joint modelling approach directly to the classical approach.

Both IQSP and joint modelling deal with dropout prior to death by analysing the dropout-free survivor function rather than the survivor function. Patients are defined as dropping out of the quality of life study either at their time of death or at the time of dropout, however that is defined, if dropout occurs prior to death. Unlike the health state approach, this does not distinguish between dropout prior to death and dropout directly due to death. This may not always be appropriate and further research to extend the joint models to incorporate a model for the dropout prior to death process as well as the death process is required.

11.2.3 Overall Conclusions from the MIC Study

The analysis of the MIC study was restricted to 15 weeks from entry to trial as this was the period of time during which quality of life data were collected. This study design does not permit inferences regarding the long-term effect of treatments on quality of life but enables the treatments to be compared during probably the most crucial part of their survival time. Even during this short period of time, the number of deaths was large enough to potentially create a biased treatment comparison in terms of quality of life, especially since the hazard of death was greater with standard palliative care than chemotherapy. In addition, examining dropout-free survival showed a differential dropout rate between the two treatment arms. Simultaneous analysis of quality of life

and survival was essential to provide a more unbiased treatment comparison and extending the analysis to dropout-free survival should reduce the potential bias further.

The analysis of the MIC study considered two different global measures of quality of life; the overall mean response to all items on the study questionnaire, *GQS*, and the single item that assesses patient malaise, *MAL*. Although *GQS* is made up of a number of ordinal responses it is reasonably continuous and is treated as such in the analysis presented. Each method of simultaneous analysis makes different assumptions about the measure. Quality-adjusted survival analysis assumes that *GQS* is a utility measure, multistate modelling assumes that summary measure values will differentiate patients into distinct health states and the joint modelling presented here assumes that *GQS* is normally distributed. All of these assumptions are potentially valid but debatable. As an ordinal variable, *MAL* lends itself to the more health-state based approaches and the analysis here applies the health-state based approaches to quality-adjusted survival analysis and multistate modelling to this measure of quality of life. The joint modelling presented in this thesis could be extended to accommodate ordinal variables such as *MAL*. Due to the small numbers of patients in the MIC study, the frequencies for the higher levels of malaise were small and the measure was treated as a binary variable and as such, the detail about the level of malaise that the patient experienced was not included. In terms of quality-adjusted survival analysis, the key problem was that utility values were not available for the health states defined by malaise.

The analysis of quality of life data as a separate endpoint suggested that quality of life improved slightly over time on chemotherapy and deteriorated slightly with standard palliative care. The summary measures analysis, using 0- to 6-week change in *GQS* and standardised area under the curve, and random effects modelling showed that this difference between treatments was of borderline statistical significance at the 5% level. This treatment comparison however is potentially biased due to the differential rates of informative dropout. Quality-adjusted survival analysis showed that the quality-adjusted survival time gained during the first 15 weeks from trial entry was greater on the chemotherapy arm than the standard palliative arm and these differences were statistically significant at the 5% level. Survival-adjusted and dropout-adjusted quality-of-life showed a trend for chemotherapy to give higher levels of quality of life than standard palliative care but this was not statistically significant. Multistate modelling

suggested that compared to standard palliative care, chemotherapy increased the hazard of transition from both a well to an ill state and also an ill to a well state, with the hazard of transition from ill to well being statistically significant at the 10% level. Chemotherapy was also shown to significantly reduce the hazard of death from the ill state compared to standard palliative care. In the joint models for quality of life and either survival or dropout-free survival, the rates of change of quality of life over time were reduced, but the difference between treatments that gave an advantage for chemotherapy remained stable with 95% credible interval for the treatment by slope interaction term just containing zero. In conclusion, simultaneous analysis of quality of life and either survival or dropout-free survival continued to show a trend for a benefit of chemotherapy for quality of life.

11.2.4 Overall Conclusions from the ESPAC Study

The design of the ESPAC study was such that quality of life data were collected routinely from trial entry to death. As the time from trial entry increases however, the quality of life data become more sparse, partly due to the overall reduction in patients due to death but also because compliance is reduced for those still alive. The analysis was therefore restricted to 24 months from trial entry, which will still provide useful inferences as most patients will not survive longer than this period. As with the MIC study, the hazard of death or dropout during this time was greater for the control arm than for chemotherapy and this differential dropout from the quality of life study needs to be accounted for in any analysis. Simultaneous analysis of quality of life with either survival or dropout-free survival was required to provide an unbiased treatment comparison.

The analysis of the ESPAC study considered a single global measure of quality of life that is widely used in cancer clinical trials, the global health status score (*GHSS*) from the EORTC QLQ-C30 questionnaire. The measure is derived from two items with 7-level ordinal responses and this produces a measure with 13 possible distinct values whose distribution is non-normal with scores of discordant pairs having low frequencies compared to scores of matching responses. Also in this study the individual patient profiles were very erratic with no clear trend at the individual level and this makes *GHSS* a difficult outcome to model. The distribution is likely to be a generally true of

this response in any study but changes over time may be clearer in other studies making modelling more feasible. Although joint modelling was feasible, the assumption of normality is likely to be invalid and methods of analysis that do not make any parametric assumptions about *GHSS* may be preferable. Multistate modelling may mask the finer detail in the data by categorising patients into a limited number of health states. Quality-adjusted survival analysis using the integrated quality-survival product may be the best approach as it makes full use of the data and does not require any distributional assumptions, although the method assumes *GHSS* is a utility measure.

The analysis of quality of life data as a separate endpoint suggested that quality of life improved initially up until 6 months from trial entry and then deteriorated slightly over time but there was no evidence of a difference between treatment arms in terms of this trend. This treatment comparison however is potentially biased due to the differential rates of informative dropout. Quality-adjusted survival analysis showed a trend for a greater number of quality-adjusted life years on chemotherapy compared to control and also better survival-adjusted and dropout-adjusted quality-of-life but these differences were relatively small. Multistate modelling showed virtually no difference between treatments in terms of hazards of transitions between the alive states defined by quality of life but there was consistent evidence of a reduction in hazard for chemotherapy compared to control in terms of the transition from ill to dead. Joint modelling gave an increased effect of treatment on the deterioration in quality of life from 6 months onwards with deterioration greater for the control arm compared to chemotherapy but the 95% credible interval for the estimated treatment effect contained zero. In conclusion, although the trend in the results is generally in favour of chemotherapy, compared to control, in terms of the effect on quality of life, the differences are small with considerable associated uncertainty.

11.3 Further Research

11.3.1 Extension of Methods to Multiple Quality of Life Measures

One of the key features about quality of life data, especially that collected in clinical trials is that they are multivariate in nature. The questionnaires used to assess quality of

life generally yield measures on a number of different dimensions of quality of life, in some cases a large number. In this thesis, the analysis has assumed a single measure of quality of life. This may be valid for some situations such as those where only one measure of quality of life is available for analysis or those which pre-specify a single measure as the focus of analysis, with the remaining measures analysed descriptively as secondary endpoints. In other situations, where a small number of quality of life outcomes require analysis, it may be feasible to analyse and report conclusions on each outcome separately. Problems of multiple testing will occur and no acknowledgement is made in such an analysis of the possible correlation between the different measures. The extension, therefore, of the methods presented in this thesis to accommodate multiple measures of quality of life are required.

The health-state based approaches can incorporate multiple measures by basing the health state definitions on the outcomes to multiple measures. This approach is relatively crude and does not properly account for the correlation between measures or the measurement error. Further research is required to develop a more sophisticated approach, particularly for multistate models. It may be possible to extend the methods of IQSP and joint modelling to accommodate multiple measures by using multilevel models for the quality of life function. This approach has previously been shown to be useful for modelling multiple measures of quality of life over time (Beacon 1996, Beacon and Thompson 1996) but inclusion of such models in IQSP and joint models for quality of life and survival requires further research.

11.3.2 Quality of Life Studies in Subgroups of Patients

Many clinical trials include quality of life as a secondary endpoint for treatment comparison and due to the labour intensive nature of data collection and limited resources, quality of life studies are often only carried out on a subgroup of trial patients. It is essential that the quality of life subgroup is representative of trial patients as a whole as this may invalidate any inferences and this should always be checked prior to any analysis. With the reduced number of patients, there may be problems with reduced statistical power to detect differences between treatments in terms of quality of life. Simultaneous analysis of quality of life and survival presented in this thesis uses the survival data for the quality of life patients only. Survival data however are available

for all trial patients and it may be preferable to include the survival data on all patients in order to maximise the amount of the data included in the simultaneous analysis and providing an improved estimation of the survival. In terms of IQSP, the extension of the method to combine the survivor function for all trial patients with the quality of life function for just the subgroup seems feasible but requires further investigation. The potential to extend the joint modelling within a Bayesian framework to incorporate survival data for all trial patients has already been demonstrated for a cost-effectiveness analysis, in which cost data are collected on just a subgroup of trial patients (Lambert et al 2003) and this approach should naturally extend to the situation with quality of life.

11.3.3 Integration of Quality of Life and Survival Data with Cost Data

In assessing which treatments should be routinely used in clinical practice, health-care decision makers are required to balance the clinical effectiveness against the costs in a cost-effectiveness analysis. Effectiveness can be measured in terms of survival benefits but often quality of life is also accounted for as part of the effectiveness measure by comparing treatments in terms of cost per quality-adjusted life year.

It is becoming more commonplace for clinical trials to collect resource use data on patients as part of the data collection. These resource use data can be combined with unit costs to give a total cost per patient. This cost data can be analysed as an outcome in its own right (Barber and Thompson 2000) or can be integrated into an analysis with the survival data collected in the trial in a cost-effectiveness analysis (Briggs and Gray 1998) and further integration with quality of life data for a cost-utility analysis. Methods for the simultaneous analysis of quality of life and survival data need to be extended to also incorporate the patient-level cost data.

A study of costs was carried out for patients in the MIC trial. Data was retrospectively collected on a subgroup of the trial patients to determine the health care resources used by patients from trial entry to death. Unit costs were allocated to these data to give a cost endpoint for each patient. Chemotherapy was shown to be more costly with a mean increased cost of approximately £3000, which for 2 months additional survival time may be considered by some to be cost effective (Billingham et al 2002). Missing data was a problem with the cost outcome and methods of imputation have been investigated

to overcome the problem (Burton et al 2003). In addition, using a Bayesian approach to simultaneously model the cost and survival (Lambert et al 2003b) has not only provided a means for overcoming the problem of missing data within the subgroup of patients in the cost study and but has also enabled the survival data from all trial patients to be incorporated with the cost data from just the subgroup. The extension of this methodology to incorporate the quality of life data in a cost-utility analysis requires further research. The analysis is complicated by the fact that the quality of life and cost data are collected on different but overlapping subgroups of trial patients and the survival data are collected on all patients. The aim will be to incorporate the Bayesian joint model for quality of life and survival developed in this thesis with the Bayesian cost-effectiveness model to provide inferences about cost-utility for all trial patients, taking into account informative dropout.

11.4 Conclusions

In clinical trials, the collection of quality of life data in conjunction with survival has become widespread. Although practitioners are aware of the problem of informative dropout in such studies, the appropriate method for analysis is still not clear. The results from this research provide statisticians analysing quality of life data with a variety of possible methods for the analysis of such data that should yield unbiased results.

Quality-adjusted survival analysis is the simplest and most accessible approach for the simultaneous analysis of quality of life and survival data. In particular, the more recently proposed integrated quality-survival product formulation provides a means of directly combining longitudinal quality of life and survival data collected on patients. It has the flexibility to be completely non-parametric, partly parametric or fully parametric. It produces clinically intuitive results in the form of either quality-adjusted life years or survival-adjusted quality-of-life and can be used also to estimate dropout-adjusted quality-of-life. This is likely to become a standard approach for the analysis of quality of life data in clinical trials.

Multistate modelling and joint modelling provide more sophisticated methods for analysis, allowing greater insight into the quality of life, survival and dropout processes

and facilitating the inclusion of covariates. The Bayesian joint modelling approach in particular offers an attractive future direction for simultaneous analysis of quality of life and survival giving greater scope for development to accommodate multiple measures of quality of life in a single analysis and extension to cost-utility analysis. It may also provide a means of analysing quality of life data from a subgroup of trial patients with the survival data from all patients. Such models can also be generalised to other situations where time-to-event and longitudinal data, such as tumour markers, need to be analysed simultaneously. Further research in this emerging area to address some of the salient points identified in this thesis should be pursued.

APPENDIX I

RELEVANT PUBLISHED PAPERS

Cullen MH, **Billingham LJ**, Woodroffe CM, Chetiyawardana AD, Gower NH, Joshi R, Ferry DR, Rudd RM, Spiro SG, Cook JE, Trask C, Bessell E Connolly CK, Tobias J, Souhami RL (1999) Mitomycin, ifosfamide and cisplatin in unresectable non-small cell lung cancer: effects on survival and quality of life. Journal of Clinical Oncology 1999; 17: 3188-3194.

Billingham LJ and Abrams KR (2002) Simultaneous analysis of quality of life and survival data. Statistical Methods in Medical Research; 11: 25-48.

Billingham LJ and Cullen MH (2003) Potential hazards in the analysis of symptom improvement. Lung Cancer; 40: 201-202.

Mitomycin, Ifosfamide, and Cisplatin in Unresectable Non–Small-Cell Lung Cancer: Effects on Survival and Quality of Life

By M.H. Cullen, L.J. Billingham, C.M. Woodroffe, A.D. Chetiyawardana, N.H. Gower, R. Joshi, D.R. Ferry, R.M. Rudd, S.G. Spiro, J.E. Cook, C. Trask, E. Bessell, C.K. Connolly, J. Tobias, and R.L. Souhami

Purpose: **Chemotherapy for non–small-cell lung cancer (NSCLC) remains controversial. We describe the two largest reported, randomized, parallel trials designed to determine whether the addition of chemotherapy influences duration and quality of life in localized, unresectable (mitomycin, ifosfamide, cisplatin [MIC]1 trial) and extensive (MIC2 trial) disease.**

Patients and Methods: **Ambulatory patients with NSCLC, aged 75 years or younger, with localized disease, were randomized in MIC1 to receive up to four cycles of chemotherapy (CT: mitomycin 6 mg/m², ifosfamide 3 g/m², and cisplatin 50 mg/m²) every 21 days, followed by radical radiotherapy (CT + RT) or radiotherapy (RT) alone. Extensive-stage patients were randomized in MIC2 to identical chemotherapy plus palliative care (CT + PC) or palliative care (PC) alone. Short-term change in quality of life (QOL) was assessed in a subgroup of patients. Data from the two trials were combined to allow multivariate and stratified survival analyses.**

Results: **Seven hundred ninety-seven eligible patients were randomized, 446 in MIC1 and 351 in MIC2. MIC CT improved survival in both trials (significantly in MIC2). The median survival time in MIC1 was 11.7 months (CT + RT) versus 9.7 months (RT alone) ($P = .14$); whereas in MIC2, median survival time was 6.7 months (CT + PC) compared with 4.8 months (PC alone) ($P = .03$). QOL, assessed in 134 patients from start of trial to week 6, showed improvement with chemotherapy and deterioration with standard treatment. In the combined analysis of 797 randomized patients, the positive effect of MIC on survival was significant overall ($P = .01$) and after adjusting for prognostic factors ($P = .01$).**

Conclusion: **MIC chemotherapy prolongs survival in unresectable NSCLC without compromising QOL.**

J Clin Oncol 17:3188-3194. © 1999 by American Society of Clinical Oncology.

DESPITE SOME EVIDENCE supporting a role for chemotherapy (CT) in inoperable non–small-cell lung cancer (NSCLC), routine management practice varies widely across the developed world. In the United States, for example, the use of CT is standard for good-performance-status patients with stage III or IV disease, whereas in the United Kingdom and much of mainland Europe, CT is still not routinely offered, even in stage III disease. Instead, radical radiotherapy (RT) alone is used for the minority of patients with localized disease, and palliative care (PC) alone is used for those with more advanced or metastatic disease. Some trials evaluating CT have shown no survival benefit.[1,2] However, others, particularly trials including cisplatin, have demonstrated an advantage.[3,4] These trials have been small and, hence, lacked the power to effect a widespread change in management practice across the developed world.[5] Furthermore, the side effects of treatment have been regarded by many as outweighing the benefit of a modest extension of life.[6] In 1995, a meta-analysis of trials concluded that cisplatin-based CT does confer a small survival benefit.[7] However, meta-analyses have been criticized on a number of counts[8]; they give no clear guidance on choice of regimen, toxicity, and quality-of-life (QOL) outcomes and are no substitute for large, randomized trials.

In 1988, we reported a phase II study of mitomycin, ifosfamide, and cisplatin (MIC) in NSCLC, with a high objective response rate, good side-effect profile, and improvement in performance status (PS) in responding patients.[9] In two parallel phase III trials (MIC1 and MIC2), we have compared for our study, MIC chemotherapy plus standard treatment (ST) with ST alone in 820 randomized patients with unresectable NSCLC to examine the effects primarily on survival but also, in a subgroup of trial patients, QOL. In the United Kingdom, ST consists of RT for patients with localized unresectable disease (MIC1 trial) and PC for patients with extensive disease (MIC2 trial). The trials had identical design, were conducted at the same time by the

From the Queen Elizabeth Centre for the Treatment of Cancer, University Hospital Birmingham, Birmingham; Cancer Research Campaign Institute for Cancer Studies, University of Birmingham, Birmingham; University College Hospitals Trust, London; Royal Hospitals Trust, London; Walsall Manor Hospital Trust, Walsall; Southend Hospital Trust, Southend; Nottingham City Trust, Nottingham; and Darlington Hospital Trust, Darlington, United Kingdom.

Submitted September 14, 1998; accepted June 2, 1999.

Supported by grants from the Cancer Research Campaign, Asta Medica AG (Frankfurt am Main), Kyowa Hakko United Kingdom Ltd, and the Oliver Prenn Foundation.

Address reprint requests to M.H. Cullen, MD, Queen Elizabeth Centre for the Treatment of Cancer, University Hospital Birmingham, B15 2TH, United Kingdom; email michael.cullen@university-b.wmids.nhs.uk.

0732-183X/99/1710-3188

same collaborative groups, and, apart from stage of disease, had identical eligibility criteria. As well as reporting the trials separately, the data were combined to enable the overall effect of CT in unresectable NSCLC to be assessed in comparison with nonchemotherapeutic treatment.

PATIENTS AND METHODS

Eligibility

Trial patients were required to have previously untreated, unresectable, histologically or cytologically proven NSCLC (squamous, adeno-, or large-cell carcinoma) that could be measured or assessed. Patients were 75 years of age or less, with World Health Organization performance scores of 0, 1, or 2. Patients with symptomatic superior vena caval obstruction, cerebral metastases, previous malignancy, spinal cord compression, or impaired renal function (glomerular filtration rate < 70 mL/min or serum creatinine > 115 mmol/L) were excluded. Patients with no clinical evidence of metastatic disease (except ipsilateral supraclavicular lymphadenopathy) and no pleural effusion and tumors encompassable in a radical RT volume were eligible for the MIC1 trial; the remainder were eligible for the MIC2 trial. Staging was essentially clinical, based on history, physical examination, chest radiograph, blood chemistry profile, and full blood count. More detailed staging investigations were performed when indicated. For instance, patients with symptoms suggesting skeletal metastases had bone scans, and those with biochemical evidence of liver secondary tumors had computed tomography or ultrasound scans.

Treatment

CT (MIC1 and MIC2 trials). The treatment regimen is listed in Table 1. In the early years of the trial, antiemetic therapy consisted of high-dose metoclopramide via intravenous infusion with intravenous dexamethasone. After the introduction of 5-hydroxytryptamine-3 antagonists, the metoclopramide was replaced by a single intravenous dose of granisetron 3 mg or ondansetron 8 mg.

CT treatment cycles were repeated every 21 days, with a maximum of four cycles. CT was discontinued if there was progressive disease at any stage, static disease after two cycles, unrelieved local symptoms, or unacceptable toxicity. During the period of these trials, CT was rarely used in advanced NSCLC in the United Kingdom. Hence, it was regarded as quite ethical and scientifically desirable not to allow its use at any stage in the control arms, and second-line or nonprotocol CT was not permitted.

Table 1. MIC Schedule

Dexamethasone 4 mg IV bolus
Granisetron 3 mg IV bolus
Mitomycin 6 mg/m^2 IV bolus
Ifosfamide 3 g/m^2 plus mesna 1 g/m^2 in 1 L 0.9% saline IV over 3 hours
1 L 0.9% saline plus 20 mmol KCl IV over 3 hours
Mesna 500 mg/m^2 in 50 mL 0.9% saline IV over 15 minutes
Dexamethasone 4 mg IV bolus
Cisplatin 50 mg/m^2 IV in 250 mL 0.9% saline IV over 1 hour
1 L 0.9% saline plus 20 mmol KCl IV over 6 hours
Mesna 500 mg/m^2 in 50 mL 0.9% saline IV over 15 minutes
Dexamethasone 4 mg IV bolus

Abbreviation: IV, intravenous.

RT (MIC1 trial). Dose and fractionation schedules for RT in limited NSCLC vary in different United Kingdom centers. To encourage maximum participation, a single dose/fractionation scheme was not specified, but patients were required to receive a total dose equivalent to not less than 40 Gy in 15 fractions. In the control arm, RT was commenced after randomization and planning. In the CT arm, RT was commenced 3 weeks after completion of four courses of MIC but could be given early if CT failed. Patients who developed distant metastatic disease were withdrawn from further protocol treatment but continued on trial follow-up, receiving palliative therapy, including, where appropriate, RT.

Palliative care, (MIC2 trial). PC, including RT, antibiotics, cough suppressants, analgesics, and so on, was given to all patients without restriction according to the standard practice of the collaborating centers. After treatment, all patients were seen for routine review with chest radiograph at 2-month intervals until the end of the first year after diagnosis and 3-month intervals, thereafter.

Response Criteria

Objective response assessment was based on physical examination and chest radiograph. Routine rescanning was not required because response rate was not a primary outcome. It is possible that response defined in this way may give a higher value than when the definition requires rescanning, as in phase II studies. The decision to continue with a second and subsequent cycles of MIC was further supported by subjective assessment of symptomatic improvement and toxicity.

Final response to CT was assessed on completion of all CT treatment, however many courses patients had received. Assessable and measurable lesions were considered for treatment response.[10] Response was defined as follows: complete response, complete clinical and radiologic disappearance of measurable or assessable lesions; partial response, greater than 50% reduction in size of measurable lesions or regression of assessable lesions.

Trial Design

Both trials were multicenter, prospective, randomized studies. The Birmingham and London Lung Cancer Groups collaborated in the MIC1 trial. The MIC2 trial involved only the Birmingham Group. The designs of the two trials were deliberately identical and ran concurrently to enable a prospective pooling of the data. Randomization was by telephone call to one of the two randomization centers, the CRC Trials Unit (Birmingham, United Kingdom) or the Clinical Trials Office of the London Lung Cancer Group. In MIC1, randomization was stratified by radiotherapist to ensure that variation in clinical practice was distributed evenly between the two arms. The primary end point for the studies was survival, with toxicity, response to treatment, and QOL as secondary end points. Patient sample sizes of 500 (MIC1) and 300 (MIC2) were planned on the basis of detecting an improvement from 5% to 10% in the 5-year survival rate for MIC1 and from 10% to 20% in the 1-year survival rate for MIC2, with 80% power and 5% significance level.

QOL Study Design

QOL data were collected in an unselected subgroup of trial patients. These were from three centers where a QOL research nurse was available to minimize noncompliance and maximize completeness of data. The QOL research nurses were based in clinics where CT was administered, which facilitated access to patients in the CT arms. Consequently, there were more CT patients in the QOL study than controls. The questionnaire, based on the lung cancer module of the European Organization for Research and Treatment of Cancer (EORTC)

QOL questionnaire (EORTC QLQ-LC13),[11] had 12 questions assessing symptoms and toxicity, cough, breathlessness (at three different exercise levels), hemoptysis, pain, appetite, anxiety, depression, dysphagia, nausea, and malaise. Responses were given on a category-rated scale, as none, a little, quite a bit, and very much, and were scored 0, 1, 2, and 3, respectively. Patients completed questionnaires at approximately 3-week intervals from the first pretreatment questionnaire to a maximum of five on the CT arm and four on ST. A minimum of 3 weeks was allowed after radical RT in MIC1 for the 6-week assessment.

Statistics

Methodology for trial data. Response rates were compared using a χ^2 test and were calculated for assessable patients and those unassessable because of disease or treatment-related factors (ie, died before CT started). Survival was measured from the date of randomization to either the date of death (all causes), censor date for the analysis (January 1, 1997), or, if not observed up to that time, date last seen alive. The Kaplan-Meier[12] method was used to estimate survival curves, and differences between treatments were assessed using the log-rank test. In the survival analysis of the combined data, treatment comparisons were adjusted for the separate effects (stratified log-rank tests) and combined effects (Cox proportional hazards regression) of various patient characteristics, ie, sex, age, PS, histologic diagnosis, and trial (representing stage of disease).[13]

Methodology for QOL study. A mean QOL score was calculated for each individual from responses to the questionnaire. The first and third questionnaires were used to calculate the 0- to 6-week change in QOL score. Treatment groups were compared using *t* tests together with multiple regression analysis to adjust for imbalances in distribution of sex, PS, age, and histologic diagnosis.

RESULTS

Patient Characteristics

Between March 1988 and March 1996, 820 patients were randomized, 461 in MIC1 and 359 in MIC2. Fifteen patients in MIC1 and eight in MIC2 were ineligible. These patients were excluded, and sensitivity analysis confirmed that the conclusions were unaffected. MIC1 protocol violations included two RT patients given CT, seven patients who received nonprotocol RT (all nine excluded from response to RT), and 10, mainly RT patients, who had CT on relapse (included in all analyses). In MIC2, two patients had CT on progression (included in all analyses). The treatment groups in both trials were well balanced with respect to patient characteristics (Table 2).

CT Treatment Details and Response

Sixty-two percent of eligible patients (138 of 222; one protocol violation) in MIC1 and 50% of eligible patients (88 of 175) in MIC2 had four courses of MIC. CT was discontinued early in 35 patients in MIC1 and in 34 in MIC2 as a result of nonresponse. Toxicity contributed to early cessation of CT in 28 and 24 patients in MIC1 and MIC2, respectively. In addition, four patients in MIC1 and three in MIC2 requested early discontinuation of CT. No treatment-related deaths occurred in MIC2, but there were three CT-related deaths (all with neutropenic sepsis, one complicated by renal failure) and three deaths from the combined effects of CT + RT in MIC1 (pneumonitis in three patients, complicated by infection in one).

Of the potentially delayable (ie, second, third, and fourth) CT cycles, only 13% (65 of 491) and 9% (27 of 315) in MIC1 and MIC2, respectively, were postponed by more than 2 days. Delays were mainly because of hematologic toxicity or infection. Early cessation and treatment delays (as indicators of tolerance to CT) were observed no more frequently in patients with a PS of 2 than in patients with better PS scores.

Table 2. Patient Characteristics on Entry to Study

Characteristic	MIC1: CT + RT (n = 223)		MIC1: RT (n = 223)		MIC2: CT + PC (n = 175)		MIC2: PC (n = 176)	
	No. of Patients	%	No. of Patients	%	No. of Patients	%	No. of Patients	%
Sex								
Male	170	76	175	78	132	75	122	69
Female	53	24	48	22	43	25	54	31
Histologic diagnosis								
Squamous	151	68	158	71	94	54	103	59
Adenocarcinoma	34	15	25	11	51	29	42	24
Large-cell undifferentiated	34	15	34	15	19	11	6	3
NSCLC unspecified	4	2	6	3	11	6	25	14
PS								
0	90	40	93	42	41	23	28	16
1	97	44	101	45	65	37	85	48
2	34	15	25	11	55	31	45	26
Unknown	2	1	4	2	14	9	18	10
Age, years								
Median	64		64		62		64	
Interquartile range	57-68		58-68		56-68		58-69	
Full range	37-75		35-75		41-75		47-75	

Objective response to CT was assessed in 88% of patients in MIC1 and 89% in MIC2. The remainder of patients either had missing or inassessable x-rays, refused CT, violated protocol, or died from nondisease/nontreatment-related cause before assessment. Of the assessable patients, the objective response rates (complete response plus partial response) were 54% (105 of 196; 95% confidence interval [CI], 47% to 61%) in MIC1 and 32% (49 of 155; 95% CI, 24% to 39%) in MIC2, with 10% and 2% complete responses in each trial, respectively.

RT Treatment Details and Response

In the MIC1 trial, 68% of patients (152 of 223) on the CT + RT arm received their planned radical dose of RT compared with 85% (189 of 223) on the RT arm. Patients did not receive radical RT because they were too sick, developed progressive disease, or died. Although the proportions of patients having radical RT were different on the two arms, the received doses and schedules were very similar (Table 3).

Eighty-three percent of CT + RT patients and 81% of RT patients in MIC1 were assessable for response after all treatment. The objective response rate to CT + RT was 53% (98 of 185; 95% CI, 46% to 60%) on the combined-modality arm and 45% (83 of 185; 95% CI, 38% to 52%) on the RT alone arm ($\chi^2 = 2.43$, $P = .12$), with 20% and 9% complete responses in each arm, respectively.

In the MIC2 trial, 40% (70 of 175) of patients on the CT + PC arm compared with 68% (120 of 176) on the PC arm received RT, which was thoracic in 51 and 99 patients, respectively. The median thoracic RT dose of 30 Gy and interquartile range of 20 Gy to 35 Gy were identical for the two arms of the trial.

Survival

On January 1, 1997, eight patients were lost to follow-up. Of the remainder, 33 patients in MIC1 were still alive (median follow-up, 31 months; range, 12 to 102 months) and six in MIC2 remained alive (median follow-up, 26 months; range, 17 to 45 months).

In both trials, survival was longer on the CT arm compared with ST (Fig 1; Table 4). The median survival in MIC1 was 11.7 months (95% CI, 9.5 to 14.0) on the CT + RT arm compared with 9.7 months (95% CI, 8.0 to 11.4) on RT alone, while in MIC2 median survival was 6.7 (95% CI, 5.4 to 8.0) months (CT + PC) compared with 4.8 months (95% CI, 4.0 to 5.7) on PC alone. In MIC1, the differences in survival did not reach conventional levels of statistical significance ($\chi^2 = 2.20$, $P = .14$), but they were statistically significant in MIC2 ($\chi^2 = 4.87$, $P = .03$).

Table 3. Thoracic Radical RT Received in MIC1 Trial

	CT + RT	RT
Dose, Gy		
Median	50	50
Range	40-60	40-64
Fractions, n		
Median	15	15
Range	10-30	12-32
Duration, days		
Median	28	26
Range	19-52	18-56

Combined Analysis of Survival

The two concurrently run trials, in a stage continuum of 797 patients with inoperable NSCLC, had identical designs that involved randomization between ST alone versus MIC plus ST. Furthermore, because the observed treatment effect is very similar in the two trials, combining the data is valid. When combined, survival was statistically superior on the CT arm compared with ST ($\chi^2 = 6.66$, $P = .01$). Trial (ie, stage: localized or extensive), PS, and histology were significant prognostic factors for survival; sex and age were not (Table 5). Cox regression analysis showed that trial, PS, and histology were independent prognostic factors for survival, and after adjusting for these, the effect of MIC was still significant ($P = .01$), with a 21% (95% CI, 4% to 41%) increased hazard of death on ST (Table 6).

QOL

The QOL subgroup comprised 67 patients from MIC1 (42 from CT + RT arm and 25 from RT arm) and 109 patients from MIC2 (67 from CT + PC arm and 42 from PC arm). The number of patients in the MIC1 and MIC2 trials responding to the third questionnaire at 6 weeks from baseline was reduced to 50 and 84, respectively, with similar dropout rates on both arms of the trials (29% compared to 20% for CT + RT and RT arms, respectively, in MIC1, and 22% compared to 24% for CT + PC and PC arms, respectively, in MIC2). In MIC1, the mean 0- to 6-week change in QOL score was −0.22 (95% CI, −0.36 to −0.08) on the CT arm compared with 0.28 (95% CI, 0.03 to 0.53) on ST, whereas in MIC2 the corresponding QOL figures were −0.09 for CT (95% CI, −0.21 to 0.03) and 0.20 for ST (95% CI, 0.01 to 0.4). Negative values imply that the level of symptom scores reduced, on average, over the 0- to 6-week time period, thus indicating an improvement in QOL. Positive values indicate deterioration. The results for both trials show that, during the 6 weeks from starting treatment, QOL improved on CT (significantly in MIC1) but deteriorated on ST (significantly in both trials). The difference between treatment arms was statistically significant ($P = .0002$ for MIC1 and $P = .007$ for MIC2) and, after adjusting for imbalances in patient characteristics between

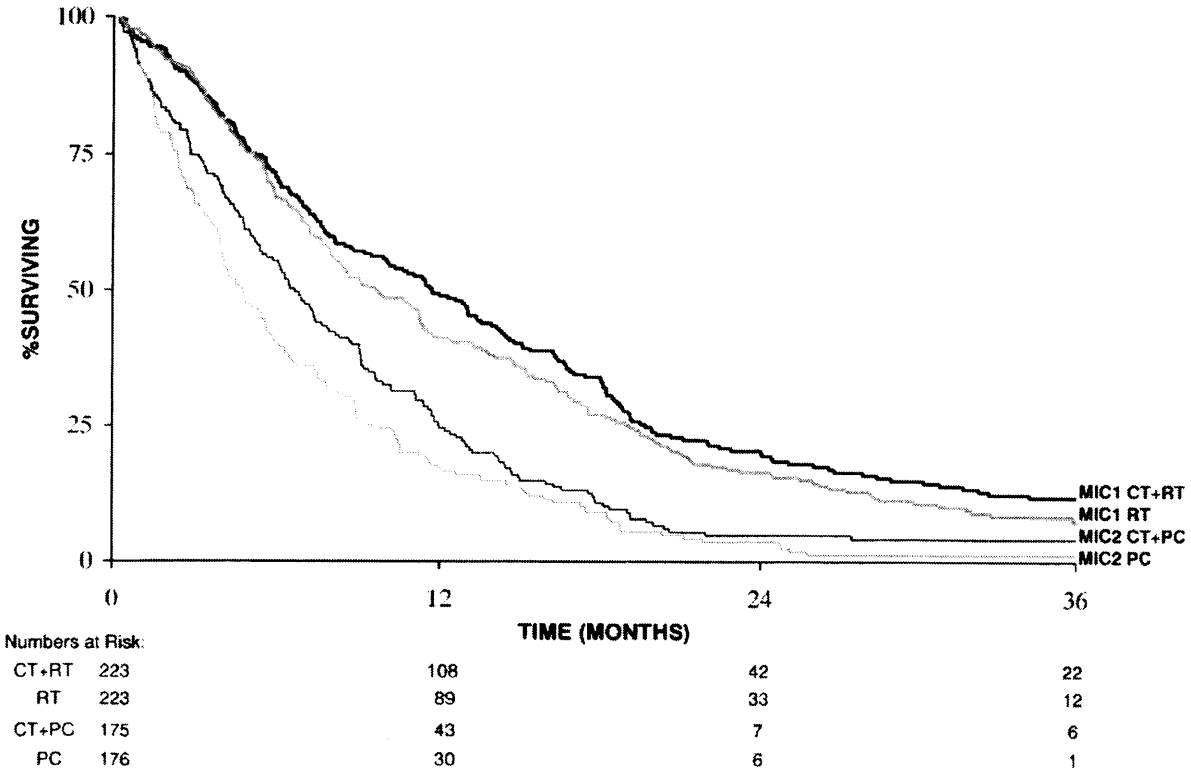


Fig 1. Survival by treatment in each trial.

the treatment groups, remained highly significant for MIC1 and became borderline significant for MIC2 ($P = .0003$ for MIC1 and $P = .06$ for MIC2).

DISCUSSION

Real interest in CT for NSCLC began in the early 1980s with the demonstration that several cisplatin-based regimens could induce objective responses in up to 50% of cases.[14] The next step was to quantify the clinical value, if any, of such intervention. Attempts so far have focused largely on duration of life in randomized trials, which frequently have

Table 4. Survival in MIC1 and MIC2 Trials

Trial Arm	Survival Time				
	Median (months)	1 Year (%)	95% CI	2 Years (%)	3 Years (%)
MIC1 CT + RT	11.7	49	43-56	20	12
MIC1 RT	9.7	41	35-48	16	8
MIC2 CT + PC	6.7	25	18-32	5	4
MIC2 PC	4.8	17	12-23	4	1

Table 5. Survival by Potential Prognostic Factors in Combined Group (n = 797)

Characteristic	No. of Patients	Observed/ Expected Deaths	Median Survival Time (months)	95% CI	χ^2	P
Sex						
Male	599	1.01	7.7	6.8-8.8	0.10	.75
Female	198	0.98	8.1	7.2-9.5		
Age						
< 65 years	446	0.95	8.2	7.2-9.4	2.18	.14
65+ years	351	1.06	7.4	6.3-8.5		
PS						
0	252	0.81	11.2	9.4-12.3	33.0	.0001
1	348	1.01	7.8	6.8-9.0		
2	159	1.46	5.1	4.0-6.8		
Histologic diagnosis						
Squamous	506	0.94	9.0	7.7-10.4	9.20	.01
Adenocarcinoma	152	1.25	5.8	4.6-7.7		
Large-cell	93	1.04	7.9	6.7-9.9		
Trial						
MIC1	446	0.79	11.0	9.0-11.9	76.1	.0001
MIC2	351	1.46	5.6	4.8-6.5		

Table 6. Results From Cox Regression Analysis*

	Hazard Ratio	95% CI	P
Trial, MIC1 v MIC2	1.74	1.48-2.04	.0001
Performance status, 0 v 1 v 2	1.20	1.08-1.34	.0008
Histology, squamous v nonsquamous	1.25	1.07-1.47	.006
Treatment, MIC v ST	1.21	1.04-1.41	.01

*n = 731 with complete data.

been too small to detect the modest improvements likely with current regimens. In our two large trials, MIC has been simultaneously evaluated in terms of both quantity and quality of life. Mitomycin, ifosfamide, and cisplatin are three of the most active single agents in NSCLC, and phase II data for the MIC combination were first reported in the late 1980s.[9,15] Since then, MIC has been tested and its activity confirmed in various contexts in randomized trials in stage IIIA,[16] IIIB, and IV disease.[17] The objective response rates of 54% in localized, unresectable disease and 32% in extensive disease imply that MIC is among the most effective regimens in NSCLC and is at least as active as more recently described combinations like paclitaxel/carboplatin,[18] gemcitabine/cisplatin,[19] and vinorelbine/cisplatin.[20] The objective response rate in the control arm of the MIC1 trial is similar to that reported by others for radical RT in localized, inoperable NSCLC.[21,22]

In localized disease (MIC1), we found no statistically significant survival advantage with the addition of MIC, although there was a trend in favor of CT. This result is almost identical to that reported by Le Chevalier et al[23] in the only other trial of similar design with more than 300 cases. Survival at 1 year was 41% for RT alone in both trials compared with 50% and 49% in the French and MIC trials, respectively, in the CT + RT arms. At 2 years, the corresponding figures for RT alone were 14% (French trial) and 16% (MIC) compared with 21% and 20%, respectively, for CT + RT in the two trials. A later analysis of the French trial reported a statistically significant benefit for CT + RT versus RT alone.[24] Preliminary results from an intergroup trial in the United States, which included only good-risk patients, also show a benefit from the addition of cisplatin-based chemotherapy.[25]

In advanced disease (MIC2), the picture is clearer, with a significant prolongation of life in the CT arm. Other smaller trials with cisplatin-based CT have come to a similar conclusion,[4,26] but there has not been widespread adoption of CT for the management of stage IV NSCLC. Raby et al[6] reported that, although a majority of Canadian clinicians involved in lung cancer therapy believed CT prolonged median survival in stage IV NSCLC, only 20% would recommend it for an asymptomatic patient. The authors believed that, although randomized trials may demonstrate that a treatment works, they often fail to show that it is worthwhile.

Toxicity is frequently cited as a reason to withhold CT. The QOL component of the present trials was incorporated to quantify symptom relief and toxicity formally. A short-term assessment was chosen to minimize the effect of subject dropout, which complicates QOL analyses in situations where patients deteriorate rapidly and allows evaluation of rapid symptom relief, which is vital in patients with short life expectancy. Drop out rates were similar in both arms of each trial, validating treatment comparisons. These comparisons showed an improvement in QOL in patients undergoing CT, implying palliation outweighed short-term toxicity as well as a deterioration in patients on ST. A more detailed analysis examining other time points, with similar overall conclusions, will be reported separately.

Others have reported symptom improvement resulting from CT treatment in NSCLC, even in the absence of objective tumor response.[27] In an interesting, recent American study, NSCLC patients who had experienced cisplatin-based CT were asked to indicate the minimum survival benefit required to accept the side effects of CT for advanced disease. For a realistic survival benefit of 3 months, only 22% chose CT against supportive care. However, 68% chose CT if it substantially reduced symptoms without prolonging life.[28]

The identical basic design and similar treatment effects of these trials permitted amalgamation of the survival data, allowing a comparison of MIC with ST in almost 800 randomized patients with unresectable NSCLC. In common with other trials, we found PS, stage, and histologic diagnosis to be independent prognostic factors for survival,[29] and, having adjusted for the effects of these, the positive impact of MIC on survival remained significant. An examination of CT effect across strata defined by a number of pretreatment patient characteristics is being prepared for separate publication.

There is now a considerable body of evidence that supports a small beneficial effect of cisplatin-based CT on survival in advanced NSCLC. The results presented here, from the two largest reported trials of one of the most active regimens,[30] are fully consistent with the meta-analysis of smaller trials.[7] The effect of MIC on survival, seen in each trial separately, is reinforced by the consistently significant treatment effect observed in the combined data. We have also shown that the treatment effect is not achieved at the expense of short-term QOL. Thus, MIC is an important comparator candidate for future trials aiming to identify regimens with greater impact on duration of life, without compromising quality.

ACKNOWLEDGMENT

We thank Janet Dunn, Janet Woods, Angela Murphy, Lindsay James, Ruth Allen, Jane Drake, Lillian Daniels, Helen Bartram, and Sue Alcock.

APPENDIX

This work resulted from a collaboration between the Birmingham and London Lung Cancer Study Groups. The principle contributors other than the listed authors were: C. Skinner, J. Ayres, D. Stableforth, F. Collins, S. Burge (from Heartlands Hospital Trust, Birmingham, United Kingdom); A. Goodman, G. Blackledge, N. James, D. Spooner, H. Earl (from Queen Elizabeth Centre for the Treatment of Cancer, University Hospital Trust, Birmingham, United Kingdom); B. Mantell, N.C. Barnes, N. Plowman (Royal Hospitals Trust, London, United Kingdom); W. Pratt, P. Murray (Essex County Hospital Trust, Colchester, United Kingdom); A.C. Robinson, A. Lamont, C.D. Eraut (Southend Hospital Trust, Southend); F. Macbeth (Beatson Oncology Centre, Glasgow, United Kingdom); C. Macmillan (Northampton General Hospital Trust, Northampton); M. Henk (Royal Marsden Hospital Trust, London, United Kingdom).

