Incomplete data in event history analysis

Christopher Julian Sutton[1]

A thesis submitted in partial fulfilment of the
requirements of the University of Leicester for
the degree of Doctor of Philosophy

Faculty of Science

October 1996

[1]Supervisors: Dr. M. J. Phillips and Dr. C. Jagger

UMI Number: U087270

All rights reserved

INFORMATION TO ALL USERS
The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript
and there are missing pages, these will be noted. Also, if material had to be removed,
a note will indicate the deletion.

UMI U087270
Published by ProQuest LLC 2015. Copyright in the Dissertation held by the Author.
Microform Edition © ProQuest LLC.
All rights reserved. This work is protected against
unauthorized copying under Title 17, United States Code.

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

X753639911

To my mother

and the

memory of my father

Acknowledgements

I wish to thank Michael and Carol for their unstinting guidance, support and good humour throughout the research that led to this thesis. I would also like to thank the many people who have provided computer support during the past seven years, both in Leicester and Preston, but especially John Woods and Tony Willows for their advice and help in overcoming the seemingly endless problems I encountered, not *all* of which were self-inflicted.

I would also like to thank my family, friends and colleagues for their patience, support and interest. Special thanks go to Anna Hart, for her help, advice and reassurance, Mike Grannell for his support and encouragement and Ray Harris for his backing during the early days of my work.

Thanks also to the University of Central Lancashire (and also its former guise of Lancashire Polytechnic) for the provision of financial assistance and computing facilities, and to the University of Leicester for the provision of computing facilities.

Christopher Julian Sutton

Incomplete data in event history analysis

Abstract

Incomplete data present a serious problem in the modelling of event histories. Two particular forms of incompleteness are in evidence for data of this form. The first is due to recording of event times in interval-censored form. For single non-repeatable events this can be accommodated by using methods for modelling grouped survival times, such as those of Prentice and Gloeckler (1978) and Finkelstein (1986). The other, more serious, problem relates to incomplete recording of follow-up measurements which would typically be included as time-dependent covariates in survival models. A number of methods exist for handling incomplete data. These include multiple imputation for variables subject to incompleteness and the application of iterative algorithms such as EM and the data augmentation algorithm.

In this thesis, a method for handling both these types of incompleteness is derived based on multiple imputation combined with an adaptation of Finkelstein's method to handle time-varying covariates. This method is then investigated via Monte Carlo simulation and applied to data arising from the annual screening of those aged 75 years and over in the town of Melton Mowbray, as performed through the local general practice. Its performance is compared with that of more traditional approaches to modelling data collected in studies of this type. It is shown that parameter estimates can be considerably affected by the choice of approach to modelling. Whilst there are some problems with the implementation of this technique, particularly with reference to the model for the multiple imputation of the repeated risk factor values, it shows promise for application to studies of this form, particularly if combined with improved models for multiple imputations. The data from the annual screenings are assumed missing at random, but the techniques used could be extended to cover non-ignorable missing data mechanisms of known form.

Contents

Chapter 1

Introduction

Incomplete data in event history analysis is a topic which has aspects of repeated measurements, due to the accumulation of covariate information from an individual's 'history', survival data, due to the interest in one or more 'events', and missing data, because the information on both the covariate history and, sometimes, the survival times is incomplete. Whilst there is extensive literature on each of the topics of *repeated measures data*, *survival data* and *missing data* individually, that dealing with their intersection is considerably more limited. This thesis covers these three subject areas.

Sections 1.1-1.3 briefly review the extensive literature on the three areas. The remainder of the thesis aims to draw together these three areas with particular emphasis of the analysis of event history data subject to various forms of incompleteness. The practical rationale for the coverage of this theoretical area is information accruing from health checks, regular screenings or long-term follow-ups. Examples of relevant types of study include the annual screening of the elderly by General Practitioners and the Framingham heart study.

Particular emphasis is placed on the analysis of data resulting from the General Practitioners' screening of the elderly in the United Kingdom. The offer of such

screening is now compulsory under the revision of the GPs' contracts in 1990. The screening is offered triennially to patients between the ages of 65 and 74 and annually from the age of 75.

In assessing the effect of various health measures on survival or other changes of state, it is desirable to include any relevant information on changing values of risk factors. Many procedures currently widely applied to event history data fail to make any use of the changing risk factors whilst few attempt to make full and appropriate use of these. A variety of methods presently applied to event history data with repeated risk factor measurements will be reviewed in Chapter 2. Some of those applied in growth curve analysis and survival analysis will then be investigated in detail in Chapters 3, 4 and 5 with a view to their application to event history data with incomplete recording of risk factors and event times.

1.1 Repeated Measures Data

Repeated measures is a general term for data consisting of two or more measurements of one or more characteristics on the same experimental unit. Many designs of experiments and studies result in the collection of such data, ranging from a simple paired design (analysed using a paired t-test in the case of normally distributed data) to designs where there are several measurements of a number of 'response' variables of a variety of types (eg. categorical, continuous). Longitudinal studies monitoring the time to death may be viewed as a special form of the latter.

The resulting data are usually in the form of a three-dimensional array, consisting of T 'measurements' of p characteristics on n experimental or observational

units. In many, but not all designs, either $T = 1$ or $p = 1$. If $T = 1$ the data are not 'repeated measures' and if $p = 1$, the data are not 'multivariate' in the usual sense. These data are usually viewed as *responses* and therefore y_{ijk} will be used to denote the j^{th} repeated measurement of the k^{th} characteristic on the i^{th} unit.

Much of the data and resulting methodology arising from longitudinal studies can be classified as **quantitative repeated measures**. Methods of analysis range from simple univariate and graphical methods to mathematically sophisticated techniques. A number of books and review papers have been published in recent years. Most of these have concentrated largely or entirely on quantitative data under the assumption of normality. More recently, these methods have been extended to cover other members of the exponential family of distributions. For a comprehensive coverage the reader is directed to the books by Crowder and Hand (1990), Lindsey (1993) and Diggle, Liang and Zeger (1994).

1.1.1 Basic methods

As with all data analysis, the first step should be to use some appropriate diagrammatic representation. This will help the analyst visualise the structure of the data and, unless suggested by some previously known theory, lead to the proposal of one or more models. The standard technique used is to plot the individuals' profiles against time. Plots of response, corrected for time trends, against potentially important explanatory variables are often useful for identifying the form, if any, of the relationship. Another useful form of plot is that of the individuals' residual profiles following removal of the mean profile. Such a plot can be useful in highlighting various features of individuals such as consistently high or low profiles, a feature known as **tracking**. Non-parametric curve fitting methods such as kernel estimation, splines and lowess may be used to highlight the underlying

pattern of the mean over time. This has the advantage over individual profiles that the diagram is less cluttered.

Although, on occasions, repeated t-tests or other basic statistical methods are applied at each time point, these tests are not independent and therefore the usual forms of drawing inferences are invalid. An improvement to repeated t-tests is the use of **response feature analysis**. This technique involves summarising a small number of important features of an individual's profile and performing a univariate analysis on each. Response feature analysis has the advantage that the experimenter must consider which is the most important feature of a profile or, on occasions, the two or three most important features. The most commonly used response feature is the **area under the curve** (AUC). Other frequently used features are the maximum (or minimum) value, the time to maximum (or minimum) response, the regression coefficient, the final value, the increase (or decrease) or percentage change between first and last value and time to reach a particular value. The choice of feature should always depend on the question, or questions, of interest to the investigator. Important additional advantages of the use of response features analysis are that they result in the use of statistically valid techniques and, to a certain degree, can accommodate irregular or missing observations (Matthews, 1993).

1.1.2 Analysis of Variance Methods

Standard analysis of variance techniques, incorporating both fixed and random effects, can be applied to repeated measures data in certain circumstances. However, the F-tests used are only valid if certain conditions hold. These conditions are the usual condition of normality of the response variable together with that of **sphericity** of the covariance matrix of $p-1$ orthogonal contrasts between the

p repeated measurements. One way of specifiying this condition is that, for two measurements X_j and X_k, $\mathrm{Var}(X_j - X_k)$ is constant for all j and k (Crowder and Hand, 1990). A sufficient condition, however, is that the covariance matrix of the repeated measures has **compound symmetry**. A covariance matrix Σ for a vector of repeated measures $\underline{\boldsymbol{Y}}$ is said to have *compound symmetry* if:

$$\begin{aligned} \mathrm{Var}(Y_j) &= \sigma^2 & \forall j \\ \mathrm{Cov}(Y_j, Y_k) &= \sigma^2 + \sigma_p^2 & \forall j,\, k \quad \text{s.t.} \quad j \neq k\,. \end{aligned}$$

Various correction factors for the univariate F-tests have been widely employed. These adjust the degrees of freedom of both numerator and denominator by the same correction prior to computation of the p-value for each test. An alternative approach, which has the main advantage that no assumption is made about the form of the covariance matrix, is that of multivariate analysis.

In situations where a t-test (either one-sample or two-sample) would be used if the data consisted of a single measurement, **Hotelling's T^2** is usually applied. In the single group case, the test statistic for testing $\mathcal{H}_0 : \underline{\boldsymbol{\mu}} = \underline{\boldsymbol{\mu}}_0$ has the form

$$T^2 = n(\underline{\bar{\boldsymbol{y}}} - \underline{\boldsymbol{\mu}}_0)^T \boldsymbol{S}^{-1} (\underline{\bar{\boldsymbol{y}}} - \underline{\boldsymbol{\mu}}_0)$$

where n is the sample size, $\underline{\bar{\boldsymbol{y}}}$ is the sample mean and $\boldsymbol{S}$ is the sample covariance matrix.

In the two group case the form when testing $\mathcal{H}_0 : \underline{\boldsymbol{\mu}}_1 = \underline{\boldsymbol{\mu}}_2$ is

$$T^2 = (n_1^{-1} + n_2^{-1})(\underline{\bar{\boldsymbol{y}}}_1 - \underline{\bar{\boldsymbol{y}}}_2)^T \boldsymbol{S}^{-1} (\underline{\bar{\boldsymbol{y}}}_1 - \underline{\bar{\boldsymbol{y}}}_2)$$

where n_1 and n_2 are the samples sizes for the two groups, $\underline{\bar{\boldsymbol{y}}}_1$ and $\underline{\bar{\boldsymbol{y}}}_2$ are the respective sample mean vectors and $\boldsymbol{S}$ is the pooled covariance matrix. In cases where ANOVA would be used in the univariate case, the extension to multivariate

analysis of variance (MANOVA) is appropriate. Various test criteria have been developed based on the eigenvalues of the appropriate sums of squares and products matrices. These include Wilk's lambda, Roy's largest root and the Lawley-Hotelling trace.

1.1.3 General Linear Modelling

As in the case of a univariate response, the traditional ANOVA and regression methods can be unified under a linear modelling framework. The model is known as the **general linear model** where the observed vector of responses $\underline{\boldsymbol{y}}$ has mean vector $\underline{\boldsymbol{\mu}} = \boldsymbol{X}\underline{\boldsymbol{\beta}}$ and covariance matrix $\boldsymbol{\Sigma}$, where $\boldsymbol{X}$ is a matrix of explanatory variables (potentially including indicator variables) and $\underline{\boldsymbol{\beta}}$ is a *fixed* but *unknown* vector of parameters. The simplest situation, corresponding to the MANOVA mentioned in Section 1.1.2, has an unstructured form for $\boldsymbol{\Sigma}$. The methods of **generalised least squares** and **maximum likelihood** both lead to the estimator

$$\underline{\hat{\boldsymbol{\beta}}} = \left(\sum_{i=1}^{n} \boldsymbol{X}_i{}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{X}_i\right)^{-1} \left(\sum_{i=1}^{n} \boldsymbol{X}_i{}^T \boldsymbol{\Sigma}^{-1} \underline{\boldsymbol{Y}_i}\right)$$

where $\boldsymbol{\Sigma}$ is assumed to be known. In practice, $\boldsymbol{\Sigma}$ is unknown and replaced by an appropriate estimator. This model can be extended to cover missing values in the $\underline{\boldsymbol{Y}}_i$ and estimated using maximum likelihood, although it is necessary to use an iterative procedure to estimate $\underline{\boldsymbol{\beta}}$ and $\boldsymbol{\Sigma}$.

The general linear model can be extended to cover structured covariance matrices. If there are p repeated measurements on n individuals, rather than estimating $\frac{1}{2}p(p+1)$ covariance parameters which leads to inefficient estimation unless n is

large compared with p, an appropriate structure may be assumed for $\boldsymbol{\Sigma}$.

Both the independence model, with $\boldsymbol{\Sigma} = \sigma^2 \boldsymbol{I}$, and the compound symmetry model with $\boldsymbol{\Sigma} = \sigma^2 \boldsymbol{I} + \sigma_p^2 \boldsymbol{J}$ and previously discussed in Section 1.1.2, are simple forms of a general linear model with structured covariance matrix. Other possible forms include the equicorrelated model, where $\boldsymbol{\Sigma} = \boldsymbol{S}\left[(1-\rho)\,\boldsymbol{I} + \rho \boldsymbol{J}\right]\boldsymbol{S}$ with $\boldsymbol{S} = \text{diag}(\sigma_1, \ldots, \sigma_p)$, autoregressive models, for example the first-order autoregressive process with $\boldsymbol{\Sigma} = (\sigma_{ij})$ where $\sigma_{ij} = \sigma^2 \rho^{|i-j|}$ and a random effects model where $\boldsymbol{\Sigma} = \boldsymbol{Z}\boldsymbol{D}\boldsymbol{Z}^T + \sigma^2 \boldsymbol{I}$ where $\boldsymbol{Z}$ is a known matrix of explanatory variables and $\boldsymbol{D}$ is the covariance matrix of the subject-specific random effects. This random effects model has been widely applied over recent years. Laird and Ware (1982) discussed a general family of such models, using a combination of empirical Bayes and maximum likelihood estimation via the EM algorithm (see Section 1.3.2). It is commonly used in growth curve modelling where each individual is assumed to have its own, often linear, growth curve over time and where each individual's parameter vector is assumed to be a random vector from a distribution, usually multivariate normal, with unknown but fixed mean vector and covariance matrix. Other similar random effects models can be specified using multilevel modelling techniques (Goldstein, 1995). In growth curve modelling, models of the family defined by Laird and Ware (1982) are usually known as the **random coefficients model** or a **two-stage random-effects model**.

Jennrich and Schluchter (1985) considered the computation of maximum likelihood estimates under a model for which the expected responses are described as 'arbitrary functions of unknown regression parameters' and the within-subject covariances are modelled as 'arbitrary functions of a set of unknown covariance parameters'. This, in fact, is a very general form of covariance structure, incorporating all the specific structures described in the preceding paragraph. Three algorithms were proposed for computing the estimates: Newton-Raphson, Fisher scoring and a generalised expectation maximisation (EM) algorithm (see Section

1.3.2) combined with scoring. The last of these is used only with incomplete data and is particularly useful if there is a large number of parameters. Jennrich and Schluchter (1985) suggested that the use of the inverse of the empirical information matrix may be preferable to the use of the Fisher (expected) information matrix with some forms of missingness. It is stated (Jennrich and Schluchter, 1985) that, despite their methods not explicitly considering the missing data mechanism, they are valid providing the mechanism is ignorable.

An interesting application of a two-stage random effects model was given by Hughes and Pocock (1992). They modelled diastolic blood pressure X_t at time t for individuals amongst men free, at the start of the study, of diagnosed cardiovascular disease and living in one of the cities of Edinburgh, Budapest and Prague. For their modelling they selected those 11299 men randomised into the study who attended for four consecutive blood pressure screens during the eight year period of the study. They were interested in assessing the variability in annual measurements within an indivdual and the resultant accuracy of categorisation into the traditional categories of 'mild', 'moderate' and 'severe' hypertension, corresponding to diastolic pressure ranges of 90-105, 105-120 and >120 mmHg respectively. The model they fitted had:

$$X_t \sim \mathcal{N}(\xi_t, \sigma_W^2(\mu))$$

where $\xi_t = \mu + \beta(t - \bar{t})$ and $\sigma_W(\mu) = \sigma_W [1 + \gamma (\mu - \bar{\mu})]$

for the within-subject effects and

$$\mu \sim \mathcal{N}(\mu^*, \sigma^2) \qquad \text{and} \qquad \beta \sim \mathcal{N}(\beta^*, \sigma_\beta^2) .$$

Here, μ is the underlying mean pressure for an individual during the four year period of the study, β is the rate of change per year in the blood pressure for the

same individual. It can be seen that the within-individual variance is dependent on their underlying mean level, individuals with higher mean levels also having higher within-subject variance (as the parameter γ is positive). They also found that this model, fitted using maximum likelihood methodology, fitted their data well. They found within-individual standard deviations of magnitude comparable with previous studies. The estimate of σ_W was 6.5 mm Hg in Edinburgh, 7.4 mm Hg in Budapest and 7.1 mm Hg in Prague. The authors attributed much of the estimated differences to the recording of blood pressures to different degrees of accuracy in the different cities.

Modelling growth curve data subject to individuals dropping out prior to completion of the study or experiment involves repeated measures analysis in the presence of missing data and will be covered briefly in Section 1.3.2 and then in more detail in Chapter 3. Moreover, such a data structure has much in common with certain aspects of the analysis of interval-censored survival times (see Chapter 4) and, as such, has potential for modelling data arising from screening studies.

1.1.4 Generalised linear modelling

As with linear models, generalised linear models can be extended to handle repeated measures data. The main difference with models with a non-linear link function is that mis-specification of the covariance structure is a far more serious problem because the parameter estimates for the linear predictor have different interpretations under different assumptions about this structure. Diggle, Liang and Zeger (1994) provide a detailed discussion of three different extensions of generalised linear models to handle repeated measures data. This subsection will give a relatively brief explanation of these three approaches, explaining the differences in their interpretations.

The first type of model is known as a **marginal model**. This is because the marginal expectation of the response is modelled separately from the covariance structure. Such models should be used when interest is concentrated on population mean effects rather than effects relating to individuals. They correspond directly to the extension of the general linear model for normally distributed responses to cover responses from the exponential family of distributions.

The second form of model is a **random effects model**. This is simply an extension of the random effects model, introduced in the preceding section, to responses from the exponential family. In these models it is assumed that the correlation between repeated measurements on an individual is due entirely to their own 'random effect' or, equivalently, that the repeated measurements are *conditionally independent*, given their particular effect. Random effects models are of particular use when the aim is to make inferences at an individual rather than population level (Diggle, Liang and Zeger, 1994).

The final type of model is a **transition** (or **Markov**) model. A transition model is specified in terms of the distribution of the present response *conditional on the preceding responses*. Therefore, previous responses enter the generalised linear model as additional predictors.

In general linear modelling using either random effects or transition models, the regression parameters $\underline{\beta}$ may also be interpreted marginally as the use of the identity link function leads to the expected response equating to the linear predictor in both cases. However, when the models are subject to a non-linear link function, this convenience does not apply. Relationships between the parameters of marginal models and those of random effects models can be established, but those between marginal model parameters and transition model parameters only exist in certain circumstances (Diggle, Liang and Zeger, 1994). Hence, it is vital to correctly specify the form of model when considering repeated realisations of

dichotomous, categorical or count responses.

For random effects models, traditional maximum likelihood methods can be used for parameter estimation. However, as the likelihood function involves integration over the unknown random effects, numerical methods are often needed in order to evaluate the likelihood. Maximum likelihood estimation is somewhat problematical for transition models as the marginal distribution of the first response y_{i1} is not always determinable without additional assumptions. For this reason, a conditional approach is often taken, the conditioning being on the problematical y_{i1}. However, with the marginal model, problems with the likelihood function are even greater as only the first two moments are specified and, for some non-normal distributions, this is not sufficient to fully specify the likelihood.

A **quasi-likelihood** approach is, therefore, often taken for parameter estimation in the marginal model. This involves use of a version of the quasi-score function

$$\underline{\boldsymbol{S}}(\underline{\boldsymbol{\beta}}) = \sum_{i=1}^{T} \left(\frac{\partial \underline{\boldsymbol{\mu}}_i}{\partial \underline{\boldsymbol{\beta}}} \right)^T \boldsymbol{\Sigma}_i^{-1} \left(\underline{\boldsymbol{Y}}_i - \underline{\boldsymbol{\mu}}_i \right) = \underline{\mathbf{0}}$$

(where $\boldsymbol{\Sigma}_i = \mathrm{Var}(\underline{\boldsymbol{Y}}_i)$) for the estimation of the regression coefficients, $\underline{\boldsymbol{\beta}}$, and in which the parameters specifying $\boldsymbol{\Sigma}$ are replaced by appropriate estimates. These equations are known as **generalised estimating equations (GEEs)** and may also be used for fitting transition models (Diggle, Liang and Zeger, 1994).

The theory behind the development of GEEs is given in Liang and Zeger (1986). With regards to missing data values, it is stated that if the between-individual correlation structure is mis-specified, the data must be **missing completely at random** (see Section 1.3) for the parameter estimation to be consistent but that certain forms of **missing at random** data (see Section 1.3) will lead to consistent estimates for normal or dichotomous responses. Zeger and Liang (1986) illustrated the use of GEEs in fitting a marginal model to data from a study of

the association between mothers' stress and children's mortality, showing that the qualitative conclusions were the same for three different forms of correlation matrix, with little difference in the parameter estimates. The model was applied under the assumption that the logit of the probability that a child was ill on day t was a linear function of the mother's stress on the previous three days and certain time independent covariates. It was also shown, by grossly mis-specifying the correlation structure, that, overall, the inferences were not very sensitive to correct specification of the correlation matrices. In fact, the only parameter estimates that appeared even moderately sensitive to the mis-specification were the time dependent mother's stress covariates and the authors believed that this was due to the limited number of women in the study. Also noted was that the sensitivity of inferences about $\underline{\beta}$ to mis-specification of $\boldsymbol{R}$ is likely also to depend on 'the degree and pattern of incomplete data', although that was not a problem in this application.

Zhao and Prentice (1990) developed a variant on GEEs in which, rather than estimating Σ and $\underline{\beta}$ separately they are estimated jointly. Liang, Zeger and Qaqish (1992) compared the original GEE approach, which they termed GEE1, with Zhao and Prentice's variant, which they termed GEE2. They concluded that GEE1, but not GEE2, is consistent if Σ is mis-specified. However, GEE2 was found to be the more efficient if the correct specification is used for Σ. Moreover, if interest is centred on Σ, it is stated that GEE2 should be used in conjunction with sensitivity analysis to check how inference on $\underline{\beta}$ is affected by changes in the specification of Σ.

Under any of these three modelling strategies, logistic models for dichotomous responses and log-linear models for count data can be fitted. A particularly useful transition model for ordered categorical data, based on the **proportional odds model** (McCullagh, 1980) may be developed. In the case where there are C ordered categories, the model takes the form:

$$\log \frac{\Pr\left(Y_{ij} \leq b \mid Y_{ij-1} = a\right)}{\Pr\left(Y_{ij} > b \mid Y_{ij-1} = a\right)} = \theta_{ab} + \underline{\boldsymbol{x}}_{ij}^T \underline{\boldsymbol{\beta}}_a$$

$$a = 0, \ldots C-1, \qquad b = 0, \ldots, C-2\,,$$

where $\underline{\boldsymbol{x}}_{ij}$ is the covariate vector for the i^{th} individual at the j^{th} measurement time, θ_{ab} is a transition-specific parameter and $\underline{\boldsymbol{\beta}}_a$ is a parameter vector which allows the covariate effects to vary depending on the state at the previous measurement, Y_{ij-1}.

This model can be fitted using conditional maximum likelihood or, almost fully efficiently, using GEEs to simultaneously fit logistic models to dummy variables Y^*_{ijk} where

$$Y^*_{ijk} = \begin{cases} 1 & \text{if } Y_{ij} \leq k \\ 0 & \text{if } Y_{ij} > k \end{cases}$$

(Diggle, Liang and Zeger, 1994).

1.2 Survival Data

Survival data are characterised by the observation of the times to occurrence of one (or more) events. In some cases, these events are potentially *repeatable* (eg. heart attacks, replacement of artificial hip joints). In others, there may be several events of interest, each corresponding to a separate change of state. This thesis will largely concentrate on the observation of times to a single, non-repeatable change of state or event.

Censoring is a general term covering cases in which the exact time to the occurence of the event is not observed. *Right-cenoring* is where the occurrence (possibly hypothetical) is *after* the time recorded, *left-censoring* is where the occurrence is *before* the time recorded and *interval-censoring* is where the occurrence is *between* two recorded times. By far the commonest of these is *right-censoring*. This is usually caused by the termination of a study (or experiment) or the loss of individual units due to some known or unknown cause. Censored survival times are a special form of *missing data* and, as such, will be covered further in Section 1.3.

To attempt to distiguish between the precise times of an event of interest and simply the recorded time (or interval of time), the former will be referred to as the **failure time** and the latter as the **survival time**. As outlined below, time measurements will always be rounded to a certain degree, so it may be argued that survival data always consist of survival times, recorded as intervals, rather than failure times.

In survival studies subject to right-censoring, survival times are usually denoted by an ordered pair (t_i, d_i) where t_i is the recorded survival time and d_i is a variable indicating whether the time t_i represents the observation of the event of interest ($d_i = 1$) or *right-censoring* ($d_i = 0$).

Interval-censored survival times are frequent occurrences in longitudinal studies and experiments in which the units are only followed-up periodically to assess their survival status. In many such studies, it is likely that other characteristics will be measured or assessed at these follow-up times. There is therefore a need to incorporate any information on repeated assessment of potential risk factors in an appropriate manner when modelling survival data. A general definition of an **interval-censored** survival time is one where the failure is known only to have occurred within an interval of time $(L_i, R_i]$. Left- and right-censored data

are, in fact, special cases of interval-censoring, in which the intervals are $(0, R_i]$ and (L_i, ∞) respectively. Another special case of interval censoring is where the failure times are grouped. In practice, failure times are grouped in all survival studies as there is always a certain degree of rounding. In most cases, however, the rounding is not sufficiently coarse to cause a substantial number of ties and therefore standard survival models are employed. If the rounding is coarse, there are likely to be many multiple ties and methods derived specifically for analysing grouped or interval-censored survival data are necessary. It should be noted that the intervals may themselves be random in cases where they are not defined at the start of the study. With *grouped* data, however, the accuracy with which the failure times are recorded will be decided at the design stage and the intervals are therefore fixed.

There are two functions of particular interest when describing or modelling survival data. These are the **survivor function** $S(t)$ and the **hazard function** $\lambda(t)$ and are defined as follows:

$$\begin{aligned} S(t) &= \mathrm{P}(T \geq t) \\ \lambda(t) &= -\frac{\mathrm{d}}{\mathrm{dt}}\left[\log S(t)\right] = \lim_{\delta t \to 0} \frac{\mathrm{P}\left(t \leq T < t + \delta t \mid T \geq t\right)}{\delta t}. \end{aligned}$$

The survivor function is the probability of surviving to time t and the hazard function is the rate of death at time t conditional upon survival to that point in time.

There are three main approaches to the analysis of survival data, namely the use of **non-parametric methods**, **parametric methods** and **semi-parametric methods**. The remainder of this section will provide a brief overview of the most important aspects of and methodologies for these approaches.

1.2.1 Non-parametric methods

Methods for estimating or comparing hazard or survivor functions for groups of individuals that do not rely upon specification of the form of the distribution are termed *non-parametric*. Many of these methods evolved during the first part of this century, prior to widespread use of computers. However, non-parametric methods are still applied in instances when it is preferred not to make specific and potentially invalid assumptions about the underlying distribution of survival times.

Most non-parametric methods were derived to estimate or compare survivor functions, although similar approaches can be used to estimate hazard functions. The most widely applied non-parametric estimation method in medical studies is the **Kaplan-Meier** (or product-limit) **estimate** (Kaplan and Meier, 1958) as it allows right-censored observations. Supposing there are n individuals in the study at the start and that, having ordered their death and censoring times, t_i there are k distinct death times, $t_{(1)}, \ldots, t_{(k)}$. Then, supposing that there are $n_{(j)}$ individuals at risk of death at time $t_{(j)}$ (including any indivduals censored at that precise time) and $d_{(j)}$ individuals who actually die at that time, the Kaplan-Meier estimate is defined as:

$$\hat{S}(t) = \prod_{j=1}^{r} \left(\frac{n_{(j)} - d_{(j)}}{n_{(j)}} \right) \qquad \text{for} \quad t_{(r)} \leq t < t_{(r+1)}$$

where $t_{(k+1)} = \infty$. The usual estimate of the variance of the Kaplan-Meier estimate is given by Greenwood's formula

$$\text{Var}(\hat{S}(t)) \simeq \left[\hat{S}(t)\right]^2 \left[\sum_{j=1}^{r} \frac{d_{(j)}}{n_{(j)}\left(n_{(j)} - d_{(j)}\right)}\right] \qquad \text{for} \quad t_{(r)} \leq t < t_{(r+1)}$$

which can be used, if desired, to give symmetric confidence limits for the survivor function or, alternatively, may be used in conjunction with a transformation and the usual first order Taylor's series approximation for a variance of a function of a random variable to give asymmetric confidence limits of a desired form.

1.2.2 Parametric methods

Parametric methods for survival analysis involve full specification of the survivor or hazard function. The two most widely applied parametric methods are based around the **exponential distribution** and the **Weibull distribution**. The exponential distribution has

$$\begin{aligned} \lambda(t) &= \lambda \\ S(t) &= \lambda \exp(-\lambda t) \end{aligned}$$

and the Weibull distribution has

$$\begin{aligned} \lambda(t) &= \lambda\gamma t^{\gamma-1} \\ S(t) &= \exp(-\lambda t^{\gamma}) \end{aligned}$$

where γ is a *shape* parameter and λ is a *scale* parameter.

Because of the assumption that the hazard is constant, the exponential distribution is rarely applicable in its basic form. For this reason, only the **piecewise**

exponential distribution, where the hazard is only assumed constant within distinct intervals of time, and the Weibull distribution, where the hazard is changing with time in a well-defined way, are frequently used in practice. These distributions may be fitted to data using maximum likelihood methods. Covariates may be adjusted for using either method, usually incorporating an exponential relationship between hazard and covariates.

It should be noted that the exponential, Weibull and piecewise exponential distributions satisfy the **proportional hazards** property, in that the ratio of the hazards of two individuals is independent of time, providing any covariates included in the model specification are time independent.

A useful informal method of checking the appropriateness of the assumption of a Weibull model is to estimate the survivor function, usually using Kaplan-Meier, and plotting $\log(-\log \hat{S}(t))$ against $\log t$. If the Weibull model is appropriate, the plot should approximate a straight line. Moreover, if the slope of the line is about 1, the exponential distribution will provide a simpler model for the data.

However, as the distributions specified in this section are very limited, other more flexible distributions have been applied. However, in most medical studies, the precise form of the relationship between time and survival is not of paramount interest, the main interest being the relationship between survival and particular covariates. For this reason, parametric methods are not often applied due to the greater flexibility of the semi-parametric Cox model, as described in Section 1.2.3

1.2.3 Semi-parametric methods

In arguably the most important journal publication in the last 25 years, Cox (1972) led to the application of a proportional hazards model under a semi-parametric

specification. In particular, the model for the hazard function is of the form:

$$\lambda(t) = \lambda_0(t)\,\psi(\underline{\boldsymbol{\beta}})$$

where $\psi(\underline{\boldsymbol{\beta}})$ is a parametric model, where, usually, $\psi(\underline{\boldsymbol{\beta}}) = \exp(\underline{\boldsymbol{x}}^T\underline{\boldsymbol{\beta}})$ for covariate vector $\underline{\boldsymbol{x}}$ for the relationship between the covariates and the hazard and $\lambda_0(t)$ is the hazard for an individual with covariate vector $\underline{\mathbf{0}}$, known as the **baseline hazard**, and is not subject to parametric specification.

Cox (1972) proposed a *partial likelihood* which is, assuming distinct failure times, a product of the conditional probabilities of each observed death given exactly one death amongst those at risk at the time of death. Due to the fact these are conditional probabilities at particular points in time, the baseline hazard is the same for all individuals and will therefore cancel. It is known as a *partial* likelihood because it is **not** the probability of any particular event due to the changing nature of the risk set over time, but is merely *part* of the likelihood.

Cox's original model has been developed to allow for time dependent covariates, thus allowing the hazards to be non-proportional in certain specified ways, to allow for different baseline hazards for different strata and for grouped or interval-censored survival times. Extensions to interval-censored survival times with time-dependent covariates will be considered in Chapter 4.

1.3 Missing Data

Following Rubin (1976), it is useful to categorise the missing-data mechanism in the following way, by considering a simple case where there are two characteristics X and Y, each to be observed on n units. These categories extend to cover

more complex data structures. However, as considered in Section 3.1, with more complex data structures the complexity necessitates the sub-division of some of these basic categories.

Suppose the variable X is truly observed on all n units but that Y is observed on $m < n$ units and is *missing* on the remaining $n - m$ units. Then if the probability of non-response is dependent on neither X nor Y then the data are called **missing completely at random (MCAR)**. If the probability of non-response is dependent on X but not Y then the data are called **missing at random (MAR)**. Data which are either MCAR or MAR are subject to non-response mechanisms which may be **ignored**, subject to the additional condition of **distinct parameters (DP)** (described later in this section), when making likelihood-based inferences. Data for which the probability of non-response is dependent on Y (and possibly also on X) and are therefore neither MCAR nor MAR are subject to an **informative** missing data mechanism. The implication of such a mechanism is that, in order to obtain valid inferences for either likelihood-based or non-likelihood-based inferences, the non-response mechanism must be considered. Such a non-response mechanism is, therefore, often referred to as a **non-ignorable** mechanism. Concentrating on likelihood-based inferences, there are clear problems in that inferences will be dependent on the form of the non-response model, but the appropriateness of such a non-response model cannot be tested due to the fact that the values of Y subject to non-response almost always remain unobserved. Only if a random sample of the Y is obtained at a later stage can the appropriateness of the non-response model be tested. An example of such is in a survey subject to item non-response. If a random sample of the item non-respondents is contacted and the relevant item obtained, then the non-response model can be tested. However, by specifying a variety of plausible models for the non-response mechanism, the sensitivity of inferences to the model specification can be assessed. This is known as **sensitivity analysis** and is an important technique in analysing data subject to potentially non-ignorable miss-

ing data mechanisms.

Extending these definitions to situations involving repeated measurements of a response variable, including at least some missing-values, we may form a $T \times 1$ vector of response variables $\underline{\boldsymbol{Y}}$ consisting of T equally spaced observations for each of n individuals. Suppose further that, for each individual, there is a design matrix $\mathbf{X}$ (which may include individual covariates, the design on time and possibly 'earlier' values of the response variable), a vector of unknown risk factor parameters $\underline{\boldsymbol{\beta}}$, a vector of unknown parameters of the non-response process $\underline{\boldsymbol{\alpha}}$ and a matrix $\mathbf{Z}$ of unknown additional covariates for the non-response process. Moreover, in order to consider the non-response process, a $T \times 1$ vector of indicator variables $\underline{\boldsymbol{R}}$ is introduced where:

$$\left.\begin{array}{rcll} R_i & = & 1 & \text{if } Y_i \text{ is observed} \\ R_i & = & 0 & \text{if } Y_i \text{ is missing} \end{array}\right\} \quad i = 1, \ldots, T\,.$$

The parameter vectors $\underline{\boldsymbol{\alpha}}$ and $\underline{\boldsymbol{\beta}}$ are assumed to be *distinct*. This allows the response to depend on $\underline{\boldsymbol{\beta}}$ but not $\underline{\boldsymbol{\alpha}}$ and the non-response process to depend on $\underline{\boldsymbol{\alpha}}$ but not $\underline{\boldsymbol{\beta}}$. If the parameters are not distinct, the following does not hold and the non-response process is informative. The papers by Wu and Carroll (1988) and Wu and Bailey (1988, 1989), discussed in detail in Chapter 3 consider cases where the data satisfy the MAR but not the **distinct parameter** (DP) requirement. The importance of the DP condition is emphasised by Shih (1992) who suggests that 'it may be possible that in certain situations when MAR holds without DP, ignoring the missing data process still leads to consistent but inefficient estimates'.

Given $\underline{\boldsymbol{R}}$, the vector $\underline{\boldsymbol{Y}}$ can be partitioned into two components $\underline{\boldsymbol{Y}}_o$ and $\underline{\boldsymbol{Y}}_m$ where $\underline{\boldsymbol{Y}}_o$ contains the observed part of the vector $\underline{\boldsymbol{Y}}$ ($R_i = 1$) and $\underline{\boldsymbol{Y}}_m$ ($R_i = 0$) the missing part. The respective lengths of the vectors $\underline{\boldsymbol{Y}}_o$ and $\underline{\boldsymbol{Y}}_m$

will vary with the different experimental units. It then follows that the density of the observed data is given by:

$$f(\underline{\boldsymbol{Y}}_o, \underline{\boldsymbol{R}} \mid \mathbf{X}, \mathbf{Z}, \underline{\boldsymbol{\beta}}, \underline{\boldsymbol{\alpha}}) = \int f(\underline{\boldsymbol{Y}} \mid \mathbf{X}, \underline{\boldsymbol{\beta}})\, f(\underline{\boldsymbol{R}} \mid \underline{\boldsymbol{Y}}, \mathbf{X}, \mathbf{Z}, \underline{\boldsymbol{\alpha}})\, d\underline{\boldsymbol{Y}}_m \quad (1.1)$$

where the integration is over the sample space of $\underline{\boldsymbol{Y}}_m$.

If the data are **MCAR** then the density of the missing data indicator $\underline{\boldsymbol{R}}$ is of the form:

$$f(\underline{\boldsymbol{R}} \mid \underline{\boldsymbol{Y}}, \mathbf{X}, \mathbf{Z}, \underline{\boldsymbol{\alpha}}) = f(\underline{\boldsymbol{R}} \mid \mathbf{X}, \mathbf{Z}, \underline{\boldsymbol{\alpha}}) .$$

This simply means that the non-response mechanism does not depend on the vector of responses $\underline{\boldsymbol{Y}}$. From an analysis point of view the only difficulty is how to implement a standard analysis with missing data. Some analysis methods employ casewise deletion for all cases with any missing observations (i.e. use $\underline{\boldsymbol{Y}}_i$ if and only if $\underline{\boldsymbol{Y}}_{oi} = \underline{\boldsymbol{Y}}_i$), others take account of any observed data values (i.e. uses $\underline{\boldsymbol{Y}}_{oi}$). The former approach is usually termed **complete case analysis (CCA)** and the latter termed **available case analysis (ACA)**. Both types of analysis will yield valid inferences if any missing data are **MCAR**.

If the non-response mechanism is *ignorable* its density is of the form

$$f(\underline{\boldsymbol{R}} \mid \underline{\boldsymbol{Y}}, \mathbf{Z}, \mathbf{X}, \underline{\boldsymbol{\alpha}}) = f(\underline{\boldsymbol{R}} \mid \underline{\boldsymbol{Y}}_o, \mathbf{Z}, \mathbf{X}, \underline{\boldsymbol{\alpha}}) ,$$

which means that the data are MAR (or possibly even MCAR). In this case the non-response mechanism may depend on the observed part of the response vector but not on the missing part. By definition,

$$f(\underline{\boldsymbol{Y}} \mid \mathbf{X}, \underline{\boldsymbol{\beta}}) = f(\underline{\boldsymbol{Y}}_m \mid \underline{\boldsymbol{Y}}_o, \mathbf{X}, \underline{\boldsymbol{\beta}})\, f(\underline{\boldsymbol{Y}}_o \mid \mathbf{X}, \underline{\boldsymbol{\beta}})$$

and so (1.1) becomes:

$$f(\underline{\boldsymbol{Y}}_o, \underline{\boldsymbol{R}} \mid \mathbf{X}, \mathbf{Z}, \underline{\boldsymbol{\beta}}, \underline{\boldsymbol{\alpha}})$$

$$= \int f(\underline{\boldsymbol{Y}}_m \mid \underline{\boldsymbol{Y}}_o, \mathbf{X}, \underline{\boldsymbol{\beta}})\, f(\underline{\boldsymbol{Y}}_o \mid \mathbf{X}, \underline{\boldsymbol{\beta}})\, f(\underline{\boldsymbol{R}} \mid \underline{\boldsymbol{Y}}_o, \mathbf{X}, \mathbf{Z}, \underline{\boldsymbol{\alpha}})\, d\underline{\boldsymbol{Y}}_m$$

or, equivalently,

$$f(\underline{\boldsymbol{Y}}_o, \underline{\boldsymbol{R}} \mid \mathbf{X}, \mathbf{Z}, \underline{\boldsymbol{\beta}}, \underline{\boldsymbol{\alpha}})$$

$$= f(\underline{\boldsymbol{Y}}_o \mid \mathbf{X}, \underline{\boldsymbol{\beta}})\, f(\underline{\boldsymbol{R}} \mid \underline{\boldsymbol{Y}}_o, \mathbf{X}, \mathbf{Z}, \underline{\boldsymbol{\alpha}}) \int f(\underline{\boldsymbol{Y}}_m \mid \underline{\boldsymbol{Y}}_o, \mathbf{X}, \underline{\boldsymbol{\beta}})\, d\underline{\boldsymbol{Y}}_m\,.$$

This shows that the contribution to the likelihood from the observed data ($\underline{\boldsymbol{Y}}_o$) can be factorized into two components, one of which depends on $\underline{\boldsymbol{\beta}}$ and the other on $\underline{\boldsymbol{\alpha}}$. If inferences are required solely about $\underline{\boldsymbol{\beta}}$ then the term $f(\underline{\boldsymbol{R}} \mid \underline{\boldsymbol{Y}}_o, \mathbf{X}, \mathbf{Z}, \underline{\boldsymbol{\alpha}})$ can be ignored. However, the non-response mechanism cannot be ignored when it comes to determining the asymptotic variance-covariance matrix of $\hat{\underline{\boldsymbol{\beta}}}$.

If the non-response mechanism is *informative*, that is if:

$$f(\underline{\boldsymbol{R}} \mid \underline{\boldsymbol{Y}}, \mathbf{Z}, \mathbf{X}, \underline{\boldsymbol{\alpha}}) \neq f(\underline{\boldsymbol{R}} \mid \underline{\boldsymbol{Y}}_o, \mathbf{Z}, \mathbf{X}, \underline{\boldsymbol{\alpha}})$$

it means that the non-response mechanism depends on the **unobserved** part of the response vector ($\underline{\boldsymbol{Y}}_m$) and possibly also on the observed part ($\underline{\boldsymbol{Y}}_o$).

In the case of informative non-response, valid likelihood-based inferences can only be made on specification of a **non-response distribution**

$$f(\underline{\boldsymbol{R}} \mid \underline{\boldsymbol{Y}}, \mathbf{Z}, \mathbf{X}, \underline{\boldsymbol{\alpha}})\,.$$

Although modelling can be performed upon specifying a non-response distribution, assessing the goodness-of-fit of the model will not be possible due to its dependence on the unobservable missing data, although sensitivity analysis can be used to assess the dependency of the inferences on the non-response model specified.

Censoring in survival studies results in **missing failure times**. For example, if the failure times are **right-censored** (see Section 1.2), the precise times are not recorded and may therefore be viewed as **missing**. There is, however, some recorded information relating to the failure times, namely that they exceed a particular value. This value is then recorded as the survival time, together with a value of 0 for the censoring indicator. Moreover, the right-censoring process is almost always *informative* under the usual definition of the term because the probability of being right-censored depends on the true failure time and, as such, tends to increase as the true failure time increases. In this situation, conditioning on survival to time t, the usual approach in survival analysis, removes this problem, effectively making the right-censoring process *non-informative* and therefore ignorable under likelihood-based inferences. The problem persists, however, if the censoring process is dependent upon any variable, either fully, partially or not observed. Traditionally the censoring is referred to as **informative** in this latter situation only. In event history analysis, problems are potentially more serious as missing data processes can affect both the recording of the failure time and the repeated covariate measurements.

The terminology **dropout** rather than **right-censoring** is used when the interest is in one or more repeatedly-measured characteristics rather than survival time. This is because all forms of 'dropout' are treated the same, rather than one, caused by a particular event (or events), being of interest and all other causes of 'dropout' being of nuisance value. The terminology **informative dropout** is

used if the dropout process is not random, depending on the values that would have been observed if the dropout had not occurred.

Prior to consideration of a variety of general analysis methods, it is appropriate to note that many such methods are only appropriate or easily implementable if the missing-data patterns are of monotone form. This can be explained in the following way. Suppose there are two variables, $\underline{\boldsymbol{x}}$ and $\underline{\boldsymbol{y}}$ subject to missingness. Then the missing data are of **monotone form** if

$$x_i \text{ is missing} \Rightarrow y_i \text{ is missing}$$

or

$$y_i \text{ is missing} \Rightarrow x_i \text{ is missing.}$$

In other words, x_i is observed at least in the cases when y_i is observed **or** vice-versa (but not both). In repeated measures data, the common situation where individuals drop out of a study or experiment prior to completion, the missing data will be of a monotone form. However, intermittent missing data will not generally be of this form and must either be subject to deletion of all responses following an intermediate missing response or must be analysed using methods not reliant upon a monotone missing data structure.

The remainder of this section will consider three particularly important approaches to the analysis of data subject to missing data mechanisms. These approaches are **imputation** of the missing data, the **expectation-maximisation (EM) algorithm** and the **data augmentation algorithm**.

1.3.1 Imputation methods

Imputation methods are a class of methods used for explicitly filling-in missing data to enable standard data analysis methods to be used. They have largely, but

not exclusively, been used to form databases resulting from large-scale surveys to enable secondary data analysts to access and analyse the completed data. In many models, it is assumed that the missing-data mechanism is *non-informative* although some imputation methods have been applied under specific informative models for the non-response process.

A number of simple imputation techniques can be applied in practice, the most commonly applied being:

a) Mean imputation

Missing-values in a particular variable are replaced by the sample mean from the responding units.

b) Hot-deck imputation

This covers a relatively broad class of methods. They involve the random choice of a value from an estimated distribution for each missing value. In many applications, the distribution is not modelled formally but is estimated empirically from the observed data. The missing-values are thus filled-in by randomly selecting from amongst responding individuals with *similar* values of other appropriate variables. A great variety of **similarity measures** is potentially available.

c) Cold-deck imputation

Unlike hot-deck imputation, cold deck imputation does not use information from the present survey. For each missing-value, a constant value from an external source is imputed. Often, this external source is a previous survey. This method will have poor statistical properties as it takes no account of potential variability nor of any other consequences of the imputation, nor has much evaluation been made of the seriousness of these problems.

d) Regression imputation

Missing-values are replaced by the conditional mean estimated by regression

methods.

e) Stochastic regression imputation

This is an extension of regression imputation to allow for uncertainty in the imputed values. For continuous responses, where linear modelling is traditionally used, a residual is added to the estimated conditional mean. This residual may be a random element from a normal distribution with zero mean and constant variance estimated by the mean square error from the fitted model or, using hot-deck imputation ideas, a randomly selected value from amongst the residuals from the fitted model. For binary responses, where logistic models are usually employed, the probability of a positive response is estimated for the missing-value and then, by randomly selecting a value from the uniform distribution, a positive response is imputed if this value is less than the estimated probability and a negative value imputed otherwise.

Appropriate imputations may lead to relatively unbiased parameter estimates if the imputation technique closely mirrors the model generating the data and takes account of the missing-data mechanism. Any bias introduced via model mis-specification will increase as the proportion of data missing increases. The methods described above implicitly asssume that the data are non-informative. Certain methods, such as stochastic regression imputation, may be adapted to incorporate informative missing-data mechanisms. These methods will not, however, automatically lead to unbiased estimates of the variance of the resulting estimates, tending to underestimate the variance due to the treatment of the missing-values as 'known' and, except in b) and e), fixed for a given set of covariates. Whilst theoretical adjustments for this bias have been developed for some methods, a scheme known as **multiple imputation** may be applied in general.

Multiple imputation

With **multiple imputation**, rather than imputing a single value for each missing-data item to create a single complete data set, I complete data sets are created by drawing I values for each missing-data item from its estimated **predictive distribution**. The first draw is used to aid in the completion of the first data set, the second draw in the completion of the second data set, and so on.

Now, inference would usually use either likelihood-based or Bayesian methodology. In either case, if the data are MCAR, it would be based on a function of the form $f(\underline{\boldsymbol{\beta}} \mid \underline{\boldsymbol{Y}}_o, \boldsymbol{X})$ where $f(\underline{\boldsymbol{\beta}})$ represents the **likelihood function** or the **posterior density** for $\underline{\boldsymbol{\beta}}$ as appropriate. This function will, in general, be difficult to obtain due to the lack of standard analytical techniques in the absence of the missing data $\underline{\boldsymbol{Y}}_m$. However, by considering the joint distribution or likelihood of $\underline{\boldsymbol{\beta}}$ and $\underline{\boldsymbol{Y}}_m$, this problem can be overcome.

As multiple imputation is more naturally viewed in a Bayesian context, this subsection will refer to $\underline{\boldsymbol{\beta}}$ from a Bayesian viewpoint.

Now,

$$f(\underline{\boldsymbol{Y}}_m, \underline{\boldsymbol{\beta}} \mid \underline{\boldsymbol{Y}}_o, \boldsymbol{X}) = f(\underline{\boldsymbol{\beta}} \mid \underline{\boldsymbol{Y}}_o, \underline{\boldsymbol{Y}}_m, \boldsymbol{X})\, f(\underline{\boldsymbol{Y}}_m \mid \underline{\boldsymbol{Y}}_o, \boldsymbol{X})\,.$$

Integrating over $\underline{\boldsymbol{Y}}_m$ gives,

$$\int f(\underline{\boldsymbol{Y}}_m, \underline{\boldsymbol{\beta}} \mid \underline{\boldsymbol{Y}}_o, \boldsymbol{X})\, d\underline{\boldsymbol{Y}}_m = \int f(\underline{\boldsymbol{\beta}} \mid \underline{\boldsymbol{Y}}_o, \underline{\boldsymbol{Y}}_m, \boldsymbol{X})\, f(\underline{\boldsymbol{Y}}_m \mid \underline{\boldsymbol{Y}}_o, \boldsymbol{X})\, d\underline{\boldsymbol{Y}}_m\,.$$

So,

$$f(\underline{\boldsymbol{\beta}} \mid \underline{\boldsymbol{Y}}_o, \boldsymbol{X}) = \int f(\underline{\boldsymbol{\beta}} \mid \underline{\boldsymbol{Y}}_o, \underline{\boldsymbol{Y}}_m, \boldsymbol{X})\, f(\underline{\boldsymbol{Y}}_m \mid \underline{\boldsymbol{Y}}_o, \boldsymbol{X})\, d\underline{\boldsymbol{Y}}_m \,. \qquad (1.2)$$

This means that the required density may be obtained by averaging the density $f(\underline{\boldsymbol{\beta}} \mid \underline{\boldsymbol{Y}}_o, \underline{\boldsymbol{Y}}_m, \boldsymbol{X})$ over the **predictive density** $f(\underline{\boldsymbol{Y}}_m \mid \underline{\boldsymbol{Y}}_o, \boldsymbol{X})$.