REFERENCES

1. Mattson K, Holsti LR, Holsti P, et al: Inoperable non-small cell lung cancer: Radiation with or without chemotherapy. Eur J Cancer 24:477-482, 1988
2. Morton RF, James R, Jett JR, et al: Thoracic radiation therapy alone compared with combined chemoradiotherapy for locally unresectable non-small cell lung cancer. Ann Intern Med 115:681-686, 1991
3. Dillman RO, Herndon J, Seagren SL, et al: Improved survival in stage III non-small cell lung cancer: Seven year follow-up of Cancer and Leukemia Group B (CALGB) 8433 trial. J Natl Cancer Inst 88:1210-1215, 1996
4. Rapp E, Pater JL, Willan A, et al: Chemotherapy can prolong survival in patients with advanced non-small-cell lung cancer: Report of a Canadian multicentre randomized trial. J Clin Oncol 6:633-641, 1988
5. Johnson DH: Chemotherapy for metastatic non-small cell lung cancer: Can that dog hunt? J Natl Cancer Inst 85:766-767, 1993
6. Raby B, Pater J, Mackillop WJ: Does knowledge guide practice? Another look at the management of non-small-cell lung cancer. J Clin Oncol 13:1904-1911, 1995
7. Non-Small Cell Lung Cancer Collaborative Group: Chemotherapy in non-small cell lung cancer: A meta-analysis using updated data on individual patients from 52 randomised trials. Br Med J 311:899-909, 1995
8. Le Lorier J, Grégoire G, Benhaddad A, et al: Discrepancies between meta-analyses and subsequent large randomized, controlled trials. N Engl J Med 337:536-542, 1997
9. Cullen MH, Joshi R, Chetiyawardana AD, et al: Mitomycin, ifosfamide and cisplatin in non-small cell lung cancer: Treatment good enough to compare. Br J Cancer 58:359-361, 1988
10. Eagan RT, Fleming TR, Schoonover V: Evaluation of response criteria in advanced lung cancer. Cancer 44:1125-1128, 1979
11. Aaronson NK, Ahmedzai S, Bergman B, et al: The European Organization for Research and Treatment of Cancer QLQ-C30: A quality of life instrument for use in international clinical trials in oncology. J Natl Cancer Inst 85:365-376, 1993
12. Kaplan EL, Meier P: Non-parametric estimation from incomplete observations. Am Stat Assoc J 53:457-481, 1958
13. Cox DR: Regression models and life tables (with discussion). J R Stat Soc B 34:187-220, 1972
14. Minna JD, Higgins GA, Glatstein EJ: Cancer of the lung, in DeVita VT, Hellman S, Rosenberg SA (eds): Principles and Practice of Oncology. Philadelphia, PA, JB Lippincott Co, 1982, pp 396-474
15. Giron CG, Ordonez A, Jalon JI, et al: Combination chemotherapy with ifosfamide, mitomycin and cisplatin in advanced non-small cell lung cancer. Cancer Treat Rep 71:851-853, 1987
16. Rosell R, Gomez-Codina J, Camps C, et al: A randomized trial comparing preoperative chemotherapy plus surgery with surgery alone in patients with non-small-cell lung cancer. N Engl J Med 330:153-158, 1994
17. Crino L, Clerici M, Figoli F, et al: Chemotherapy of advanced non-small cell lung cancer: A comparison of three active regimens—A randomized trial of the Italian Oncology Group for Clinical Research (GOIRC). Ann Oncol 6:347-353, 1995
18. Johnson DH, Paul DM, Hande KR, et al: Paclitaxel plus carboplatin in advanced non-small-cell lung cancer: A phase II trial. J Clin Oncol 14:2054-2060, 1996
19. Abratt RP, Bezwoda WR, Goedhals L, et al: Weekly gemcitabine with monthly cisplatin: Effective chemotherapy for advanced non-small-cell lung cancer. J Clin Oncol 15:744-749, 1997
20. Le Chevalier T, Brisgand D, Douillard J-Y, et al: Randomized study of vinorelbine and cisplatin versus vindesine and cisplatin versus vinorelbine alone in advanced non-small-cell lung cancer: Results of a European multicenter trial including 612 patients. J Clin Oncol 12:360-367, 1994
21. Cox JD, Komaki R, Byhardt RW: Is immediate chest radiotherapy obligatory for any or all patients with limited-stage non-small cell carcinoma of the lung? Yes. Cancer Treat Rep 67:327-331, 1983
22. Perez CA, Stanley K, Grunday G, et al: Impact of irradiation technique and tumor extent in tumor control and survival of patients with unresectable non-oatcell carcinoma of the lung: Report by the Radiation Therapy Oncology Group. Cancer 50:1091-1099, 1982
23. Le Chevalier T, Arriagada R, Quoix E, et al: Radiotherapy alone versus combined chemotherapy and radiotherapy in unresectable non-small cell lung carcinoma. J Natl Cancer Inst 83:417-423, 1991
24. Le Chevalier T, Arriagada R, Tarayre: Significant effect of adjuvant therapy on survival in locally advanced non-small cell lung cancer. J Natl Cancer Inst 84:58, 1992 (abstr)
25. Sause WT, Scott C, Taylor S, et al: Radiation Therapy Oncology Group (RTOG) 88-08 and Eastern Cooperative Oncology Group (ECOG) 4588: Preliminary results of a phase III trial in regionally advanced, unresectable non-small cell lung cancer. J Natl Cancer Inst 87:198-205, 1995
26. Cartei G, Cartei F, Cantone A, et al: Cisplatin-cyclophosphamide-mitomycin combination chemotherapy with supportive care versus supportive care alone for treatment of metastatic NSCLC. J Natl Cancer Inst 85:794-800, 1993
27. Hardy JR, Noble T, Smith IE: Symptom relief with moderate dose chemotherapy (mitomycin-C, vinblastine and cisplatin) in advanced non-small cell lung cancer. Br J Cancer 60:764-766, 1989
28. Silvestri G, Pritchard R, Welch HG: Preferences for chemotherapy in patients with advanced non-small cell lung cancer: Descriptive study based on scripted interviews. Br Med J 317:771-775, 1998
29. Coy P, Elwood JM, Coldman AJ: Clinical indicators of prognosis in unresected lung cancer. Chest 80:453-458, 1981
30. Giron CG, Perez YF, Navarro JG, et al: Revision del tratamiento del carcinoma no microcitico de pulmon con la combinacion mitomicina-ifosfamida-cisplatino (MIP). Oncologia 19:377-387, 1996

Statistical Methods in Medical Research 2002; **11**: 25–48

Simultaneous analysis of quality of life and survival data

LJ Billingham Cancer Research Campaign Trials Unit, Institute for Cancer Studies, The Medical School, University of Birmingham, Birmingham, UK and **KR Abrams** Department of Epidemiology and Public Health, University of Leicester, Leicester, UK

In many phase III clinical trials, particularly in the field of cancer, the comparison of treatments is based on both length of survival and quality of life. Subjects are followed over time until death and during this period, quality of life is assessed on a number of occasions. Simultaneous analysis of these two outcomes supplements the comparison of treatments in terms of each outcome independently with an assessment of the net effect. In addition, it provides a means of accounting for the informative dropout due to death of patients within the time frame of the quality of life study. The methods also have the potential to be extended to allow for informative dropout from the quality of life study prior to death. There are a number of broad approaches for the simultaneous analysis of quality of life and survival data. The most widely used approach in clinical research is quality-adjusted survival analysis, where treatments are compared in terms of a composite measure of quality and quantity of life. The paper reviews the different techniques for quality-adjusted survival analysis, illustrating the methodology by application to data from a phase III clinical trial in pancreatic cancer. In addition, alternative approaches using multistate survival analysis and joint modelling methods are also discussed.

1 Introduction

The ideal context for assessing a new treatment is a randomized phase III clinical trial. The new treatment is compared to the 'standard' treatment using a number of pre-defined clinically relevant outcomes. In the field of cancer, the primary outcome measure is usually length of survival, with quality of life often included as a secondary outcome. Patients are followed over time until they are observed to die or until the time of analysis. The patient's quality of life is measured at various points during this period of study.

Quality of life is generally measured using a patient-completed questionnaire. There are a number of standard instruments available for this purpose.[1] The choice of instrument depends on the disease and treatment being studied and the aspects of quality of life that are of particular interest. The instruments generally measure a number of different dimensions of quality of life. Methods of analysis that deal with the issue of multiple dimensions are discussed elsewhere.[2] For the purposes of this paper we assume that there is one, ideally generic, measure of quality of life for analysis. The application of methods outlined in this paper in a multiple dimensions setting is addressed in the discussion.

Address for correspondence: Lucinda Billingham, Cancer Research Campaign Trials Unit, Institute for Cancer Studies, The Medical School, University of Birmingham, Birmingham B15 2TT, UK.

10.1191/0962280202sm269ra

The assessment of treatments in terms of quality of life is usually based on their effect on this outcome over time. Some studies may be interested in the effects during a fixed time frame of interest, such as the treatment period or the first year from entry to trial, for example. In other studies the focus may be on the quality of life during a patient-related time frame, such as from entry to trial until death. In this case, the length of study will vary for each patient but the overall time frame of the longitudinal analysis will be defined by the longest survival time. For both the fixed and overall patient-related time frame, even though patients may have a complete set of quality of life measures until death, a standard longitudinal analysis of the quality of life outcome will treat the non-existent data at all times within the time frame after death as missing and will assume that they are missing at random. Clearly, because they are missing due to death this assumption is invalid and any longitudinal analysis needs to account for this. Patients who die within the fixed or overall patient-related time frame are said to have 'dropped out' of the quality of life study and because the resulting missing data are not missing at random, the dropout is said to be 'informative'.

Within both types of quality of life study time frame, patients can also drop out of the study prior to death for a number of reasons. Dropout could be purely random or may be due to reasons related to disease or treatment. In longitudinal studies where patients are sufficiently ill for length of survival also to be an outcome, subjects will often drop out due to illness. In this case, since the reason for dropping out is directly related to what is being measured, the dropout is informative and the resulting missing values are not missing at random. Application of standard longitudinal data analysis in this context could give biased results. The point at which a patient ceases to be a participant in the quality of life study and instead becomes a dropout will not necessarily be clear-cut and the choice will be subjective.

The analysis of quality of life and survival data simultaneously has a number of advantages. Firstly, it supplements the comparison of treatments in terms of each outcome independently with an assessment of the net effect, which will assist the clinical decision regarding the choice of treatment for patients. This decision may also need to be made in conjunction with other outcomes such as toxicity and cost. Secondly, it allows a more valid assessment of quality of life since the longitudinal analysis of quality of life data can account for data that are missing within the study time frame as a result of informative dropout due to death. Thirdly, the methods have the potential to be extended to allow for informative dropout from the quality of life study prior to death, resulting in a further improvement in the validity of the quality of life analysis.

There are a number of different broad approaches that can be used to analyse quality of life and survival data simultaneously. The most widely used approach in clinical research is quality-adjusted survival analysis. This approach combines the amount of time patients spend in a number of different health states with weights reflecting the quality of life of those health states to create a composite measure of quality and quantity of life. More recently methods have been proposed that directly incorporate the longitudinal quality of life collected on patients with the survival data and this alternative to the health-state based method is far more disposed to the analysis of data collected in clinical trials.

Other methods of simultaneous analysis include multistate survival analysis and joint modelling. Multistate survival analysis models the movement of patients between

various health states, defined by levels of quality of life and death, and explores how treatments differ in terms of the transition rates between health states. Joint modelling considers quality of life and survival as two simultaneous processes occurring in patients and describes the data in terms of two inter-linked models.

The paper reviews in detail the different methods for quality-adjusted survival analysis, illustrating the methodology by application to data from a phase III clinical trial in pancreatic cancer. In addition, the alternative approaches of multistate survival analysis and joint modelling are also discussed.

2 Quality-adjusted survival analysis

Quality-adjusted survival analysis is based on the concept of quality-adjusted life years (QALYs)[3] where quality and quantity of survival are combined into a composite measure. In the general QALY model, survival time is split into periods spent in different health states where different quality of life is experienced. QALYs are calculated, as shown in (1), by summing these times, t_j ($j = 1, \ldots, J$), over the J different time periods with weights, u_j ($j = 1, \ldots, J$), attached to each period reflecting the quality of life experienced during that time.

$$\text{QALY} = \sum_{j=1}^{J} u_j t_j \tag{1}$$

The weights range from 0 to 1, with 0 representing quality of life equivalent to death and 1 representing perfect health. Negative weights can be used if the quality of life is thought to be worse than death. These weights are intended to reflect the relative desirability of the state and are usually referred to as 'health state utilities'.[4] The most widely used QALY models are TWiST (Time Without Symptoms or Toxicity)[5,6] and Q-TWiST (Quality-adjusted TWiST).[7,8] For TWiST, all periods of survival time with symptoms of disease or toxicity resulting from treatment are given a weight of 0 whilst all other time periods are given a weight of 1, so that, as the term suggests, TWiST counts only time without symptoms or toxicity. With the Q-TWiST endpoint, originally developed for a breast cancer application,[7] periods of survival time spent with symptoms of disease and toxicity resulting from treatment are each given weights between 0 and 1, rather than being ignored as they are in TWiST.

The periods of time that patients spend in the health states involved in the Q-TWiST model are generally determined using clinical outcomes collected in trials, rather than quality of life data. In many studies of quality of life, the Q-TWiST model is not relevant because patients never experience the TWiST health state. For example, in studies of palliative care where quality of life is an important outcome in its own right, the patients are never in a disease-free state and are therefore generally never without symptoms. In other situations, quality of life studies will often focus on the experience of patients during treatment, and in this case patients are generally never in a toxicity-free state. The inapplicability of the Q-TWiST model to longitudinal quality of life data collected in clinical trials has been recognized and alternative methods have been

developed that directly incorporate the patients' quality of life data in a quality-adjusted survival analysis.[9] It is these methods that are reviewed here.

There are two main approaches to quality-adjusted survival analysis depending whether the level of aggregation of the quality of life and survival data is at the subject or the group level. The subject-based approach combines quality of life and survival at the patient level, thus creating a single endpoint for each subject on which to compare treatments. The group-based approach aggregates quality of life and survival at a (pre-defined) group level. Both approaches will be discussed and illustrated using data from a previously conducted phase III trial in pancreatic cancer.

2.1 Background to the pancreatic cancer trial

The aim of the trial was to investigate the use of chemotherapy compared to a control in patients with pancreatic cancer in terms of both length of survival and quality of life. For practical reasons, quality of life was assessed in only a subgroup of trial patients. The results from the trial are currently unpublished and hence, for the purposes of this paper, the trial will be anonymous.

Quality of life was measured using the EORTC QLQ-C30, the standard instrument developed by the European Organisation for Research and Treatment of Cancer,[10] which measures 15 different dimensions including a measure of 'global health status'. This single measure of quality of life is selected as the variable for analysis and, although simplistic, it nonetheless illustrates the essential elements of the various methods for quality-adjusted survival analysis and allows insight into the associated problems. The global health status score, which is constructed from two questions each with a seven-level ordinal response (see Figure 1), ranges from 0 to 100 and takes 13 equally spaced possible values in this range.

The quality of life study was designed to assess patients every 3 months from trial entry to death. In practice, there was some variation in the timing of assessments and this paper uses actual assessment times rather than planned ones in the analyses that follow. In addition, many of the patients were not followed up until death and the length of follow-up in terms of quality of life and survival is variable across patients. In an attempt to equalize and at the same time maximize the length of follow-up for quality of life and survival for all patients, the analysis is restricted to the 2 year period from entry to trial. The reason that equal follow-up is important is discussed later. The data used in this paper is from 140 patients who all had a baseline quality of life measure and at least one other quality of life assessment during this time period.

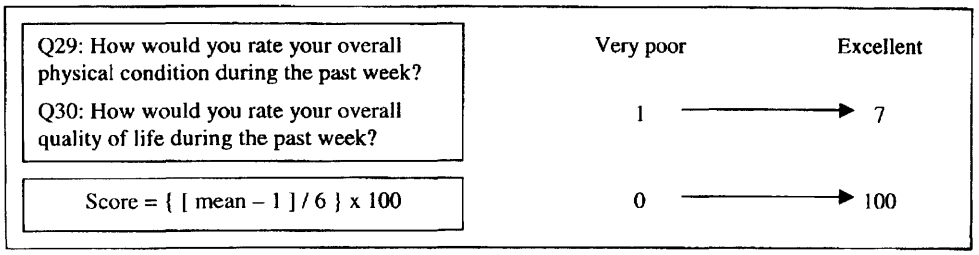


Figure 1 Global health status score from the EORTC QLQ-C30 [where mean=(Q29+Q30)/2].

Figure 2 shows both the longitudinal quality of life in terms of global health status score and the survival data of patients over the first 2 years. Note that the graph of quality of life values over time assumes a linear change between assessments (see below for further discussion of plotting curves of quality of life over time). All except 18 patients had full follow-up in terms of survival for the 2 years from entry, resulting in just 18 patients with a censored survival time during the 2 year study period. Eighty patients died within 2 years of entry to the trial, resulting in a large number of dropouts due to death within the 2 year time frame for the quality of life study, thus making it a realistic example with which to illustrate the simultaneous analysis of quality of life and survival data.

2.2 Subject-based approach

The simplest approach to quality-adjusted survival analysis is to use the quality of life data collected over time to calculate the number of QALYs for each subject.[11,12] The measures of quality of life taken at discrete points over time for an individual can be used to create a 'curve' that describes the quality of life from entry to trial to either death or to a fixed end-of-study time. When the quality of life measure is a utility, the

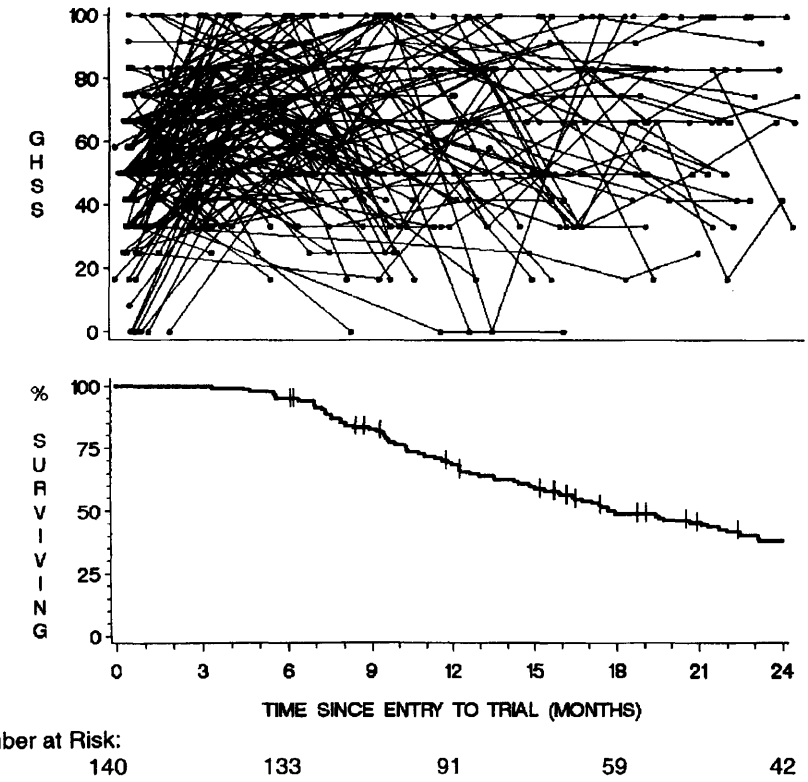
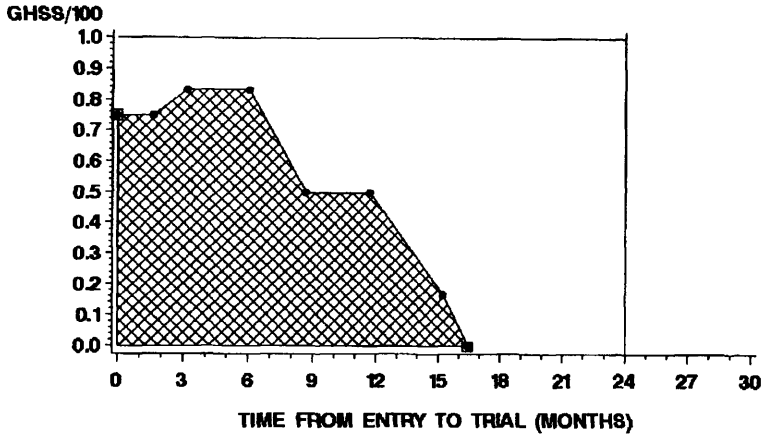


Figure 2 Global health status score (GHSS) interpolated over time and Kaplan–Meier survival curve for all patients in the pancreatic cancer study (censored survival times shown by dashes).

area under this curve represents the QALYs for that subject. If the quality of life measure is not a utility then it may be possible to transform it to a 0–1 scale, to create a pseudo utility, and the resulting area under the curve may be interpreted like a QALY. There are a number of practical issues to consider when describing a patient's quality of life, measured at a number of discrete time points, by a curve over continuous time.

- The measures at discrete time points can be connected in a number of ways. A step function can be used if it is thought that the quality of life remains constant until the next measure. In some cases it may be appropriate to assume that the change happens midway between two time points. Another option is to assume that there is a linear change from one time point to the next (see examples in Figure 3). By connecting measures at discrete time points in this way, all quality of life values between the actual assessment times are effectively being imputed.
- Often the quality of life study is designed to assess patients at fixed time points but there is usually some variation in the timings of the actual assessments around these fixed time points. The quality of life curve can be defined using either the planned fixed timings or the actual timings (latter used in examples in Figure 3).
- If actual assessment times are being used and the first assessment is some time after the trial entry time, then the value of quality of life at entry needs to be imputed and connected to the value at the first assessment. The simplest approach is to carry the first value back to the trial entry time (see examples in Figure 3).
- If the assessments are made at fixed times and there are intermittent missing assessments, a decision needs to be made as to whether it is reasonable to assume that such missing data are missing at random. If so, then the missing data can be ignored and the quality of life values on either side can be connected as normal (as done for example in Figure 3d). If not, then an approach that accounts for informative intermittent missing data needs to be used; however, such methodology is not well developed[13] for applications in this context.
- If a patient dies then their measure on the utility-type scale becomes zero from that point onwards. The curve between the last measure before death and the zero measure can be completed either as a step function or as a linear decline (see example in Figure 3a).
- If a patient's survival time is censored then the quality of life curve will stop at the point when the patient was last known to be alive and the QALY will be censored (see Figure 3b). This can cause problems for the analysis as discussed later.
- If quality of life is being studied for a fixed period of time and a patient is known to survive the full duration of the quality of life study period, then the curve must be completed by connecting the last measure taken within the study period to the fixed end-of-study time. This can be done in a number of ways depending on the nature of the post-study information. In some situations, it may be necessary or desirable to use purely the information within the study period for the analysis but if post-study information is available then it may enable the values imputed by the quality of life curve at the end of the study period to be more accurate in the following way. If a post-study quality of life assessment is available then the last measure before the fixed cut-off can be connected to the first measure after the cut-off using either a step function or linear change (see Figure 3c). Otherwise, if there is no post-study quality

(a) Death within the 2 y study period



(b) Censored survival within the 2 y study period

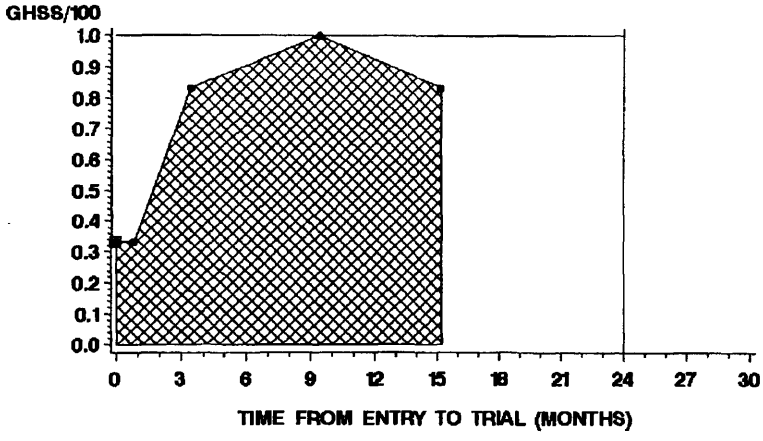
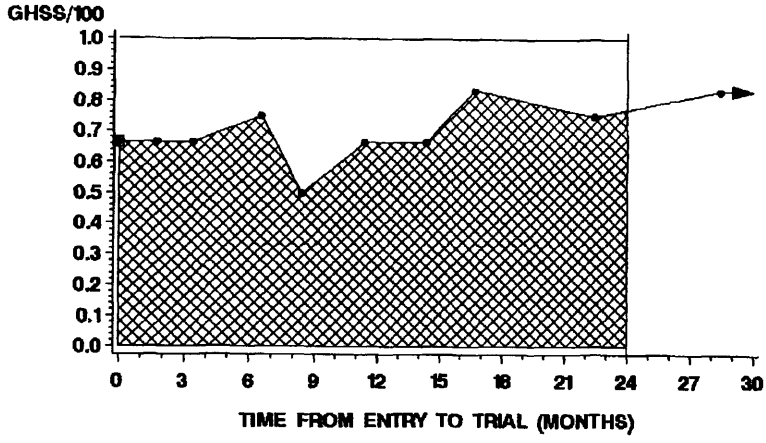


Figure 3 Examples of curves describing the transformed global health status score (GHSS/100) over time within 2 years of entry to trial for patients in the pancreatic cancer study (squares around dots indicate imputed values). (a) Death within the 2 year study period. (b) Censored survival within the 2 year study period. (c) Survival greater than 2 years with post-2 year quality of life assessment. (d) Death after the 2 year study period. (e) Censored survival greater than 2 years.

of life data available but the date of death is known, the last measure can be connected to the value 0 at the time of death in the most appropriate way (see Figure 3d). In both cases the part of the curve within the study period completes the quality of life curve, with the value at the cut-off time calculated using interpolation. If neither of these events occur then the curve between the last measure and the cut-off time can be completed by carrying this value forward (see Figure 3e).

(c) Survival greater than 2 y with post-2 y quality of life assessment



(d) Death after the 2 y study period

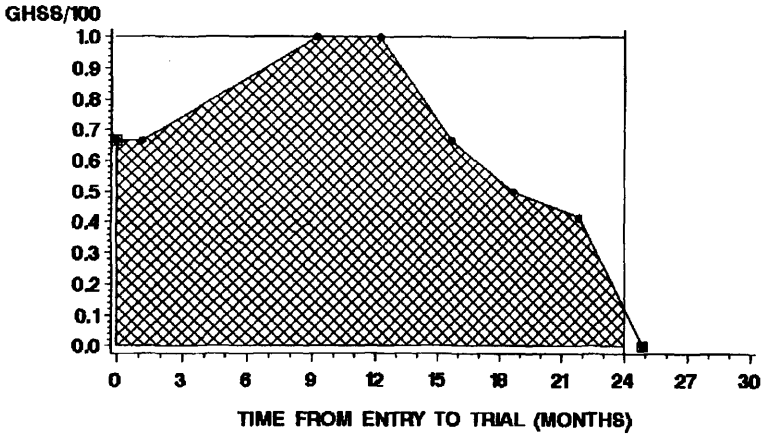


Figure 3 (*continued*)

- If a patient drops out of the quality of life study but has continued follow-up in terms of survival then either the curve could stop at the last recorded assessment giving a censored QALY, which as discussed later can cause problems in the analysis, or the curve could continue to the last survival follow-up point (whether it be dead or alive) by imputing the missing values. There are a number of options available for imputation,[14] the validity of which may be compromised if the time from last quality of life assessment to last survival follow-up is very long. The

(e) Censored survival greater than 2 y

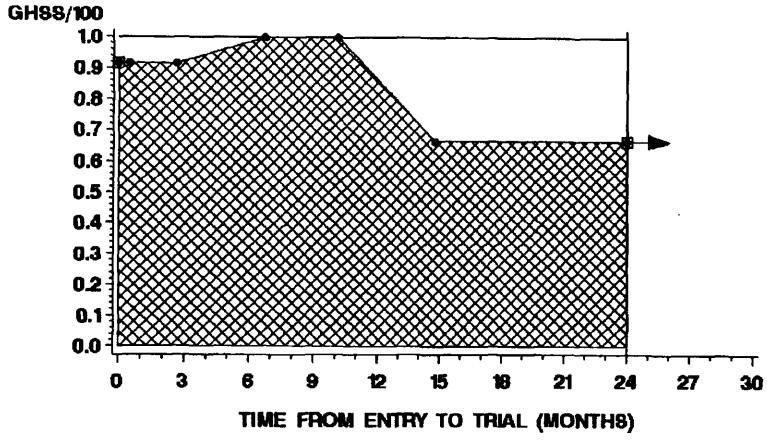


Figure 3 (*continued*)

simplest approach is to carry the last value forward. Alternatively, the worst or best value could be carried forward from some appropriate point after the last assessment. If the approximate health state of the patient during the dropout time can be ascertained from other clinical information such as clinically assessed performance status then, at some appropriate point after the last assessment, the curve could take on a value that reflects the estimated health state of the patient. If the date of death of the patient is known then the curve could linearly decrease to zero over the dropout time. More sophisticated methods such as multiple imputation should be considered for imputing values at the scheduled times after dropout (whenever that is defined to be) for the formation of the quality of life curves.

If all patients in the study are followed up until death and quality of life is assessed for the full duration of this time, then a QALY can be calculated for each patient and treatments can be compared using standard analytical methods. In general, however, not all subjects will be followed up until death and for those patients with censored survival times, their QALY will also be censored. It would seem sensible to analyse such censored QALY data using standard survival analysis techniques, but the use of quality weightings in calculating the QALY endpoint creates an informative censoring and thus renders such methods invalid.[3]

One option is to fix the period of study, say to time L, such that all subjects either die within this time or are known to survive this time and hence, in all cases, survival times within the restricted period will be uncensored. The QALYs achieved within this study time, QALY(L), can be calculated and as long as quality of life values are available for the full duration of each patient's survival time, there will be no censored values of QALY(L) and the comparison of treatments in terms of this measure using standard analytical methods will be valid.[11] If, for some patients, quality of life values are not available for the full duration of the survival time then to avoid censored values of

QALY(L) it will be necessary to impute quality of life values for the periods of survival with no measures of quality of life in the way described in the last bullet point given above.

If there are a number of survival times censored at an early time point then choosing a fixed time period with no censored survival times may not be a practical approach but using survival analysis techniques even with just a small number of censored values could give biased results.[5] An alternative option is therefore to replace the censored QALYs with imputed values and analyse using standard statistical methods for non-censored data. Censored data can be replaced with a range of possible values.[5] The minimum possible QALY would occur if the patient died immediately after the last follow-up. The maximum possible QALY would occur if the patient survived the rest of the study period with perfect quality of life. A mean value is obtained by assuming the patient survived the rest of the study period with quality of life valued at 0.5. This is illustrated and discussed further in the pancreatic cancer example. A more appropriate solution to overcome the problem of censored QALYs is to use a group-based approach as discussed later.

2.3 Subject-based approach illustrated with data from the pancreatic cancer study

In the pancreatic cancer study, the quality of life curves for each subject were created by plotting the global health status score (divided by 100 to put it on a 0–1 scale) against actual quality of life assessment times, assuming a linear change between assessments and ignoring intermittent missing values. A linear decrease in quality of life was assumed from the value at the last assessment to the value of 0 at the time of death (regardless of the time span). For patients with censored survival times, values of quality of life at the last assessment were carried forward until the time last seen alive (regardless of the time span).

Quality of life curves were restricted to the 2 year study period and post-study information in terms of quality of life where available and survival otherwise was used to complete the curves within the analysis period. For all subjects with survival follow-up greater than the study period, the 2 year value is interpolated from the linear change between the two values that straddle the 2 year cut-off. Examples of subjects' quality of life curves illustrating some of these assumptions are shown in Figure 3 and the formation of curves based on these assumptions allows the quality of life at all times during an individual's survival time to be estimated. Any values between the actual assessment times are effectively imputed by the formation of the quality of life curve.

Table 1 shows five different categories of patient in the study according to follow-up in terms of both survival and quality of life. The table shows the amount of time between the last quality of life assessment and the 2 year cut-off for those who survive the 2 year study time, and between the last quality of life assessment and the last survival follow-up time for those who die or are censored within this period. From this the dropouts from the quality of life study can be subjectively defined. Of the 140 patients, 23 had a post-2 year quality of life assessment and therefore, despite the fact that some had a long period of time with no assessments within the 2 year study time, they were not deemed to have dropped out of the quality of life study. Of the remaining 117 patients, 19 were known to have survived the 2 year study time, but only six of

Table 1 Distribution of time from last quality of life (QoL) assessment to either 2 year cut-off for the 2 year survivors or last survival follow-up for those who died or were censored within 2 years (dropouts underlined)

	Time from last QoL assessment				Total
	<3	3–<6	6–<12	12+	
To 2 year cut-off for those who:					
• survived 2 years and had post 2 year QoL assessment	15	4	2	2	23
• survived 2 years, had no post 2 year QoL assessment but had known death date	5	1	3	4	13
• survived 2 years, had no post 2 year QoL assessment or known death date	1	0	2	3	6
To last survival follow-up for those who:					
• died within 2 year study period	46	17	13	4	80
• censored within 2 year study period	13	3	1	1	18
Total	80	25	21	14	140

these were considered to be non-dropouts as they had a quality of life assessment within 3 months prior to the 2 year cut-off. This leaves 111 patients who dropped out of the quality of life study within the 2 year analysis period. Of these, 46 could be attributed directly to death in that they died within 3 months of the last completed assessment and 13 directly due to censoring. The reason for dropout in the remaining 52 was not known, although dropout within 3–6 months of dying could probably be attributed to illness. There were 60 patients whose curves between the last assessment and either the last survival follow-up or the fixed end-of-study time were based on imputed values for greater than 3 months and 14 of these had values imputed for greater than one year (Table 1). The imputed quality of life values become more questionable the larger the time span.

For each subject, the area under the quality of life curve within 2 years was calculated to give quality-adjusted life months achieved within this time, i.e., QALM(24). There were 18 censored QALM(24) values which were dealt with by (i) retaining them as censored values and using Kaplan–Meier estimates of QALM(24), and (ii) imputing extreme and average values for them and using standard estimates of QALM(24). These individual QALM(24) were compared across the two treatment arms (Table 2). In all cases the chemotherapy arm achieved, in terms of means, approximately one extra month quality-adjusted survival time within 2 years, whilst the medians showed a difference of 2–3 months between the treatment arms. None of these approaches deals with the censored data adequately and the alternative group-based approach, shown below, is preferred.

2.4 Group-based approaches

In the group-based approach, rather than combining quality of life and survival into a composite QALY measure for each subject and averaging them across each treatment group, average quality of life and survival are combined at the group level. In the Q-TWiST model, the mean amount of time a group spends in each health state is calculated using partitioned survival analysis.[3] The amount of Q-TWiST is then

Table 2 Quality-adjusted life months within 2 years by treatment group

	Control arm (n = 67)	Chemotherapy arm (n = 73)
Subject-based approach using Kaplan–Meier estimates		
Median	7.7	10.0
Mean (standard error)	9.7 (0.79)	10.7 (0.72)
Subject-based approach using best values for censored data		
Median	8.0	10.4
Mean (standard error)	10.1 (0.80)	11.1 (0.74)
Subject-based approach using worst values for censored data		
Median	6.8	9.5
Mean (standard error)	8.8 (0.68)	10.0 (0.67)
Subject-based approach using average values for censored data		
Median	7.4	10.4
Mean (standard error)	9.4 (0.71)	10.6 (0.67)
Group-based approach		
Mean (standard error)	9.4 (0.73)	10.4 (0.65)

cal-culated by combining this time with utilities reflecting the average quality of life of the group in each health state. The analysis generally uses utilities obtained from external valuational studies or considers the full range of possible values in a threshold utility analysis.[3] The application of partitioned survival analysis in a Q-TWiST-type model using quality of life data to define the progressive health states has been illustrated and discussed elsewhere.[15] This method, however, is based on the ability to define a meaningful set of *progressive* health states and this is often not possible with quality of life data. A simple methodology that directly combines longitudinal quality of life data with survival data at a group level has been proposed by a number of authors, each advocating the same model but suggesting alternative approaches to estimation.[9,11,16] The method has been referred to as the *integrated quality–survival product* and applied to data from cancer clinical trials.[17] It is comparable to that proposed by Lin and colleagues to analyse censored cost data.[18]

The method multiplies the survival function $S(t)$ by the quality of life function $Q(t)$ for the group, where $S(t)$ is the proportion of subjects that survive to time t and $Q(t)$ is the quality of life of those survivors. In this way a quality-adjusted survival curve is created for the group. The area under this curve, usually calculated for a restricted time period, say up to time L, gives the mean QALY for the group for this period, thus

$$QALY(L) = \int_0^L Q(t)S(t)\,dt \qquad (2)$$

The key decision is what estimators to use for the quality of life and survival functions, $Q(t)$ and $S(t)$, respectively, in (2). The survival function can be estimated from the sample of survival data using standard methods such as the Kaplan–Meier product-limit estimator, the life-table method or by fitting a parametric model. There are a number of options for the quality of life function. One option is to estimate it using a model for quality of life over time fitted to the observed data. For example, a simple linear regression model fitted to all the available data could be used as an estimate of the quality of life function.[17] Alternatively a lowess or kernel-type smoother could be

applied to the sample to estimate mean quality of life for the group over time.[16,17] A similar but simpler approach is to calculate the mean quality of life of survivors at the fixed assessment times and the quality of life function over continuous time can be created by connecting these estimates using either a step function or by assuming a linear change.[9] If quality of life assessments are not taken at fixed times or the use of actual rather than fixed assessment times is preferable then the quality of life curves for each individual, as formed for the subject-based analysis, can be used to determine the quality of life at any time point t_j and the mean of these values across all individuals that are alive and uncensored at this time gives an estimate of the quality of life function at time t_j, that is

$$\hat{Q}(t_j) = 1/n_j \sum_{i=1}^{n_j} q_i(t_j) \tag{3}$$

where $q_i(t_j)$ is the estimate of quality of life at time t_j for surviving and uncensored individual i ($i = 1$ to n_j). Having used individual values to calculate $\hat{Q}(t_j)$ at a number of discrete time points, it is still necessary to join these values using say a step or linear function so that $\hat{Q}(t)$ is a continuous function over time.

Having estimated the survival and quality of life functions, the quality-adjusted survival curve is created by calculating their product and the area beneath this curve gives an estimate of the mean QALY(L). Clearly if the estimators for both survival and quality of life are simple functions of time then they can be multiplied and the integral of the quality-survival product can be calculated for time from 0 to L to give an estimate of the mean QALY(L). However, the most obvious estimator for the survival function is the Kaplan–Meier estimate, which is non-parametric and cannot be expressed as a simple function of time. The Kaplan–Meier method estimates the survival function $S(t_j)$ at each death time t_j ($j = 1$ to k) and connects these estimates using a step function, i.e., it assumes the survival function remains constant until the next death time. With this as a survival function, the estimator for QALY(L) becomes

$$\hat{QALY}(L) = \sum_{j=0}^{k} \left[\hat{S}(t_j) \int_{t_j}^{t_{j+1}} \hat{Q}(t)dt \right] \tag{4}$$

where $t_0 = 0$ and $t_{k+1} = L$.

If the estimator for the quality of life function is also piecewise constant between death times t_j, then the estimator for QALY(L) becomes

$$\hat{QALY}(L) = \sum_{j=0}^{k} \left[\hat{S}(t_j)\hat{Q}(t_j)(t_{j+1} - t_j) \right] \tag{5}$$

Thus having established the death times t_j ($j = 1$ to k) of the patients, individual quality of life curves can be used to estimate the quality of life at each death time $q_i(t_j)$ ($j = 1$ to k) and $\hat{Q}(t_j)$ can then be estimated as the mean of the values for the survivors at these times and, by assuming a step function, the quality of life estimator will be piecewise constant between death times. A valid alternative to using step functions would be to assume a linear change between time points for both of the functions.[16] Combining the stepped quality of life function with the stepped Kaplan–

Meier survival function in the integrated quality–survival product given in (5) gives quality-adjusted survival time for the group as illustrated in the next section.

In calculating the quality of life function, the problem of missing values of quality of life for survivors that have dropped out needs to be considered and dealt with appropriately. If the missing data are missing at random then they can be ignored and the quality of life function estimated from the survivors with observed data will be unbiased. Otherwise, as with the creation of individual quality of life curves in the subject-based approach, the unobserved values of quality of life from the last assessment time to the last survival follow-up can be imputed. Choosing how to deal with these missing data is not a trivial issue and may have implications on the results. Imputing single values for missing data has the advantage of being a simple approach but underestimates the variability of parameter estimates and methods such as multiple imputation would help to address this and may therefore be preferable.

The standard error for the mean $\hat{QALY}(L)$ is mathematically complicated and bootstrapping is recommended as the method for estimation,[9,16,19] enabling confidence intervals to be calculated and hypothesis tests to be carried out.

2.5 Group-based approach illustrated with data from the pancreatic cancer study

The integrated quality-survival product was calculated for each treatment arm separately to give the estimated mean QALM(24) for each treatment. The Kaplan–Meier estimate of the survival function was calculated for each death time and, in the standard way, these estimates were assumed to remain constant until the next death time creating a stepped function as shown in Figure 4a. There were 41 distinct death times within the 2 year analysis period on the control arm and 37 on the chemotherapy arm. The quality of life at each death time within each group was calculated on an individual basis using their quality of life curves as defined for the subject-based approach previously described. The mean quality of life at each death time was then calculated from the estimates of those individuals who are known to be alive and uncensored at that time. Due to the various assumptions made to create the individual quality of life curves, all patients who are known to be alive at a particular time point will have a value even if they have dropped out of the quality of life study for some reason. The quality of life function over continuous time was estimated by assuming the quality of life remained constant between death times, thus creating a stepped function as shown in Figure 4b. The quality-adjusted survival functions for each group, as shown in Figure 4c, were created by multiplying the survivor and quality of life functions together and the areas under these curves calculated using (5) give the estimated mean QALM(24) for each group.

To estimate the standard error of mean QALM(24) for each treatment arm, 500 bootstrap samples were taken from the observed data on each treatment arm and the mean QALM(24) was estimated for each sample. From the distribution of sample means, the standard errors were estimated.

The mean QALM(24) for the chemotherapy arm was 10.4 (standard error = 0.65) compared to 9.4 (standard error = 0.73) for the control group, so within 2 years of entry to the trial, chemotherapy gives patients on average one extra quality-adjusted life

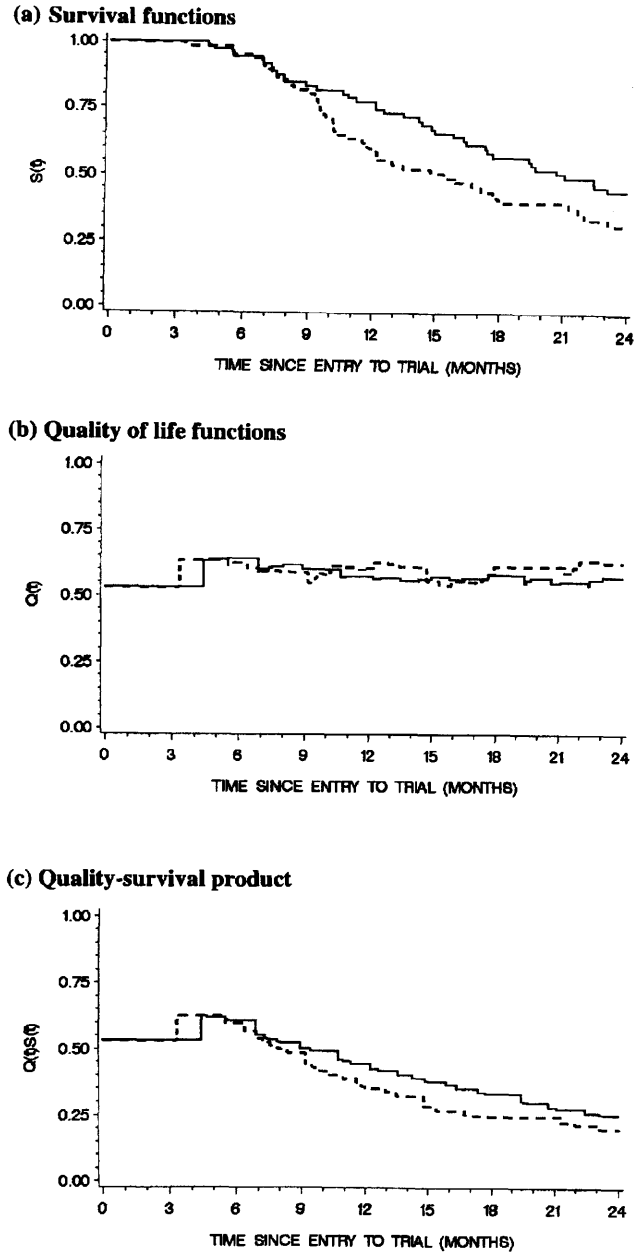


Figure 4 Group-based approach to quality-adjusted survival analysis for chemotherapy arm (solid line) and control arm (dotted line) in the pancreatic cancer study. (a) Survival functions. (b) Quality of life functions. (c) Quality–survival product.

month. In this example, the results obtained with the group approach are comparable to the results obtained using the subject-based approach (Table 2).

3 Alternative approaches

3.1 Multistate survival analysis

Multistate survival models have been advocated as a possible means for analysing quality of life and survival data simultaneously.[20,21] Although they have been applied in a variety of clinical settings,[22–26] they have not been widely used for quality of life data.

The multistate model is defined by a finite number of health states, including death, together with the possible transitions between them. The health states can be described in terms of the nature and levels of quality of life experienced by patients during the study and should be defined such that the number of patients passing from one state to another is sufficient for adequate modelling of the data. The inclusion of death as a health state in the model deals with the situation when a patient ceases participation in the quality of life study due to death. In addition, the inclusion of a 'dropout' state allows the model to incorporate the time when patients drop out of the quality of life study prior to death and enables the dropout process to be explored. When the reasons for dropout are known, it may be appropriate to include a number of different dropout states related to the different reasons.

The movement of patients between health states is governed by transition rates. A transition rate is the instantaneous potential of transition at any point in time and is equivalent to the standard hazard rate function for a survival time distribution. The transition rates in a multistate model can be represented by Cox regression models[27] where the underlying baseline transition rate may be left unspecified or modelled parametrically by assuming the transition times follow a specific distribution. The most commonly used distributions are the exponential and the Weibull distributions, the exponential being a special form of Weibull distribution. If an exponential is assumed, then the underlying baseline transition rate is assumed to be constant, ie time-homogeneous, otherwise for a Weibull distribution, or in fact any distribution in which the hazard is a function of time, the underlying baseline transition rate is allowed to change over time, ie time-inhomogeneous. In modelling the transition rate from one state to another, simplifying assumptions that may be made include (a) a Markov process or (b) a semi-Markov process. In a Markov process the transition rate from state to state is dependent only upon the present state occupied; in a semi-Markov process the transition rate is also dependent upon the *duration* of time in that state.[28]

Model parameters are estimated using the observed transition times for patients. Exact transition times tend to be available for the transition to death but transition times from one alive state to another need to be estimated from the timing of the quality of life assessments. When a dropout state is included in the model as a transient state on the way to death, the time after their last quality of life assessment at which a patient is deemed to enter this state needs to be defined.

The application of such models to quality of life data has been illustrated elsewhere.[15] Multistate survival analysis allows the effect of treatments on different health state transitions to be explored. If a single overall statement of the superior treatment is

required explicitly then some sort of trade-off between the separate results would be needed.

3.2 Joint modelling

The change in quality of life over time and the time to death can be considered as two simultaneous processes occurring in patients, and can be modelled as such. Models are assumed for each process and the parameters can be estimated simultaneously in a single analysis. There is an increasing literature on the joint modelling of repeated measures and survival outcomes.[29–32] The majority of examples have used classical approaches, such as selection models and pattern-mixture models, and there are only a limited number of examples of the application of such methods to longitudinal quality of life data.

One example[33] simultaneously analyses the quality of life and survival data using a selection model originally proposed by Schluchter[34] in which the intercept and slope from a linear random effects model of quality of life over time together with the log survival time are assumed to follow a trivariate normal distribution. Estimation of the parameters of this joint distribution is problematic when the data includes censored survival times, but these difficulties may be overcome by using the EM algorithm. This method could be extended to encompass a more complex random effects model, but convergence problems for parameter estimation could occur unless there is a large number of subjects.[35]

Fairclough and colleagues[35] not only illustrate the use of Schluchter's model in a quality of life and survival application but also implement a pattern-mixture model. In this latter approach patients are stratified according to different missing data patterns and within each strata the parameters in the model of quality of life over time are estimated. These individual strata estimates are then combined in a weighted average to give estimates of the overall population parameters. Types of missing data patterns need to be defined such that there are sufficient patients within each stratum for adequate modelling of the parameters. The most simplistic categorization is to create two strata, one consisting of 'completers' and a second consisting of dropouts at any time. Depending on the number of subjects available, additional strata can be added according to time of, and reason for, dropout. An alternative, and arguably more clinically relevant approach (see Discussion), has been proposed that uses the pattern-mixture models in a composite approach rather than averaging the parameter estimates over the different dropout patterns.[36]

A Bayesian approach to simultaneous modelling of repeated measures and survival data has also been considered[37,38] and the application of these methods to longitudinal quality of life has recently been illustrated.[39] As with the classical approaches two interlinked models are specified for the quality of life and survival processes. For the survival process, Cox regression models, accelerated failure time models and parametric models, either over all time or piecewise, should all be considered. For the quality of life process, a mixed model that adequately describes the pattern of quality of life over time should be considered. Markov Chain Monte Carlo methods, in particular Gibbs sampling, can then be used to fit the models within a single analysis. This gives the joint posterior distribution for *all* unknown parameters in both models. In this way, the parameter

estimates in the quality of life model will account for informative dropout due to death and the parameter estimates in the survival model will be based on the true underlying value of quality of life and will thus allow for potential measurement error. By using general linear mixed models, non-normal quality of life data can also be accommodated.[39]

The joint modelling approach offers an attractive future direction for the simultaneous analysis of quality of life and survival data and further research in this area is in progress.

4 Discussion

4.1 Missing data and multiple measures

In clinical trials where quality of life and survival are both important outcomes, it may be advantageous to assess treatments in terms of both outcomes simultaneously. The analysis and reporting of quality of life and survival as separate outcomes is important to allow any conflict between them in terms of treatment effect to be apparent,[20] but the simultaneous analysis can supplement this and allow the net benefit to be assessed. The reporting of quality of life as a separate outcome, however, needs considerable care as the problem of informative dropout will generally need to be accounted for and simultaneous analysis of quality of life with survival can at least account for the informative dropout due to death of the patients within the quality of life study time frame. One view held is that the time-dependent structure of the individual quality of life process can best be accounted for when quality and quantity of survival are analysed simultaneously.[21]

The simultaneous analysis of quality of life and survival addresses the issue of missing data within the quality of life study time frame beyond death, but the problem of missing data due to dropout from the quality of life study prior to death has not been fully addressed here. Imputation may be one approach that could be adopted to deal with the missing data or it may be preferable to consider the dropout time as a quality of life health state (or maybe several) that is experienced prior to death. As well as deciding how to deal with the missing data created by dropout, a decision as to how to define when a patient becomes a dropout needs to be made and different definitions may result in different conclusions.

The problem of intermittent missing quality of life data is a further issue that needs consideration. In many cases the intermittent missing data will be missing at random and therefore ignorable, but if the reason for missingness is known to be related to the health of the patient then an analysis that ignores this may give biased results. Methodology for dealing with informative intermittent missing values is not well developed. One simple option could be to impute values such that, if the patient is known to have missed quality of life assessments due to illness, an appropriate quality of life value that reflects this could be used to impute the missing values.

Other missing data issues that have not been addressed are the problem of missing baseline values and how to handle the situation when a patient has completed only one quality of life assessment. In the illustrative example used here we excluded such patients but this could introduce bias and approaches, such as the imputation of missing

data or the use of model-based approaches, which allow these patients to be included should be considered.

The methods for simultaneous analysis are generally based on a single measure of quality of life. In reality, there will be a number of quality of life measures, and there are a number of approaches that can be taken to deal with the multiple outcomes. It may be appropriate to analyse a single quality of life measure, either in the form of a summary score created from various measures, or some global score used as a summary measure or a single measure identified as the one of particular relevance, ideally decided at the study design stage. Other measures could undergo a purely descriptive analysis. Alternatively, if several measures are of interest the simultaneous analysis could be repeated for each one. This obviously causes problems of multiple testing, takes no account of correlations between quality of life measures, and conflicting conclusions could prove difficult to interpret. Multi-level modelling has been proposed to analyse multiple measures of quality of life over time[2] and this approach has been extended to incorporate survival.[33] Further work is needed to extend the methods discussed in this paper to account for multiple quality of life outcomes.

4.2 Quality-adjusted survival analysis

Quality-adjusted survival analysis provides a relatively straightforward approach for the simultaneous analysis of quality of life and survival data. Ideally it requires the quality of life measures over time to be utilities but the approach is applicable to any measure from a quality of life instrument provided that it is transformed onto a 0–1 scale. The choice of transformation will generally be linear but in some cases it may be more appropriate to use one that is non-linear.[17] The interpretation of the resultant measure as a QALY however may be questionable if the quality of life measures are not true utilities.

The subject-based approach of combining quality and quantity of life into individual QALYs using the area under the utility-type curve is the simplest form of quality-adjusted survival analysis. A number of assumptions need to be made to create a quality of life curve over time from which QALYs can be calculated. Different assumptions will produce different curves and give different results[40] and the assumptions may be questionable if the time spans between assessments are large. Although the method deals with the problem of missing quality of life data due to death by allocating a value of 0 to all time beyond death, it does not deal with missing data resulting from dropout from the quality of life study prior to death. Values for these missing data can be imputed, in particular if the reasons for dropout are known or clinical data are available an appropriate value can be allocated to the time spent as a dropout to reflect the quality of life during that time. Other simple methods of imputation such as last value carried forward and linear interpolation that have been used here may not be adequate and more sophisticated methods such as multiple imputation need to be considered and their use within this context needs to be assessed. Some authors who use an area under the curve approach to analyse longitudinal quality of life data, deal with dropout, including that due to death, by dividing the area by the length of the observation time from trial entry to last assessment and compare treatments using a standardized area under the curve.[41] This is effectively an analysis of the distribution of quality of life

during the time that they participate in the study. The problem with this approach is that individuals with a short follow-up time will have equal weighting to those with a long follow-up time.[11] In particular, if death is not accounted for then, for example, an individual with a quality of life value of 0.7 for 6 months who then dies will be treated with equal weight as someone who survives for the whole 2 year analysis period with a quality of life value of 0.7.

The subject-based approach may be problematic if there are subjects with censored survival times. This may be overcome by restricting the analysis to an upper time limit, which may automatically happen if the collection of quality of life data is restricted to a limited time period. The choice of upper time limit is subjective, and the possible inclusion of available quality of life and survival information after this time needs to be considered carefully to ensure that no bias is introduced. In many studies restricting the analysis to an upper time limit will merely minimize, rather than eliminate, the number of censored QALYs and in this case, although imputing values for the censored data in a sensitivity analysis is an option, in general a group-based approach will be preferable if censored data are present.

The integrated quality-survival product provides a simple method for combining quality of life and survival data at the group level whilst dealing with censored survival times. It is preferable to the method of partitioned survival analysis that was recommended for Q-TWiST analysis since it is not based on progressive health states, which are generally difficult to define in relation to quality of life data. Instead the method directly incorporates longitudinal quality of life data collected on patients in a quality-adjusted survival analysis.

There are number of choices to be made when implementing this method, in particular estimates for the quality of life and survival functions need to be chosen and the analysis in general will need to be restricted to an upper time limit. Sensitivity analysis is recommended to assess the robustness of the results to these choices. The most obvious choice for the survival function is the Kaplan–Meier survival curve. For the quality of life function, using individual's functions of quality of life over time to estimate the average quality of life of survivors at a number of given time points is the most flexible approach in that it allows for differing assessment times. The problems associated with creating individual quality of life curves in the subject-based approach are applicable here. For both the quality of life and survival functions, assumptions are required to realistically map the values estimated at discrete time points to their estimated course in continuous time. The major problem with the group-based approach is the difficulty in calculating standard errors, although software to perform bootstrapping is becoming more readily available.

4.3 Multistate survival analysis

Multistate survival analysis provides a flexible approach for the simultaneous assessment of quality of life and survival data, and may provide a greater insight into the effect of treatments on quality of life than quality-adjusted survival analysis. The method, however, can be problematic since it requires a number of simplifying assumptions to be made and extensive data for adequate modelling of the transition rates.

The method can be used with any quality of life instrument, providing they can be used to define meaningful health states. The quality of life measures do not need to be utilities, and for multidimensional instruments, combinations of values for the different dimensions could be used to define the health states. Defining health states using quality of life data is subjective, and the effects of using different definitions should be considered as part of a sensitivity analysis. Definitions will differ depending on the single quality of life variable or combinations of variables to be used, the number of health states to be included, and the cut-off values used to discriminate between health states. At the one extreme, the model needs to be sufficiently complex in order to be clinically meaningful and to also ensure that information from the data is utilized to a maximum. At the other extreme, a simple model is needed to ease the interpretation of the analysis and to allow a sufficient number of transitions between health states for adequate estimation of transition rates.

There are a number of choices to be made regarding the appropriate model for the transition rates. The Cox regression model is an attractive option due to its lack of assumptions regarding the baseline transition rates, but it requires that the hazards of transition are proportional, and this should be checked. If the assumption of a particular probability distribution for the data is valid then a parametric model may be a more appropriate choice, but the goodness-of-fit should be checked. A choice also has to be made as to whether the underlying process is Markov or semi-Markov.

To be able to model the transition rates accurately, the 'exact' times of transition between health states are required. These can be estimated and the accuracy is determined by the frequency of the quality of life assessments. Alternative methods that do not require exact transition times are available,[42] but these methods require specialized software that is not readily accessible to researchers.

One of the advantages of the method is that it provides a means of dealing with the problem of informative dropout. The inclusion of death as a health state in the model deals with dropout from the quality of life study due to death and the inclusion of one or more dropout states allows dropout prior to death to be incorporated and enables the dropout process to be explored.

4.4 Joint modelling

Joint modelling is the most complex approach but enables the separate quality of life and survival processes to be modelled whilst accounting for the other process and allows the interrelationship between the two processes to be examined. The appropriateness of both selection and pattern mixture models for jointly modelling quality of life data over time with survival has been questioned since the aim of these analyses is to estimate the hypothetical complete-data marginal means, that is the quality of life that would have been observed had the patients not died. This is not clinically relevant and a composite approach has been proposed that estimates more relevant measures, namely the conditional mean of the completers together with the probability of completion.[36]

Joint modelling, unlike quality-adjusted survival analysis and multistate survival analysis, requires distributional assumptions for the quality of life measures. The majority of examples of joint modelling for repeated measures over time and survival assume that the repeated measures have a normal distribution, which will often not be appropriate for quality of life outcomes. The Bayesian approach, however, could easily

be adapted to be applicable to any type of distribution. In general, the Bayesian approach to joint modelling allows a greater degree of flexibility. The models for each process can be as complex as required, though a balance has to be maintained between complexity and ease of interpretation, and it may be possible to incorporate a model for the missing data process to account for data missing for reasons other than death.[43] There is also potential to extend the method to include models of a hierarchical nature that would account for the multidimensionality of quality of life data.

5 Summary

In summary, quality-adjusted survival analysis is the simplest and most accessible approach for the simultaneous analysis of quality of life and survival data. In particular the more recently proposed integrated quality–survival product formulation provides a means for directly combining longitudinal quality of life and survival data collected on patients. We believe that this will probably become the standard approach for quality-adjusted survival analysis in clinical trials. Although quality-adjusted survival analysis overcomes the problem of missing quality of life data due to death, it does not adequately deal with missing data resulting from dropout from the quality of life study prior to death. Alternative modelling-based approaches, such as multistate survival analysis or joint modelling, may provide the means to model the dropout processes explicitly. Currently, these methods have not been developed fully for quality of life data and their application in this field needs further evaluation. We believe that the joint modelling approach in particular offers an attractive future direction for the simultaneous analysis of quality of life and survival data and further research in this emerging area is actively being pursued.

Acknowledgements

We would like to thank the National Health Service Health Technology Assessment Methodology Programme (project 93/50/04) for funding the research in the initial stages and the Cancer Research Campaign for their continued support. We would also like to thank all our colleagues at Birmingham and Leicester for their support and advice, in particular Deborah Stocken and Nicola Cooper for their help in reviewing the paper together with Professor David R Jones for his invaluable comments and continued contributions to the ongoing research.

References

1 Bowling A. *Measuring health: a review of quality of life measurement scales.* Milton Keynes: Open University Press, 1991.

2 Beacon HJ, Thompson S. Multi-level models for repeated measurement data: application to quality of life data in clinical trials. *Statistics in Medicine* 1996; 15: 2717–32.

3 Glasziou PP, Simes RJ, Gelber RD. Quality adjusted survival analysis. *Statistics in Medicine* 1990; 9: 1259–76.

4 Torrance GW. Utility approach to measuring health-related quality of life. *Journal of Chronic Diseases* 1987; 40 (6): 593–600.

5 Gelber RD, Gelman RS, Goldhirsch A. A quality-of-life-oriented endpoint for comparing therapies. *Biometrics* 1989; **45**: 781–95.

6 Gelber RD, Goldhirsch A. A new endpoint for the assessment of adjuvant therapy in postmenopausal women with operable breast cancer. *Journal of Clinical Oncology* 1986; **4**: 1772–79.

7 Goldhirsch A, Gelber RD, Simes RJ, Glasziou P, Coates AS. Costs and benefits of adjuvant therapy in breast cancer: a quality-adjusted survival analysis. *Journal of Clinical Oncology* 1989; **7**: 36–44.

8 Gelber RD, Cole BF, Gelber S, Goldhirsch A. Comparing treatments using quality-adjusted survival: the Q-TWiST method. *American Statistician* 1995; **49**: 161–69.

9 Glasziou PP, Cole BF, Gelber RD, Hilden J, Simes RJ. Quality adjusted survival analysis with repeated quality of life measures. *Statistics in Medicine* 1998; **17**: 1215–29.

10 Aaronson NK, Ahmedzai S, Bergman B *et al.* The European Organization for Research and Treatment of Cancer QLQ-C30: A quality of life instrument for use in international clinical trials in oncology. *Journal of the National Cancer Institute* 1993; **85**: 365–76.

11 Ganiats TG *et al.* Comparing the quality-adjusted life-year output of two treatment arms in a randomised trial. *Medical Care* 1995; **33**: AS245–AS254.

12 Korn EL. On estimating the distribution function for quality of life in cancer clinical trials. *Biometrika* 1993; **80**: 535–42.

13 Diggle PJ, Liang K, Zeger SL. *Analysis of longitudinal data.* Oxford: Clarendon Press, 1994.

14 Fayers PM, Machin D. *Quality of life: assessment, analysis and interpretation.* John Wiley: Chichester, 2000.

15 Billingham LJ, Abrams KR, Jones DR. Methods for the analysis of quality-of-life and survival data in health technology assessment. *Health Technology Assessment* 1999; **3** (10): 1–152.

16 Hwang JS, Tsauo JY, Wang JD. Estimation of expected quality adjusted survival by cross-sectional survey. *Statistics in Medicine* 1996; **15**: 93–102.

17 Beacon HJ. The statistical analysis of self assessed quality of life data in cancer clinical trials, PhD thesis. London School of Hygiene and Tropical Medicine, University of London, 1996.

18 Lin DY, Feuer EJ, Etzioni R, Wax Y. Estimating medical costs from incomplete follow-up data. *Biometrics* 1997; **53**: 419–34.

19 Efron B, Tibshirani RJ. *An introduction to the bootstrap.* San Francisco, CA: Chapman and Hall, 1993.

20 Cox DR, Fitzpatrick R, Fletcher AE, Gore SM, Spiegelhalter DJ, Jones DR. Quality-of-life assessment: can we keep it simple? *Journal of the Royal Statistical Society (A)* 1992; **155**: 353–93.

21 Olschewski M, Schumacher M. Statistical analysis of quality of life data in cancer clinical trials. *Statistics in Medicine* 1990; **9**: 749–63.

22 Kay R. Multistate survival analysis: an application in breast cancer. *Methods of Information in Medicine* 1984; **23**: 157–62.

23 Hansen BE, Thorogood J, Hermans J, Ploeg RJ, van Bockel JH, van Houwelingen JC. Multistate modeling of liver transplantation data. *Statistics in Medicine* 1994; **13**: 2517–29.

24 Andersen PK. Multistate models in survival analysis: a study of nephropathy and mortality in diabetes. *Statistics in Medicine* 1988; **7**: 661–70.

25 Marshall G, Jones RH. Multistate models and diabetic retinopathy. *Statistics in Medicine* 1995; **14**: 1975–83.

26 Gentleman RC, Lawless JF, Lindsey JC, Yan P. Multistate Markov models for analyzing incomplete disease history data with illustrations for HIV disease. *Statistics in Medicine* 1994; **13**: 805–21.

27 Kay R. The analysis of transition times in multistate stochastic processes using proportional hazard regression models. *Communications in Statistics—Theory and Methodology* 1982; **11**: 1743–56.

28 Andersen PK, Borgan O, Gill RD, Keiding N. *Statistical Models Based on Counting Processes.* New York: Springer, 1993.

29 Henderson R, Diggle P, Dobson A. Joint modelling of longitudinal measurements and event time data. *Biostatistics* 2000; **1**: 465–80.

30 Hogan JW, Laird NM. Mixture models for the joint distribution of repeated measures and event times. *Statistics in Medicine* 1997; **16**: 239–57.

31 Hogan JW, Laird NM. Model-based approaches to analysing incomplete

longitudinal and failure time data. *Statistics in Medicine* 1997; **16**: 259–72.
32 Wulfsohn MS, Tsiatis AA. A joint model for survival and longitudinal data measured with error. *Biometrics* 1997; **53**: 330–39.
33 Ribaudo HJ, Thompson SG, Allen-Mersh TG. A joint analysis of quality of life and survival using a random effect selection model. *Statistics in Medicine* 2000; **19**: 3237–50.
34 Schluchter MD. Methods for the analysis of informatively censored longitudinal data. *Statistics in Medicine* 1992; **11**: 1861–70.
35 Fairclough DL, Peterson HF, Cella D *et al.* Comparison of several model-based methods for analysing incomplete quality of life data in cancer clinical trials. *Statistics in Medicine* 1998; **17**: 781–96.
36 Shih WJ, Quan H. Testing for treatment differences with dropouts present in clinical trials—a composite approach. *Statistics in Medicine* 1997; **16**: 1225–39.
37 Faucett CL, Thomas DC. Simultaneously modelling censored survival data and repeatedly measured covariates: a Gibbs sampling approach. *Statistics in Medicine* 1996; **15**: 1663–85.
38 Berzuini C. Medical monitoring. In: Gilks WR, Richardson S, Spiegelhalter DJ, eds. *Markov chain Monte Carlo in practice*, Chap 18. London: Chapman and Hall, 1995.
39 Xu J, Zeger SL. Joint analysis of longitudinal data compromising repeated measures and time to events. *Applied Statistics* 2001; **50**: 375–87.
40 Ganiats TG, Browner DK, Kaplan RM. Comparison of two methods of calculating quality-adjusted life years. *Quality of Life Research* 1996; **5**: 162–64.
41 Qian W, Parmar MKB, Sambrook RJ, Fayers PM, Girling DJ, Stephens RJ. Analysis of messy longitudinal data from a randomized clinical trial. *Statistics in Medicine* 2000; **19**: 2657–74.
42 Kay R. A Markov model for analysing cancer markers and disease states in survival studies. *Biometrics* 1986; **42**: 855–65.
43 Lambert PC, Abrams KR, Sanso B, Jones DR. Synthesis of incomplete data using Bayesian hierarchical models: an illustration based on data describing survival from neuroblastoma, Technical Report 93-03. University of Leicester, Epidemiology and Public Health, 1999.

ELSEVIER

Lung Cancer 40 (2003) 201–202

LUNG CANCER

IASLC

www.elsevier.com/locate/lungcan

Editorial Comment

Potential hazards in the analysis of symptom improvement☆

So far systemic treatments for non-small cell lung cancer (NSCLC) have yielded only small survival benefits. Consequently, the balance between additional quantity of life and the quality of that added survival is critical. Treatment for advanced lung cancer can influence quality of life (QL) both positively through alleviation of symptoms and negatively through toxicity. Monitoring these effects and analysing and interpreting the data collected are far more difficult than the relatively standard examinations of survival within clinical trials.

In studies of QL, assessments are generally made at a number of time points during a patient's follow-up giving rise to what are commonly known as longitudinal data. A simple approach to the analysis of such data is to summarise each patient's results over time into one single summary measure [2]. One measure that could be used is the occurrence or not of some QL-related event during their follow-up. For example, Vansteenkiste et al. [1] use the occurrence of symptom improvement as their endpoint on which to compare treatments. There are a number of problems that can occur when analysing such an endpoint and these are discussed below.

The definition of any QL-related event is subjective and potentially problematic and this should always be acknowledged in the reporting of results. For example, Vansteenkiste et al. define symptom improvement as a 'lower score sustained for at least 8 weeks'. This definition can encompass a wide range of different possibilities. On the one hand, a patient could have a fractionally lower score for 8 weeks, possibly just as a result of variability in how they mark the visual analogue scale, and this would be deemed an improvement. On the other hand, someone could have a substantial decrease sustained for 7 weeks and 6 days and this would not be deemed to be an improvement. The variability of patient outcomes within the definition of symptom improvement should be checked and reported. Also, the analysis should be re-run using alternative definitions to assess if the results are sensitive to the definition of symptom improvement.

It is widely known that the only truly unbiased way to compare treatments in a clinical trial is to use an intention-to-treat analysis. This means that all randomised patients should be included in the analysis [3]. Intention-to-treat analysis for the survival endpoint is reasonably straightforward and hence commonplace but such an analysis is problematic and rarely used for the QL endpoint. For this reason QL analyses may give biased results. In particular, with the occurrence of symptom improvement as an endpoint, patients are excluded for a number of different reasons, each of which causes possible bias.

In studies of symptom improvement patients may be excluded from entry or from analysis if they do not exhibit specific symptoms at baseline. This denies the possibility of showing treatment: (i) preventing or delaying the onset of symptoms; and (ii) causing new symptoms. Furthermore, this practice may lead to bias. In the Vansteenkiste study, for example, the analysis of each symptom excludes patients who do not have that symptom at baseline. The exclusion of such patients could cause bias if the treatment groups differ with respect to the presence of baseline symptoms. Patients with missing baseline assessments could also cause problems for analysis. In addition, since patients with the most extreme symptom levels at baseline have greater potential for improvement, any differences between treatment groups in terms of baseline symptom levels could give biased results.

When using longitudinal data to determine whether or not patients have experienced a symptom improvement, the opportunity to observe the event is related to the length of follow-up time and hence in order to handle all patients equivalently, they need to have been followed-up for exactly the same amount of time. Certainly, the follow-up time on the two treatment arms, which should always be reported, needs to be comparable for an unbiased treatment comparison. This is particularly important when the definition of symptom improvement also requires the patient to maintain the improvement for a specified period of time. For example, in the Vansteenkiste study, patients who have a symptom improvement but do not have a further 8 weeks of follow-up do not conform to the definition of symptom improvement and are excluded from the analysis. This exclusion of patients could cause bias.

☆ PII of linked article S0169-5002(03)00515-9

0169-5002/03/$ - see front matter
doi:10.1016/S0169-5002(03)00074-6

Differential follow-up is caused by the early dropout of patients from the study. The number of patients who dropout over time and if possible the reason for dropout should be reported. If a patient drops out due to death or disease progression prior to obtaining a symptom improvement then they can be included in the analysis by categorising their outcome as no symptom improvement. If a patient drops out for other or unknown reasons then it may be more difficult to include these patients. The inclusion of all patients in an intention-to-treat analysis would require the outcomes for these patients to be imputed by using other known clinical data or by making assumptions about such patients. If the dropout rate and the reasons for dropout are comparable for the different treatment arms then the exclusion of such patients may not bias the treatment comparison. Otherwise, more complex statistical methods that analyse the longitudinal data whilst accounting for the dropout of patients should be employed [4].

The exclusion of patients not only lead to bias but also results in small numbers for analysis, and hence less precise estimates and reduced power to detect treatment differences. In addition, the conclusions will only be valid for the specific group included. For example, the analysis presented by Vansteenkiste et al. in Figures 1 and 2 relates to a very select and small group of patients. The confidence intervals for the estimates shown will be wide and inferences from the analysis can only be made to those patients who are well enough to undergo the full duration of treatment.

In general, a quality of life questionnaire assesses a number of different symptoms and the analysis will often consider each symptom separately. This can result in large numbers of hypothesis tests. For example, Vansteenkiste et al. analyse nine different measures of quality of life. The hazards of multiple testing are well known [5], with the chance of incorrectly rejecting the null hypothesis increasing as the number of tests increases. Methods, such as Bonferroni adjustments, are available to adjust for multiple testing [6] and, at a minimum, the implications of multiple testing should be discussed in any report.

The detailed reporting of QL data from clinical trials, such as the paper by Vansteenkiste et al. included in this journal, is important in aiding clinical decisions about treatments. The analysis of longitudinal QL data is notoriously difficult and will usually entail many challenging decisions regarding the most appropriate methods. Here we have discussed the possibility of summarising the data as the occurrence or not of symptom improvement, which creates a simple outcome on which to compare treatments but can still be problematic to analyse. The exclusion of patients, which can reduce numbers for analysis and create less generalisable and possibly biased results, is one of the greatest problems. An intention-to-treat analysis, which includes all randomised patients, will always guarantee an unbiased comparison of treatments but is not always possible without making sweeping assumptions or employing complex statistical methods. The difficulties and potential biases should always be acknowledged in the analysis, reporting and interpretation of QL data.

References

[1] Vansteenkiste JF, Vandebroek J, Nackaerts et al., The Leuven Lung Cancer Group. Influence of cisplatin-use, age, performance status and duration of chemotherapy on symptom control in advanced non-small cell lung cancer: detailed symptom analysis of a randomised study comparing cisplatin-vindesine to gemcitabine. Lung Can. 2003; 40: 191–199.

[2] Matthews JNS, Altman DG, Campbell MJ, Royston P. Analysis of serial measurements in medical research. Br Med J 1990;300:230–5.

[3] Altman DG. Practical statistics for medical research. London: Chapman and Hall, 1991.

[4] Fairclough DL, Peterson HF, Cella D, Bonomi P. Comparison of several model-based methods for analysing incomplete quality of life data in cancer clinical trials. Stat Med 1998;17:781–96.

[5] Pocock SJ, Hughes MD, Lee RJ. Statistical problems in the reporting of clinical trials. N Engl J Med 1987;317:426–32.

[6] Zhang J, Quan H, Ng J, Stepanavage ME. Some statistical methods for multiple endpoints in clinical trials. Control Clin Trials 1997;18:204–21.

Lucinda Billingham
Cancer Research UK Clinical Trials Unit,
Institute for Cancer Studies,
University of Birmingham,
Vincent Drive, Edgbaston,
Birmingham B15 2TT,
UK
E-mail address: l.j.billingham@bham.ac.uk

Michael Cullen
Cancer Centre at the Queen Elizabeth Hospital,
University Hospitals Birmingham,
NHS Trust,
Edgbaston,
Birmingham B15 2TH,
UK

APPENDIX II

WINBUGS CODE FOR JOINT MODELLING OF QUALITY OF LIFE AND SURVIVAL DATA FROM THE MIC AND ESPAC STUDIES

MIC Quality of Life Study
Joint model for quality of life and survival or dropout
Linear random effects for GQS over time
Piecewise exponential for survival or dropout

MODEL

Using Product Normal formulation

```
model
{
for (j in 1:NQOL) {
   score[j]   ~ dnorm(mu[j],tau)
   mu[j] <- theta[pat[j],1] + theta[pat[j],2]*time[j] + delta1*trt[j] + delta2*trt[j]*time[j]
   }
for (i in 1:N) {
   theta[i,1] ~ dnorm(mu.theta1,tau1)
   theta[i,2] ~ dnorm(mu.theta2[i],tau2)
   mu.theta2[i] <- lambda0 + lambda1*(theta[i,1] - 78)
   }
mu.alpha1 <- mu.theta1
mu.alpha2 <- lambda0 + lambda1*(mu.theta1 - 78)
var.alpha1 <- 1/tau1
var.alpha2 <- (1/tau2) + (var.alpha1*pow(lambda1,2))
cov.a1a2 <- lambda1*var.alpha1
var <- 1/tau

mu.theta1 ~ dnorm(0.0,0.0001)
lambda0  ~ dnorm(0,0.0001)
lambda1 ~ dnorm(0,0.0001)
delta1 ~ dnorm(0,0.0001)
delta2 ~ dnorm(0,0.0001)

# Gamma priors for precision
 tau1 ~ dgamma(0.001,0.001)
 tau2  ~ dgamma(0.001,0.001)
 tau  ~ dgamma(0.001,0.001)

# Half normal priors for standard deviations
 tau1<-1/pow(sigma1,2)
 tau2<-1/pow(sigma2,2)
 tau<-1/pow(sigma,2)
 sigma1~dnorm(0,0.0001)I(0,)
 sigma2~dnorm(0,0.0001)I(0,)
 sigma~dnorm(0,0.0001)I(0,)

# Uniform priors for standard deviations
# tau1<-1/pow(sigma1,2)
# tau2<-1/pow(sigma2,2)
# tau<-1/pow(sigma,2)
# sigma1~dunif(0,10000)
# sigma2~dunif(0,10000)
# sigma~dunif(0,10000)

for (p in 1:NSURV) {

# Choice of predictive value for QOL
   score.start[p] <- theta[patid[p],1] + theta[patid[p],2]*timest[p] + delta1*trtarm[p] +delta2*trtarm[p]*timest[p]
```

```
    score.end[p] <- theta[patid[p],1] + theta[patid[p],2]*timeend[p] + delta1*trtarm[p]
                                                                    + delta2*trtarm[p]*timeend[p]
    score.pred[p] <- score.start[p] -78
#   score.pred[p] <- ((score.start[p]+score.end[p])/2)-78
#   score.pred[p] <- (theta[patid[p],2] + delta2*trtarm[p])+0.06

    dintv[p] ~ dpois(nu[p])
    log(nu[p]) <- offset[p] + gamma0[nintv[p]] + gamma1*trtarm[p] + gamma2*score.pred[p]
    offset[p]<-log(survtime[p])
    }
for (k in 1:F) {
    gamma0[k] ~ dnorm(0,0.0001)
    h0[k]<- exp(gamma0[k])
    }
gamma1 ~ dnorm(0,0.0001)
gamma2 ~ dnorm(0,0.0001)
}
```

Using Wishart Prior

```
model
{
for (j in 1:NQOL) {
    score[j]   ~ dnorm(mu[j],tau)
    mu[j] <- theta[pat[j],1] + theta[pat[j],2]*time[j] + delta1*trt[j] + delta2*trt[j]*time[j]
    }
for (i in 1:N){
    theta[i,1:2] ~ dmnorm(mu.alpha[], R[,])
    }

tau  ~ dgamma(0.001,0.001)
var <- 1/tau
delta1 ~ dnorm(0,0.0001)
delta2 ~ dnorm(0,0.0001)

mu.alpha[1] ~ dnorm(0,0.0001)
mu.alpha[2] ~ dnorm(0,0.0001)

R[1:2 , 1:2] ~ dwish(Omega[ , ], 2)

for (k in 1:2) {
     for (l in 1:2){
        COV[k,l]<-inverse(R[,],k,l)
        }
    }

for (p in 1:NSURV) {

# Choice of predictive value for QOL
    score.start[p] <- theta[patid[p],1] + theta[patid[p],2]*timest[p] + delta1*trtarm[p] +delta2*trtarm[p]*timest[p]
    score.end[p] <- theta[patid[p],1] + theta[patid[p],2]*timeend[p] + delta1*trtarm[p]
                                                                    + delta2*trtarm[p]*timeend[p]
#   score.pred[p] <- score.start[p] -78
#   score.pred[p] <- ((score.start[p]+score.end[p])/2)-78
    score.pred[p] <- theta[patid[p],2] + delta2*trtarm[p]

    dintv[p] ~ dpois(nu[p])
    log(nu[p]) <- offset[p] + gamma0[nintv[p]] + gamma1*trtarm[p] + gamma2*score.pred[p]
                  + gamma3*score.pred[p]*trtarm[p]
```

```
    offset[p]<-log(survtime[p])
    }
for (k in 1:F) {
    gamma0[k] ~ dnorm(0,0.0001)
    h0[k]<- exp(gamma0[k])
    }
gamma1 ~ dnorm(0,0.0001)
gamma2 ~ dnorm(0,0.0001)
gamma3 ~ dnorm(0,0.0001)
}
```

DATA

QOL Data for all models

list(NQOL=392,N=109)

pat[]	trt[]	time[]	score[]
1	0.5	-5	77.083
1	0.5	-2	80.208
1	0.5	1	87.5
1	0.5	4	95.833
1	0.5	7	81.25

Extra QOL data for Wishart prior

list(Omega = structure(.Data = c(100, 0, 0, 0.5), .Dim = c(2,2)))

Survival data

list(F=4,NSURV=393)

patid[]	trtarm[]	dintv[]	nintv[]	timest[]	timeend[]	survtime[]
1	0.5	0	1	-6	0	6
1	0.5	0	2	0	3	3
1	0.5	0	3	3	6	3
1	0.5	0	4	6	9	3
2	-0.5	0	1	-6	0	6
2	-0.5	0	2	0	3	3
2	-0.5	1	3	3	6	1.8571
3	0.5	0	1	-6	0	6
3	0.5	0	2	0	3	3
3	0.5	0	3	3	6	3

Dropout data

list(F=4,NSURV=349)

patid[]	trtarm[]	dintv[]	nintv[]	timest[]	timeend[]	survtime[]
1	0.5	0	1	-6	0	6
1	0.5	0	2	0	3	3
1	0.5	0	3	3	6	3
1	0.5	0	4	6	9	3
2	-0.5	0	1	-6	0	6
2	-0.5	0	2	0	3	3
2	-0.5	1	3	3	6	1.8571
3	0.5	0	1	-6	0	6

INITIAL VALUES

Inital values for Product-Normal formulation

list(mu.theta1 = 78, lambda0 = -0.06, lambda1=0, delta1=6, delta2=0.5)

With Gamma priors on precision

list(tau = 0.02, tau1 = 0.01, tau2 = 2.4)

With Half Normal and Uniform priors on sds

list(sigma = 7, sigma1 = 10, sigma2 = 0.6)

Inital values for Wishart prior formulation

list(mu.alpha=c(78, -0.06), delta1=6, delta2=0.5, tau=0.02, R = structure(.Data = c(0.01,0,0,2.4), .Dim = c(2,2)))

Initial values for all models from classical estimates of random effects

theta[,1]	theta[,2]
80.92	0.21
82.62	-0.54
81.12	-0.22
80.08	0.24
85.06	0.87
78.41	0.06

For piecewise exponential with 4 time intervals

list(gamma0 = c(-4,-4,-4,-4), gamma1=0, gamma2=0)
list(gamma0 = c(-4,-4,-4,-4), gamma1=0, gamma2=0, gamma3=0)

ESPAC Quality of Life Study
Joint model for quality of life and survival or dropout
Piecewise linear random effects model for GHSS over time
Piecewise exponential model for survival

MODEL

Using Wishart Prior

```
model
{
for (j in 1:NQOL) {
   score[j]   ~ dnorm(mu[j],tau)
   mu[j] <- theta[pat[j],1] + theta[pat[j],2]*timeint1[j] + theta[pat[j],3]*timeint2[j]
            + delta1*trt[j] + delta2*trt[j]*timeint1[j] + delta3*trt[j]*timeint2[j]
   }
for (i in 1:N){
   theta[i,1:3] ~ dmnorm(mu.alpha[], R[,])
   }

tau  ~ dgamma(0.001,0.001)
var <- 1/tau
delta1 ~ dnorm(0,0.0001)
delta2 ~ dnorm(0,0.0001)
delta3 ~ dnorm(0,0.0001)

mu.alpha[1] ~ dnorm(0,0.0001)
mu.alpha[2] ~ dnorm(0,0.0001)
mu.alpha[3] ~ dnorm(0,0.0001)

R[1:3 , 1:3] ~ dwish(Omega[ , ], 3)

for (k in 1:3) {
    for (l in 1:3){
       COV[k,l]<-inverse(R[,],k,l)
       }
   }
for (p in 1:NSURV) {

# Choice of predictive value for QOL
   score.start[p] <- theta[patid[p],1] + theta[patid[p],2]*timest1[p] + theta[patid[p],3]*timest2[p]
                     + delta1*trtarm[p] + delta2*trtarm[p]*timest1[p] + delta3*trtarm[p]*timest2[p]
   score.end[p] <- theta[patid[p],1] + theta[patid[p],2]*timeend1[p] + theta[patid[p],3]*timeend2[p]
                     + delta1*trtarm[p] + delta2*trtarm[p]*timeend1[p] + delta3*trtarm[p]*timeend2[p]
   score.pred[p] <- score.start[p]-62
#   score.pred[p] <- ((score.start[p]+score.end[p])/2)-62

   dintv[p] ~ dpois(nu[p])
   log(nu[p]) <- offset[p] + gamma0[nintv[p]] + gamma1*trtarm[p] + gamma2*score.pred[p]
   offset[p]<-log(survtime[p])
   }
for (k in 1:F) {
   gamma0[k] ~ dnorm(0,0.001)
   h0[k]<- exp(gamma0[k])
   }
gamma1 ~ dnorm(0,0.001)
gamma2 ~ dnorm(0,0.001)
}
```

DATA

QoL data for models

list(NQOL=710,N=175)

pat[]	trt[]	timeint1[]	timeint2[]	score[]
1	0.5	1.1184	0	83.3333
1	0.5	2.5987	0	58.3333
1	0.5	5.6908	0	41.6667
1	0.5	6	19.625	50
1	0.5	6	22.8487	50
1	0.5	6	25.5789	50

Extra data for Wishart prior

list(Omega = structure(.Data = c(200, 0, 0, 0, 10, 0, 0, 0, 1), .Dim = c(3,3)))

Survival within 24 months

list(F=8, NSURV = 950)

patid[]	trtarm[]	dintv[]	nintv[]	timest1[]	timest2[]	timeend1[]	timeend2[]	survtime[]
1	0.5	0	1	0	0	3	0	3
1	0.5	0	2	3	0	6	0	3
1	0.5	0	3	6	0	6	3	3
1	0.5	0	4	6	3	6	6	3
1	0.5	0	5	6	6	6	9	3
1	0.5	0	6	6	9	6	12	3
1	0.5	0	7	6	12	6	15	3
1	0.5	0	8	6	15	6	18	3
2	-0.5	0	1	0	0	3	0	3
2	-0.5	0	2	3	0	6	0	3
2	-0.5	0	3	6	0	6	3	3
2	-0.5	0	4	6	3	6	6	3
2	-0.5	1	5	6	6	6	9	1.4539

Dropout within 24 months

id	trtc	eventint	interval	timest1	timest2	timeend1	timeend2	survintrnd
1	0.5	0	1	0	0	3	0	3
1	0.5	0	2	3	0	6	0	3
1	0.5	0	3	6	0	6	3	3
1	0.5	0	4	6	3	6	6	3
1	0.5	0	5	6	6	6	9	3
1	0.5	0	6	6	9	6	12	3
1	0.5	0	7	6	12	6	15	3
1	0.5	0	8	6	15	6	18	3
2	-0.5	0	1	0	0	3	0	3
2	-0.5	0	2	3	0	6	0	3
2	-0.5	0	3	6	0	6	3	3
2	-0.5	0	4	6	3	6	6	3
2	-0.5	1	5	6	6	6	9	0.4408

INITIAL VALUES

Inital values for Wishart prior formulation

list(mu.alpha=c(51, 2.8, -0.9), delta1=-1, delta2=0.3, delta3=0.5, tau=0.01, R = structure(.Data = c(0.01,0,0,0, 0.09, 0, 0, 0, 0.8), .Dim = c(3,3)))

Initial values for all models from classical estimates of random effects

theta[,1]	theta[,2]	theta[,3]
-0.6251	-1.1075	0.7908
-2.0914	-1.9224	-0.02311
-20.743	5.4189	1.2106
5.1243	-2.1766	-0.451
11.3461	0.869	-0.4925
6.2172	-0.6869	-0.341

Survival/Dropout within 24 months

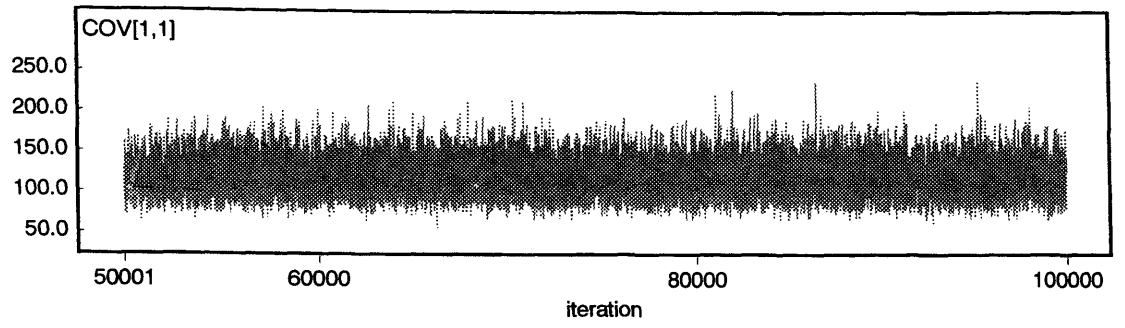
list(gamma0 = c(-3,-3,-3,-3,-3,-3,-3,-3), gamma1=0, gamma2=0)