However, this predictive density is also unknown as

$$f(\underline{\boldsymbol{Y}}_m \mid \underline{\boldsymbol{Y}}_o, \boldsymbol{X}) = \int f(\underline{\boldsymbol{Y}}_m \mid \underline{\boldsymbol{\beta}}, \underline{\boldsymbol{Y}}_o, \boldsymbol{X})\, f(\underline{\boldsymbol{\beta}} \mid \underline{\boldsymbol{Y}}_o, \boldsymbol{X})\, d\underline{\boldsymbol{\beta}} \qquad (1.3)$$

but can itself be obtained by averaging $f(\underline{\boldsymbol{Y}}_m \mid \underline{\boldsymbol{\beta}}, \underline{\boldsymbol{Y}}_o, \boldsymbol{X})$ over $f(\underline{\boldsymbol{\beta}} \mid \underline{\boldsymbol{Y}}_o, \boldsymbol{X})$.

This, however, requires, $f(\underline{\boldsymbol{\beta}} \mid \underline{\boldsymbol{Y}}_o, \boldsymbol{X})$ for use in the averaging, and so leads to a procedure of the following form:

i) start with the initial approximation for $f(\underline{\boldsymbol{\beta}} \mid \underline{\boldsymbol{Y}}_o, \boldsymbol{X})$;

ii) using the approximation to $f(\underline{\boldsymbol{\beta}} \mid \underline{\boldsymbol{Y}}_o, \boldsymbol{X})$ in conjunction with the 'model' $f(\underline{\boldsymbol{Y}}_m \mid \underline{\boldsymbol{\beta}}, \underline{\boldsymbol{Y}}_o, \boldsymbol{X})$ to approximate the **predictive density** $f(\underline{\boldsymbol{Y}}_m \mid \underline{\boldsymbol{Y}}_o, \boldsymbol{X})$ (as per 1.3);

iii) estimate the **posterior density** $f(\underline{\boldsymbol{\beta}} \mid \underline{\boldsymbol{Y}}_o, \boldsymbol{X})$ by averaging the **conditional density** $f(\underline{\boldsymbol{\beta}} \mid \underline{\boldsymbol{Y}}_o, \underline{\boldsymbol{Y}}_m, \boldsymbol{X})$ over the approximation of the predictive density $f(\underline{\boldsymbol{Y}}_m \mid \underline{\boldsymbol{Y}}_o, \boldsymbol{X})$ (as per 1.2).

In practice, this procedure takes the following form:

i) (a) assume a distributional form, often multivariate normal, for $f(\underline{\boldsymbol{\beta}} \mid \underline{\boldsymbol{Y}}_o, \boldsymbol{X})$;

(b) estimate the mean and variance (or alternative parameters) of the above conditional density of $\underline{\boldsymbol{\beta}}$. This is usually performed using either a complete-case or available-case analysis method in a standard software package;

ii) using the density estimated in i), take I random draws of $\underline{\boldsymbol{\beta}}$ and, in conjunction with the model $f(\underline{\boldsymbol{Y}}_m \mid \underline{\boldsymbol{\beta}}, \underline{\boldsymbol{Y}}_o, \boldsymbol{X})$, make predictions of $\underline{\boldsymbol{Y}}_{1m}, \ldots, \underline{\boldsymbol{Y}}_{Im}$;

iii) using a standard procedure for complete-cases, assuming the choice of a likelihood approach, obtain estimates of $\mathrm{E}[\underline{\boldsymbol{\beta}}]$ and $\mathrm{Var}(\underline{\boldsymbol{\beta}})$ from each of the data sets $(\underline{\boldsymbol{Y}}_o, \underline{\boldsymbol{Y}}_{1m}, \boldsymbol{X}), \ldots, (\underline{\boldsymbol{Y}}_o, \underline{\boldsymbol{Y}}_{Im}, \boldsymbol{X})$.

The estimates obtained in iii) are then combined by averaging the individual components of the estimates of $\mathrm{E}[\underline{\boldsymbol{\beta}}]$ over I, i.e. $\mathrm{E}[\hat{\underline{\boldsymbol{\beta}}}]^T = (\mathrm{E}[\hat{\beta}_1], \ldots, \mathrm{E}[\hat{\beta}_p])$ where

$$\mathrm{E}[\hat{\beta}_j] = \frac{\sum_{k=1}^{I} \mathrm{E}[\hat{\beta}_j]_k}{I}$$

and

$$\mathrm{E}[\hat{\underline{\boldsymbol{\beta}}}]_k^T = (\mathrm{E}[\hat{\beta}_1]_k, \ldots, \mathrm{E}[\hat{\beta}_p]_k) .$$

This is effected using the relationship

$$\mathrm{E}(\underline{\boldsymbol{\beta}} \mid \underline{\boldsymbol{Y}}_o, \boldsymbol{X}) = \mathrm{E}\left[\mathrm{E}\left(\underline{\boldsymbol{\beta}} \mid \underline{\boldsymbol{Y}}_o, \underline{\boldsymbol{Y}}_m, \boldsymbol{X}\right)\right]$$

which follows from Equation 1.2 upon taking expectations.

The analogous result for variances is

$$\mathrm{Var}\left(\underline{\boldsymbol{\beta}} \mid \underline{\boldsymbol{Y}}_o, \boldsymbol{X}\right) = \mathrm{Var}\left[\mathrm{E}\left(\underline{\boldsymbol{\beta}} \mid \underline{\boldsymbol{Y}}_o, \underline{\boldsymbol{Y}}_m, \boldsymbol{X}\right)\right] + \mathrm{E}\left[\mathrm{Var}\left(\underline{\boldsymbol{\beta}} \mid \underline{\boldsymbol{Y}}_o, \underline{\boldsymbol{Y}}_m, \boldsymbol{X}\right)\right]$$

and the diagonal terms of this variance-covariance matrix may be found as:

$$\mathrm{Var}\left(\beta_j \mid \underline{\boldsymbol{Y}}_o, \boldsymbol{X}\right) = \mathrm{Var}\left[\mathrm{E}\left(\beta_j \mid \underline{\boldsymbol{Y}}_o, \underline{\boldsymbol{Y}}_m, \boldsymbol{X}\right)\right] + \mathrm{E}\left[\mathrm{Var}\left(\beta_j \mid \underline{\boldsymbol{Y}}_o, \underline{\boldsymbol{Y}}_m, \boldsymbol{X}\right)\right]$$

and the off-diagonal terms can be found using the corresponding result for covariances.

The first of the terms in the equation for $\mathrm{Var}(\beta_j)$ is obtained from the variability in the estimates of $\mathrm{E}[\beta_j]$ between the I imputed data sets, this variability simply being estimated as

$$B_I = \sum_{k=1}^{I} \frac{\left(\mathrm{E}[\hat{\beta}_j]_k - \mathrm{E}[\hat{\beta}_j]\right)^2}{I-1}$$

and the second term is obtained from averaging the estimates of the variance of β_j, namely

$$W_I = \sum_{k=1}^{I} \frac{\mathrm{Var}(\hat{\beta}_j)_k}{I} .$$

However, as the number of imputations, I, is finite, a minor adjustment is necessary to eliminate bias in using B_I to estimate $\mathrm{Var}\left[\mathrm{E}\left(\beta_j \mid \underline{\boldsymbol{Y}}_o, \underline{\boldsymbol{Y}}_m, \boldsymbol{X}\right)\right]$ and so the total variance for β_j is estimated as

$$T_I = W_I + \frac{I+1}{I} B_I .$$

The reference distribution for β_j is, approximately, the t-distribution on ν degrees of freedom, where

$$\nu = (I-1)\left[1 + \frac{1}{I+1}\frac{W_I}{B_I}\right]^2$$

although multivariate analogues are preferable if $p > 1$ (Little and Rubin, 1987).

Afifi and Elashoff (1969a, 1969b), considered several methods of estimation in the simple linear regression setting with missing data on both the xs and the ys. The asymptotic distribution of the resulting estimators was derived and, when necessary, bias-corrected estimators were provided. Under MCAR mechanisms for x and y, ordinary least-squares estimation was compared with both two-stage and three-stage (first-order) methods. In the two-stage method, the first stage involved imputing conditional means for the xs based on regression on the ys and the second simply applied ordinary least squares estimation to the non-missing ys. The three-stage method used ordinary least-squares estimation on the ys based on the complete cases, then imputing the conditional mean for the missing ys to complete the first stage. The second stage involved completing the observations missing on the xs. Ordinary least squares estimation was used on the complete and completed cases and then the conditional mean was imputed. At the third stage ordinary least-squares was applied to the ys using the now fully completed data set. Additionally, asymptotically bias-corrected forms of the two- and three-stage estimators were derived (1969a) and compared (1969a,b).

It was concluded that the performance of the least squares estimator is generally worse than its competitors in terms of both asymptotic and finite sample efficiencies. The relative merits of the first-order and two-stage methods depended on

the relative proportions of data missing in the x and y and the correlation between the x and y. In general, the two-stage estimator was found to be better than the first-order method in the estimation of $\mu_{y|x}$. Whilst the performance was slightly worse than maximum likelihood, providing there were only moderate amounts ($\leq$30%) of missing data the asymptotic relative efficiency was always at least 94.5%, although this declined for large amounts of missing data. Moreover, the estimators with good asymptotic efficiencies generally had small bias, even with sample sizes of 20 and 40% of xs and ys missing and their ranking in terms of small sample efficiency generally concurred with that based on asymptotic efficiency (1969b).

Little (1992a) presented a review of regression methods used with missing values in the predictor variables. It was suggested that most methods can be classified into one of three 'direct' methods: complete case (CC) analysis; available case (AC) analysis and least squares (LS) on imputed data or one of three model-based methods: maximum likelihood (ML), Bayesian methods and multiple imputation. Complete case and available case analyses were discussed briefly at the start of this section and are considered in Chapter 5 with reference to interval-censored survival data. Maximum likelihood methods are based on the assumption of bivariate normality and the estimates are found by appropriate iterative search techniques. Least squares methods impute the missing values and then use ordinary least squares or weighted least squares, downweighting the incomplete cases, for parameter estimation. Three imputation methods are considered. The first is unconditional mean imputation, where missing values in a predictor variable are replaced by the unconditional mean for that predictor variable. Whilst bias is introduced into the regression parameter estimates and the covariance matrix estimate for all types of missing data mechanisms, this can be adjusted for if the data are MCAR. Adjustment of the variance-covariance matrix results in equivalence to the AC method.

Whilst unconditional mean imputation leads to bias and overstated precision, an improvement is provided by conditioning on either the observed predictor values for that case or, if the partial correlation of the missing predictor and the response given the remaining predictors is high, by conditioning on both the observed predictor values and the response value for that case. In the former, weighted least squares is recommended. Various weights based on the value of the partial correlation of the missing predictor and response given the observed predictors have been proposed. When conditioning is on the response as well as the predictors, biased parameter estimates result although these can be corrected for bias as discussed by Afifi and Elashoff (1969a, 1969b) in the case of a single predictor or using Buck's method (Buck, 1960). Using these methods, the downward bias in the standard errors is less marked than when unconditional means are imputed, but this is still a problem. Although corrections can be derived for certain missing data patterns, more general derivations present more difficult problems. It is therefore suggested (Little, 1992a) that it is more sensible to adjust for the bias in standard error estimation by using appropriate multiple imputation methods than by using a variety of theoretically complex bias corrections for different situations.

Maximum likelihood has certain attractive features such as the potential for extension to non-ignorable missing data mechanisms. However, it still suffers from the necessity to consider the missing data pattern and, moreover, the use of iterative methods, such as the EM algorithm (discussed further in Section 1.3.2), may be difficult to implement. The bias in the estimation of the standard errors is less than that when using CC, AC or simple imputation but is still considerable in some situations. Moreover the usual problem that maximum likelihood does not perform well in small samples persists. Little (1992a) also suggested a Bayesian extension in which a prior is added to the likelihood and inference is then based on the resulting posterior. For complex missing-data situations the use of the data augmentation algorithm, Gibbs sampling and importance sampling were recommended in order to simulate values from the posterior distribution.

Wang, Sedransk and Jinn (1992) considered informative missing data mechanisms where the model under consideration was the simple linear regression model and missing values were restricted to the responses. Seven imputation methods, plus a standard complete case (N) method were considered. The imputation methods were: mean imputation overall (MO) where the unconditional mean response was imputed for all missing responses; random imputation overall (R) where for each missing response a value was randomly imputed from the observed responses; simple regression imputation (RG) where the missing responses were imputed as their conditional expectations which were estimated using the ordinary least squares estimates provided by the complete cases; random regression imputation (RRN) where random residuals were added to the RG imputation; another random regression imputation (RSS) where the residuals, rather than being simulated from a normal distribution with zero mean and variance estimated as the mean square error from OLS on the complete cases were randomly chosen from the observed residuals; random imputation within adjustment cells (RC) where method R was applied within an appropriate class interval of the predictor variable; multiple imputation (MI) where an appropriate model for the missing data mechanism was introduced to estimate the predictive distribution for the missing data.

In terms of bias, the MO and R methods were found to be highly unsatisfactory with RG and RRN being notably better. In terms of confidence interval coverage, MO and R were again unacceptable, and whilst RG and RRN were a slight improvement, they were still considerably worse than the N method. MI provided coverage probabilities close to their nominal levels, with no real overall difference in performance across the simulations between two, five or twenty-five imputations. The authors suggest that using only the observed values may be satisfactory for certain objectives (e.g. inferences about β_0 and β_1) but unsatisfactory for others (e.g. inferences about the overall mean). Likelihood methodology is to

be preferred although it is very difficult to implement in some circumstances. In cases where likelihood methods are deemed impractical, the use of multiple imputation based around some correctly specified non-response model is advised. If, however, the data are MAR or nearly so, it was suggested that standard software be used when available.

A number of practical applications of multiple imputation have been presented in the literature. Many of these relate to the analysis of large scale surveys although a number of examples have occurred in recent years in relation to clinical trials and health-related applications, two of which will now be discussed in some detail.

Reilly (1993) presented an application of hot deck multiple imputation. This method has the advantage over parametric multiple imputation that it does not rely on correct specification of the model for the missing data mechanism providing this is non-informative. The application cited related to the occurrence of acute graft-versus-host disease in 97 female patients who received bone marrow transplants from female sibling donors. Logistic regression methods were used to model the occurrence of the disease. Complete covariates considered were age and prophylactic regime. All these covariates were considered as categorical. A single incomplete covariate, donor pregnancy status was present. The results of the analysis showed that, whilst logistic regression of 66 complete cases failed to show strong evidence of any of the covariates being true risk factors, the hot deck multiple imputation showed strong evidence of both donor pregnancy and prophylactic regime being associated with disease occurrence, despite an increase in standard error for the donor pregnancy parameter estimate. The paper presents results for 3, 10 and 100 imputations. There was very good agreement between the estimates and standard errors when there were 10 and 100 imputations and the agreement was still 'remarkably good' with only three imputations.

Dorey, Little and Schenker (1993) considered multiple imputation methods ap-

plied to threshold-crossing times. The aim was to use repeated measures in screening studies to improve inference whilst not resorting to specialist modelling. It is important to note that the response is a threshold-crossing time of a continuous variable and the value of this continuous variable is measured routinely at each screen. Moreover, inclusion of this continuous variable as a model covariate was considered, the authors stating that this is 'a problem which is not addressed in the current literature on interval-censoring'. The analysis presented was under the assumption that the interval-censoring process was non-informative.

Dorey et al. (1993) considered two problems. The first was a hip-replacement problem where radiographs were used to detect radiolucent lines between the prosthesis (artificial hip insertion) and cement interface. The threshold-crossing was defined as when the radiolucent lines covered 100% of the prosthesis with a maximum width of at least 2mm. Five imputation methods were considered in conjunction with this problem. Three methods were deterministic. These involved imputation of the right interval end-point, imputation of the interval mid-point and linear-interpolation imputation. Two imputation methods were random and ten multiple imputations were used. These were both based on the 'crossing fraction', namely the proportion of the interval during which the crossing was known to have occurred prior to the imputed crossing time. The first involved imputing a random value from the uniform distribution on (0,1) for the crossing fraction. The second used the beta distribution on (0,1) for the crossing fraction. As this beta distribution has two parameters, μ and κ, it was necessary to estimate these prior to the imputation. The parameter μ was simply estimated as the linear interpolation imputation value, as described above, whereas κ was estimated using this estimate of μ together with the data for the 22 complete cases plus the knowledge that the variance of the beta distribution is $\kappa\mu(1-\mu)$. Moreover, for multiple imputation, it is necessary to simulate draws of κ from its posterior distribution. The draws were made using bootstrap estimates of κ, having repeatedly sampled *with replacement* 22 complete cases from the 22 complete cases. It is noted in the

paper that this will only result in an approximate draw from the predictive distribution of the missing data (given the observed data) due to using only a bootstrap estimate for κ rather than incorporating all potentially important covariate information. The survival methodologies employed were Kaplan-Meier for the survivor function and Cox modelling for the hazard function. The survival time for the Kaplan-Meier analysis was defined as the time from the threshold-crossing (radiographic loosening) to revision surgery. In the Kaplan-Meier analysis, right-point imputation resulted in considerably lower estimated probabilities of survival to three and five years. Mid-point and linear interpolation methods yielded similar point estimates at three and five years, as did the two multiple imputation methods. The multiple imputation methods gave noticeably higher point estimates than the mid-point and linear interpolation methods at both three and five years. The suggestion was made that this is due to the Kaplan-Meier estimate not being linear in the crossing-times. The Cox regression analysis used the time from initial surgery to revision surgery as the survival time. The occurrence of radiographic loosening was included as a time-dependent binary covariate and was imputed using each of the five methods described earlier. With right-point imputation a substantially larger estimate of the risk of revision surgery due to radiographic loosening was obtained than with the other four methods, all of which provided similar estimates. All five methods yielded similar standard error estimates.

The second problem considered involved prostate cancer and the use of radioimmunoassay, the prostate specific antigen (PSA) assay, to detect cancer recurrence. The aim of the analysis was to model the time taken until clinically detected cancer recurrence following a positive PSA result. The definition of a positive PSA result was based on the crossing of a threshold. The imputation methods under consideration again included right-point and mid-point imputation and uniform multiple imputation. Linear interpolation imputation was, however, inappropriate for this analysis due to the non-linear form of the PSA assay response curve over time. It was believed that the growth curve could be adequately modelled as

a linear calibration problem relating the log of PSA assay value to time, allowing the intercept and gradient to vary between individuals.

Two imputation methods resulted from this approach. These were a deterministic-calibration method, imputing a single assay threshold-crossing time via a relatively complex calibration method and a random-calibration method, multiply imputing threshold-crossing times by randomly drawing a slope from the set of slope estimates for all cases plus a random error whose variance was also drawn (once for each data set). Appropriate corrections were introduced to ensure the imputed value lay inside the interval during which the censoring occurred. Kaplan-Meier estimation was again performed with estimates of survival to two and four years being presented. Right-point imputation again resulted in much lower survival probability estimates, the other four methods giving similar estimates in most cases.

The authors made the suggestion that the previously more commonly applied right-point method is considerably biased and, whilst multiple imputation may seem unnecessary in some cases, it does provide appropriate standard error estimates and, also, information on the fractional increase in variance due to the missing data. It was noted that the actual fraction of missing information tends to be lower than the proportion of missing data for a variable if there is auxiliary information available which can be used in the imputation process.

An alternative and theoretically preferable alternative to explicitly imputing the missing observations is to implicitly impute them via the EM algorithm. This involves imputation of expected values of sufficient statistics rather than imputation of the values for the missing data and will be covered in more detail in the next section.

1.3.2 The expectation-maximisation (EM) algorithm

The EM algorithm, first described under this name by Dempster, Laird and Rubin (1977) as a generalisation of an algorithm proposed by Orchard and Woodbury (1972), is a general purpose iterative procedure for computing maximum-likelihood estimates in a wide variety of incomplete data settings. Amongst these settings are not only many cases of traditional missing data in both ignorable and non-ignorable situations, but also less obvious applications including interval-censored survival data. Various applications will be described following a broad description of the algorithm.

When using the EM algorithm, rather than using the **observed data** to draw inferences based on $f(\underline{\boldsymbol{\beta}} \mid \underline{\boldsymbol{Y}}_o, \boldsymbol{X})$ using the **observed data** likelihood which is of the form

$$L(\underline{\boldsymbol{\beta}} \mid \underline{\boldsymbol{Y}}_o, \boldsymbol{X}) = \int f(\underline{\boldsymbol{Y}}_o, \underline{\boldsymbol{Y}}_m \mid \underline{\boldsymbol{\beta}}, \boldsymbol{X})\, d\underline{\boldsymbol{Y}}_m$$

if the data are MCAR or MAR, and of the form

$$L(\underline{\boldsymbol{\beta}}, \underline{\boldsymbol{\alpha}} \mid \underline{\boldsymbol{R}}, \underline{\boldsymbol{Y}}_o, \boldsymbol{X}) = \int f(\underline{\boldsymbol{Y}}_o, \underline{\boldsymbol{Y}}_m \mid \underline{\boldsymbol{\beta}}, \boldsymbol{X})\, f(\underline{\boldsymbol{R}} \mid \underline{\boldsymbol{Y}}_o, \underline{\boldsymbol{Y}}_m, \boldsymbol{X}, \underline{\boldsymbol{\alpha}})\, d\underline{\boldsymbol{Y}}_m$$

if the data are not MAR but the DP condition holds, the **complete data** likelihood $L(\underline{\boldsymbol{\beta}} \mid \underline{\boldsymbol{Y}}, \boldsymbol{X})$ (or $L(\underline{\boldsymbol{\beta}}, \underline{\boldsymbol{\alpha}} \mid \underline{\boldsymbol{R}}, \underline{\boldsymbol{Y}}_o, \boldsymbol{X})$) is used.

This likelihood function may be decomposed as

$$\begin{aligned} L(\underline{\boldsymbol{\beta}} \mid \underline{\boldsymbol{Y}}, \boldsymbol{X}) &= L(\underline{\boldsymbol{\beta}} \mid \underline{\boldsymbol{Y}}_o, \underline{\boldsymbol{Y}}_m, \boldsymbol{X}) \\ &= L(\underline{\boldsymbol{\beta}} \mid \underline{\boldsymbol{Y}}_o, \boldsymbol{X})\, f(\underline{\boldsymbol{Y}}_m \mid \underline{\boldsymbol{Y}}_o, \boldsymbol{X}, \underline{\boldsymbol{\beta}}) \end{aligned}$$

if the missing data mechanism is ignorable, or as

$$\begin{aligned} L(\underline{\boldsymbol{\beta}}, \underline{\boldsymbol{\alpha}} \mid \underline{\boldsymbol{Y}}, \boldsymbol{X}, \underline{\boldsymbol{R}}) &= L(\underline{\boldsymbol{\beta}} \mid \underline{\boldsymbol{Y}}, \boldsymbol{X}, \underline{\boldsymbol{R}})\, L(\underline{\boldsymbol{\alpha}} \mid \underline{\boldsymbol{Y}}_o, \underline{\boldsymbol{Y}}_m, \boldsymbol{X}, \underline{\boldsymbol{R}}) \\ &= L(\underline{\boldsymbol{\beta}} \mid \underline{\boldsymbol{Y}}_o, \underline{\boldsymbol{Y}}_m, \boldsymbol{X})\, L(\underline{\boldsymbol{\alpha}} \mid \underline{\boldsymbol{Y}}_o, \underline{\boldsymbol{Y}}_m, \boldsymbol{X}, \underline{\boldsymbol{R}}) \\ &= L(\underline{\boldsymbol{\beta}} \mid \underline{\boldsymbol{Y}}_o, \boldsymbol{X})\, f(\underline{\boldsymbol{Y}}_m \mid \underline{\boldsymbol{Y}}_o, \boldsymbol{X}, \underline{\boldsymbol{\beta}}) \\ &\qquad \times L(\underline{\boldsymbol{\alpha}} \mid \underline{\boldsymbol{Y}}_o, \underline{\boldsymbol{Y}}_m, \boldsymbol{X}, \underline{\boldsymbol{R}}) \end{aligned}$$

if the data are subject to a non-ignorable missing-data mechanism. In the latter case, as the likelihood for $\underline{\boldsymbol{\alpha}}$ depends on $\underline{\boldsymbol{Y}}_m$, it will be necessary to propose a non-response distribution for $\underline{\boldsymbol{R}}$ which depends on the missing-data vector $\underline{\boldsymbol{Y}}_m$.

The algorithm involves maximisation of the complete-data log-likelihood and, ideally, writing the resulting estimates of $\underline{\boldsymbol{\beta}}$ (or $\underline{\boldsymbol{\beta}}$ and $\underline{\boldsymbol{\alpha}}$) in closed-form as functions of sufficient statistics. The **expectation (E) step** then consists of taking expectations of these sufficient statistics, conditional upon the data and the present approximations to the parameter estimates. The **maximisation (M) step** then requires solution for the roots of the score equations, treating the expectations of the sufficient statistics as their observed values. Iteration through these two steps continues until stability of the solutions is achieved. In cases where the maximum likelihood estimates cannot be written in closed-form as functions of sufficient statistics, it may still be possible to use **EM** but its implementation will be considerably more difficult (Laird, 1988). **EM** has the desirable property that it has reliable convergence properties in that the log-likelihood is guaranteed to increase with each successive iteration. The downside is that its convergence is generally slow and can become painfully so in the presence of moderate amounts of missing data.

Laird, Lange and Stram (1987) considered the application of the EM algor-

ithm to the general linear mixed model for repeated measures of Laird and Ware (1982). The model is of the form

$$\underline{\boldsymbol{Y}}_i = \boldsymbol{X}_i\underline{\boldsymbol{\alpha}} + \boldsymbol{Z}_i\underline{\boldsymbol{b}}_i + \underline{\boldsymbol{e}}_i$$

where $\underline{\boldsymbol{Y}}_i$ is an $n_i \times 1$ vector of observations on the i^{th} unit, $\boldsymbol{X}_i$ and $\boldsymbol{Z}_i$ are known $n_i \times p$ and $n_i \times q$ design matrices, $\underline{\boldsymbol{\alpha}}$ is a vector of fixed effects to be estimated and $\underline{\boldsymbol{b}}_i$ and $\underline{\boldsymbol{e}}_i$ are independent vectors of random variables with multivariate normal distributions with zero mean vector. Moreover, $\underline{\boldsymbol{b}}_i$ has covariance matrix $\boldsymbol{D}$ with parameters to be estimated and $\text{Var}(\underline{\boldsymbol{e}}_i) = \sigma^2\boldsymbol{I}$. Detailed discussion of various forms for $\boldsymbol{D}$ may be found in Schluchter (1988), in the general linear model setting. Laird, Lange and Stram (1987) described the computational formulae for implementing the EM algorithm to find maximum likelihood (ML) or restricted maximum likelihood (REML) estimates using equations derived in Laird and Ware (1982). Special cases of the general linear mixed model were covered, including the growth curve formulation, the choice of starting values and techniques to accelerate convergence.

Vacek, Mickey and Bell (1989) compared maximum likelihood estimates obtained using this procedure with those obtained using two alternative estimation procedures. These were empirical Bayes and a non-iterative two-step approach with ordinary least squares at the first step and generalised least squares at the second. The application used for illustration was the modelling of longitudinal pulmonary function data from sarcoidosis patients. The repeated response 'measurements' were of the patients' forced vital capacity, expressed as a percentage of a predicted value for a person of that age, sex and height. This response variable was referred to as **FVC%**. At the first stage **FVC%** was modelled as a linear function of time for each individual and, at the second stage, the individual's intercept and 'growth rate' parameter were modelled as functions of particular patient charac-

teristics. The parameter estimates obtained using these three methods were very similar but the estimates of the variability of the various random effects differed by a moderate extent, potentially leading to different inferences. If these latter estimates are of interest, it was suggested that empirical Bayes may be preferable to maximum likelihood as the maximum likelihood estimates are known to be biased downwards.

Schluchter and Jackson (1989) contrasted the use of the EM algorithm with Newton-Raphson in computing maximum likelihood estimates in a log-linear model for hazard including only categorical covariates. These covariates may, however, be grouped continuous variables, thus extending the applicability of this model. The covariates may also be only partially observed providing any censoring is non-informative, the process causing the missingness in the covariates is ignorable and the random censoring variable does not depend on any covariate that is missing.

Schluchter (1992) considered the analysis of informatively censored longitudinal data where the primary outcome was the rate of change of a continuous variable. Various previously developed methods for the analysis of such data were compared and an extension to the linear random effects model was proposed in which it was assumed that the true intercept, slope and log-survival time follow a trivariate normal distribution. The resulting estimation is performed using maximum likelihood via the EM algorithm. Advantages of this model, as described by the author are:

- it allows arbitrary unbalanced data and takes account of staggered patient entry;
- asymptotically efficient estimates are obtained using all the data, including patients with only a single measurement;
- likelihood ratio tests, including tests of informative censoring, can be constructed.

The disadvantages are complications in programming and the requirement of large amounts of data to obtain stable parameter estimates and to avoid potential convergence problems.

Little and Schluchter (1985) applied the EM algorithm to a complete data model for the joint distribution of continuous variables ($\boldsymbol{X}$) and categorical variables ($\boldsymbol{W}$) in terms of the conditional distribution of $\boldsymbol{X}$ given $\boldsymbol{W}$ and the marginal distribution of $\boldsymbol{W}$. The following were suggested as applications:

- imputation of missing values;
- logistic regression and discriminant analysis with missing predictors and unclassified observations;
- linear regression with missing continuous and categorical predictors;
- parametric cluster analysis with incomplete data.

Tu, Meng and Pagano (1993) reported on estimating survival after AIDS diagnosis, using US surveillance data. The application considered related to the problems associated with the delay in reporting deaths to the surveillance system. Separate estimation of the delay and survival distributions was discussed. The survival distribution was modelled as a discrete-time proportional hazards model using the EM algorithm in conjunction with the complementary-log-log approach which can be applied using standard generalised linear model software (e.g. GLIM, SAS PROC LOGISTIC) and is discussed further in Section 4.1. In order to implement this on the data subject to reporting delays, unreported deaths were multiply imputed using independent draws from the predictive distribution, following modelling of the delay distribution. The delay distribution itself was also modelled via this discrete-time proportional hazards model using information available on the delay in reporting the more recent deaths.

Another form of the EM algorithm in the analysis of grouped survival data was used by Sinha, Tanner and Hall (1994). In this case, rather than the discrete-time proportional hazards model being used, Peto's marginal likelihood (Peto, 1972) based on the incomplete ranks of the failure times was used. Whilst this likelihood is inherently very difficult to maximise in normal circumstances, the authors showed how the Monte Carlo EM algorithm can be used to obtain the maximum likelihood estimates of the model parameters. The Monte Carlo EM algorithm is where Monte Carlo integration (or summation), rather than direct integration or a numerical technique, is used in the E-step.

Lindsey and Ryan (1993) used a non-homogeneous Markov three-state illness-death model, with the states of 'tumour-free', 'tumour' and 'death', to model rodent tumorigenicity experiments. A multiplicative relationship between death rates with and without tumour was assumed together with a piecewise exponential model for the baseline transition rates. State information was obtained from death times or sacrifice. Although the state transition times were interval-censored, continuous rather than discrete-time hazard models were used. As the tumour onset times were not observed, the *observed* data log-likelihood was difficult to work with directly. However, as the likelihood can be shown to take a simple form when the data are complete, the EM algorithm provided an appropriate numerical method for maximising the likelihood. Moreover, it was shown that this approach could be extended to incorporate covariate information in the model for the transition rates.

Kuk and Chen (1994) presented quite a different view of survival data, but again employing the EM algorithm. The view was that some individuals would **never** experience the event of interest. A logistic model was used to model the probability that an individual would ever suffer the event. A Weibull model was then used for the hazard function, conditional upon membership of the group who would eventually experience the event. A marginal likelihood approach was

used for this hazard model to eliminate the necessity to model the baseline hazard at this stage. However, due to right-censoring, it is not feasible to compute the marginal likelihood in practice. This is because it would require summation of relatively complex likelihood terms over all the different possible combinations for group membership of the censored individuals. To overcome this problem, Monte Carlo methods were used to approximate the marginal likelihood function and the corresponding logarithm of the averaged likelihood was then maximised using existing software. Finally, the EM algorithm was used to estimate the conditional baseline survivor function.

Whilst this review of applications has concentrated on areas directly relevant to this thesis, it should be emphasised that the EM algorithm has been applied in a wide variety of alternative incomplete data situations to facilitate maximum likelihood estimation.

An alternative to EM, which can be viewed as combining the more desirable features of multiple imputation and the EM algorithm, is the **data augmentation algorithm**. This is essentially an algorithm for estimating posterior densities which uses both multiple imputation and iteration to optimise the estimation and is covered further in Section 1.3.3.

1.3.3 The data augmentation algorithm

This is an algorithm, proposed by Tanner and Wong (1987a), to calculate *posterior distributions* in incomplete data situations. It is a fully Bayesian approach to such problems, unlike multiple imputation which applies Bayesian methodology to create the multiple completed data sets but then allows 'standard analysis techniques', often likelihood-based, in making inferences. Moreover, unlike multiple

imputation, it necessarily involves repeated drawing from successive approximations to the predictive density. The algorithm has two steps, which are repeated in turn until the posterior density $f(\underline{\boldsymbol{\beta}} \mid \underline{\boldsymbol{Y}}_o, \boldsymbol{X})$ stabilises.

These steps are:

a) The Imputation (I) Step

- draw a value of $\underline{\boldsymbol{\beta}}$ from the current estimate of $f(\underline{\boldsymbol{\beta}} \mid \underline{\boldsymbol{Y}}_o, \boldsymbol{X})$;
- generate $\boldsymbol{m}$ draws of $\underline{\boldsymbol{Y}}_m$ from $f(\underline{\boldsymbol{Y}}_m \mid \underline{\boldsymbol{\beta}}, \underline{\boldsymbol{Y}}_o, \boldsymbol{X})$ to approximate the predictive density $f(\underline{\boldsymbol{Y}}_m \mid \underline{\boldsymbol{Y}}_o, \boldsymbol{X})$.

b) The Posterior (P) Step

Update the current approximation of $f(\underline{\boldsymbol{\beta}} \mid \underline{\boldsymbol{Y}}_o, \boldsymbol{X})$ by mixing the conditional densities of $\underline{\boldsymbol{\beta}}$ over the missing-data patterns generated in a).

The P-step involves Monte Carlo approximation of the integral

$$\int f(\underline{\boldsymbol{\beta}} \mid \underline{\boldsymbol{Y}}_o, \underline{\boldsymbol{Y}}_m, \boldsymbol{X})\, f(\underline{\boldsymbol{Y}}_m \mid \underline{\boldsymbol{Y}}_o, \boldsymbol{X})\, d\underline{\boldsymbol{Y}}_m$$

to give the new approximation to $f(\underline{\boldsymbol{\beta}} \mid \underline{\boldsymbol{Y}}_o, \boldsymbol{X})$.

The I-step is merely an application of multiple imputation whilst the P-step maintains the Bayesian approach, rather than potentially returning to frequentist or likelihood-based inference as chosen by the secondary data analyst. Although convergence to $f(\underline{\boldsymbol{\beta}} \mid \underline{\boldsymbol{Y}}_o, \boldsymbol{X})$ is guaranteed, the implementation of either step may be difficult. Tanner and Wong (1987a) note that this mirrors the potential problems in applying the EM algorithm where either the E-step or the M-step may be difficult to implement in particular incomplete data situations.

Tanner and Wong (1987b) showed how the algorithm may be extended to the non-parametric estimation of the hazard function in the presence of interval-censored survival times. The algorithm works by using a starting estimate of the failure hazard to generate m multiple imputations of the failure times. This is the I-step. This improves in two distinct ways on the estimation when the conditional mean failure time is imputed. It produces a smoother non-parametric estimator and a better approximation to its variance. The P-step involves estimating the hazard function using kernel smoothing methods on the empirical hazard, as proposed in an earlier paper (Tanner and Wong, 1984), for each of the m data sets generated in the I-step. The algorithm then involves iterating between the I-step and the P-step until the estimated hazard curve stabilises. In a simulation study, they showed that the performance of the estimator is reasonable, in terms of the bias of both the estimate and its standard error, providing the sample size (n) and the number of imputations (m) are not both small. There was, however, clear skew in the estimate of the hazard function with $n = 25$ and $m = 5$, although this improved sufficiently to be deemed 'reasonable' in terms of bias upon increasing either the sample size three-fold or the number of imputations four-fold.

1.3.4 Overview of the three approaches

The three categories of methods for handling missing data are similar in that they aim to 'fill-in' values for the missing-data, based around a model for the missing-data. They differ, however, in that whilst the imputation methods and data augmentation algorithm *explicitly* fill-in the data values (to allow standard data analysis methods to be used), the EM algorithm *implicitly* fills-in by 'imputing' the expected values of sufficient statistics. Moreover, whilst the EM algorithm and data augmentation algorithm *iterate* by repeatedly imputing data and estimating parameters, imputation methods simply pass through the appropriate algorithm

once.

The EM algorithm is used in likelihood estimation whereas the data augmentation algorithm is used in posterior density estimation. Multiple imputation, the most useful of the imputation methods, is itself based on Bayesian ideas. However, the multiple imputation algorithm simply leads to the creation of several data sets which can then be analysed using standard frequentist, likelihood or Bayesian approaches.

1.4 Summary

This chapter briefly reviewed the literature on the three areas of repeated measures data, survival data and missing data.

Random coefficients models (see Section 1.1.3) used for growth curve modelling will be investigated is Chapter 3 and the transition models of Section 1.1.4 will be used in implementations of multiple imputation (see Section 1.3.1) in Chapters 4 and 5.

Survival modelling is restricted to semi-parametric modelling and concetrates largely on interval-censored data (see Section 1.2.3). Extensive coverage of techniques for analysing interval-censored survival data can be found in Chapter 4 and the application of these methods can be found in both Chapter 4 and Chapter 5.

Whilst the only type of technique for analysing incomplete data considered in the practical implementations of Chapters 4 and 5 is imputation, with particular emphasis on multiple imputation (see Section 1.3.1), the use of iterative tech-

niques of the form of the data augmentation algorithm is appealing with a view to improving inferences, particularly in the presence of non-monotone missing data patterns. Of the imputation techniques applied in the multiple imputation, model-based methods corresponding to *stochastic regression imputation* are used extensively, as is a special form of cold-deck imputation in which measurements from earlier realisations of the same survey are used for imputation.

Other techniques have been included in this introduction for completeness and also to provide a basis for extension or improvement of the techniques proposed in the later chapters.

Chapter 2

Event history analysis and screening studies

The main aim of this thesis is to explore the analysis of forms of longitudinal studies in which the occurrence or non-occurrence of one or more events of primary interest are recorded and investigated. This is referred to as *event history analysis*. In investigating these events, it is usually desirable to collect information on potential risk factors, preferably at frequent and regular time points. Within the class of event history studies may be found screening studies in which individuals in a cohort are invited to attend regular 'check-ups' in which certain measurements and assessments are made. Many, but not necessarily all, of the data collected will relate to health. Other aspects of screens may be sociological or psychological. Because screening studies are longitudinal and usually relate to moderately large cohorts, they are frequently subject to moderate amounts of missing data. Moreover, in many cases, individuals missing a screen will not usually be subjected to repeated reminders and therefore detection of informative missing-data mechanisms is made unfeasible without additional untestable assumptions about the nature of the informativeness. In such cases it is advisable to consider various plausible forms of the missing-data process and perform sensitivity analysis. In addition to this, in some screening studies the subset of

measurements collected at the follow-up times varies, as in the Framingham Heart Study. This leads to additional missing data, although missing data caused in this way will be less problematical due to to their MCAR nature. The remainder of this thesis will be directed towards the analysis of event history data, with specific emphasis on screening studies. In the Department of Epidemiology and Public Health at the University of Leicester there is a particular interest in screening the *elderly* for various state changes. Application in the context of screening studies of the elderly will therefore be considered in detail, especially in Chapter 5.

The main aim of screens is usually to detect a change of state although some more complex programmes may incorporate screening for a number of multi-state outcomes. Moreover, although invitation to screens will sometimes be of the same regularity and form for all cohort members, this will not always be the case. If screening is of an irregular nature, it will require the data analyst to carefully consider its likely effect on the form of the missing data mechansism. In screening studies, changes of state will generally only be detectable at the screens and will therefore lead to recording of the corresponding transition times in interval-censored form. This will be referred to as **screen-dependent event collection** in the remainder of this thesis. If changes of state are detected by some process independent of the screens, this will be termed **screen-independent event collection**. Screen-independent event collection may lead to either interval-censored or right-censored event times, depending on the nature of the event collection mechanism. The impact of a missed screen will be different under screen-dependent and screen-independent event collection. In the former, both covariate information and change-of-state information will be missing whereas in the latter only the covariate information will go unrecorded due to the missed screen. Both screen-dependent and screen-independent event collection will be considered in more detail in Chapters 4 and 5.

The next section will consider various forms of changes of state and forms of ex-

amination schemes for their detection. As risk factors will tend to vary with time it is important to account for the way in which the accumulation of a patient's risk factor history interacts with any change-of-state process and an understanding of this interaction should assist in the planning of preventive medicine. In particular, knowledge of which individuals will benefit most from preventive measures and at what stage such intervention should be made may be gleaned from this interaction.

It is clear, therefore, that in order to consider modelling based on data arising from screening studies, it is necessary to explore methods which make use of the repeated measures nature of any covariate information in modelling changes of state whilst also being sufficiently flexible to allow for incomplete information on both covariate information and times to change of state. The final section of this chapter will review modelling techniques applied or applicable to event history analysis with a view to suggesting which approaches hold most promise for application to screening studies. Some of the more promising approaches will be considered in more detail in Chapters 3 and 4 and then, where appropriate, applied to data collected from the annual screening of the elderly in Melton Mowbray in Chapter 5.

2.1 Detection of changes of state in screening studies

2.1.1 Forms of change of state

There are many possible forms of change of state which might be collected in screening studies. Moreover, in some cases, several assessments are made as part

of the same programme. For example, an annual screening of the elderly might incorporate assessment of:

(i) residential status, with states:

- own/family home;
- hospitalization and/or residential home/permanent nursing care;
- dead;

(ii) physical ability status, with states:

- not disabled;
- disabled;
- dead;

(iii) mental status, with states:

- no dementia;
- possible dementia;
- dementia;

whereas in a study of animal tumorigenicity or patient survival following surgical removal of a tumour a single outcome might be of interest, where the states may be:

- alive and tumour free;
- alive and tumour present;
- dead.

Hsieh, Crowley and Tormey (1983) discuss the application of Cox's proportional hazards model in semi-Markov models for the transition rates between states, showing that the partial likelihood may be factorised into components corresponding to the individual states. It follows, therefore, that each transition may be analysed separately whilst treating the other types of transition as a form of censoring. This thesis concentrates on the modelling of two-state processes.

In many screening studies, a state transition may only be observed to occur between two screening times, resulting in an interval-censored observation (see Section 1.2). For the second set of states in the annual screening of elderly people introduced earlier, transitions into the state '*dead*' will generally be known 'exactly', whereas transitions between '*not disabled*' and '*disabled*' states will generally only be known to have taken place between successive screens. An added complication is that the **exact** date of transitions may be known for some individuals and the use of such information would be desirable as this would allow the form of the baseline hazard (or survivor) function to be investigated in more detail and may also increase the precision of the estimation of the model parameters. Moreover, it is possible for two transitions to have occurred between consecutive screening times, further complicating the modelling process. Whilst one or more of these forms of state transitions will be encountered in many screening studies, only interval-censored event times will be considered in detail in this thesis.

2.1.2 Types of examination scheme

With interval-censored data from screening studies, it is important to consider the examination scheme process which underlies the interval-censoring. As with right-censoring or dropout processes, the usual likelihood-based methods require the examination scheme to be **non-informative**. This means that the full

likelihood of survival statuses at the respective screening times and the number and times of examinations is proportional to the likelihood based simply on the, potentially multi-state, survival statuses at the screening times. By a suitable factorization of the full likelihood, Grüger, Kay and Schumacher (1991) showed that the examination times which satisfy the following criterion are non-informative:

- The choice of the next examination time may only be based on the entire patient history: namely the times of examinations and any covariates measured and statuses observed at these times.

This criterion may not be satisfied in all screening studies. To illustrate this, Grüger et al. (1991) considered four simple models to describe how examination schemes may be categorised:

(i) examination at regular intervals;

(ii) random sampling;

(iii) 'doctor's care';

(iv) patient self-selection.

Using the criterion given above, the first three models satisfy non-informativeness of examination scheme. Model (iii) satisfies this providing the doctor chooses the time of next examination based on information available at the present examination time rather than on any *subsequent* patient information. The patient self-selection process is, however, an informative scheme as the examination times are decided by the patients. These will usually depend on the present disease state and/or covariates, rather than their past history, thus failing to satisfy non-informativeness. If the analysis is to be performed on multi-state 'survival' data collected from the annual screenings of the elderly, the main model for the examination will be (i). It may be, however, that the healthiest patients are the least likely to attend regular screenings by personal choice, thus forcing the type (iv) model to apply and requiring any likelihood-based analyses to take account

of the missing-data mechanism. In all cases it is useful, whenever possible, to collect the information on the patient's survival state at their missed screening time, thus avoiding any problems associated specifically with screen-dependent event collection whilst providing more information and hence more precision for the modelling. As the covariate information for missed screening times will be unavailable, if likelihood-based methods are to be used it is important to consider the missing data mechanism corresponding to the examination scheme. Moreover, if the examination or screening process is potentially informative, the specification of appropriate models for this process will be necessary if unbiased likelihood-based inferences are to be made. This will entail either the collection of covariate values for a random sample of those missing their scheduled screen, knowledge of the form of the model for the missing data mechanism or the use of sensitivity analysis for a number of plausible such models. Grüger et al. (1991) illustrated that bias could be a serious problem in the case of patient self-selection via a simulation study.

2.2 Modelling in screening studies

Several distinct approaches to modelling data arising from screening studies exist. Widely recommended by a number of authors in the literature (e.g. Thomas (1992), Manton (1993), Dwyer (1992)) is the application of what have been termed **'strong models'** (Thomas, 1992) as they exploit biological knowledge about the relationship between measurements, usually of a physiological nature, and state transition processes. These models are usually expressed in continuous-time and use either deterministic or stochastic differential equations, often termed *diffusion equations.* However, there is considerably wider application of the alternative **'weak models'** which are based on discrete-time and can be viewed as integrated

equations. These methods are criticised by the proponents of the 'strong models' in that they fail to take into account any knowledge of the underlying biological processes in the modelling of state transitions. In this thesis, whilst concentrating mainly on the 'weak models' widely applied in survival analysis, there will be some further discussion of diffusion models both later in this section and in Chapter 3.

In terms of 'weak models' there are two distinct classes of techniques that have been developed for handling data with a similar structure to that typically arising from screening studies. One class corresponds to growth curve models in which, using screening study terminology, the primary interest is in modelling the profile of the variable or variables measured at each screen and their effect on change of state is of secondary interest. These and associated methods will be discussed in Chapter 3.

The alternative is where the primary interest is in the effect of the risk factors on change of state and uses methodology developed for modelling grouped or interval-censored survival data. For efficient modelling, however, some account should be taken of the development of risk factors over time. Most methods that have been developed are for data which are essentially the exact failure times, albeit subject to some degree of rounding. These methods are also applicable when the times are interval-censored, providing the potential censoring interval boundaries and risk factor measurement times coincide and do not vary between individuals. Data of this form will be referred to as **coincident interval** data and correspond, in terms of screening studies, to screen-dependent event collection with no missing screens. Methods for modelling interval-censored survival data with repeated risk factor measurement are therefore required. Such methods are not widely available. A particular problem occurs when the data are not coincident interval. In survival analysis it is usual to condition on the covariates, but this is not trivial when their values are time-dependent, especially if the data are not coincident interval. Methods for the analysis of interval-censored survival

data of both coincident interval and non-coincident interval forms will be considered in detail in Chapter 4.

If there are repeated assessments of risk factors and a number of states, including an absorbing state such as death, several authors (e.g. Lindsey (1993), Andersen (1992), Hsieh, Crowley and Tormey (1983)) suggest the use of Markov and semi-Markov models, using in particular the Cox regression model (1972, 1975) with time-dependent strata. However, these models have a restrictive structure and are limited to 'few time-dependent covariates' (Andersen, 1992) which again emphasises the lack of available methods for the analysis of event history data with incomplete repeated assessment of states together with discrete and continuous risk factors.

A number of 'strong models' have been described in the literature, although much of their application has been in the field of mathematical biology rather than epidemiology. This may well be due to the widespread availability of survival and, to a certain extent, growth curve software for the implementation of 'weak models' compared with the more specialist software required for the 'strong models' which, due to their direct link with biological theory, are necessarily application-specific. Manton (1993) described three estimation strategies for modelling multi-state processes, placing particular emphasis on the interaction between human ageing and mortality. Two of these, one based on multivariate diffusion processes and the other based on fuzzy state spaces, have been used for modelling event history data (Manton, Stallard and Liu, 1993). The former, discussed in an extensive series of papers by Manton and various colleagues (e.g. Woodbury and Manton (1977), Woodbury, Manton and Stallard (1979), Manton, Stallard and Woodbury (1986), Yashin, Manton and Stallard (1986), Manton, Woodbury and Stallard (1988), Manton, Dowd and Stallard (1993), Manton, Stallard and Liu (1993)) is appropriate when there are several continuous physiological measurements made at each follow-up point. Fuzzy state models have been suggested as 'a new approach to

multi-state modelling' (Manton, 1993) and were illustrated in an application to active life expectancy (ALE) with a classification of people into six fuzzy states based on 27 measures of their physical functioning (Manton, Stallard and Liu, 1993). Following the classification of individuals at various time points, these states can be related to mortality and to risk factors through ancillary equations (Manton, Stallard and Liu, 1993).

These, and other 'strong models', are useful additions to the literature on event history analysis, although they are relatively complex mathematically. They have the advantage over many other approaches in that they are biologically and medically plausible as they can model the interaction of risk factors, multiple state transitions (such as levels of disability) and mortality. However, these methods do not readily incorporate missing covariate values of the form frequently encountered in screening studies. In fact, Manton, Woodbury and Stallard (1988) imputed missing risk factor measurements into the Framingham data set by simple between-wave interpolation prior to the use of a multivariate diffusion model. If the values were MCAR, the standard errors would be underestimated, and in other cases there is also the potential for the estimation to be biased. Moreover, in many screening studies, a number of the risk factors will be categorical and therefore will not lend themselves to interpolation-based imputation. Further discussion of the stochastic diffusion modelling of the type first proposed by Woodbury and Manton (1977) is included in Chapter 3 as this is one of the few 'strong' approaches which seems to have permeated the statistical literature.

Chapter 3

Analysing repeated measurements subject to dropouts

As outlined in Chapter 2, within the class of 'weak models' there are two distinct views of repeated measures of one or more characteristics when units are lost to follow-up prior to completion of a study or experiment. The first, which is considered in this chapter, considers the modelling of the repeated measurements as of primary interest with the process leading to loss of units, called **dropout**, being largely of nuisance value. This is the view taken in the literature on *repeated measures subject to dropouts* and, in particular, that on *growth curves*. In these cases the parameters of major interest are those of the repeated measures mean profile or the growth curve rather than those of the dropout process.