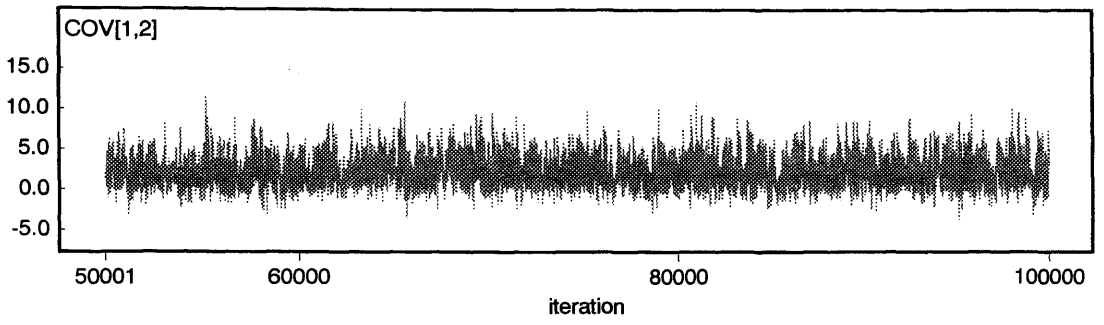
APPENDIX III A

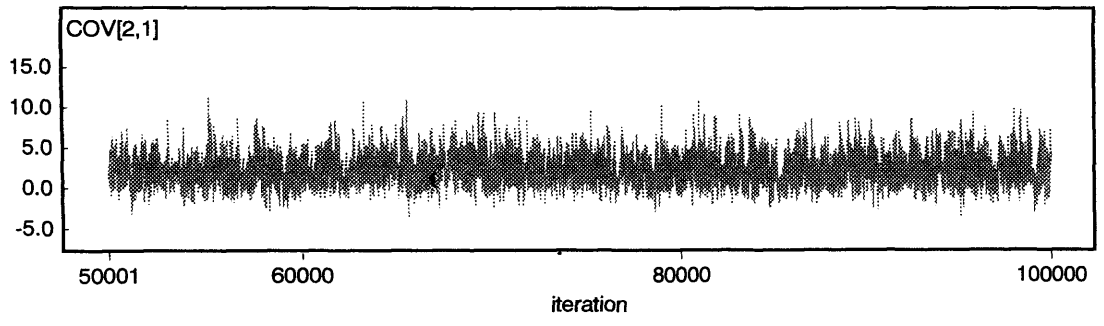
OUTPUT FROM JOINT MODELLING OF MIC DATA IN WINBUGS

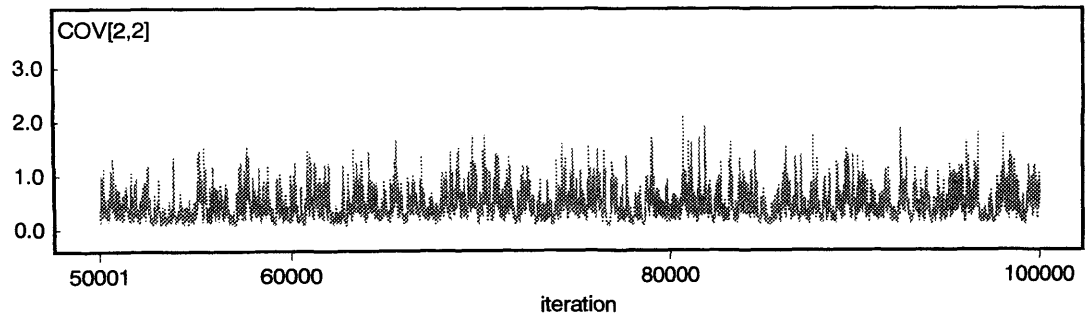
Linear random effects model for *GQS* over time with treatment and treatment by time covariates and Wishart prior distribution on precision matrix

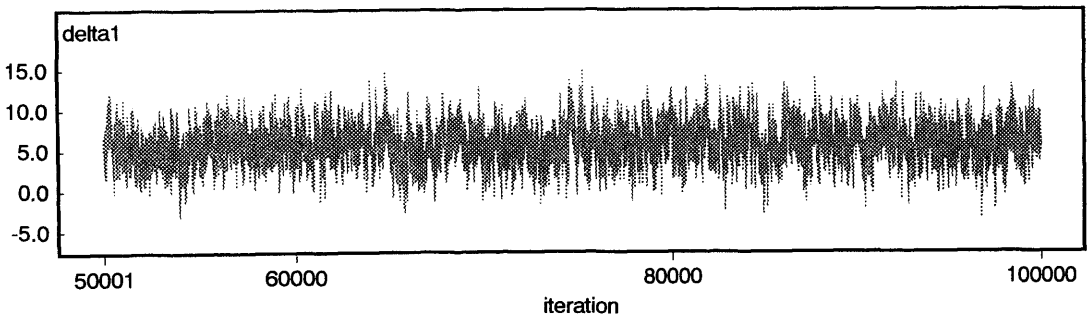
Piecewise exponential model for survival with treatment and *GQS at start of interval* as covariates

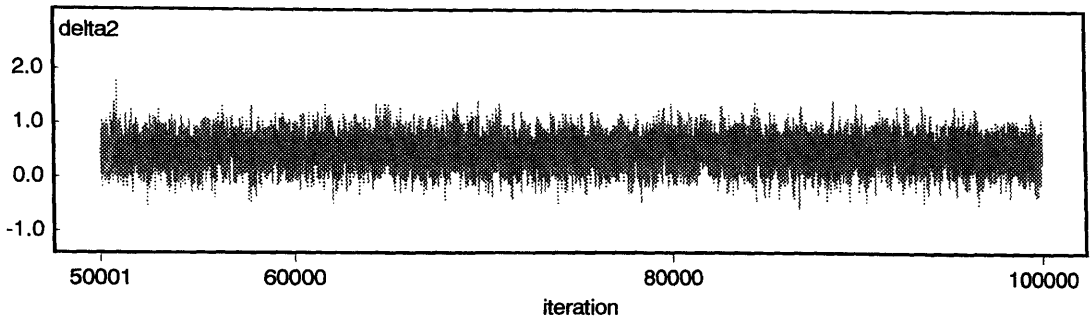

COV[1,1]
250.0
200.0
150.0
100.0
50.0
50001
60000
80000
100000
iteration

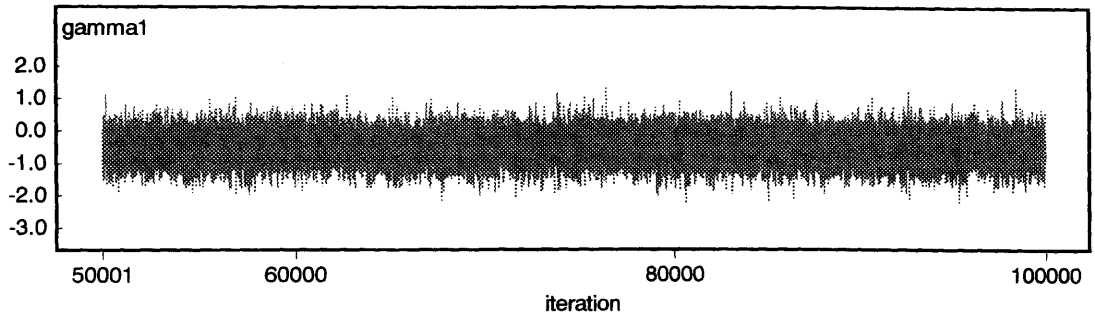

COV[1,2]
15.0
10.0
5.0
0.0
-5.0
50001
60000
80000
100000
iteration

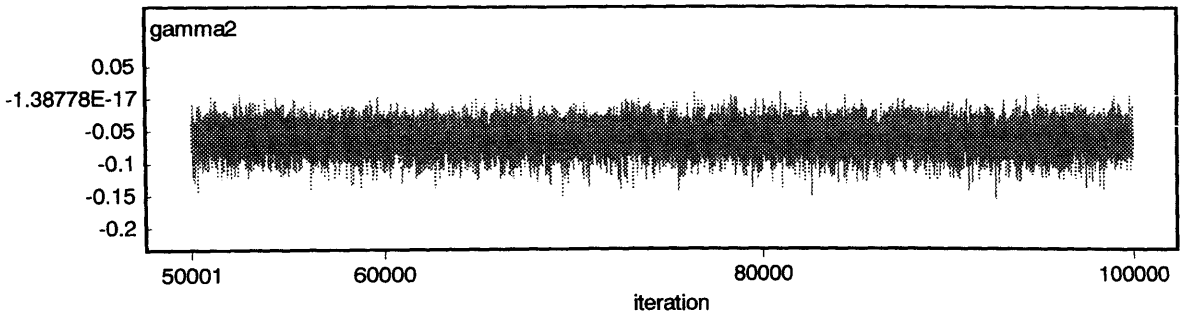

COV[2,1]
15.0
10.0
5.0
0.0
-5.0
50001
60000
80000
100000
iteration

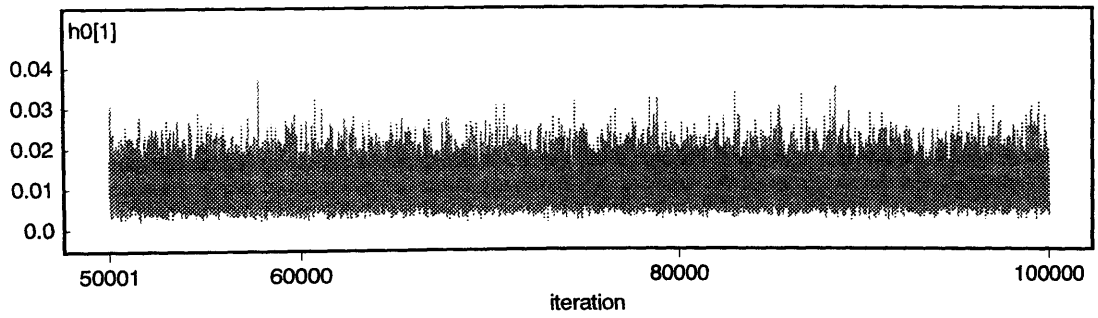

COV[2,2]
3.0
2.0
1.0
0.0
50001
60000
80000
100000
iteration

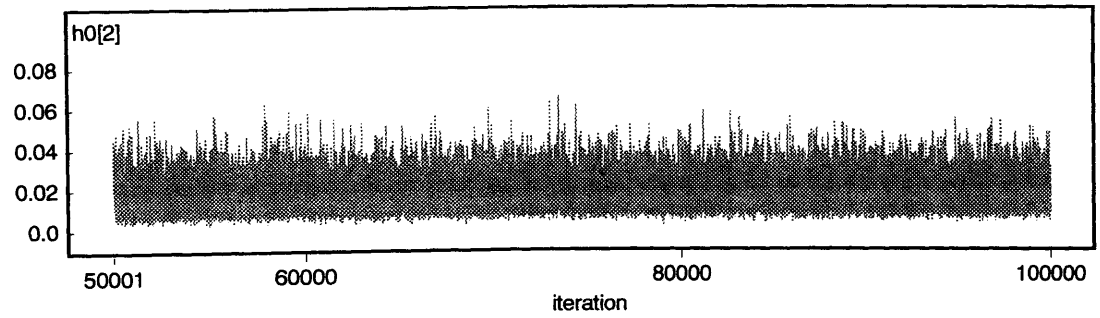

delta1
15.0
10.0
5.0
0.0
-5.0
50001
60000
80000
100000
iteration

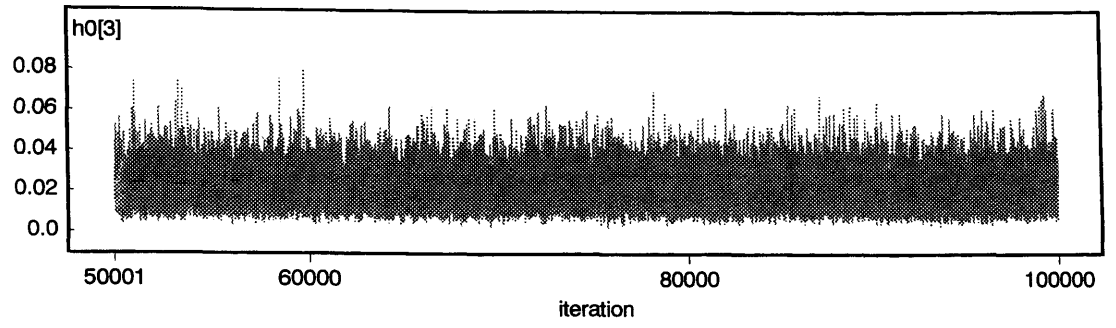

delta2
2.0
1.0
0.0
-1.0
50001
60000
80000
100000
iteration

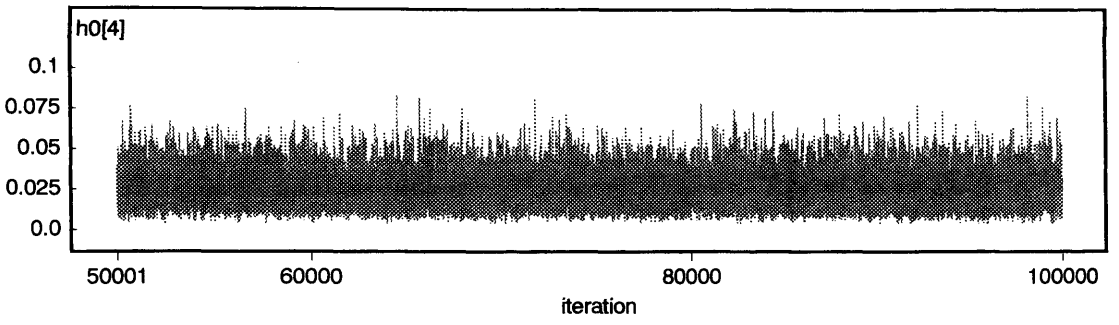

gamma1
2.0
1.0
0.0
-1.0
-2.0
-3.0
50001
60000
80000
100000
iteration

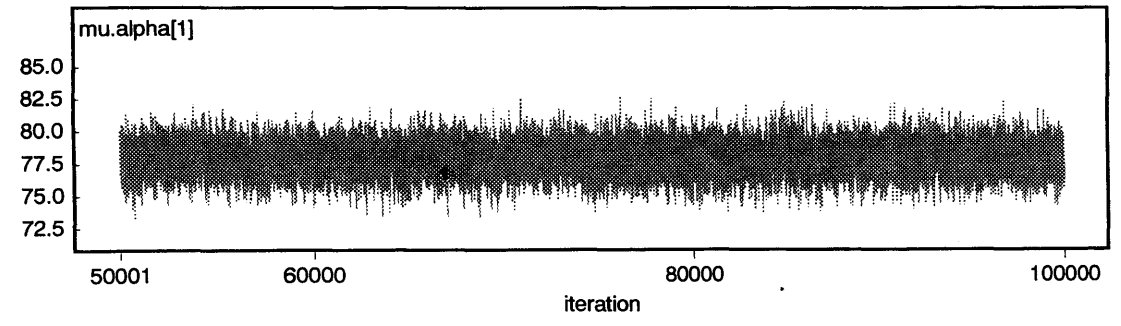

gamma2
0.05
-1.38778E-17
-0.05
-0.1
-0.15
-0.2
50001
60000
80000
100000
iteration

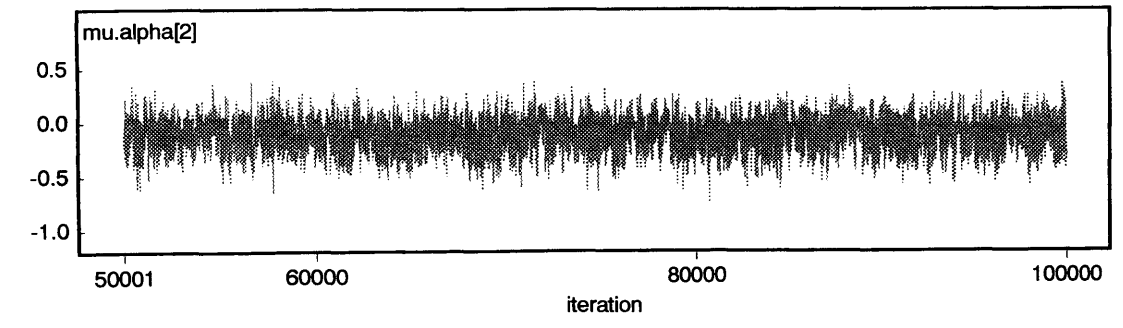

h0[1]
0.04
0.03
0.02
0.01
0.0
50001
60000
80000
100000
iteration

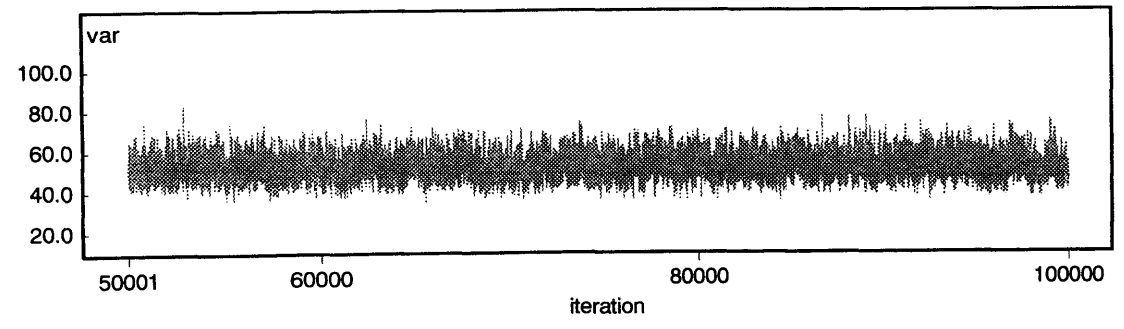

h0[2]
0.08
0.06
0.04
0.02
0.0
50001
60000
80000
100000
iteration

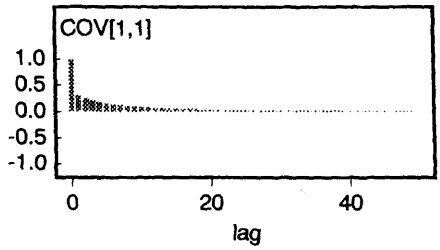

h0[3]
0.08
0.06
0.04
0.02
0.0
50001
60000
80000
100000
iteration

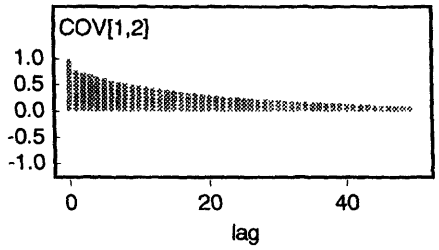

h0[4]
0.1
0.075
0.05
0.025
0.0
50001
60000
80000
100000
iteration

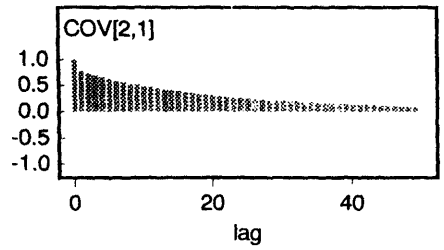

mu.alpha[1]
85.0
82.5
80.0
77.5
75.0
72.5
50001
60000
80000
100000
iteration

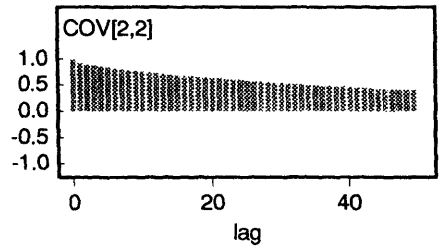

mu.alpha[2]
0.5
0.0
-0.5
-1.0
50001
60000
80000
100000
iteration

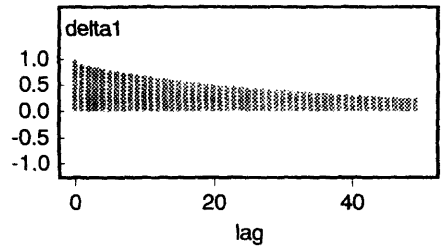

var
100.0
80.0
60.0
40.0
20.0
50001
60000
80000
100000
iteration

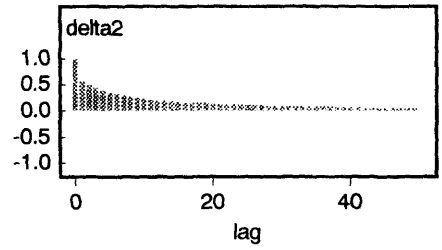

COV[1,1]
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag

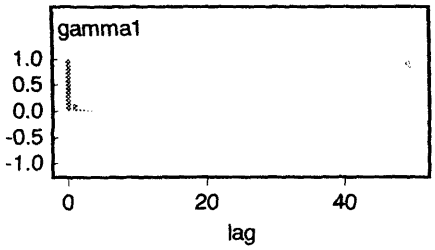

COV[1,2]
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag

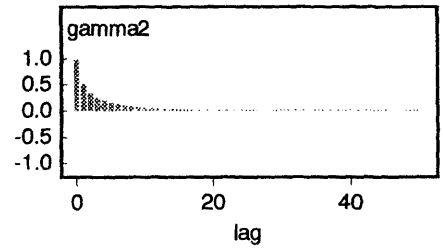

COV[2,1]
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag

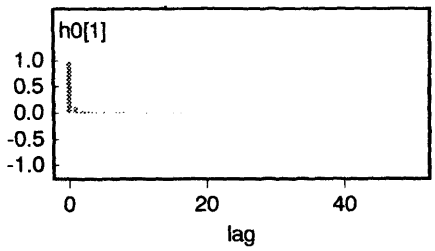

COV[2,2]
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag

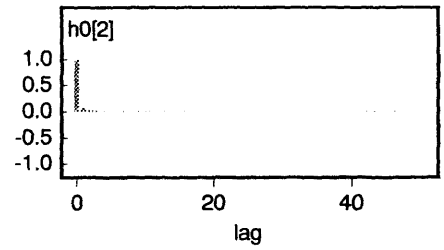

delta1
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag

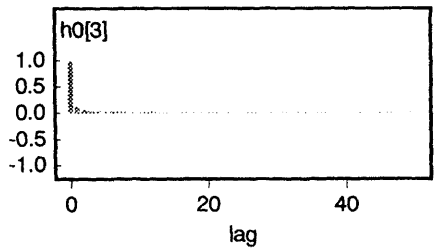

delta2
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag

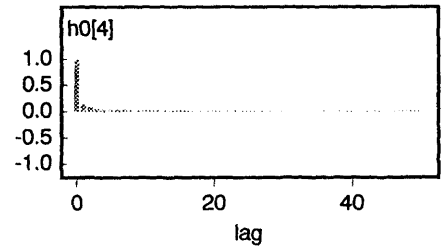

gamma1
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag

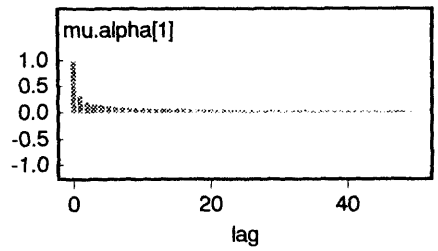

gamma2
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag

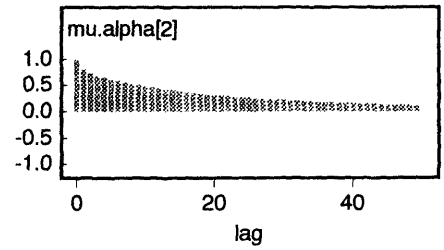

h0[1]
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag

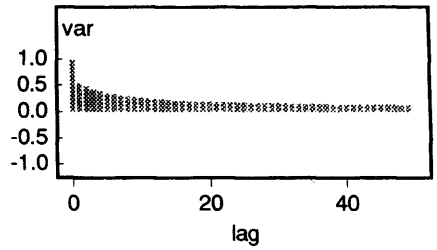

h0[2]
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag

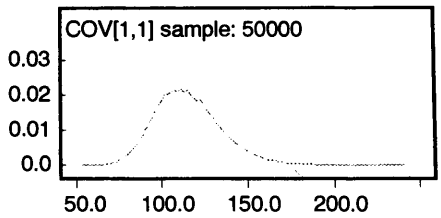

h0[3]
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag

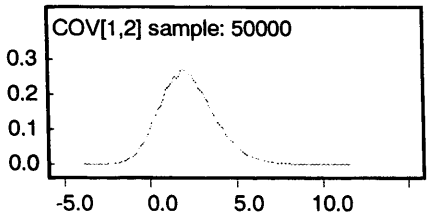

h0[4]
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag

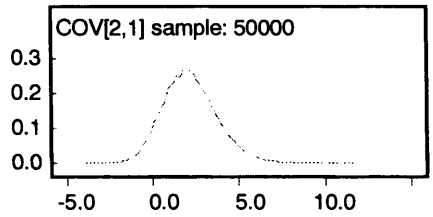

mu.alpha[1]
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag

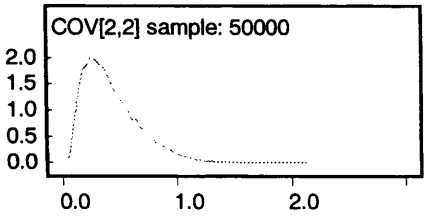

mu.alpha[2]
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag

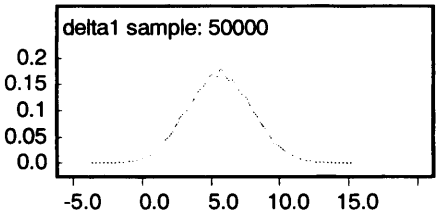

var
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag

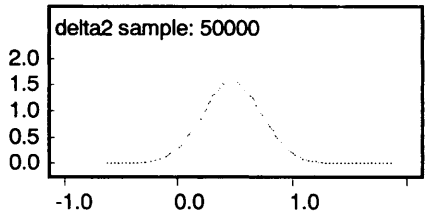

COV[1,1] sample: 50000
0.03
0.02
0.01
0.0
50.0
100.0
150.0
200.0

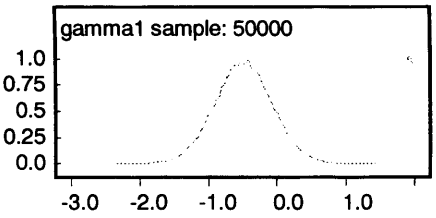

COV[1,2] sample: 50000
0.3
0.2
0.1
0.0
-5.0
0.0
5.0
10.0

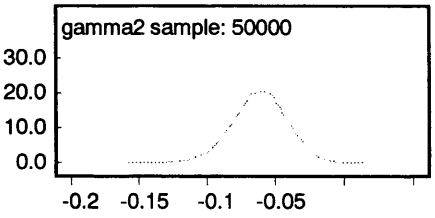

COV[2,1] sample: 50000
0.3
0.2
0.1
0.0
-5.0
0.0
5.0
10.0

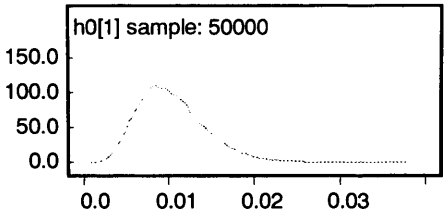

COV[2,2] sample: 50000
2.0
1.5
1.0
0.5
0.0
0.0
1.0
2.0

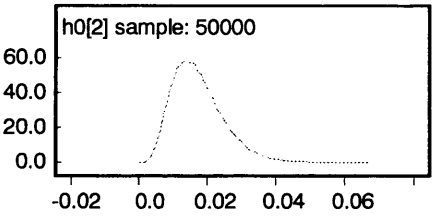

delta1 sample: 50000
0.2
0.15
0.1
0.05
0.0
-5.0
0.0
5.0
10.0
15.0

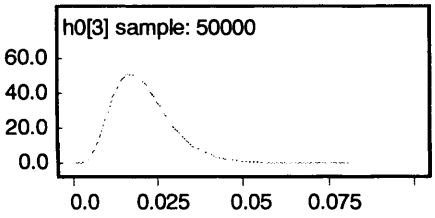

delta2 sample: 50000
2.0
1.5
1.0
0.5
0.0
-1.0
0.0
1.0

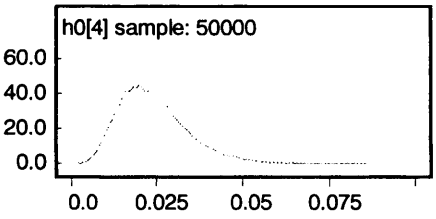

gamma1 sample: 50000
1.0
0.75
0.5
0.25
0.0
-3.0
-2.0
-1.0
0.0
1.0

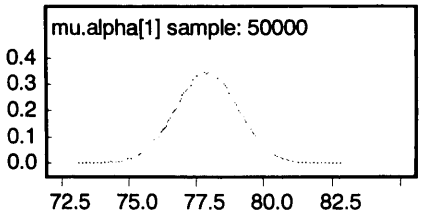

gamma2 sample: 50000
30.0
20.0
10.0
0.0
-0.2
-0.15
-0.1
-0.05

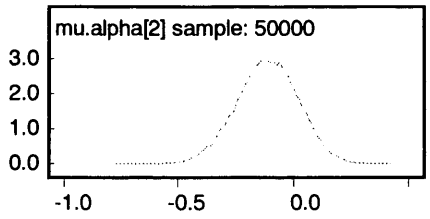

h0[1] sample: 50000
150.0
100.0
50.0
0.0
0.0
0.01
0.02
0.03

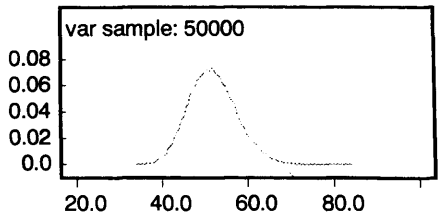

h0[2] sample: 50000
60.0
40.0
20.0
0.0
-0.02
0.0
0.02
0.04
0.06


h0[3] sample: 50000
60.0
40.0
20.0
0.0
0.0
0.025
0.05
0.075


h0[4] sample: 50000
60.0
40.0
20.0
0.0
0.0
0.025
0.05
0.075


mu.alpha[1] sample: 50000
0.4
0.3
0.2
0.1
0.0
72.5
75.0
77.5
80.0
82.5


mu.alpha[2] sample: 50000
3.0
2.0
1.0
0.0
-1.0
-0.5
0.0

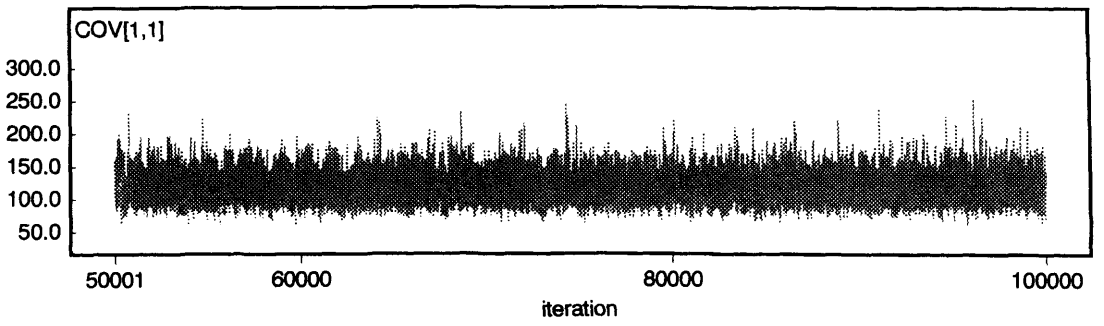

var sample: 50000
0.08
0.06
0.04
0.02
0.0
20.0
40.0
60.0
80.0

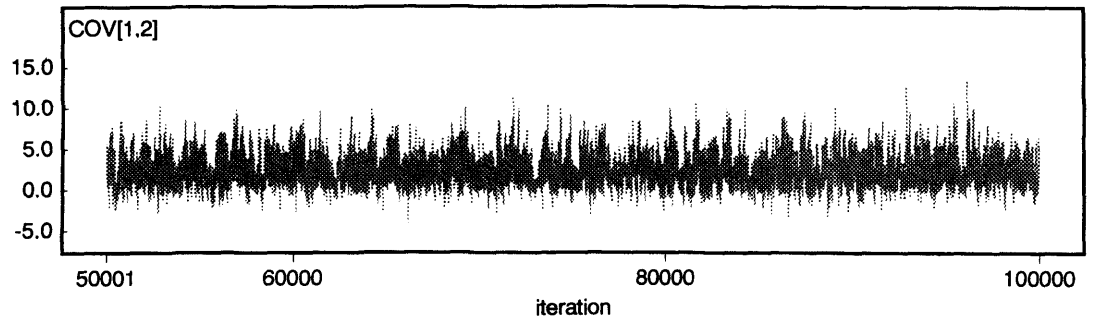
APPENDIX III B

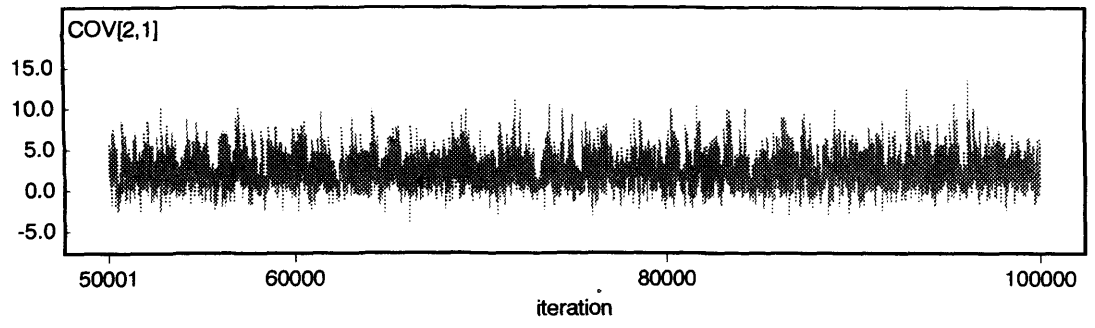
OUTPUT FROM JOINT MODELLING OF MIC DATA IN WINBUGS

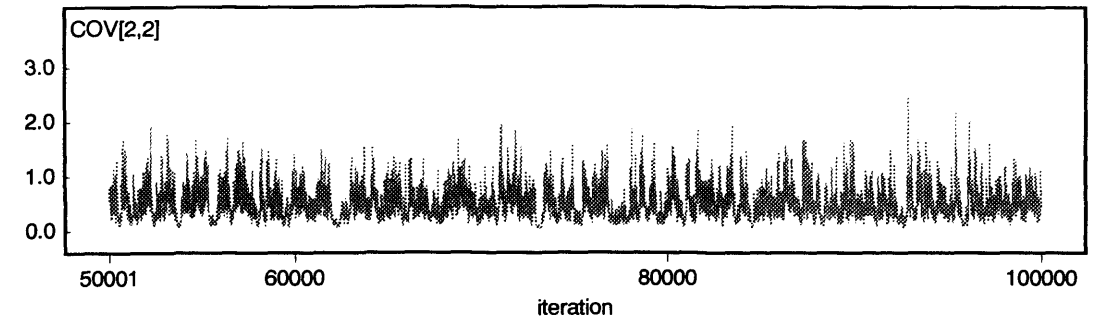
Linear random effects model for *GQS* over time with treatment and treatment by time covariates and Wishart prior distribution on precision matrix

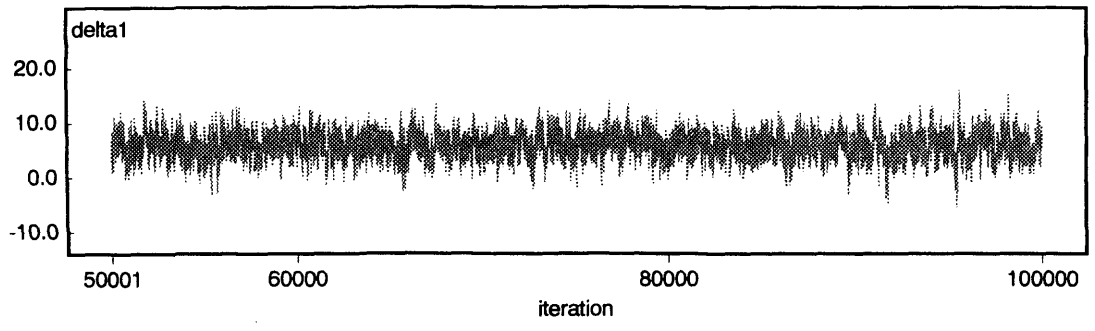
Piecewise exponential model for survival with treatment and *mean GQS for interval* as covariates

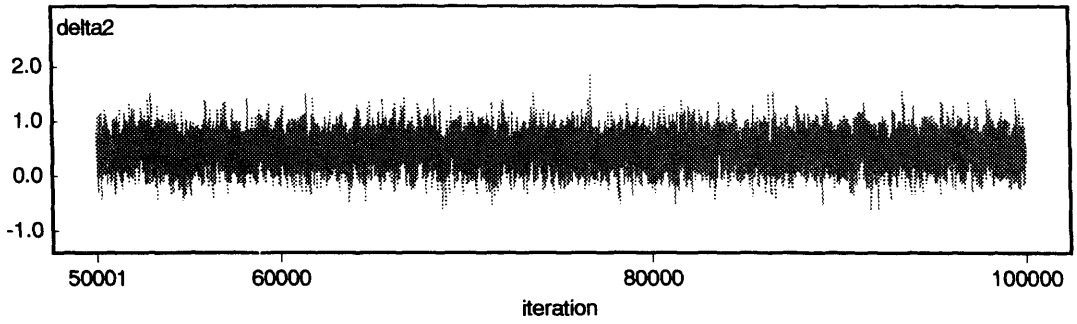

COV[1,1]
300.0
250.0
200.0
150.0
100.0
50.0
50001
60000
80000
100000
iteration

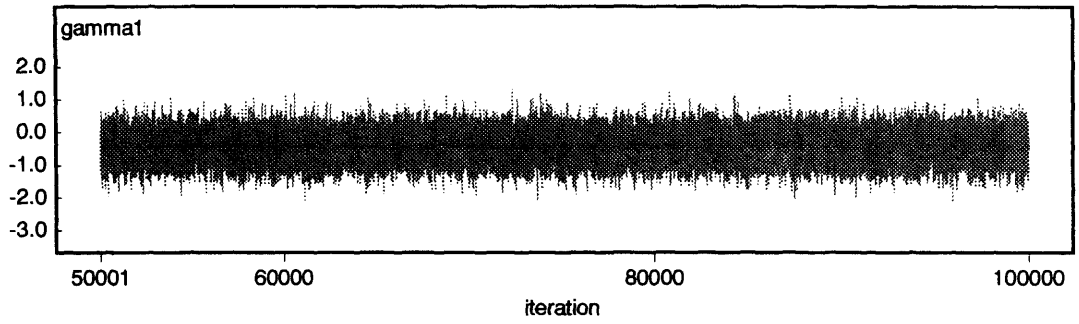

COV[1,2]
15.0
10.0
5.0
0.0
-5.0
50001
60000
80000
100000
iteration

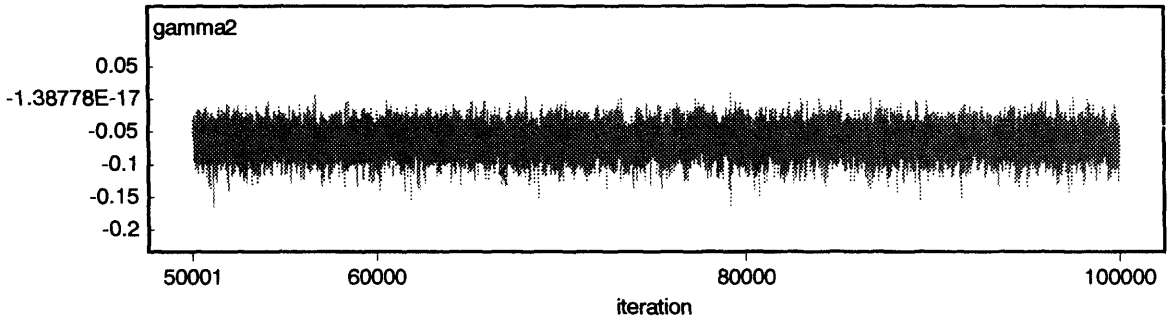

COV[2,1]
15.0
10.0
5.0
0.0
-5.0
50001
60000
80000
100000
iteration

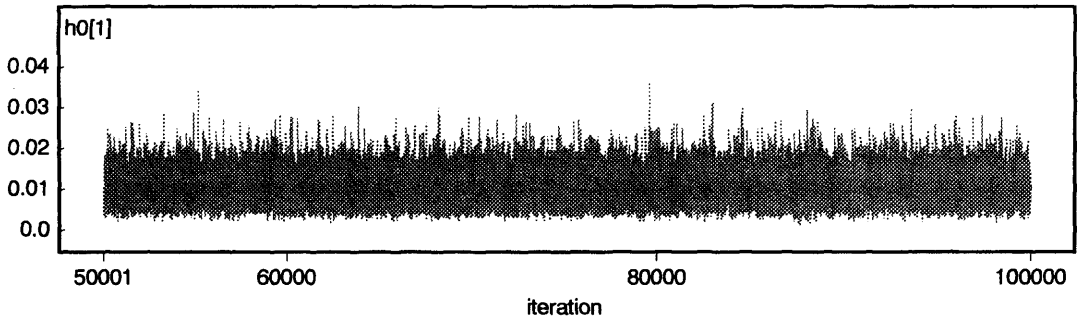

COV[2,2]
3.0
2.0
1.0
0.0
50001
60000
80000
100000
iteration

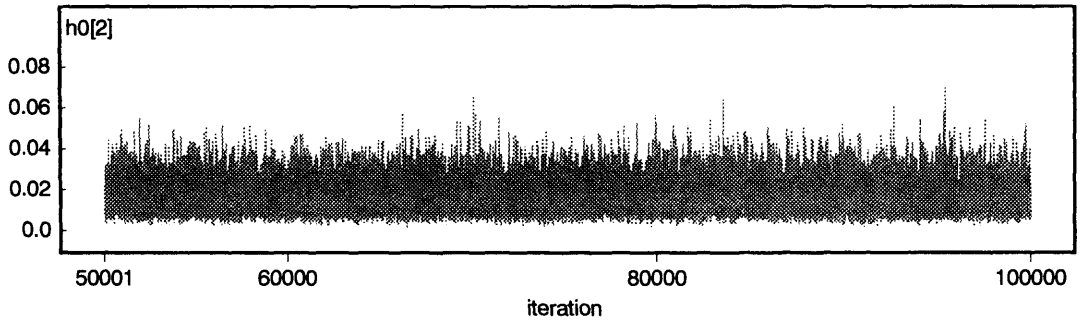

delta1
20.0
10.0
0.0
-10.0
50001
60000
80000
100000
iteration

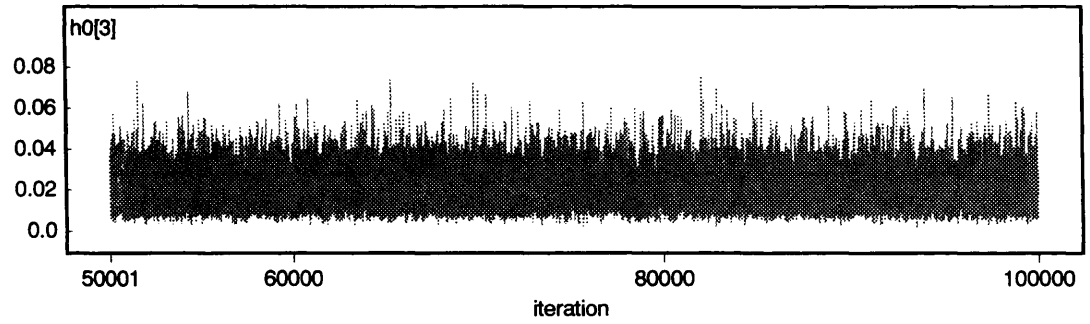

delta2
2.0
1.0
0.0
-1.0
50001
60000
80000
100000
iteration

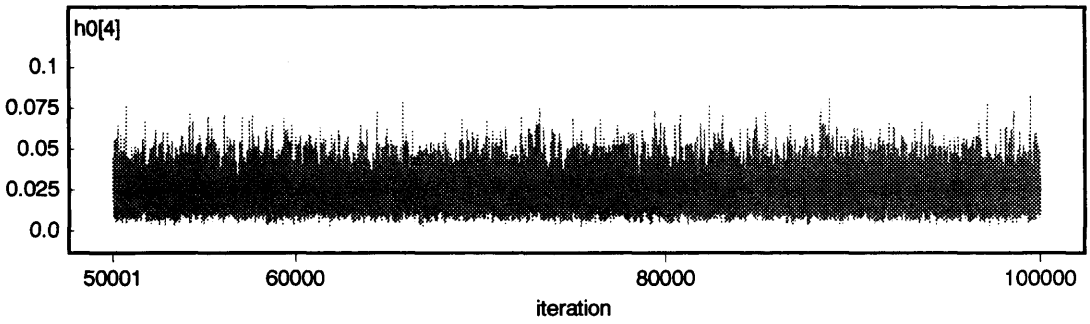

gamma1
2.0
1.0
0.0
-1.0
-2.0
-3.0
50001
60000
80000
100000
iteration

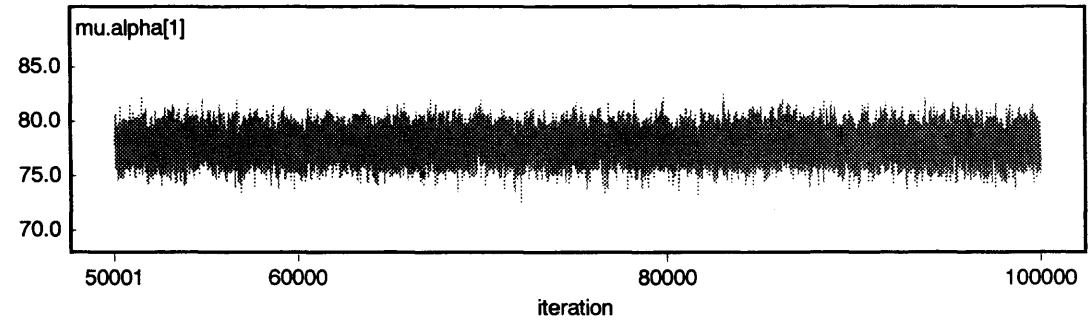

gamma2
0.05
-1.38778E-17
-0.05
-0.1
-0.15
-0.2
50001
60000
80000
100000
iteration