The alternative approach is to view the process describing loss of units due to one particular event as of paramount interest and the use or modelling of the repeated measurements can be used to improve inferences relating to this process. In addition, there is frequently loss of units due to other reasons which is referred

to as *right-censoring* (rather than dropout). This approach is taken in *survival analysis* and will be considered in the remainder of the thesis.

There are two other major differences between the approaches generally used in growth curve analysis and survival analysis. Only a single continuous variable plus the measurement (and dropout) times are generally recorded and used in growth curve analysis whereas a number of both continuous and categorical potential risk factors are usually assessed in survival analysis. Additionally, in survival or event history analysis, the risk factors are generally introduced into the survival or change of state model as covariates or, if the risk factor distribution is modelled, its coefficients are *usually* assumed to be *fixed* effects. In growth curve modelling, conversely, these are *usually* treated as *random* effects.

3.1 Repeated measures, growth curves and dropout processes

A number of papers appearing in journals during the last dozen years extend the methodology applicable to repeated measures to include data subject to dropout. Little (1995) discusses 'methods which simultaneously model the data and the dropout process within a unified model-based framework'. It was suggested that models can be classified in two ways. One type of classification is into models where the dropout process depends directly on the repeated measures and those in which it depends on them indirectly though random coefficients. It is this latter class which is generally used in the growth curve literature. The other classification which can be made is into **selection models** and **pattern-mixture models**. In selection models, which are the more common, detailed specification of the non-response mechanism (i.e. the distribution or density of the vector of indicator

variables $\underline{\boldsymbol{R}}_i$) is required as this is used to explain how the data are *selected* from the complete data model $f(\underline{\boldsymbol{Y}}_i \mid \boldsymbol{X}_i, \underline{\boldsymbol{\beta}}_i)$ to be missing. For mixture models the distribution of the $\underline{\boldsymbol{Y}}_i$ is viewed as a mixture over the different patterns of missing data, having stratified over these by considering $f(\underline{\boldsymbol{Y}}_i \mid \boldsymbol{X}_i, \underline{\boldsymbol{\beta}}_i, \underline{\boldsymbol{R}}_i)$ rather than the complete data model given above.

It is clear that the terminology used to refer to the dropout process is extremely inconsistent in usage and frequently depends on an author's preference and the specific situation under consideration. In addition to completely random dropout which corresponds directly to MCAR data, Little (1995) introduces the terminology *covariate dependent dropout* if dropout is dependent on a set of fixed covariates but not on the repeated measures of the response nor, in the case of random coefficient models, on the random coefficients. When the units are subjected to this form of dropout, complete case analysis (CCA) is inefficient but not biased. Another distinct form of dropout is the standard *missing at random (MAR) dropout* in which dependence is on the observed values of the response but not on those unobserved. Little (1995) also sub-divided the class of non-ignorable dropout into *non-ignorable outcome-based dropout* and *non-ignorable random-coefficient-based dropout.* The distinction between these final two categories is that, in the former,

$$f(\underline{\boldsymbol{R}}_i \mid \boldsymbol{X}_i, \underline{\boldsymbol{Y}}_{oi}, \underline{\boldsymbol{Y}}_{mi}, \underline{\boldsymbol{\beta}}_i) = f(\underline{\boldsymbol{R}}_i \mid \boldsymbol{X}_i, \underline{\boldsymbol{Y}}_{oi}, \underline{\boldsymbol{Y}}_{mi})$$

whereas, in the latter,

$$f(\underline{\boldsymbol{R}}_i \mid \boldsymbol{X}_i, \underline{\boldsymbol{Y}}_{oi}, \underline{\boldsymbol{Y}}_{mi}, \underline{\boldsymbol{\beta}}_i) = f(\underline{\boldsymbol{R}}_i \mid \boldsymbol{X}_i, \underline{\boldsymbol{Y}}_{oi}, \underline{\boldsymbol{\beta}}_i)\,.$$

In the former case, the non-response mechanism is non-ignorable in the traditional sense, due to the unobservable $\underline{\boldsymbol{Y}}_{mi}$, whereas in the latter it is due to its failure to satisfy the distinct parameter (DP) requirement (Shih, 1992). Prior to the contribution from Shih (1992), many authors had failed to make the important

distinction between these situations although the distinction has since been addressed (e.g. Diggle and Kenward (1994)).

Whilst modern sofware, in particular SAS PROC MIXED (SAS, 1996) and BMDP5V (BMDP, 1990), handle arbitrary missing data, it is under the assumption that the non-response mechanism is ignorable. If this is not the true mechanism, bias will be introduced into the estimation. It is widely and sensibly recommended (e.g. Little (1995)) that any model fitted should take into account any available knowledge of the form of the missing data process as well as the model for the mean and covariance structure. Many of the recent approaches to the analysis of repeated measures data subject to dropout and irregularly spaced measurements times follow this recommendation and will be discussed in the remainder of this section.

Vonesh and Carter (1987) used non-iterative estimation and comparison of location parameters as an alternative to the computationally expensive maximum likelihood (ML) or restricted maximum likelihood (REML) estimation. This approach uses estimated generalised least squares (EGLS) and is shown to be both consistent and asymptotically efficient. Vonesh and Carter (1987) contrasted their approach with the likelihood-based techniques of Laird and Ware (1982) and Jennrich and Schluchter (1986) which were briefly described in Section 1.3.2. Carter, Resnick, Ariet and Shieh (1992) illustrated the use of this EGLS estimator on a model relating a measure of pre-school cognitive development to age in four race-by-sex groups of low birthweight infants, also demonstrating some techniques for validation of model assumptions. The study group comprised 375 infants observed at a minimum of three ages and had no missing risk factors. The model was used under the assumption that any dropout was ignorable and it was recognised that the EGLS estimator would be biased if this were not the case. Other than this untestable assumption, the model assumptions appeared to be valid.

Crépeau, Koziol, Reid and Yuh (1985) used a score test for comparison of repeated measures data subject to ignorable dropout and under the assumption of multivariate normality of the responses. Diggle (1988) improved on this approach in terms of efficiency, illustrating this by re-analysing data from an experiment to determine the effect of halothane at different doses on the blood pressure of rats. He concluded that there was a marginally significant effect of treatment at the 5% level in contrast to the conclusion of Crépeau et al. (1985) that there was no significant effect of treatment (p=0.28) although they showed that a linear trend in dose was only marginally non-significant at the 5% level (p=0.06). Crépeau et al. (1985) based their score test on maximum likelihood estimates from a model with unstructured covariance matrix whereas Diggle (1988) reduced the number of parameters in the modelling of the same data set from 45 to firstly four and then, following simplification of the covariance structure to that of a uniform correlation model, to two. This was achieved using a model-fitting process comprising the following three steps:

(i) formulation of a provisional, possibly over-parametrised, model for the mean structure;

(ii) formulation of a model for the covariance structure using the residuals from an ordinary least squares fit of the model for the means;

(iii) use of a generalised likelihood-ratio test for the model parameters, treating the parameters for the mean and covariances structures separately.

Whilst Diggle's general model, with four parameters for the covariance structure, is somewhat specialised, it is still reasonably widely applicable due to the sensible choice of parametrisation corresponding to typical designs for experiments of this type.

Further work by Diggle (1989) led to a class of tests for randomness of dropouts for repeated measures data. This can be particularly useful as there is a wide range of techniques available if dropouts are random whereas the range of techniques is

far more limited if dropouts are non-random. The terminology 'non-random' here was used to correspond to dependence on previous measurements of the response and, potentially, missing values of the response, thus corresponding to missing at random or informative dropout, between which it would not be possible to distinguish. As non-random dropouts imply the possibility of the dropout being informative, this requires the adoption of more complex techniques to limit bias. When testing for random dropout, Diggle (1989) proposed the choice of a test statistic as a suitable function of the observed responses and then to perform a test of the hypothesis that, of those with complete observations up to and including that at a particular time-point, those with no further observations are a random sample. This is repeated for each time-point and then the p-values thus obtained are tested to see whether they represent a random sample from a uniform distribution on (0, 1). The results of tests on the data analysed earlier by Crépeau et al. (1985) and Diggle (1988) supported these authors' treatment of the dropout as random in all but the highest dose group which, in each case, had been removed from the data set prior to analysis.

Diggle and Kenward (1994) extended the model from Diggle (1988) to cover situations where the dropout mechanism may be informative. A logistic model was proposed for dropout at time t in which the linear predictor could depend on the prior history, the observation at time t, neither or both. Likelihood estimation was then used under the assumption that the distinct parameter (DP) condition (Shih, 1992) held. Simulation studies showed that biases in models and estimation strategies that fail to account for informative dropout are eliminated by this approach. The authors suggested that, whilst inferences could be based on the information matrix, alternative strategies using likelihood-ratio tests and likelihood-based confidence intervals are preferable.

More central to the growth curve literature covering situations where units are subject to dropout is a series of papers by Wu and her collaborators. Wu and

Carroll (1988) proposed a random coefficient *selection* model in which parameters of the repeated measures growth curve model and those of the dropout model are jointly estimated via maximisation of a *marginal* likelihood function. They termed their method pseudo-maximum likelihood estimation (PMLE) although, as only application of a probit model for dropout was considered by Wu and Carroll (1988), it will be referred to in the remainder of this thesis as **probit pseudo-maximium likelihood estimation (PPMLE)**. Even in the case of a probit dropout model, the use of this form of model requires substantial programming and the use of iterative methods. For other dropout models, numerical integration would also be required for evaluation of the likelihood or its logarithm. However, this form of model would appear to show some potential for extending its use to event history or survival modelling due to the natural appeal of selecting for death based on individual's parameters from a repeated measures risk factor model. It will, however, only be applicable when risk is related directly to the parameters of the repeated measures model.

Wu and Bailey (1989) used a random coefficient *pattern-mixture* model, which they called a *conditional linear model*, for data of the form previously considered by Wu and Carroll (1988). It related the conditional expectation of the slope for the repeated measures of the response for an individual, given the dropout time, as a linear function of dropout time, thus leading to different expectations for the different 'patterns' of dropout. The individual slopes were estimated via ordinary least squares, following which two different methods were used for estimation of the population slope coefficient. The first used weighted least squares to obtain the **linear minimum variance unbiased (LMVUB)** estimator which is a linear combination of the parameters of the linear model for dropout time. The other approach was designed to minimise the mean square error under the conditional linear model by suitable choice of weights in forming a weighted average of the individual ordinary least squares slopes for the response. This was termed the **linear minimum mean square error (LMMSE)** estimator. Wu

and Bailey (1988, 1989) provided comparisons of PPMLE, the LMVUB estimator and the LLMSE estimator with the more traditional **unweighted least squares estimator (UWLE)** and **weighted least squares estimator (WLSE)**. They concluded that the former trio of techniques were preferable to the WLSE in terms of bias and to the UWLE in terms of mean square error and, potentially, bias if there was informative dropout. In terms of dropout, it was concluded that at least one of the LMVUB and LMMSE estimators was comparable with or better than PPMLE, the preference between the LMVUB and LMMSE estimators being dependent on the form of dropout process. As LMVUB and LMMSE are simple non-iterative techniques, providing the form of dropout is known, the use of the one with the superior performance with that form of dropout can avoid the necessity to use highly complex iterative techniques such as PPMLE if the aim is to estimate population slope coefficients.

Wu and Lan (1992) used results of simulations from Wu and Carroll (1988) and Wu and Bailey (1988) to propose techniques for sequential monitoring in the comparison of changes between two groups based on the generalised least squares estimator (GLSE) and the UWLE, based on areas under expected response curves and using the spending function approach of Lan and DeMets (1983). They considered a specific situation in which dropout was non-informative (by design) although they also derived the technique under a more general approach for which estimates of the population mean for the individual slopes for each group could be made using alternatives such as the LMMSE estimator, the LMVUB estimator or PPMLE.

As described in Section 1.3.2, Schluchter (1992) investigated the same data structure as Wu and Carroll (1988) and Wu and Bailey (1988, 1989) using a random coefficient selection model, based on a trivariate normal model for the coefficients. This avoided the usual problem with random coefficient models of the evaluation of a potentially analytically intractable integral for a marginal likelihood. Schluchter

(1992) performed no investigation into the sensitivity to the model specification. However the discussion recommended further investigation of the use of models such as this trivariate model, Wu and Carroll's probit dropout model and Wu and Bailey's approaches based on a conditional linear model with a view to the adoption of the most appropriate of these or, alternatively, one of the simpler UWLE, GLSE or WLSE, depending on the sensititivity to plausible dropout proceses. As always, knowledge of the type of dropout process would enable the most appropriate of the techniques to be used in a particular application although the typical case is, of course, ignorance of the true nature of the dropout process although in most cases an educated proposal will usually provide an adequate approximation.

In the remainder of this chapter, a more detailed investigation of Wu and Carroll's model is performed, with specific emphasis on its performance in different circumstances, difficulty in implementation and achieving convergence plus its potential for extension to include more than a single repeated 'measurement' of various distributional forms and application to event history analysis. The reason for choice of this method was it is believed to have been the first model proposed that jointly modelled growth curve data subject to dropout in a manner with some potential for application to survival or event history analysis with repeated assessment of risk factors, although Schluchter (1992) has since proposed an alternative with similar properties.

3.2 Wu and Carroll's model

Wu and Carroll (1988) developed a relatively complex random coefficient selection model for comparing the rates of change between two groups in the presence of informative dropout using *probit pseudo-maximum likelihood estimation* (PPMLE). The potential this model may have for wider application is that it can combine in-

ferences about the informativeness of a dropout process, albeit under an assumed and untestable model structure, with those for the parameters of the growth process.

Wu and Carroll (1988) assumed that the participants in a longitudinal study are randomly divided into two treatment groups although their approach would be generally applicable in other cases. The sample sizes in the two groups are denoted by n_k $(k = 1, 2)$ and the notation $i \epsilon k$ is used to denote group membership for the i^{th} individual. In the derivations and simulations in the remainder of this section, it is assumed that there is no staggered entry, although the theoretical development by Wu and Carroll (1988) demonstrated that the method could be extended to allow for this. The restriction employed in this thesis is merely to simplify the development and resulting simulation. It is also assumed in this section, as in Wu and Carroll (1988), that there are J identical follow-up time points to determine mortality and withdrawal status $(t_1, \ldots, t_J)$ and T times at which a single response is measured. It should be noted that this latter assumption implies that the follow-up times may differ from the measurement times, although the last follow-up and measurement time coincide. Hence, letting $v_i (\leq T)$ denote the total number of observations recorded for the i^{th} individual and Y_{iv} and t_{iv} denote, respectively, the v^{th} response and corresponding measurement time where $t_{i1} = 0$, the following notation may be adopted:

$Z(i,j) = 1$ and $t_{iv_i} < t_J$ if death or withdrawal occurred in the j^{th} interval i.e. between time t_j and t_{j+1} ;

$Z(i,j) = 0$ otherwise,

with $t_{iv_i} = t_J$ if the individual was still in the study at termination.

The i^{th} individual therefore has $T - v_i$ measurements missing due to dropout.

The assumed model for the response variable is that the repeated measurements

follow a linear function of *time* for each individual where:

$$\underline{\boldsymbol{\beta}}_i = (\beta_{i1}, \beta_{i2})^T \sim \mathcal{N}(\underline{\boldsymbol{B}}_k, \boldsymbol{\Sigma}_\beta)$$

represents the unobservable vector of the individual's intercept and slope, i.e. that:

$$\underline{\boldsymbol{Y}}_i = \boldsymbol{X}_i \underline{\boldsymbol{\beta}}_i + \underline{\boldsymbol{\varepsilon}}_i \quad \text{with } \underline{\boldsymbol{\varepsilon}}_i \sim \mathcal{N}(\underline{\mathbf{0}}, \sigma_\varepsilon^2 \boldsymbol{I}), \quad \text{for } i \epsilon k \quad \text{and} \quad k = 1, 2,$$

with $\boldsymbol{X}_i^T = \begin{bmatrix} 1 & \dots & 1 \\ t_{i1} & \dots & t_{iv_i} \end{bmatrix}$, $\boldsymbol{\Sigma}_\beta = \begin{bmatrix} \sigma_{\beta_1}^2 & \sigma_{\beta_1\beta_2} \\ \sigma_{\beta_1\beta_2} & \sigma_{\beta_2}^2 \end{bmatrix}$, $\underline{\boldsymbol{B}}_k{}^T = (B_{k1}, B_{k2})$.

It is assumed that the process generating dropouts has a distribution function of the form $M(\underline{\boldsymbol{\alpha}}^T \underline{\boldsymbol{\beta}}_i, \alpha_{0j})$ where the α_{0j}, $j = 2, \dots, J$ are dropout time parameters for the different intervals and $\underline{\boldsymbol{\alpha}}^T = (\alpha_1, \alpha_2)$. The dropout process is therefore *directly* related to each individual's intercept and slope.

Now for each $\underline{\boldsymbol{\beta}}_i$ the following are sufficient statistics:

(i) $\hat{\underline{\boldsymbol{\beta}}}_i = (\boldsymbol{X}_i^T \boldsymbol{X}_i)^{-1} (\boldsymbol{X}_i^T \underline{\boldsymbol{Y}}_i)$; and

(ii) the dropout time.

Hence, viewing $\underline{\boldsymbol{\beta}}_i$ as a nuisance parameter, estimation of the remaining parameters can be made from the marginal likelihood function, which is given by

$$L(\underline{\boldsymbol{B}}_1, \underline{\boldsymbol{B}}_2, \underline{\boldsymbol{\alpha}}, \alpha_{02}, \dots, \alpha_{0J}) = \prod_{i=1}^{n} L_i$$

where

$$
\begin{aligned}
L_i \;=\; & D \int \phi_2(\hat{\underline{\boldsymbol{\beta}}}_i,\, \underline{\boldsymbol{\beta}}_i,\, \boldsymbol{C}_{1i})\, \phi_2(\underline{\boldsymbol{\beta}}_i,\, \underline{\boldsymbol{B}}_k,\, \boldsymbol{\Sigma}_\beta) \quad \times \\
& \prod_{j=2}^{J} \left[\left(M\left(\underline{\boldsymbol{\alpha}}^T \underline{\boldsymbol{\beta}}_i,\, \alpha_{0j}\right) - M\left(\underline{\boldsymbol{\alpha}}^T \underline{\boldsymbol{\beta}}_i,\, \alpha_{0j-1}\right) \right)^{Z(i,j-1)} \right] \quad \times \\
& \left(1 - M\left(\underline{\boldsymbol{\alpha}}^T \underline{\boldsymbol{\beta}}_i,\, \alpha_{0J}\right)\right)^{(1-\sum_j Z(i,j))}\, d\underline{\boldsymbol{\beta}}_i \,,
\end{aligned}
$$

and where

$$
\begin{aligned}
M\left(\underline{\boldsymbol{\alpha}}^T \underline{\boldsymbol{\beta}}_i,\, \alpha_{01}\right) \;&=\; 0 \qquad\qquad \text{by definition, and} \\
\boldsymbol{C}_{1i} \;&=\; \sigma_\varepsilon^2 \left(\boldsymbol{X}_i^T \boldsymbol{X}_i\right)^{-1},
\end{aligned}
$$

and the notation $\phi_2(\underline{\boldsymbol{Y}},\, \underline{\boldsymbol{\beta}},\, \boldsymbol{\Sigma})$ represents the bivariate normal probability density function with mean vector $\underline{\boldsymbol{\beta}}$ and covariance matrix $\boldsymbol{\Sigma}$. In the above marginal likelihood, the first two terms under the integral sign represent the conditional density of an individual's estimated random effect (given their true random effect) and the density of their true random effect, respectively. The next $J-1$ terms under the integral sign represent the probability of death in each of the $J-1$ intervals and the final term represents the probability of an individual being in the study at termination, each of these J terms being raised to an appropriate power to indicate the interval of death or that an individual survived the whole study. In addition, D is constant with respect to the parameters to be estimated. Note that it is assumed here that $\boldsymbol{\Sigma}_\beta$ and σ_ε^2 are known whereas in reality they are unknown and replaced by unbiased estimates.

If the form of the dropout distribution is assumed to be **probit**, i.e. of the form given by

$$
M\left(\underline{\boldsymbol{\alpha}}^T \underline{\boldsymbol{\beta}}_i,\, \alpha_{0j}\right) \;=\; \Phi\left(\underline{\boldsymbol{\alpha}}^T \underline{\boldsymbol{\beta}}_i + \alpha_{0j}\right),
$$

where $\Phi\left(\underline{\boldsymbol{\alpha}}^T \underline{\boldsymbol{\beta}}_i + \alpha_{0j}\right)$ is the cumulative distribution function of the standard

normal distribution, then the marginal likelihood function is given by

$$\begin{aligned} L_i &= D \int \phi_2(\hat{\underline{\beta}}_i, \underline{\beta}_i, \boldsymbol{C}_{1i})\, \phi_2(\underline{\beta}_i, \underline{\boldsymbol{B}}_k, \boldsymbol{\Sigma}_\beta) \\ &\times \prod_{j=2}^{J} \left[\left(\Phi\left(\underline{\alpha}^T \underline{\beta}_i + \alpha_{0j}\right) - \Phi\left(\underline{\alpha}^T \underline{\beta}_i + \alpha_{0j-1}\right) \right)^{Z(i,j-1)} \right] \\ &\times \left(1 - \Phi\left(\underline{\alpha}^T \underline{\beta}_i + \alpha_{0J}\right)\right)^{(1-\sum_j Z(i,j))} d\underline{\beta}_i \,. \end{aligned}$$

The integration can be performed analytically, as shown below. If the dropout distribution proposed were other than probit, numerical integration would be necessary. In practice, the maximisation of the log-likelihood is difficult but feasible under the assumption of a probit dropout distribution, this being the main reason why Wu and Carroll (1988) used a probit model.

For $i \epsilon k$ and $k = 1, 2$ an individual's contribution to the marginal likelihood is given by

$$\begin{aligned} L_i &= \frac{D}{(2\pi)^2 \mid \boldsymbol{C}_{1i} \mid^{1/2} \mid \boldsymbol{\Sigma}_\beta \mid^{1/2}} \int \exp\left[-\frac{1}{2} \left(\hat{\underline{\beta}}_i - \underline{\beta}_i\right)^T \boldsymbol{C}_{1i}{}^{-1} \left(\hat{\underline{\beta}}_i - \underline{\beta}_i\right) \right. \\ &\quad \left. - \frac{1}{2} \left(\underline{\beta}_i - \underline{\boldsymbol{B}}_k\right)^T \boldsymbol{\Sigma}_\beta{}^{-1} \left(\underline{\beta}_i - \underline{\boldsymbol{B}}_k\right) \right] \\ &\times \prod_{j=2}^{J} \left[\left(\Phi\left(\underline{\alpha}^T \underline{\beta}_i + \alpha_{0j}\right) - \Phi\left(\underline{\alpha}^T \underline{\beta}_i + \alpha_{0j-1}\right) \right)^{Z(i,j-1)} \right] \\ &\times \left(1 - \Phi\left(\underline{\alpha}^T \underline{\beta}_i + \alpha_{0J}\right)\right)^{(1-\sum_j Z(i,j))} d\underline{\beta}_i \quad . \end{aligned}$$

It can be shown (see Appendix A) that this can be written as

$$
\begin{aligned}
L_i &= \frac{D \exp\left[-\frac{1}{2}\left(\underline{\hat{\boldsymbol{\beta}}}_i - \underline{\boldsymbol{B}}_k\right)^T \boldsymbol{C}_{2i}{}^{-1}\left(\underline{\hat{\boldsymbol{\beta}}}_i - \underline{\boldsymbol{B}}_k\right)\right]}{2\pi \mid \boldsymbol{C}_{2i} \mid^{1/2}} \\
&\times \left[\prod_{j=2}^{J}\left\{\Phi\left[\left(\underline{\boldsymbol{d}}_{ik}{}^T \boldsymbol{C}_{3i}\underline{\boldsymbol{\alpha}} + \alpha_{0j}\right)\left(1 + \underline{\boldsymbol{\alpha}}^T \boldsymbol{C}_{3i}\underline{\boldsymbol{\alpha}}\right)^{-\frac{1}{2}}\right]\right.\right. \\
&- \left.\left.\Phi\left[\left(\underline{\boldsymbol{d}}_{ik}{}^T \boldsymbol{C}_{3i}\underline{\boldsymbol{\alpha}} + \alpha_{0j-1}\right)\left(1 + \underline{\boldsymbol{\alpha}}^T \boldsymbol{C}_{3i}\underline{\boldsymbol{\alpha}}\right)^{-\frac{1}{2}}\right]\right\}^{Z(i,j-1)}\right] \\
&\times \left\{1 - \Phi\left[\left(\underline{\boldsymbol{d}}_{ik}{}^T \boldsymbol{C}_{3i}\underline{\boldsymbol{\alpha}} + \alpha_{0J}\right)\left(1 + \underline{\boldsymbol{\alpha}}^T \boldsymbol{C}_{3i}\underline{\boldsymbol{\alpha}}\right)^{-\frac{1}{2}}\right]\right\}^{(1-\sum_j Z(i,j))} ,
\end{aligned}
$$

and taking logs this gives

$$
\begin{aligned}
\log L_i &= \log D + \log\left[\left(2\pi \mid \boldsymbol{C}_{2i} \mid^{1/2}\right)^{-1}\right] \\
&- \frac{1}{2}\left(\underline{\hat{\boldsymbol{\beta}}}_i - \underline{\boldsymbol{B}}_k\right)^T \boldsymbol{C}_{2i}{}^{-1}\left(\underline{\hat{\boldsymbol{\beta}}}_i - \underline{\boldsymbol{B}}_k\right) \\
&+ \left[\sum_{j=2}^{J} Z(i,j-1)\log\left\{\Phi\left[\left(\underline{\boldsymbol{d}}_{ik}{}^T \boldsymbol{C}_{3i}\underline{\boldsymbol{\alpha}} + \alpha_{0j}\right)\left(1 + \underline{\boldsymbol{\alpha}}^T \boldsymbol{C}_{3i}\underline{\boldsymbol{\alpha}}\right)^{-\frac{1}{2}}\right]\right.\right. \\
&- \left.\left.\Phi\left[\left(\underline{\boldsymbol{d}}_{ik}{}^T \boldsymbol{C}_{3i}\underline{\boldsymbol{\alpha}} + \alpha_{0j-1}\right)\left(1 + \underline{\boldsymbol{\alpha}}^T \boldsymbol{C}_{3i}\underline{\boldsymbol{\alpha}}\right)^{-\frac{1}{2}}\right]\right\}\right] \\
&+ (1 - \sum_j Z(i,j)) \\
&\times \log\left\{1 - \Phi\left[\left(\underline{\boldsymbol{d}}_{ik}{}^T \boldsymbol{C}_{3i}\underline{\boldsymbol{\alpha}} + \alpha_{0J}\right)\left(1 + \underline{\boldsymbol{\alpha}}^T \boldsymbol{C}_{3i}\underline{\boldsymbol{\alpha}}\right)^{-\frac{1}{2}}\right]\right\} ,
\end{aligned}
$$

where

$$
\begin{aligned}
\boldsymbol{C}_{2i} &= \boldsymbol{C}_{1i} + \boldsymbol{\Sigma}_\beta , \\
\boldsymbol{C}_{3i} &= (\boldsymbol{C}_{1i}{}^{-1} + \boldsymbol{\Sigma}_\beta{}^{-1})^{-1} , \quad \text{and} \\
\underline{\boldsymbol{d}}_{ik} &= \boldsymbol{C}_{1i}{}^{-1}\underline{\boldsymbol{\beta}}_i + \boldsymbol{\Sigma}_\beta{}^{-1}\underline{\boldsymbol{B}}_k \ .
\end{aligned}
$$

Hence

$$\log L_i \;=\; \log D \;+\; \log A_i \;-\; \frac{1}{2}\left(\underline{\hat{\boldsymbol{\beta}}}_i \;-\; \underline{\boldsymbol{B}}_k\right)^T \boldsymbol{C}_{2i}{}^{-1}\left(\underline{\hat{\boldsymbol{\beta}}}_i \;-\; \underline{\boldsymbol{B}}_k\right) \;+\; T_i\,,$$

where

$$A_i \;=\; \left(2\pi \;|\; \boldsymbol{C}_{2i} \;|^{1/2}\right)^{-1}$$

and

$$\begin{aligned} T_i \;&=\; \sum_{j=2}^{J} \{Z(i,\, j-1)\, \log\left[\Phi(U_{ij}) - \Phi(U_{i\,j-1})\right]\} \\ &+\; \left\{1 - \sum_j Z(i,\, j)\right\} \log\left[1 \;-\; \Phi(U_{iJ})\right]\,, \end{aligned}$$

where $\qquad U_{ij} \;=\; \left(\underline{\boldsymbol{d}}_{ik}{}^T \boldsymbol{C}_{3i}\underline{\boldsymbol{\alpha}} \;+\; \alpha_{0j}\right)\left(1 \;+\; \underline{\boldsymbol{\alpha}}^T \boldsymbol{C}_{3i}\underline{\boldsymbol{\alpha}}\right)^{-\frac{1}{2}}.$

It should be noted that $\boldsymbol{C}_{1i}$, and hence $\boldsymbol{C}_{2i}$ and $\boldsymbol{C}_{3i}$, differ between individuals unless all have the same number of measurements at the same time points. Even in the simplest situation of only two dropout follow-up times, there is joint estimation of seven parameters - the number of parameters increasing by one for each additional follow-up time point. This is likely to lead to a variety of practical problems in the numerical maximisation of the log-likelihood. It may, however, be possible to introduce appropriate constraints on the α_{0j}. For example, if a linear function of time is deemed appropriate, this will limit the number of dropout time

parameters to two and, hence, the total number of parameters to eight.

Wu and Carroll (1988) maximised the log-likelihood function using a Newton-Raphson procedure requiring first and second partial derivatives of this function. However, although it is presumed that these were correctly derived and used by the authors in their analyses and simulations, there were many errors when they were included as the Appendix to the paper. It should be noted that many, but not all, of these errors are simple and obviously typographical. In order to carry out simulations to assess Wu and Carroll's method employing a quasi-Newton-Raphson algorithm, as described in Section 3.2.2, it was therefore necessary to derive both first and second order partial derivatives, with the following results.

Using the notation

$$\boldsymbol{C}_{2i} = \begin{pmatrix} \sigma_{2i1}^2 & \sigma_{2i12} \\ \sigma_{2i12} & \sigma_{2i2}^2 \end{pmatrix}, \quad \boldsymbol{C}_{3i} = \begin{pmatrix} \sigma_{3i1}^2 & \sigma_{3i12} \\ \sigma_{3i12} & \sigma_{3i2}^2 \end{pmatrix},$$

$$\underline{\boldsymbol{d}}_{ik}^T = (d_{ik1}, d_{ik2}), \quad \rho_i = \sigma_{2i12}/(\sigma_{2i1}\sigma_{2i2}),$$

$$\phi_{ij} = \phi(U_{ij}), \qquad \Phi_{ij} = \Phi(U_{ij}), \quad \text{for } j = 2, \ldots, J$$

$$\text{and} \quad \phi_{i1} = \Phi_{i1} = 0 \ ,$$

U_{ij} can be written in the form

$$U_{ij} = (\alpha_{0j} + \alpha_1 d_{ik1}\sigma_{3i1}^2 + \sigma_{3i12}(\alpha_1 d_{ik2} + \alpha_2 d_{ik1}) + \alpha_2 d_{ik2}\sigma_{3i2}^2) \,/\, D^{1/2},$$

$$\text{for } j = 2, \ldots, J \,,$$

where $$D = (1 + \sigma_{3i1}^2\alpha_1^2 + 2\sigma_{3i12}\alpha_1\alpha_2 + \sigma_{3i2}^2\alpha_2^2) \ .$$

The vector of parameters to be estimated is

$$\underline{\boldsymbol{\theta}}^T = (\theta_1, \ldots, \theta_{J+5}) = (B_{11}, B_{12}, B_{21}, B_{22}, \alpha_1, \alpha_2, \alpha_{02}, \ldots, \alpha_{0J}),$$

and the partial derivatives of the log-likelihood with respect to these parameters are as follows:

$$\frac{\partial \log(L_i)}{\partial B_{kl}} = \begin{cases} \dfrac{[(\hat{\beta}_{il} - B_{kl})/\sigma_{2il}^2 \; - \; \rho_i(\hat{\beta}_{im} - B_{km})/(\sigma_{2i1}\sigma_{2i2})]}{(1-\rho_i^2)} + \dfrac{\partial T_i}{\partial B_{kl}} & \text{for } i \,\epsilon\, k; \\ & k = 1,\, 2;\; l = 1, 2;\; m = 3 - l; \\ 0 & \text{otherwise}; \end{cases}$$

$$\frac{\partial^2 \log(L_i)}{\partial B_{k_1 l}\, \partial B_{k_2 p}} = \begin{cases} \dfrac{(-1)^{(l-p+1)} \rho_i^{|l-p|}}{\sigma_{2il}\sigma_{2ip}(1-\rho^2)} + \dfrac{\partial^2 T_i}{\partial B_{k_1 l} \partial B_{k_2 p}} & \text{for } i\epsilon k; \\ & k_1 = k_2 = k; \\ & l = 1, 2;\; p = 1, 2; \\ 0 & \text{otherwise}; \end{cases}$$

$$\frac{\partial \log(L_i)}{\partial \theta_l} = \frac{\partial T_i}{\partial \theta_l} \qquad \text{for} \quad l = 5, \ldots, J+5;$$

$$\frac{\partial^2 \log(L_i)}{\partial \theta_l \partial \theta_m} = \frac{\partial^2 T_i}{\partial \theta_l \partial \theta_m} \qquad \text{for} \quad l = 5, \ldots, J+5; \quad m = 1, \ldots, J+5.$$

When there is no staggered entry,

$$\frac{\partial T_i}{\partial \theta_l} = \sum_{j=2}^{J} \left\{ \frac{Z(i,j-1)[\phi_{ij}(\partial U_{ij}/\partial \theta_l) - \phi_{ij-1}(\partial U_{ij-1}/\partial \theta_l)]}{(\Phi_{ij} - \Phi_{ij-1})} \right\}$$

$$- \left(1 - \sum_{j=2}^{J} Z(i,j-1)\right) \left[\frac{\phi_{iJ}\partial U_{iJ}/\partial \theta_l}{(1-\Phi_{iJ})} \right],$$

$$l = 1, \ldots, J+5;$$

$$\frac{\partial^2 T_i}{\partial \theta_l \partial \theta_m} = \sum_{j=2}^{J} \left\{ Z(i,j-1) \left\{ (\Phi_{ij} - \Phi_{ij-1}) \left[-U_{ij}\phi_{ij} \left(\frac{\partial U_{ij}}{\partial \theta_l}\right) \left(\frac{\partial U_{ij}}{\partial \theta_m}\right) \right.\right.\right.$$

$$+U_{ij-1}\phi_{ij-1} \left(\frac{\partial U_{ij-1}}{\partial \theta_l}\right) \left(\frac{\partial U_{ij-1}}{\partial \theta_m}\right) + \phi_{ij} \left(\frac{\partial^2 U_{ij}}{\partial \theta_l \partial \theta_m}\right) - \phi_{ij-1} \left(\frac{\partial^2 U_{ij-1}}{\partial \theta_l \partial \theta_m}\right) \Big]$$

$$- \left[\phi_{ij} \left(\frac{\partial U_{ij}}{\partial \theta_l}\right) - \phi_{ij-1} \left(\frac{\partial U_{ij-1}}{\partial \theta_l}\right) \right] \left[\phi_{ij} \left(\frac{\partial U_{ij}}{\partial \theta_m}\right) - \phi_{ij-1} \left(\frac{\partial U_{ij-1}}{\partial \theta_m}\right) \right] \Big\}$$

$$/ \left(\Phi_{ij} - \Phi_{ij-1}\right)^2 \Big\} - \left[1 - \sum_{j=2}^{J} Z(i,j-1) \right] [(1-\Phi_{iJ})$$

$$\left\{ -U_{iJ}\phi_{iJ} \left(\frac{\partial U_{iJ}}{\partial \theta_l}\right) \left(\frac{\partial U_{iJ}}{\partial \theta_m}\right) + \phi_{iJ} \left(\frac{\partial^2 U_{iJ}}{\partial \theta_l \partial \theta_m}\right) \right\} + \phi_{iJ}^2 \left(\frac{\partial U_{iJ}}{\partial \theta_l}\right) \left(\frac{\partial U_{iJ}}{\partial \theta_m}\right) \Big]$$

$$/ \left(1 - \Phi_{iJ}\right)^2,$$

$$\text{for} \quad l = 1, \ldots, J+5 \quad \text{and} \quad m = 1, \ldots, J+5;$$

with

$$\frac{\partial U_{ij}}{\partial B_{kl}} = \begin{cases} [\frac{\partial d_{ik1}}{\partial B_{kl}}(\alpha_1\sigma_{3i1}^2 + \alpha_2\sigma_{3i12}) + \frac{\partial d_{ik2}}{\partial B_{kl}}(\alpha_1\sigma_{3i12} + \alpha_2\sigma_{3i2}^2)\,]/D^{1/2} \\ \qquad\qquad \text{for } i\,\epsilon k;\ l = 1,2; \\ 0 \qquad\qquad \text{otherwise;} \end{cases}$$

$$\partial U_{ij}/\partial \alpha_l = [D^{1/2}(d_{ikl}\sigma_{3il}^2 + d_{ik(3-l)}\sigma_{3i12}) - U_{ij}\,C_l]\,/\,D \qquad \text{for } i\,\epsilon\,k;\ l = 1,2;$$

$$\partial U_{ij}/\partial \alpha_{0l} = \begin{cases} D^{-1/2} & \text{for } l = j; \\ 0 & \text{otherwise;} \end{cases}$$

$$\frac{\partial^2 U_{ij}}{\partial B_{kl}\partial \alpha_m} = \begin{cases} \dfrac{D^{1/2}\left(\dfrac{\partial d_{ikm}}{\partial B_{kl}}\,\sigma_{3im}^2 + \dfrac{\partial d_{ik(3-m)}}{\partial B_{kl}}\,\sigma_{3i12}\right) - C_m\dfrac{\partial U_{ij}}{\partial B_{kl}}}{D} \\ \qquad\qquad \text{for } i\,\epsilon\,k;\ l = 1,2; \quad \text{and } m = 1,2; \\ 0 \qquad\qquad \text{otherwise;} \end{cases}$$

$$\frac{\partial^2 U_{ij}}{\partial \alpha_l \partial \alpha_m} = \left\{ D\left[\left(D^{-1/2}C_m\right)\left(d_{ikl}\sigma_{3il}^2 + d_{ik(3-l)}\sigma_{3i12}\right) - \frac{\partial U_{ij}}{\partial \alpha_m}C_l - U_{ij}a_{lm}\right]\right.$$

$$\left. - 2\left[D^{1/2}\left(d_{ikl}\sigma_{3il}^2 + d_{ik(3-l)}\sigma_{3i12}\right) - C_l U_{ij}\right]C_m\right\}\,/\,D^2;$$

$$l = 1,2;\ m = 1,2;$$

$$\frac{\partial^2 U_{ij}}{\partial \alpha_l \partial \alpha_{0j}} = -C_l/D^{3/2} \qquad \text{for } l = 1,2;$$

$$\frac{\partial^2 U_{ij}}{\partial B_{kl}\partial B_{km}} = \frac{\partial^2 U_{ij}}{\partial B_{kl}\partial \alpha_{0j}} = \frac{\partial^2 U_{ij}}{\partial \alpha_{0j_1}\partial \alpha_{0j_2}} = 0$$

for $l = 1, 2$; $m = 1, 2$; $i \epsilon k$; $j_1 = 2, \ldots, J$;

$j_2 = 2, \ldots, J$; and $j = 2, \ldots, J$;

where $C_\gamma = \sigma^2_{3i\gamma}\alpha_\gamma + \sigma_{3i12}\alpha_{3-\gamma}$ for $\gamma = 1, 2$;

$$\text{and} \quad a_{lm} = \begin{cases} \sigma^2_{3il} & \text{for } l = m; \\ \sigma_{3i12} & \text{for } l \neq m. \end{cases}$$

3.2.1 Testing the significance of parameters

There will be little or no interest in the α_{0j}s, so significance testing will be restricted to the $\underline{\boldsymbol{B}}_k$ (k=1, 2) and $\underline{\boldsymbol{\alpha}}$. In all practical situations, the dropout process will be non-informative with respect to $\underline{\boldsymbol{B}}_k$ providing that $\alpha_1 = \alpha_2 = 0$. Otherwise, this process will be informative with respect to $\underline{\boldsymbol{B}}_k$. Of particular interest in most applications is the informativeness with respect to B_{k2}, that is whether $\alpha_2 = 0$.

To test for informativeness of the dropout process, likelihood-ratio tests can be performed. The following four hypotheses may be constructed:

$$\boldsymbol{H}_0 \;:\; \alpha_1 = 0,\ \alpha_2 = 0$$

$$\boldsymbol{H}_1 \;:\; \alpha_1 \neq 0,\ \alpha_2 = 0$$

$$\boldsymbol{H}_2 \;:\; \alpha_1 = 0,\ \alpha_2 \neq 0$$

$$\boldsymbol{H}_3 \;:\; \alpha_1 \neq 0,\ \alpha_2 \neq 0$$

In simulations perfomed by Wu and Carroll, and those performed by the author and discussed later in this chapter, the testing is of hypothesis $\boldsymbol{H}_0$ against $\boldsymbol{H}_1$ and of $\boldsymbol{H}_0$ against $\boldsymbol{H}_2$. If $\boldsymbol{H}_0$ is true, the dropout process will be *non-informative*, otherwise it will be *informative* with respect to $\underline{\boldsymbol{B}}_k$. Moreover, if $\boldsymbol{H}_0$ is true, the term T_i does not depend on the $\underline{\boldsymbol{B}}_k$ and so maximising L reduces, in terms of $\underline{\boldsymbol{B}}_k$, to solving the equation:

$$\begin{aligned} \frac{\partial \log L}{\partial \underline{\boldsymbol{B}}_k} &= \sum_{i \epsilon k} {\boldsymbol{C}_{2i}}^{-1} \underline{\boldsymbol{B}}_k - \sum_{i \epsilon k} {\boldsymbol{C}_{2i}}^{-1} \underline{\hat{\boldsymbol{\beta}}}_i \\ &= 0\,, \qquad \text{for a maximum.} \end{aligned}$$

So

$$\left(\sum_{i \epsilon k} {\boldsymbol{C}_{2i}}^{-1}\right) \underline{\hat{\boldsymbol{B}}}_k = \sum_{i \epsilon k} \left({\boldsymbol{C}_{2i}}^{-1} \underline{\hat{\boldsymbol{\beta}}}_i\right),$$

or

$$\underline{\hat{\boldsymbol{B}}}_k = \left(\sum_{i \epsilon k} {\boldsymbol{C}_{2i}}^{-1}\right)^{-1} \sum_{i \epsilon k} \left({\boldsymbol{C}_{2i}}^{-1} \underline{\hat{\boldsymbol{\beta}}}_i\right).$$

This is the weighted or *generalized least squares estimator* (GLSE) of $\underline{\boldsymbol{B}}_k$ and will be referred to using the notation $\underline{\hat{\boldsymbol{B}}}_{GL,k}$, as used by Wu and Carroll. If all

individuals have the same number of measurements, the above expression reduces to simply

$$\underline{\hat{\boldsymbol{B}}}_k = \sum_{i \epsilon k} \frac{\underline{\hat{\boldsymbol{\beta}}}_i}{n_k} \quad ,$$

which is the *unweighted least squares estimator* (UWLE) of $\underline{\boldsymbol{B}}_k$ and will be denoted by $\underline{\hat{\boldsymbol{B}}}_{UW,k}$. Their respective covariance matrices can be shown to be given by

$$\boldsymbol{C}_{GL,k} = \left[\sum_{i \epsilon k} \boldsymbol{C}_{2i}{}^{-1}\right]^{-1} \quad \text{and} \quad \boldsymbol{C}_{UW,k} = \left[\sum_{i \epsilon k} \boldsymbol{C}_{2i}\right] n_k^{-2} \quad .$$

In carrying out tests of equality of initial values in the two groups ($B_{11} = B_{21}$) and equality of rates of change in the two groups ($B_{12} = B_{22}$) the test statistic is

$$\frac{\hat{B}_{1l} - \hat{B}_{2l}}{\left[\sigma^2_{\hat{B}_{1l}} + \sigma^2_{\hat{B}_{2l}}\right]^{\frac{1}{2}}} \sim \mathcal{N}(0,1), \qquad \begin{array}{l}\text{under the null hypothesis} \\ \text{of no difference .}\end{array} \tag{3.1}$$

In practice, the variances $\sigma^2_{\hat{B}_{1l}}$ and $\sigma^2_{\hat{B}_{2l}}$ may be replaced by their large sample estimates. Adopting the notation $\underline{\hat{\boldsymbol{B}}}_{PM,k}$ for the probit pseudo-maximum likelihood estimates, $\underline{\hat{\boldsymbol{B}}}_k$ can be any one of $\underline{\hat{\boldsymbol{B}}}_{PM,k}$, $\underline{\hat{\boldsymbol{B}}}_{GL,k}$ and $\underline{\hat{\boldsymbol{B}}}_{UW,k}$ in the above. In the case of $\underline{\hat{\boldsymbol{B}}}_{PM,k}$, the estimates of $\sigma^2_{\hat{B}_{kl}}$ ($l = 1, 2$) are obtained from the inverse of the sample information matrix.

3.2.2 Comparing Wu and Carroll's model with standard estimation methods

In order to gain an appreciation of the feasibility and usefulness of Wu and Carroll's method for extending to modelling event history data, it was decided to use Monte Carlo simulation to carry out a comparison of their method with unweighted least squares and generalised least squares estimation in a similar way to that used in the original paper. As no standard software was available for implementing Wu and Carroll's method, it was necessary to write a program to perform the simulations. This was done in Fortran 77 using double-precision arithmetic. The data simulated were intended to correspond roughly to repeated measurements of diastolic blood pressures for two treatment groups, each group consisting of 100 individuals. NAG subroutines and functions were employed as follows:

G05CBF & G05DDF	to simulate standard normal random variables;
G05CAF	to simulate uniform random variables on [0, 1];
S15ABF	to calculate cumulative normal probabilities;
G02CAF	to calculate the least squares regression coefficients for the individuals' growth curves;
F02ABF	to calculate the eigenvalues and eigenvectors of the sample variance-covariance matrix for the individual's intercept and slope and then used to check that the matrix is positive-definite;
E04LAF	a quasi-Newton-Raphson minimisation algorithm used to *minimise* the *negative* of the log marginal likelihood function using both first and second partial derivatives;

F01AAF an approximate matrix inversion procedure to calculate the sample variance-covariance matrix by inversion of the observed information matrix.

Five hundred such simulations were performed within each program. There were three basic sets of random effects' mean parameters. The first was where the first group had blood pressures such that $\underline{\boldsymbol{B}}_1^T = (75, 1.5)$ and the second group had $\underline{\boldsymbol{B}}_2^T = (75, 1.0)$, the units being millimetres of mercury (mm Hg). This set was intended to represent a two treatment randomised controlled designed experiment where the two groups had equal expected initial blood pressures but were subject to different expected treatment effects.

The second set was where the first group had blood pressures such that $\underline{\boldsymbol{B}}_1^T = (75, 1.5)$ and the second group had $\underline{\boldsymbol{B}}_2^T = (100, 1.0)$. The intention here was to allow for a more general case where a study was being performed to investigate differences between two populations, as will frequently be the case with survival data (eg. differences between the sexes).

The third set was where $\underline{\boldsymbol{B}}_1^T = (75, 1.0)$ and $\underline{\boldsymbol{B}}_2^T = (75, 1.0)$, covering the situation where there was no real difference between the groups under comparison.

In each case there were six equally spaced time-points at which blood pressure measurements were simulated ($T = 6$). The variance-covariance matrix for the individual's intercept and slope equalled:

$$\boldsymbol{\Sigma}_\beta = \begin{bmatrix} 100 & -3.5 \\ -3.5 & 0.49 \end{bmatrix}.$$

The within-individual measurement error standard deviation, σ_ε, was assumed to be 1 mm Hg.

The follow-up times were chosen as identical to the measurement times for the purpose of simulation but, for estimation purposes, the second dropout follow-up point was chosen to be the third measurement time, thus making the first follow-up interval twice the width of the other three. This was due to the necessary elimination of any individual with only one measurement as the least squares estimation of such an individual's slope was unfeasible in these circumstances. Appropriate values of the interval-specific dropout process parameters, the α_{0j}, were chosen to ensure sufficient individuals dropped out during each of the time intervals to provide reasonable power for the testing for informativeness of the probit dropout process. The values chosen for $\underline{\boldsymbol{\alpha}}^T$ were (0.02, 0.5) and (0.015, 0.75) which, in conjunction with the α_{0j} gave expected percentages of individuals dropping out during each of the four follow-up time periods as given in Table 3.1.

For each set of parameter values, the simulated power for testing for the significance of each of α_1 and α_2 was estimated by calculating the observed proportion of significant results for each of the chosen values of $\underline{\boldsymbol{\alpha}}$. The simulated significance level was estimated by setting the value of the parameter of interest to zero and calculating the proportion of likelihood-ratio test statistics exceeding the 5% value from the χ^2-distribution on one degree of freedom. The results are given in Table 3.2 and show that the simulated significance level is in the region of the desired true value of 0.05 in each case. The 95% confidence intervals for the significance level are consistent with the true level being 5%. It should be noted that, as a repeatable random number generator was used, the six confidence intervals presented in Table 3.2 are not independent.

Estimated power curves together with approximate 95% confidence intervals for α_1 are given in Figures 3.1 and 3.2 for the cases where $\underline{\boldsymbol{B}}_1$ and $\underline{\boldsymbol{B}}_2$ were different. The curves were approximated using cubic splines. The power reaches 90% for a

Non-informative dropout $\alpha_1 = \alpha_2 = 0$
14.92
29.91
25.37
19.78
10.03

$\underline{B}_k^T$	Informative dropout $\underline{\alpha}^T$	
	(0.02, 0.5)	(0.015, 0.75)
(75, 1.5)	11.76	11.76
	41.36	41.36
	29.73	29.73
	13.72	13.72
	3.42	3.42
(75, 1.0)	8.21	6.73
	35.80	32.73
	32.17	32.86
	18.14	20.47
	5.68	7.21
(100, 1.0)	16.07	11.76
	45.78	41.36
	26.25	29.73
	9.92	13.72
	1.98	3.42

Table 3.1: Expected percentages dropping out in follow-up periods 1, 2, 3 and 4 followed by those completing the study for different values of $\underline{B}_k$ under a non-informative or informative probit dropout mechanism, where $\alpha_{02} = -3.20$, $\alpha_{03} = -2.10$, $\alpha_{04} = -1.26$ and $\alpha_{05} = -0.40$.

value of α_1 around 0.017 if $B_{11} = 75$ and $B_{21} = 100$ whereas a value of around 0.025 will be required to achieve 90% power if $B_{11} = B_{21} = 75$. The power curves for α_2, again just considering the two cases where $\underline{\boldsymbol{B}}_1$ and $\underline{\boldsymbol{B}}_2$ were different, are more similar (Figures 3.3 and 3.4). For power of 90% a value of α_2 of around 0.42 would be necessary when $B_{21} = 1.5$ and $B_{22} = 1.0$ whereas, to achieve the same power when $B_{21} = B_{22} = 1.0$, a value of α_2 of around 0.45 would be required.