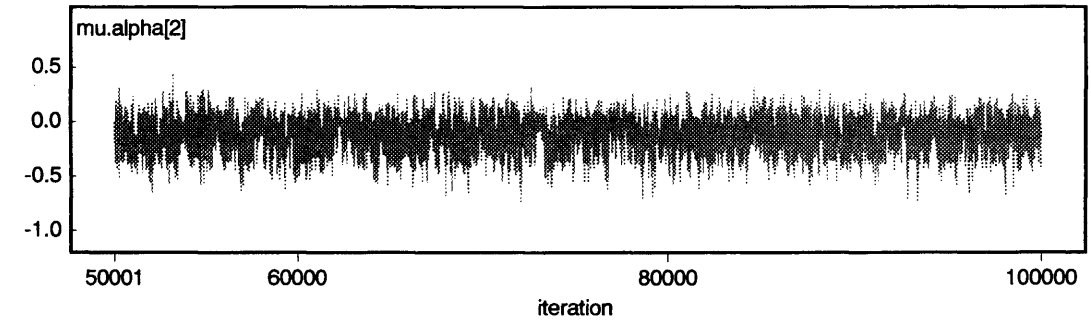

h0[1]
0.04
0.03
0.02
0.01
0.0
50001
60000
80000
100000
iteration

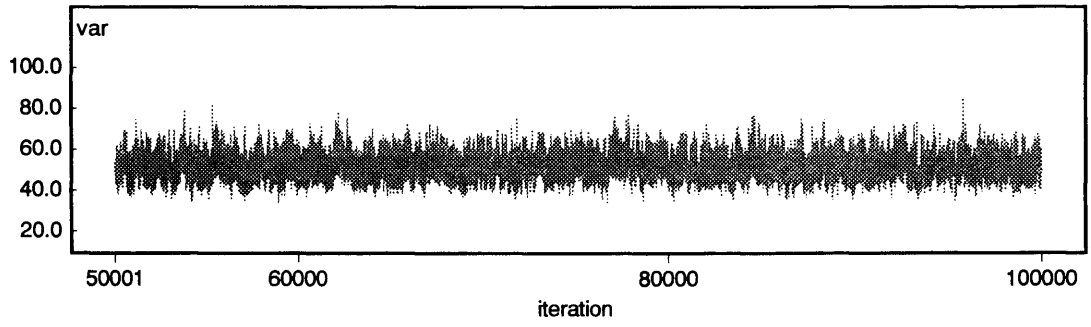

h0[2]
0.08
0.06
0.04
0.02
0.0
50001
60000
80000
100000
iteration

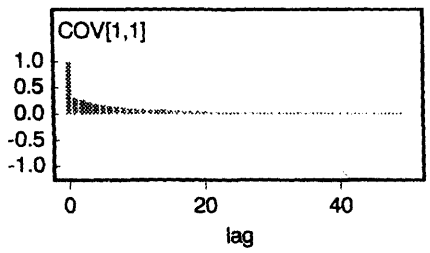

h0[3]
0.08
0.06
0.04
0.02
0.0
50001
60000
80000
100000
iteration

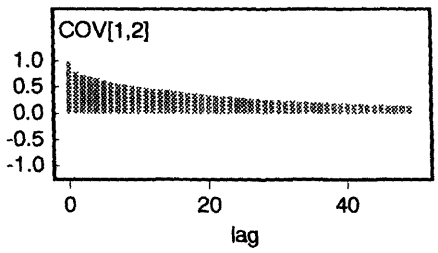

h0[4]
0.1
0.075
0.05
0.025
0.0
50001
60000
80000
100000
iteration

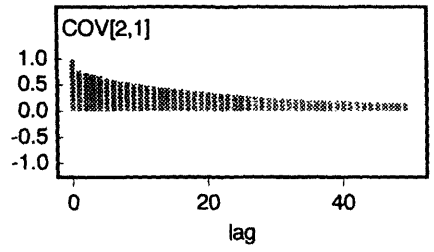

mu.alpha[1]
85.0
80.0
75.0
70.0
50001
60000
80000
100000
iteration

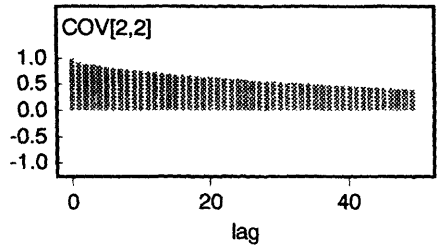

mu.alpha[2]
0.5
0.0
-0.5
-1.0
50001
60000
80000
100000
iteration

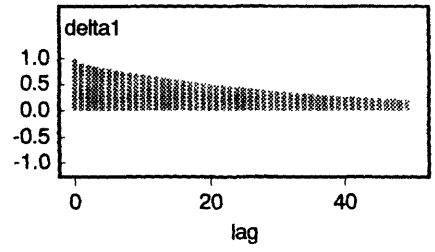

var
100.0
80.0
60.0
40.0
20.0
50001
60000
80000
100000
iteration

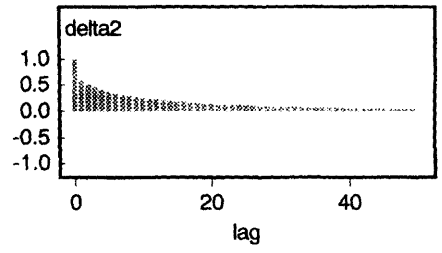

COV[1,1]
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag

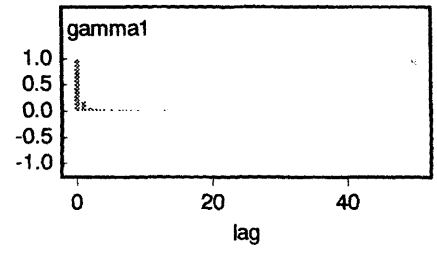

COV[1,2]
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag

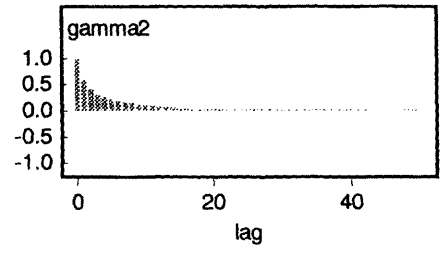

COV[2,1]
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag

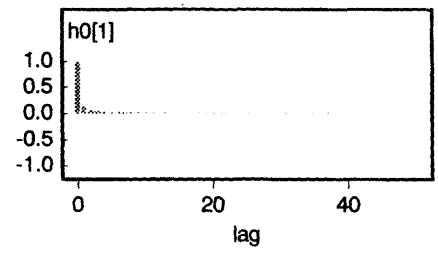

COV[2,2]
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag

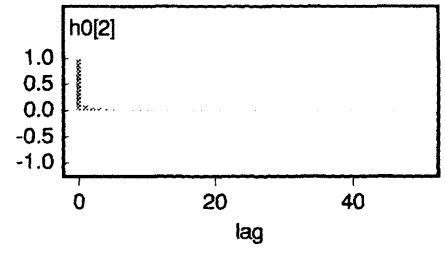

delta1
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag

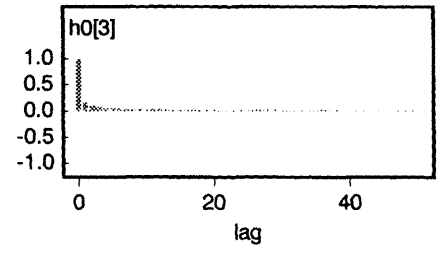

delta2
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag

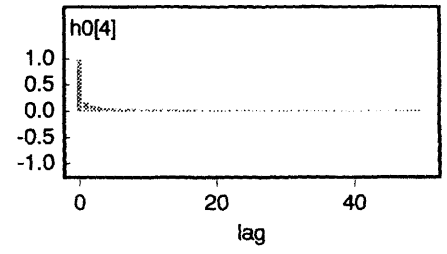

gamma1
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag

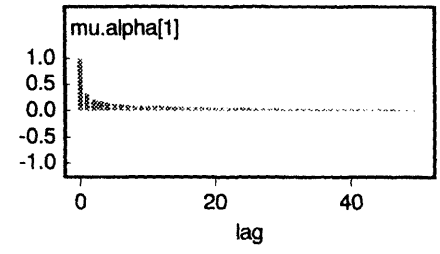

gamma2
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag

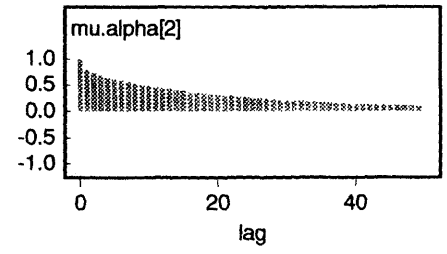

h0[1]
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag

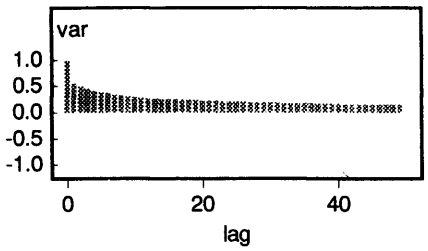

h0[2]
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag

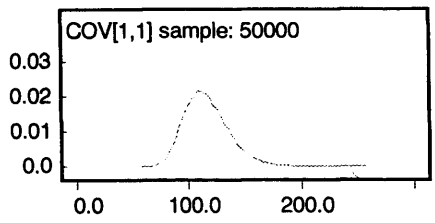

h0[3]
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag

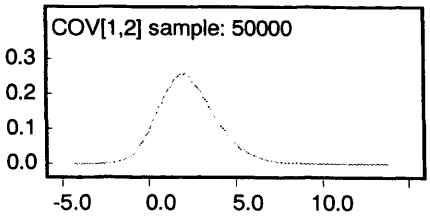

h0[4]
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag

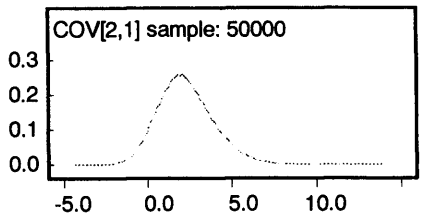

mu.alpha[1]
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag

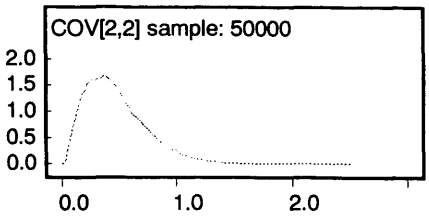

mu.alpha[2]
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag

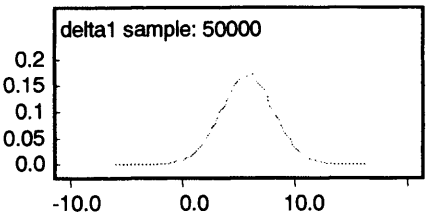

var
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag

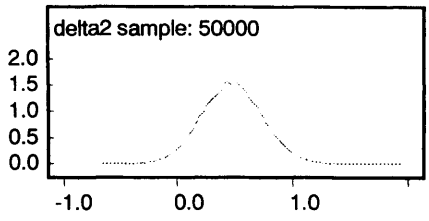

COV[1,1] sample: 50000
0.03
0.02
0.01
0.0
0.0
100.0
200.0

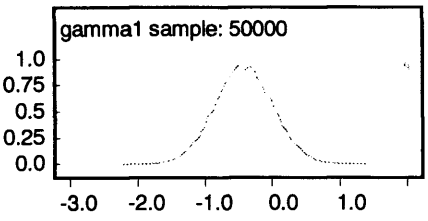

COV[1,2] sample: 50000
0.3
0.2
0.1
0.0
-5.0
0.0
5.0
10.0

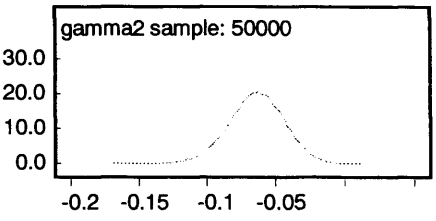

COV[2,1] sample: 50000
0.3
0.2
0.1
0.0
-5.0
0.0
5.0
10.0

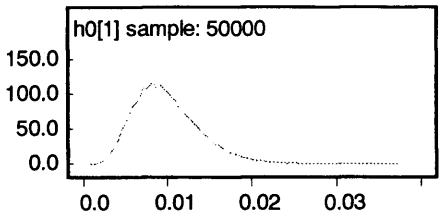

COV[2,2] sample: 50000
2.0
1.5
1.0
0.5
0.0
0.0
1.0
2.0

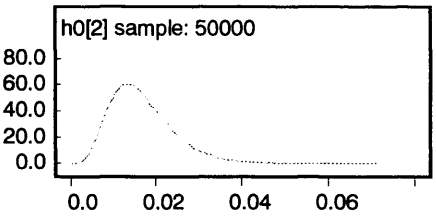

delta1 sample: 50000
0.2
0.15
0.1
0.05
0.0
-10.0
0.0
10.0

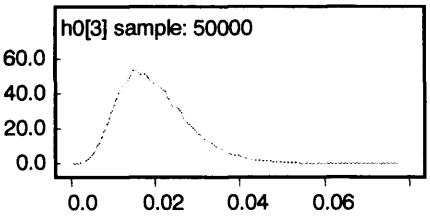

delta2 sample: 50000
2.0
1.5
1.0
0.5
0.0
-1.0
0.0
1.0

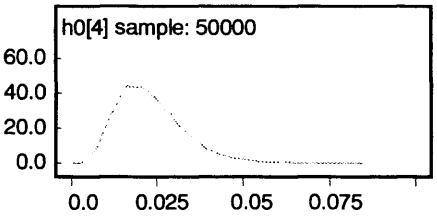

gamma1 sample: 50000
1.0
0.75
0.5
0.25
0.0
-3.0
-2.0
-1.0
0.0
1.0

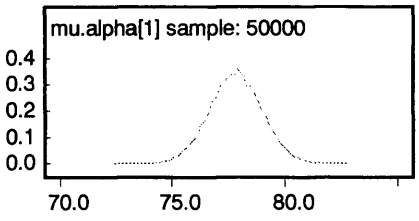

gamma2 sample: 50000
30.0
20.0
10.0
0.0
-0.2
-0.15
-0.1
-0.05

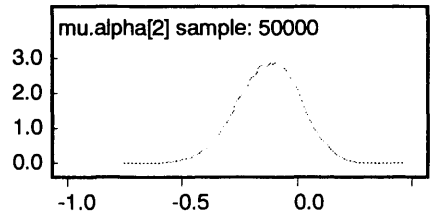

h0[1] sample: 50000
150.0
100.0
50.0
0.0
0.0
0.01
0.02
0.03

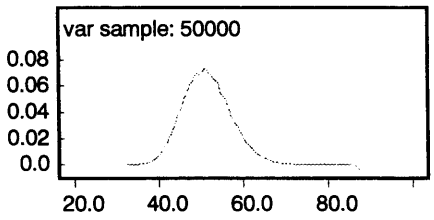

h0[2] sample: 50000
80.0
60.0
40.0
20.0
0.0
0.0
0.02
0.04
0.06


h0[3] sample: 50000
60.0
40.0
20.0
0.0
0.0
0.02
0.04
0.06


h0[4] sample: 50000
60.0
40.0
20.0
0.0
0.0
0.025
0.05
0.075


mu.alpha[1] sample: 50000
0.4
0.3
0.2
0.1
0.0
70.0
75.0
80.0


mu.alpha[2] sample: 50000
3.0
2.0
1.0
0.0
-1.0
-0.5
0.0

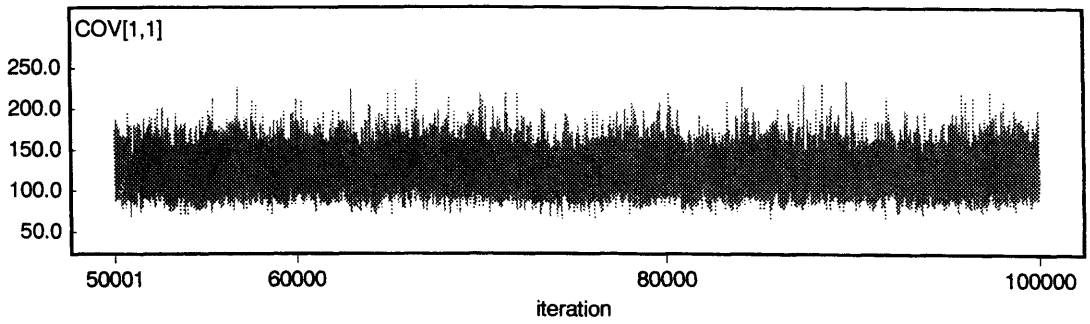

var sample: 50000
0.08
0.06
0.04
0.02
0.0
20.0
40.0
60.0
80.0

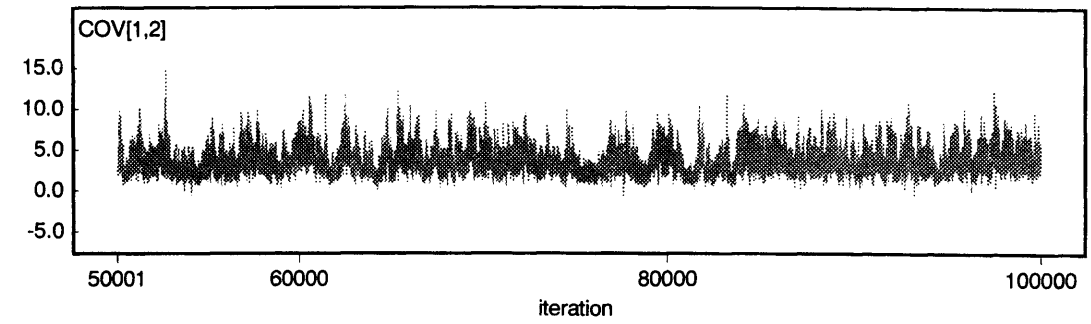
APPENDIX III C

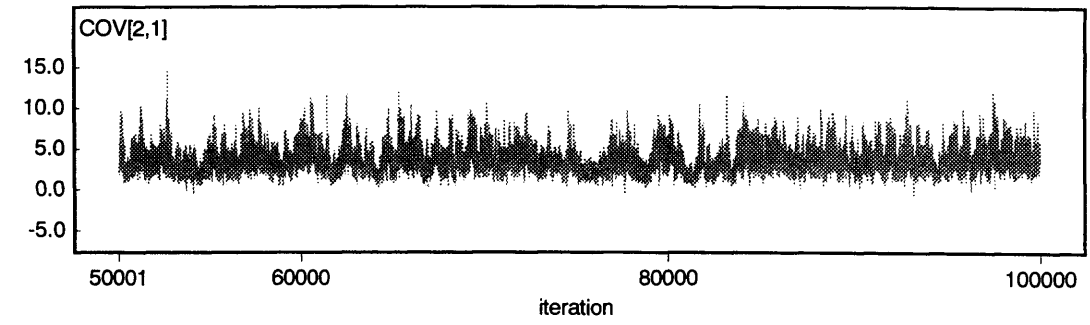
OUTPUT FROM JOINT MODELLING OF MIC DATA IN WINBUGS

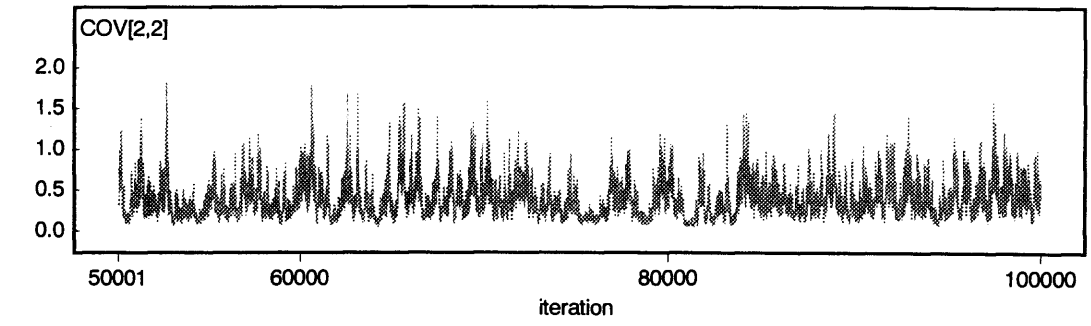
Linear random effects model for *GQS* over time with treatment and treatment by time covariates and Wishart prior distribution on precision matrix

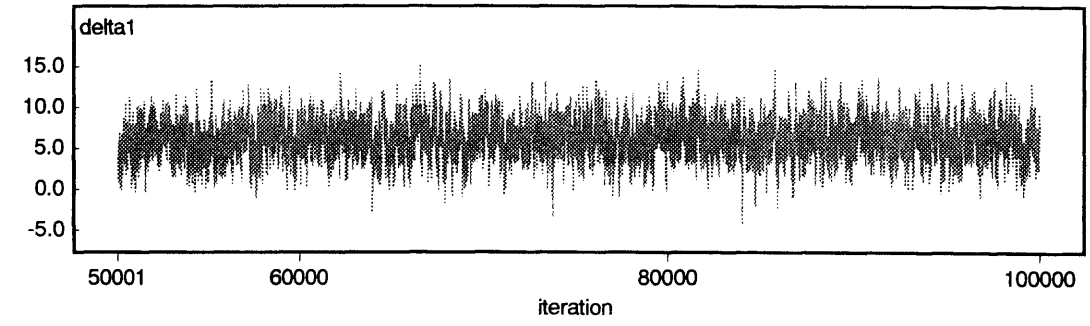
Piecewise exponential model for survival with treatment and *slope of GQS over time* as covariates