The estimates and mean squared errors for α_1 and α_2 are given in Table 3.3 for the three basic sets of random effects' parameters. In the presence of informative dropout, the magnitude of the observed percentage bias in estimating α_1 ranged from approximately 0.5% to 10.0% and that in estimating α_2 from around 0.9% to 4.0%, demonstrating that, although the estimation of $\underline{\boldsymbol{\alpha}}$ is relatively unbiased in many cases, the bias could sometimes be moderately large, especially when the random effects' mean parameters were lower and particularly for α_1. The bias in the estimation of each of α_1 and α_2 was negative when $\underline{\boldsymbol{B}}_1^T = (75, 1.5)$ and $\underline{\boldsymbol{B}}_2^T = (100, 1.0)$. When $\underline{\boldsymbol{B}}_1^T = (75, 1.5)$ and $\underline{\boldsymbol{B}}_2^T = (75, 1.0)$, α_2 was slightly underestimated but α_1 was slightly overestimated. The bias was, however, positive for the estimation of both α_1 and α_2 when $\underline{\boldsymbol{B}}_1^T = \underline{\boldsymbol{B}}_2^T = (75, 1.0)$.

	Significance level			
	H_1 v H_0; (for α_1)		H_2 v H_0; (for α_2)	
	Estimate	95% C.I.	Estimate	95% C.I.
$\underline{B}_1^T$ = (75, 1.5) $\underline{B}_2^T$ = (75, 1.0)	0.052	(0.033, 0.071)	0.046	(0.028, 0.064)
$\underline{B}_1^T$ = (75, 1.5) $\underline{B}_2^T$ = (100, 1.0)	0.046	(0.028, 0.064)	0.052	(0.033, 0.071)
$\underline{B}_1^T$ = (75, 1.0) $\underline{B}_2^T$ = (75, 1.0)	0.064	(0.043, 0.085)	0.038	(0.021, 0.055)

Table 3.2: Simulated significance levels (including 95% confidence intervals) for testing hypotheses relating to dropout parameters α_1 and α_2 for three sets of random effects' mean parameters using PPMLE.

	Non-informative dropout			
	$\alpha_1 = \alpha_2 = 0$			
	$\hat{\alpha}_1$	MSE	$\hat{\alpha}_2$	MSE
$\underline{B}_1^T = (75, 1.5)$ $\underline{B}_2^T = (75, 1.0)$	-3.55×10^{-4}	9.29×10^{-5}	-9.87×10^{-3}	2.68×10^{-2}
$\underline{B}_1^T = (75, 1.5)$ $\underline{B}_2^T = (100, 1.0)$	-2.00×10^{-5}	3.95×10^{-5}	-7.99×10^{-3}	3.21×10^{-2}
$\underline{B}_1^T = (75, 1.0)$ $\underline{B}_2^T = (75, 1.0)$	-3.96×10^{-4}	1.02×10^{-4}	-1.07×10^{-2}	3.55×10^{-2}

	Informative dropout							
	$\alpha_1 = 0.02 \quad \alpha_2 = 0.5$				$\alpha_1 = 0.015 \quad \alpha_2 = 0.75$			
	$\hat{\alpha}_1$	MSE	$\hat{\alpha}_2$	MSE	$\hat{\alpha}_1$	MSE	$\hat{\alpha}_2$	MSE
$\underline{B}_1^T = (75, 1.5)$ $\underline{B}_2^T = (75, 1.0)$	0.0199	1.01×10^{-4}	0.492	3.50×10^{-2}	0.0152	8.68×10^{-5}	0.726	3.64×10^{-2}
$\underline{B}_1^T = (75, 1.5)$ $\underline{B}_2^T = (100, 1.0)$	0.0194	7.03×10^{-5}	0.492	6.71×10^{-2}	0.0147	5.62×10^{-5}	0.734	5.36×10^{-2}
$\underline{B}_1^T = (75, 1.0)$ $\underline{B}_2^T = (75, 1.0)$	0.0211	1.18×10^{-4}	0.520	5.05×10^{-2}	0.0165	1.23×10^{-4}	0.757	5.39×10^{-2}

Table 3.3: Results of PPMLE of $\underline{\alpha}^T = (\alpha_1, \alpha_2)$ for simulations for various combinations of $\underline{\alpha}$ and $\underline{\beta}$.

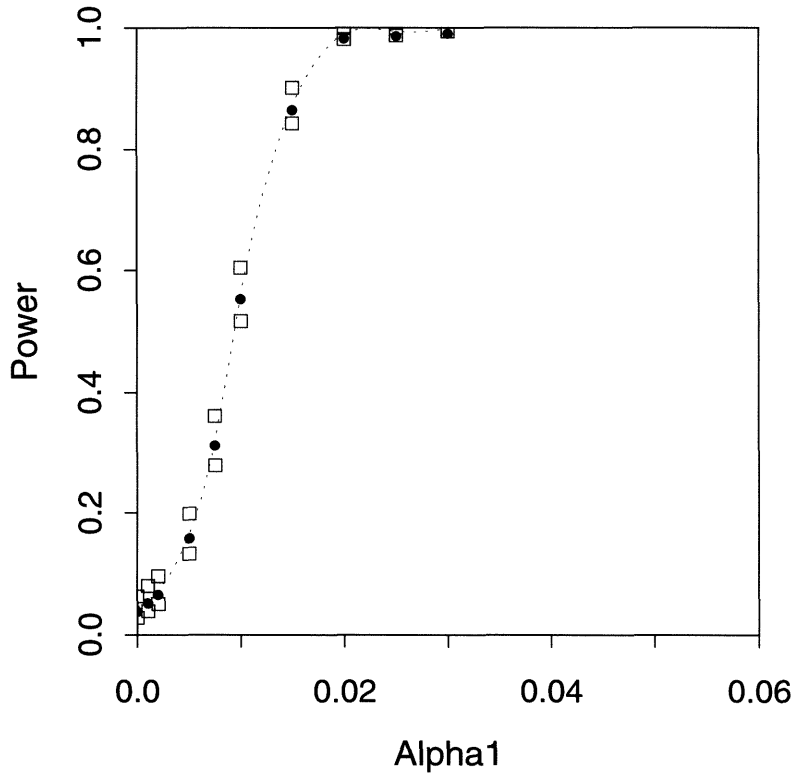

Figure 3.1: Power estimates via PPMLE (•) and corresponding 95% confidence limits (□) for chosen values of α_1 from simulations with $\underline{B}_1^T = (75, 1.5)$ and $\underline{B}_2^T = (100, 1.0)$: power curve fitted to estimates using a cubic spline.

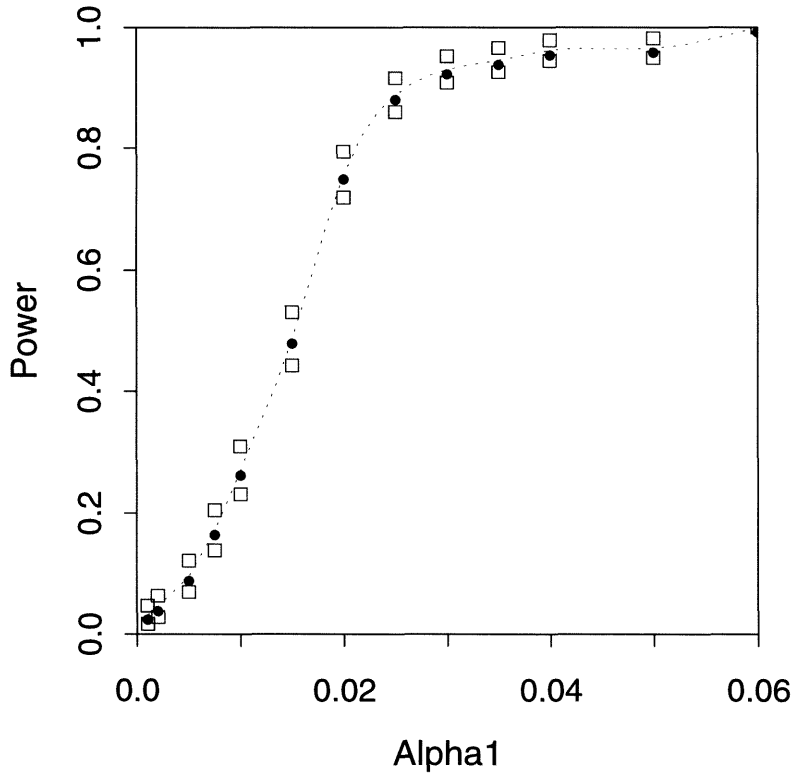

Figure 3.2: Power estimates via PPMLE (•) and corresponding 95% confidence limits (□) for chosen values of α_1 from simulations with $\underline{B}_1^T = (75, 1.5)$ and $\underline{B}_2^T = (75, 1.0)$: power curve fitted to estimates using a cubic spline.

The three different methods used for the estimation of $\underline{\boldsymbol{B}}_k (k = 1, 2)$ were PPMLE, the GLSE and the UWLE. The estimates, averaged over the five hundred simulations performed for each, and their mean square errors are given in Table 3.4 for equal true group intercepts of 75 mm Hg and true group slopes of 1.5 and 1.0 mm Hg for Groups 1 and 2 respectively. The estimates and MSEs of the between-group differences are also given. Table 3.5 contains the corresponding values for the second set of group mean intercepts and slopes. Results from the third set of simulations are not included as these were similar to the results from the second group of the second set. This is because both groups in the third set had the same parameter values as this group. With regard to the estimation of the intercept, negligible differences between the performance of the techniques nor any bias are observed in the case of non-informative dropout. In the presence of informative dropout, the bias in the estimation of the intercept is closely linked to that in the estimation of the slope as the estimators are negatively correlated. From the first set of group mean intercepts and slopes, the percentage bias in estimating the slope is low for each of PPMLE (0.02%), the GLSE (0.05%) and the UWLE (0.04%) when there is non-informative dropout. As the individuals who are censored early will tend to have larger slopes but lower weights in the GLSE, it is clear that the GLSE will be biased downwards when the dropout is informative. This downward bias can be observed to be of magnitude 0.0769 (5.1%) when $\alpha_2 = 0.5$, $B_{k2} = 1.5$, magnitude 0.0733 (7.3%) when $\alpha_2 = 0.5$, $B_{k2} = 1.0$, magnitude 0.1212 (8.1%) when $\alpha_2 = 0.75$, $B_{k2} = 1.5$ and magnitude 0.1070 (10.7%) when $\alpha_2 = 0.75$, $B_{k2} = 1.0$. In comparing the UWLE and PPMLE, the bias is of similar magnitude when there is informative dropout, the bias being slightly smaller in magnitude for the UWLE than PPMLE. This is offset, however, by the lower mean square error under PPMLE. The performance is generally slightly worse in the second set of simulations, where the second group intercept was 100 rather than 75, due to a higher level of dropout in the early follow-up periods.

In terms of the power for each of the estimation approaches in detecting differences between B_{11} and B_{12} of given magnitudes, there appears to be a very slight decline in power for increasing values of α_1, as shown in Figures 3.5-3.8 but little or no differences in power between the UWLE, the GLSE and PPMLE. In the estimation of differences between B_{21} and B_{22} the power appears to decline slightly with increasing the value of α_2 for PPMLE and the GLSE, as shown in Figures 3.9-3.12, although use of the UWLE led to a loss of power of up to around 10%. When $\alpha_1 = \alpha_2 = 0$ the power was 80% for detecting differences between B_{12} and B_{11} of approximately 3.3 and 90% for detecting corresponding differences of around 3.8 (Figure 3.5). For the same value of $\underline{\boldsymbol{\alpha}}$, the power was 80% for detecting differences between B_{21} and B_{22} of approximately 0.31 and the power was 90% for detecting corresponding differences of approximately 0.38 (Figure 3.9).

In many applications, the primary interest will be in making inferences regarding between-group differences. In making such inferences, much of the bias in the individual estimates cancels, so the main disadvantage of the GLSE all but vanishes. In fact, although the GLSE still results in larger bias in estimating between-slope differences, the mean square error is consistently lower than that of either the UWLE or PPMLE. Whilst Wu and Carroll (1988) observed PPMLE to lead to lower MSEs than the GLSE in their simulations, there was little relative difference between them, as seen here also, so there would appear to be little advantage in using PPMLE if the sole or main aim is to make inferences about between-group differences. Likewise, in making inferences about the individual intercepts and slopes, the UWLE provides estimates comparable, in terms of both bias and mean square error, with PPMLE. Thus the sole realistic advantage of PPMLE is the knowledge gained regarding the informativeness of the dropout process. If this is not of interest, then there would appear to be little gain in the computationally complex and expensive PPMLE over the GLSE. However, as the information obtained about the dropout process is dependent upon the correctness of the assumed model and this can never be fully validated due to the

missing measurements following dropout, it would appear that there is little to be gained by using PPMLE for *growth curve analysis*. However, it should be recalled at this stage that the rationale for this investigation was to consider the potential for application of PPMLE to event history modelling. For this purpose, the UWLE and the GLSE would be inappropriate as they do not attempt to consider or model the dropout process.

The method for checking the goodness-of-fit of the model proposed by Wu and Carroll (1988) is to simply compare the modelled survival distribution with the empirical survival distribution. In the simulations with a probit dropout process, the model fitted the data well.

	Non-informative dropout $\alpha_1 = \alpha_2 = 0$			
	Intercept	MSE	Slope	MSE
$\underline{B}_1^T = (75, 1.5)$				
UWLE	74.957	0.9192	1.4994	9.912×10^{-3}
GLSE	74.956	0.9176	1.5008	7.036×10^{-3}
PPMLE	74.957	0.9180	1.4997	7.419×10^{-3}
$\underline{B}_2^T = (75, 1.0)$				
UWLE	74.993	1.0120	1.0047	10.161×10^{-3}
GLSE	74.993	1.0131	1.0039	7.876×10^{-3}
PPMLE	74.994	1.0120	1.0020	8.295×10^{-3}
$(\underline{B}_1 - \underline{B}_2)^T = (0, 0.5)$				
UWLE	-0.036	1.6767	0.4947	0.2058
GLSE	-0.037	1.6753	0.4969	0.1498
PPMLE	-0.037	1.6771	0.4977	0.1503

	Informative dropout $\alpha_1 = 0.02 \quad \alpha_2 = 0.5$				$\alpha_1 = 0.015 \quad \alpha_2 = 0.75$			
	Intercept	MSE	Slope	MSE	Intercept	MSE	Slope	MSE
$\underline{B}_1^T = (75, 1.5)$								
UWLE	74.946	1.0369	1.4814	10.596×10^{-3}	75.150	0.9461	1.4590	10.892×10^{-3}
GLSE	74.981	1.0324	1.4231	14.098×10^{-3}	75.199	0.9635	1.3788	22.471×10^{-3}
PPMLE	74.949	1.0334	1.4795	8.999×10^{-3}	75.154	0.9487	1.4577	10.075×10^{-3}
$\underline{B}_2^T = (75, 1.0)$								
UWLE	74.945	1.1617	0.9837	9.704×10^{-3}	75.115	0.9802	0.9690	8.422×10^{-3}
GLSE	74.981	1.1603	0.9267	13.518×10^{-3}	75.165	0.9959	0.8930	18.092×10^{-3}
PPMLE	74.946	1.1631	0.9810	9.618×10^{-3}	75.116	0.9804	0.9669	8.268×10^{-3}
$(\underline{B}_1 - \underline{B}_2)^T = (0, 0.5)$								
UWLE	0.001	3.0271	0.4977	0.2582	0.035	2.8847	0.4900	0.2443
GLSE	0.000	3.0293	0.4964	0.2547	0.035	2.8830	0.4858	0.2370
PPMLE	0.003	3.0293	0.4985	0.2567	0.038	2.8762	0.4908	0.2420

Table 3.4: Results of simulations with different estimation procedures under a linear random effects model with non-informative or probit informative dropout with various dropout parameter values and with $\underline{B}_1^T = (75, 1.5)$ and $\underline{B}_2^T = (75, 1.0)$.

	Non-informative dropout $\alpha_1 = \alpha_2 = 0$			
	Intercept	MSE	Slope	MSE
$\underline{B}_1^T = (75, 1.5)$				
UWLE	74.957	0.9192	1.4994	9.912×10^{-3}
GLSE	74.956	0.9176	1.5008	7.036×10^{-3}
PPMLE	74.957	0.9182	1.4997	7.419×10^{-3}
$\underline{B}_2^T = (100, 1.0)$				
UWLE	99.993	1.0131	1.0047	10.161×10^{-3}
GLSE	99.993	1.0120	1.0039	7.876×10^{-3}
PPMLE	99.994	1.0125	1.0029	8.520×10^{-3}
$(\underline{B}_1 - \underline{B}_2)^T = (-25, 0.5)$				
UWLE	-25.037	1.6767	0.4946	0.2058
GLSE	-25.037	1.6753	0.4969	0.1498
PPMLE	-25.037	1.6765	0.4966	0.1501

	Informative dropout $\alpha_1 = 0.02 \quad \alpha_2 = 0.5$				$\alpha_1 = 0.015 \quad \alpha_2 = 0.75$			
	Intercept	MSE	Slope	MSE	Intercept	MSE	Slope	MSE
$\underline{B}_1^T = (75, 1.5)$								
UWLE	74.930	0.9919	1.4814	11.07×10^{-3}	75.179	1.0193	1.4529	11.71×10^{-3}
GLSE	74.965	1.0003	1.4214	14.54×10^{-3}	75.229	1.0414	1.3687	25.04×10^{-3}
PPMLE	74.932	0.9979	1.4779	9.69×10^{-3}	75.182	1.0210	1.4498	11.04×10^{-3}
$\underline{B}_2^T = (100, 1.0)$								
UWLE	99.946	1.0657	0.9656	13.00×10^{-3}	100.13	0.9622	0.9568	11.18×10^{-3}
GLSE	99.979	1.0690	0.9069	17.72×10^{-3}	100.18	0.9781	0.8778	22.31×10^{-3}
PPMLE	99.947	1.0667	0.9676	10.93×10^{-3}	100.13	0.9643	0.9588	9.72×10^{-3}
$(\underline{B}_1 - \underline{B}_2)^T = (-25, 0.5)$								
UWLE	-25.016	1.9532	0.5158	0.2356	-24.951	1.8939	0.4961	0.1918
GLSE	-25.014	1.9480	0.5144	0.1764	-24.948	1.8966	0.4910	0.1512
PPMLE	-25.014	1.9478	0.5104	0.1775	-24.949	1.8975	0.4910	0.1482

Table 3.5: Results of simulations with different estimation procedures under a linear random effects model with non-informative or probit informative dropout with various dropout parameter values and with $\underline{B}_1^T = (75, 1.5)$ and $\underline{B}_2^T = (100, 1.0)$.

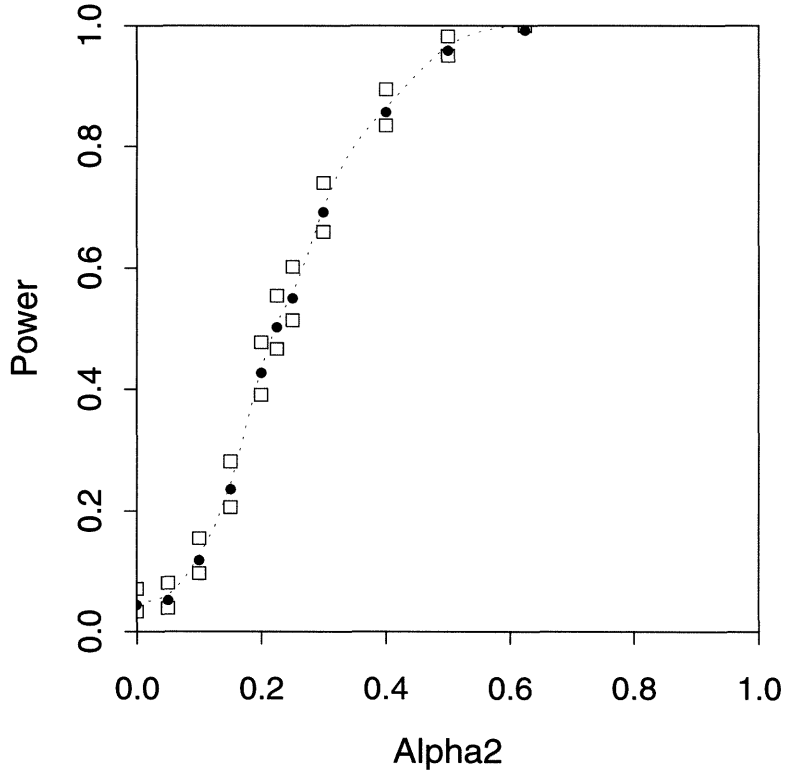

Figure 3.3: Power estimates via PPMLE (•) and corresponding 95% confidence limits (□) for chosen values of α_2 from simulations with $\underline{B}_1^T = (75, 1.5)$ and $\underline{B}_2^T = (100, 1.0)$: power curve fitted to estimates using a cubic spline.

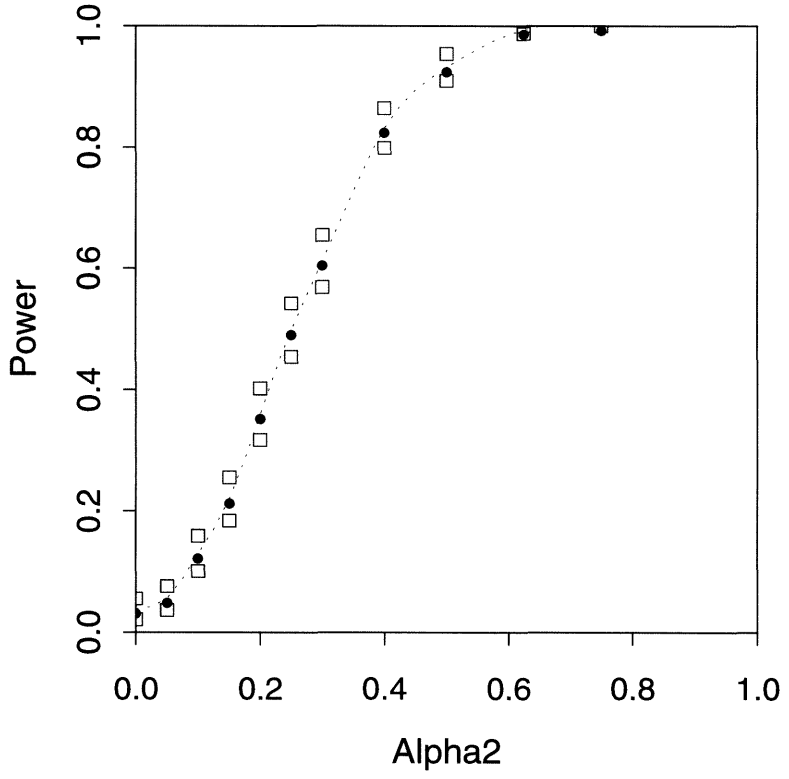

Figure 3.4: Power estimates (•) via PPMLE and corresponding 95% confidence limits (□) for chosen values of α_2 from simulations with $\underline{B}_1^T = (75, 1.0)$ and $\underline{B}_2^T = (75, 1.0)$: power curve fitted to estimates using a cubic spline.

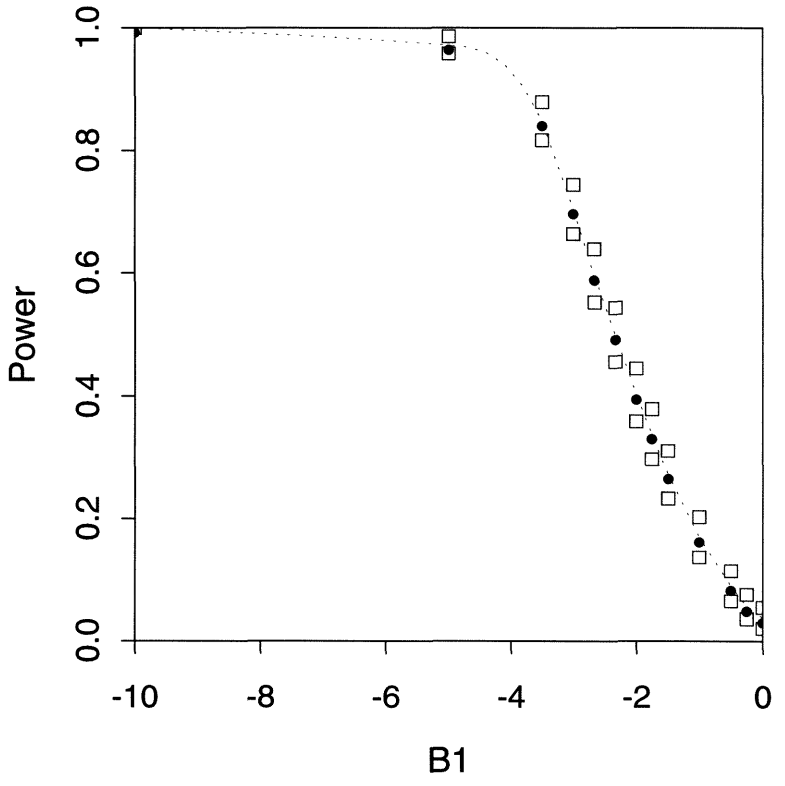

Figure 3.5: Power estimates for PPMLE (•) and corresponding 95% confidence limits (□) for chosen values of $B_{11} - B_{12} = B1$ from simulations with $\underline{\alpha}^T = (0, 0)$: power curve fitted to estimates using a cubic spline.

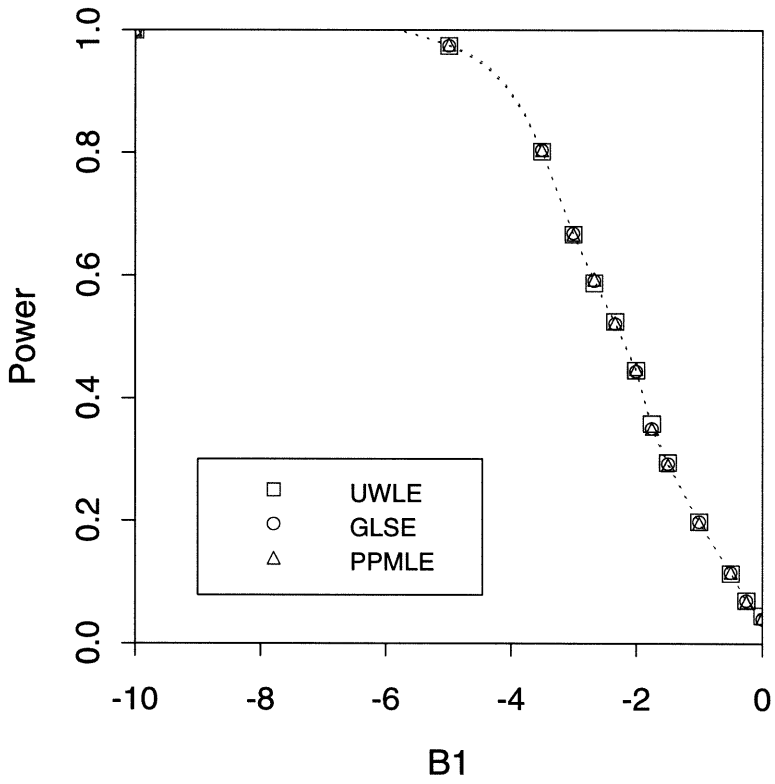

Figure 3.6: Power curve estimates for $B_{11} - B_{12} = B1$ for the UWLE, the GLSE and PPMLE from simulations with $\underline{\alpha}^T = (0.02, 0.5)$: power curve fitted to estimates using a cubic spline.

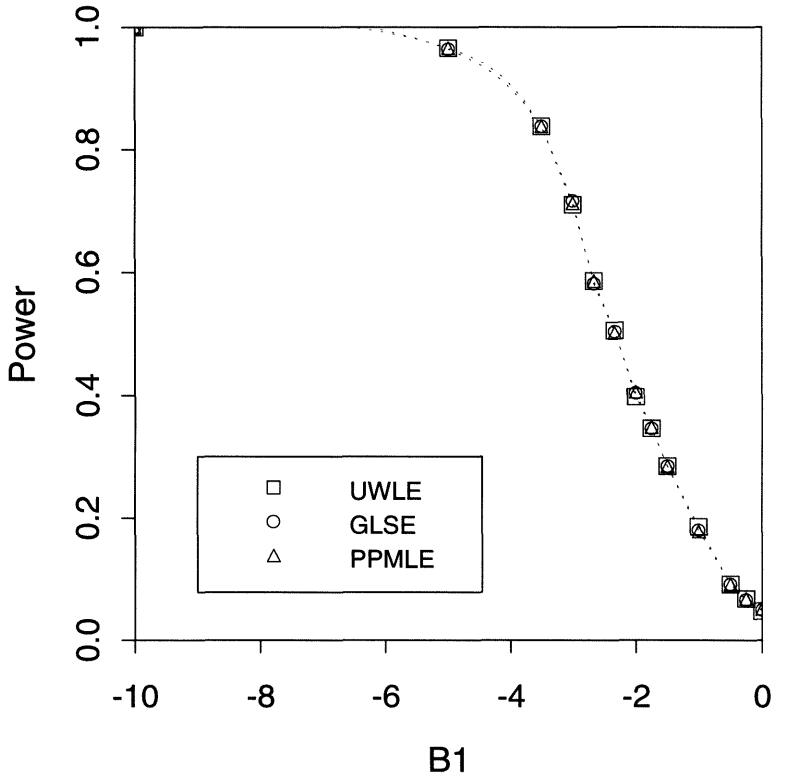

Figure 3.7: Power curve estimates for $B_{11} - B_{12} = B1$ for the UWLE, the GLSE and PPMLE from simulations with $\underline{\alpha}^T = (0.015, 0.75)$: power curve fitted to estimates using a cubic spline.

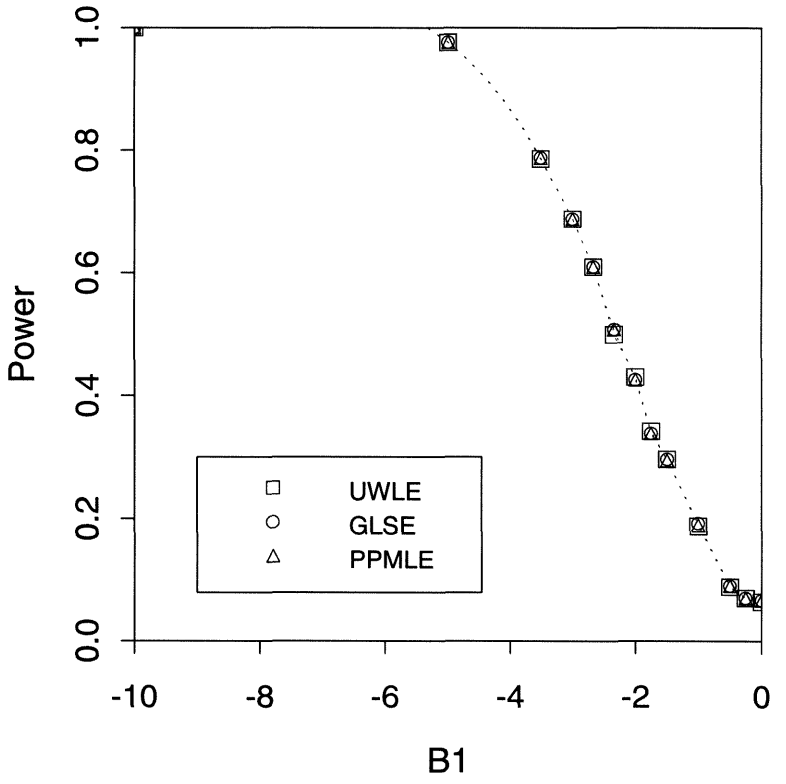

Figure 3.8: Power curve estimates for $B_{11} - B_{12} = B1$ for the UWLE, the GLSE and PPMLE from simulations with $\underline{\alpha}^T = (0.025, 0.25)$: power curve fitted to estimates using a cubic spline.

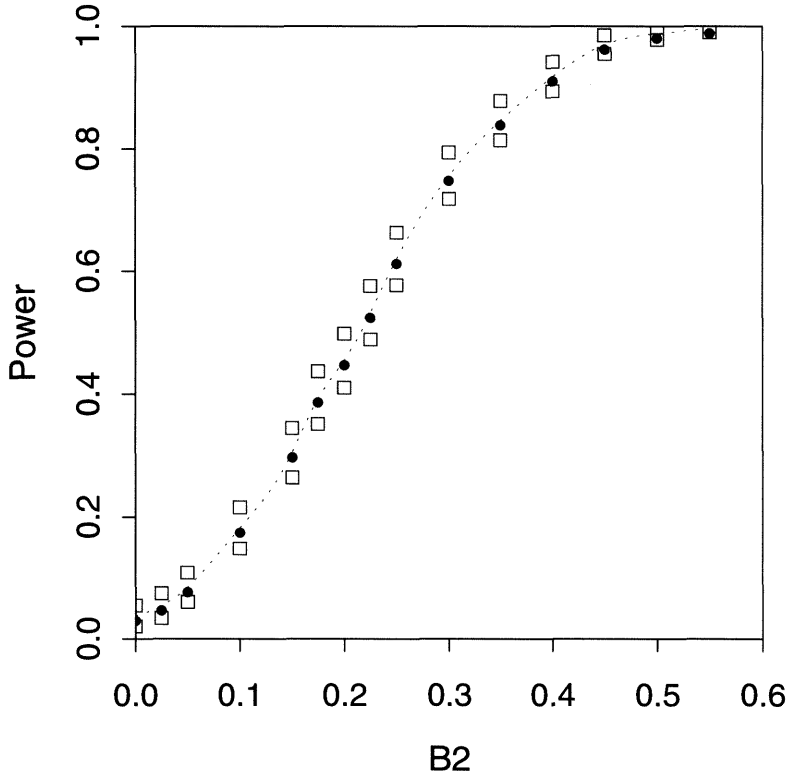

Figure 3.9: Power estimates for PPMLE (•) and corresponding 95% confidence limits (□) for chosen values of $B_{21} - B_{22} = B2$ from simulations with $\underline{\alpha}^T = (0, 0)$: power curve fitted to estimates using a cubic spline.

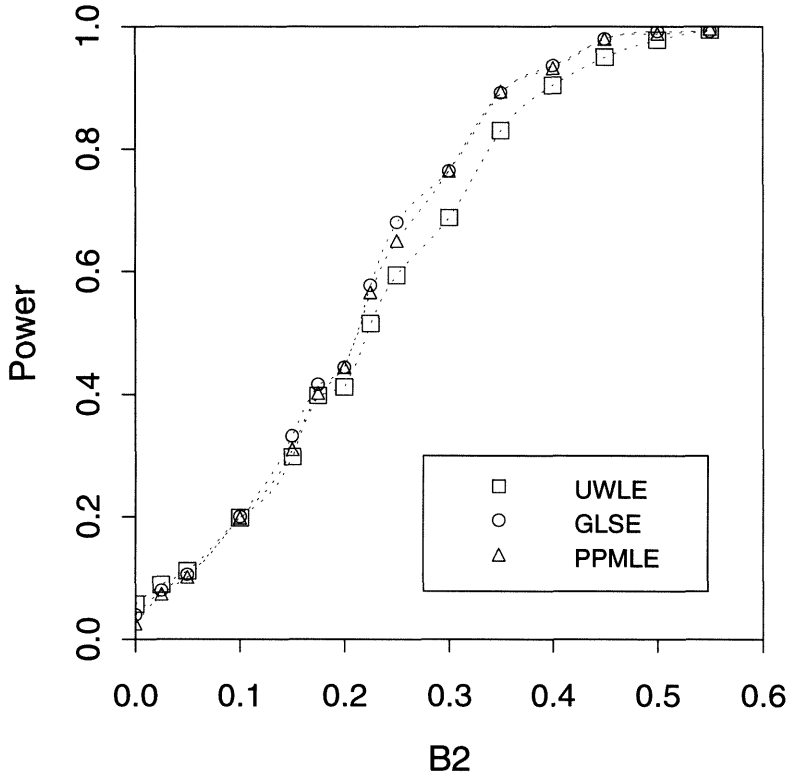

Figure 3.10: Power curve estimates for $B_{21} - B_{22} = B2$ for the UWLE, the GLSE and PPMLE from simulations with $\underline{\alpha}^T = (0.02, 0.5)$: power curve fitted to estimates using a cubic spline.

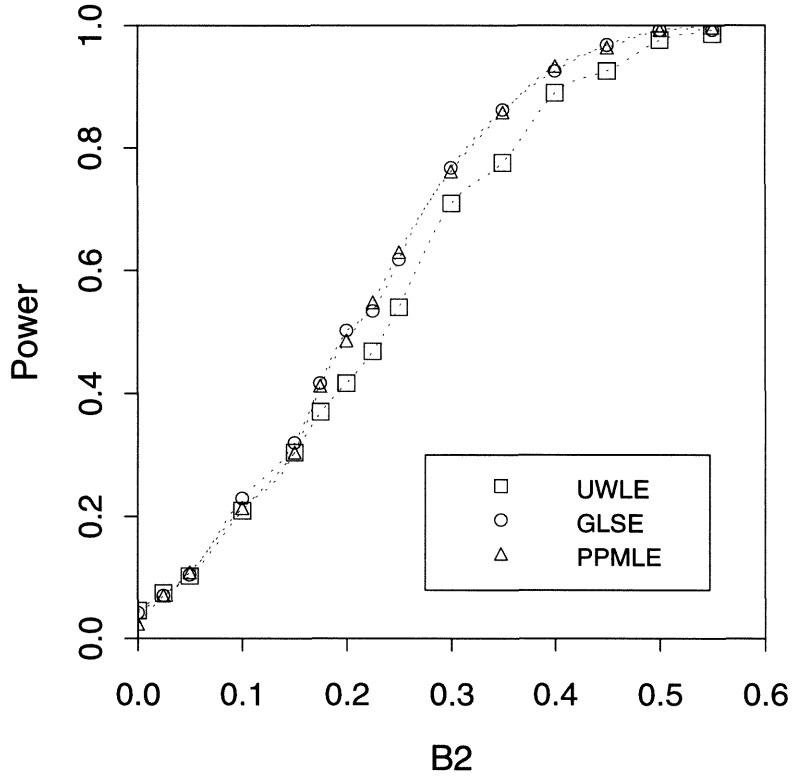

Figure 3.11: Power curve estimates for $B_{21} - B_{22} = B2$ for the UWLE, the GLSE and PPMLE from simulations with $\underline{\alpha}^T = (0.015, 0.75)$: power curve fitted to estimates using a cubic spline.

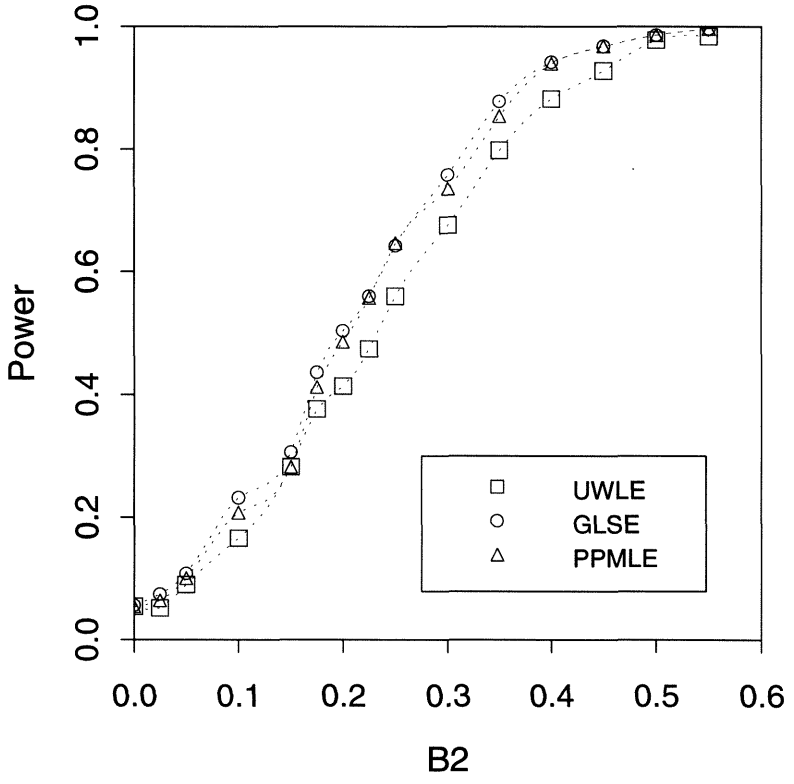

Figure 3.12: Power curve estimates for $B_{21} - B_{22} = B2$ for the UWLE, the GLSE and PPMLE from simulations with $\underline{\alpha}^T = (0.025, 0.25)$: power curve fitted to estimates using a cubic spline.

To test the sensitivity of the estimates of the population intercept and slope parameters to the validity of the probit assumption for the dropout process, the simulations were repeated using a Cox dropout process of form:

$$M\left(\underline{\boldsymbol{\alpha}}^T\underline{\boldsymbol{\beta}}_i,\, \alpha_{0j}\right) \;=\; \alpha_{0j}^{\exp\left(\underline{\boldsymbol{\alpha}}^T\underline{\boldsymbol{\beta}}_i\right)}$$

using $\underline{\boldsymbol{\alpha}}^T = (0.005,\, 0.125)$ and $(0.00375,\, 0.1875)$. The percentages dropping out in each of the follow-up periods, together with those surviving to the end of the study, are given in Table 3.6.

The probit dropout model was then fitted to the simulated data. Whilst the fit was a little worse than when the dropout process was truly probit, it was still generally acceptable. The relative sizes of the bias in the estimation of the group intercept and slope for the UWLE and PPMLE were both similar under the Cox dropout process when compared with those observed under the earlier probit dropout process (Tables 3.7 and 3.8 compared with Tables 3.4 and 3.5). This would appear to indicate that any worsening of the performance of PPMLE in the presence of a Cox rather than a probit dropout process will be slight. This observation is consistent with that made in Wu and Bailey (1988). They do, however, observe that the performance of PPMLE is adversely affected by the presence of truncated rather than shifted treatment effects.

The quasi-Newton-Raphson algorithm implemented via the NAG subroutine E04LAF showed various convergence problems during the maximisation process, despite the choice of starting values equal to the true parameter values. The problems were of a reasonably consistent form across the simulations. By suitable scaling of the log-likelihood function and parameters, the problems were limited to its returning the value 5 for the IFAIL parameter in, typically, 1-2% of the simulations. An IFAIL value of 5, which has a common interpretation in the various NAG quasi-Newton-Raphson minimisation procedures, indicates that there is

some doubt about whether the point found is a minimum because not all not all necesary conditions have been met. However it is deemed 'probable that the value returned gives a good estimate of the position of a minimum' (NAG, 1991). Moreover, on comparison of parameter estimates obtained when the IFAIL parameter was returned with the value 5 with those from the fully successful minimisations of the negative of the log-likelihood (with IFAIL returned as 0), there were no obvious differences. For these reasons, combined with a desire to retain comparability through consistent usage of random numbers across different simulations, those simulations for which IFAIL values of 5 had been obtained were retained.

$\underline{B}_k^T$	$\underline{\alpha}^T$ (0.005, 0.125)	$\underline{\alpha}^T$ (0.00375, 0.1875)
(75, 1.5)	18.14	18.14
	21.83	21.83
	23.48	23.48
	21.43	21.43
	15.12	15.12
(75, 1.0)	16.25	15.34
	21.42	21.19
	23.94	24.15
	22.38	22.85
	16.01	16.48
(100, 1.0)	20.12	18.14
	22.13	21.83
	22.97	23.48
	20.51	21.43
	14.27	15.12

Table 3.6: Expected percentages dropping out in follow-up periods 1, 2, 3 and 4 followed by those completing the study for different values of $\underline{B}_k$ under a Cox informative dropout mechanism.

	$\alpha_1 = 0.005 \quad \alpha_2 = 0.125$ (where $\alpha_{01} = 0$, $\alpha_{02} = 0.05$, $\alpha_{03} = 0.2$, $\alpha_{04} = 0.45$, $\alpha_{05} = 0.75$)				$\alpha_1 = 0.00375 \quad \alpha_2 = 0.1875$			
	Intercept	MSE	Slope	MSE	Intercept	MSE	Slope	MSE
$\underline{B}_1^T = (75, 1.5)$								
UWLE	75.010	0.8823	1.5055	10.032×10^{-3}	75.049	0.9506	1.4943	10.126×10^{-3}
GLSE	75.018	0.8840	1.4889	7.635×10^{-3}	75.061	0.9532	1.4704	8.052×10^{-3}
PPMLE	75.011	0.8832	1.5019	8.211×10^{-3}	75.049	0.9511	1.4905	8.100×10^{-3}
$\underline{B}_2^T = (75, 1.0)$								
UWLE	75.030	0.9481	0.9976	9.406×10^{-3}	74.972	0.9394	1.0013	10.096×10^{-3}
GLSE	75.038	0.9499	0.9817	7.651×10^{-3}	74.985	0.9391	0.9770	8.206×10^{-3}
PPMLE	75.032	0.9493	0.9929	7.941×10^{-3}	74.974	0.9394	0.9944	8.280×10^{-3}
$(\underline{B}_1 - \underline{B}_2)^T = (0, 0.5)$								
UWLE	-0.020	1.8954	0.5079	19.457×10^{-3}	0.077	1.9887	0.4930	20.512×10^{-3}
GLSE	-0.020	1.8954	0.5072	14.772×10^{-3}	0.076	1.9892	0.4934	14.893×10^{-3}
PPMLE	-0.021	1.8934	0.5090	14.898×10^{-3}	0.075	1.9890	0.4961	14.981×10^{-3}

Table 3.7: Results of simulations with different estimation procedures under a linear random effects model with Cox informative dropout with various dropout parameter values and with $\underline{B}_1^T = (75, 1.5)$ and $\underline{B}_2^T = (75, 1.0)$.

	$\alpha_1 = 0.005 \quad \alpha_2 = 0.125$ (where $\alpha_{01} = 0$, $\alpha_{02} = 0.05$, $\alpha_{03} = 0.2$, $\alpha_{04} = 0.45$, $\alpha_{05} = 0.75$)				$\alpha_1 = 0.00375 \quad \alpha_2 = 0.1875$			
	Intercept	MSE	Slope	MSE	Intercept	MSE	Slope	MSE
$\underline{B}_1^T = (75, 1.5)$								
UWLE	75.084	0.9075	1.4947	9.362×10^{-3}	74.968	0.9391	1.5040	10.794×10^{-3}
GLSE	75.092	0.9103	1.4790	7.374×10^{-3}	74.980	0.9391	1.4794	8.627×10^{-3}
PPMLE	75.085	0.9081	1.4916	7.729×10^{-3}	74.971	0.9381	1.4959	8.960×10^{-3}
$\underline{B}_2^T = (100, 1.0)$								
UWLE	99.930	0.9646	1.0085	11.760×10^{-3}	100.02	0.9711	1.0006	10.087×10^{-3}
GLSE	99.939	0.9648	0.9902	7.955×10^{-3}	100.03	0.9714	0.9752	7.781×10^{-3}
PPMLE	99.932	0.9646	1.0037	8.775×10^{-3}	100.02	0.9713	0.9918	8.463×10^{-3}
$(\underline{B}_1 - \underline{B}_2)^T = (-25, 0.5)$								
UWLE	-24.846	1.9752	0.4862	21.529×10^{-3}	-25.05	2.0512	0.5034	18.715×10^{-3}
GLSE	-24.847	1.9761	0.4888	14.922×10^{-3}	-25.05	2.0478	0.5042	14.009×10^{-3}
PPMLE	-24.847	1.9760	0.4879	15.153×10^{-3}	-25.05	2.0474	0.5041	13.943×10^{-3}

Table 3.8: Results of simulations with different estimation procedures under a linear random effects model with Cox informative dropout with various dropout parameter values and with $\underline{B}_1^T = (75, 1.5)$ and $\underline{B}_2^T = (100, 1.0)$.

3.3 Summary

The aim of this chapter was to gain insight into the workings of random coefficient models and into their potential for extension to situations where the 'dropout' process is of primary (rather than secondary) interest and the profile of repeated measurements of secondary importance, namely event history or survival analysis. The method introduced by Wu and Carroll (1988) was chosen as it was a selection model and thus jointly modelled the repeated measures profile and dropout process. In its present form, however, it only includes a single repeatedly measured variable and, despite this, it was both inherently complex to implement, even under the requirement that the dropout process be of probit form for an analystically tractable integration of the marginal likelihood, and subject to a certain degree of convergence problems, even when the starting values were chosen to coincide with the true parameter values.

The results of the simulations in Section 3.2 showed that the performance of PPMLE is comparable with the UWLE in terms of bias if the dropout process is probit and non-ignorable although PPMLE is considerably more efficient. The generalised least squares estimator (GLSE), however, is biased if the dropout process is informative (i.e. non-ignorable random-coefficient-based). However, if interest is solely in between-group differences, the bias in the GLSE for each group was found to be of very similar magnitude in all cases and led to similar levels of bias and similar MSEs to PPMLE. When the dropout process was of Cox (or proportional hazards) form, PPMLE appeared to be affected little in terms of bias compared with the UWLE which is not dropout model specific. With non-informative dropout all three methods were unbiased in their estimation of the growth curve parameters although both the GLSE and PPMLE were more efficient than the UWLE, with the GLSE tending to lead to a slightly lower mean square error than PPMLE.

As indicated by Little (1995), the use of random coefficient selection models such as that proposed by Wu and Carroll (1988), can be viewed as preferable to random coefficient pattern-mixture models as they are conceptually more in line with the expected interaction between the repeated measures process and the dropout or survival process. Selection models of this type can be viewed as somewhat 'stronger' models than the traditional survival models considered in the next Chapter whilst not fully exploiting biological knowledge in the manner of the stochastic diffusion models described in Chapter 2. The performance of random coefficient selection models in estimating the growth curve parameters can be matched or exceeded by pattern-mixture models, especially when the form of non-ignorability of the dropout process is known or correctly specified. If the form of dropout process is subject to uncertainty, sensitivity analysis is recommended (Little, 1995).

In considering the potential for extension of random coefficient selection models to event history analysis a number of problems were perceived. In general, and with the data collected for analysis in Chapter 5 in particular, there will be a moderate number of variables of various types repeatedly measured rather than the single continuous variable considered by Wu and Carroll (1988). Additionally, the interpretation of the α_l parameters in the selection model of Wu and Carroll (1988) is not in terms of risks or hazards. For these reasons, together with the problems in implementation of even this restrictive case of a random coefficient selection model, it was decided that the scope for extension to handle event history data was limited.