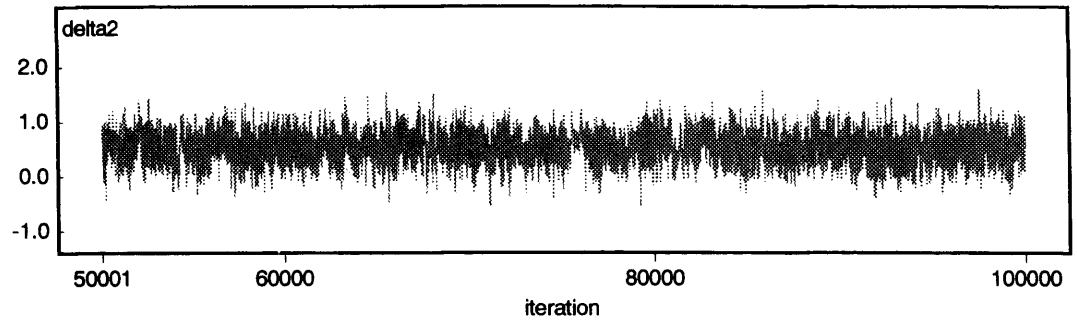

COV[1,1]
250.0
200.0
150.0
100.0
50.0
50001
60000
80000
100000
iteration

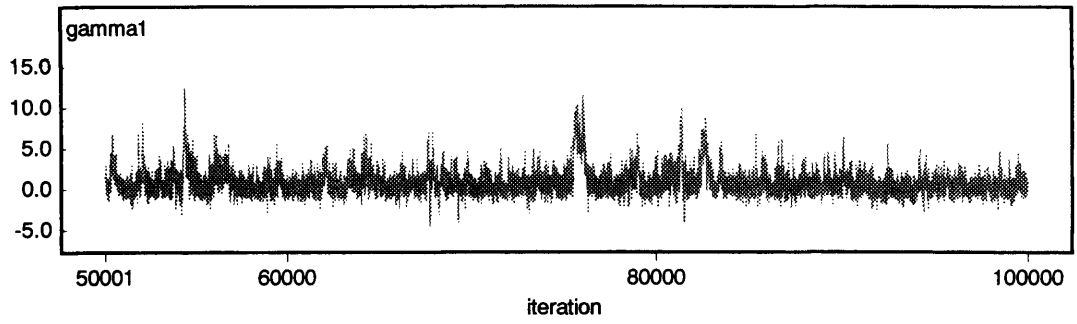

COV[1,2]
15.0
10.0
5.0
0.0
-5.0
50001
60000
80000
100000
iteration

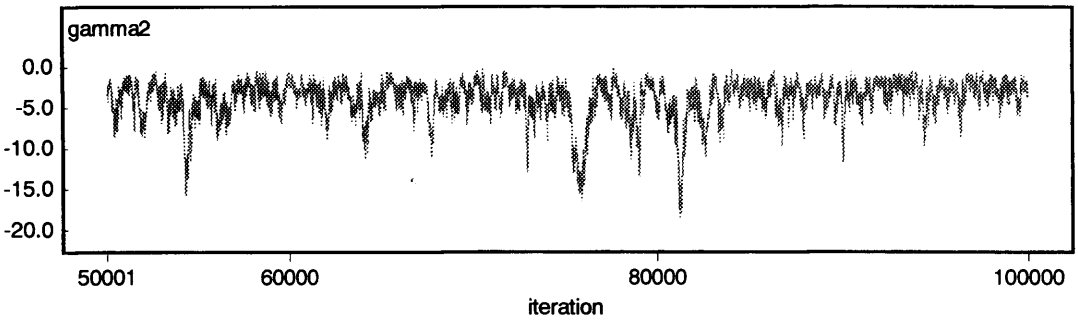

COV[2,1]
15.0
10.0
5.0
0.0
-5.0
50001
60000
80000
100000
iteration

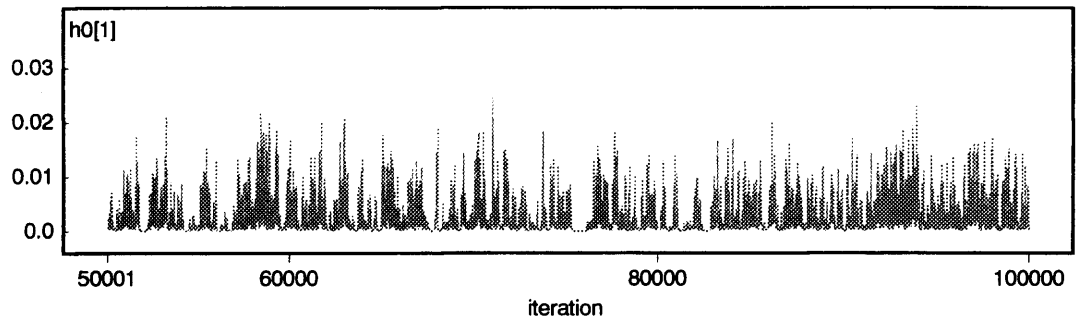

COV[2,2]
2.0
1.5
1.0
0.5
0.0
50001
60000
80000
100000
iteration

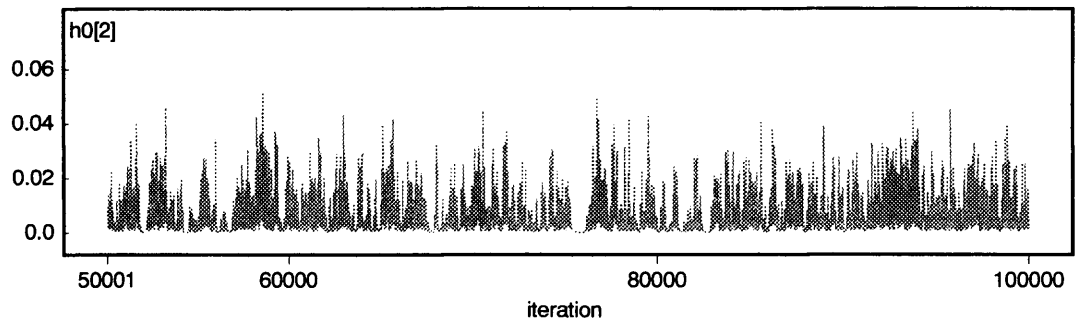

delta1
15.0
10.0
5.0
0.0
-5.0
50001
60000
80000
100000
iteration

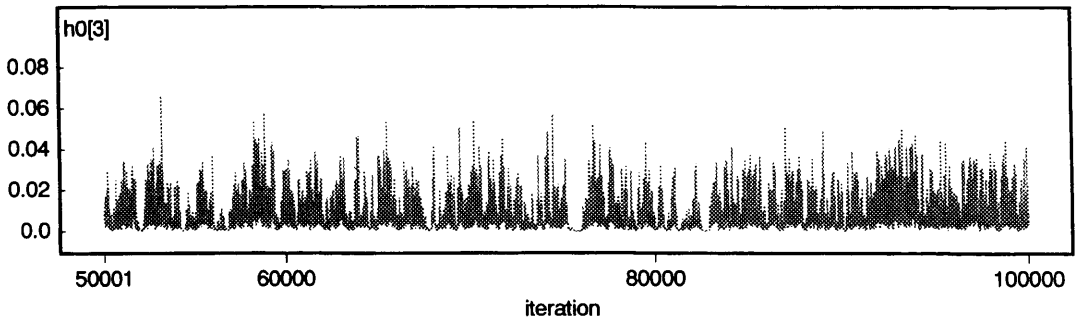

delta2
2.0
1.0
0.0
-1.0
50001
60000
80000
100000
iteration

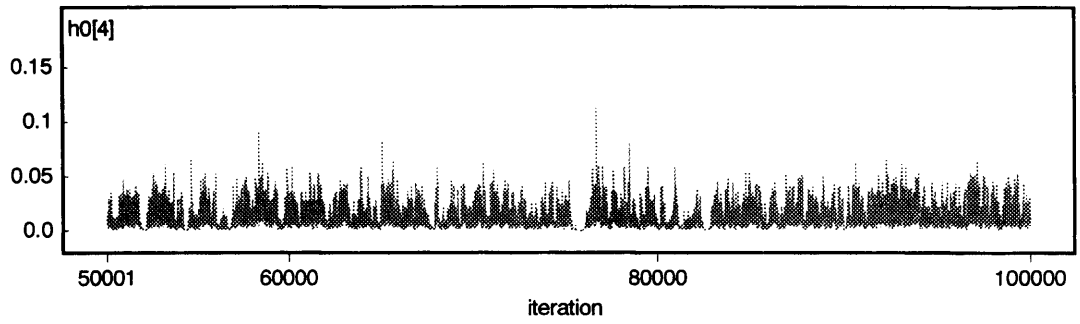

gamma1
15.0
10.0
5.0
0.0
-5.0
50001
60000
80000
100000
iteration

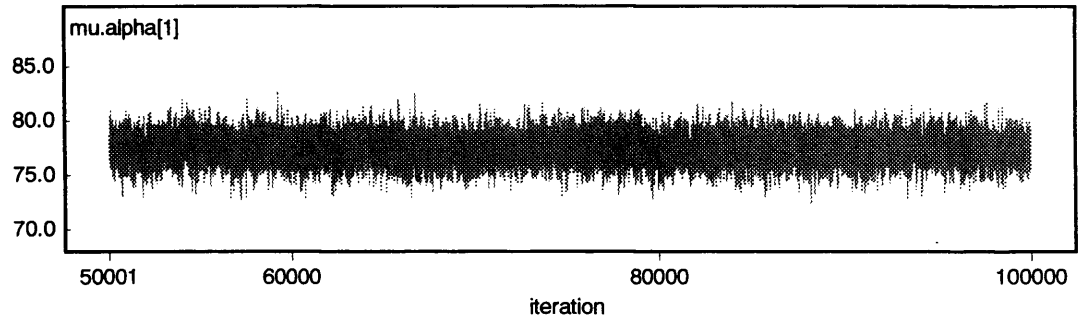

gamma2
0.0
-5.0
-10.0
-15.0
-20.0
50001
60000
80000
100000
iteration

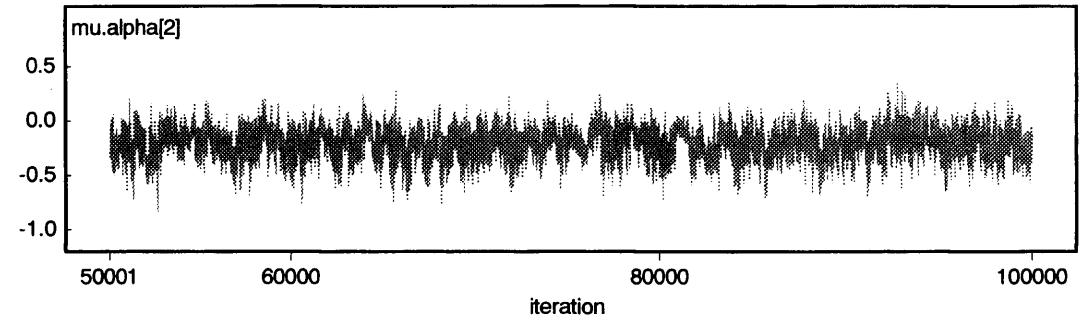

h0[1]
0.03
0.02
0.01
0.0
50001
60000
80000
100000
iteration

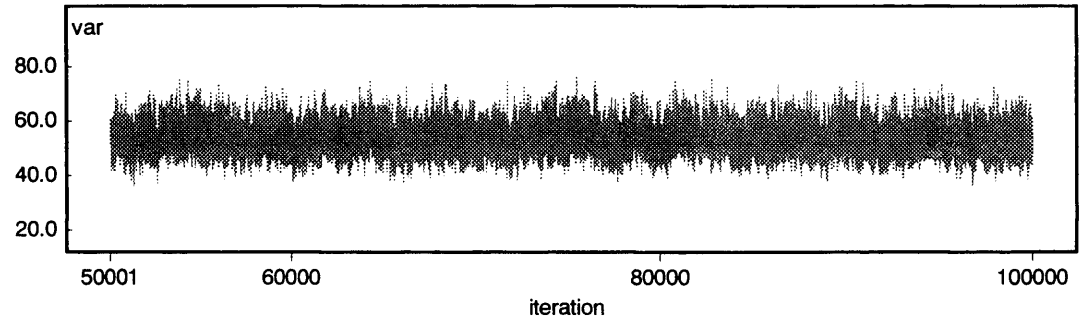

h0[2]
0.06
0.04
0.02
0.0
50001
60000
80000
100000
iteration

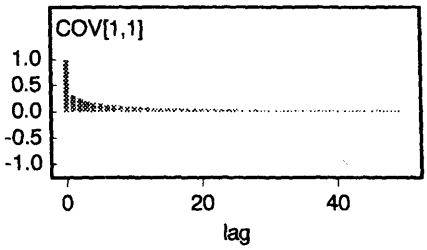

h0[3]
0.08
0.06
0.04
0.02
0.0
50001
60000
80000
100000
iteration

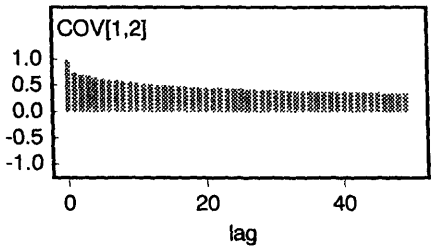

h0[4]
0.15
0.1
0.05
0.0
50001
60000
80000
100000
iteration

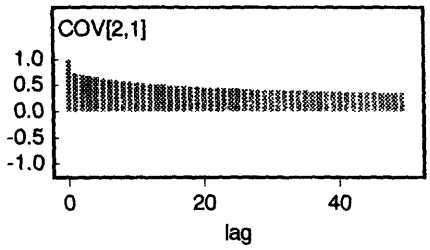

mu.alpha[1]
85.0
80.0
75.0
70.0
50001
60000
80000
100000
iteration

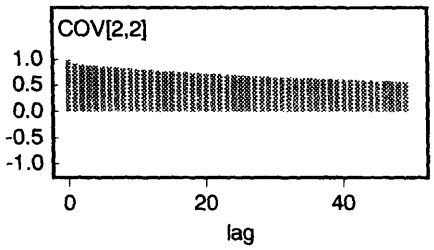

mu.alpha[2]
0.5
0.0
-0.5
-1.0
50001
60000
80000
100000
iteration

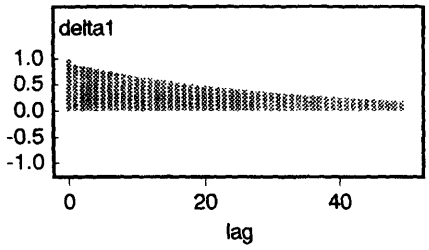

var
80.0
60.0
40.0
20.0
50001
60000
80000
100000
iteration

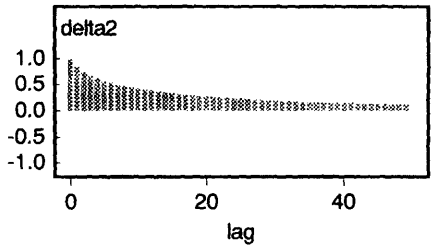

COV[1,1]
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag

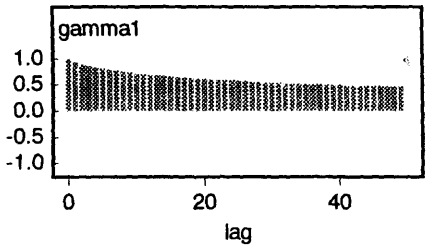

COV[1,2]
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag

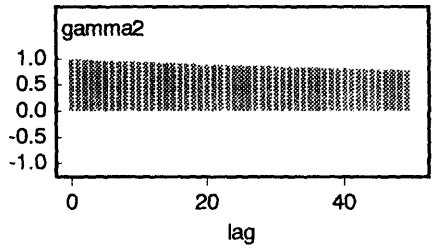

COV[2,1]
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag

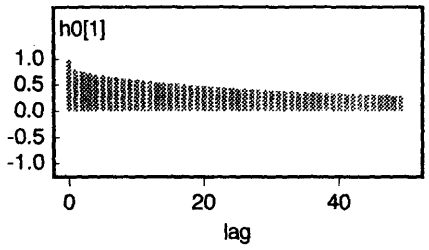

COV[2,2]
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag

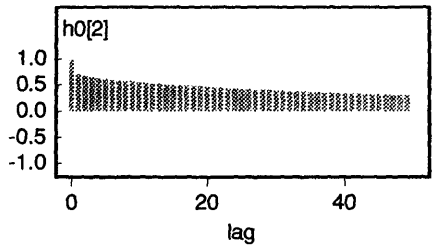

delta1
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag

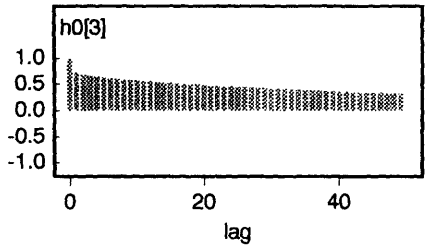

delta2
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag

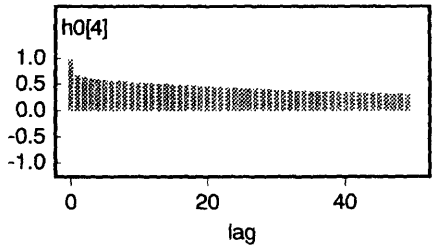

gamma1
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag

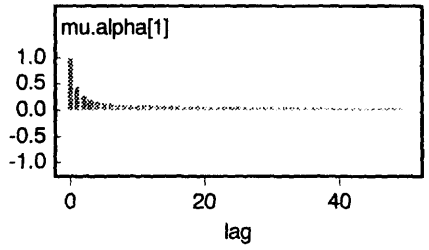

gamma2
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag

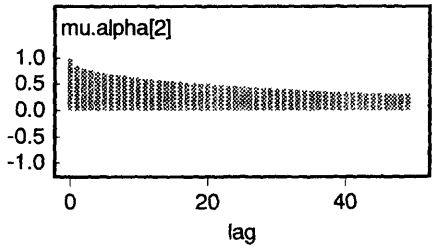

h0[1]
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag

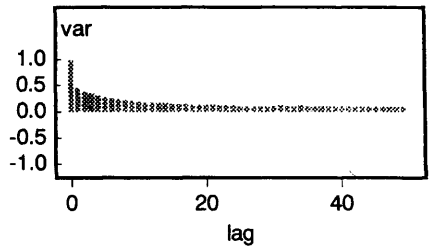

h0[2]
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag

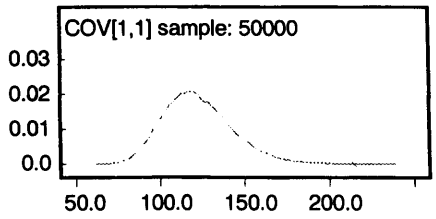

h0[3]
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag

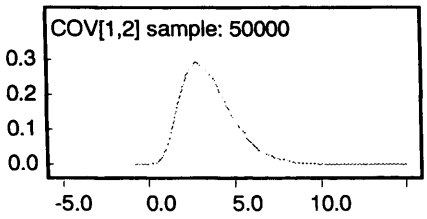

h0[4]
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag

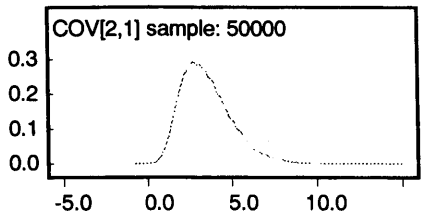

mu.alpha[1]
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag

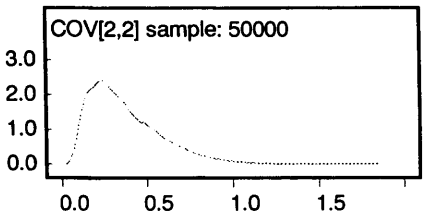

mu.alpha[2]
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag

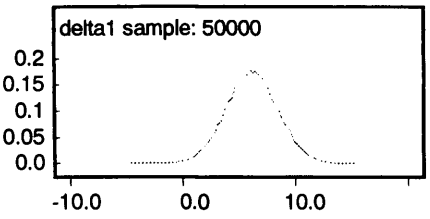

var
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag

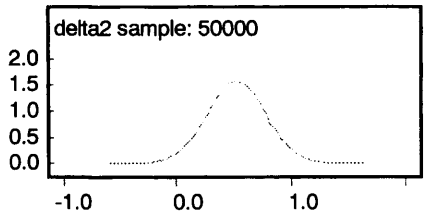

COV[1,1] sample: 50000
0.03
0.02
0.01
0.0
50.0
100.0
150.0
200.0

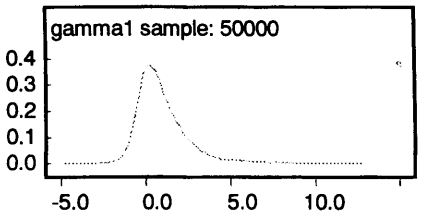

COV[1,2] sample: 50000
0.3
0.2
0.1
0.0
-5.0
0.0
5.0
10.0

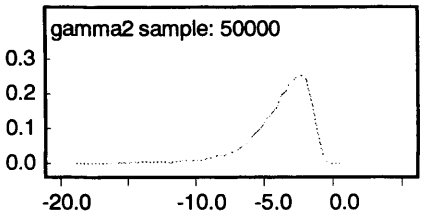

COV[2,1] sample: 50000
0.3
0.2
0.1
0.0
-5.0
0.0
5.0
10.0

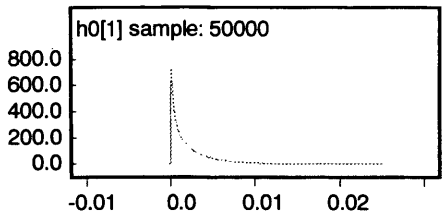

COV[2,2] sample: 50000
3.0
2.0
1.0
0.0
0.0
0.5
1.0
1.5

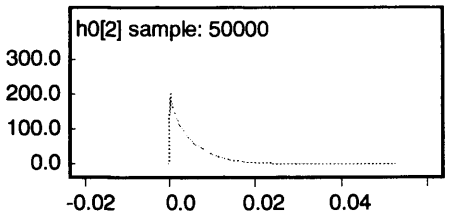

delta1 sample: 50000
0.2
0.15
0.1
0.05
0.0
-10.0
0.0
10.0

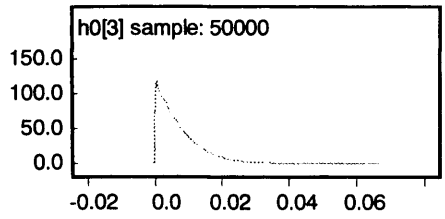

delta2 sample: 50000
2.0
1.5
1.0
0.5
0.0
-1.0
0.0
1.0

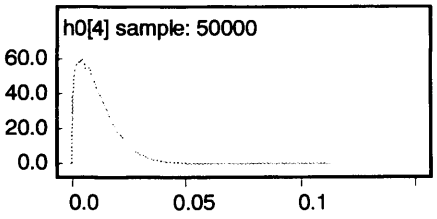

gamma1 sample: 50000
0.4
0.3
0.2
0.1
0.0
-5.0
0.0
5.0
10.0

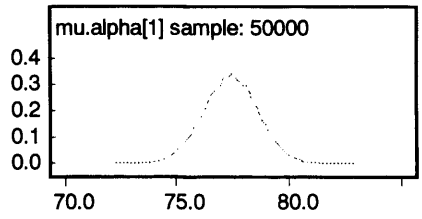

gamma2 sample: 50000
0.3
0.2
0.1
0.0
-20.0
-10.0
-5.0
0.0

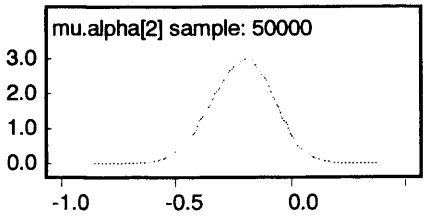

h0[1] sample: 50000
800.0
600.0
400.0
200.0
0.0
-0.01
0.0
0.01
0.02

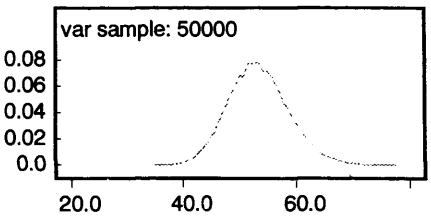

h0[2] sample: 50000
300.0
200.0
100.0
0.0
-0.02
0.0
0.02
0.04


h0[3] sample: 50000
150.0
100.0
50.0
0.0
-0.02
0.0
0.02
0.04
0.06


h0[4] sample: 50000
60.0
40.0
20.0
0.0
0.0
0.05
0.1


mu.alpha[1] sample: 50000
0.4
0.3
0.2
0.1
0.0
70.0
75.0
80.0


mu.alpha[2] sample: 50000
3.0
2.0
1.0
0.0
-1.0
-0.5
0.0

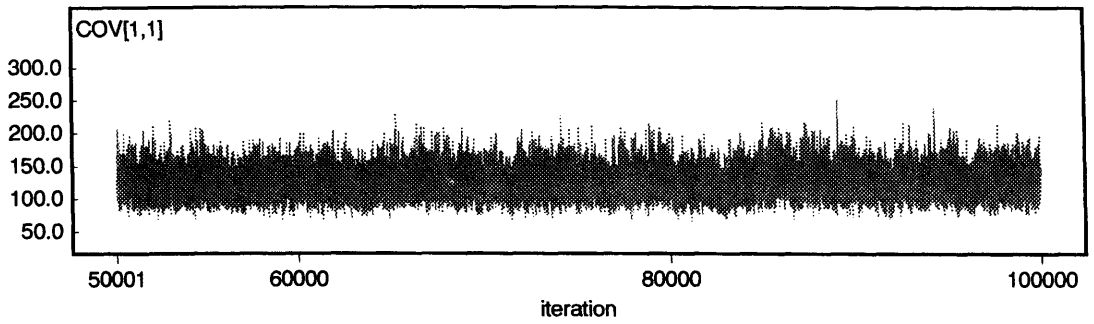

var sample: 50000
0.08
0.06
0.04
0.02
0.0
20.0
40.0
60.0

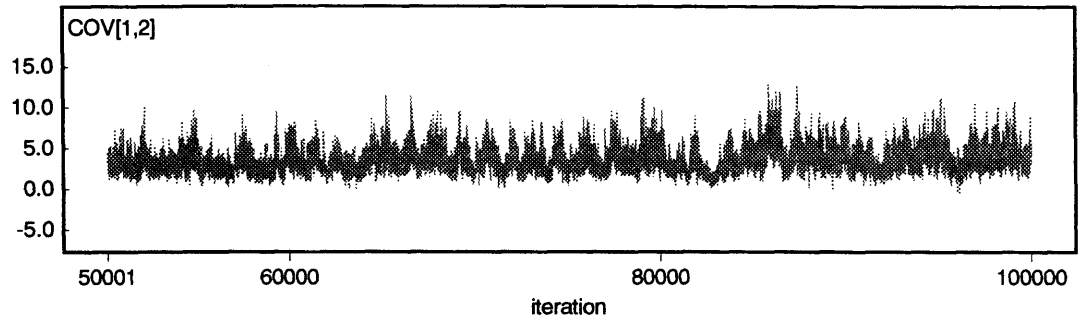
APPENDIX III D

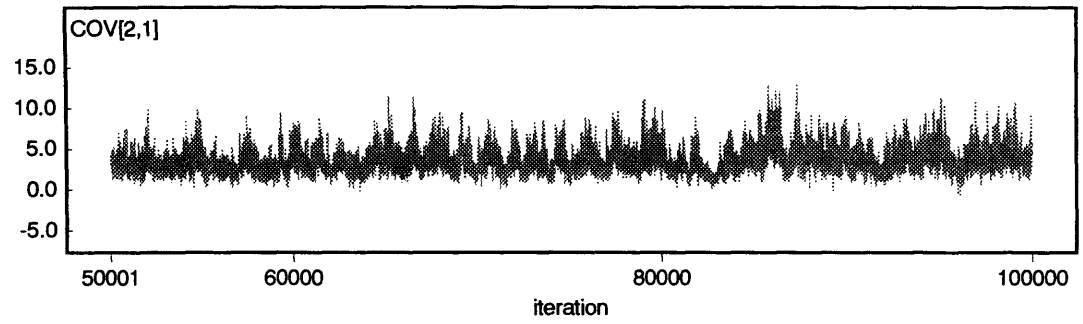
OUTPUT FROM JOINT MODELLING OF MIC DATA IN WINBUGS

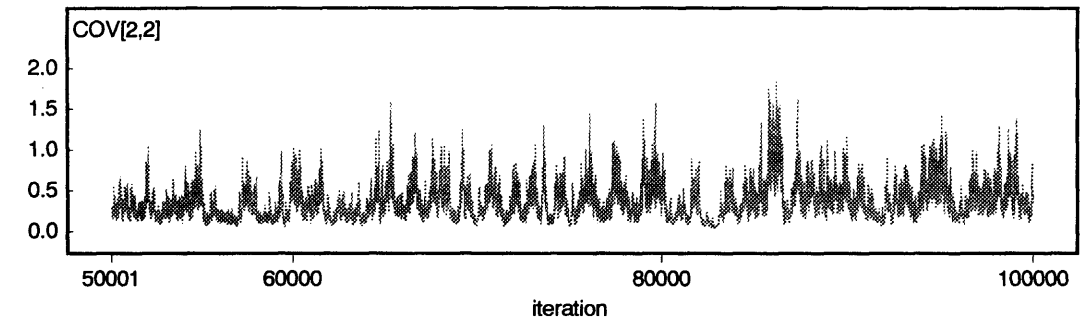
Linear random effects model for *GQS* over time with treatment and treatment by time covariates and Wishart prior distribution on precision matrix

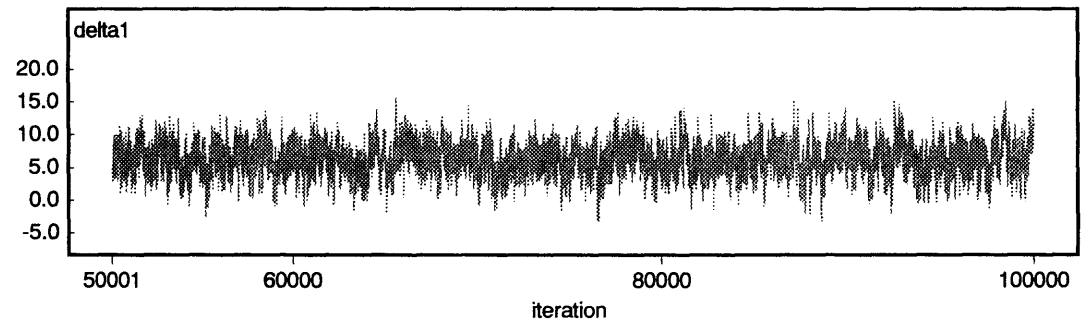
Piecewise exponential model for survival with treatment and *slope of GQS over time* *and interaction between treatment and slope* as covariates

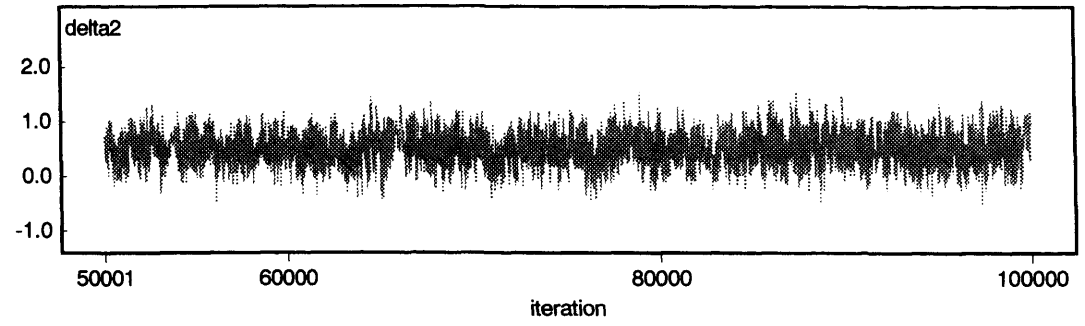

COV[1,1]
300.0
250.0
200.0
150.0
100.0
50.0
50001
60000
80000
100000
iteration

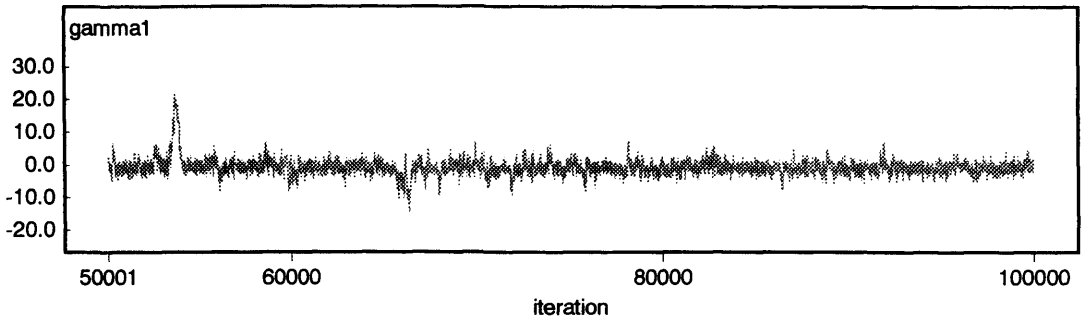

COV[1,2]
15.0
10.0
5.0
0.0
-5.0
50001
60000
80000
100000
iteration

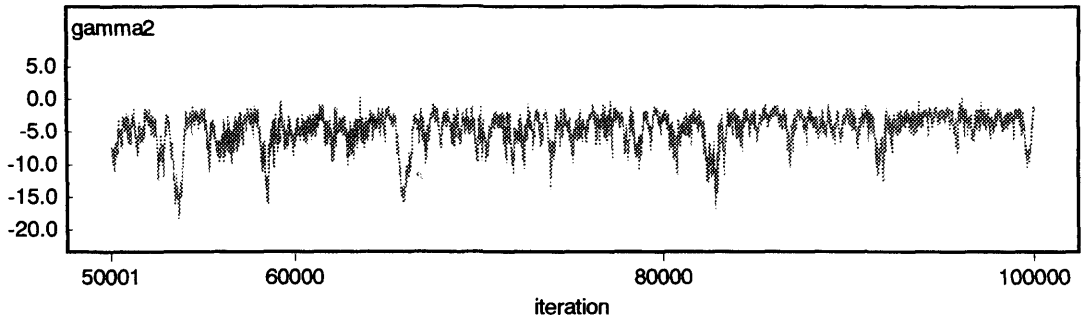

COV[2,1]
15.0
10.0
5.0
0.0
-5.0
50001
60000
80000
100000
iteration

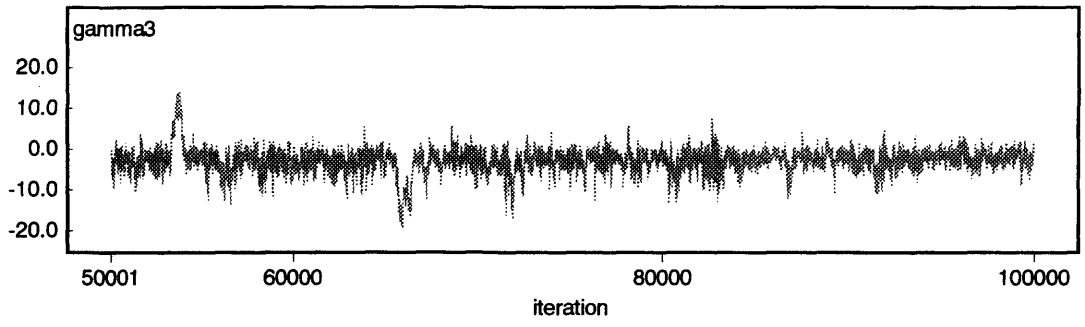

COV[2,2]
2.0
1.5
1.0
0.5
0.0
50001
60000
80000
100000
iteration

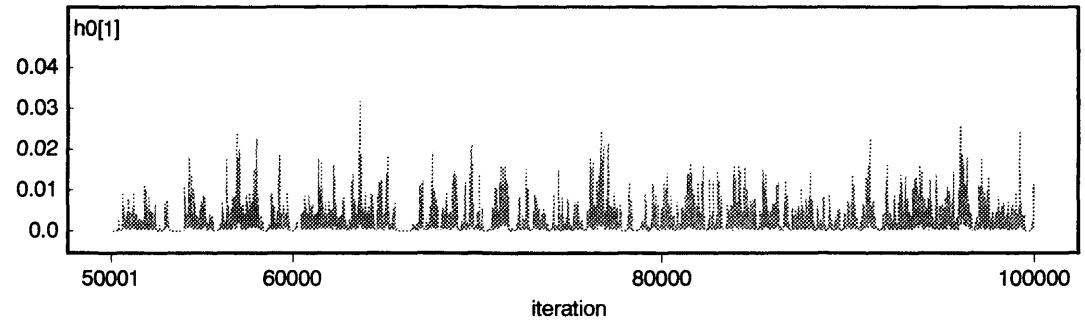

delta1
20.0
15.0
10.0
5.0
0.0
-5.0
50001
60000
80000
100000
iteration

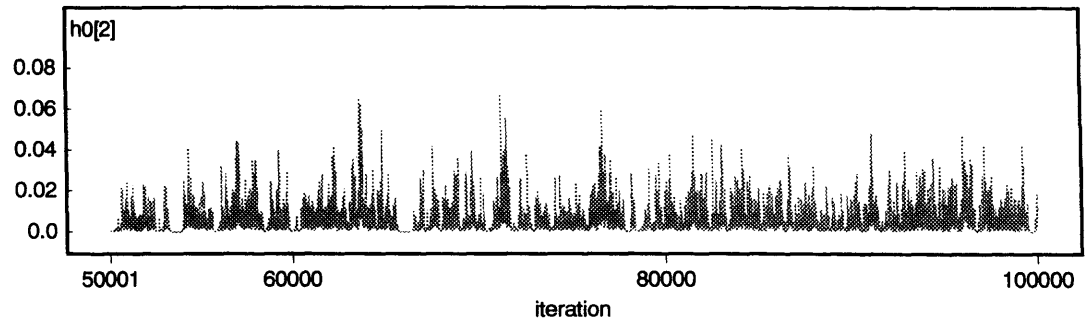

delta2
2.0
1.0
0.0
-1.0
50001
60000
80000
100000
iteration

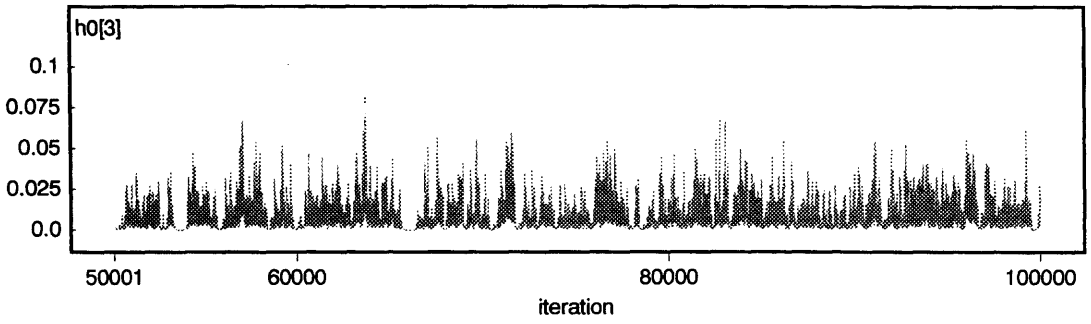

gamma1
30.0
20.0
10.0
0.0
-10.0
-20.0
50001
60000
80000
100000
iteration

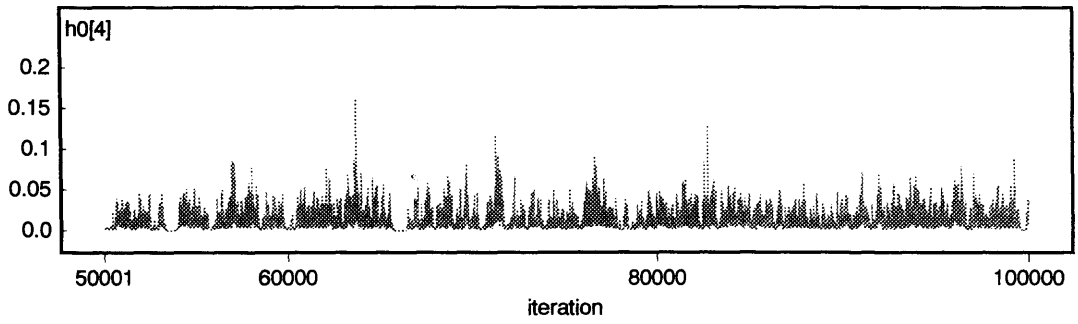

gamma2
5.0
0.0
-5.0
-10.0
-15.0
-20.0
50001
60000
80000
100000
iteration

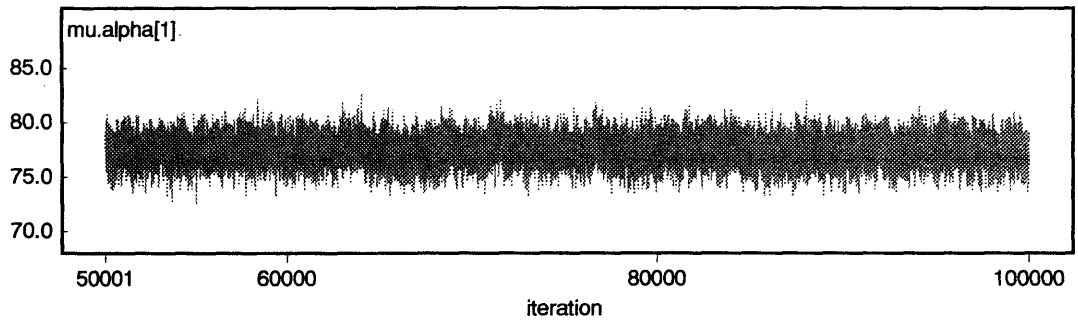

gamma3
20.0
10.0
0.0
-10.0
-20.0
50001
60000
80000
100000
iteration

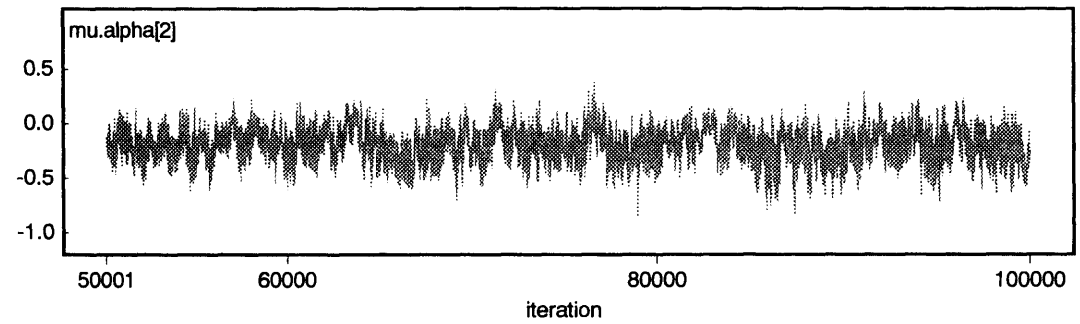

h0[1]
0.04
0.03
0.02
0.01
0.0
50001
60000
80000
100000
iteration

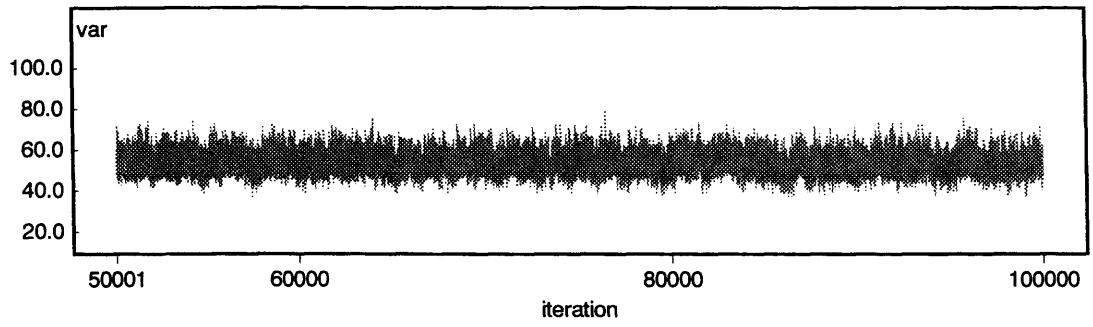

h0[2]
0.08
0.06
0.04
0.02
0.0
50001
60000
80000
100000
iteration

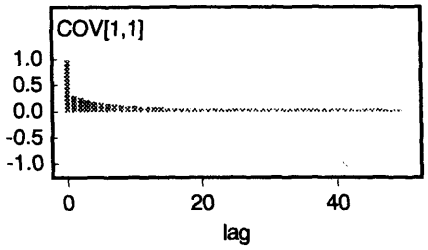

h0[3]
0.1
0.075
0.05
0.025
0.0
50001
60000
80000
100000
iteration

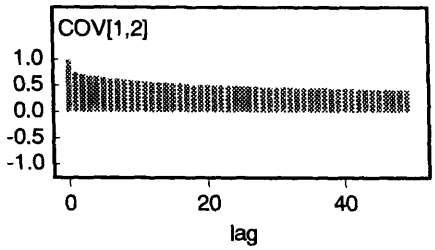

h0[4]
0.2
0.15
0.1
0.05
0.0
50001
60000
80000
100000
iteration

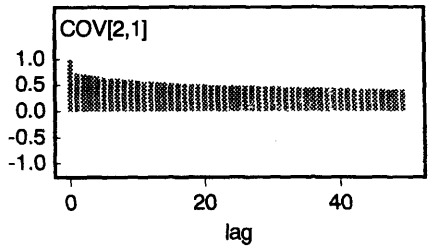

mu.alpha[1]
85.0
80.0
75.0
70.0
50001
60000
80000
100000
iteration

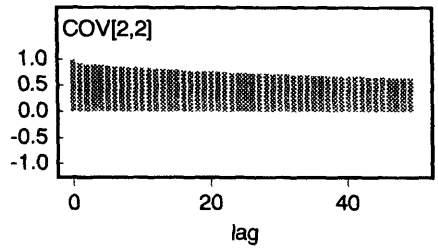

mu.alpha[2]
0.5
0.0
-0.5
-1.0
50001
60000
80000
100000
iteration

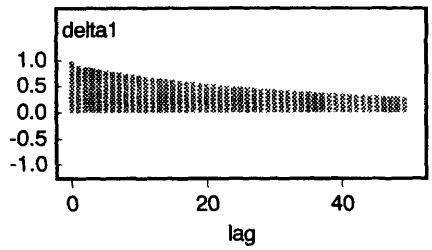

var
100.0
80.0
60.0
40.0
20.0
50001
60000
80000
100000
iteration

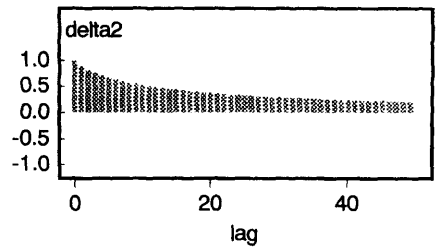

COV[1,1]
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag

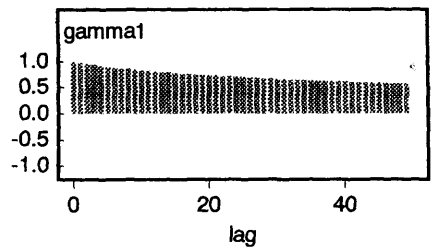

COV[1,2]
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag

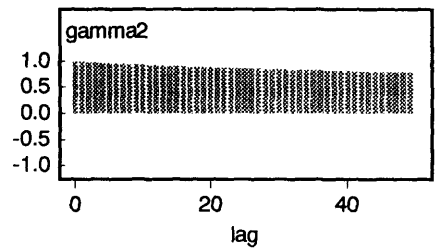

COV[2,1]
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag

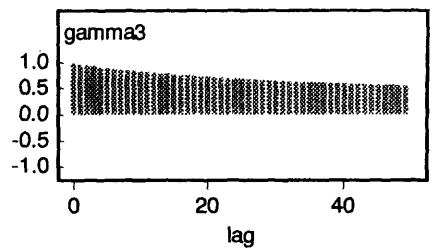

COV[2,2]
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag

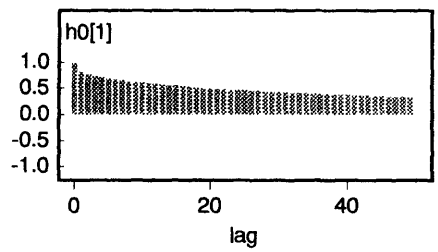

delta1
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag

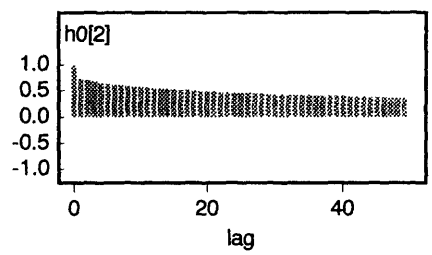

delta2
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag

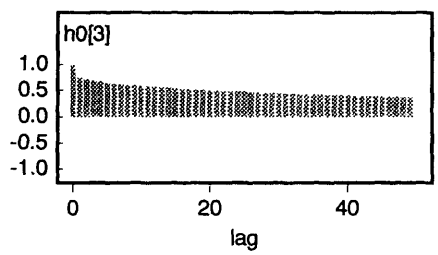

gamma1
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag

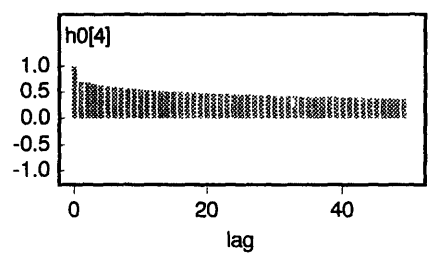

gamma2
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag

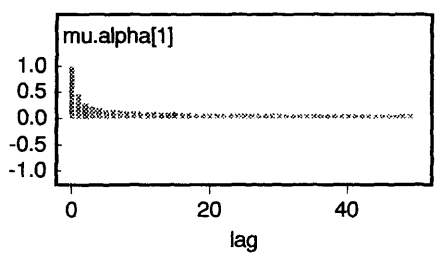

gamma3
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag

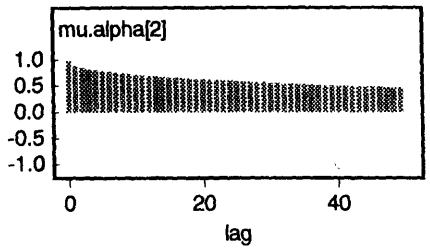

h0[1]
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag

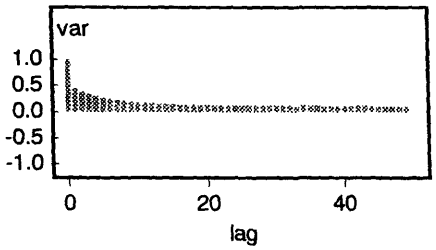

h0[2]
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag

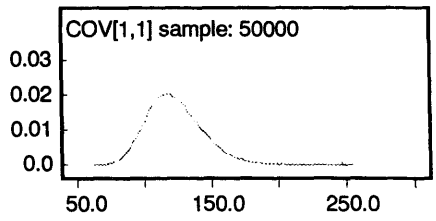

h0[3]
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag

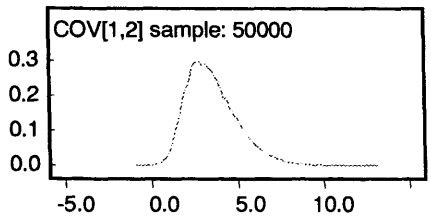

h0[4]
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag

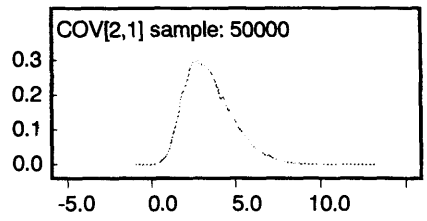

mu.alpha[1]
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag

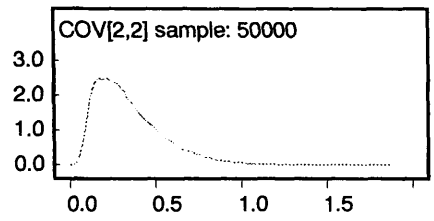

mu.alpha[2]
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag

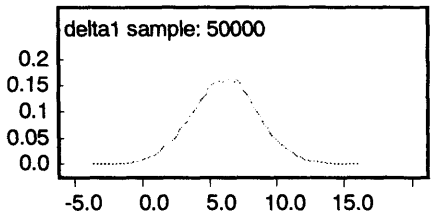

var
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag

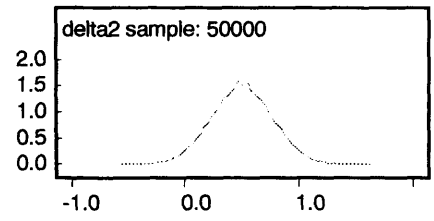

COV[1,1] sample: 50000
0.03
0.02
0.01
0.0
50.0
150.0
250.0

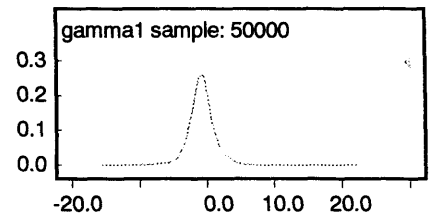

COV[1,2] sample: 50000
0.3
0.2
0.1
0.0
-5.0
0.0
5.0
10.0

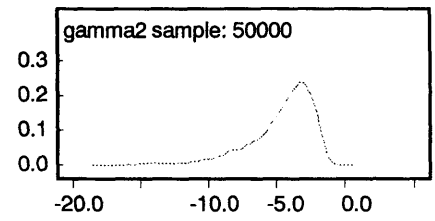

COV[2,1] sample: 50000
0.3
0.2
0.1
0.0
-5.0
0.0
5.0
10.0

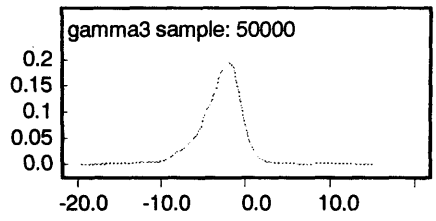

COV[2,2] sample: 50000
3.0
2.0
1.0
0.0
0.0
0.5
1.0
1.5

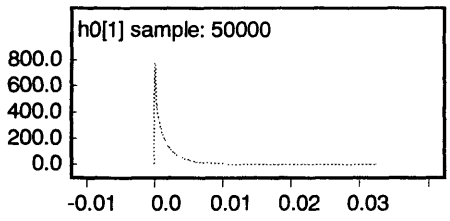

delta1 sample: 50000
0.2
0.15
0.1
0.05
0.0
-5.0
0.0
5.0
10.0
15.0

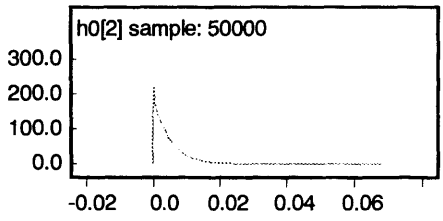

delta2 sample: 50000
2.0
1.5
1.0
0.5
0.0
-1.0
0.0
1.0

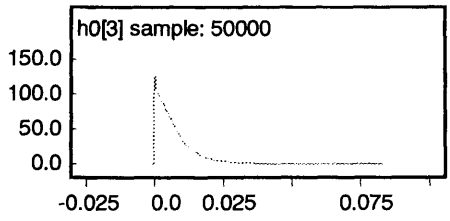

gamma1 sample: 50000
0.3
0.2
0.1
0.0
-20.0
0.0
10.0
20.0

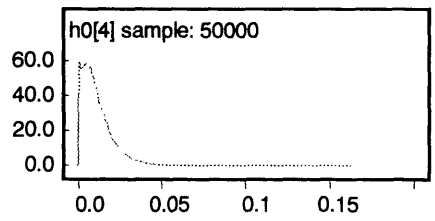

gamma2 sample: 50000
0.3
0.2
0.1
0.0
-20.0
-10.0
-5.0
0.0

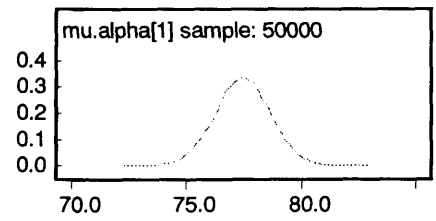

gamma3 sample: 50000
0.2
0.15
0.1
0.05
0.0
-20.0
-10.0
0.0
10.0

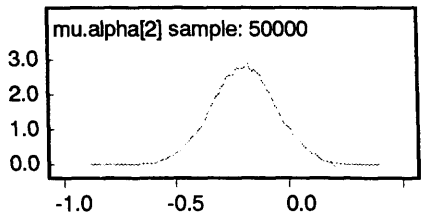

h0[1] sample: 50000
800.0
600.0
400.0
200.0
0.0
-0.01
0.0
0.01
0.02
0.03

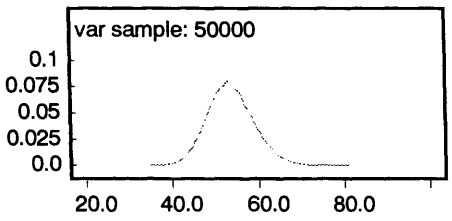

h0[2] sample: 50000
300.0
200.0
100.0
0.0
-0.02
0.0
0.02
0.04
0.06


h0[3] sample: 50000
150.0
100.0
50.0
0.0
-0.025
0.0
0.025
0.075


h0[4] sample: 50000
60.0
40.0
20.0
0.0
0.0
0.05
0.1
0.15


mu.alpha[1] sample: 50000
0.4
0.3
0.2
0.1
0.0
70.0
75.0
80.0


mu.alpha[2] sample: 50000
3.0
2.0
1.0
0.0
-1.0
-0.5
0.0

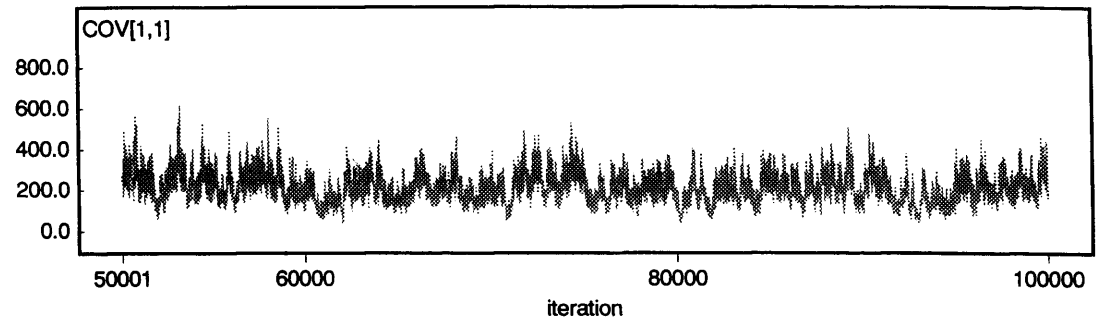

var sample: 50000
0.1
0.075
0.05
0.025
0.0
20.0
40.0
60.0
80.0

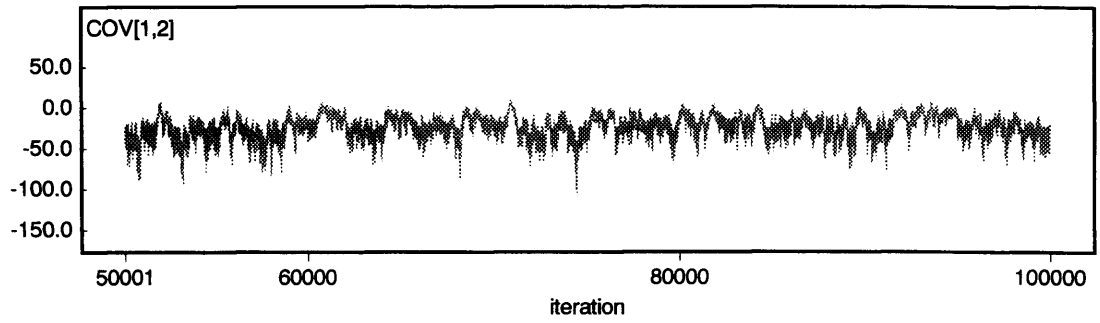
APPENDIX IV

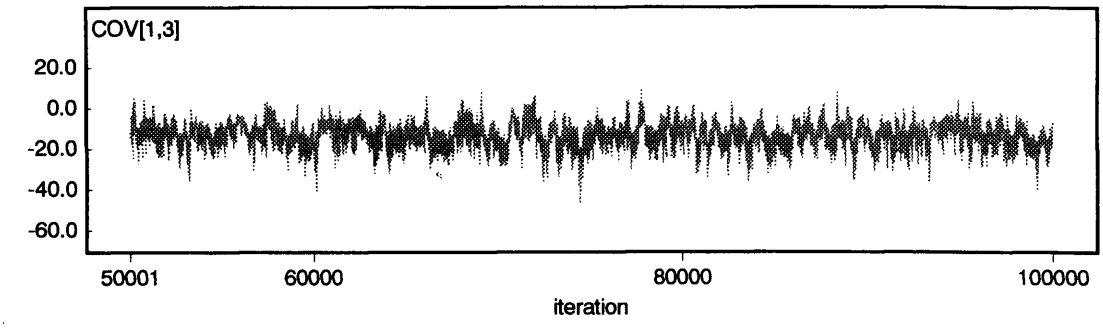
OUTPUT FROM JOINT MODELLING OF ESPAC DATA IN WINBUGS

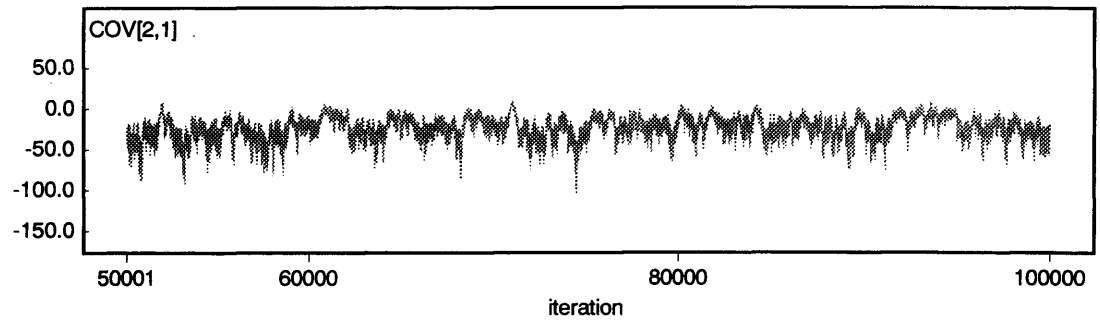
Piecewise linear random effects model for *GHSS* over time with treatment and treatment by time covariates and Wishart prior distribution on precision matrix

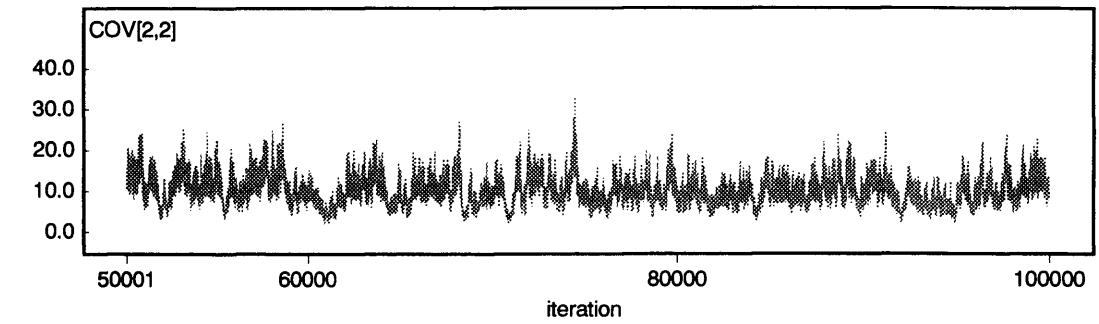
Piecewise exponential model for survival with treatment and *GHSS at start of interval* as covariates