It was therefore decided to concentrate on extending existing survival models to cover event history data, paying particular attention to the specific problems inherent in screening studies. Research into and extensions to existing models will be performed in Chapter 4. Resulting methods will then be applied to recently-

collected data from a screening study of the health of an elderly population in Chapter 5.

Chapter 4

Analysing interval-censored survival data in the presence of a risk factor history

An alternative approach to analysing repeated measurements subject to dropouts applies survival analysis methodology in conjunction with appropriate use of the repeated measurements as covariates. This may involve formal modelling of the process generating the covariates as well as that of the survival process although, as explained in Chapter 2, any modelling should ideally consider these processes jointly. However, although 'strong models' such as the stochastic diffusion models described in Chapters 2 and 3 may be desirable in some contexts, they are highly complex and, moreover, do not consider problems caused by missing data. Therefore, the use of more traditional survival analysis methods with extensions to handle missing risk factor values will be relatively simple and yet may lead to inferences with acceptable properties. However, in such studies, it is common for the 'survival' status to be recorded only at the measurement or screening times, thus leading to incomplete information in the form of interval-censored survival times. There have been a number of methods developed over the past twenty-five years to allow for such incomplete information. Many of these methods have been

derived or adapted specifically to allow for grouped or interval-censored survival times, some of which allow for updating covariate information as it accumulates.

In Section 4.1, several of these methods are presented and then, in the remaining sections of this chapter, two of these methods are combined and the perfomance of the resulting method is assessed under various forms of missing covariate information using Monte Carlo simulation. In addition, the common approach of simply using baseline covariate values is compared with the use of updated covariate values which, in most situations, will be preferable.

4.1 Review of methods for analysing interval-censored survival data

One method suitable for modelling interval-censored survival data, providing it is of the **coincident interval** form (defined in Section 2.2), is that proposed by Prentice and Gloeckler (1978) in which an extension of the grouped-data version of the proportional hazards model (Kalbfleisch and Prentice, 1973) is proposed. As usual, it must be assumed that any right-censoring mechanism is independent of the mechanism causing the failures. This can be a problem with screening studies, as described in Section 2.1.2. The extension involves allowing the covariate vector to vary between intervals but to be fixed within intervals, that is, to be of step-functional form:

$$\underline{\boldsymbol{x}}_i(t) = \underline{\boldsymbol{x}}_i(t_j) \qquad t_j \leq t < t_{j+1}$$

This allows the accumulating risk factor history to be used in the analysis, although only the previous values of each covariate can be used in this manner as

the total number of covariates must be fixed.

Prentice and Gloeckler's likelihood function is:

$$L = \prod_{i=1}^{n} \left[\left\{ 1 - \alpha_k^{\exp\{\underline{\boldsymbol{x}}_i(t_k)^T \underline{\boldsymbol{\beta}}\}} \right\}^{\delta_i} \prod_{j=1}^{k-1} \alpha_j^{\exp\{\underline{\boldsymbol{x}}_i(t_j)^T \underline{\boldsymbol{\beta}}\}} \right] \qquad i = 1, \ldots, n$$

where the survival data on the i^{th} individual is recorded as a pair $(I_{ik},\ \delta_i)$ and $I_{ik} = (t_k,\ t_{k+1}]$ represents the k^{th} interval during which the i^{th} individual failed or was censored,

$$\delta_i = \begin{cases} 1 & \text{if the individual failed during } I_{ik}\,; \\ 0 & \text{if the individual was right-censored during } I_{ik}\,, \end{cases}$$

and $\alpha_j = \exp\left\{ - \int_{t_j}^{t_{j+1}} \lambda_0(u)\, du \right\} = \exp\left\{ - \int_{I_{ij}} \lambda_0(u)\, du \right\}$ is the conditional probability that an individual with $\underline{\boldsymbol{x}}_i(t_j) = \underline{\mathbf{0}}$ survives the interval $(t_j,\ t_{j+1}]$.

Prentice and Gloeckler (1978), although developing theory applicable where time-dependent covariates of this step-functional form exist, illustrated their method using a data set without any repeatedly measured covariates. Thompson (1981) described how a generalized linear model with binomial error structure and complementary-log-log link function can be used to fit Prentice and Gloeckler's model, albeit using a different parametrisation as shown later in this section. This method can be applied using GLIM upon construction of individual binary response units for each member of the risk set at the start of each interval and is covered further in Whitehead (1989) and Collett (1994).

Wu and Ware (1979) showed how the requirement that a regression model has a fixed number of covariates could be overcome when there is an accumulating risk factor history by suitable construction of an *augmented covariate vector*. This allows more flexibility in the use of risk factor measurements as covariates, removing

the restrictiveness of using the first or the last measurement. Although a logistic rather than a complementary-log-log model was used, this method of covariate vector construction can equally well be applied when Prentice and Gloeckler's method is used to fit a proportional hazards model. The augmented covariate vector has the relevant data in the appropriate position and zeros elsewhere for each individual in each period. Depending on the form of the parameter vector, some covariates in the augmented data vector may be defined as linear combinations of the values of the repeated measurements of a covariate from the individual histories. This method will allow the fitting of various models of a fully or semi-parametric form to see how the accumulating risk factor history can be used to best model survival.

Cupples, D'Agostino, Anderson and Kannel (1988) described how the methods proposed by Wu and Ware (1979) and by Prentice and Gloeckler (1978) are equivalent to an extension of their PRO (Pooling of Repeated Observations) method. In the standard PRO method, a proportional hazards model is used to model the risk of an event during a standard length observation period, pooling these observation periods in a way that an individual can contribute several 'independent' periods to the pooled information. This standard method also assumes that the underlying risk does not change over time, but an extension by including additional parameters to allow this risk to vary with time reverts to Prentice and Gloeckler's model or a particular parsimonious form of Wu and Ware's model, depending on the formulation.

Finkelstein (1986) derived maximum likelihood estimates of the parameters of a proportional hazards model when the data are not necessarily **coincident interval**. This method allows for censoring into overlapping and non-disjoint intervals but does not allow *directly* for time-dependent covariates as the survivor function is not factored as survival probabilites conditional upon survival to the start of each interval. Finkelstein's likelihood is expressed in terms of proba-

bilities of failure during the individual elements of an interval A_i which comprises a set of intervals, where each element of this set of the form $(t_j, t_{j+1}]$, $j = 1, \ldots, m+1$, where $t_1 = 0$ and $t_{m+1} = \infty$. This leads to a contribution from the i^{th} observation to the likelihood of:

$$\sum_{j=1}^{m} \delta_{ij} \left[S(t_j \mid \underline{\boldsymbol{x}}_i) - S(t_{j+1} \mid \underline{\boldsymbol{x}}_i) \right]$$

where $\delta_{ij} = 1$ if $(t_j, t_{j+1}]$ is a subset of A_i and 0 otherwise
and $S(t_j \mid \underline{\boldsymbol{x}}) = P(T > t \mid \underline{\boldsymbol{X}} = \underline{\boldsymbol{x}})$ is the survivor function for an individual with covariate vector $\underline{\boldsymbol{x}}$.

If proportionality of hazards is assumed, that is

$$\lambda(t) = \lambda_0(t) \exp(\underline{\boldsymbol{x}}_i{}^T \underline{\boldsymbol{\beta}})$$

then

$$\begin{aligned} S(t_j \mid \underline{\boldsymbol{x}}_i) - S(t_{j+1} \mid \underline{\boldsymbol{x}}_i) &= \exp(-H(t_j)) - \exp(-H(t_{j+1})) \\ &\quad \text{where } H(t) \text{ represents the} \\ &\quad \textit{integrated} \text{ hazard function} \\ &= \exp\left[- \int_0^{t_j} \lambda_0(u) \exp\left(\underline{\boldsymbol{x}}_i{}^T \underline{\boldsymbol{\beta}}\right) du \right] \\ &\quad - \exp\left[- \int_0^{t_{j+1}} \lambda_0(u) \exp\left(\underline{\boldsymbol{x}}_i{}^T \underline{\boldsymbol{\beta}}\right) du \right] \\ &= \exp\left[- \exp\left(\underline{\boldsymbol{x}}_i{}^T \underline{\boldsymbol{\beta}}\right) \int_0^{t_j} \lambda_0(u)\, du \right] \\ &\quad - \exp\left[- \exp\left(\underline{\boldsymbol{x}}_i{}^T \underline{\boldsymbol{\beta}}\right) \int_0^{t_{j+1}} \lambda_0(u)\, du \right] \end{aligned}$$

$$= \exp\left[-\exp\left(\underline{\boldsymbol{x}}_i{}^T\underline{\boldsymbol{\beta}} + \log\left[\int_0^{t_j} \lambda_0(u)\,du\right]\right)\right]$$
$$- \exp\left[-\exp\left(\underline{\boldsymbol{x}}_i^T\underline{\boldsymbol{\beta}} + \log\left[\int_0^{t_{j+1}} \lambda_0(u)\,du\right]\right)\right]$$

Then, letting $\zeta_j = \log\left[\int_0^{t_{j+1}} \lambda_0(u)\,du\right] \quad j = 1, \ldots, m-1$, Finkelstein obtains maximum-likelihood estimates for $\underline{\boldsymbol{\beta}}$ and $\underline{\boldsymbol{\zeta}}$.

It should be noted that the traditional right-censored observations are allowed for by the inclusion of $t_{m+1} = \infty$ and so $S(t_{m+1} \mid \underline{\boldsymbol{x}}) = 0, \quad \forall \underline{\boldsymbol{x}}$.

Assuming **coincident interval data** and no dependency of covariates, individuals with $\delta_{ik} = 1$ (and hence $\delta_{ij} = 0\,, \forall j \neq k$) simply contribute a term:

$$\exp\left[-\exp\left(\underline{\boldsymbol{x}}_i{}^T\underline{\boldsymbol{\beta}} + \log\left[\int_0^{t_k} \lambda_0(u)\,du\right]\right)\right]$$
$$-\exp\left[-\exp\left(\underline{\boldsymbol{x}}_i{}^T\underline{\boldsymbol{\beta}} + \log\left[\int_0^{t_{k+1}} \lambda_0(u)\,du\right]\right)\right]$$
$$= \exp\left[-\exp\left(\underline{\boldsymbol{x}}_i{}^T\underline{\boldsymbol{\beta}} + \zeta_{k-1}\right)\right] - \exp\left[-\exp\left(\underline{\boldsymbol{x}}_i{}^T\underline{\boldsymbol{\beta}} + \zeta_k\right)\right] \qquad (4.1)$$

to the likelihood, if failure is during the k^{th} interval. Similar contributions are made by individuals lost from the study due to right-censoring, except that the second term will then represent the zero survival probability at time infinity.

A contribution from such an individual could be parametrised somewhat differently by splitting the integrated baseline hazard function into separate components for each period, so equation 4.1 can be written:

$$\exp\left[-\exp(\underline{\boldsymbol{x}}_i{}^T\underline{\boldsymbol{\beta}})\int_0^{t_k}\lambda_0(u)\,du\right]$$
$$-\exp\left[-\exp(\underline{\boldsymbol{x}}_i{}^T\underline{\boldsymbol{\beta}})\left[\int_0^{t_k}\lambda_0(u)\,du+\int_{t_k}^{t_{k+1}}\lambda_0(u)\,du\right]\right]$$

$$=\exp\left[-\exp(\underline{\boldsymbol{x}}_i{}^T\underline{\boldsymbol{\beta}})\int_0^{t_k}\lambda_0(u)\,du\right]$$
$$\times\left[1-\exp\left[-\exp(\underline{\boldsymbol{x}}_i{}^T\underline{\boldsymbol{\beta}})\left\{\int_{t_k}^{t_{k+1}}\lambda_0(u)\,du\right\}\right]\right]$$

$$=\exp\left[-\exp(\underline{\boldsymbol{x}}_i{}^T\underline{\boldsymbol{\beta}})\sum_{j=1}^{k-1}\left(\int_{t_j}^{t_{j+1}}\lambda_0(u)\,du\right)\right]$$
$$\times\left[1-\exp\left[-\exp(\underline{\boldsymbol{x}}_i{}^T\underline{\boldsymbol{\beta}})\left\{\int_{t_k}^{t_{k+1}}\lambda_0(u)\,du\right\}\right]\right]$$

Now, letting $\gamma_j = \log\left(\int_{t_j}^{t_{j+1}}\lambda_0(t)\,dt\right) = \log\left(-\log\left(\alpha_j\right)\right)$ from Prentice and Gloeckler's original parametrisation, the likelihood contribution can be written:

$$\exp\left[-\exp(\underline{\boldsymbol{x}}_i^T\underline{\boldsymbol{\beta}})\sum_{j=1}^{k-1}\exp(\gamma_j)\right]\left[1-\exp\left(-\exp(\underline{\boldsymbol{x}}_i{}^T\underline{\boldsymbol{\beta}})\exp(\gamma_k)\right)\right]$$

$$=\exp\left[-\sum_{j=1}^{k-1}\exp\left(\underline{\boldsymbol{x}}_i{}^T\underline{\boldsymbol{\beta}}+\gamma_j\right)\right]\left[1-\exp\left(-\exp\left(\underline{\boldsymbol{x}}_i{}^T\underline{\boldsymbol{\beta}}+\gamma_k\right)\right)\right]$$

Hence, for coincident interval data with no time-dependency of covariates, Finkelstein's method is equivalent to Thompson's implementation of Prentice and Gloeckler's but with a different parametrisation for the interval-specific parameters ζ_j and $\gamma_j \quad j = 1, \ldots, m-1$, where:

$$\exp(\zeta_j) = \sum_{l=1}^{j}\exp(\gamma_l)$$

It is possible to modify Finkelstein's method to allow for time-dependent covariates of the type suggested by Prentice and Gloeckler. This will require the imputation of 'missing' covariate values if the measurement times vary between individuals. It will also be necessary to use Prentice and Gloeckler's parametrisation.

An alternative approach, introduced briefly in Chapter 2, was firstly proposed by Woodbury and Manton (1977) and applied in various papers from within their research group. They developed a model for the interaction of risk factor development and survival status. This incorporates two linked processes. The first describes the evolution of potential risk factors and is based on an autoregressive process. In their applications, this is a first-order process. The second process describes the form of the hazard function. This is specified as a quadratic function of an individual's present covariate values, where these may be estimated from the first process. A quadratic function is necessary to preserve multivariate normality of the full process (Woodbury and Manton, 1977) but also has the advantage of proposing a potential optimum level of each covariate, corresponding to the minimum of its particular quadratic, subject to its existence.

The general form of the likelihood function proposed may be factorised in the following way:

$$L = \prod_{i \in S_0} f_0(\underline{\boldsymbol{x}}_{io}) \prod_{t=0}^{\tau-1} \left[\prod_{i \in S_{t+1}} f_{t+1}\left(\underline{\boldsymbol{x}}_{i\,t+1} \mid \underline{\boldsymbol{x}}_{it}\right) \exp\left[-\mu_t\left(\underline{\boldsymbol{x}}_{it}\right)\right] \times \prod_{i \in \bar{S}_{t+1}} \left\{1 - \exp\left[\mu_t\left(\underline{\boldsymbol{x}}_{it}\right)\right]\right\} \right] \qquad t = 1, \ldots, \tau$$

where S_t denotes the set of individuals alive at time t and $\bar{S}_{t+1}$ denotes the set of individuals whose time of death is in the interval $(t,\, t+1]$; and where

$$f_t(\underline{\boldsymbol{x}}_{it}) \equiv f(\underline{\boldsymbol{x}}_i(t) \mid T_i > t)$$

is the conditional multivariate normal density function given survival to time t; and

$$f_{t+1}(\underline{\boldsymbol{x}}_{i\,t+1} \mid \underline{\boldsymbol{x}}_{it}) \equiv f(\underline{\boldsymbol{x}}_i(t+1) \mid \underline{\boldsymbol{x}}_i(t),\, T_i > t+1)$$

is the multivariate normal density function at time $t+1$, conditional on the covariate history. This expression may easily be factorised into three independent components, one involving $f_0(\underline{\boldsymbol{x}}_{io})$, one involving $f_{t+1}(\underline{\boldsymbol{x}}_{i\,t+1} \mid \underline{\boldsymbol{x}}_{it})$ and the third involving the hazard function $\mu_t(\underline{\boldsymbol{x}}_{it})$. Estimation of the corresponding parameters may then be performed separately for each component. For the first two components the estimation is using standard regression techniques, subject to appropriate assumptions. The parameters of the mortality component are then estimated via maximum likelihood. This third component may be specified in various ways, to allow for single or competing risks. It can also allow for exact death times as well as the more usual interval-censored data. This stochastic diffusion model for suvival has not been widely employed in epidemiology. Its sole users appear to have been the research team at Duke University who first proposed it. This limited level of adoption may be due to the general complexity of the modelling and the lack of availability of software together with the wide use of proportional hazard models for survival data.

4.2 Summary of methods and proposals

A general principle in statistical modelling is to use a model that is as simple as possible but which provides a satisfactory description of the processes relating the

variables involved. Considering the methods discussed in Section 4.1, whilst the stochastic diffusion models of Manton et al. may both provide an adequate summary of the data and take into account the likely relationship between underlying biological processes and mortality, they cannot be viewed as 'simple'. Therefore, if alternative strategies can satisfactorily explain the structure of data subject to interval-censoring and repeated risk factor assessment, they can be preferable, especially if they can handle certain forms of missing data processes without the necessity for highly complex derivations for standard errors of model parameter estimates. To this effect, it was decided to investigate further the methods of Prentice and Gloeckler and Finkelstein, paying particular attention to combining these two methods via the alternative parametrisation of Finkelstein's method which allows time dependent covariates of the step-functional form used by Prentice and Gloeckler to be incorporated into the modelling. This combination of methods will hereafter be referred to as *'adapted Finkelstein'*.

In the following sections, simulation is used to investigate the performance of Prentice and Gloeckler's method when the interval-censored survival times are of coincident-interval form and that of 'adapted Finkelstein' otherwise, in each case subjecting the repeated risk factor measurement process to various types of missing data processes.

Section 4.3 will compare two common situations in longitudinal studies using Prentice and Gloeckler's method. The first is where, rather than collecting repeated covariate measurements and using either the most recently recorded or, potentially, all measurements in an individual's history (Wu and Ware, 1979) as covariates, only the baseline covariates are included in the model. The other is where repeated covariate measurements are available at each follow-up point and treated as time-dependent covariates updated at each point. In this section there are repeated measurements of a single continuous variable and it is assumed there are no missing measurements. In Section 4.4, data sets with repeated covariate

measurements as created in Section 4.3 will then be subjected to various types of missingness in the repeated measurements. In this section, adaptated Finkelstein and Prentice and Gloeckler's method will each be used as, in some study designs, missing covariate measurements will coincide with missing information on the event of interest, thus rendering Prentice and Gloeckler's method unfeasible. A comparison will be made between the performance of this method under three different methods of imputation: imputation of the most recently recorded value; imputation of an interpolated value where possible - otherwise imputing the most recently recorded value; multiple imputation based on a simple linear regression model. Finally, in Section 4.5 similar comparisons will be made in the presence of repeated 'measurements' of one continuous and one ordinal variable, each subject to incomplete observation. In Section 4.5, however, the imputation of an interpolated value is not investigated.

4.3 A single repeatedly measured continuous variable without missing values

The aim of this section is to compare the performance of Prentice and Gloeckler's model in the presence of repeated measurements of a single continuous variable, using as covariate:

(i) a single covariate, measured at the baseline screen only;

(ii) a single covariate, measured at the last screen;

via simulations performed in GLIM. Eleven annual repeated measurements of diastolic blood pressure were simulated using the following autoregressive process of order one:

$$x_{ij} = 1.02\,(x_{ij-1} - 40) + 40 + \varepsilon_{ij} \qquad j = 1, \ldots 10$$

where $\varepsilon_{ij} \sim \mathcal{N}(0, \sigma_W^2)$ and $x_{i0} \sim \mathcal{N}(80, \sigma_B^2)$.

A two-parameter Weibull process was used for the baseline mortality process. This had a hazard function of the form:

$$\lambda_0(t) = \gamma\lambda t^{\gamma-1}$$

where γ is a *shape* parameter and λ is a *scale* parameter. The full mortality process then has hazard function of the form:

$$\lambda(t) = \lambda_0(t)\,\exp(\beta x_{ij})$$

where x_{ij} is the value of the i^{th} individual's diastolic blood pressure at their last screen, that is, at the start of the $j + 1^{\text{th}}$ year. It should be noted that the 'correct model' is therefore (ii) above and that the values of j are equal to the time in years (t) since entry to the study. The values of the parameters β, γ and σ_W^2/σ_B^2 were varied and those of the parameter λ was chosen to balance the effect of β on the initial death rates. A standard program was used in which $\beta = 0.07$, $\gamma = 1.5$ (baseline hazard increasing with the square root of time) and $\sigma_W^2/\sigma_B^2 = 0.25$ (with $\sigma_W = 4$ and $\sigma_B = 8$). These parameters gave an increasing hazard with time, with a relative risk of 2.01 for an increase in 10 mm Hg in diastolic blood pressure. The effect of the choice of baseline mortality process parameters was that an individual following the expected diastolic blood pressure profile had an interval-specific conditional probability of death increasing with time and which ranged from 0.0273 in the first interval to 0.2181 in the last.

For comparison of estimation properties, β took alternative values 0.035 (relative risk 1.42 for a 10 mm Hg increase), 0.14 (relative risk 4.06 for a 10 mm Hg increase) and 0.21 (relative risk 8.17 for a 10 mm Hg increase); γ took alternative values 1.0 (constant baseline hazard), 2.0 (baseline hazard increasing linearly with time) and 2.5 (hazard increasing with time$^{\frac{3}{2}}$); and σ_W^2/σ_B^2 took alternative values 0.0625, 1 and 4. Only a single parameter was altered for each program, except that λ was changed in accordance with β. One hundred simulations were performed for each set of parameter values. It should be noted that, although there was a single parameter γ in the baseline hazard function, separate estimates of a function of this parameter were made for each of the ten intervals between screens. The functional relationship between these parameters, the γ_js, and the Weibull shape parameter γ is:

$$\gamma_j = \log(\lambda) + \log(j^\gamma - (j-1)^\gamma) \qquad j = 1, \ldots, 10\,.$$

The results from the standard program may be found in Table 4.1 on page 130. There was little or no bias if the last measurement is used, as would be hoped. With the exception of the parameter γ_1, all the 95% confidence intervals constructed for the true bias include the value zero. If the baseline measurement is used as the covariate, the parameter β is underestimated and the baseline hazard function parameters, the γ_js, are resultantly over-estimated, this over-estimation increasing through the intervals. Similarly, the mean squared error (MSE) of the estimates is considerably lower when the last measurement is used as the covariate. On varying the ratio of within-patient variance to between-patient variance, the performance of the model using the last covariate measurement remained relatively consistent whereas that of the model using the first covariate measurement deteriorated in terms of both bias and mean square error as the variance ratio increases (Tables 4.2 - 4.4). When the Weibull shape parameter was varied, al-

though the direction of bias changed, there was little effect on the estimation. It did appear, however, that the performance was somewhat worse when $\gamma = 1.0$ (Table 4.5) than when it took values 1.5, 2.0 or 2.5 (Tables 4.1, 4.6 and 4.7). Finally, on varying the regression parameter β, the performance of the model using the last covariate measurement showed a slight deterioration as β increased (Tables 4.1 and 4.8 - 4.10). The performance of the model using the first covariate measurement showed a considerable worsening as β increased, to the extent that a relative bias in the range 30-35% was observed for all the parameters with $\beta = 0.21$.

In the vast majority of survival studies, only baseline measurements are ever made. These simulations support the hypothesis that the magnitude of the effects of risk factors will then tend to be under-estimated. This under- estimation may be considerable in the presence of strong covariate effects. If the hazard depends only upon the present value of an individual's risk factor, this should be included as a covariate. However, it may be that the hazard depends upon the covariate in some more complex way, for example its value some time in the past or the length of time it has exceeded some critical value. In such cases it will be vital to investigate and model the true nature of the type of effect of this risk factor, having collected as much of the risk factor history as is practical.

As noted by Altman and de Stavola (1994), there are two important differences between, using terminology introduced in this review, *time-fixed* and *updated covariate* models. The first is in the interpretation of the parameters. When using the time-fixed model, a parameter β_j represents the difference in relative risk on unit difference in a covariate z_j at time $t = 0$. However, if the updated covariates model is applied, a parameter β_j represents the difference in relative risk on unit increase in a covariate z_j at *any time*. The other difference is that the cumulative survival probability can be estimated directly from the estimated cumulative survival probability at time t, given knowledge of the covariate value at time $t = 0$,

for the time-fixed model. For the survival probability estimates for the updated covariate model, account must be taken of all values of z_j prior to time t, which will generally be unknown or, at best, be known only approximately or at certain time points.

Parameter	True Value	BASELINE Bias	BASELINE 95% C.I. for bias	LAST Bias	LAST 95% C.I. for bias
β	0.07	-2.938×10^{-3}	$(-3.976\times10^{-3}, -1.899\times10^{-3})$	0.242×10^{-3}	$(-0.485\times10^{-3}, 0.969\times10^{-3})$
γ_1	-9.2000	0.1496	(0.0587, 0.2405)	-0.1181	(-0.1898, -0.0465)
γ_2	-8.5965	0.3173	(0.2287, 0.4059)	-0.0521	(-0.1164, 0.0123)
γ_3	-8.3381	0.4327	(0.3435, 0.5219)	-0.0237	(-0.0915, 0.0442)
γ_4	-8.1690	0.5012	(0.4128, 0.5897)	-0.0319	(-0.0985, 0.0347)
γ_5	-8.0430	0.5822	(0.4943, 0.6700)	-0.0071	(-0.0752, 0.0609)
γ_6	-7.9425	0.6057	(0.5203, 0.6911)	-0.0201	(-0.0851, 0.0449)
γ_7	-7.8589	0.6141	(0.5272, 0.7009)	-0.0319	(-0.1004, 0.0366)
γ_8	-7.7873	0.6399	(0.5552, 0.7247)	-0.0126	(-0.0817, 0.0566)
γ_9	-7.7247	0.6264	(0.5395, 0.7133)	-0.0239	(-0.0936, 0.0459)
γ_{10}	-7.6690	0.5998	(0.5117, 0.6878)	-0.0332	(-0.1064, 0.0401)

Parameter	True Value	BASELINE MSE	BASELINE Standard Error of MSE	LAST MSE	LAST Standard Error of MSE
β	0.07	3.64×10^{-5}	4.38×10^{-6}	1.37×10^{-5}	2.13×10^{-6}
γ_1	-9.2000	0.2352	0.0287	0.1462	0.0194
γ_2	-8.5965	0.3029	0.0366	0.1093	0.0157
γ_3	-8.3381	0.3922	0.0422	0.1191	0.0199
γ_4	-8.1690	0.4530	0.0470	0.1153	0.0175
γ_5	-8.0430	0.5378	0.0539	0.1193	0.0191
γ_6	-7.9425	0.5548	0.0522	0.1093	0.0167
γ_7	-7.8589	0.5714	0.0575	0.1220	0.0198
γ_8	-7.7873	0.5946	0.0565	0.1233	0.0188
γ_9	-7.7247	0.5869	0.0582	0.1259	0.0180
γ_{10}	-7.6690	0.5596	0.0575	0.1392	0.0217

Table 4.1: Comparison of bias and MSE from simulations of a Weibull mortality process with shape parameter 1.5 and $\sigma_W^2/\sigma_B^2 = 0.25$ using Prentice and Gloeckler's method with baseline versus last covariate measurement with β=0.07.

Parameter	True Value	BASELINE Bias	BASELINE 95% C.I. for bias	LAST Bias	LAST 95% C.I. for bias
β	0.07	-1.805×10^{-3}	$(-3.842\times10^{-3}, 0.232\times10^{-3})$	0.559×10^{-3}	$(-0.250\times10^{-3}, 1.367\times10^{-3})$
γ_1	-9.2000	0.0778	(-0.0896, 0.2451)	-0.1133	(-0.1841, -0.0424)
γ_2	-8.5965	0.2067	(0.0408, 0.3726)	-0.0817	(-0.1556, -0.0078)
γ_3	-8.3381	0.3405	(0.1777, 0.5033)	-0.0375	(-0.1060, 0.0310)
γ_4	-8.1900	0.4233	(0.2605, 0.5860)	-0.0367	(-0.1064, 0.0330)
γ_5	-8.0430	0.4869	(0.3204, 0.6535)	-0.0368	(-0.1110, 0.0374)
γ_6	-7.9425	0.4956	(0.3316, 0.6597)	-0.0749	(-0.1467, -0.0031)
γ_7	-7.8589	0.5371	(0.3676, 0.7065)	-0.0600	(-0.1360, 0.0160)
γ_8	-7.7873	0.5604	(0.3977, 0.7231)	-0.0460	(-0.1218, 0.0298)
γ_9	-7.7247	0.5547	(0.3902, 0.7191)	-0.0447	(-0.1232, 0.0338)
γ_{10}	-7.6690	0.5136	(0.3449, 0.6823)	-0.0636	(-0.1421, 0.0148)

Parameter	True Value	BASELINE MSE	BASELINE Standard Error of MSE	LAST MSE	LAST Standard Error of MSE
β	0.07	1.10×10^{-4}	1.48×10^{-5}	0.17×10^{-4}	0.23×10^{-5}
γ_1	-9.2000	0.7277	0.0959	0.1420	0.0177
γ_2	-8.5965	0.7517	0.0997	0.1475	0.0184
γ_3	-8.3381	0.7988	0.1034	0.1224	0.0177
γ_4	-8.1690	0.8617	0.1041	0.1265	0.0167
γ_5	-8.0430	0.9518	0.1117	0.1433	0.0185
γ_6	-7.9425	0.9392	0.1088	0.1384	0.0187
γ_7	-7.8589	1.0282	0.1196	0.1525	0.0230
γ_8	-7.7873	1.0391	0.1208	0.1502	0.0199
γ_9	-7.7247	1.0043	0.1155	0.1610	0.0226
γ_{10}	-7.6690	0.9973	0.1219	0.1626	0.0204

Table 4.2: Comparison of bias and MSE from simulations of a Weibull mortality process with shape parameter 1.5 and $\sigma_W^2/\sigma_B^2 = 1.0$ using Prentice and Gloeckler's method with baseline versus last covariate measurement with β=0.07.

Parameter	True Value	BASELINE Bias	BASELINE 95% C.I. for bias	LAST Bias	LAST 95% C.I. for bias
β	0.07	-0.0175	(-0.0193, -0.0157)	6.02×10^{-4}	$(0.40\times10^{-4}, 11.65\times10^{-4})$
γ_1	-9.2000	1.3478	(1.2033, 1.4922)	-0.1157	(-0.1805, -0.0510)
γ_2	-8.5965	1.5896	(1.4439, 1.7352)	-0.0830	(-0.1373, -0.0287)
γ_3	-8.3381	1.7716	(1.6212, 1.9220)	-0.0638	(-0.1218, -0.0057)
γ_4	-8.1900	1.8775	(1.7310, 2.0240)	-0.0575	(-0.1124, -0.0027)
γ_5	-8.0430	1.8993	(1.7544, 2.0441)	-0.0619	(-0.1197, -0.0041)
γ_6	-7.9425	1.8782	(1.7306, 2.0257)	-0.0450	(-0.1049, 0.0149)
γ_7	-7.8589	1.7728	(1.6236, 1.9219)	-0.0698	(-0.1273, -0.0122)
γ_8	-7.7873	1.7095	(1.5612, 1.8578)	-0.0278	(-0.0831, 0.0274)
γ_9	-7.7247	1.5799	(1.4314, 1.7283)	-0.0405	(-0.1001, 0.0192)
γ_{10}	-7.6690	1.4489	(1.3010, 1.5968)	-0.0569	(-0.1182, 0.0044)

Parameter	True Value	BASELINE MSE	BASELINE Standard Error of MSE	LAST MSE	LAST Standard Error of MSE
β	0.07	3.902×10^{-4}	3.266×10^{-5}	0.085×10^{-4}	0.123×10^{-5}
γ_1	-9.2000	2.3538	0.2036	0.1214	0.0177
γ_2	-8.5965	3.0735	0.2309	0.0829	0.0116
γ_3	-8.3381	3.7218	0.2739	0.0908	0.0153
γ_4	-8.1690	4.0783	0.2794	0.0809	0.0120
γ_5	-8.0430	4.1478	0.2813	0.0899	0.0129
γ_6	-7.9425	4.0888	0.2827	0.0945	0.0131
γ_7	-7.8589	3.7159	0.2755	0.0902	0.0125
γ_8	-7.7873	3.4893	0.2553	0.0796	0.0110
γ_9	-7.7247	3.0641	0.2399	0.0932	0.0141
γ_{10}	-7.6690	2.6633	0.2203	0.1001	0.0133

Table 4.3: Comparison of bias and MSE from simulations of a Weibull mortality process with shape parameter 1.5 and $\sigma_W^2/\sigma_B^2 = 2.0$ using Prentice and Gloeckler's method with baseline versus last covariate measurement with β=0.07.

Parameter	True Value	BASELINE Bias	BASELINE 95% C.I. for bias	LAST Bias	LAST 95% C.I. for bias
β	0.07	3.864×10^{-3}	$(2.805\times10^{-3}, 4.923\times10^{-3})$	0.281×10^{-3}	$(-0.611\times10^{-3}, 1.172\times10^{-3})$
γ_1	-9.2000	-0.3682	(-0.4573, -0.2790)	-0.0651	(-0.1416, 0.0113)
γ_2	-8.5965	-0.2931	(-0.3846, -0.2016)	-0.0638	(-0.1422, 0.0147)
γ_3	-8.3381	-0.1835	(-0.2724, -0.0946)	-0.0248	(-0.1025, 0.0530)
γ_4	-8.1690	-0.1269	(-0.2149, -0.0389)	-0.0372	(-0.1151, 0.0406)
γ_5	-8.0430	-0.0323	(-0.1208, 0.0563)	-0.0059	(-0.0835, 0.0717)
γ_6	-7.9425	-0.0031	(-0.0904, 0.0843)	-0.0350	(-0.1146, 0.0445)
γ_7	-7.8589	0.0421	(-0.0480, 0.1323)	-0.0424	(-0.1242, 0.0395)
γ_8	-7.7873	0.0938	(0.0037, 0.1839)	-0.0373	(-0.1207, 0.0461)
γ_9	-7.7247	0.1562	(0.0615, 0.2509)	-0.0170	(-0.1051, 0.0711)
γ_{10}	-7.6690	0.1874	(0.0963, 0.2785)	-0.0217	(-0.1058, 0.0624)

Parameter	True Value	BASELINE MSE	BASELINE Standard Error of MSE	LAST MSE	LAST Standard Error of MSE
β	0.07	4.38×10^{-5}	5.77×10^{-6}	2.06×10^{-5}	2.72×10^{-6}
γ_1	-9.2000	0.3401	0.0430	0.1548	0.0196
γ_2	-8.5965	0.3017	0.0426	0.1626	0.0237
γ_3	-8.3381	0.2375	0.0338	0.1565	0.0210
γ_4	-8.1690	0.2155	0.0307	0.1576	0.0217
γ_5	-8.0430	0.2030	0.0270	0.1552	0.0198
γ_6	-7.9425	0.1966	0.0285	0.1643	0.0238
γ_7	-7.8589	0.2110	0.0287	0.1744	0.0234
γ_8	-7.7873	0.2180	0.0306	0.1806	0.0234
γ_9	-7.7247	0.2556	0.0368	0.2002	0.0299
γ_{10}	-7.6690	0.2491	0.0345	0.1829	0.0258

Table 4.4: Comparison of bias and MSE from simulations of a Weibull mortality process with shape parameter 1.5 and $\sigma_W^2/\sigma_B^2 = 0.0625$ using Prentice and Gloeckler's method with baseline versus last covariate measurement with β=0.07.

Parameter	True Value	BASELINE Bias	BASELINE 95% C.I. for bias	LAST Bias	LAST 95% C.I. for bias
β	0.07	2.407×10^{-3}	$(1.009\times10^{-3}, 3.804\times10^{-3})$	0.983×10^{-3}	$(0.004\times10^{-3}, 1.923\times10^{-3})$
γ_1	-9.2000	-0.2307	(-0.3551, -0.1064)	-0.1098	(-0.2026, -0.0171)
γ_2	-9.2000	-0.1094	(-0.2350, 0.0162)	-0.0909	(-0.1786, -0.0031)
γ_3	-9.2000	-0.0049	(-0.1251, 0.1153)	-0.0858	(-0.1696, -0.0020)
γ_4	-9.2000	-0.0309	(-0.0882, 0.1500)	-0.1467	(-0.2355, -0.0580)
γ_5	-9.2000	0.2052	(0.0820, 0.3283)	-0.0647	(-0.1555, 0.0262)
γ_6	-9.2000	0.2376	(0.1148, 0.3604)	-0.1163	(-0.2074, -0.0251)
γ_7	-9.2000	0.3227	(0.1988, 0.4466)	-0.1113	(-0.2055, -0.0171)
γ_8	-9.2000	0.3680	(0.2435, 0.4925)	-0.1396	(-0.2349, -0.0444)
γ_9	-9.2000	0.4624	(0.3462, 0.5786)	-0.1090	(-0.2043, -0.0136)
γ_{10}	-9.2000	0.5420	(0.4235, 0.6605)	-0.0836	(-0.1797, 0.0124)

Parameter	True Value	BASELINE MSE	BASELINE Standard Error of MSE	LAST MSE	LAST Standard Error of MSE
β	0.07	5.61×10^{-5}	8.65×10^{-6}	2.37×10^{-5}	3.37×10^{-6}
γ_1	-9.2000	0.4518	0.0647	0.2337	0.0303
γ_2	-9.2000	0.4186	0.0605	0.2066	0.0266
γ_3	-9.2000	0.3722	0.0524	0.1883	0.0259
γ_4	-9.2000	0.3667	0.0505	0.2245	0.0316
γ_5	-9.2000	0.4329	0.0536	0.2169	0.0308
γ_6	-9.2000	0.4450	0.0593	0.2276	0.0303
γ_7	-9.2000	0.4999	0.0664	0.2411	0.0347
γ_8	-9.2000	0.5348	0.0658	0.2533	0.0336
γ_9	-9.2000	0.5618	0.0704	0.2462	0.0364
γ_{10}	-9.2000	0.6555	0.0736	0.2447	0.0341

Table 4.5: Comparison of bias and MSE from simulations of a Weibull mortality process with shape parameter 1.0 and $\sigma_W^2/\sigma_B^2 = 0.25$ using Prentice and Gloeckler's method with baseline versus last covariate measurement with β=0.07.

Parameter	True Value	BASELINE Bias	BASELINE 95% C.I. for bias	LAST Bias	LAST 95% C.I. for bias
β	0.07	-5.814×10^{-3}	$(-6.751\times10^{-3}, -4.878\times10^{-3})$	0.683×10^{-3}	$(0.007\times10^{-3}, 1.358\times10^{-3})$
γ_1	-9.2000	0.4943	(0.4080, 0.5806)	-0.0527	(-0.1182, 0.0128)
γ_2	-8.1014	0.5619	(0.4812, 0.6426)	-0.0815	(-0.1434, -0.0196)
γ_3	-7.5906	0.6519	(0.5706, 0.7331)	-0.0627	(-0.1247, -0.0008)
γ_4	-7.2541	0.6910	(0.6124, 0.7696)	-0.0632	(-0.1245, -0.0018)
γ_5	-7.0028	0.6711	(0.5948, 0.7474)	-0.0851	(-0.1437, -0.0266)
γ_6	-6.8021	0.6574	(0.5814, 0.7335)	-0.0662	(-0.1257, -0.0067)
γ_7	-6.6351	0.6056	(0.5291, 0.6820)	-0.0591	(-0.1191, 0.0009)
γ_8	-6.4920	0.5200	(0.4386, 0.6013)	-0.0659	(-0.1262, -0.0056)
γ_9	-6.3668	0.4536	(0.3811, 0.5260)	-0.0387	(-0.0978, 0.0204)
γ_{10}	-6.2556	0.3410	(0.2615, 0.4204)	-0.0540	(-0.1140, 0.006)

Parameter	True Value	BASELINE MSE	BASELINE Standard Error of MSE	LAST MSE	LAST Standard Error of MSE
β	0.07	5.64×10^{-5}	6.75×10^{-6}	1.22×10^{-5}	1.63×10^{-6}
γ_1	-9.2000	0.4362	0.0532	0.1133	0.0169
γ_2	-8.1014	0.4834	0.0554	0.1054	0.0144
γ_3	-7.5906	0.5950	0.0612	0.1029	0.0136
γ_4	-7.2541	0.6367	0.0647	0.1010	0.0121
γ_5	-7.0028	0.6004	0.0581	0.0956	0.0119
γ_6	-6.8021	0.5813	0.0584	0.0957	0.0116
γ_7	-6.6351	0.5173	0.0550	0.0963	0.0122
γ_8	-6.4920	0.4408	0.0500	0.0981	0.0128
γ_9	-6.3668	0.3411	0.0359	0.0915	0.0116
γ_{10}	-6.2556	0.2788	0.0386	0.0957	0.0133

Table 4.6: Comparison of bias and MSE from simulations of a Weibull mortality process with shape parameter 2.0 and $\sigma_W^2/\sigma_B^2 = 0.25$ using Prentice and Gloeckler's method with baseline versus last covariate measurement with β=0.07.

Parameter	True Value	BASELINE Bias	BASELINE 95% C.I. for bias	LAST Bias	LAST 95% C.I. for bias
β	0.07	-5.619×10^{-3}	$(-6.590\times10^{-3}, -4.648\times10^{-3})$	0.612×10^{-3}	$(-0.156\times10^{-3}, 1.380\times10^{-3})$
γ_1	-9.2000	0.4628	(0.3681, 0.5574)	-0.0621	(-0.1448, 0.0205)
γ_2	-7.6617	0.5543	(0.4742, 0.6343)	-0.0625	(-0.1285, 0.0035)
γ_3	-6.9043	0.5979	(0.5163, 0.6795)	-0.0672	(-0.1348, 0.0005)
γ_4	-6.4020	0.6014	(0.5224, 0.6805)	-0.0509	(-0.1179, 0.0161)
γ_5	-6.0261	0.5263	(0.4525, 0.6000)	-0.0535	(-0.1144, 0.0073)
γ_6	-5.7256	0.4226	(0.3476, 0.4976)	-0.0383	(-0.0992, 0.0227)
γ_7	-5.4753	0.2659	(0.1859, 0.3458)	-0.0472	(-0.1154, 0.0210)
γ_8	-5.2608	0.0718	(-0.0100, 0.1536)	-0.0822	(-0.1495, -0.0149)
γ_9	-5.0732	-0.0736	(-0.1622, 0.0149)	-0.0545	(-0.1303, 0.0214)
γ_{10}	-4.9064	-0.2408	(-0.3449, -0.1368)	-0.0396	(-0.1346, 0.0554)

Parameter	True Value	BASELINE MSE	BASELINE Standard Error of MSE	LAST MSE	LAST Standard Error of MSE
β	0.07	5.59×10^{-5}	6.20×10^{-6}	1.56×10^{-5}	2.30×10^{-6}
γ_1	-9.2000	0.4449	0.0523	0.1798	0.0262
γ_2	-7.6617	0.4724	0.0476	0.1162	0.0178
γ_3	-6.9043	0.5290	0.0536	0.1225	0.0191
γ_4	-6.4020	0.5229	0.0506	0.1183	0.0177
γ_5	-6.0261	0.4172	0.0419	0.0983	0.0149
γ_6	-5.7556	0.3237	0.0364	0.0972	0.0131
γ_7	-5.4753	0.2354	0.0329	0.1220	0.0168
γ_8	-5.2608	0.1775	0.0244	0.1234	0.0143
γ_9	-5.0732	0.2076	0.0290	0.1512	0.0201
γ_{10}	-4.9064	0.3372	0.0458	0.2343	0.0366

Table 4.7: Comparison of bias and MSE from simulations of a Weibull mortality process with shape parameter 2.5 and $\sigma_W^2/\sigma_B^2 = 0.25$ using Prentice and Gloeckler's method with baseline versus last covariate measurement with β=0.07.

Parameter	True Value	BASELINE Bias	BASELINE 95% C.I. for bias	LAST Bias	LAST 95% C.I. for bias
β	0.035	0.0026	(0.0016, 0.0035)	-0.0001	(-0.0007, 0.0005)
γ_1	-6.4000	-0.2652	(-0.3593, -0.1712)	-0.0492	(-0.1150, 0.0166)
γ_2	-5.7965	-0.2000	(-0.2789, -0.1211)	-0.0243	(-0.0751, 0.0265)
γ_3	-5.5381	-0.1199	(-0.2047, -0.0351)	0.0169	(-0.0381, 0.0718)
γ_4	-5.3690	-0.0849	(-0.1661, -0.0036)	0.0145	(-0.0405, 0.0694)
γ_5	-5.2430	-0.0562	(-0.1393, 0.0269)	0.0093	(-0.0484, 0.0670)
γ_6	-5.1425	-0.0462	(-0.1267, 0.0343)	-0.0114	(-0.0694, 0.0465)
γ_7	-5.0589	-0.0181	(-0.1046, 0.0684)	-0.0120	(-0.0740, 0.0500)
γ_8	-4.9873	0.0080	(-0.0732, 0.0893)	-0.0094	(-0.0682, 0.0493)
γ_9	-4.9247	0.0355	(-0.0472, 0.1182)	-0.0002	(-0.0582, 0.0578)
γ_{10}	-4.8690	0.0522	(-0.0243, 0.1288)	0.0044	(-0.0491, 0.0580)

Parameter	True Value	BASELINE MSE	BASELINE Standard Error of MSE	LAST MSE	LAST Standard Error of MSE
β	0.035	3.07×10^{-5}	4.05×10^{-6}	0.93×10^{-5}	1.26×10^{-6}
γ_1	-6.4000	0.2983	0.0405	0.1140	0.0176
γ_2	-5.7965	0.2005	0.0276	0.0671	0.0099
γ_3	-5.5381	0.1999	0.0245	0.0780	0.0086
γ_4	-5.3690	0.1773	0.0249	0.0779	0.0119
γ_5	-5.2430	0.1810	0.0243	0.0858	0.0106
γ_6	-5.1425	0.1691	0.0222	0.0867	0.0119
γ_7	-5.0589	0.1932	0.0295	0.0993	0.0129
γ_8	-4.9873	0.1702	0.0247	0.0891	0.0127
γ_9	-4.9247	0.1776	0.0253	0.0867	0.0125
γ_{10}	-4.8690	0.1537	0.0228	0.0739	0.0101

Table 4.8: Comparison of bias and MSE from simulations of a Weibull mortality process with shape parameter 1.5 and $\sigma_W^2/\sigma_B^2 = 0.25$ using Prentice and Gloeckler's method with baseline versus last covariate measurement with β=0.035.

Parameter	True Value	BASELINE Bias	BASELINE 95% C.I. for bias	LAST Bias	LAST 95% C.I. for bias
β	0.14	-0.0314	(-0.0324, -0.0303)	0.0018	(0.0008, 0.0028)
γ_1	-14.8000	2.6861	(2.5930, 2.7793)	-0.2131	(-0.3029, -0.1233)
γ_2	-14.1965	2.9286	(2.8381, 3.0190)	-0.1751	(-0.2630, -0.0872)
γ_3	-13.9381	3.0461	(2.9517, 3.1405)	-0.1744	(-0.2641, -0.0848)
γ_4	-13.7690	3.1124	(3.0241, 3.2007)	-0.1652	(-0.2543, -0.0761)
γ_5	-13.6430	3.1283	(3.0444, 3.2122)	-0.1632	(-0.2500, -0.0765)
γ_6	-13.5425	3.1019	(3.0115, 3.1922)	-0.1652	(-0.2502, -0.0803)
γ_7	-13.4589	3.0403	(2.9516, 3.1291)	-0.1794	(-0.2695, -0.0893)
γ_8	-13.3873	2.9974	(2.9061, 3.0887)	-0.1748	(-0.2647, -0.0849)
γ_9	-13.3246	2.9527	(2.8643, 3.0410)	-0.1574	(-0.2450, -0.0699)
γ_{10}	-13.2690	2.8641	(2.7757, 2.9526)	-0.1859	(-0.2734, -0.0984)

Parameter	True Value	BASELINE MSE	BASELINE Standard Error of MSE	LAST MSE	LAST Standard Error of MSE
β	0.14	1.013×10^{-3}	3.452×10^{-5}	0.027×10^{-3}	0.330×10^{-5}
γ_1	-14.8000	7.4389	0.2631	0.2533	0.0308
γ_2	-14.1965	8.7872	0.2732	0.2297	0.0291
γ_3	-13.9381	9.5082	0.2955	0.2374	0.0287
γ_4	-13.7690	9.8878	0.2833	0.2320	0.0289
γ_5	-13.6430	9.9676	0.2700	0.2207	0.0281
γ_6	-13.5425	9.8319	0.2890	0.2133	0.0249
γ_7	-13.4589	9.4465	0.2767	0.2413	0.0285
γ_8	-13.3873	9.1991	0.2860	0.2389	0.0282
γ_9	-13.3247	8.9194	0.2713	0.2223	0.0286
γ_{10}	-13.2690	8.4050	0.2567	0.2319	0.0314

Table 4.9: Comparison of bias and MSE from simulations of a Weibull mortality process with shape parameter 1.5 and $\sigma_W^2/\sigma_B^2 = 0.25$ using Prentice and Gloeckler's method with baseline versus last covariate measurement with β=0.14.

Parameter	True Value	BASELINE Bias	BASELINE 95% C.I. for bias	LAST Bias	LAST 95% C.I. for bias
β	0.21	-0.0731	(-0.0744, -0.0718)	0.0011	(-0.0002, 0.0025)
γ_1	-20.4000	6.4660	(6.3530, 6.5790)	-0.1071	(-0.2361, 0.0219)
γ_2	-19.7965	6.6251	(6.5155, 6.7347)	-0.1206	(-0.2456, 0.0044)
γ_3	-19.5381	6.6928	(6.5897, 6.7959)	-0.0924	(-0.2135, 0.0286)
γ_4	-19.3690	6.6670	(6.5599, 6.7740)	-0.1013	(-0.2209, 0.0183)
γ_5	-19.2430	6.6050	(6.4993, 6.7108)	-0.1159	(-0.2419, 0.0101)
γ_6	-19.1425	6.5502	(6.4440, 6.6564)	-0.0934	(-0.2160, 0.0293)
γ_7	-19.0589	6.4523	(6.3440, 6.5606)	-0.1100	(-0.2289, 0.0089)
γ_8	-18.9873	6.3801	(6.2722, 6.4881)	-0.1007	(-0.2283, 0.0269)
γ_9	-18.9247	6.2811	(6.1739, 6.3882)	-0.1026	(-0.2240, 0.0188)
γ_{10}	-18.8690	6.2055	(6.1007, 6.3103)	-0.0961	(-0.2212, 0.0290)

Parameter	True Value	BASELINE MSE	BASELINE Standard Error of MSE	LAST MSE	LAST Standard Error of MSE
β	0.21	5.384×10^{-3}	9.514×10^{-5}	0.050×10^{-3}	0.698×10^{-5}
γ_1	-20.4000	42.1379	0.7358	0.4403	0.0617
γ_2	-19.7965	44.2014	0.7296	0.4175	0.0595
γ_3	-19.5381	45.0679	0.6907	0.3861	0.0528
γ_4	-19.3690	44.7436	0.7123	0.3790	0.0500
γ_5	-19.2430	43.9147	0.7061	0.4225	0.0607
γ_6	-19.1425	43.1955	0.6980	0.3964	0.0581
γ_7	-19.0589	41.9339	0.7003	0.3764	0.0499
γ_8	-18.9873	41.0067	0.6887	0.4296	0.0613
γ_9	-18.9247	39.7479	0.6814	0.3904	0.0550
γ_{10}	-18.8690	38.7909	0.6478	0.4124	0.0596

Table 4.10: Comparison of bias and MSE from simulations of a Weibull mortality process with shape parameter 1.5 and $\sigma_W^2/\sigma_B^2 = 0.25$ using Prentice and Gloeckler's method with baseline versus last covariate measurement with β=0.21.

4.4 Imputation for incomplete repeated measurements of a continuous risk factor

The form of data set with a single continuous repeatedly measured variable introduced in Section 4.3 is again used in this section, but here is subjected to various types of missing-data mechanisms. A Weibull model is again used in the simulation of the survival process. The relevant parameters are $\gamma = 1.15$, $\log \lambda = -9.2$, $\beta = 0.07$, $\sigma_W = 1.5$ and $\sigma_B = 8.0$. The effects on the performance of the imputation methods on increasing the within patient variability, σ_W, will be investigated.

Observations are deleted using various missing data mechanisms, with nominal percentages of observations missing of 5%, 10%, 20% and 50%. The three methods used for imputation are:

a) imputation of last recorded measurement;

b) linear interpolation between recorded measurements combined with extrapolation from last pair of recorded measurements if no further measurements were available and imputation of last recorded measurement if only a single measurement was recorded;

c) multiple imputation using a ordinary linear regression model for $\underline{Y}_m$.

The ideas of multiple imputation can be applied to repeated measurements of the form discussed in this chapter in the following way. Firstly, it is assumed that the parameters of the survival model $\underline{\gamma}$ are distinct from those of the repeated measures model $\underline{\beta}$. Then 'adapted Finkelstein' can be repeatedly fitted to the imputed data sets, to average the distribution of interval-censored survival times over the predictive distribution of the missing data, $f(x_{m0}, \ldots, x_{m\,10} \mid x_{o0}, \ldots, x_{o\,10})$. It should be noted that the imputation of $x_{m\,10}$ is optional as this observation is

not used in 'adaptated Finkelstein' because there is no further follow-up. Also, if x_{o0} is defined as the observation at the first *attended* screen, then there are no data missing at the 'first' screen, although this approach will require both the screening process to be non-informative and the absence of period effects.

Suppressing the indexing for the i^{th} individual, the predictive distribution may be written:

$$f(x_{m1}, \ldots, x_{m\,10} \mid \underline{\boldsymbol{x}}_o) = \left[\int f(x_{m1} \mid \underline{\boldsymbol{\beta}}, \underline{\boldsymbol{x}}_o)\, f(\underline{\boldsymbol{\beta}} \mid \underline{\boldsymbol{x}}_o)\, d\underline{\boldsymbol{\beta}}\right] \times$$
$$\left[\prod_{j=1}^{9} \int f(x_{m\,11-j} \mid \underline{\boldsymbol{\beta}}, x_{m1}, \ldots, x_{m\,10-j}, \underline{\boldsymbol{x}}_o)\, f(\underline{\boldsymbol{\beta}} \mid x_{m1}, \ldots, x_{m\,10-j}, \underline{\boldsymbol{x}}_o)\, d\underline{\boldsymbol{\beta}}\right]$$

where $\underline{\boldsymbol{x}}_o^T = (x_{o0}, \ldots, x_{o\,10})$.

The model for x_{imj} is,

$$x_{imj} = \beta_0 + \beta_1 x_{i\bullet\, j-1} + \varepsilon_{ij} \qquad j = 1, \ldots, 10$$

where $\varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$ and ε_{ij} is independent of ε_{kl}, $\forall\ (i, j) \neq (k, l)$, and $x_{i\bullet\, j-1}$ is $x_{io\,j-1}$ or $x_{im\,j-1}$, as appropriate.

Although in the simulations that follow diagnostics will not be performed, it should be noted that, unless the data used in the model has a **monotone pattern**, diagnostics will be difficult to perform and in some cases will be inappropriate. Moreover, if the non-response mechanism is non-ignorable, the estimation of $f(\underline{\boldsymbol{\beta}} \mid \underline{\boldsymbol{x}}_o)$ is likely to be biased. Hence, in estimating $f(\underline{\boldsymbol{\beta}} \mid \underline{\boldsymbol{x}}_o)$ or $f(\underline{\boldsymbol{\beta}} \mid x_{m1}, \ldots, x_{m\,10-j}, \underline{\boldsymbol{x}}_o)$, only complete data pairs $(x_{i\bullet j}, x_{i\bullet\, j-1})$ will be used. Moreover, rather than using $f(\underline{\boldsymbol{\beta}} \mid x_{m1}, \ldots, x_{m\,10-j}, \underline{\boldsymbol{x}}_o)$ in this scheme, $f(\underline{\boldsymbol{\beta}} \mid \underline{\boldsymbol{x}}_o)$ is used in its place. An improvement to this scheme might be achieved

by using the multiply imputed values of earlier measurement times, that is using estimates of $f(\underline{\beta} \mid x_{m1}, \ldots, x_{m\,10-j}, \underline{x}_o)$ for each j and, moreover, using iteration as per the data augmentation algorithm to improve the imputations, in line with the suggestion made by Rubin (1987).

As the performance of multiple imputation is known to be adequate even for moderate multiples and the simulation programs took up considerable amounts of CPU time, it was decided to use five imputations for each missing covariate value. This led to the construction of five separate completed data sets for each of the simulations performed in each program run.

As described in Chapter 2, there are, additionally, two possible forms for the information relating to the interval containing the event of interest. If a missing value in the covariate has no effect on the recording of the information relating to the interval there is *screen-independent event collection* and may occur for events such as hospitalisation, removal into institutional care or death. Alternatively, if change of state is only to be recorded at the next *attended* screen there is *screen-dependent event collection.* An example of this is in assessing the onset of dementia, disability or general disease.

Screen-independent event collection will result in coincident interval data and hence Thompson's GLIM implementation of Prentice and Gloeckler's method could be used. With screen-dependent event collection, however, Prentice and Gloeckler's method is not applicable and therefore 'adapted Finkelstein' is required. As it was desirable to use both the same parametrisation and the same basic program for all simulations, a purpose-written double-precision Fortran 77 program was used to carry out all the simulations.

Five hundred simulations were performed for each combination of missing-data percentage, imputation method and screen independent/dependent event collec-

tion. In all cases, a small percentage of simulations, typically 0.5-2% led to convergence problems in the NAG quasi-Newton-Raphson algorithm E04JAF used in the maximisation process. In all such cases, the IFAIL parameter returned by the subroutine took the value 5, meaning that, although not all conditions for a maximum had been met, the parameter estimates returned should be good approximations to the maximum likelihood estimates.

4.4.1 Data MCAR

Initially, a MCAR mechanism was used to create the missing data. For this form of missing data mechanism, the nominal missing data percentage was equal to the expected missing data percentage, any discrepancies observed in the data sets being purely due to random variation.

The results from these simulations are summarised in Tables 4.13 and 4.14 (pages 155 and 156). More detailed tables which include the perfomance in estimating the nuisance parameters γ_1-γ_{10} are included in Appendix B as Tables B.1-B.13 and in which the mean observed bias, a 95% confidence interval for the bias, the estimated mean square error (MSE) and its estimated standard error are given for each of the eleven parameters estimated.

Table B.1 shows that, in the absence of missing covariate data, there is no significant bias in the estimation of either the parameter of interest (β) or the nuisance parameters (γ_js).

Imputation of last measurement

The main results from these simulations are summarised in Tables 4.13 and 4.14. More detailed exposition, including reference to the nuisance parameters, may be found in Appendix B as Tables B.2-B.5. These tables show that, if event collection is screen-independent and is always non-missing, imputation of the last measurement performs well, both in terms of bias and mean square error, even in the presence of moderate amounts of missing-data (20%). With 50% missing, its performance is reasonably good, with no significant bias in the estimation of β and little bias in the estimation of the γ_j parameters. The estimated mean square error for $\hat{\beta}$ increased from 3.149×10^{-5} with no missing data to 3.193×10^{-5} with 50% missing, an increase of 1.38%. The corresponding estimated mean square error increased by a lesser percentage for all ten $\hat{\gamma}_j$s.

However, if event collection is screen-dependent, the performance of this simple imputation worsens steadily as the proportion of missing data increases. The percentage bias observed in the estimation of the parameter β is of the order of an eighth of the percentage of missing data, ranging from a non-significant 0.67% with 5% missing to a very highly significant 6.48% with 50% missing. The estimated mean square error and its estimated standard error also increased steadily with the percentage of data missing. The increase in estimated mean square error for $\hat{\beta}$ relative to that when there were no data missing ranged from 0.14% with 5% missing to 73.42% when there were 50% missing, with a relative increase in its standard error of similar magnitude. The effect of increased amounts of missing data was not quite as detrimental on the estimation of the γ_js, with the relative increase in estimated mean square error ranging from 7.01% to 43.36%, the greatest increases being observed during the middle periods.

The observed bias in the estimation of β was positive in seven of the eight combinations of percentages of missing data and forms of event collection. The

exception was with 50% missing when the event collection was screen-dependent, although the bias observed here was not significantly negative. The estimator of β will be negatively correlated with each of the estimators for the γ_js because, for a given amount of information on mortality, an increase in the estimate of the contribution of an individual's diastolic blood pressure will result in a decrease in the estimate of the corresponding contribution from the baseline mortality process. This resulted in consistently negative bias being observed in the estimates of γ_1 to γ_{10} when there were 5%, 10% and 20% data missing, for both forms of event collection. When there were 50% data missing, the expected pattern was generally observed, with positive estimates of the bias in $\hat{\gamma}_j$ with the negative estimate of the bias in $\hat{\beta}$ when there was screen-independent event collection and negative estimates of the bias in the estimation of γ_j with the positive estimate of that in the estimation of β when there was screen-dependent event collection. The only departure from this expected pattern was in the estimation of γ_1 when, in both situations, the estimate was biased in the opposite direction to that expected, albeit only by a small amount. This bias was significant in the case of screen-dependent event collection but non-significant for screen-independent event collection. It is worth remarking here that γ_1 is affected in a different way to the other nuisance parameters as all individuals have a first screen. The effect of screen-dependent event collection is to transfer the possible intervals during which a death occurs back through the time periods, thus increasing the apparent risk of death in the first period and slightly decreasing the apparent risk of death in the later periods, thus causing the effect observed.

Imputation by interpolation

The main results from these simulations are summarised in Tables 4.13 and 4.14. Tables B.6-B.9 may be found in Appendix B and also include the results for the γ_j parameters.

Overall, the performance of imputation by interpolation was better than that when imputing the last observed value, notably so when there was screen-dependent event collection. However, the simulations indicate that the performance of imputation by interpolation is slightly worse than last measurement imputation when the data were 50% missing in conjunction with screen-independent event collection, both in terms of bias and mean square error. This can be attributed to a far greater degree of bias (4.36%) observed using the interpolation method, than that observed under last measurement imputation (0.18%). This may well simply be a chance observation, especially as the observed bias was lower with screen-independent event collection than for screen-dependent event collection. When the event collection was screen-independent, the estimated mean square error was of comparable magnitude to that for last measurement imputation but when there was screen-dependent event collection, the use of interpolated values produced estimates with considerably better properties than the use of the last observed measurement. In fact, the estimated mean square errors for $\hat{\beta}$ were of similar magnitude for the two forms of event collection, although those for the $\hat{\gamma}_j$s showed some relative deterioration with increasing amounts of missing data.

Multiple Imputation

The results from these simulations are summarised in Tables 4.13 and 4.14, with details for the nuisance parameters included in Tables B.10-B.13. When multiple imputation was compared with imputation by linear interpolation, the performance was similarly good with regards to bias, with possibly a slightly lesser deterioration in performance as the amount of missing data increased. This slightly better performance was consistent across the estimation of β and the γ_js under both screen-independent and screen-dependent event collection. Likewise, in terms of the mean square error, the deterioration of multiple imputation was

less marked than linear interpolation imputation as the percentage of missing data increased.

One unusual effect of using multiple imputation was to lead to a lower mean square error than with complete data, except when 50% of the covariate data were missing. Moreover, the mean square error was lower when the event collection was screen-dependent than when it was screen-independent. This effect was observed, although to a far lesser degree, when imputation was by linear interpolation. It is believed that this is caused by the increased variability leading to reduction of the MSE in a similar way to that observed later in this section on 'increasing the within-patient variability'.

A comparison of the imputation methods

Whilst imputation of the last observed covariate value had good estimation properties when there was little missing data, its performance deteriorated as the amount of missing data increased, especially when the event collection was screen-dependent. Linear interpolation imputation (or imputation of the last measurement where this was not possible) performed better overall, the relative improvement being more marked as the degree of incompleteness increased, including that relating to the event of interest. Multiple imputation performed marginally better than linear interpolation imputation across the various missing data percentages and event collection strategies, although not sufficiently so to merit its use on these grounds alone. As described earlier, the real advantage of multiple imputation is that it will allow standard error estimates to be constructed from the information matrices obtained from the multiple data sets by including adjustments estimated from their between imputation variability. This was not pursued at this stage as the standard errors could be estimated from the variability in parameter estimates between simulations.

The main disadvantage of multiple imputation is the increase in processing time, the relative increase being approximately equal to the number of imputations made. With 5% missing data, these simulations, this led to an increase from a little over 30 minutes CPU time for both last measurement imputation and linear interpolation imputation, to around $2\frac{1}{2}$ hours for the regression-based multiple imputation. These processing times were around 20% higher when there were 50% of the data missing. The machine specification was:

12 150 MHz IP19 Processors
CPU: MIPS R4400 Processor Chip Revision: 5.0
FPU: MIPS R4010 Floating Point Chip Revision: 0.0
Data cache size: 16 Kbytes
Instruction cache size: 16 Kbytes
Secondary unified instruction size: 1 Mbyte
Main memory size: 512 Mbytes, 2-way interleaved.

Increasing the within-patient variability

The estimation using the three imputation methods was then repeated using data simulated under the same models and using a MCAR mechanism except that the within-patient standard deviation, σ_W, was increased from 1.5 to 4.0, the value of σ_B remaining at 8.0. A similar pattern in the bias and MSEs in the estimates of both β and the γ_js was observed as with the lower within-patient standard deviation. However, the magnitude of the estimated bias was greater and significantly non-zero for all eleven parameters. Moreover, the MSE decreases significantly when the within-patient standard deviation is increased, the estimated relative decrease being 44% for β and ranging from 32% to 40% for the γ_js.

For β, the results of the simulations with a nominal 5%-50% of covariate values MCAR can be found in Tables 4.15 and 4.16. As the parameter β was the only parameter of major interest, the bias and MSE for the γ_j parameter estimates are not included.

When compared with the results from the simulations with lower within-patient variability, there was an increased magnitude of bias in most combinations of imputation method, missing data percentage and event collection strategy but especially in the cases where there had been significant bias with the lower value of σ_W.

Both the estimate and estimated standard error of the MSE were lower for simulations except when there were 50% of covariate values missing and interpolation or multiple imputation was used with screen-independent event collection. With screen-dependent event collection, there was a large and relatively consistent decrease in the observed standard error of the MSE but the estimated MSE still decreased except when there was 50% of missing covariate values. In this case, a decrease in estimated MSE was still observed when imputation was via interpolation but an increase was observed when each of the other two imputation methods were used. With screen-independent event collection, the increase in the mean square error under the regression-based multiple imputation was considerable due to the larger magnitude of the bias. The increase with the screen-dependent event collection was less marked. In the former case there was a moderate increase in the standard error of the MSE whereas in the latter, in common with all other cases, a decrease was observed.

This general decrease in both the MSE and its standard error upon increasing the within-individual variability is caused by the reduction in standard error of the bias estimate which, in turn, is due to the more precise nature of the information about the effect of the covariate on the event and, hence, also about the

interval-specific nuisance parameters.

4.4.2 Data MAR

The estimation of survival parameters was then repeated on data sets constructed using a MAR mechanism. A logistic function was used for the probability of a missing covariate measurement. This took the form:

$$p_{ij} = \frac{e^{\eta_{ij}}}{1 + e^{\eta_{ij}}}$$

where

$$\eta_{ij} = \log\left(\frac{\pi}{1-\pi}\right) + 0.075(x_{ij-1} - 84.25)$$

where the value of π corresponds to the nominal proportion of missing values (0.05, 0.1, 0.2 or 0.5). The value 84.25 was chosen as baseline because it is the mean value over the eleven repeated measurements for an individual following the expected profile. The value 0.075 was chosen to give a moderate increase in the probability of missingness upon a moderate increase in covariate value. For example, the probability of missingness increases from the nominal 0.05 to 0.100 upon increasing the previous covariate value from 84.25 to 94.25. As individuals with higher covariate values had increased susceptibility to the event of interest, the nominal percentage of missing values was always higher than the true percentage missing. For example, in the first of the five hundred data sets constructed, the percentages missing were as given in Table 4.11.

With regards to β, the results of these simulations can be found in Tables 4.17 and 4.18 (on pages 159 and 160). The results for the nuisance parameters are not included. The results for β show that both linear interpolation imputation and regression-based multiple imputation are preferable to imputation of the last recorded measurement in terms of both bias and mean square error for both forms of event collection. There was little to choose between imputation by linear interpolation and multiple imputation in terms of bias and MSE, although the observed performance of multiple imputation tended to be marginally (but not significantly) better. The performance of all three methods of imputation was relatively poor with screen-dependent event collection in the presence of a nominal 50% missing at screens.

4.4.3 Informative missing data

Finally, the simulations were repeated on data sets subject to informative missing-data mechanisms. Two such mechanisms were adopted. The first was where the probability of a covariate value being missing was directly related, via a logistic function, to the value that would be observed. This will be termed a **value-based missing data mechanism**. The second was where their probability was related, again via a logistic function, to the difference between the value that would be observed and the value at the previous screen. This will be termed a **difference-based missing data mechanism**. These choices were made because they were reasonably simple to interpret and correspond to the risk of a missing measurement increasing as an individual's blood pressure on the day of a potential screen increases and the risk increasing as the change in blood pressure between screens changes, potentially indicating those whose physical condition was worsening most rapidly. In both cases, however, it is important to note that the risk of an individual missing a screen is related to their health at that time and corresponds

to *patient self-selection* in terms of the types of examination schemes discussed in Chapter 2.

For the value-based mechanism, the function used was as for the MAR mechanism but replacing the previous measurement (x_{ij-1}) by the 'present' measurement (x_{ij}).

For the difference-based mechanism, the function η_{ij} was replaced by:

$$\eta_{ij} = \log\left(\frac{\pi}{1-\pi}\right) + 0.4(x_{ij} - x_{ij-1} - 0.885).$$

The value 0.885 was the expected increase for an individual following the mean profile. The value 0.4 was chosen to give a moderate increase in the probability of missingness for a moderately large increase in covariate value between screens. For example, with a nominal 5% missing, an increase in 1 standard deviation (1.5 mm Hg) more than that expected between screens would lead to a probability of 0.0875 of the value at the latter screen being missing. In both cases, the nominal missing data percentages were greater than those observed due to patients with higher or rapidly increasing blood pressures being more likely to be subject to the event of interest. The percentages observed to be missing in the first of the 500 simulated data sets were as shown in Table 4.12.

With regards to the parameter of interest, β, the results of these simulations are given in Tables 4.19 and 4.20 for the value-based missing data mechanism and in Tables 4.21 and 4.22 for the difference-based missing data mechanism. Results for the nuisance parameters are not included.

Inspection of these tables shows a similar pattern to that seen under the MAR

mechanism, with both linear interpolation imputation and regression-based multiple imputation proving superior to imputation of the last recorded measurement. Likewise, the performance of all three imputation methods was relatively poor with 50% missing attendance at screens with screen-dependent event collection when the missingness was dependent on the value that would have been observed. However, although the performance of last measurement imputation showed a moderate deterioration in these circumstances when the difference-based missing data mechanism was applied, no serious deterioration was observed in the performance of either linear interpolation imputation or the regression-based multiple imputation.

Nominal (%)	Actual (%)
5	4.30
10	8.56
20	17.32
50	40.38

Table 4.11: Comparison of actual and nominal percentages of data missing under a MAR mechanism.

Nominal (%)	Actual (%) (value-based)	Actual (%) (difference-based)
5	4.62	4.22
10	9.24	8.65
20	18.46	18.11
50	42.39	44.65

Table 4.12: Comparison of actual and nominal percentages of data missing under the chosen informative missing data mechanisms.

Percent missing	Imputation method	BIAS (units 10^{-3}) Estimate	95% C.I.	MSE Estimate (units 10^{-5})	S.E. (units 10^{-6})
0	N/A	0.148	(-0.344, 0.640)	3.149	1.96
5	LAST	0.155	(-0.338, 0.647)	3.150	1.96
	INTERPOLATION	0.079	(-0.413, 0.572)	3.151	1.96
	MULTIPLE	0.255	(-0.232, 0.742)	3.084	1.78
10	LAST	0.144	(-0.346, 0.635)	3.129	1.93
	INTERPOLATION	-0.009	(-0.500, 0.483)	3.138	1.93
	MULTIPLE	0.053	(-0.541, 0.436)	3.099	1.93
20	LAST	0.170	(-0.322, 0.663)	3.157	1.94
	INTERPOLATION	-0.277	(-0.766, 0.213)	3.119	1.92
	MULTIPLE	-0.492	(-0.944, -0.039)	2.684	1.64
50	LAST	-0.126	(-0.622, 0.369)	3.193	1.97
	INTERPOLATION	-3.053	(-3.545, -2.562)	4.065	2.44
	MULTIPLE	-1.665	(-2.149, -1.180)	3.326	2.12

Table 4.13: Performance of three imputation methods for 'adapted Finkelstein' under a Weibull mortality process with shape parameter 1.15 and $\sigma_W = 1.5$ and $\sigma_B = 8.0$ under screen-independent event collection with a single continuous repeated measurement and various percentages of missing data generated via a MCAR mechanism.

Percent missing	Imputation method	BIAS (units 10^{-3}) Estimate	95% C.I.	MSE Estimate (units 10^{-5})	S.E. (units 10^{-6})
0	N/A	0.148	(-0.344, 0.640)	3.149	1.96
5	LAST	0.471	(-0.023, 0.965)	3.192	2.00
	INTERPOLATION	0.139	(-0.354, 0.633)	3.163	1.97
	MULTIPLE	0.133	(-0.335, 0.602)	2.850	1.66
10	LAST	0.808	(0.314, 1.302)	3.231	2.01
	INTERPOLATION	0.123	(-0.370, 0.616)	3.153	1.94
	MULTIPLE	0.350	(-0.113, 0.814)	2.803	1.78
20	LAST	1.612	(1.113, 2.112)	3.500	2.18
	INTERPOLATION	0.066	(-0.429, 0.561)	3.178	1.94
	MULTIPLE	0.149	(-0.357, 0.656)	3.336	1.88
50	LAST	4.536	(4.023, 5.047)	5.461	3.24
	INTERPOLATION	-1.146	(-1.648, -0.644)	3.405	2.03
	MULTIPLE	-0.882	(-1.367, -0.397)	3.132	1.91

Table 4.14: Performance of three imputation methods for 'adapted Finkelstein' under a Weibull mortality process with shape parameter 1.15 and $\sigma_W = 1.5$ and $\sigma_B = 8.0$ under screen-dependent event collection with a single continuous repeated measurement and various percentages of missing data generated via a MCAR mechanism.

Percent missing	Imputation method	BIAS (units 10^{-3}) Estimate	95% C.I.	MSE Estimate (units 10^{-5})	S.E. (units 10^{-6})
0	N/A	0.707	(0.342, 1.071)	1.775	1.11
5	LAST	0.653	(0.286, 1.020)	1.796	1.10
	INTERPOLATION	0.646	(0.278, 1.015)	1.805	1.11
	MULTIPLE	-0.139	(-0.513, 0.234)	1.816	1.20
10	LAST	0.588	(0.222, 0.953)	1.773	1.07
	INTERPOLATION	0.545	(0.175, 0.915)	1.807	1.09
	MULTIPLE	-0.591	(-0.951, -0.232)	1.710	1.17
20	LAST	0.429	(0.060, 0.798)	1.789	1.09
	INTERPOLATION	0.146	(-0.228, 0.520)	1.817	1.08
	MULTIPLE	-1.990	(-2.363, -1.617)	2.204	1.42
50	LAST	-0.717	(-1.082, -0.352)	1.781	1.10
	INTERPOLATION	-6.069	(-6.466, -5.673)	5.725	2.83
	MULTIPLE	-7.658	(-7.981, -7.334)	7.224	2.55

Table 4.15: Performance of three imputation methods for 'adapted Finkelstein' under a Weibull mortality process with shape parameter 1.15 and $\sigma_W = 4.0$ and $\sigma_B = 8.0$ under screen-independent event collection with a single continuous repeated measurement and various percentages of missing data generated via a MCAR mechanism.

Percent missing	Imputation method	BIAS (units 10^{-3}) Estimate	95% C.I.	MSE Estimate (units 10^{-5})	S.E. (units 10^{-6})
0	N/A	0.707	(0.342, 1.071)	1.775	1.11
5	LAST	0.995	(0.627, 1.362)	1.853	1.14
	INTERPOLATION	0.798	(0.431, 1.166)	1.820	1.13
	MULTIPLE	0.132	(-0.244, 0.507)	1.833	1.23
10	LAST	1.292	(0.922, 1.662)	1.946	1.19
	INTERPOLATION	0.872	(0.501, 1.244)	1.867	1.14
	MULTIPLE	-0.017	(-0.380, 0.347)	1.714	1.21
20	LAST	1.985	(1.609, 2.362)	2.239	1.42
	INTERPOLATION	1.015	(0.639, 1.391)	1.942	1.23
	MULTIPLE	-0.862	(-1.246, -0.479)	1.987	1.34
50	LAST	4.249	(3.846, 4.652)	3.915	2.16
	INTERPOLATION	-0.128	(-0.527, 0.271)	2.066	1.24
	MULTIPLE	-4.327	(-4.679, -3.975)	3.481	1.75

Table 4.16: Performance of three imputation methods for 'adapted Finkelstein' under a Weibull mortality process with shape parameter 1.15 and $\sigma_W = 4.0$ and $\sigma_B = 8.0$ under screen-dependent event collection with a single continuous repeated measurement and various percentages of missing data generated via a MCAR mechanism.

Percent missing	Imputation method	BIAS (units 10^{-3}) Estimate	95% C.I.	MSE Estimate (units 10^{-5})	S.E. (units 10^{-6})
0	N/A	0.148	(-0.344, 0.640)	3.149	1.96
5	LAST	0.512	(0.017, 1.007)	3.204	1.98
	INTERPOLATION	0.143	(-0.351, 0.636)	3.166	1.97
	MULTIPLE	-0.103	(-0.579, 0.372)	2.941	1.76
10	LAST	0.852	(0.351, 1.352)	3.328	2.07
	INTERPOLATION	0.082	(-0.411, 0.575)	3.159	1.95
	MULTIPLE	0.087	(-0.387, 0.561)	2.916	1.88
20	LAST	1.437	(0.935, 1.940)	3.482	2.21
	INTERPOLATION	-0.310	(-0.802, 0.182)	3.149	1.93
	MULTIPLE	-0.598	(-1.073, -0.123)	2.965	1.88
50	LAST	2.801	(2.284, 3.318)	4.255	2.65
	INTERPOLATION	-3.580	(-4.065, -3.095)	4.334	2.54
	MULTIPLE	-2.046	(-2.524, -1.567)	3.389	1.95

Table 4.17: Performance of three imputation methods for 'adapted Finkelstein' under a Weibull mortality process with shape parameter 1.15 and $\sigma_W = 1.5$ and $\sigma_B = 8.0$ under screen-independent event collection with a single continuous repeated measurement and various percentages of missing data generated via a MAR mechanism.

Percent missing	Imputation method	BIAS (units 10^{-3}) Estimate	95% C.I.	MSE Estimate (units 10^{-5})	S.E. (units 10^{-6})
0	N/A	0.148	(-0.344, 0.640)	3.149	1.96
5	LAST	2.215	(1.716, 2.715)	3.735	2.34
	INTERPOLATION	1.362	(0.866, 1.858)	3.380	2.12
	MULTIPLE	1.172	(0.693, 1.651)	3.117	1.95
10	LAST	4.212	(3.704, 4.721)	5.132	3.10
	INTERPOLATION	2.499	(2.000, 2.998)	3.859	2.43
	MULTIPLE	2.551	(2.070, 3.032)	3.655	2.35
20	LAST	8.212	(7.693, 8.732)	10.254	5.06
	INTERPOLATION	4.672	(4.169, 5.175)	5.473	3.29
	MULTIPLE	4.272	(3.782, 4.763)	4.952	2.90
50	LAST	20.890	(20.325, 21.454)	47.78	12.6
	INTERPOLATION	10.336	(9.813, 10.859)	14.232	6.14
	MULTIPLE	9.881	(9.353, 10.409)	13.383	6.10

Table 4.18: Performance of three imputation methods for 'adapted Finkelstein' under a Weibull mortality process with shape parameter 1.15 and $\sigma_W = 1.5$ and $\sigma_B = 8.0$ under screen-dependent event collection with a single continuous repeated measurement and various percentages of missing data generated via a MAR mechanism.

Percent missing	Imputation method	BIAS (units 10^{-3}) Estimate	95% C.I.	MSE Estimate (units 10^{-5})	S.E. (units 10^{-6})
0	N/A	0.148	(-0.344, 0.640)	3.149	1.96
5	LAST	0.593	(0.096, 1.090)	3.240	2.01
	INTERPOLATION	0.096	(-0.398, 0.591)	3.176	1.98
	MULTIPLE	0.465	(-0.001, 0.931)	2.842	1.86
10	LAST	1.029	(0.529, 1.530)	3.358	2.11
	INTERPOLATION	-0.030	(-0.523, 0.462)	3.148	1.94
	MULTIPLE	-0.136	(-0.593, 0.320)	2.709	1.67
20	LAST	1.749	(1.245, 2.253)	3.602	2.26
	INTERPOLATION	-0.624	(-1.117, -0.130)	3.206	1.96
	MULTIPLE	-0.179	(-0.628, 0.270)	2.625	1.74
50	LAST	3.187	(2.665, 3.708)	4.545	2.81
	INTERPOLATION	-4.633	(-5.123, -4.143)	5.265	2.95
	MULTIPLE	-1.564	(-2.040, -1.087)	3.195	2.00

Table 4.19: Performance of three imputation methods for 'adapted Finkelstein' under a Weibull mortality process with shape parameter 1.15 and $\sigma_W = 1.5$ and $\sigma_B = 8.0$ under screen-independent event collection with a single continuous repeated measurement and various percentages of missing data generated via a value-based non-ignorable mechanism.

Percent missing	Imputation method	BIAS (units 10^{-3}) Estimate	95% C.I.	MSE Estimate (units 10^{-5})	S.E. (units 10^{-6})
0	N/A	0.148	(-0.344, 0.640)	3.149	1.96
5	LAST	2.542	(2.040, 3.043)	3.912	2.45
	INTERPOLATION	1.517	(1.020, 2.014)	3.441	2.17
	MULTIPLE	1.927	(1.455, 2.399)	3.265	2.07
10	LAST	4.927	(4.416, 5.438)	5.819	3.45
	INTERPOLATION	2.858	(2.357, 3.359)	4.074	2.59
	MULTIPLE	2.758	(2.297, 3.219)	3.522	2.23
20	LAST	9.693	(9.171, 10.216)	12.941	5.78
	INTERPOLATION	5.486	(4.982, 5.991)	6.317	3.62
	MULTIPLE	5.726	(5.258, 6.193)	6.118	3.53
50	LAST	24.327	(23.765, 24.889)	63.29	14.5
	INTERPOLATION	12.866	(12.340, 13.392)	20.145	7.45
	MULTIPLE	13.994	(13.470, 14.517)	23.146	7.92

Table 4.20: Performance of three imputation methods for 'adapted Finkelstein' under a Weibull mortality process with shape parameter 1.15 and $\sigma_W = 1.5$ and $\sigma_B = 8.0$ under screen-dependent event collection with a single continuous repeated measurement and various percentages of missing data generated via a value-based non-ignorable mechanism.

Percent missing	Imputation method	BIAS (units 10^{-3}) Estimate	95% C.I.	MSE Estimate (units 10^{-5})	S.E. (units 10^{-6})
0	N/A	0.148	(-0.344, 0.640)	3.149	1.96
5	LAST	0.153	(-0.339, 0.645)	3.148	1.96
	INTERPOLATION	0.040	(-0.453, 0.533)	3.158	1.96
	MULTIPLE	0.465	(-0.001, -0.931)	2.842	1.86
10	LAST	0.144	(-0.347, 0.635)	3.135	1.96
	INTERPOLATION	-0.120	(-0.613, 0.372)	3.154	1.95
	MULTIPLE	0.234	(-0.220, 0.688)	2.682	1.60
20	LAST	0.115	(-0.377, 0.608)	3.156	1.98
	INTERPOLATION	-0.541	(-1.034, -0.049)	3.177	1.97
	MULTIPLE	-0.101	(-0.603, 0.402)	3.283	1.95
50	LAST	-0.244	(-0.741, 0.253)	3.212	2.03
	INTERPOLATION	-3.757	(-4.243, -3.270)	4.484	2.59
	MULTIPLE	-1.674	(-2.145, -1.202)	3.168	1.95

Table 4.21: Performance of three imputation methods for 'adapted Finkelstein' under a Weibull mortality process with shape parameter 1.15 and $\sigma_W = 1.5$ and $\sigma_B = 8.0$ under screen-independent event collection with a single continuous repeated measurement and various percentages of missing data generated via a difference-based non-ignorable mechanism.

Percent missing	Imputation method	BIAS (units 10^{-3}) Estimate	95% C.I.	MSE Estimate (units 10^{-5})	S.E. (units 10^{-6})
0	N/A	0.148	(-0.344, 0.640)	3.149	1.96
5	LAST	0.496	(0.003, 0.989)	3.181	2.00
	INTERPOLATION	0.119	(-0.374, 0.612)	3.156	1.97
	MULTIPLE	0.175	(-0.285, 0.636)	2.757	1.58
10	LAST	0.861	(0.369, 1.353)	3.220	2.04
	INTERPOLATION	0.063	(-0.429, 0.555)	3.142	1.96
	MULTIPLE	0.424	(-0.029, 0.878)	2.691	1.61
20	LAST	1.685	(1.187, 2.183)	3.504	2.25
	INTERPOLATION	-0.073	(-0.568, 0.421)	3.178	1.99
	MULTIPLE	0.313	(-0.193, 0.819)	3.336	1.94
50	LAST	4.792	(4.281, 5.304)	5.696	3.40
	INTERPOLATION	-1.427	(-1.925, -0.929)	3.424	2.16
	MULTIPLE	-0.371	(-0.861, 0.119)	3.134	1.99

Table 4.22: Performance of three imputation methods for 'adapted Finkelstein' under a Weibull mortality process with shape parameter 1.15 and $\sigma_W = 1.5$ and $\sigma_B = 8.0$ under screen-dependent event collection with a single continuous repeated measurement and various percentages of missing data generated via a difference-based non-ignorable mechanism.

4.5 Imputation for incomplete repeated measurements of continuous and ordinal covariates

In modelling data arising from the annual screening of the elderly or similar screening studies, there will typically be a number of time-varying covariates of both ordinal and continuous nature. Therefore, to make progress towards both modelling such data and gaining further understanding of the likely effects of missing data on the usefulness of the modelling methods considered in Section 4.4, it was decided to extend the previous model to one where the hazard depended on two covariates, one ordinal and one continuous. The continuous variable was simulated in the same manner as in Section 4.4. The ordinal variable, which has four levels, was simulated using a transition proportional odds model (McCullagh (1980), Diggle, Liang and Zeger (1994)). The variable simulated was designed to represent a disability scale, in which the first level represented 'no disability' and the fourth level represented 'severe disability'. As there were four levels, and none of the sixteen possible transitions was deemed impossible, there were twelve model parameters. Initially the continuous and ordinal variables were simulated independently. The probabilities of membership of each level at the first time-point were:

$$1 : 0.25 \quad 2 : 0.40 \quad 3 : 0.20 \quad 4 : 0.15$$

and the transition probability matrix used to generate the values at the ten later time-points was:

$$\begin{pmatrix} 0.60 & 0.25 & 0.09 & 0.06 \\ 0.20 & 0.50 & 0.18 & 0.12 \\ 0.04 & 0.11 & 0.60 & 0.25 \\ 0.01 & 0.03 & 0.08 & 0.88 \end{pmatrix}.$$

In addition, simulations were performed subjecting the initial value of the ordinal variable to rank correlations of approximately 0.3 and 0.6 with the initial value of the continuous variable. This was done using a logistic model. The values of each of the variables at the later time points were then generated independently from their initial value in an identical manner to when the initial values were uncorrelated. A Weibull model was once more used to simulate the survival probabilities. The values of the parameters common to both this model and that used in Section 4.4 were the same except for $\log\lambda$ which was reduced from -9.2 to -9.7. The additional parameters corresponding to the ordinal variable were $\alpha_1 = 0.5$, $\alpha_2 = 1.0$, $\alpha_3 = 2.0$ and where the linear predictor for the hazard function (excluding the contribution from the baseline hazard) was of the form:

$$\beta x_{ij} + \alpha_1 z_{ij2} + \alpha_2 z_{ij3} + \alpha_3 z_{ij4} \qquad i = 1, \ldots, 1000, \qquad j = 0, \ldots, 9,$$

where z_{ijk} is the dummy variable indicating membership of the k^{th} level of the ordinal variable for the i^{th} individual at the start of the $j+1^{\text{th}}$ year. The reduction of $\log\lambda$ by 0.5 corresponds to level 2 of the ordinal variable having the survival probabilities considered in Section 4.4 (for given x_{ij}). Missing values were simulated in a similar way to that used in Section 4.4. In this case a logistic function was again used for the probabilities of the observations being missing but the linear predictor was chosen to be:

$$\eta_{ij} = \log\left(\frac{\pi}{1-\pi}\right) + 0.05\,(x_{ij-1} - 84.25) - 0.75z_{ij-11} - 0.25z_{ij-12}$$

$$+0.25z_{ij-13} + 0.75z_{ij-14} \qquad i = 1, \ldots, 1000, \quad j = 1, \ldots 10\,,$$

for the MAR mechanism,

$$\eta_{ij} = \log\left(\frac{\pi}{1-\pi}\right) + 0.05\,(x_{ij} - 84.25) - 0.75z_{ij1} - 0.25z_{ij2}$$

$$+0.25z_{ij3} + 0.75z_{ij4} \qquad i = 1, \ldots, 1000, \quad j = 1, \ldots 10\,,$$

for the value-based non-ignorable mechanism, and

$$\eta_{ij} = \log\left(\frac{\pi}{1-\pi}\right) + 0.3\,(x_{ij} - x_{ij-1} - 0.885) + 0.5\,(z_{ij2} - z_{ij-12})$$

$$+1.0\,(z_{ij3} - z_{ij-13}) + 1.5\,(z_{ij4} - z_{ij-14}) \quad i = 1, \ldots, 1000, \; j = 1, \ldots 10,$$

the difference-based non-ignorable mechanism.

Only cases where the two variables were jointly missing were considered. The two imputation methods considered were last measurement imputation and multiple imputation. Imputation by interpolation was not used for several reasons. Firstly, ordinal variables do not typically satisfy the general requirement of being

interval data and, additionally, would require either some degree of rounding or allocation to a level based on a suitable random process. Secondly, the variability in levels would tend to be less and any dramatic fluctuations in a value of this ordinal variable would be lost in the imputation process. This will always be a particular problem when the non-ignorable difference-based missing-data mechanism is applied. Finally, when the single continuous risk factor was investigated, little difference was observed between the performance of imputation by interpolation and that of regression-based multiple imputation. The standard errors obtained directly from the inverse of either the observed or expected information matrix would, however, be too low if imputation by interpolation were used because the imputed data would be treated as known.

For the multiple imputation process, the continuous variable and ordinal variable were modelled and imputed separately, both when they were uncorrelated and when they were correlated. Both imputation procedures would have allowed the flexibility for any of the earlier realisations of the other covariate to have been used. For the continuous variable, the multiple imputation process was the same as in Section 4.4 whereas that for the ordinal variable was based on a transition ordinal proportional odds regression model (Diggle, Liang and Zeger, 1994). The model used was of the form:

$$\log\left[\frac{P\left(Z_{ij} \leq k \mid Z_{ij-1} = z_{ij-1}\right)}{P\left(Z_{ij} > k \mid Z_{ij-1} = z_{ij-1}\right)}\right] = \theta_k + \sum_{l=1}^{3} \tau_{lk} z^*_{ij-1\,l}$$

$$i = 1, \ldots, 1000, \quad j = 1, \ldots, 10, \quad k = 1, 2, 3,$$

where $$z^*_{ij-1\,l} = \begin{cases} 1 & \text{if } z_{ij-1} \leq l \\ 0 & \text{if } z_{ij-1} > l \end{cases}.$$

The parameter vector

$$\underline{\theta}^T = (\theta_1, \theta_2, \theta_3, \tau_{11}, \tau_{12}, \tau_{13}, \tau_{21}, \tau_{22}, \tau_{23}, \tau_{31}, \tau_{32}, \tau_{33})$$

was estimated using the maximum likelihood and $\text{Var}(\hat{\underline{\theta}})$ was estimated using the inverse of the Fisher (expected) information matrix. The estimation was performed using the quasi-Newton-Raphson NAG subroutine E04KAF which uses the first partial derivatives of the log-likelihood function. The equations given in the Appendix of McCullagh (1980) were used in an appropriate manner (see Appendix C). Fisher's information matrix was also obtained by applying the appropriate equations given in McCullagh (1980) (see Appendix C). This matrix was inverted using the NAG Subroutine F01ABF. Approximate multivariate normality of the vector of parameter estimates was then assumed and five values of $\underline{\theta}$ were simulated using the point estimate of $\underline{\theta}$ as mean and the inverse of the Fisher's information matrix evaluated at $\underline{\theta}$ as covariance matrix. These values of $\underline{\theta}$ were then used to complete for the ordinal variable in the creation of the five multiples of the data set. Likewise, five separate data values were simulated for the missing values of the continuous variable in the manner described in Section 4.4 to complete the five data sets. The parameters of the survival model were then estimated using 'adapted Finkelstein' for each of the five completed data sets and the resulting estimates and their estimated variances were combined to give estimates and standard errors corrected for the degree of missingness, as described earlier.

4.5.1 Results and Conclusions

As previously, the results presented concentrate on the parameters of interest, namely β, α_1, α_2 and α_3. Whilst the nuisance parameters are not discussed here, the patterns of bias and MSE tend to reflect those of the estimates of the parameters of interest. Tables 4.23-4.30 illustrate the performance of last measurement imputation and the model-based multiple imputation, previously described in the estimation of β, over different missing data percentages for a limited selection of combinations of missing data mechanisms and initial value rank correlations. Although the results for the estimation of the γ_j are not shown, they again tend to reflect the negative correlation between $\hat{\beta}$ and each of the $\hat{\gamma}_j$.

Last measurement imputation

Considering firstly the estimation of β, the patterns in terms of bias and mean square error were consistent over the different initial value correlations considered. For screen-dependent event collection and for all missing data mechanisms considered, the bias was always significantly positive and became larger as the amount of data missing increased. The bias became particularly large with the missing at random and value-based non-ignorable mechanisms. The largest bias was with the MAR data for which it reached 0.0153 with 95% confidence interval $(0.0148, 0.0157)$ with 50% missing, an over-estimation of around 22%. For screen-independent event collection, the bias was generally slightly, but not always significantly, positive when there were 5% of the data missing and decreased as the percentage of data missing increased, becoming significantly negative with 50% missing. The exception to this was when the missing at random mechanism was applied. In this case, the bias started positive and became slightly larger in magnitude as the percentage of missing data increased, but still led to a relative

bias of only 1.5% when there were 50% of data missing. The pattern in MSE reflected the size of the bias as the effect on the variance of parameter estimates is simply that of an increase with increasing amounts of data missing.

For the situations where the two variables were uncorrelated there was also a consistent pattern in the bias in the estimation of the ordinal variable parameters when the data were MCAR or subjected to a non-ignorable missing data mechanism. The bias was significantly negative (at the 5% level) for all three parameters for both screen-dependent and screen-independent event collection, except that α_1 and sometimes α_2 were not significantly different from zero when there was a low percentage (5% or 10%) of data missing. The degree of bias was usually a little less, and sometimes significantly so, when the event collection was screen-dependent rather than screen-independent. As expected, the size of the bias increased as the amount of missing data increased and was also larger when the within-patient variability was increased under the MCAR mechanism.

With the MAR data, although the parameter α_3 remained significantly negative over the range of percentages of missing-data under screen-independent event collection, the pattern of bias was less clear cut. With 5% missing data, the biases in α_1 and α_2 were not significantly different from zero under either form of event collection, nor was it for α_3 under the screen-dependent event collection. Under the screen-independent event collection, the bias in α_1 remained non-significant as the percentage of missing data increased. This was also observed for the bias in α_2 until there were 50% data missing. For the larger percentages of missing data with the screen-dependent event collection, however, there was significantly positive bias in the estimation of α_1, α_2 and α_3, except in the case of α_3 with 50% missing when the bias was non-significant at the 5% level.

On increasing the rank correlation between the initial values of the two predictor variables from 0 to approximately 0.3 and then to approximately 0.6, the

relative deterioration in the estimation of the ordinal variable parameters as the percentage of data missing increased became less marked except when the missing at random mechanism was applied in which case no specific trends could be detected. However, when there were no data missing, the performance of 'adapted Finkelstein' deteriorated consistently by a small amount as the rank correlation increased. This trend was reversed with the highest percentage missing data, again with the exception of the missing at random mechanism where no general pattern was observable.

Multiple imputation

The application of the multiple imputation technique described earlier in this section is more problematical when there is an ordinal variable. Firstly, transition models such as the proportional odds model are susceptible to convergence problems due to observed zeros for certain transitions if the respective probabilities are small and insufficient data are collected or simulated. To avoid this problem, all twelve transition probabilities were chosen to be at least 0.01. This will, in turn, lead to imputations which will tend to be inconsistent with the next observed value of the ordinal variable as this value is not taken into account in the imputation model. It is therefore plausible, in certain situations, that the performance of this multiple imputation method may be worse than that of last measurement imputation. It should be noted, however, that the relative ease of obtaining appropriate standard errors may outweigh a slight deterioration in performance. In the analysis of real data, a more appropriate multiple imputation model and, ideally, sensitivity analysis should be used.

The performance of multiple imputation was generally disappointing, tending to result in under-estimation of the parameters of interest. The only exceptions to

this were the over-estimation of β for the MAR and non-ignorable mechanisms and also for α_1 and α_2 for the MAR mechanism. The degree of bias consistently increased as the missing data percentage was increased for all situations investigated. With 50% data missing, the relative bias in the estimation of the α_s varied from 17-54% of the true parameter value, with typical under-estimation of around 30%, with the exception of screen-dependent event collection with either a missing at random mechanism or value-based missing data mechanism in which cases the performance was considerably better. When the estimation of β resulted in negative bias, this was between 5% and 10% of the true value of 0.07.

Comparison of the imputation methods

The parameter estimates obtained following multiple imputation were consistently lower than those from last measurement imputation, indicating a weakening of the effects detected. This is to be expected, particularly in cases where a relatively naive multiple imputation process is applied to processes which are not truly missing completely at random and, in fact, is less worrying than an over-estimation of the effects. However the degree of the under-estimation is far larger than would be hoped. The situations where the multiple imputation performed well in comparison with last measurement imputation were those where last measurement imputation led to large positive bias, as in the estimation of β with screen-dependent event collection and data either missing at random or subjected to a value-based non-ignorable missing data mechanism. Tables 4.23-4.30 show the performance of the two imputation methods under both screen-independent and screen-dependent event collection for both correlated and uncorrelated MCAR data and for uncorrelated MAR data and value-based non-ignorable missing data.

One possible explanation for the disappointing performance of the multiple im-

putation method in this instance is the moderately sized probabilities in the strict lower triangle of the transition matrix leading to imputations which deviate from the true pattern of the repeated ordinal observations by a large extent for a moderate proportion of individuals. It was therefore decided to investigate the relative performances of the two imputation methods with an alternative transition matrix for the repeated observations of the ordinal variable in which extreme or unexpected changes of the ordinal variable were made less likely. This will be reported in Section 4.5.2.

Percent missing	Imputation method	Parameter	BIAS (β: units 10^{-3}) Estimate	95% C.I.	MSE (β: units 10^{-6}) Estimate	S.E.
0	**N/A**	β	**0.202**	(-0.193, 0.596)	**20.23**	1.29
		α_1	**0.0120**	(-0.0051, 0.0291)	**0.0381**	0.0025
		α_2	**0.0002**	(-0.0149, 0.0152)	**0.0295**	0.0020
		α_3	**0.0083**	(-0.0056, 0.0221)	**0.0250**	0.0016
5	**LAST**	β	**0.094**	(-0.304, 0.491)	**20.53**	1.33
		α_1	**-0.0005**	(-0.0172, 0.0161)	**0.0360**	0.0024
		α_2	**-0.0187**	(-0.0337, -0.0037)	**0.0297**	0.0020
		α_3	**-0.0301**	(-0.0438, -0.0165)	**0.0251**	0.0015
	MULTIPLE	β	**-0.409**	(-0.814, -0.004)	**21.42**	1.30
		α_1	**-0.0109**	(-0.0269, 0.0051)	**0.0334**	0.0021
		α_2	**-0.0291**	(-0.0437, -0.0144)	**0.0286**	0.0019
		α_3	**-0.0473**	(-0.0610, -0.0336)	**0.0266**	0.0016
10	**LAST**	β	**-0.038**	(-0.434, 0.357)	**20.36**	1.30
		α_1	**-0.0125**	(-0.0290, 0.0040)	**0.0354**	0.0024
		α_2	**-0.0392**	(-0.0540, -0.0244)	**0.0299**	0.0019
		α_3	**-0.0695**	(-0.0829, -0.0560)	**0.0284**	0.0017
	MULTIPLE	β	**-0.269**	(-0.657, 0.120)	**19.66**	1.26
		α_1	**-0.0353**	(-0.0505, -0.0201)	**0.0313**	0.0019
		α_2	**-0.0608**	(-0.0756, -0.0461)	**0.0319**	0.0020
		α_3	**-0.1051**	(-0.1185, -0.0918)	**0.0341**	0.0019
20	**LAST**	β	**-0.433**	(-0.827, -0.040)	**20.30**	1.29
		α_1	**-0.0402**	(-0.0558, -0.0246)	**0.0332**	0.0022
		α_2	**-0.0839**	(-0.0981, -0.0696)	**0.0334**	0.0021
		α_3	**-0.1594**	(-0.1723, -0.1466)	**0.0470**	0.0024
	MULTIPLE	β	**-1.885**	(-2.284, -1.487)	**24.17**	1.44
		α_1	**-0.0889**	(-0.1033, -0.0745)	**0.0348**	0.0020
		α_2	**-0.1297**	(-0.1435, -0.1159)	**0.0416**	0.0022
		α_3	**-0.2331**	(-0.2458, -0.2205)	**0.0751**	0.0031
50	**LAST**	β	**-2.595**	(-2.990, -2.199)	**27.07**	1.59
		α_1	**-0.1371**	(-0.1504, -0.1238)	**0.0418**	0.0022
		α_2	**-0.2581**	(-0.2711, -0.2450)	**0.0887**	0.0036
		α_3	**-0.4909**	(-0.5030, -0.4788)	**0.2600**	0.0059
	MULTIPLE	β	**-5.516**	(-5.892, -5.140)	**48.80**	2.50
		α_1	**-0.2265**	(-0.2393, -0.2137)	**0.0726**	0.0032
		α_2	**-0.3356**	(-0.3478, -0.3233)	**0.1321**	0.0043
		α_3	**-0.6451**	(-0.6562, -0.6339)	**0.4322**	0.0071

Table 4.23: Performance of two imputation methods under 'adapted Finkelstein' with a Weibull mortality process with shape parameter 1.15 and **screen-independent** event collection with two **uncorrelated** repeated measurements (one continuous with $\sigma_W = 1.5$ and $\sigma_B = 8.0$ and one ordinal with the original transition matrix) and various percentages of jointly missing data generated via a **MCAR** mechanism.