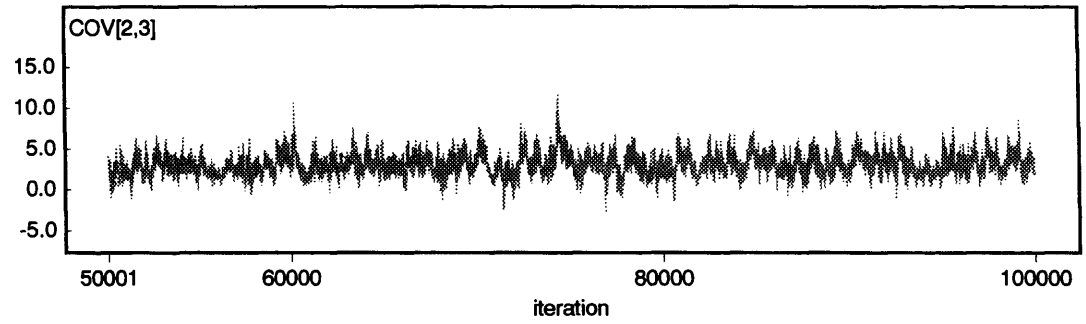

COV[1,1]
800.0
600.0
400.0
200.0
0.0
50001
60000
80000
100000
iteration

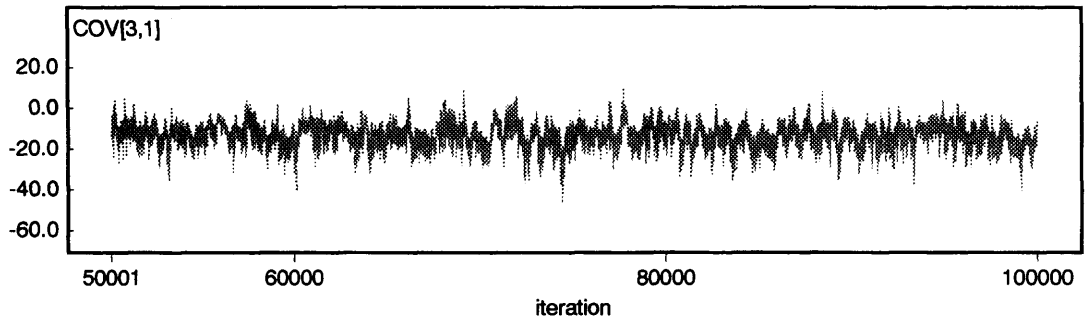

COV[1,2]
50.0
0.0
-50.0
-100.0
-150.0
50001
60000
80000
100000
iteration

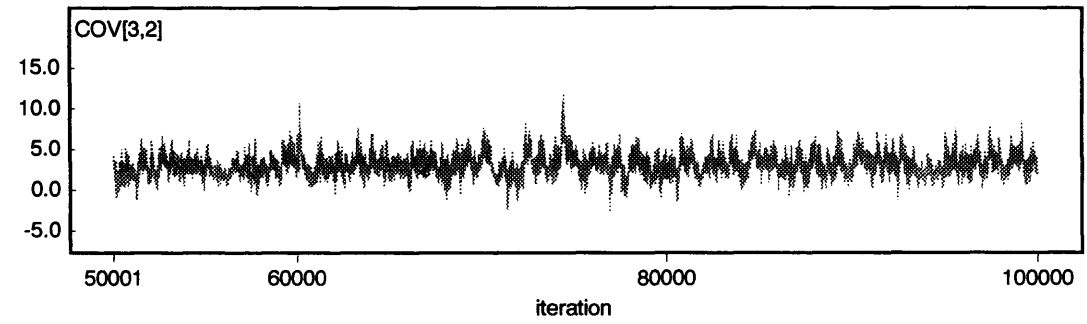

COV[1,3]
20.0
0.0
-20.0
-40.0
-60.0
50001
60000
80000
100000
iteration

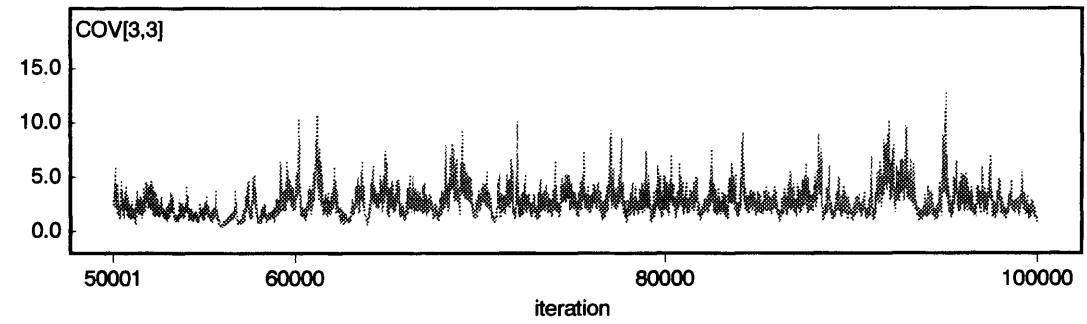

COV[2,1]
50.0
0.0
-50.0
-100.0
-150.0
50001
60000
80000
100000
iteration

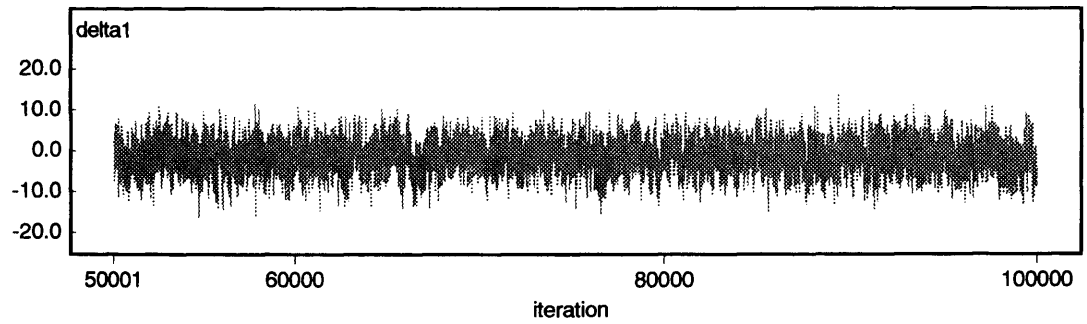

COV[2,2]
40.0
30.0
20.0
10.0
0.0
50001
60000
80000
100000
iteration

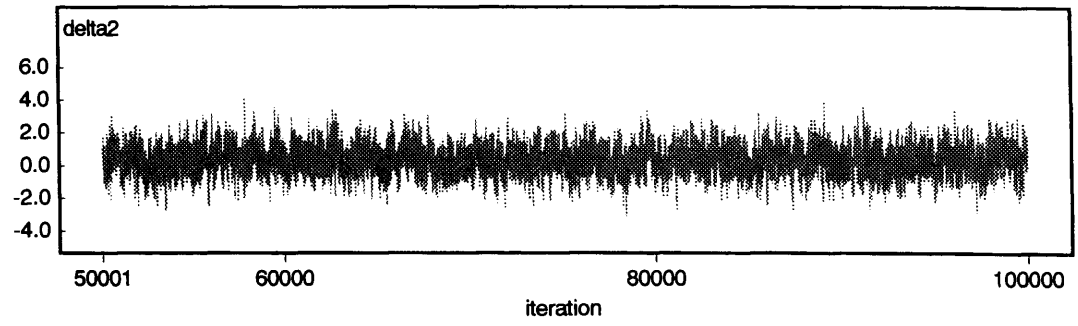

COV[2,3]
15.0
10.0
5.0
0.0
-5.0
50001
60000
80000
100000
iteration

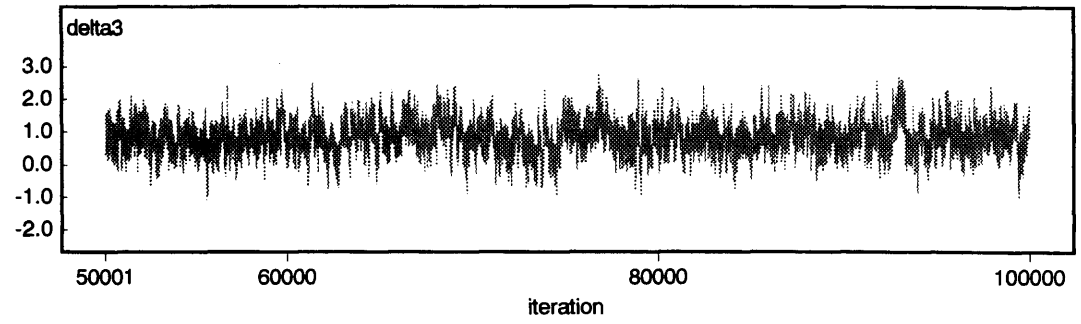

COV[3,1]
20.0
0.0
-20.0
-40.0
-60.0
50001
60000
80000
100000
iteration

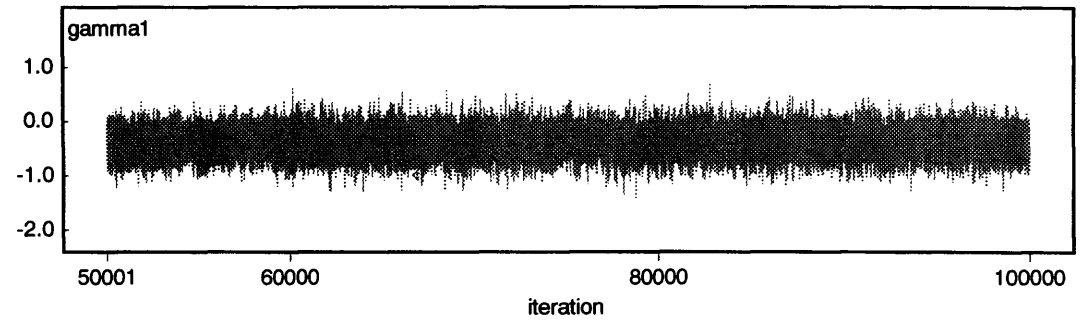

COV[3,2]
15.0
10.0
5.0
0.0
-5.0
50001
60000
80000
100000
iteration

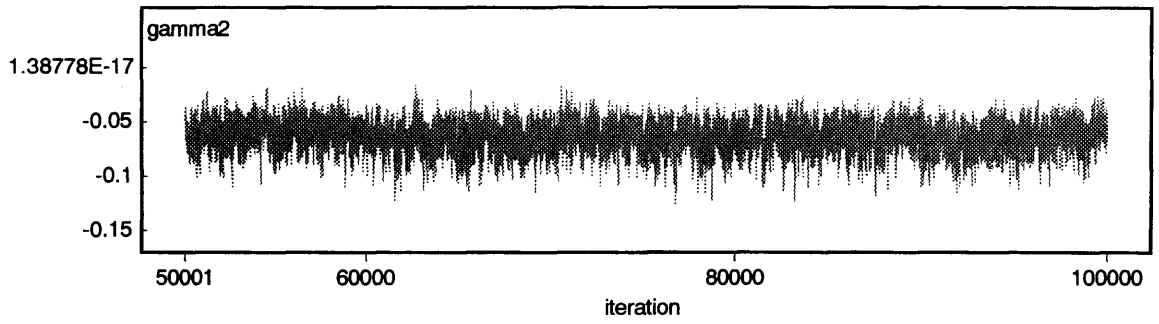

COV[3,3]
15.0
10.0
5.0
0.0
50001
60000
80000
100000
iteration

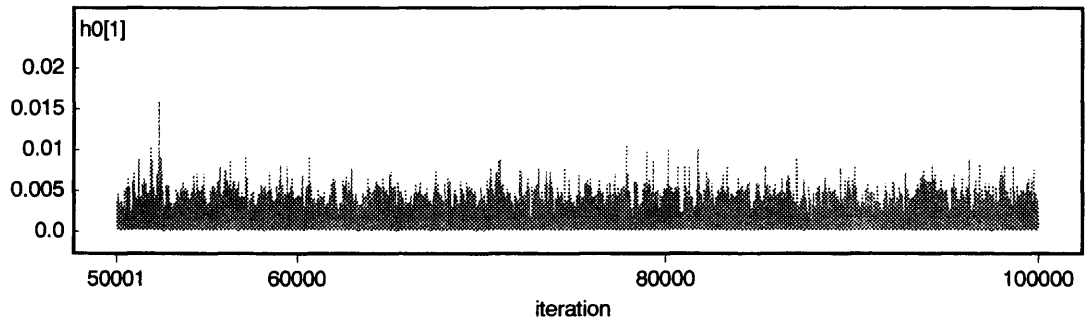

delta1
20.0
10.0
0.0
-10.0
-20.0
50001
60000
80000
100000
iteration

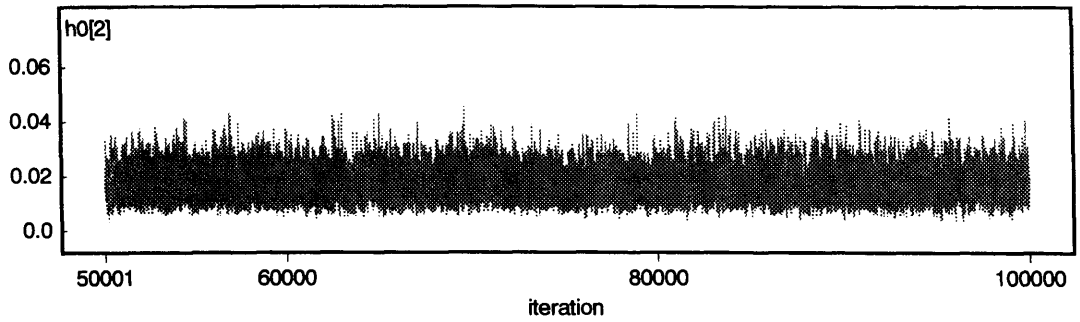

delta2
6.0
4.0
2.0
0.0
-2.0
-4.0
50001
60000
80000
100000
iteration

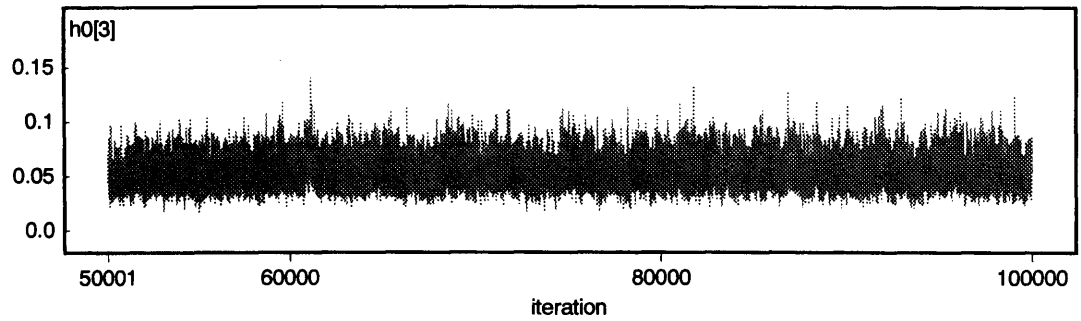

delta3
3.0
2.0
1.0
0.0
-1.0
-2.0
50001
60000
80000
100000
iteration

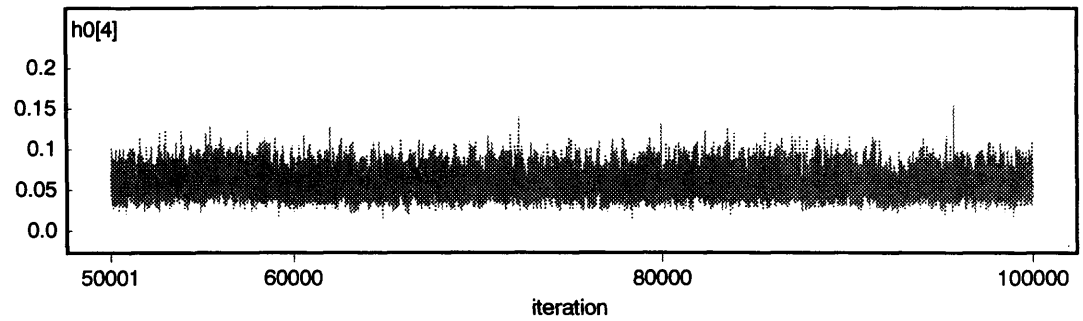

gamma1
1.0
0.0
-1.0
-2.0
50001
60000
80000
100000
iteration

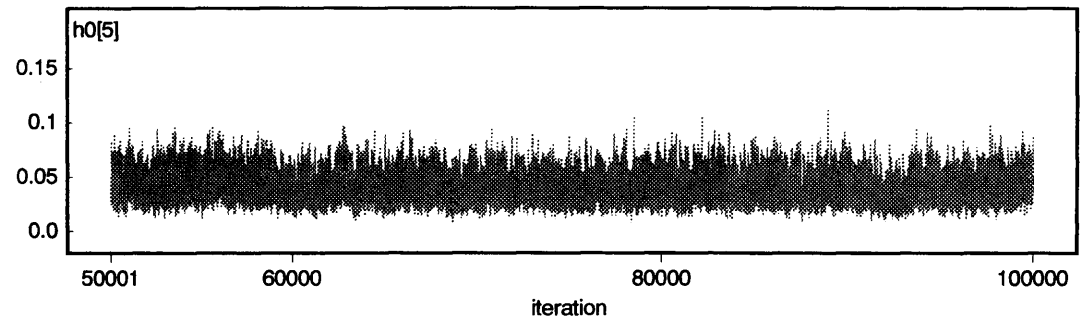

gamma2
1.38778E-17
-0.05
-0.1
-0.15
50001
60000
80000
100000
iteration

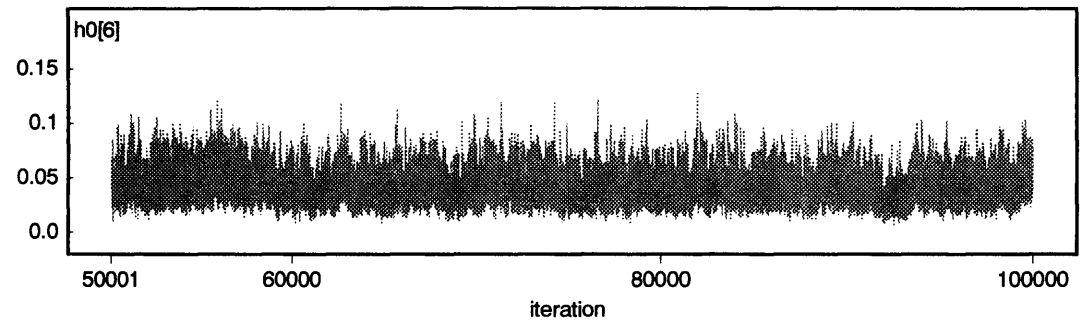

h0[1]
0.02
0.015
0.01
0.005
0.0
50001
60000
80000
100000
iteration

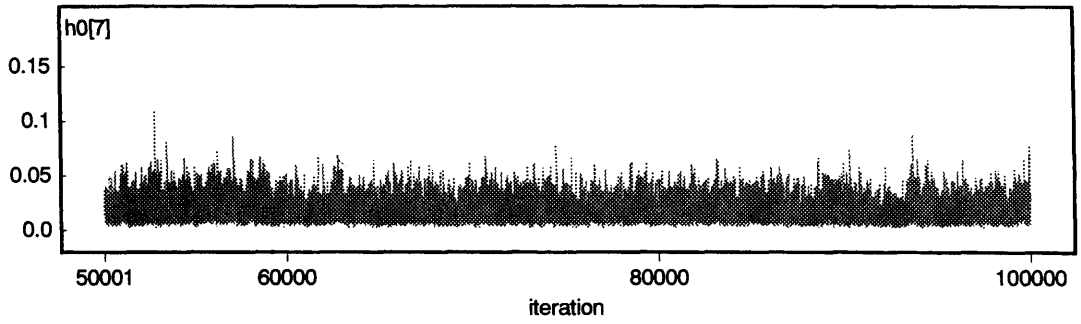

h0[2]
0.06
0.04
0.02
0.0
50001
60000
80000
100000
iteration

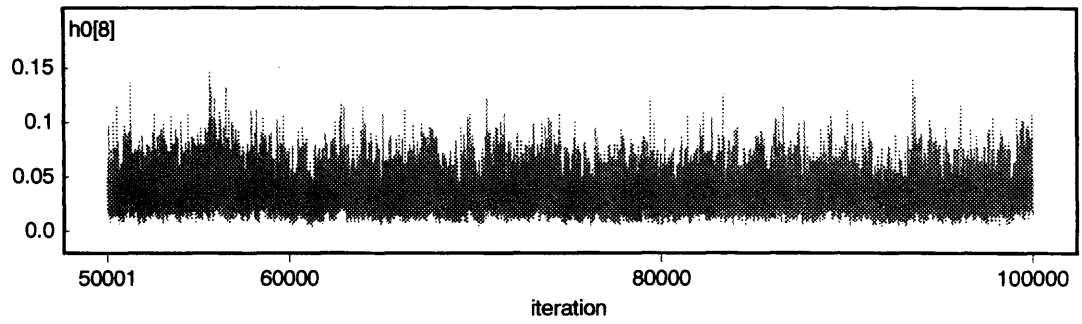

h0[3]
0.15
0.1
0.05
0.0
50001
60000
80000
100000
iteration

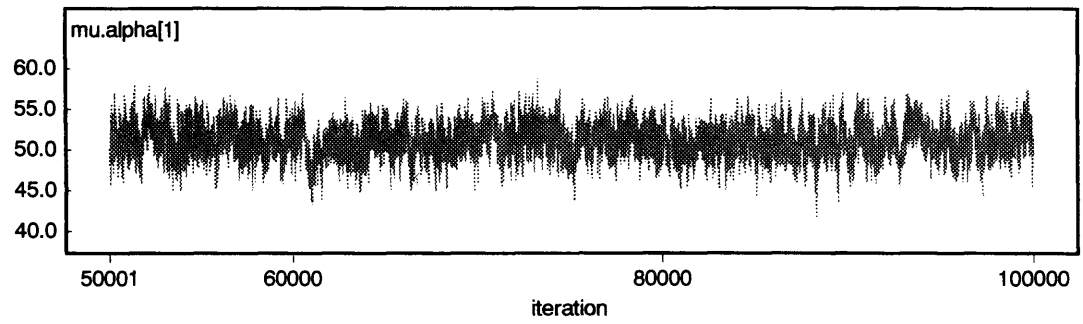

h0[4]
0.2
0.15
0.1
0.05
0.0
50001
60000
80000
100000
iteration

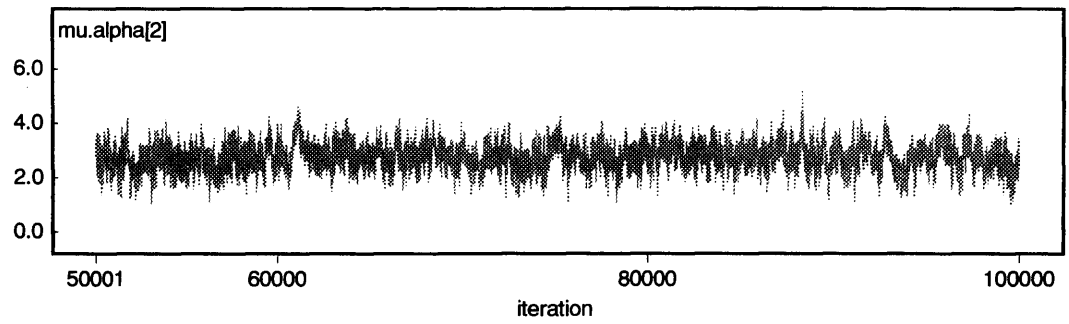

h0[5]
0.15
0.1
0.05
0.0
50001
60000
80000
100000
iteration

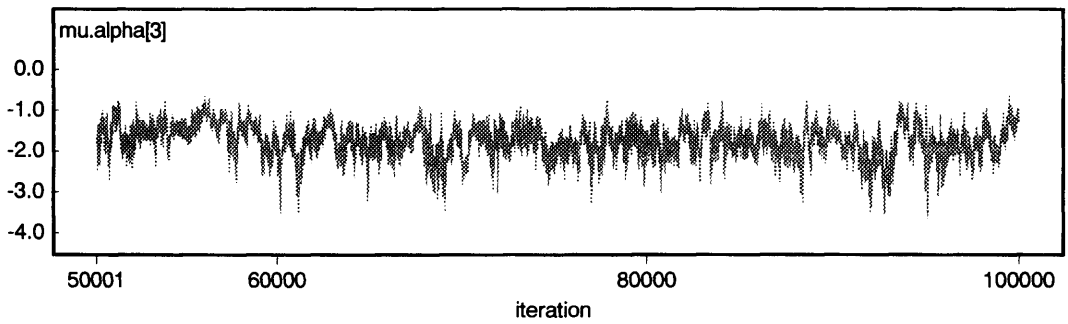

h0[6]
0.15
0.1
0.05
0.0
50001
60000
80000
100000
iteration

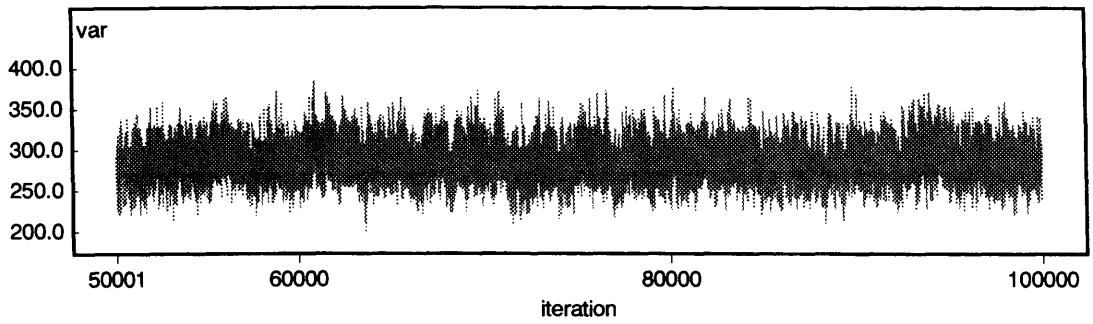

h0[7]
0.15
0.1
0.05
0.0
50001
60000
80000
100000
iteration

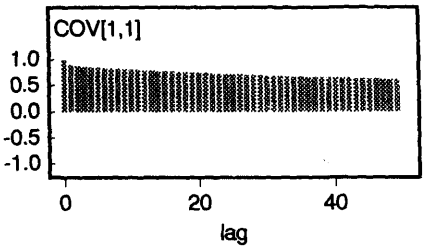

h0[8]
0.15
0.1
0.05
0.0
50001
60000
80000
100000
iteration

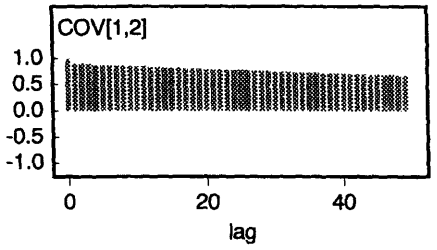

mu.alpha[1]
60.0
55.0
50.0
45.0
40.0
50001
60000
80000
100000
iteration

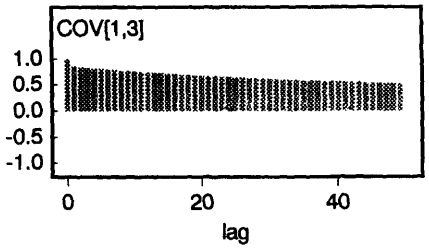

mu.alpha[2]
6.0
4.0
2.0
0.0
50001
60000
80000
100000
iteration

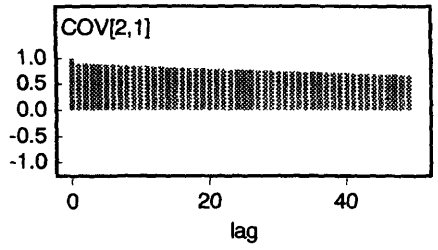

mu.alpha[3]
0.0
-1.0
-2.0
-3.0
-4.0
50001
60000
80000
100000
iteration

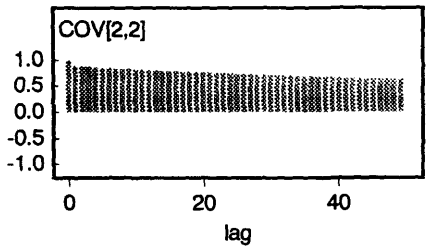

var
400.0
350.0
300.0
250.0
200.0
50001
60000
80000
100000
iteration

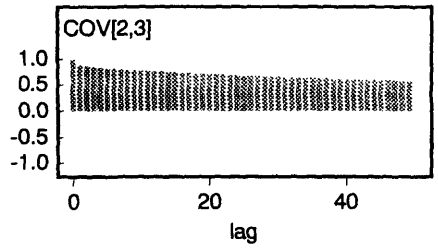

COV[1,1]
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag

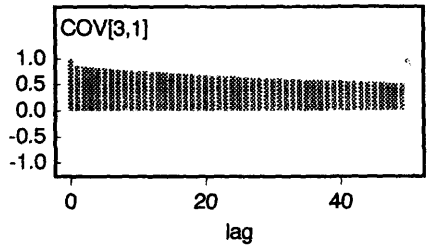

COV[1,2]
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag

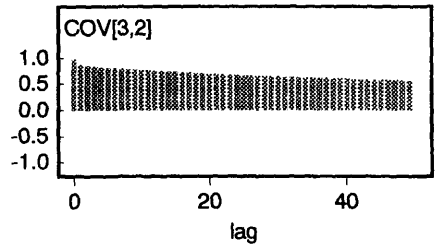

COV[1,3]
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag

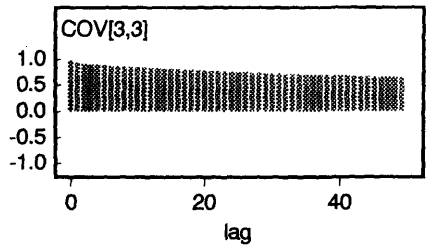

COV[2,1]
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag

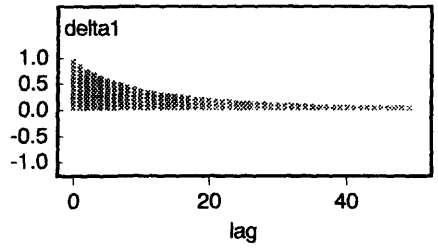

COV[2,2]
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag

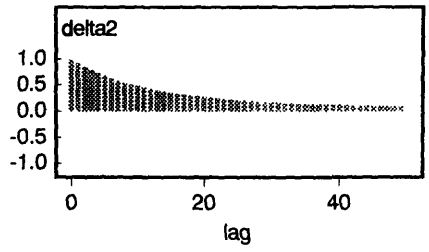

COV[2,3]
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag

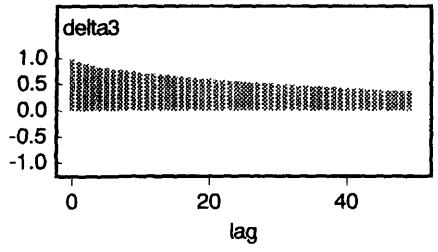

COV[3,1]
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag

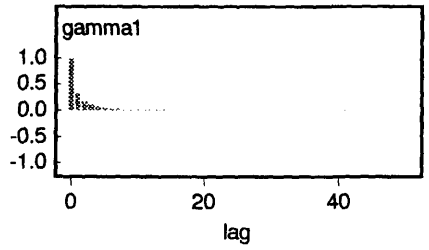

COV[3,2]
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag

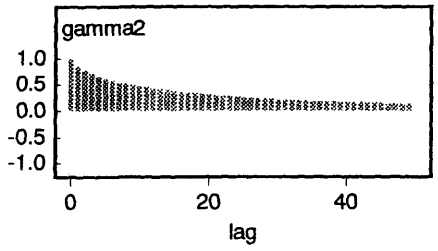

COV[3,3]
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag

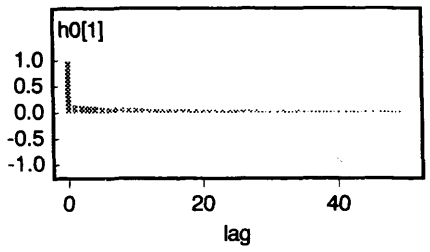

delta1
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag

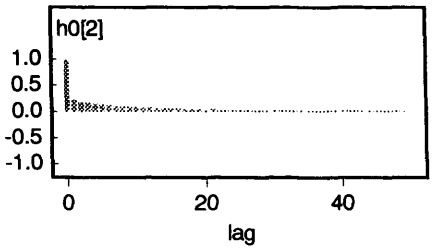

delta2
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag

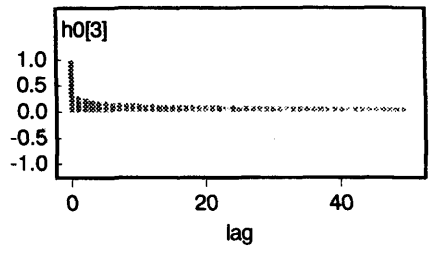

delta3
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag

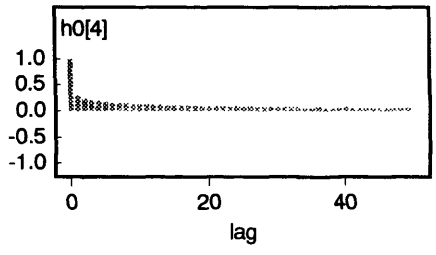

gamma1
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag

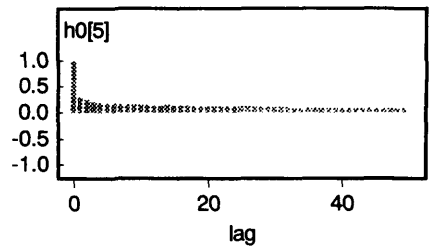

gamma2
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag

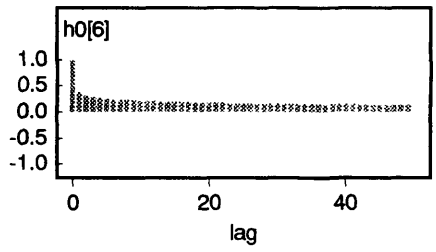

h0[1]
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag

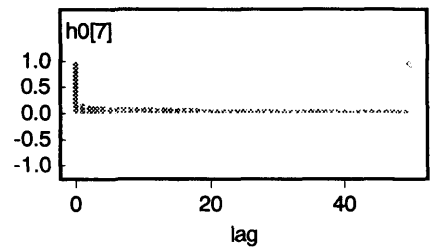

h0[2]
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag

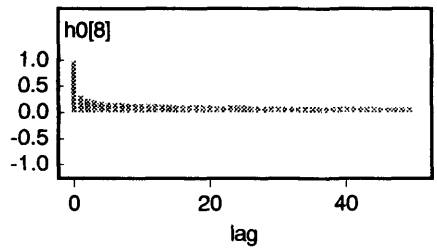

h0[3]
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag

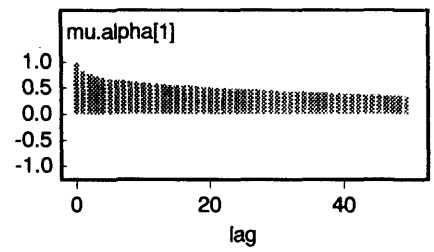

h0[4]
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag

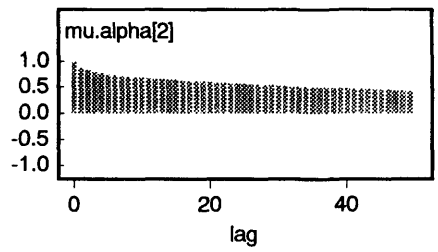

h0[5]
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag

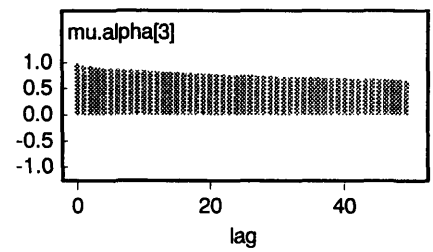

h0[6]
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag

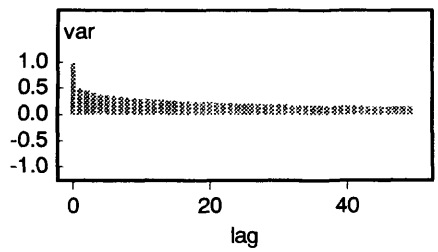

h0[7]
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag

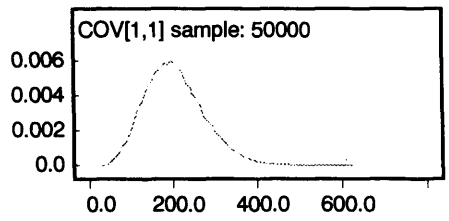

h0[8]
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag

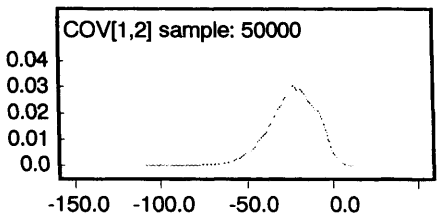

mu.alpha[1]
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag

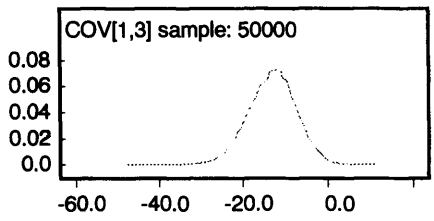

mu.alpha[2]
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag

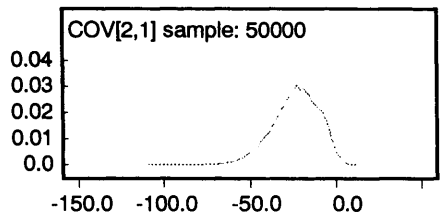

mu.alpha[3]
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag

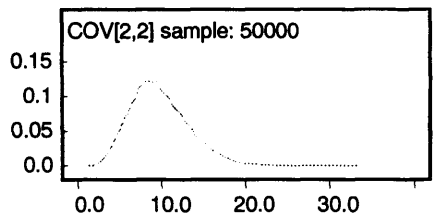

var
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag

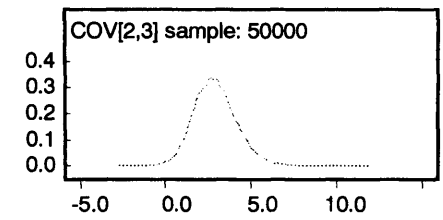

COV[1,1] sample: 50000
0.006
0.004
0.002
0.0
0.0
200.0
400.0
600.0

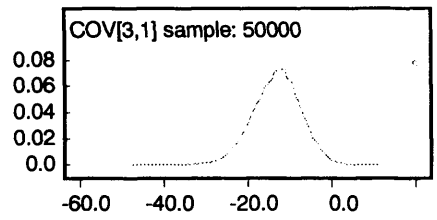

COV[1,2] sample: 50000
0.04
0.03
0.02
0.01
0.0
-150.0
-100.0
-50.0
0.0

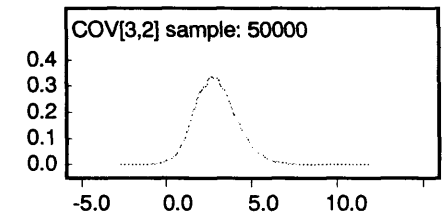

COV[1,3] sample: 50000
0.08
0.06
0.04
0.02
0.0
-60.0
-40.0
-20.0
0.0

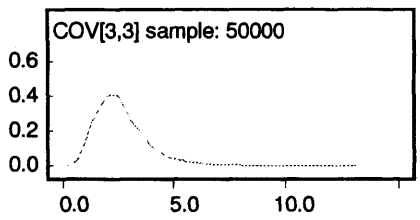

COV[2,1] sample: 50000
0.04
0.03
0.02
0.01
0.0
-150.0
-100.0
-50.0
0.0

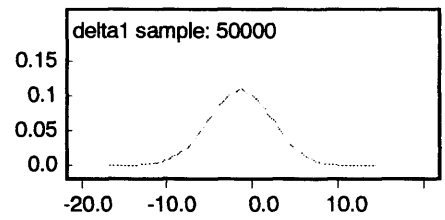

COV[2,2] sample: 50000
0.15
0.1
0.05
0.0
0.0
10.0
20.0
30.0

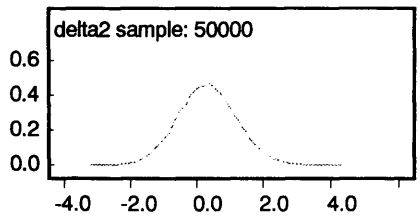

COV[2,3] sample: 50000
0.4
0.3
0.2
0.1
0.0
-5.0
0.0
5.0
10.0

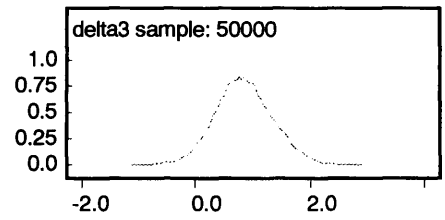

COV[3,1] sample: 50000
0.08
0.06
0.04
0.02
0.0
-60.0
-40.0
-20.0
0.0

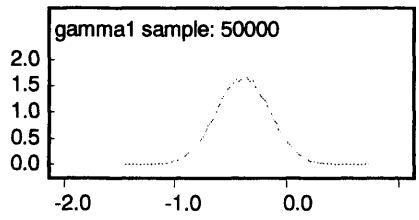

COV[3,2] sample: 50000
0.4
0.3
0.2
0.1
0.0
-5.0
0.0
5.0
10.0

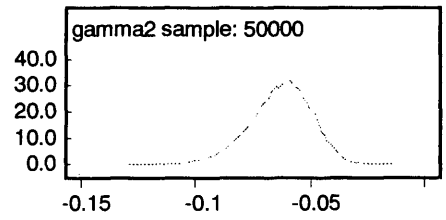

COV[3,3] sample: 50000
0.6
0.4
0.2
0.0
0.0
5.0
10.0

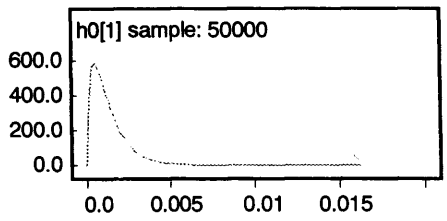

delta1 sample: 50000
0.15
0.1
0.05
0.0
-20.0
-10.0
0.0
10.0

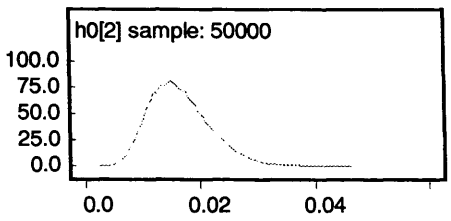

delta2 sample: 50000
0.6
0.4
0.2
0.0
-4.0
-2.0
0.0
2.0
4.0

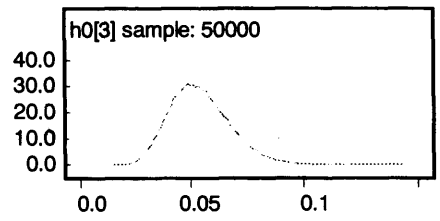

delta3 sample: 50000
1.0
0.75
0.5
0.25
0.0
-2.0
0.0
2.0

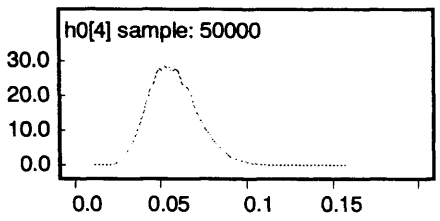

gamma1 sample: 50000
2.0
1.5
1.0
0.5
0.0
-2.0
-1.0
0.0

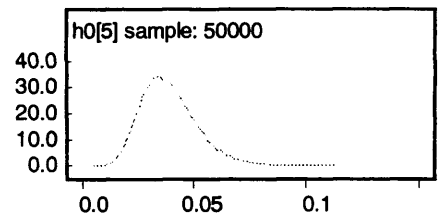

gamma2 sample: 50000
40.0
30.0
20.0
10.0
0.0
-0.15
-0.1
-0.05

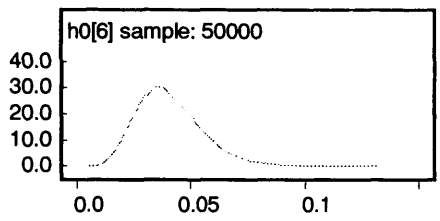

h0[1] sample: 50000
600.0
400.0
200.0
0.0
0.0
0.005
0.01
0.015

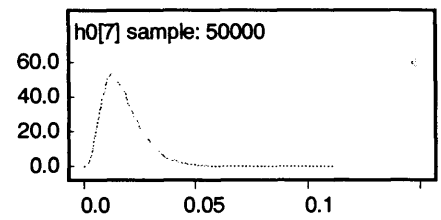

h0[2] sample: 50000
100.0
75.0
50.0
25.0
0.0
0.0
0.02
0.04

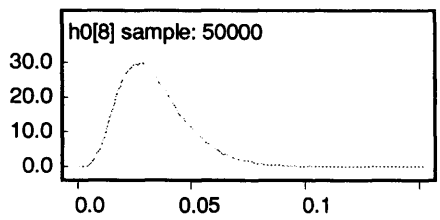

h0[3] sample: 50000
40.0
30.0
20.0
10.0
0.0
0.0
0.05
0.1

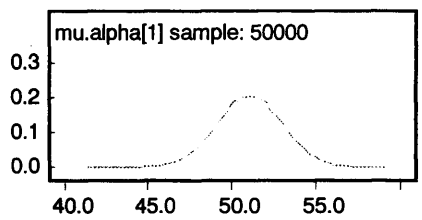

h0[4] sample: 50000
30.0
20.0
10.0
0.0
0.0
0.05
0.1
0.15

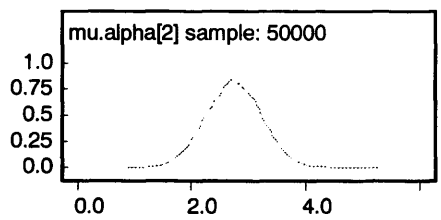

h0[5] sample: 50000
40.0
30.0
20.0
10.0
0.0
0.0
0.05
0.1

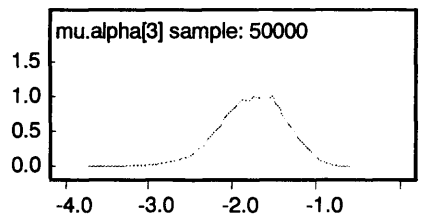

h0[6] sample: 50000
40.0
30.0
20.0
10.0
0.0
0.0
0.05
0.1

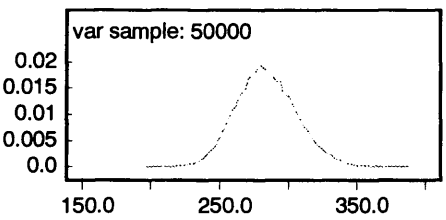

h0[7] sample: 50000
60.0
40.0
20.0
0.0
0.0
0.05
0.1


h0[8] sample: 50000
30.0
20.0
10.0
0.0
0.0
0.05
0.1


mu.alpha[1] sample: 50000
0.3
0.2
0.1
0.0
40.0
45.0
50.0
55.0


mu.alpha[2] sample: 50000
1.0
0.75
0.5
0.25
0.0
0.0
2.0
4.0


mu.alpha[3] sample: 50000
1.5
1.0
0.5
0.0
-4.0
-3.0
-2.0
-1.0


var sample: 50000
0.02
0.015
0.01
0.005
0.0
150.0
250.0
350.0

BIBLIOGRAPHY

Aaronson NK (1989) Quality of life assessment in clinical trials: methodologic issues. Controlled Clinical Trials; 10: S195-208.