Percent missing	Imputation method	Parameter	BIAS (β: units 10^{-3}) Estimate	95% C.I.	MSE (β: units 10^{-6}) Estimate	S.E.
0	**N/A**	β	**0.202**	(-0.193, 0.596)	**20.23**	1.29
		α_1	**0.0120**	(-0.0051, 0.0291)	**0.0381**	0.0025
		α_2	**0.0002**	(-0.0149, 0.0152)	**0.0295**	0.0020
		α_3	**0.0083**	(-0.0056, 0.0221)	**0.0250**	0.0016
5	**LAST**	β	**0.465**	(0.064, 0.866)	**21.09**	1.38
		α_1	**-0.0013**	(-0.0180, 0.0154)	**0.0363**	0.0024
		α_2	**-0.0151**	(-0.0302, 0.0000)	**0.0298**	0.0020
		α_3	**-0.0229**	(-0.0366, -0.0091)	**0.0249**	0.0015
	MULTIPLE	β	**-0.313**	(-0.720, 0.095)	**21.66**	1.33
		α_1	**0.0011**	(-0.0151, 0.0173)	**0.0342**	0.0022
		α_2	**-0.0193**	(-0.0343, -0.0044)	**0.0295**	0.0020
		α_3	**-0.0398**	(-0.0537, -0.0258)	**0.0270**	0.0016
10	**LAST**	β	**0.741**	(0.337, 1.144)	**21.65**	1.40
		α_1	**-0.0143**	(-0.0309, 0.0023)	**0.0359**	0.0024
		α_2	**-0.0322**	(-0.0471, -0.0173)	**0.0299**	0.0020
		α_3	**-0.0543**	(-0.0678, -0.0407)	**0.0268**	0.0016
	MULTIPLE	β	**-0.051**	(-0.445, 0.344)	**20.22**	1.30
		α_1	**-0.0137**	(-0.0300, 0.0027)	**0.0349**	0.0021
		α_2	**-0.0436**	(-0.0591, -0.0282)	**0.0329**	0.0021
		α_3	**-0.0925**	(-0.1064, -0.0785)	**0.0340**	0.0019
20	**LAST**	β	**1.191**	(0.783, 1.599)	**23.00**	1.47
		α_1	**-0.0444**	(-0.0602, -0.0286)	**0.0343**	0.0022
		α_2	**-0.0686**	(-0.0830, -0.0542)	**0.0317**	0.0020
		α_3	**-0.1269**	(-0.1399, -0.1138)	**0.0382**	0.0021
	MULTIPLE	β	**-1.555**	(-1.964, -1.147)	**24.13**	1.46
		α_1	**-0.0443**	(-0.0602, -0.0285)	**0.0346**	0.0019
		α_2	**-0.0968**	(-0.1121, -0.0815)	**0.0398**	0.0021
		α_3	**-0.2132**	(-0.2269, -0.1994)	**0.0699**	0.0031
50	**LAST**	β	**2.024**	(1.600, 2.448)	**27.45**	1.74
		α_1	**-0.1555**	(-0.1695, -0.1415)	**0.0497**	0.0026
		α_2	**-0.2161**	(-0.2297, -0.2024)	**0.0708**	0.0032
		α_3	**-0.3970**	(-0.4094, -0.3845)	**0.1778**	0.0050
	MULTIPLE	β	**-5.294**	(-5.688, -4.900)	**48.23**	2.56
		α_1	**-0.1070**	(-0.1252 -0.0887)	**0.0547**	0.0032
		α_2	**-0.3065**	(-0.3230, -0.2900)	**0.1291**	0.0054
		α_3	**-0.6777**	(-0.6917, -0.6638)	**0.4846**	0.0096

Table 4.24: Performance of two imputation methods under 'adapted Finkelstein' with a Weibull mortality process with shape parameter 1.15 and **screen-dependent** event collection and two **uncorrelated** repeated measurements (one continuous with $\sigma_W = 1.5$ and $\sigma_B = 8.0$ and one ordinal with the original transition matrix) and various percentages of jointly missing data generated via a **MCAR** mechanism.

Percent missing	Imputation method	Parameter	BIAS (β: units 10^{-3}) Estimate	95% C.I.	MSE (β: units 10^{-6}) Estimate	S.E.
0	N/A	β	**0.456**	(0.045, 0.867)	**22.12**	1.48
		α_1	**0.0113**	(-0.0061, 0.0288)	**0.0397**	0.0025
		α_2	**0.0101**	(-0.0061 0.0262)	**0.0340**	0.0022
		α_3	**0.0090**	(-0.0054, 0.0233)	**0.0269**	0.0017
5	LAST	β	**0.381**	(-0.030, 0.792)	**22.06**	1.47
		α_1	**0.0019**	(-0.0148, 0.0186)	**0.0363**	0.0023
		α_2	**-0.0083**	(-0.0240, 0.0073)	**0.0319**	0.0021
		α_3	**-0.0290**	(-0.0430, -0.0150)	**0.0262**	0.0016
	MULTIPLE	β	**0.214**	(-0.197, 0.625)	**22.02**	1.38
		α_1	**-0.0164**	(-0.0327, -0.0002)	**0.0347**	0.0023
		α_2	**-0.0271**	(-0.0420 -0.0123)	**0.0294**	0.0019
		α_3	**-0.0502**	(-0.0643, -0.0361)	**0.0283**	0.0017
10	LAST	β	**0.282**	(-0.130, 0.693)	**22.07**	1.46
		α_1	**-0.0107**	(-0.0269, 0.0056)	**0.0343**	0.0022
		α_2	**-0.0300**	(-0.0454, -0.0146)	**0.0317**	0.0021
		α_3	**-0.0683**	(-0.0818, -0.0548)	**0.0283**	0.0017
	MULTIPLE	β	**-0.005**	(-0.435, 0.426)	**24.06**	1.44
		α_1	**-0.0340**	(-0.0497, -0.0183)	**0.0333**	0.0019
		α_2	**-0.0538**	(-0.0685, -0.0392)	**0.0308**	0.0020
		α_3	**-0.0971**	(-0.1104, -0.0838)	**0.0324**	0.0019
20	LAST	β	**-0.045**	(-0.457, 0.367)	**22.03**	1.43
		α_1	**-0.0390**	(-0.0544, -0.0236)	**0.0323**	0.0020
		α_2	**-0.0768**	(-0.0917, -0.0619)	**0.0348**	0.0021
		α_3	**-0.1574**	(-0.1704, -0.1444)	**0.0468**	0.0023
	MULTIPLE	β	**-0.646**	(-1.073, -0.220)	**24.05**	1.59
		α_1	**-0.0764**	(-0.0912, -0.0617)	**0.0342**	0.0020
		α_2	**-0.1104**	(-0.1248, -0.0960)	**0.0390**	0.0022
		α_3	**-0.2164**	(-0.2291, -0.2037)	**0.0678**	0.0028
50	LAST	β	**-1.763**	(-2.182, -1.343)	**25.95**	1.57
		α_1	**-0.1379**	(-0.1509, -0.1250)	**0.0410**	0.0022
		α_2	**-0.2457**	(-0.2588, -0.2326)	**0.0826**	0.0033
		α_3	**-0.4783**	(-0.4897, -0.4669)	**0.2457**	0.0056
	MULTIPLE	β	**-3.168**	(-3.576, -2.760)	**31.65**	1.70
		α_1	**-0.2251**	(-0.2384, -0.2119)	**0.0735**	0.0034
		α_2	**-0.3273**	(-0.3395, -0.3151)	**0.1264**	0.0041
		α_3	**-0.6291**	(-0.6403, -0.6178)	**0.4122**	0.0071

Table 4.25: Performance of two imputation methods under 'adapted Finkelstein' with a Weibull mortality process with shape parameter 1.15 and **screen-independent** event collection with two **correlated** repeated measurements with inital value rank correlation $\simeq 0.3$ (one continuous with $\sigma_W = 1.5$ and $\sigma_B = 8.0$ and one ordinal with the original transition matrix) and various percentages of jointly missing data generated via a **MCAR** mechanism.

Percent missing	Imputation method	Parameter	BIAS (β: units 10^{-3}) Estimate	95% C.I.	MSE (β: units 10^{-6}) Estimate	S.E.
0	**N/A**	β	**0.456**	(0.045, 0.867)	**22.12**	1.48
		α_1	**0.0113**	(-0.0061, 0.0288)	**0.0397**	0.0025
		α_2	**0.0101**	(-0.0061, 0.0262)	**0.0340**	0.0022
		α_3	**0.0090**	(-0.0054, 0.0233)	**0.0269**	0.0017
5	**LAST**	β	**0.658**	(0.244, 1.071)	**22.64**	1.51
		α_1	**0.0007**	(-0.0161, 0.0174)	**0.0366**	0.0023
		α_2	**-0.0062**	(-0.0220, 0.0095)	**0.0321**	0.0021
		α_3	**-0.0238**	(-0.0379, -0.0098)	**0.0262**	0.0017
	MULTIPLE	β	**0.226**	(-0.188, 0.639)	**22.25**	1.41
		α_1	**-0.0034**	(-0.0197, 0.0129)	**0.0346**	0.0023
		α_2	**-0.0150**	(-0.0301 0.0002)	**0.0301**	0.0020
		α_3	**-0.0407**	(-0.0550, -0.0264)	**0.0282**	0.0017
10	**LAST**	β	**0.868**	(0.452, 1.284)	**23.22**	1.55
		α_1	**-0.0130**	(-0.0294, 0.0033)	**0.0349**	0.0022
		α_2	**-0.0259**	(-0.0414, -0.0104)	**0.0318**	0.0021
		α_3	**-0.0570**	(-0.0706, -0.0434)	**0.0273**	0.0017
	MULTIPLE	β	**0.065**	(-0.366, 0.497)	**24.20**	1.46
		α_1	**-0.0098**	(-0.0263, 0.0068)	**0.0357**	0.0021
		α_2	**-0.0320**	(-0.0473, -0.0167)	**0.0314**	0.0021
		α_3	**-0.0808**	(-0.0945, -0.0670)	**0.0311**	0.0019
20	**LAST**	β	**1.204**	(0.784, 1.625)	**24.43**	1.63
		α_1	**-0.0443**	(-0.0599, -0.0288)	**0.0335**	0.0021
		α_2	**-0.0677**	(-0.0828, -0.0526)	**0.0343**	0.0021
		α_3	**-0.1334**	(-0.1466, -0.1203)	**0.0403**	0.0021
	MULTIPLE	β	**-0.556**	(-0.988, -0.125)	**24.46**	1.64
		α_1	**-0.0292**	(-0.0455, -0.0129)	**0.0355**	0.0022
		α_2	**-0.0731**	(-0.0883, -0.0578)	**0.0356**	0.0022
		α_3	**-0.1910**	(-0.2045, -0.1774)	**0.0602**	0.0028
50	**LAST**	β	**1.527**	(1.080, 1.973)	**28.26**	1.85
		α_1	**-0.1609**	(-0.1746, -0.1472)	**0.0503**	0.0026
		α_2	**-0.2226**	(-0.2362, -0.2090)	**0.0735**	0.0032
		α_3	**-0.4136**	(-0.4256, -0.4017)	**0.1896**	0.0050
	MULTIPLE	β	**-3.401**	(-3.819, -2.984)	**34.23**	1.88
		α_1	**-0.0883**	(-0.1081, -0.0686)	**0.0583**	0.0033
		α_2	**-0.2733**	(-0.2889, -0.2578)	**0.1061**	0.0045
		α_3	**-0.6409**	(-0.6545, -0.6274)	**0.4347**	0.0089

Table 4.26: Performance of two imputation methods under 'adapted Finkelstein' with a Weibull mortality process with shape parameter 1.15 and **screen-dependent** event collection with two **correlated** repeated measurements with inital value rank correlation $\simeq 0.3$ (one continuous with $\sigma_W = 1.5$ and $\sigma_B = 8.0$ and one ordinal with the original transition matrix) and various percentages of jointly missing data generated via a **MCAR** mechanism.

Percent missing	Imputation method	Parameter	BIAS (β: units 10^{-3}) Estimate	95% C.I.	MSE (β: units 10^{-6}) Estimate	S.E.
0	N/A	β	**0.202**	(-0.193, 0.596)	**20.23**	1.29
		α_1	**0.0120**	(-0.0051, 0.0291)	**0.0381**	0.0025
		α_2	**0.0002**	(-0.0149, 0.0152)	**0.0295**	0.0020
		α_3	**0.0083**	(-0.0056, 0.0221)	**0.0250**	0.0016
5	LAST	β	**0.460**	(0.062, 0.857)	**20.77**	1.34
		α_1	**0.0104**	(-0.0063, 0.0271)	**0.0362**	0.0024
		α_2	**0.0017**	(-0.0134, 0.0168)	**0.0296**	0.0020
		α_3	**-0.0176**	(-0.0314, -0.0039)	**0.0248**	0.0015
	MULTIPLE	β	**-0.468**	(-0.853, -0.084)	**19.41**	1.22
		α_1	**-0.0116**	(-0.0279, 0.0048)	**0.0348**	0.0021
		α_2	**-0.0188**	(-0.0340, 0.0036)	**0.0304**	0.0019
		α_3	**-0.0567**	(-0.0705, -0.0429)	**0.0280**	0.0016
10	LAST	β	**0.668**	(0.270, 1.066)	**21.02**	1.35
		α_1	**0.0128**	(-0.0038, 0.0293)	**0.0357**	0.0024
		α_2	**0.0071**	(-0.0080, 0.0223)	**0.0298**	0.0020
		α_3	**-0.0402**	(-0.0540, -0.0264)	**0.0263**	0.0016
	MULTIPLE	β	**-1.094**	(-1.480, -0.708)	**20.55**	1.20
		α_1	**-0.0233**	(-0.0383, -0.0083)	**0.0298**	0.0019
		α_2	**-0.0440**	(-0.0583, -0.0298)	**0.0283**	0.0017
		α_3	**-0.1174**	(-0.1304, -0.1044)	**0.0357**	0.0019
20	LAST	β	**0.985**	(0.580, 1.390)	**22.29**	1.44
		α_1	**0.0152**	(-0.0010, 0.0314)	**0.0342**	0.0024
		α_2	**0.0078**	(-0.0069, 0.0224)	**0.0279**	0.0020
		α_3	**-0.0930**	(-0.1065, -0.0796)	**0.0321**	0.0019
	MULTIPLE	β	**-2.565**	(-2.979, -2.151)	**28.84**	1.70
		α_1	**-0.0711**	(-0.0856, -0.0567)	**0.0322**	0.0019
		α_2	**-0.0892**	(-0.1032, -0.0753)	**0.0333**	0.0018
		α_3	**-0.2418**	(-0.2548, -0.2289)	**0.0802**	0.0032
50	LAST	β	**1.043**	(0.625, 1.461)	**23.81**	1.49
		α_1	**-0.0057**	(-0.0200, 0.0085)	**0.0265**	0.0017
		α_2	**-0.0526**	(-0.0668, -0.0384)	**0.0290**	0.0019
		α_3	**-0.3203**	(-0.3337, -0.3070)	**0.1257**	0.0045
	MULTIPLE	β	**-7.050**	(-7.430, -6.670)	**68.47**	3.12
		α_1	**-0.1803**	(-0.1943, -0.1663)	**0.0579**	0.0031
		α_2	**-0.2562**	(-0.2687, -0.2438)	**0.0859**	0.0035
		α_3	**-0.6196**	(-0.6316, -0.6075)	**0.4027**	0.0075

Table 4.27: Performance of two imputation methods under 'adapted Finkelstein' with a Weibull mortality process with shape parameter 1.15 and **screen-independent** event collection with two **uncorrelated** repeated measurements (one continuous with $\sigma_W = 1.5$ and $\sigma_B = 8.0$ and one ordinal with the original transition matrix) and various percentages of jointly missing data generated via a **MAR** mechanism.

Percent missing	Imputation method	Parameter	BIAS (β: units 10^{-3}) Estimate	95% C.I.	MSE (β: units 10^{-6}) Estimate	S.E.
0	**N/A**	β	**0.202**	(-0.193, 0.596)	**20.23**	1.29
		α_1	**0.0120**	(-0.0051, 0.0291)	**0.0381**	0.0025
		α_2	**0.0002**	(-0.0149, 0.0152)	**0.0295**	0.0020
		α_3	**0.0083**	(-0.0056, 0.0221)	**0.0250**	0.0016
5	**LAST**	β	**2.108**	(1.702, 2.514)	**25.88**	1.67
		α_1	**0.0123**	(-0.0044, 0.0290)	**0.0365**	0.0024
		α_2	**0.0143**	(-0.0009, 0.0294)	**0.0301**	0.0021
		α_3	**0.0108**	(-0.0030, 0.0246)	**0.0249**	0.0016
	MULTIPLE	β	**0.781**	(0.390, 1.171)	**20.45**	1.31
		α_1	**0.0171**	(0.0003, 0.0339)	**0.0368**	0.0023
		α_2	**0.0236**	(0.0079, 0.0394)	**0.0329**	0.0020
		α_3	**-0.0033**	(-0.0176, 0.0110)	**0.0267**	0.0017
10	**LAST**	β	**3.864**	(3.448, 4.279)	**37.32**	2.22
		α_1	**0.0170**	(0.0003, 0.0336)	**0.0363**	0.0024
		α_2	**0.0330**	(0.0178, 0.0483)	**0.0313**	0.0022
		α_3	**0.0169**	(0.0031, 0.0308)	**0.0252**	0.0017
	MULTIPLE	β	**1.201**	(0.802, 1.601)	**22.16**	1.47
		α_1	**0.0299**	(0.0137, 0.0460)	**0.0347**	0.0023
		α_2	**0.0387**	(0.0237, 0.0537)	**0.0307**	0.0020
		α_3	**-0.0181**	(-0.0318, -0.0045)	**0.0246**	0.0016
20	**LAST**	β	**7.133**	(6.704, 7.562)	**74.75**	3.54
		α_1	**0.0246**	(0.0082, 0.0410)	**0.0355**	0.0025
		α_2	**0.0630**	(0.0481, 0.0778)	**0.0325**	0.0024
		α_3	**0.0238**	(0.0102, 0.0374)	**0.0246**	0.0018
	MULTIPLE	β	**1.444**	(1.008, 1.881)	**26.85**	1.68
		α_1	**0.0462**	(0.0299, 0.0625)	**0.0367**	0.0023
		α_2	**0.0763**	(0.0603, 0.0922)	**0.0388**	0.0023
		α_3	**-0.0541**	(-0.0687, -0.0394)	**0.0307**	0.0017
50	**LAST**	β	**15.283**	(14.827, 15.739)	**260.59**	7.49
		α_1	**0.0301**	(0.0152, 0.0450)	**0.0297**	0.0020
		α_2	**0.1169**	(0.1020, 0.1318)	**0.0425**	0.0026
		α_3	**0.0079**	(-0.0060, 0.0217)	**0.0250**	0.0017
	MULTIPLE	β	**0.937**	(0.529, 1.345)	**22.53**	1.55
		α_1	**0.1328**	(0.1145, 0.1511)	**0.0610**	0.0039
		α_2	**0.1275**	(0.1117, 0.1433)	**0.0486**	0.0028
		α_3	**-0.2475**	(-0.2625, -0.2326)	**0.0901**	0.0051

Table 4.28: Performance of two imputation methods under 'adapted Finkelstein' with a Weibull mortality process with shape parameter 1.15 and **screen-dependent** event collection with two **uncorrelated** repeated measurements (one continuous with $\sigma_W = 1.5$ and $\sigma_B = 8.0$ and one ordinal with the original transition matrix) and various percentages of jointly missing data generated via a **MAR** mechanism.

Percent missing	Imputation method	Parameter	BIAS (β: units 10^{-3}) Estimate	95% C.I.	MSE (β: units 10^{-6}) Estimate	S.E.
0	N/A	β	**0.202**	(-0.193, 0.596)	**20.23**	1.29
		α_1	**0.0120**	(-0.0051, 0.0291)	**0.0381**	0.0025
		α_2	**0.0002**	(-0.0149, 0.0152)	**0.0295**	0.0020
		α_3	**0.0083**	(-0.0056, 0.0221)	**0.0250**	0.0016
5	LAST	β	**0.811**	(0.414, 1.208)	**21.12**	1.39
		α_1	**0.0021**	(-0.0142, 0.0184)	**0.0346**	0.0023
		α_2	**-0.0217**	(-0.0363, -0.0071)	**0.0281**	0.0018
		α_3	**-0.0585**	(-0.0720, -0.0451)	**0.0268**	0.0016
	MULTIPLE	β	**-0.121**	(-0.486, 0.244)	**17.30**	1.07
		α_1	**-0.0248**	(-0.0401, -0.0094)	**0.0312**	0.0019
		α_2	**-0.0405**	(-0.0546 -0.0265)	**0.0272**	0.0016
		α_3	**-0.0916**	(-0.1044, -0.0788)	**0.0296**	0.0016
10	LAST	β	**1.116**	(0.721, 1.511)	**21.50**	1.38
		α_1	**-0.0087**	(-0.0246, 0.0071)	**0.0327**	0.0022
		α_2	**-0.0447**	(-0.0588, -0.0306)	**0.0278**	0.0018
		α_3	**-0.1260**	(-0.1389, -0.1130)	**0.0376**	0.0020
	MULTIPLE	β	**-0.302**	(-0.700, 0.097)	**20.72**	1.31
		α_1	**-0.0464**	(-0.0603, -0.0324)	**0.0273**	0.0017
		α_2	**-0.0827**	(-0.0957, -0.0698)	**0.0286**	0.0017
		α_3	**-0.1880**	(-0.2001, -0.1759)	**0.0545**	0.0025
20	LAST	β	**1.327**	(0.927, 1.727)	**22.58**	1.43
		α_1	**-0.0310**	(-0.0457, -0.0163)	**0.0290**	0.0020
		α_2	**-0.0955**	(-0.1093, -0.0817)	**0.0340**	0.0020
		α_3	**-0.2494**	(-0.2621, -0.2367)	**0.0831**	0.0033
	MULTIPLE	β	**-1.625**	(-2.015, -1.235)	**22.39**	1.39
		α_1	**-0.1216**	(-0.1347, -0.1085)	**0.0371**	0.0021
		α_2	**-0.1822**	(-0.1944, -0.1700)	**0.0526**	0.0024
		α_3	**-0.3761**	(-0.3873, -0.3648)	**0.1578**	0.0043
50	LAST	β	**-0.992**	(-1.410, -0.573)	**23.71**	1.40
		α_1	**-0.1139**	(-0.1248, -0.1030)	**0.0284**	0.0015
		α_2	**-0.2658**	(-0.2770, -0.2547)	**0.0869**	0.0032
		α_3	**-0.6171**	(-0.6277, -0.6064)	**0.3954**	0.0066
	MULTIPLE	β	**-6.304**	(-6.684, -5.923)	**58.55**	2.69
		α_1	**-0.2688**	(-0.2811, -0.2565)	**0.0919**	0.0035
		α_2	**-0.4172**	(-0.4285, -0.4059)	**0.1907**	0.0048
		α_3	**-0.8213**	(-0.8316, -0.8111)	**0.6883**	0.0085

Table 4.29: Performance of two imputation methods under 'adapted Finkelstein' with a Weibull mortality process with shape parameter 1.15 and **screen-independent** event collection with two **uncorrelated** repeated measurements (one continuous with $\sigma_W = 1.5$ and $\sigma_B = 8.0$ and one ordinal with the original transition matrix) and various percentages of jointly missing data generated via a **value-based non-ignorable** missing data mechanism.

Percent missing	Imputation method	Parameter	BIAS (β: units 10^{-3}) Estimate	95% C.I.	MSE (β: units 10^{-6}) Estimate	S.E.
0	**N/A**	β	**0.202**	(-0.193, 0.596)	**20.23**	1.29
		α_1	**0.0120**	(-0.0051, 0.0291)	**0.0381**	0.0025
		α_2	**0.0002**	(-0.0149, 0.0152)	**0.0295**	0.0020
		α_3	**0.0083**	(-0.0056, 0.0221)	**0.0250**	0.0016
5	**LAST**	β	**2.672**	(2.265, 3.079)	**28.66**	1.83
		α_1	**0.0034**	(-0.0130, 0.0199)	**0.0350**	0.0023
		α_2	**-0.0087**	(-0.0234, 0.0060)	**0.0281**	0.0019
		α_3	**-0.0317**	(-0.0453, -0.0182)	**0.0248**	0.0015
	MULTIPLE	β	**1.209**	(0.837, 1.581)	**19.42**	1.24
		α_1	**0.0066**	(-0.0095, 0.0227)	**0.0337**	0.0022
		α_2	**-0.0010**	(-0.0159, 0.0138)	**0.0286**	0.0017
		α_3	**-0.0497**	(-0.0632, -0.0363)	**0.0259**	0.0015
10	**LAST**	β	**4.670**	(4.258, 5.083)	**43.94**	2.47
		α_1	**-0.0059**	(-0.0220, 0.0102)	**0.0336**	0.0023
		α_2	**-0.0188**	(-0.0331, -0.0045)	**0.0270**	0.0018
		α_3	**-0.0737**	(-0.0868, -0.0607)	**0.0276**	0.0016
	MULTIPLE	β	**2.292**	(1.877, 2.707)	**27.66**	1.65
		α_1	**0.0140**	(-0.0010, 0.0291)	**0.0296**	0.0020
		α_2	**-0.0064**	(-0.0204, 0.0076)	**0.0255**	0.0017
		α_3	**-0.1066**	(-0.1197, -0.0936)	**0.0335**	0.0018
20	**LAST**	β	**7.967**	(7.546, 8.389)	**86.57**	3.86
		α_1	**-0.0261**	(-0.0412, -0.0111)	**0.0301**	0.0021
		α_2	**-0.0445**	(-0.0586, -0.0304)	**0.0279**	0.0018
		α_3	**-0.1495**	(-0.1625, -0.1366)	**0.0442**	0.0023
	MULTIPLE	β	**2.901**	(2.487, 3.315)	**30.69**	1.91
		α_1	**-0.0105**	(-0.0252, 0.0042)	**0.0282**	0.0018
		α_2	**-0.0469**	(-0.0607, -0.0330)	**0.0271**	0.0016
		α_3	**-0.2339**	(-0.2464, -0.2214)	**0.0750**	0.0032
50	**LAST**	β	**13.220**	(12.767, 13.673)	**201.47**	6.50
		α_1	**-0.1117**	(-0.1234, -0.1000)	**0.0303**	0.0017
		α_2	**-0.1573**	(-0.1690, -0.1455)	**0.0426**	0.0021
		α_3	**-0.4015**	(-0.4127, -0.3902)	**0.1777**	0.0046
	MULTIPLE	β	**3.004**	(2.576, 3.431)	**32.76**	1.94
		α_1	**-0.0406**	(-0.0560, -0.0251)	**0.0327**	0.0019
		α_2	**-0.1601**	(-0.1735, -0.1467)	**0.0490**	0.0025
		α_3	**-0.5782**	(-0.5916, -0.5648)	**0.3576**	0.0097

Table 4.30: Performance of two imputation methods under 'adapted Finkelstein' with a Weibull mortality process with shape parameter 1.15 and **screen-dependent** event collection with two **uncorrelated** repeated measurements (one continuous with $\sigma_W = 1.5$ and $\sigma_B = 8.0$ and one ordinal with the original transition matrix) and various percentages of jointly missing data generated via a **value-based non-ignorable** missing data mechanism.

4.5.2 The effect of amending the transition matrix

It was decided to use a banded transition matrix to generate the alternative data set which would then be subjected to the same forms of missing data processes as for the previous transition matrix. Moreover, most of the probabilities in the strict lower triangle were reduced to lower the chance of an improvement in disability condition. Minor changes were made to the major diagonal of the matrix and the probabilities on the diagonal representing a worsening of disability by one category between consecutive screens were increased. Corresponding minor amendments were also made to the other entries in the matrix but the probabilities of membership of each level at the first time point were left unchanged. The actual transition matrix used was

$$\begin{pmatrix} 0.55 & 0.30 & 0.10 & 0.05 \\ 0.07 & 0.53 & 0.30 & 0.10 \\ 0.03 & 0.07 & 0.60 & 0.30 \\ 0.01 & 0.03 & 0.07 & 0.89 \end{pmatrix}.$$

Comparison of the imputation methods

On altering the transition matrix, little general change in the patterns of either bias or mean square error of any of the non-nuisance parameters was observed. Tables 4.31-4.38 illustrate the relative performance of the two imputation techniques over different percentages of missing data under selected missing data mechanisms. Over the two imputation methods, the tendency was for little consistent effect on the estimation of β in terms of either bias or mean square error. With the α_s, there was a tendency for both the biases and mean square errors to be larger than previously, especially when the performance was only fair or poor in

the simulations under the previous transition matrix. This led to the performance of multiple imputation deteriorating a little further in relation to that of the last measurement imputation.

Percent missing	Imputation method	Parameter	BIAS (β: units 10^{-3}) Estimate	95% C.I.	MSE (β: units 10^{-6}) Estimate	S.E.
0	N/A	β	**0.222**	(-0.180, 0.624)	**21.04**	1.36
		α_1	**0.0043**	(-0.0170, 0.0255)	**0.0586**	0.0038
		α_2	**-0.0034**	(-0.0231, 0.0163)	**0.0506**	0.0034
		α_3	**0.0030**	(-0.0155, 0.0216)	**0.0447**	0.0029
5	LAST	β	**0.095**	(-0.310, 0.500)	**21.34**	1.39
		α_1	**-0.0059**	(-0.0267, 0.0149)	**0.0561**	0.0037
		α_2	**-0.0231**	(-0.0423, -0.0039)	**0.0486**	0.0032
		α_3	**-0.0395**	(-0.0577, -0.0213)	**0.0447**	0.0028
	MULTIPLE	β	**-0.133**	(-0.542, 0.276)	**21.74**	1.34
		α_1	**-0.0063**	(-0.0266, 0.0139)	**0.0533**	0.0036
		α_2	**-0.0190**	(-0.0386, 0.0005)	**0.0501**	0.0033
		α_3	**-0.0416**	(-0.0595, -0.0237)	**0.0433**	0.0029
10	LAST	β	**-0.056**	(-0.459, 0.348)	**21.15**	1.40
		α_1	**-0.0130**	(-0.0334, 0.0075)	**0.0546**	0.0035
		α_2	**-0.0389**	(-0.0577, -0.0200)	**0.0475**	0.0030
		α_3	**-0.0803**	(-0.0980, -0.0626)	**0.0470**	0.0028
	MULTIPLE	β	**-0.729**	(-1.121, -0.338)	**20.43**	1.25
		α_1	**-0.0389**	(-0.0574, -0.0203)	**0.0463**	0.0031
		α_2	**-0.0637**	(-0.0811, -0.0463)	**0.0435**	0.0028
		α_3	**-0.1171**	(-0.1336, -0.1006)	**0.0491**	0.0026
20	LAST	β	**-0.457**	(-0.860, -0.054)	**21.27**	1.44
		α_1	**-0.0410**	(-0.0602, -0.0218)	**0.0498**	0.0031
		α_2	**-0.0880**	(-0.1057, -0.0704)	**0.0483**	0.0028
		α_3	**-0.1831**	(-0.2000, -0.1662)	**0.0707**	0.0036
	MULTIPLE	β	**-1.536**	(-1.941, -1.130)	**23.75**	1.51
		α_1	**-0.0696**	(-0.0881, -0.0511)	**0.0493**	0.0031
		α_2	**-0.1178**	(-0.1353, -0.1004)	**0.0535**	0.0029
		α_3	**-0.2207**	(-0.2372, -0.2042)	**0.0841**	0.0040
50	LAST	β	**-2.660**	(-3.065, -2.255)	**28.36**	1.87
		α_1	**-0.1500**	(-0.1650, -0.1349)	**0.0519**	0.0029
		α_2	**-0.2948**	(-0.3099, -0.2798)	**0.1165**	0.0046
		α_3	**-0.5703**	(-0.5843, -0.5564)	**0.3506**	0.0080
	MULTIPLE	β	**-5.087**	(-5.474, -4.701)	**45.25**	2.34
		α_1	**-0.2066**	(-0.2217, -0.1915)	**0.0724**	0.0037
		α_2	**-0.3264**	(-0.3398, -0.3130)	**0.1299**	0.0045
		α_3	**-0.6505**	(-0.6634, -0.6375)	**0.4448**	0.0084

Table 4.31: Performance of two imputation methods under 'adapted Finkelstein' with a Weibull mortality process with shape parameter 1.15 and **screen-independent** event collection with two **uncorrelated** repeated measurements (one continuous with $\sigma_W = 1.5$ and $\sigma_B = 8.0$ and one ordinal with the revised transition matrix) and various percentages of jointly missing data generated via a **MCAR** mechanism.

Percent missing	Imputation method	Parameter	BIAS (β: units 10^{-3}) Estimate	95% C.I.	MSE (β: units 10^{-6}) Estimate	S.E.
0	**N/A**	β	**0.222**	(-0.180, 0.624)	**21.04**	1.36
		α_1	**0.0043**	(-0.0170, 0.0255)	**0.0586**	0.0038
		α_2	**-0.0034**	(-0.0231, 0.0163)	**0.0506**	0.0034
		α_3	**0.0030**	(-0.0155, 0.0216)	**0.0447**	0.0029
5	**LAST**	β	**0.468**	(0.062, 0.875)	**21.67**	1.40
		α_1	**-0.0038**	(-0.0246, 0.0170)	**0.0562**	0.0037
		α_2	**-0.0151**	(-0.0343, 0.0042)	**0.0484**	0.0032
		α_3	**-0.0273**	(-0.0455, -0.0090)	**0.0440**	0.0028
	MULTIPLE	β	**-0.062**	(-0.472, 0.349)	**21.86**	1.34
		α_1	**0.0051**	(-0.0155, 0.0258)	**0.0552**	0.0035
		α_2	**-0.0101**	(-0.0300, 0.0098)	**0.0516**	0.0033
		α_3	**-0.0371**	(-0.0552, -0.0190)	**0.0439**	0.0029
10	**LAST**	β	**0.720**	(0.313, 1.128)	**22.10**	1.44
		α_1	**-0.0091**	(-0.0296, 0.0115)	**0.0550**	0.0035
		α_2	**-0.0224**	(-0.0413, -0.0035)	**0.0468**	0.0030
		α_3	**-0.0552**	(-0.0729, -0.0374)	**0.0439**	0.0027
	MULTIPLE	β	**-0.547**	(-0.942, -0.153)	**20.52**	1.25
		α_1	**-0.0190**	(-0.0384, 0.0005)	**0.0496**	0.0032
		α_2	**-0.0483**	(-0.0666, -0.0301)	**0.0456**	0.0029
		α_3	**-0.1108**	(-0.1279, -0.0936)	**0.0505**	0.0026
20	**LAST**	β	**1.173**	(0.763, 1.583)	**23.18**	1.47
		α_1	**-0.0333**	(-0.0527, -0.0139)	**0.0501**	0.0031
		α_2	**-0.0539**	(-0.0718, -0.0361)	**0.0443**	0.0027
		α_3	**-0.1302**	(-0.1472, -0.1131)	**0.0548**	0.0031
	MULTIPLE	β	**-1.237**	(-1.649, -0.826)	**23.53**	1.49
		α_1	**-0.0455**	(-0.0646, -0.0263)	**0.0498**	0.0031
		α_2	**-0.1083**	(-0.1262, -0.0904)	**0.0535**	0.0030
		α_3	**-0.2315**	(-0.2486, -0.2144)	**0.0915**	0.0043
50	**LAST**	β	**1.894**	(1.469, 2.319)	**27.06**	1.69
		α_1	**-0.1270**	(-0.1429, -0.1112)	**0.0487**	0.0028
		α_2	**-0.1912**	(-0.2070, -0.1754)	**0.0689**	0.0034
		α_3	**-0.4111**	(-0.4259, -0.3964)	**0.1972**	0.0062
	MULTIPLE	β	**-4.917**	(-5.322, -4.512)	**45.48**	2.45
		α_1	**-0.1206**	(-0.1382, -0.1030)	**0.0548**	0.0033
		α_2	**-0.3509**	(-0.3664, -0.3354)	**0.1543**	0.0059
		α_3	**-0.7590**	(-0.7753, -0.7427)	**0.6107**	0.0146

Table 4.32: Performance of two imputation methods under 'adapted Finkelstein' with a Weibull mortality process with shape parameter 1.15 and **screen-dependent** event collection with two **uncorrelated** repeated measurements (one continuous with $\sigma_W = 1.5$ and $\sigma_B = 8.0$ and one ordinal with the revised transition matrix) and various percentages of jointly missing data generated via a **MCAR** mechanism.

Percent missing	Imputation method	Parameter	BIAS (β: units 10^{-3}) Estimate	95% C.I.	MSE (β: units 10^{-6}) Estimate	S.E.
0	**N/A**	β	**0.295**	(-0.102, 0.692)	**20.54**	1.43
		α_1	**0.0069**	(-0.0144, 0.0282)	**0.0588**	0.0036
		α_2	**0.0027**	(-0.0173, 0.0226)	**0.0517**	0.0034
		α_3	**0.0079**	(-0.0112, 0.0269)	**0.0474**	0.0032
5	**LAST**	β	**0.215**	(-0.181, 0.612)	**20.49**	1.43
		α_1	**-0.0012**	(-0.0224, 0.0199)	**0.0581**	0.0035
		α_2	**-0.0145**	(-0.0342, 0.0052)	**0.0508**	0.0032
		α_3	**-0.0329**	(-0.0518, -0.0140)	**0.0475**	0.0029
	MULTIPLE	β	**0.325**	(-0.106, 0.756)	**24.22**	1.74
		α_1	**0.0004**	(-0.0212, 0.0220)	**0.0606**	0.0042
		α_2	**-0.0137**	(-0.0331, 0.0057)	**0.0491**	0.0032
		α_3	**-0.0369**	(-0.0557, -0.0180)	**0.0474**	0.0033
10	**LAST**	β	**0.111**	(-0.285, 0.508)	**20.45**	1.43
		α_1	**-0.0063**	(-0.0272, 0.0146)	**0.0570**	0.0033
		α_2	**-0.0286**	(-0.0482, -0.0089)	**0.0509**	0.0030
		α_3	**-0.0723**	(-0.0910, -0.0535)	**0.0511**	0.0029
	MULTIPLE	β	**0.353**	(-0.058, 0.764)	**22.06**	1.58
		α_1	**-0.0404**	(-0.0595, -0.0213)	**0.0490**	0.0029
		α_2	**-0.0651**	(-0.0831 -0.0471)	**0.0465**	0.0028
		α_3	**-0.1133**	(-0.1303, -0.0964)	**0.0501**	0.0030
20	**LAST**	β	**-0.156**	(-0.553, 0.240)	**20.43**	1.41
		α_1	**-0.0333**	(-0.0533, -0.0132)	**0.0534**	0.0033
		α_2	**-0.0766**	(-0.0958, -0.0574)	**0.0538**	0.0030
		α_3	**-0.1717**	(-0.1899, -0.1535)	**0.0723**	0.0037
	MULTIPLE	β	**-0.148**	(-0.559, 0.263)	**21.97**	1.70
		α_1	**-0.0844**	(-0.1038, -0.0650)	**0.0560**	0.0033
		α_2	**-0.1218**	(-0.1401, -0.1034)	**0.0587**	0.0032
		α_3	**-0.2290**	(-0.2464, -0.2115)	**0.0919**	0.0041
50	**LAST**	β	**-1.834**	(-2.232, -1.436)	**23.92**	1.55
		α_1	**-0.1371**	(-0.1528, -0.1215)	**0.0508**	0.0029
		α_2	**-0.2741**	(-0.2893, -0.2589)	**0.1052**	0.0044
		α_3	**-0.5447**	(-0.5593, -0.5301)	**0.3245**	0.0080
	MULTIPLE	β	**-2.764**	(-3.150, -2.377)	**27.02**	1.58
		α_1	**-0.1968**	(-0.2119, -0.1817)	**0.0683**	0.0036
		α_2	**-0.3182**	(-0.3322, -0.3041)	**0.1267**	0.0048
		α_3	**-0.6277**	(-0.6419, -0.6135)	**0.4202**	0.0091

Table 4.33: Performance of two imputation methods under 'adapted Finkelstein' with a Weibull mortality process with shape parameter 1.15 and **screen-independent** event collection with two **correlated** repeated measurements with inital value rank correlation $\simeq 0.3$ (one continuous with $\sigma_W = 1.5$ and $\sigma_B = 8.0$ and one ordinal with the revised transition matrix) and various percentages of jointly missing data generated via a **MCAR** mechanism.

Percent missing	Imputation method	Parameter	BIAS (β: units 10^{-3}) Estimate	95% C.I.	MSE (β: units 10^{-6}) Estimate	S.E.
0	**N/A**	β	**0.295**	(-0.102, 0.692)	**20.54**	1.43
		α_1	**0.0069**	(-0.0144, 0.0282)	**0.0588**	0.0036
		α_2	**0.0027**	(-0.0173, 0.0226)	**0.0517**	0.0034
		α_3	**0.0079**	(-0.0112, 0.0269)	**0.0474**	0.0032
5	**LAST**	β	**0.483**	(0.084, 0.882)	**20.90**	1.45
		α_1	**0.0004**	(-0.0208, 0.0216)	**0.0583**	0.0035
		α_2	**-0.0081**	(-0.0278, 0.0117)	**0.0509**	0.0032
		α_3	**-0.0230**	(-0.0419, -0.0040)	**0.0472**	0.0029
	MULTIPLE	β	**0.346**	(-0.087, 0.778)	**24.43**	1.75
		α_1	**0.0127**	(-0.0091, 0.0346)	**0.0623**	0.0043
		α_2	**-0.0029**	(-0.0225, 0.0167)	**0.0499**	0.0033
		α_3	**-0.0311**	(-0.0502, -0.0120)	**0.0484**	0.0033
10	**LAST**	β	**0.676**	(0.276, 1.077)	**21.28**	1.47
		α_1	**-0.0034**	(-0.0244, 0.0177)	**0.0575**	0.0034
		α_2	**-0.0155**	(-0.0352, 0.0043)	**0.0509**	0.0031
		α_3	**-0.0520**	(-0.0708, -0.0331)	**0.0489**	0.0028
	MULTIPLE	β	**0.387**	(-0.025, 0.800)	**22.29**	1.60
		α_1	**-0.0169**	(-0.0366, 0.0029)	**0.0509**	0.0029
		α_2	**-0.0462**	(-0.0649, -0.0274)	**0.0477**	0.0028
		α_3	**-0.1058**	(-0.1232, -0.0883)	**0.0507**	0.0030
20	**LAST**	β	**1.012**	(0.605, 1.420)	**22.55**	1.55
		α_1	**-0.0277**	(-0.0480, -0.0075)	**0.0540**	0.0034
		α_2	**-0.0496**	(-0.0690, -0.0301)	**0.0515**	0.0030
		α_3	**-0.1293**	(-0.1476, -0.1110)	**0.0602**	0.0032
	MULTIPLE	β	**-0.019**	(-0.438, 0.399)	**22.73**	1.80
		α_1	**-0.0391**	(-0.0595, -0.0188)	**0.0553**	0.0032
		α_2	**-0.0964**	(-0.1158, -0.0770)	**0.0581**	0.0032
		α_3	**-0.2235**	(-0.2419, -0.2052)	**0.0938**	0.0044
50	**LAST**	β	**1.197**	(0.775, 1.619)	**24.57**	1.67
		α_1	**-0.1212**	(-0.1374, -0.1049)	**0.0489**	0.0028
		α_2	**-0.1902**	(-0.2058, -0.1746)	**0.0679**	0.0034
		α_3	**-0.4171**	(-0.4321, -0.4020)	**0.2033**	0.0064
	MULTIPLE	β	**-2.801**	(-3.204, -2.397)	**28.97**	1.64
		α_1	**-0.1308**	(-0.1503, -0.1113)	**0.0665**	0.0039
		α_2	**-0.3496**	(-0.3670, -0.3323)	**0.1614**	0.0068
		α_3	**-0.7426**	(-0.7590, -0.7261)	**0.5866**	0.0125

Table 4.34: Performance of two imputation methods under 'adapted Finkelstein' with a Weibull mortality process with shape parameter 1.15 and **screen-dependent** event collection with two **correlated** repeated measurements with inital value rank correlation $\simeq 0.3$ (one continuous with $\sigma_W = 1.5$ and $\sigma_B = 8.0$ and one ordinal with the revised transition matrix) and various percentages of jointly missing data generated via a **MCAR** mechanism.

Percent missing	Imputation method	Parameter	BIAS (β: units 10^{-3}) Estimate	95% C.I.	MSE (β: units 10^{-6}) Estimate	S.E.
0	**N/A**	β	**0.222**	(-0.180, 0.624)	**21.04**	1.36
		α_1	**0.0043**	(-0.0170, 0.0255)	**0.0586**	0.0038
		α_2	**-0.0034**	(-0.0231, 0.0163)	**0.0506**	0.0034
		α_3	**0.0030**	(-0.0155, 0.0216)	**0.0447**	0.0029
5	**LAST**	β	**0.522**	(0.115, 0.929)	**21.76**	1.39
		α_1	**0.0121**	(-0.0091, 0.0333)	**0.0587**	0.0039
		α_2	**0.0063**	(-0.0134, 0.0260)	**0.0503**	0.0034
		α_3	**-0.0207**	(-0.0393, -0.0020)	**0.0455**	0.0029
	MULTIPLE	β	**-0.291**	(-0.685, 0.103)	**20.23**	1.28
		α_1	**0.0061**	(-0.0137, 0.0260)	**0.0511**	0.0035
		α_2	**-0.0096**	(-0.0284, 0.0092)	**0.0460**	0.0033
		α_3	**-0.0475**	(-0.0653, -0.0298)	**0.0433**	0.0028
10	**LAST**	β	**0.749**	(0.341, 1.158)	**22.24**	1.44
		α_1	**0.0132**	(-0.0077, 0.0341)	**0.0570**	0.0037
		α_2	**0.0103**	(-0.0090, 0.0297)	**0.0488**	0.0033
		α_3	**-0.0496**	(-0.0679, -0.0312)	**0.0463**	0.0029
	MULTIPLE	β	**-1.034**	(-1.429, -0.638)	**21.35**	1.45
		α_1	**-0.0203**	(-0.0400, -0.0005)	**0.0509**	0.0032
		α_2	**-0.0337**	(-0.0513, -0.0160)	**0.0414**	0.0026
		α_3	**-0.1174**	(-0.1349, -0.0999)	**0.0536**	0.0029
20	**LAST**	β	**1.068**	(0.658, 1.477)	**22.93**	1.51
		α_1	**0.0191**	(-0.0011, 0.0394)	**0.0537**	0.0035
		α_2	**0.0159**	(-0.0028, 0.0347)	**0.0460**	0.0030
		α_3	**-0.1085**	(-0.1264, -0.0907)	**0.0531**	0.0032
	MULTIPLE	β	**-2.751**	(-3.144, -2.358)	**27.65**	1.90
		α_1	**-0.0587**	(-0.0771, -0.0404)	**0.0473**	0.0030
		α_2	**-0.0920**	(-0.1087, -0.0753)	**0.0447**	0.0028
		α_3	**-0.2532**	(-0.2693, -0.2370)	**0.0981**	0.0045
50	**LAST**	β	**1.127**	(0.708, 1.545)	**24.06**	1.51
		α_1	**-0.0018**	(-0.0190, 0.0155)	**0.0386**	0.0028
		α_2	**-0.0598**	(-0.0763, -0.0433)	**0.0390**	0.0024
		α_3	**-0.3729**	(-0.3890, -0.3568)	**0.1727**	0.0061
	MULTIPLE	β	**-6.370**	(-6.743, -5.996)	**58.70**	2.80
		α_1	**-0.1790**	(-0.1957, -0.1622)	**0.0685**	0.0035
		α_2	**-0.2723**	(-0.2874, -0.2571)	**0.1040**	0.0042
		α_3	**-0.6402**	(-0.6549, -0.6256)	**0.4379**	0.0091

Table 4.35: Performance of two imputation methods under 'adapted Finkelstein' with a Weibull mortality process with shape parameter 1.15 and **screen-independent** event collection with two **uncorrelated** repeated measurements (one continuous with $\sigma_W = 1.5$ and $\sigma_B = 8.0$ and one ordinal with the revised transition matrix) and various percentages of jointly missing data generated via a **MAR** mechanism.