Aaronson NK, Ahmedzai S, Bergman B et al (1993) The European Organization for Research and Treatment of Cancer QLQ-C30: a quality of life instrument for use in international clinical trials in oncology. Journal of the National Cancer Institute; 85: 365-76.

Abrams KR (1992a) In discussion of Cox et al: Quality-of-life assessment: can we keep it simple? Journal of the Royal Statistical Society (A); 155: 353-93.

Abrams KR (1992b) Bayesian Survival Analysis. PhD thesis; University of Liverpool.

Abrams KR (1998) Regression models for survival data. In Everitt BS, Dunn G (Eds) Statistical Analysis of Medical Data. Arnold: 13-58.

Albert PS (1999) Longitudinal data analysis (repeated measures) in clinical trials. Statistics in Medicine; 18: 1707-32.

Allen-Mersh TG, Earlham S, Fordy C, Abrams K, Houghton J (1994) Quality of life and survival with continuous hepatic-artery floxuridine infusion for colorectal liver metastases. Lancet; 344: 1255-60.

Allison PD (1995) Survival Analysis Using the SAS System: A Practical Guide. Cary, NC: SAS Institute Inc.

Andersen PK (1988) Multistate models in survival analysis: a study of nephropathy and mortality in diabetes. Statistics in Medicine; 7: 661-70.

Andersen PK, Borgan O (1985) Counting process models for life history data: a review. Scandinavian Journal of Statistics; 12: 97-158.

Andersen PK, Keiding N (2002) Multi-state models for event history analysis. Statistical Methods in Medical Research; 11: 91-115.

Andersen PK, Hansen LS, Keiding N (1991) Assessing the influence of reversible disease indicators on survival. Statistics in Medicine; 10: 1061-7.

Andersen PK, Borgan O, Gill RD, Keiding N (1993) Statistical Models Based on Counting Processes. New York: Springer-Verlag.

Andersen PK, Esbjerg S, Sorensen TIA (2000) Multi-state models for bleeding episodes and mortality in liver cirrhosis. Statistics in Medicine; 19: 587-99.

Armitage P, Berry G (1987) Statistical Methods in Medical Research. Blackwell Scientific Publications.

Bailey AJ, Parmar MK, Stephens RJ (1998) Patient-reported short-term and long-term physical and psychological symptoms: results of the continuous hyperfractionated accelerated radiotherapy (CHART) randomized trial in non-small-cell lung cancer. CHART Steering Committee. Journal of Clinical Oncology; 16: 3083-93.

Barber JA, Thompson SG (2000) Analysis of cost data in randomised trials: application of the non-parametric bootstrap. Statistics in Medicine; 19: 3219-3236.

Beacon HJ (1996) The Statistical Analysis of Self Assessed Quality of Life Data in Cancer Clinical Trials. PhD thesis; London School of Hygiene and Tropical Medicine, University of London.

Beacon HJ, Thompson S (1996) Multi-level models for repeated measurement data: application to quality of life data in clinical trials. Statistics in Medicine; 15: 2717-32.

Beck GJ (1979) Stochastic survival models with competing risks and covariates. Biometrics; 35: 427-38.

Bergner M, Babbitt RA, Carter WB, Gilson BS (1981) The sickness impact profile: development and final revision of a health status measure. Medical Care; 19: 787-8

Bernardo JM, Smith AFM (1994) Bayesian Theory. Chichester: Wiley.

Bernhard J, Gelber RD (Eds) (1998) Workshop on missing data in quality of life research in cancer clinical trials: practical and methodological issues. Statistics in Medicine; Vol 17: Nos 5-7.

Berzuini C (1995) Medical monitoring. In Gilks WR, Richardson S, Spiegelhalter DJ (Eds) Markov Chain Monte Carlo in Practice. London: Chapman and Hall; Chapter 18.

Best NG, Spiegelhalter DJ, Thomas A, Brayne CEG (1996) Bayesian analysis of realistically complex models. Journal of the Royal Statistical Society (A); 159: 323-42.

Billingham LJ, Abrams KR (2002) Simultaneous analysis of quality of life and survival data. Statistical Methods in Medical Research; 11: 25-48.

Billingham LJ, Cullen M (2003) Potential hazards in the analysis of symptom improvement. Lung Cancer; 40: 201-2.

Billingham LJ, Cullen MH, Woods J et al (1997) Mitomycin, ifosfamide and cisplatin (MIC) in non-small cell lung cancer (NSCLC): 3. Results of a randomised trial evaluating palliation and quality of life. Lung Cancer; 18 Suppl: abstract 26.

Billingham LJ, Abrams KR, Jones DR (1999) Methods for the analysis of quality-of-life and survival data in health technology assessment. Health Technology Assessment; Vol3; No10.

Billingham LJ, Bathers S, Burton A, Bryan S, Cullen MH (2002) Patterns, costs and cost-effectiveness of care in a trial of chemotherapy for advanced non-small cell lung cancer. Lung Cancer; 37: 219-225.

Bland JM, Altman DG. (1998) Bayesians and frequentists. British Medical Journal; 317: 1151.

Bosch JL, Hunink MGM (1996) The relationship between descriptive and valuational quality of life measures in patients with intermittent claudication. Medical Decision Making; 16: 217-25.

Bowling A (1991) Measuring Health: A Review of Quality of Life Measurement Scales. Open University Press.

Bowling A (1996) Measuring Disease: A Review of Disease-Specific Quality of Life Measurement Scales. Open University Press.

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: 271-92.

Breslow NE (1974) Covariance analysis of censored survival data. Biometrics; 30: 89-99.

Breslow N, Crowley J (1974) A large sample study of the life table and product limit estimates under random censorship. The Annals of Statistics; 2: 437-53.

Briggs A, Gray A (1998) The distribution of health care costs and their statistical analysis for economic evaluation. Journal of Health Service Research Policy; 3: 233-245.

Brunner KW (1989) Adjuvant therapies in breast cancer and quality of life: a critical review of the TWiST concept. In Senn HJ, Goldhirsch A, Gelber RD, Osterwalder B (Eds) Adjuvant Therapy of Primary Breast Cancer. Springer-Verlag: 239-43.

Burton A, Billingham LJ, Bryan S (2003) Cost-effectiveness in clinical trials: using imputation to deal with incomplete cost data. Submitted to Health Economics.

Burton P, Gurrin L, Sly P (1998) Extending the simple linear regression model to account for correlated responses: an introduction to generalized estimating equations and multi-level mixed modelling. Statistics in Medicine; 17: 1261-91.

Campbell M, Gibbard N (1998) Frequency of use of quality of life instruments. Health Technology Assessment.

Cantor AB (1997) Extending SAS Survival Techniques for Medical Research. Cary, NC: SAS Institute Inc.

Carpenter J, Pocock S, Lamm CJ (2002) Coping with missing data in clinical trials: a model-based approach applied to asthma trials. Statistics in Medicine; 21: 1043-66.

Chancellor JVM, Coyle D, Drummond MF (1997) Constructing health state preference values from descriptive quality of life outcomes: mission impossible? Quality of Life Research; 6: 159-68.

Charlesworth A, Skene AM (1997) A multi-state Markov process model for quality of life and survival in large heart failure clinical trials. Controlled Clinical Trials; 18: 165S.

Chevret S, Leporrier M, Chastang C (2000) Measures of treatment effectiveness on tumour response and survival: a multi-state approach. Statistics in Medicine; 19: 837-48.

Chiang CL (1980) An Introduction to Stochastic Processes and Their Applications. Robert E Krieger Pub. Co.

Clayton D (1988) The analysis of event history data: a review of progress and outstanding problems. Statistics in Medicine; 7: 819-41.

Cnaan A, Laird NM, Slasor P (1997) Using the general linear mixed model to analyse unbalanced repeated measures and longitudinal data. Statistics in Medicine; 6: 2349-80.

Cole BF, Gelber RD, Anderson KM (1994) Parametric approaches to quality-adjusted survival analysis. Biometrics; 50: 621-31.

Cole BF, Gelber RD, Goldhirsch A (1995) A quality-adjusted survival meta-analysis of adjuvant chemotherapy for premenopausal breast cancer. Statistics in Medicine; 14: 1771-84.

Collett D (1994) Modelling Survival Data in Medical Research. London: Chapman and Hall.

Cox DR (1972) Regression models and life tables. Journal of the Royal Statistical Society (B); 34: 187-220.

Cox DR (1984) Some remarks on semi-Markov processes in medical statistics. International Symposium on Semi-Markov Processes and their Applications.

Cox DR, Miller HD (1965) The Theory of Stochastic Process. Methuen and Co Ltd.

Cox DR, Fitzpatrick R, Fletcher AE, Gore SM, Spiegelhalter DJ, Jones DR (1992) Quality-of-life assessment: can we keep it simple? Journal of the Royal Statistical Society(A); 155: 353-93.

Cullen MH, Billingham LJ, Woodroffe CM, Chetiyawardana AD, Gower NH, Joshi R, Ferry DR, Rudd RM, Spiro SG, Cook JE et al (1999) Mitomycin, ifosfamide and cisplatin (MIC) in unresectable non-small cell lung cancer: effects on survival and quality of life. Journal of Clinical Oncology; 17: 3188-3194.

Curran D, Bacchi M, Hsu Schmitz SF, Molenberghs G, Sylvester RJ (1998a) Identifying the types of missingness in quality of life data from clinical trials. Statistics in Medicine; 17: 739-56.

Curran D, Fayers PM, Molenberghs G, Machin D (1998b) Analysis of incomplete quality of life data in clinical trials. In Staquet MJ, Hays RD, Fayers PM (Eds) Quality of Life Assessment in Clinical Trials: Methods and Practice. Oxford University Press: New York.

Curran D, Molenberghs G, Fayers PM, Machin D (1998c) Incomplete quality of life data in randomised trials: missing forms. Statistics in Medicine; 17: 697-709.

Curran D, Molenburghs G, Aaronson NK, Fossa SD, Sylvester RJ (2002) Analysing longitudinal continuous quality of life data with dropout. Statistical Methods in Medical Research; 11: 5-23.

David HA, Moeschberger ML (1978) The Theory of Competing Risks. Charles Griffin and Co. Ltd.

De Gruttola V, Tu XM (1994) Modelling progression of CD4 lymphocyte count and its relationship to survival time. Biometrics; 50: 1003-14.

De Stavola BL, Christensen E (1996) Multilevel models for longitudinal variables prognostic for survival. Lifetime Data Analysis; 2: 329-47.

Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood estimation from incomplete data via the EM algorithm. Journal of the Royal Statistical Society (B); 39:1-38.

Diggle P, Kenward MG (1994) Informative drop-out in longitudinal data analysis. Applied Statistics; 43: 49-93.

Diggle PJ, Liang K, Zeger SL (1994) Analysis of Longitudinal Data. Oxford: Clarendon Press.

Dobson AJ (1990) An Introduction to Generalized Linear Models. London: Chapman and Hall.

Draper D (1995) Assessment and propogation of model uncertainty. Journal of the Royal Statistical Society; 57: 45-97.

Efron B (1977) The efficiency of Cox's likelihood function for censored data. Journal of the American Statistical Association; 72: 557-65.

Efron B, Tibshirani RJ. (1993) An Introduction to the Bootstrap. Chapman and Hall: San Francisco.

Everitt BS, Dunn G, Holford TR (Eds) (2002) Analysis of quality of life data in controlled clinical trials. Statistical Methods in Medical Research; 11: 1-87.

Fairclough DL (1997) Summary measures and statistics for comparison or quality of life in a clinical trial of cancer therapy. Statistics in Medicine; 16: 1197-209.

Fairclough DL (2002) Design and Analysis of Quality of Life Studies in Clinical Trials. Chapman & Hall: USA.

Fairclough DL, Peterson HF, Cella D, Bonomi P (1998a) Comparison of several model-based methods for analysing incomplete quality of life data in cancer clinical trials. Statistics in Medicine; 17: 781-96.

Fairclough DL, Peterson HF, Chang V (1998b) Why are missing quality of life data a problem in clinical trials of cancer therapy? Statistics in Medicine; 17: 667-77.

Fallowfield L (1990) Quality of Life: The Missing Measurement in Health Care. London: Souvenir Press (E&A) Ltd.

Faucett CL, Thomas DC (1996) Simultaneously modelling censored survival data and repeatedly measured covariates: a Gibbs sampling approach. Statistics in Medicine; 15: 1663-85.

Faucett CL, Schenker N, Elashoff RM (1998) Analysis of censored survival data with intermittently observed time-dependant binary covariates. Journal of the American Statistical Association; 93: 427-37.

Fayers PM, Jones DR (1983) Measuring and analysing quality of life in cancer clinical trials: a review. Statistics in Medicine; 2: 429-46.

Fayers PM, Machin D (1998a) Factor analysis. In Staquet MJ, Hays RD, Fayers PM (Eds) Quality of Life Assessment in Clinical Trials: Methods and Practice; Chapter 12: 191-223.

Fayers PM, Machin D (1998b) Summarizing quality of life data using graphical methods. In Staquet MJ, Hays RD, Fayers PM (Eds) Quality of Life Assessment in Clinical Trials: Methods and Practice; Chapter 17: 315-35.

Fayers PM, Machin D (2000) Quality of Life: Assessment, Analysis and Interpretation. Chichester: John Wiley and Sons Ltd.

Fayers PM, Curran D, Machin D (1998) Incomplete quality of life data in randomised trials: missing items. Statistics in Medicine; 17: 679-96.

Feldstein ML (1991) Quality-of-life-adjusted survival for comparing cancer treatments: a commentary on TWiST and Q-TWiST. Cancer; 67: 851-4.

Fitzpatrick R, Fletcher A, Gore SM, Jones DR, Spiegelhalter DJ, Cox DR (1992) Quality of life measures in health care. I: Applications and issues in assessment. British Medical Journal; 305: 1074-77.

Fitzpatrick R, Davey C, Buxton MJ, Jones DR (1998) Evaluating patient-based outcome measures for use in clinical trials. Health Technology Assessment Vol2; No14.

Fix E, Neyman J (1951) A simple stochastic model of recovery, relapse and loss of patients. Human Biology; 23: 205-41.

Fleming TR, Harrington DP (1991) Counting Processes and Survival Analysis. John Wiley and Sons.

Fletcher A, Gore SM, Jones DR, Fitzpatrick R, Spiegelhalter DJ, Cox DR (1992) Quality of life measures in health care. II: Design, analysis and interpretation. British Medical Journal; 305: 1145-48.

Gail MH (1981) Evaluating serial cancer marker studies in patients at risk of recurrent disease. Biometrics; 37: 67-78.

Ganiats TG, Miller CJ, Kaplan RM (1995) Comparing the quality-adjusted life-year output of two treatment arms in a randomised trial. Medical Care; 33: AS245-AS254.

Ganiats TG, Browner DK, Kaplan RM (1996) Comparison of two methods of calculating quality-adjusted life years. Quality of Life Research; 5: 162-164.

Gelber RD, Goldhirsch A (1986) A new endpoint for the assessment of adjuvant therapy in postmenopausal women with operable breast cancer. Journal of Clinical Oncology; 4: 1772-9.

Gelber RD, Goldhirsch A (1989) Methods for assessing treatment efficacy in trials for adjuvant therapy for breast cancer. In Senn HJ, Goldhirsch A, Gelber RD, Osterwalder B (Eds) Adjuvant Therapy of Primary Breast Cancer: 211-19. Springer-Verlag.

Gelber RD, Goldhirsch A (1991) Quality-adjusted survival analysis - reply. Journal of Clinical Oncology; 9: 526.

Gelber RD, Goldhirsch A, Castiglione M, Price K, Isley M, Coates A (1987) Time without symptoms and toxicity (TWiST): a quality-of-life-oriented endpoint to evaluate adjuvant therapy. In Salmon SE (Ed) Adjuvant Therapy of Cancer V: 455-65. Orlando: Grune and Stratton.

Gelber RD, Gelman RS, Goldhirsch A (1989) A quality-of-life-oriented endpoint for comparing therapies. Biometrics; 45: 781-95.

Gelber RD, Goldhirsch A, Cavalli F (1991) Quality-of-life-adjusted evaluation of adjuvant therapies for operable breast cancer. Annals of Internal Medicine; 114: 621-28.

Gelber RD, Lenderking WR, Cotton DJ, Cole BF, Fischl MA, Goldhirsch A, Testa, MA (1992) Quality-of-life evaluation in a clinical trial of zidovudine therapy in patients with mildly symptomatic HIV infection. Annals of Internal Medicine; 116: 961-66.

Gelber RD, Cole BF, Gelber S, Goldhirsch A (1995) Comparing treatments using quality-adjusted survival: the Q-TWiST method. American Statistician; 49: 161-9.

Gelber RD, Goldhirsch A, Cole BF, Weiand HS, Schroeder G, Krook JE (1996) A quality-adjusted time without symptoms or toxicity (Q-TWiST) analysis of adjuvant radiation therapy and chemotherapy for resectable rectal cancer. Journal of the National Cancer Institute; 88: 1039-45.

Gelfand AE, Smith FM (1990) Sampling-based approaches to calculating marginal densities. Journal of the American Statistical Association; 85: 398-409.

Gentleman RC, Lawless JF, Lindsey JC, Yan P (1994) Multistate Markov models for analyzing incomplete disease history data with illustrations for HIV disease. Statistics in Medicine; 13: 805-21.

Gilks WR, Clayton DG, Spiegelhalter DJ, Best NG, McNeil AJ, Sharples LD, Kirby AJ (1993) Modelling complexity: applications of Gibbs sampling in medicine. Journal of the Royal Statistical Society (B); 55: 39-52.

Gilks WR, Thomas A, Spiegelhalter DJ (1994) A language and program for complex Bayesian modelling. The Statistician; 43: 169-77.

Gilks WR, Richardson S, Spiegelhalter DJ (1996) Markov Chain Monte Carlo in Practice. Chapman and Hall.

Glasziou PP, Simes RJ, Gelber RD (1990) Quality adjusted survival analysis. Statistics in Medicine; 9: 1259-76.

Glasziou PP, Cole BF, Gelber RD, Hilden J, Simes RJ (1998) Quality adjusted survival analysis with repeated quality of life measures. Statistics in Medicine; 17: 1215-29.

Glimelius B, Hoffman K, Graf W, Haglund U, Nyren O, Pahlman L, Sjoden PO (1995) Cost-effectiveness of palliative chemotherapy in advanced gastrointestinal cancer. Annals of Oncology; 6: 267-74.

Goldhirsch A, Gelber RD, Simes RJ, Glasziou P, Coates AS (1989) Costs and benefits of adjuvant therapy in breast cancer: a quality-adjusted survival analysis. Journal of Clinical Oncology; 7: 36-44.

Goldstein H (1995) Multilevel Statistical Models, 2nd Edn. London: Edward Arnold.

Goodwin PJ, Field R, Evans WK, Pater J (1988) Cost-effectiveness of cancer chemotherapy: an economic evaluation of a randomised trial in small-cell lung cancer. Journal of Clinical Oncology; 6: 1537-47.

Gottschau A, Hogh B (1995) Interval censored survival data and multistate compartmental models in the analysis of first appearance of plasmodium falciparum parasites in infants. Statistics in Medicine; 14: 2727-36.

Hand D, Crowder M (1996) Practical Longitudinal Data Analysis. London: Chapman and Hall.

Hansen BE, Thorogood J, Hermans J, Ploeg RJ, van Bockel JH, van Houwelingen JC (1994) Multistate modelling of liver transplantation data. Statistics in Medicine; 13: 2517-29.

Hays RD, Anderson R, Revicki D (1993) Psychometric considerations in evaluating health-related quality of life measures. Quality of Life Research; 2: 441-9.

Hays RD, Anderson R, Revick D (1998) Assessing reliability and validity of measurement in clinical trials. In Staquet MJ, Hays RD, Fayers PM (Eds) Quality of Life Assessment in Clinical Trials: Methods and Practice; Chapter 10: 169-82.

Henderson R, Diggle P, Dobson A (2000) Joint modelling of longitudinal measurements and event time data. Biostatistics; 1: 465-80.

Hillis A, Maguire M, Hawkins BS, Newhouse MM (1986) The Markov process as a general method for nonparametric analysis of right-censored medical data. Journal of Chronic Diseases; 39: 595-604.

Hinkley DV (1988) Bootstrap methods. Journal of the Royal Statistical Society (B); 50: 321-37.

Hogan JW, Laird NM (1997a) Mixture models for the joint distribution of repeated measures and event times. Statistics in Medicine; 16: 239-257.

Hogan JW, Laird NM (1997b) Model-based approaches to analysing incomplete longitudinal and failure time data. Statistics in Medicine; 16: 259-272.

Hollen PJ, Gralla RJ, Cox C, Eberly SW, Kris MG (1997) A dilemma in analysis: issues in the serial measurement of quality of life in patients with advanced lung cancer. Lung Cancer; 18: 119-36.

Hopwood P, Stephens RJ, Machin D (1994) Approaches to the analysis of quality of life data: experiences gained from a Medical Research Council lung cancer working party palliative chemotherapy trial. Quality of Life Research; 3: 339-52.

Hosmer DN, Lemeshow S (1999) Applied Survival Analysis: Regression Modelling of Time-to-Event Data. John Wiley & Sons.

Hsieh FY, Crowley J, Tormey DC (1983) Some test statistics for use in multistate survival analysis. Biometrika; 70: 111-19.

Hwang JS, Tsauo JY, Wang JD (1996) Estimation of expected quality adjusted survival by cross-sectional survey. Statistics in Medicine; 15: 93-102.

Ibrahim JG, Chen MH, Sinia D (2001) Baysian Survival Analysis. New York: Springer-Velag.

Islam MA (1994) Multistate survival models for transitions and reverse transitions: an application to contraceptive use data. Journal of the Royal Statistical Society (A); 157: 441-55.

Islam MA, Singh KP (1992) Multistate survival models for partially censored data. Environmetrics; 3: 223-34.

Kalbfleisch JD, Prentice RL (1980) The Statistical Analysis of Failure Time Data. John Wiley and Sons.

Kaplan R (1998) Profile versus utility based measures of outcome for clinical trials. In Staquet MJ, Hays RD, Fayers PM (Eds) Quality of Life Assessment in Clinical Trials: Methods and Practice; Chapter 5: 69–90.

Kaplan EL, Meier P (1958) Nonparametric estimation from incomplete observations. Journal of the American Statistical Association; 53: 457-81.

Karnofsky D, Burchenal JH (1949) Clinical evaluation of chemotherapeutic agents in cancer. In MacLeod CM (Ed) Evaluation of Chemotherapeutic Agents. New York: Columbia University Press.

Kay R (1982) The analysis of transition times in multistate stochastic processes using proportional hazard regression models. Communications in Statistics - Theory and Methodology; 11: 1743-56.

Kay R (1984) Multistate survival analysis: an application in breast cancer. Methods of Information in Medicine; 23: 157-62.

Kay R (1986) A Markov model for analysing cancer markers and disease states in survival studies. Biometrics; 42: 855-65.

Kenward MG (1998) Selection models for repeated measurements with non-random dropout: an illustration of sensitivity. Statistics in Medicine; 17: 2723-32.

Kenward MG, Molenberghs G (1999) Parametric models for incomplete continuous and categorical longitudinal data. Statistical Methods in Medical Research; 8: 51-83.

Klein JP, Shu Y (2002) Multi-state models for bone marrow transplantation studies. Statistical Methods in Medical Research; 11: 117-139.

Klein JP, Keiding N, Copelan EA (1993) Plotting summary predictions in multistate survival models: probabilities of relapse and death in remission for bone-marrow transplantation patients. Statistics in Medicine; 12: 2315-32.

Korn EL (1993) On estimating the distribution function for quality of life in cancer clinical trials. Biometrika; 80: 535-42.

Lagakos SW (1976) A stochastic model for censored survival data in the presence of an auxiliary variable. Biometrics; 32: 551-59.

Lagakos SW, Sommer CJ, Zelen M (1978) Semi-Markov models for partially censored data. Biometrika; 65: 311-17.

Laird NM, Ware JH (1982) Random-effects models for longitudinal data. Biometrics; 38: 963-74.

Laird NM (1988) Missing data in longitudinal studies. Statistics in Medicine; 7: 305-15.

Lambert PC, Abrams KR, Sanso B, Jones DR (1999) Synthesis of incomplete data using Bayesian hierarchical models: an illustration based on data describing survival from Neuroblastoma. University of Leicester: Epidemiology and Public Health: Tech Report 93-03.

Lambert PC, Sutton AJ, Burton PR, Abrams KR, Jones DR (2003a) How vague is vague? A simulation study to investigate the use of vague prior distributions in WinBUGS. Submitted to Statistics in Medicine.

Lambert PC, Billingham LJ, Cooper NJ, Sutton AJ, Abrams KR (2003b) Estimating the cost-effectiveness of an intervention in a clinical trial when partial cost information is available: a Bayesian approach. Submitted to JRSS.

Lee PM (1989) Bayesian Statistics: An Introduction. Arnold.

Lin DY, Feuer EJ, Etzioni R, Wax Y (1997) Estimating medical costs from incomplete follow-up data. Biometrics; 53: 419-434.

Lindsey JK (1994) Models for Repeated Measurements. New York: Oxford University Press.

Lindsey JK (1997) Modelling longitudinal measurement dropouts as a survival process with time-varying covariates. Technical Report. Limburgs Universitair Centrum, Belgium.

Lindstrom MJ, Bates DM (1988) Newton-Raphson and EM algorisms for linear mixed effects models for repeated measures data. Journal of the American Statistical Association; 83: 1014 –22.

Littell RC, Pendergast J, Natarajan R (2000) Modelling covariance structure in the analysis of repeated measures data. Statistics in Medicine; 19: 1793-819.

Little RJA (1995) Modelling the drop-out mechanism in repeated measures studies. Journal of the American Statistical Association; 90: 1112-21.

Little RJA, Rubin DB (1987) Statistical Analysis with Missing Data. New York: John Wiley and Sons.

Longini IM, Clark WS, Byers RH, Ward JW, Darrow WW, Lemp GF, Hethcote HW (1989) Statistical analysis of the stages of HIV infection using a Markov model. Statistics in Medicine; 8: 831-43.

Lu Y, Stitt FW (1994) Using Markov processes to describe the prognosis of HIV-1 infection. Medical Decision Making; 14: 266-72.

Lydick E, Epstein RS, Himmelberger D, White CJ (1995) Area under the curve: a metric for patient subjective responses in episodic diseases. Quality of Life Research; 4: 41-45.

Machin D, Weeden S (1998) Suggestions for the presentation of quality of life data from clinical trials. Statistics in Medicine; 17: 711-24.

Marshall G, Jones RH (1995) Multistate models and diabetic retinopathy. Statistics in Medicine; 14: 1975-83.

Matthews JNS (1993) A refinement to the analysis of serial data using summary measures. Statistics in Medicine; 12: 27-37.

Matthews JNS, Altman DG, Campbell MJ, Royston P (1990) Analysis of serial measurements in medical research. British Medical Journal; 300: 230-5.

Michiels B, Molenberghs G, Bijnens L, Vangeneugden T, Thijs H (2002) Selection models and pattern-mixture models to analyse longitudinal quality of life data subject to drop-out. Statistics in Medicine; 21: 1023-41.

Miller DK, Homan SM (1994) Determining transition probabilities: confusion and suggestions. Medical Decision Making; 14: 52-8.

Mori M, Woodworth GG, Woolson RF (1992) Application of empirical Bayes inference to estimation of rate of change in the presence of informative right censoring. Statistics in Medicine; 11: 621-31.

MRC Lung Cancer Working Party (1991) Inoperable non-small-cell lung cancer (NSCLC): a Medical Research Council randomised trial of palliative radiotherapy with two fractions or ten fractions. British Journal of Cancer; 63: 265-70.

MRC Lung Cancer Working Party (1992) A Medical Research Council (MRC) randomised trial of palliative radiotherapy with two fractions or a single fraction in patients with inoperable non-small-cell lung cancer (NSCLC) and poor performance status. British Journal of Cancer; 65: 934-41.

MRC Lung Cancer Working Party (1996) Randomized trial of palliative two-fraction versus more intensive 13-fraction radiotherapy for patients with inoperable non-small cell lung cancer and good performance status. Clinical Oncology; 8: 167-75.

Myers LE, Paulson DF, Berry WR, Cox EB, Laszlo J, Stanley W (1980) A time-dependent statistical model which relates current clinical status to prognosis: application to advanced prostatic cancer. Journal of Chronic Diseases; 33: 491-9.

Nabholtz JM, Gelmon K, Bontenbal M, Spielmann M, Catimel G, Conte P, Klaassen U, Namer M, Bonneterre J, Fumoleau B, Winograd, B (1996) Multicenter, randomized comparative study of 2 doses of paclitaxel in patients with metastatic breast cancer. Journal of Clinical Oncology; 14: 1858-67.

Nayfield SG, Ganz PA, Moinpour CM, Cella DF, Hailey BJ (1992) Report from a National Cancer Institute (USA) workshop on quality of life assessment in cancer clinical trials. Quality of Life Research; 1: 203-10.

Neoptolemos J, Dunn JA, Stocken DD et al (2001) Adjivant chemoradiotherapy and chemotherapy in resectable pancreatic cancer: a randomised controlled trial. Lancet; 358: 1576-85.

O'Brien PC (1984) Procedures for comparing samples with multiple endpoints. Biometrics; 40: 1079-87.

O'Brien BJ, Buxton MJ, Ferguson BA (1987) Measuring the effectiveness of heart transplant programs: quality of life data and their relationship to survival analysis. Journal of Chronic Diseases; 40: S137-S153.

Olschewski M (1984) Methods of evaluating and analyzing quality of life in clinical trials. Biometrics; 40, 264.

Olschewski M, Schumacher M (1990) Statistical analysis of quality of life data in cancer clinical trials. Statistics in Medicine; 9: 749-63.

Olschewski M, Schulgen G, Schumacher M (1992) In discussion of Cox et al: Quality-of-life assessment: can we keep it simple? Journal of the Royal Statistical Society (A); 155: 353-93.

Osoba D (1998) Guidelines for measuring health-related quality of life in clinical trials. In Staquet MJ, Hays RD, Fayers PM (Eds) Quality of Life Assessment in Clinical Trials: Methods and Practice; Chapter 2: 19-35.

Parmar MKB, Machin D (1995) Survival Analysis: A Practical Approach. England: John Wiley and Sons.

Pauler DK, Finkelstein DM (2002) Predicting time to prostate cancer recurrence based on joint models for non-linear longitudinal biomarkers and event time outcomes. Statistics in Medicine; 21: 3897-911.

Pauler DK, McCoy S, Moinpour C (2003) Pattern mixture models for longitudinal quality of life studies in advanced stage disease. Statistics in Medicine; 22: 795-809.

Pawitan Y, Self S (1993) Modelling disease marker processes in AIDS. Journal of the American Statistical Association; 88: 719-26.

Perez-Ocon R, Ruiz-Castro JE, Gamiz-Perez ML (2001a) A piecewise Markov process for analysing survival from breast cancer in different risk groups. Statistics in Medicine; 20: 109-22.

Perez-Ocon R, Ruiz-Castro JE, Gamiz-Perez ML (2001b) Non-homogeneous Markov models in the analysis of survival after breast cancer. Applied Statistics; 50: 111-24.

Petersen T (1986) Fitting parametric survival models with time-dependent covariates. Applied Statistics; 35: 281-8.

Pocock SJ (1991) A perspective on the role of quality-of-life assessment in clinical trials. Controlled Clinical Trials; 12: S257-S265.

Pocock SJ, Geller NL, Tsiatis AA (1987) The analysis of multiple endpoints in clinical trials. Biometrics; 43: 487-98.

Prentice RL (1982) Covariate measurement errors and parameter estimation in a failure time regression model. Biometrika; 69: 331-42.

Prentice RL, Williams BJ, Peterson AV (1981) On the regression analysis of multivariate failure time data. Biometrika; 68: 373-9.

Qian W, Parmar MKB, Sambrook RJ, Fayers PM, Girling DJ, Stephens RJ (2000) Analysis of messy longitudinal data from a randomized clinical trial. Statistics in Medicine; 19: 2657-74.

Rabin R, de Charro F (2001) EQ-5D: a measure of health status from EuroQol Group. Annals of Medicine; 33: 337-43.

Revicki DA, Kaplan RM (1993) Relationship between psychometric and utility-based approaches to the measurement of health-related quality of life. Quality of Life Research; 2: 477-87.

Ribaudo HJ, Thompson SG, Allen-Mersh TG (2000) A joint analysis of quality of life and survival using a random effect selection model. Statistics in Medicine; 19: 3237-50.

Rosenman J, Choi NC (1982) Improved quality of life of patients with small-cell carcinoma of the lung by elective irradiation of the brain. International Journal of Radiation Oncology Biology Physics; 8: 1041-3.

Rosenthal MA, Webster PJ, Gebski VJ, Stuartharris RC, Langlands AO, Boyages J (1992) The cost of treating small-cell lung cancer. Medical Journal of Australia; 156: 605-10.

Rubin DB (1976) Inference and missing data. Biometrika; 63: 581-92.

Rubin DB (1987) Multiple Imputation for Non-Response in Surveys. New York: John Wiley.

SAS Institute Inc (1989) SAS/STAT Users Guide, Version 6, Fourth edition, Volume 2. Cary, NC: SAS Institute Inc.

SAS Institute Inc (1992) SAS Technical Report P-229, SAS/STAT Software: Changes and Enhancements; Release 6.07. Cary, NC: SAS Institute Inc.

Schafer JL (1997) Analysis of Incomplete Multivariate Data. Chapman & Hall.

Schafer JL (2001) Multiple imputation with PAN. In Collins LM, Sayer AG (Eds) New Methods for Analysis of Change. American Psychological Association; Chapter 12: 357-77.

Schluchter MD (1992) Methods for the analysis of informatively censored longitudinal data. Statistics in Medicine; 11: 1861-70.

Schluchter MD, Greene T, Beck GJ (2001) Analysis of change in the presence of informative censoring: application to a longditudinal clinical trial of progressive renal disease. Statistics in Medicine; 20: 989-1007.

Schluchter MD, Konstan MW, Davis PB (2002) Jointly modelling the relationship between survival and pulmonary function in cystic fibrosis patients. Statistics in Medicine; 21: 1271-87.

Schumacher M, Olschewski M, Schulgen G (1991) Assessment of quality-of-life in clinical trials. Statistics in Medicine; 10: 1915-30.

Schwartz CE, Cole BF, Gelber RD (1995a) Measuring patient-centered outcomes in neurologic disease: extending the Q-TWiST method. Archives of Neurology; 52: 754-62.

Schwartz CE, Cole BF, Vickrey BG, Gelber RD (1995b) The Q-TWiST approach to assessing health-related quality of life in epilepsy. Quality of Life Research; 4: 135-41.

Shih WJ, Quan H (1997) Testing for treatment differences with dropouts present in clinical trials – a composite approach. Statistics in Medicine; 16: 1225-39.

Simes RJ, Greatorex V, Gebski VJ (1998) Practical approaches to minimize problems with missing quality of life data. Statistics in Medicine; 17: 725-37.

Singer JD (1997) Using SAS PROC MIXED to fit multilevel models, hierarchical models and individual growth models. The Journal of Educational and Behavioral Statistics.

Spiegelhalter DJ, Thomas A, Best NG, Gilks W (1995) BUGS Manual Version 0.5 Vol(i). MRC Biostatistics Unit, Cambridge.

Spiegelhalter D, Myles J, Jones D, Abrams K (1999) An introduction to Bayesian methods in health technology assessment. British Medical Journal; 319: 508-512.

Spiegelhalter DJ, Thomas A, Best NG (2000) WinBUGS Version 1.3 user manual. MRC Biostatistics Unit, Cambridge.

Spiegelhalter DJ, Abrams KR, Myles JP (2003) Bayesian Approaches to Clinical Trials and Health-Care Evaluation. Wiley (in press).

Staquet MJ, Hays RD, Fayers PM (Eds) (1998) Quality of Life Assessment in Clinical Trials: Methods and Practice. New York: Oxford University Press.

StatCorp (2002) Stata Statistical Software Release 7.0. College Station, Texas; Stata Corporation, 2002.

Stephens RJ, Hopwood P (1995) Defining and analysing the palliation of physical symptoms: a deceptively difficult exercise. Palliative Medicine; 9: 63-75.

Stephens RJ, Hopwood P, Girling DJ (1999) Defining and analysing symptom palliation in cancer clinical trials: a deceptively difficult exercise. British Journal of Cancer; 79: 538-44.

Sypsa V, Touloumi G, Kenward M, Karafoulidou A, Hatzakis A (2001) Comparison of smoothing techiniques for CD4 data in a Markov model with states defined by CD4: an example on the estimation of the HIV incubation time distribution. Statistics in Medicine; 20: 3667-76.

Tandon PK (1990) Applications of global statistics in analyzing quality of life data. Statistics in Medicine; 9: 819-27.

Tandon PK, Stander H, Schwarz RP (1989) Analysis of quality of life data from a randomized, placebo-controlled heart failure trial. Journal of Clinical Epidemiology; 42: 955-62.

Taylor JMG, Wang Y (2002) Surrogate markers and joint models for longitudinal and survival data. Controlled Clinical Trials; 23: 626-34.

Till JE, de Haes JCJM (1991) Quality-adjusted survival analysis. Journal of Clinical Oncology; 9: 525-26.

Torrance GW (1986) Measurement of health state utilities for economic appraisal. Journal of Health Economics; 5: 1-30.

Torrance GW (1987) Utility approach to measuring health-related quality of life. Journal of Chronic Diseases; 40(6): 593-600.

Touloumi G, Pocock SJ, Babiker AG, Darbyshire JH (1999) Estimation and comparison of rates of change in longitudinal studies with informative drop-outs. Statistics in Medicine; 18: 1215-33.

Troxel AB, Fairclough DL, Curran D, Hahn EA (1998) Statistical analysis of quality of life with missing data in cancer clinical trials. Statistics in Medicine; 17: 653-66.

Tsiatis AA, Davidian M (2001) A semiparametric estimator for the proportional hazards model with longitudinal covariates measured with error. Biometrika; 88: 447-58.

Tsiatis AA, de Gruttola V, Wulfsohn MS (1995) Modelling the relationship of survival to longitudinal data measured with error: applications to survival and CD4 counts in patients with AIDS. Journal of the American Statistical Association; 90: 27-37.

Von Neumann J, Morgenstern O (1953) Theory of Games and Economic Behaviour, 3rd Edn. New York: Wiley.

Waldron D, O'Boyle CA, Kearney M, Moriarty M, Carney D (1999) Quality of life measurement in advanced cancer: assessing the individual. Journal of Clinical Oncology; 17: 3603-11.

Wang Y, Taylor JMG (2001) Jointly modelling longitudinal and event time data with application to acquired immunodeficiency syndrome. Journal of the American Statistical Association; 96: 895-905.

Wang C, Douglas J, Anderson S (2002) Item response models for joint analysis of quality of life and survival. Statistics in Medicine; 21: 129-142.

Willemse PHB, van Lith J, Mulder NH, Aalders JG, Bouma J, de Vries EGE, Sleijfer DT (1990) Risks and benefits of cisplatin in ovarian cancer: a quality- adjusted survival analysis. European Journal of Cancer; 26: 345-52.

Willemse PHB, de Vries EGE, Kloppenburg M, Fontein DL, Aalders JG, Boonstra H, Mulder NH, Sleijfer DT (1992) Carboplatin with cyclophosphamide in patients with advanced ovarian cancer: an efficacy and quality-adjusted survival analysis. International Journal of Gynecological Cancer; 2: 236-43.

World Health Organization (1947) The constitution of the World Health Organization. WHO Chronicle; 1: 29.

Wu MC, Bailey K (1988) Analysing changes in the presence of informative right censoring caused by death and withdrawal. Statistics in Medicine; 7: 337-46.

Wu MC, Bailey KR (1989) Estimation and comparison of changes in the presence of informative right censoring: conditional linear model. Biometrics; 45: 939-55.

Wu MC, Carroll RJ (1988) Estimation and comparison of changes in the presence of informative right censoring by modelling the censoring process. Biometrics; 44: 175-88.

Wu S-C (1982) A semi-Markov model for survival data with covariates. Mathematical Biosciences; 60: 197-206.

Wulfsohn MS, Tsiatis AA (1997) A joint model for survival and longitudinal data measured with error. Biometrics; 53: 330-39.

Xu J, Zeger SL (2001) Joint analysis of longitudinal data compromising repeated measures and time to events. Applied Statistics; 50: 375-387.

Young PJ, Weeden S, Kirwan JR (1999) The analysis of a bivariate multi-state Markov transition model for rheumatoid arthritis with an incomplete disease history. Statistics in Medicine; 18: 1677-90.

Zhang J, Quan H, Ng J, Stepanavage ME (1997) Some statistical methods for multiple endpoints in clinical trials. Controlled Clinical Trials; 18: 204-21.

Zwinderman AH (1992) Statistical analysis of longitudinal quality of life data with missing measurements. Quality of Life Research; 1: 219-24.