Percent missing	Imputation method	Parameter	BIAS (β: units 10^{-3}) Estimate	95% C.I.	MSE (β: units 10^{-6}) Estimate	S.E.
0	**N/A**	β	**0.222**	(-0.180, 0.624)	**21.04**	1.36
		α_1	**0.0043**	(-0.0170, 0.0255)	**0.0586**	0.0038
		α_2	**-0.0034**	(-0.0231, 0.0163)	**0.0506**	0.0034
		α_3	**0.0030**	(-0.0155, 0.0216)	**0.0447**	0.0029
5	**LAST**	β	**2.235**	(1.824, 2.645)	**26.89**	1.64
		α_1	**0.0166**	(-0.0046, 0.0379)	**0.0590**	0.0040
		α_2	**0.0234**	(0.0037, 0.0431)	**0.0511**	0.0035
		α_3	**0.0131**	(-0.0056, 0.0318)	**0.0456**	0.0030
	MULTIPLE	β	**0.986**	(0.591, 1.380)	**21.18**	1.33
		α_1	**0.0423**	(0.0220, 0.0626)	**0.0555**	0.0037
		α_2	**0.0455**	(0.0261, 0.0649)	**0.0508**	0.0037
		α_3	**0.0129**	(-0.0054, 0.0313)	**0.0438**	0.0030
10	**LAST**	β	**4.054**	(3.635, 4.472)	**39.20**	2.18
		α_1	**0.0225**	(0.0015, 0.0435)	**0.0578**	0.0038
		α_2	**0.0450**	(0.0256, 0.0645)	**0.0513**	0.0035
		α_3	**0.0182**	(-0.0003, 0.0367)	**0.0448**	0.0030
	MULTIPLE	β	**1.313**	(0.910, 1.717)	**22.89**	1.54
		α_1	**0.0426**	(0.0216, 0.0636)	**0.0592**	0.0039
		α_2	**0.0676**	(0.0490, 0.0863)	**0.0498**	0.0036
		α_3	**-0.0095**	(-0.0281, 0.0091)	**0.0451**	0.0027
20	**LAST**	β	**7.413**	(6.984, 7.842)	**78.82**	3.47
		α_1	**0.0393**	(0.0189, 0.0597)	**0.0556**	0.0037
		α_2	**0.0900**	(0.0709, 0.1090)	**0.0552**	0.0037
		α_3	**0.0304**	(0.0123, 0.0485)	**0.0434**	0.0029
	MULTIPLE	β	**1.398**	(0.989, 1.807)	**23.64**	1.48
		α_1	**0.0691**	(0.0486, 0.0896)	**0.0592**	0.0042
		α_2	**0.0977**	(0.0788, 0.1165)	**0.0558**	0.0038
		α_3	**-0.0554**	(-0.0735, -0.0373)	**0.0457**	0.0029
50	**LAST**	β	**15.868**	(15.395, 16.341)	**280.83**	7.76
		α_1	**0.0718**	(0.0538, 0.0897)	**0.0468**	0.0034
		α_2	**0.1716**	(0.1543, 0.1889)	**0.0683**	0.0043
		α_3	**0.0219**	(0.0051, 0.0387)	**0.0371**	0.0028
	MULTIPLE	β	**1.558**	(1.149, 1.968)	**24.21**	1.44
		α_1	**0.1396**	(0.1204, 0.1589)	**0.0675**	0.0042
		α_2	**0.1109**	(0.0938, 0.1281)	**0.0504**	0.0031
		α_3	**-0.3081**	(-0.3284, -0.2877)	**0.1486**	0.0109

Table 4.36: Performance of two imputation methods under 'adapted Finkelstein' with a Weibull mortality process with shape parameter 1.15 and **screen-dependent** event collection with two **uncorrelated** repeated measurements (one continuous with $\sigma_W = 1.5$ and $\sigma_B = 8.0$ and one ordinal with the revised transition matrix) and various percentages of jointly missing data generated via a **MAR** mechanism.

Percent missing	Imputation method	Parameter	BIAS (β: units 10^{-3}) Estimate	95% C.I.	MSE (β: units 10^{-6}) Estimate	S.E.
0	**N/A**	β	**0.222**	(-0.180, 0.624)	**21.04**	1.36
		α_1	**0.0043**	(-0.0170, 0.0255)	**0.0586**	0.0038
		α_2	**-0.0034**	(-0.0231, 0.0163)	**0.0506**	0.0034
		α_3	**0.0030**	(-0.0155, 0.0216)	**0.0447**	0.0029
5	**LAST**	β	**0.820**	(0.416, 1.224)	**21.84**	1.40
		α_1	**-0.0017**	(-0.0219, 0.0185)	**0.0528**	0.0034
		α_2	**-0.0192**	(-0.0379, -0.0004)	**0.0461**	0.0029
		α_3	**-0.0675**	(-0.0852, -0.0498)	**0.0453**	0.0028
	MULTIPLE	β	**0.002**	(-0.389, 0.393)	**19.84**	1.22
		α_1	**-0.0044**	(-0.0244, 0.0157)	**0.0521**	0.0047
		α_2	**-0.0310**	(-0.0496 -0.0123)	**0.0460**	0.0036
		α_3	**-0.0817**	(-0.0995, -0.0639)	**0.0477**	0.0033
10	**LAST**	β	**1.171**	(0.766, 1.575)	**22.65**	1.47
		α_1	**-0.0153**	(-0.0345, 0.0039)	**0.0481**	0.0033
		α_2	**-0.0448**	(-0.0625, -0.0271)	**0.0426**	0.0027
		α_3	**-0.1428**	(-0.1596, -0.1261)	**0.0567**	0.0034
	MULTIPLE	β	**-0.388**	(-0.775, 0.000)	**19.62**	1.16
		α_1	**-0.0529**	(-0.0709, -0.0349)	**0.0448**	0.0026
		α_2	**-0.0901**	(-0.1068, -0.0734)	**0.0443**	0.0025
		α_3	**-0.1945**	(-0.2101, -0.1789)	**0.0694**	0.0034
20	**LAST**	β	**1.448**	(1.037, 1.858)	**23.96**	1.53
		α_1	**-0.0375**	(-0.0548, -0.0202)	**0.0403**	0.0027
		α_2	**-0.1006**	(-0.1172, -0.0840)	**0.0459**	0.0027
		α_3	**-0.2838**	(-0.2996, -0.2679)	**0.1131**	0.0048
	MULTIPLE	β	**-1.747**	(-2.142, -1.351)	**23.38**	1.50
		α_1	**-0.1231**	(-0.1392, -0.1071)	**0.0488**	0.0027
		α_2	**-0.1928**	(-0.2074, -0.1782)	**0.0649**	0.0031
		α_3	**-0.4000**	(-0.4145, -0.3855)	**0.1873**	0.0059
50	**LAST**	β	**-0.605**	(-1.011, -0.198)	**21.87**	1.38
		α_1	**-0.1364**	(-0.1491, -0.1236)	**0.0396**	0.0021
		α_2	**-0.3170**	(-0.3301, -0.3040)	**0.1225**	0.0043
		α_3	**-0.7159**	(-0.7285, -0.7033)	**0.5330**	0.0092
	MULTIPLE	β	**-5.882**	(-6.248, -5.516)	**51.97**	2.59
		α_1	**-0.2576**	(-0.2715, -0.2437)	**0.0915**	0.0039
		α_2	**-0.4367**	(-0.4494, -0.4240)	**0.2117**	0.0056
		α_3	**-0.8682**	(-0.8801, -0.8564)	**0.7722**	0.0104

Table 4.37: Performance of two imputation methods under 'adapted Finkelstein' with a Weibull mortality process with shape parameter 1.15 and **screen-independent** event collection with two **uncorrelated** repeated measurements (one continuous with $\sigma_W = 1.5$ and $\sigma_B = 8.0$ and one ordinal with the revised transition matrix) and various percentages of jointly missing data generated via a **value-based non-ignorable** missing data mechanism.

Percent missing	Imputation method	Parameter	BIAS (β: units 10^{-3}) Estimate	95% C.I.	MSE (β: units 10^{-6}) Estimate	S.E.
0	**N/A**	β	**0.222**	(-0.180, 0.624)	**21.04**	1.36
		α_1	**0.0043**	(-0.0170, 0.0255)	**0.0586**	0.0038
		α_2	**-0.0034**	(-0.0231, 0.0163)	**0.0506**	0.0034
		α_3	**0.0030**	(-0.0155, 0.0216)	**0.0447**	0.0029
5	**LAST**	β	**2.748**	(2.337, 3.158)	**29.44**	1.75
		α_1	**0.0037**	(-0.0165, 0.0239)	**0.0530**	0.0034
		α_2	**0.0009**	(-0.0179, 0.0197)	**0.0460**	0.0030
		α_3	**-0.0334**	(-0.0512, -0.0156)	**0.0422**	0.0026
	MULTIPLE	β	**1.383**	(0.988, 1.778)	**22.18**	1.39
		α_1	**0.0292**	(0.0085, 0.0500)	**0.0569**	0.0051
		α_2	**0.0167**	(-0.0027, 0.0360)	**0.0489**	0.0040
		α_3	**-0.0392**	(-0.0575, -0.0208)	**0.0453**	0.0035
10	**LAST**	β	**4.876**	(4.460, 5.291)	**46.18**	2.46
		α_1	**-0.0040**	(-0.0233, 0.0154)	**0.0487**	0.0033
		α_2	**-0.0043**	(-0.0222, 0.0135)	**0.0414**	0.0026
		α_3	**-0.0755**	(-0.0924, -0.0587)	**0.0427**	0.0027
	MULTIPLE	β	**2.212**	(1.814, 2.611)	**25.55**	1.53
		α_1	**0.0138**	(-0.0058, 0.0335)	**0.0504**	0.0031
		α_2	**-0.0036**	(-0.0218, 0.0146)	**0.0430**	0.0027
		α_3	**-0.1163**	(-0.1331, -0.0995)	**0.0503**	0.0027
20	**LAST**	β	**8.436**	(8.005, 8.868)	**95.40**	3.89
		α_1	**-0.0124**	(-0.0300, 0.0053)	**0.0407**	0.0028
		α_2	**-0.0202**	(-0.0372, -0.0031)	**0.0382**	0.0024
		α_3	**-0.1537**	(-0.1699, -0.1374)	**0.0579**	0.0031
	MULTIPLE	β	**2.966**	(2.556, 3.376)	**30.65**	1.95
		α_1	**-0.0017**	(-0.0196, 0.0161)	**0.0414**	0.0026
		α_2	**-0.0451**	(-0.0610, -0.0292)	**0.0349**	0.0021
		α_3	**-0.2674**	(-0.2831, -0.2516)	**0.1036**	0.0044
50	**LAST**	β	**14.522**	(14.067, 14.977)	**237.80**	6.95
		α_1	**-0.0753**	(-0.0888, -0.0618)	**0.0294**	0.0017
		α_2	**-0.1299**	(-0.1438, -0.1160)	**0.0421**	0.0024
		α_3	**-0.4265**	(-0.4400, -0.4131)	**0.2054**	0.0060
	MULTIPLE	β	**3.659**	(3.248, 4.069)	**35.28**	2.11
		α_1	**-0.0318**	(-0.0483, -0.0152)	**0.0366**	0.0024
		α_2	**-0.2198**	(-0.2349, -0.2048)	**0.0778**	0.0036
		α_3	**-0.6832**	(-0.6967, -0.6697)	**0.4904**	0.0093

Table 4.38: Performance of two imputation methods under 'adapted Finkelstein' with a Weibull mortality process with shape parameter 1.15 and **screen-dependent** event collection with two **uncorrelated** repeated measurements (one continuous with $\sigma_W = 1.5$ and $\sigma_B = 8.0$ and one ordinal with the revised transition matrix) and various percentages of jointly missing data generated via a **value-based non-ignorable** missing data mechanism.

4.6 Summary

In the analysis of interval-censored survival data in the presence of a risk factor history, a variety of methods is available. The method chosen will depend on the aims of a particular investigation, the likely means by which the risk factor is related to the event of interest and whether there are missing risk factor values. If there *are* some missing values, two other aspects of the missing data should be considered. Firstly, the percentage of the data missing will affect the choice of technique. If a low percentage of the data is missing, say 10%, then alternative techniques will tend to lead to very similar conclusions, as illustrated in Sections 4.4 and 4.5. However, as also shown in Sections 4.4 and 4.5, if a larger percentage of data is missing, different imputation techniques are likely to lead to substantially different conclusions and the validity of each technique will depend on the type and form of missing data mechanism.

When designing a study, it is important to consider the likely mechanism by which potential risk factors will affect the occurrence of the event of interest. If it is the present or a recent risk factor value which is of major importance, updated values of this risk factor should be part of the longitudinal design of the study (Altman and De Stavola (1994), Cupples et al. (1988)). This will allow investigation of the relationship between the risk factor profile and survival. In situations when 'survival' is related solely to a recent value of a risk factor, the use of the baseline risk factor measurement alone tends to lead to under-estimation of the magnitude of the true effect of the risk factor, as illustrated in Section 4.3 for Prentice and Gloeckler's model and also in Altman and De Stavola (1994), although this picture may be reversed in certain circumstances. This was observed when the within-individual variability was low compared with the between-individual variability. In this case, the unknown value could be so well-approximated by a linear function of the baseline covariate value that the parameter was over-estimated as

the last measurement tends to be relatively close to a multiple, greater than one, of the baseline value.

As described in detail in earlier chapters, when risk factor values are missing it is important to determine the mechanism which has caused their missingness. If the data are not missing completely at random or, more particularly, if it is believed that the examination scheme might be informative, it is desirable to determine the form of the mechanism causing the missingness and to use any relevant information when modelling the survival process (e.g. Grüger et al., 1991). If however, as is often the case, little is known or suspected about the form of the missing data process, sensitivity analysis should be used to investigate whether the conclusions would differ under a range of plausible forms for this process (Rubin, 1987). The results of the Monte Carlo simulations presented in Sections 4.4 and 4.5 indicate that this will be particularly important when a moderate percentage, from between 10% and 20% upwards, of the data is missing.

One general strategy, applicable when modelling data subject to a missing data mechanism, is that of imputation of missing risk factor values. The main advantage of imputation is that standard modelling techniques can be applied once the data set has been completed. One disadvantage is that imputed data are then treated as observed and can thus lead to under-estimation of standard errors. This particular disadvantage can be overcome by applying multiple imputations via an appropriate imputation strategy. In this chapter, two imputation strategies have been investigated, one applicable for continuous risk factors and the other applicable for ordinal risk factors. In each case a relatively simple approach has been used, imputing based on the previous value of that particular risk factor rather than on the profile of that and, potentially, other risk factors. This is a simple method and it was seen that, whilst it performed satisfactorily when there was a single continous risk factor as in Section 4.4, its performance for most missing data situations with moderate or large proportions missing was poor both in

terms of bias and mean square error when there was a single ordinal risk factor in addition to a single continuous one in the simulations performed in Section 4.5. This is thought to be due largely to the complexity of the ordinal response model in terms of the number of model parameters which leads to relatively imprecise estimation and hence quite a degree of variability between the imputations. This will, in turn, lead to some imputed values which would be deemed highly unlikely if the later risk factor values for that individual were considered. Moreover, the problems in estimating the parameters relating to the effect(s) of the ordinal risk factor(s) (the α_s in Section 4.5) will have a knock-on effect in the parameter estimation relating to the continuous risk factor (β in Section 4.5).

It is therefore recommended that, unless the percentage of data missing is low (under around 10%) a more sophisticated imputation stategy should be applied. Various approaches are possible, including application of a strategy which takes into account the true longitudinal nature of the data. A simple approach would be to delete observations following any missing ones in order to obtain a monotone missing data structure prior to imputing for any missing or deleted observations (Rubin, 1987). Whilst this approach might lead to a reduction in the degree of bias, it will certainly lead to a decrease in precision. If there is a low number of scheduled screens and a low proportion of data missing, such an approach might lead to an improved estimation. However, in other circumstances, any potential gain in respect of lowering the bias will be outweighed by the consequential increase in standard errors leading to an overall increase in mean square error.

Chapter 5

Modelling data arising from the annual screening of the elderly of Melton Mowbray

In modelling mortality and other changes of state arising from screening studies or in the more general topic of event history modelling, a number of problems are typically encountered. The two main problems surround the irregular and incomplete assessment of risk factor history with unknown form(s) of missing data mechanism and the recording of survival or change of state times in interval-censored form. Whilst it is recognised that determination of the risk factor profile and its interaction with the change of state process remains the ideal, the complexity of the data structure will often require the analyst to compromise his approach to enable a realistically implementable estimation procedure to be used. Methods such as those of Wu and Carroll (1988) covered in Chapter 3 and those of Manton and his collaborators discussed in Chapters 2 and 4 are themselves relatively complex without even considering their extension to allow their application to screening studies with missing risk factor values prior to censoring or the occurrence of the event of interest.

Other alternatives, including the methods of Prentice and Gloeckler (1978) and

Finkelstein (1986), are less complex and, whilst not involving direct modelling of the risk factor distribution, afford more natural extensions to account for missing values during the accumulation of an individual's risk factor history. As was shown in Chapter 4, Finkelstein's method can be adapted to include time-dependent covariates (providing they are viewed as fixed within each interval between successive measurements and follow-up) and missing values of the form described above.

In this chapter, this approach which combines the broadening properties of both methods (termed 'adapted Finkelstein') will be applied to screen and mortality data arising from the first three waves of a health screening of the elderly population of the Leicestershire market town of Melton Mowbray. Results from imputing missing values from these screens using the last recorded measurement will be constrasted with those obtained using a model-based multiple imputation approach under both the natural screen-independent recording of deaths, under which 'adapted Finkelstein' is identical to Prentice and Gloeckler's method, and also by viewing the recording as screen-dependent. This latter approach, whilst both artificial and problematical given the design of the study and nature of event under consideration, is included to illustrate 'adapted Finkelstein' in a case in which it differs from Prentice and Gloeckler's method. Additionally, the effects of using the more traditional approaches to the analysis of survival data of complete case (CC) and available case (AC) methods and those of using baseline as opposed to updated risk factor values will be investigated. Finally, the use of interval-censored survival times will be contrasted with the more common form in which the times are sufficiently accurately recorded to be viewed as 'exact'.

5.1 Studies of the elderly population of Melton Mowbray

Over the last fifteen years, a series of population studies of the elderly has been carried out by the Latham House general practice in Melton Mowbray in conjunction with the Department of Epidemiology and Public Health at the University of Leciester. The Latham House general practice is the sole provider of primary care services in Melton Mowbray and presently has 13 full-time and 3 part-time doctors together with a list size of some 33000, around 1900 of whom are over the age of 75. A very advantageous design property for population health studies holds here as a single practice covers a fairly well-defined population of a moderately large size. Cohorts of those over 75 years of age at the start of each of the years 1981 and 1988 were defined from the age-sex register at Latham House and various aspects of the health, mortality and social contacts of the elderly were studied and have been reported elsewhere (Clarke, Clarke, Odell and Jagger (1984), Jagger, Clarke and Davies (1986), Jagger and Clarke (1988), Jagger, Clarke and Cook (1989), Jagger, Clarke and Clarke (1991), Jagger and Sutton (1991), Jagger, Spiers and Clarke (1993)). In 1985 a study of the elderly living alone in the town together with a separate follow-up of those from the 1981 cohort not living alone provided longitudinal data on the 1981 cohort. This enabled some limited investigation of the changing health profile of this cohort and its resultant effect on survival to be performed (Jagger and Clarke (1988), Jagger, Clarke and Cook (1989)). The 1988 cohort provided further longitudinal data for the 1981 cohort and enabled further investigations to be performed (Jagger, Clarke and Clarke, 1991). On the introduction of the revised GP contract in April 1990, the requirement to offer an annual health screen presented an opportunity for a natural extension to the scope of the previous studies, enabling more regular and routine collection of information on the profile of risk factors which could then be used in the modelling of resultant state changes. This thesis, and this chapter in particular, concentrates

on these health screens and the modelling of survival data based on the set of potential risk factors collected at these screens.

The Latham House practice decided that the annual health screen should be carried out at the elderly person's residence. A double-sided A4 card was designed, following a pilot, to record the information collected at each screen and a single card was to be sufficient for five screens. The region covered by the practice was divided into twelve areas, the aim being to screen one area during each month of the year. Letters were then sent to all people on the practice's age-sex register who were aged over 75 years and recorded as being resident in that area, giving them details of the screen offered and informing them that a nurse would be in their neighbourhood in the near future to carry out screens. A tear-off slip and pre-paid envelope were provided for those who wished to formally refuse the offer of a screen. The date of posting this letter was recorded, is termed the *post date* and will be used in the analysis when appropriate. A minor problem with this strategy is that a small number of people will be missed at each wave of screens due to moving residence from an area of Melton in which screening has not yet been offered to an area in which screening has been completed. Although the aim was to have annual screens, practical considerations led to the first three screens covering a total period of five-and-a-half years. The first wave of screens commenced on August 6^{th} 1990, the second wave commenced on February 10^{th} 1992 and third wave began on November 18^{th} 1993. Screening of those in the third wave continued until February 19^{th} 1996.

5.2 The data set

The base data set comprised 72 data fields and 2203 observations, covering the individuals screened in at least one out of the first three screens. In terms of modelling the survival of the elderly, 33 of the 72 fields comprised assessment of 11 potential risk factors at each of the three screens. Further information on these 11 variables can be found in Table 5.1 (page 202) and Table 5.2 (page 209). The discrete 'score' variables TOTADLn and IOSCOREn (n = 1, 2, 3) were recoded as ordinal variables as follows. Those with 0 or 1 ADLs carried out independently were classified as being 'highly dependent' (and coded as '4'), those with 2, 3 or 4 independent ADLs were classified as being 'moderately dependent' (and coded as '3'), those with 5 or 6 independent ADLs were classified as being 'moderately independent' (and coded as '2') and those with 7 independent ADLs were classified as being 'fully independent' (and coded as '1'). The ordinal ADL variables thus created will be called ADLn (n = 1, 2, 3). In previous community-based studies (Clarke, Jagger, Anderson, Battcock, Kelly and Stern, 1991) it has been found that a score of 8 or less on the Information-Orientation subtest score of the CAPE assessment (Pattie and Gilleard, 1979) has the greatest sensitivity and specificity to detect moderate or severe dementia and so the IOSCOREn variables were recoded according to this criterion. The dementia indicator thus created will be called IOn (n = 1, 2, 3).

The remaining 39 fields related to dates of posting (10), dates of screening (9), date of death (3), ages at posting (3), various indicators used in merging subsidiary data sets (10), the individual's age-sex number at the practice (1) and screening indicators (3). The screening indicators showed whether an individual was screened (1), actively or passively refused (2), had died (3) or was missing (.). All these variables, with the exception of the age-sex number and the indicators used in merging, are used in the modelling in the later sections of this chapter.

A number of problems were apparent on inspecting the data. Most of these were easily rectified and are not relevant to this thesis. Four individuals screened in the first wave and one individual each for the second and third waves did not have their screen dates recorded. For the vast majority of subjects the screen date was very close to the post date. For those screened but whose screen date was not recorded, their screen date was imputed as the post date plus the median number of days between the post and screen dates for those screened in the appropriate wave. The median number of days between these dates were 26, 15 and 14 days respectively for the three waves.

Choice of time scale and definition of survival time

For the interval-censored approach to modelling survival (and similar changes of state) the obvious approach was to choose 'time in study' as the time scale. The only problem then remaining was to consider how 'entry into study' should be defined. The two possibilites were the first offered screen and the first completed screen. Due to the repeated measures nature of the covariate information and the corresponding methods discussed and investigated in Chapter 4, it was deemed appropriate to use the first completed screen as entry into the study. This approach can be viewed as shifting the problem of missing risk factor measurements due to left-truncation to a problem of missing risk factor measurements due to right-censoring. Whilst the precise nature and form of the corresponding missing data mechanism is not known, it may not be ignorable for the reasons described later in this section. The same form of time scale was chosen for the 'exact' survival times. In this case, survival time was defined to be the time between first completed screen and death, post date of the next missed screen or, for removers or those screened in wave 3, their final screen date.

Variable Name	Description	Units (continuous) or Range (discrete scores) or Codes & categories (ordinal)
AGESCn	Age at screening	Years (truncated)
DIASn	Diastolic blood pressure	millimetres Hg
DSPANn	Demi-span	centimetres
HHCOMPn	Household composition	**1** not alone; **2** alone; **3** institution.
IOSCOREn	Information/Orientation subtest score	**0-12**
SEXn	Sex	**1** male; **2** female.
SHn	Perceived health status	**1** good; **2** fair; **3** poor.
SMOKINGn	Smoking status	**1** never; **2** ex; **3** current.
SYSn	Systolic blood pressure	millimetres Hg
TOTADLn	Total number of activities of daily living independent and without difficulty	**0-7**
WTn	Weight	pounds

Table 5.1: List and description of potential risk factors for survival recorded at each screen ($n = 1, 2, 3$).

One additional problem persisted in the definition of survival time for some individuals. Those who were only *offered* one screen or only attended the last screen offered to them were censored having contributed no information relating to survival as they were not followed-up after they were entered into the study. These individuals were deleted from the data set prior to the commencement of modelling. There were 490 such individuals leaving 1713 others to make some potential contribution to the relevant likelihoods.

Amount and forms of missing data

The data from the Melton Mowbray study are subject to both item and unit screen non-response. True unit screen non-response is due to the failure to take up the offer of a screen. Jagger, Clarke, O'Shea and Gannon (1996) reported on the characteristics of those who failed to take up this offer in the first wave. It was concluded that those who missed being screened and yet had been included in the 1988 cohort were similar with regard to sex, age and their continence, disability, vision and hearing in 1988 to those who were screened. There were, however, significant differences between those screened and those not screened in terms of use of services, perceived health status and level of morale. In terms of unit screen non-response, it would appear that the type of missing data is, in terms of its severity, at least weakly missing at random. It is, however, possible that the active mechanism is non-ignorable although the detection of such a mechanism is either impractical, impossible or would require additional assumptions regarding the form of the mechanism. One such assumption would be to assume that those screened 'late' had similar characteristics to those not screened. A suitable measure of 'lateness' would be the difference between the date of screen and the date of posting (corresponding to the date when the letter informing the individual of

their pending visit was generated). A brief investigation of the link between refusal and an individual's ADL score, perceived health status and dementia status was carried out for those screened at one wave and refusing at the next. This was done separately for the first pair and second pair of waves. There was no evidence of an association between refusal and ADL dependence, perceived health status or dementia status for either the first or second pair of screens. There appeared to be no strong link between an individual's screen values and their likelihood of being screened in the next wave. This concurs with Jagger et al. (1996) to a reasonable degree except for the lack of an association between perceived health status in this study and acceptance of the later screens.

A further complication arises from the 377 people who were offered one or more screens but were never screened. As none of the potential risk factors collected during the screen will be available, these individuals were excluded from the base data set. However, with the definition of survival time given earlier, these individuals never actually entered the study nor became members of the population at risk. These individuals would, however, be important if the population under consideration were the more general 'over 75s' rather than the subset who had been screened. These 377 elderly Melton Mowbray residents eligible for screening but never screened comprised:

- 8 offered a screen in wave 3 only and found to have died prior to a screen being performed (similar individuals may have existed for the first two waves but did not appear in the data set);
- 261 who 'refused' a single screen (comprising 79 offered a screen during wave 3 only: 128 out of the remaining 182 subsequently died);
- 108 who 'refused' two screens.

Another form of unit screen non-response, discussed briefly when choosing the

time scale, is due to staggered entry. Individuals become part of the population of interest on attaining the age of 75 and are offered screens thereafter. Because the population of interest is defined as *those aged over 75 who have been screened* and the definition of survival time as *time from entry into the study*, it is more appropriate to view their third (and possibly fourth) potential screens, rather than their first (and possibly second), to be missing. This staggered entry would not usually lead to any problems as those in the study could be viewed as a random sample of the population at risk. In this case, matters are a little more complex due to the sampling mechanism. If, as in most towns, geographic areas are related to the probability of occurrence of death, as they tend to be a measure of health, prosperity and other potential risk factors or their surrogates, then it is possible that a little bias could be introduced due to missingness that is non-ignorable in terms of the models fitted.

Table 5.2 contains the screen profiles of the 2203 people in the base data set. Those with leading missing values, numbering 805 in total, are largely those attaining the age of 75 after screening commenced in their neighbourhood. The remainder are those who either moved to Melton Mowbray after the start of the study or were missed due to moving within the town during this first screening wave. Those who were offered a screen in the first wave and yet were missed in the later waves were missed due to removal within or out of the town. These number 25 for the second wave, comprising 4 refusers and 21 screened from the first wave, of whom 9 were offered the third screen. Similarly there were 81 missing for the third wave, comprising 14 who had been missing in the first wave, 16 who had been missing in the second wave, 30 who had previously been screened twice and 21 who had had one previous screen and one previous refusal. Reasons for unit screen non-response, other than that of 'late entry' discussed earlier, may be a little more problematical as they will depend on the true, but undetectable, reasons for the non-response. A few temporary hospitalisations may be included amongst those missed or recorded as refusals, although those hospitalised will tend to have

been contacted on being discharged, providing they returned home or moved to an area of Melton in which the screening for that wave had not been completed. There will also be removals to other types of temporary or permanent residence within Melton. Removals of this type could be an indication of either continuing good health or deterioration of health, depending on the type of the housing into which the subjects moved. Recent ill-health is likely also to contribute to removals out of the study area, such as those moving to live with relatives on a temporary or permanent basis. Data missing for this or other reasons for removal would be due to either a MAR or a non-ignorable missing-data mechanism, depending on whether their cause could be detected from earlier screen information. Although cases recorded as 'missing' were moderate in number, especially for the third screen, and it would have been useful to determine their survival to enable this information to be included in the analysis, this was not possible at the time. It should be noted, however, that fully or partially excluding such cases from the analysis will tend to introduce a little bias into the estimation of the model parameters.

Considering those offered a screen and yet either actively or passively refusing, the first wave refusers (numbering 325) are ignorable due to the definition of the population. They are, however, right-censored by a mechanism which will not be MCAR if the first wave refusal is linked to survival. Those refusing the third screen but having earlier been screened, numbering 146, are members of the population but their covariate information is not available for the third screen (although this is not used in the analysis except when modelling for the multiple imputation). These individuals are, however, removed either fully or partially from many of the analyses described in Sections 5.3-5.6 and so, if their reasons for refusal are related to the change of state, their exclusion will tend to introduce bias.

Those individuals of particular interest in this thesis are the 238 non-respondents in the second wave. Of these, 116 had not previously been screened and so only

their future covariate values, from the third wave, are available. As this thesis does not investigate imputation using the **individual's** *future* realisations of risk factors as predictors, these cases will be excluded, either fully or partially, from the analyses. For the remaining 122, there was information available from previous screens which could be utilised in a number of ways in order to model changes of state. These methods will be discussed in Section 5.5 and applied in Section 5.6. Of these 122, the nine who were missing at the third screen are therefore censored at their second screen, so the covariate information would not have been used (except in the modelling for the multiple imputation) even if it had been available. The remaining 113 comprise 55 screened during wave 3, 32 who refused at wave 3 and 26 who had died by wave 3.

Some insight into the pattern of unit screen non-response due to active or passive refusal can be obtained by considering those 870 screened at wave 1, offered a screen at wave 2 and not missing at wave 3. The 113 refusing a screen at wave 2 are considered above whilst, of the remaining 757, 592 were screened at wave 3, 48 refused and 117 had died. There is clearly, therefore, a tendency for multiple consecutive refusers as, of those still alive in wave 3, only 7.5% of those screened twice refused the third wave screen whereas 36.7% of those who followed a first wave screen by refusing the second wave screen also refused the third wave screen. Further insight into refusals could be gained by extending the number of 'survival' states from two (*alive* and *dead*) to three (*alive*, *refused* and *dead*) and then using multi-state modelling to investigate the probabilities of the state transitions. However, as this is an *extension* of the work presented in this thesis, it will still be subject to the problems of both missing covariate data and missing (rather than refusing) screens. Further insight into the unit screen refusal mechanism can be gained by comparing the percentages of those dying between waves 2 and 3 for the different screen participation rates over the first two waves. Excluding the 43 individuals recorded as missing for wave 3:

- 117 (15.46%) of the 757 screened twice had died;
- 69 (19.06%) of the 362 screened once had died;
- 22 (34.07%) of the 91 not screened had died.

These percentages would tend to provide support for the hypothesis that the refusal mechanism is not MCAR. In particular, on fitting a simple logistic model, there appears to be a trend in the log-odds of death which is linear in the number of refusals (χ^2=5.34, p=0.021).

The amount of *item* missing data is considerably less. Many of the variables considered had little or no item non-response. The only variables with consistently over 5% item non-response were demi-span, perceived health status and weight. The numbers and percentages of item non-response at each wave are given in Table 5.3. It is expected that item non-response is less problematical than the unit screen non-response as it is less prevalent and believed to be more likely to be due to an ignorable mechanism. On investigating the association between the item missing data for the three variables with consistently over 5% item non-response and health measurements made at the previous screen, the following observations were made. The item missingness in demi-span appeared to be associated with both ADL dependence and dementia status at the previous screen. The evidence of an association in each case was stronger for the second pair of screens than the first (p=0.115 and p=0.001 for ADL dependence and p=0.054 and p<0.001 for dementia status respectively for the first and second pairs of screens using χ^2-tests). Similarly there was very strong evidence of associations between item missing weights and both ADL dependence and dementia status (p<0.001 for ADL dependence for each pair and p=0.115 and p<0.001 for dementia status respectively for the first and second pairs of screens using χ^2-tests). The evidence of associations between item non-response for the perceived health status question is considerably less. There is no evidence of an association with ADL dependence although there is evidence of an association with dementia status, mainly for the

Screen	Screen2	Screen3	Frequency
.	.	1	328
.	1	.	14
.	1	1	310
.	1	2	38
.	1	3	59
.	2	1	56
1	.	.	16
1	.	1	3
1	.	2	1
1	.	3	1
1	1	.	30
1	1	1	592
1	1	2	48
1	1	3	117
1	2	.	9
1	2	1	55
1	2	2	32
1	2	3	26
1	3	3	143
2	.	1	4
2	1	.	12
2	1	1	179
2	1	2	27
2	1	3	43
2	2	1	60

Table 5.2: Frequencies of the different screen profiles for the 2203 individuals screened at least once: 1=screened, 2=not screened, 3=dead, .=missing (due to removal or being aged under 75).

second pair of screens (p=0.118 for the first pair and p<0.001 for the second pair using χ^2-tests). These observations indicate that the item non-response is not MCAR but it is still believed that is likely to be due to a MAR rather than a non-ignorable mechanism.

5.3 Non-imputation-based methods

As described in the introduction to this chapter, several of the methods described in earlier chapters are potentially suitable for modelling state changes for the data arising from the annual screening of the elderly in Melton Mowbray. Implementation details for a number of the more basic methods will be described in this section whereas those for imputation methods will be described in Section 5.5. Throughout this chapter, modelling will concentrate on the change of state from 'alive' to 'dead'. Whilst other changes of state, such as from 'independent' to 'dependent' for ADLs and from 'non-demented' to 'demented' would typically also be of interest in a study of this type, full consideration of their modelling will be inherently more complex as those lost to the study due to death would either require either formal multi-state modelling or the modelling of the missing-data mechanism corresponding to censoring due to death. However, some modelling of transitions between different levels of dependence for performing ADLs and between non-demented and demented states is performed in the model-based multiple imputations described in Section 5.5.

In this study, whilst the information regarding *death* is available as the *day* of occurrence, there are some instances where it will only be known that events have either 'occurred' or 'not occurred' during a specific period of time, that is the survival time will *interval-censored.* Only this form of event time would be available for modelling the changes of state relating to dependence for performing ADLs

	Wave 1		Wave 2		Wave 3	
AGESCn	7	(0.65)	1	(0.07)	1	(0.06)
DIASn	12	(1.18)	10	(0.68)	32	(2.02)
DSPANn	188	(17.52)	166	(11.30)	91	(5.73)
HHCOMPn	27	(2.52)	10	(0.68)	12	(0.76)
IOSCOREn	0	(0.00)	0	(0.00)	0	(0.00)
SEXn	0	(0.00)	0	(0.00)	0	(0.00)
SHn	40	(3.73)	63	(4.29)	36	(2.27)
SMOKINGn	11	(1.03)	10	(0.68)	6	(0.38)
SYSn	10	(0.93)	8	(0.54)	32	(2.02)
TOTADLn	0	(0.00)	0	(0.00)	0	(0.00)
WTn	65	(6.06)	80	(5.45)	88	(5.55)
TOTAL SCREENED	1073	(100)	1469	(100)	1587	(100)

Table 5.3: Numbers (%) of item non-response in each of the potential risk factors in each wave ($n = 1, 2, 3$).

and dementia outlined above. As the main objective of this thesis is to investigate the modelling of changes of state where exact change of state times are not necessarily available, survival times will generally be viewed as interval-censored in this chapter although some investigations will also be performed on the 'exact' times.

As the recording of deaths is performed independently of the screening, there is *screen-independent event collection* present and so Prentice and Gloeckler's (1978) proportional hazards model for interval-censored data can be fitted using Thompson's (1981) GLIM implementation, as described in Section 4.1. This will be done taking both complete case (CC) and available case (AC) approaches. Comparisons will then be made between inferences. The number of binary responses used in the two approaches can differ substantially if only a relatively small proportion of the potential risk factors are significant and the non-significant variables contain a reasonable amount of the item missing data. As many automatic variable selection procedures implemented in computer software use complete case methods, it is of interest to see the effect of restricting the data to those cases complete on all potential predictors has on inferences. This will give some indication as to the form of some of the active missing data mechanisms. For comparability purposes, the subset of the potential risk factors chosen using Prentice and Gloeckler's model on complete cases will be the basis for all further models. This subset is to be chosen using a stepwise backward elimination strategy, implemented via the use of likelihood-ratio tests and a 5% significance level. However, the significance of additional terms in subsequent types of models will be investigated when deemed appropriate and is reported on in Section 5.4.

As described in Chapter 4, it is still common practice in survival studies to simply measure potential risk factors at baseline and use these as fixed covariates for the duration of follow-up. None of the publications based on studies of the elderly in Melton Mowbray have used updated risk factors in the modelling of mortality

although there had been no such aim in the studies' design. It was shown both in Chapter 4 and by Altman and de Stavola (1994) that the use solely of baseline covariates will tend to lead to under-estimation of the size of the risk factor effects. Therefore, given these findings and the previous practice in the Melton studies, it is of interest to investigate the result of using regular updates of risk factor values rather than simply their baseline values on their estimated effects on survival. This will be done by comparing the effect on estimation from using baseline covariates rather than the values of covariates from the most recent screen, and will be performed for both the interval-censored and 'exact' time forms of the proportional hazards model.

It is important to note that it is believed that, for most if not all the time-dependent potential risk factors measured in this study, it is recent rather than baseline values which will have the primary effect on survival. Once further waves of screens are completed and the population has been followed-up for a longer time, it will be useful to investigate the risk factor profile in more detail. Such investigations would include more sophisticated modelling of this profile and the consideration of the introduction of an augmented covariate vector (Wu and Ware, 1979) into the survival models.

The final question to be posed in this section is whether, either through the decrease in power and precision or otherwise, the use of the interval-censored approach leads to substantively different conclusions in terms of the significance of potential risk factors and the magnitude of hazard ratios than when using the standard 'exact' times approach. To address this, proportional hazards models were fitted to the data, firstly using the interval-censored survival times and Prentice and Gloeckler's model in GLIM, and secondly using the 'exact' survival times (recorded to an accuracy of ± 1 day) via the standard implementation of Cox's proportional hazards model in SAS PROC PHREG (SAS, 1996).

5.4 Results for non-imputation-based methods

5.4.1 Complete case (CC) and available case (AC) analyses with interval-censored data

To compare the use of *complete case* rather than *available case* methods, a data set was compiled from the base data set using SAS. The complete case analysis (**CCA**) was then performed in GLIM. The data set was constructed by deleting all individuals who were not screened at least twice, constructing binary responses for each individual in each interval, subject to the individual being screened at the beginning of the interval and either being screened or dead at the end of the interval. Therefore, individuals with either a **1 1 1** or a **1 1 3** screen profile (see Table 5.2) will contribute two separate binary observations. Then all the binary observations for which at least one of the potential risk factors was missing were also deleted. Therefore, the view was taken that *complete case* meant that all potential covariates must be present, together with a screening indicator of 'dead' or 'alive and screened' at both the beginning and end of the interval. It should be noted that this set of potential covariates included age (at screen) and that this was missing for a small number of individuals. Only when considering imputation methods will these ages be imputed, despite their deterministic nature and the availability of relevant information from alternative sources. An additional factor 'PERIOD' was created to indicate the interval of interest and hence allow for potential time trends in survival, taking the value '1' if the binary observation corresponded to the interval between the first and second wave of screens and the value '2' if it corresponded to the interval between the second and third wave of screens. There were 1755 binary observations thus created. Following the

stepwise backward elimination procedure, five variables remained in the model. These were the variates measuring age (**AGE**) and weight (**WT**) and the factors classifying the level of independence in performing the activities of daily living (**ADL**), sex (**SEX**) and dementia status (**IO**).

Interactions between each pair of significant terms and quadratic terms in the continuous variables were then added, one at a time, to the main effects model. The quadratic term in weight (**WT2**) was highly significant ($\chi^2 = 9.77$, p=0.002) whereas none of the other terms was significant at the 5% level. The variable WT2 was thus included in the model and the interactions and quadratic in age were added in turn. All were highly non-significant. As the effect of weight on survival is likely to depend on an individual's size and body mass index (**BMI**), usually defined as 'weight divided by (height squared)', is often used as a more appropriate risk factor than weight, a proxy for BMI was calculated as 'weight divided by (demi-span squared)'. When linear and quadratic terms in the BMI proxy were included in place of those for weight, a slightly greater level significance was attained for the pair of weight terms (p<0.001 compared with p=0.002 for the BMI proxy pair). Moreover, a considerable amount of the item missing data was for the measurement of demi-span. It was decided, for a combination of these reasons, to include the weight terms rather than those for the BMI proxy. It should also be noted that the inclusion of the linear and quadratic terms in the BMI proxy led to a non-significant improvement in the fit of the model already including the weight terms ($\chi^2 = 2.50$, p=0.287). The model chosen as a basis for all further modelling therefore included the six variables AGE, WT, WT2, ADL, SEX and IO.

Whilst interpretation of the weight effects in terms of hazard ratios is not appropriate, the presence of a quadratic term, leading to a U-shaped risk profile over the weight range, means that estimation of an optimal weight for survival, defined as the weight corresponding to minimum risk of death, will generally be of interest.

Moreover, as the inclusion of the interactions of WT and WT2 with SEX was not significant (χ^2 = 3.33, p=0.189), the optimal weight for survival would appear to be the same for both males and females. Under CCA the optimal weight for survival was estimated to be 167.5 pounds (fractionally under 12 stone).

A data set was then compiled for the available case analysis (**ACA**) from the base data set by deleting all binary observations for which at least one of the risk factors included in the complete case model was missing. There were 2100 units of data. The near 20% increase in the number of binary observations was due largely to the absence of the demi-span variable (**DSPAN**) from the set of significant risk factors. The resulting parameter estimates, hazard ratios and the corresponding 95% confidence intervals are given in Table 5.4 (page 223). There were moderate differences between the magnitudes of the parameter estimates obtained on fitting the chosen model to the two sets of data. There was no consistent pattern in terms of the direction of the effect observed. Under ACA, the sex effect and that of moderate (as opposed to full) independence for ADLs were stronger than under CCA. The effects of age, moderate or high dependence (as opposed to full independence) for ADLs, dementia status and both linear and quadratic terms in weight were found to be less pronounced under ACA than under CCA. None of the changes in parameter estimates would lead to substantially or statistically significantly different conclusions although, in the cases of dementia status and moderate dependence for ADLs, the changes in parameter estimates were of a moderate size (7.4% and 9.9% respectively, in relative terms). The greatest change, in both relative and absolute terms, was for the parameter estimate relating to a moderate dependence for ADLs and led to a reduction in its hazard ratio estimate from 2.40 to 2.21 on changing from CCA to ACA. In terms of the estimation of optimal weight for survival, there was a moderate difference in the estimates provided by the two analyses. Using ACA, the optimal weight for survival was estimated as 176.5 pounds, a nine pound increase over that from the CCA.

The period indicator, which had been amongst the non-significant variables removed from the model during the stepwise backward elimination under CCA, was then added to the model fitted to the available cases. It remained non-significant (χ^2=0.06, df=1, p=0.806). Hence there was no evidence of any time trend in survival during the period of study under either CCA or ACA.

5.4.2 The use of baseline measurements versus last measurements

Interval-censored survival times

For the comparison of exact and interval-censored approaches to modelling survival times, appropriate data sets were constructed using SAS. With the interval-censored approach, the number of available units rose from the 2100 in the available case data set to 2421 when the baseline measurements were to be used as covariates and to 2317 when the measurements from the latest wave were used as covariates.

The parameter estimates, hazard ratios and the corresponding 95% confidence intervals from the two analyses are given in Table 5.5 (page 224). On comparing the parameter estimates from the baseline analysis with those from when the last measurements were used, their magnitudes were lower with the baseline analysis except those corresponding to age and both weight and weight-squared for which the estimates were similar. The decreases in magnitude were generally greatest for the terms that can be viewed as corresponding to serious deterioration in health, namely high dependence for ADLs (13.6%) and dementia (31.6%), although the

effect of 'sex' was also weakened by a substantial degree (18.2%). As the effects of dependence for activities of daily living and the onset of dementia can be viewed as two possible intermediate states between a healthy state and death, it is to be expected that their time-dependence is particularly important when it comes to modelling survival. As the age and weight covariates can be viewed, at least partially, as surrogates for updates of recorded risk factors (and also for unrecorded measures of health), the somewhat contrary effect on their effects due to using the baseline covariates is not unexpected.

With the baseline measurement analysis, the optimal weight for survival was estimated to be 169.4 pounds, a very similar value to that obtained using CCA, whereas that obtained using last measurement analysis led to an estimate of 172.8 pounds. However, it should be noted that, as the former estimate relates to the optimal weight for survival on entry to the study whereas the latter relates to the optimal weight for survival at any screen, they are not strictly comparable.

With the last measurement analysis, the period effect was again non-significant cant ($\chi^2 = 0.64$, df=1, p=0.424). However, due to the weakening of most of the effects of the baseline covariates by the interval between the second and third waves of screens, the inclusion of the period effect then leads to a significant improvement in the model fit ($\chi^2 = 6.46$, df=1, p=0.011) as it corresponds to a surrogate for the time-dependent effects of the baseline covariates.

The standard errors of the parameter estimates were generally slightly lower under baseline measurement analysis than under last measurement analysis. The exceptions to this were the parameter for IO(2), measuring the effect of dementia, and WT and WT2, the linear and quadratic effects of weight. As the number of observations used in the last measurement analysis was approximately 95.7% of those used in the baseline measurement analysis, the fact that these three standard errors are estimated to be slightly lower in the last measurement analysis

would tend to confirm that recent mental status and weight are more consistent predictors of mortality than recordings from the start of the study.

The results are extremely similar for ACA and the last measurement analysis. This is not unexpected as the only difference in construction of the data sets is that, for the ACA, all binary observations resulting from missed second and third screens were excluded whereas for the last measurement analysis such observations were only excluded if the covariate information from the start of the interval was either subject to item or unit screen non-response.

'Exact' survival times

As shown in Tables 5.5 and 5.6 (pages 224 and 225), when the 'exact' ($\pm$ 1 day) survival times are used as opposed to the interval-censored times, similar estimates are obtained for most parameters, although the effects detected are slightly stronger when the 'exact' times are used. With regard to the precision, slightly lower standard error estimates were obtained with the interval-censored analysis. This is due to the additional implicit parameters in the non-parametric component of Cox's proportional hazards model which, whilst not estimated directly, will reduce the precision of the estimates in the parametric component when compared with the interval-censored method in which a limited number of nuisance parameters, corresponding to the follow-up cut-points, are explicitly estimated.

The effect of using the baseline rather than the last measurements with 'exact' survival times mirrors that observed in the interval-censored approach, namely a weakening of all effects bar that of age. Similarly, the standard errors are generally slightly lower under the baseline measurement analysis, due in this case to the increased censoring rather than loss of binary observations when there was item or unit non-response in the second wave following a first wave screen. On this

occasion, only the linear and quadratic effects of weight had lower standard errors under last measurement analysis as the estimated standard error for IO(2), which was marginally lower for the last measurement analysis with interval-censored data, was now marginally higher.

When using the baseline measurements, the optimal weight for survival was estimated to be 172.6 pounds whereas it was 178.8 pounds when the last measurement analysis was used. As with the interval-censored approach, these estimates are not strictly comparable due to their different interpretations. However, both estimates are greater than their counterparts under the interval-censored approach, that under the baseline analysis by around 3 pounds and that under the last measurement analysis by around 6 pounds. Whilst differences of around a stone or a stone and a half may well be viewed as substantive, the difference between an optimal weight for survival of 12 stone 5 pounds and one of 12 stone 11 pounds would not be deemed important. Furthermore, when confidence intervals for the optimal weight for survival are constructed, the lack of precision in estimating the linear and quadratic effects of weight combine to give extremely wide intervals. For example, using Fieller's method to construct a 95% confidence interval for the optimal weight for survival gave the interval (157.7, 276.8) pounds when the last measurement analysis was used.

As the weight effects are not particularly strong it is not surprising that there was no data-based evidence of differences between its effect on survival for the two sexes when the terms were chosen for inclusion in the model earlier. However, the selection of the same optimal weight for both males and females would be likely to be viewed with scepticism for obvious reasons. As discussed earlier, the inclusion of a BMI proxy in the survival model in place of weight might alleviate this problem but this is not particularly practical due to the moderate proportion of missing demi-span measurements. Further investigation of differential effects of weight for males and females showed that, under the last measurements approach, the in-

teractions of sex with weight and weight-squared were non-significant ($\chi^2 = 2.99$, p=0.224). However, on including these interactions, there were moderate differences between the resultant optimal weights for survival of 178.1 pounds for males and 168.2 pounds for females. For males this estimate is very similar to that obtained under the homogeneous weight effects for the sexes whereas for females the estimate is a full ten pounds lower. Suggesting an optimal weight for survival of 12 stone for elderly women may, perhaps, still be viewed with a certain degree of scepticism but is still more acceptable than the earlier value of 12 stone 11 pounds suggested for both males and females. Therefore, although these interaction terms were found to be non-significant, there is a reasonable argument for their inclusion in a model that might be used for determination of optimal weights for survival as it would be viewed as inappropriate to suggest the same optimal weight for both men and women.

An additional point of note is that, with the interactions between the weight terms and sex included in the model, the estimated risk profile was considerably flatter for women than for men, the coefficient of the quadratic term for females being estimated at only a little over a half of that for males. Whilst the 95% confidence intervals for optimal weight for survival by sex under the last measurement analysis are still very wide, the interval of (163.1, 237.5) pounds for males is far narrower than the interval for males and females combined, further suggesting that, despite the lack of statistical significance, the weight by sex interaction should be included in the survival model. The corresponding interval for female optimal weight is (142.4, 591.9) pounds, illustrating the relatively ill-defined nature of the risk profile for women. It would therefore appear that an elderly individual's weight is a more important risk factor for their survival if they are male rather than female.

The inclusion of the non-significant main effects from the model selection performed using CCA was investigated. None of these terms attained significance at

the 5% level although perceived health status, which had been the last term removed in the earlier backward elimination under complete case analysis (χ^2 = 3.80, df=2, p=0.150) was now close to significance at the 5% level (χ^2 = 5.50, df=2, p=0.064).

Term	Category/*Units*	Method	PARAMETER Estimate	PARAMETER 95% C.I.	HAZARD RATIO Estimate	HAZARD RATIO 95% C.I.
ADL(2)	Moderately independent	**CC**	0.512	(0.175, 0.849)	1.669	(1.192, 2.339)
		AC	0.530	(0.236, 0.823)	1.698	(1.267, 2.277)
ADL(3)	Moderately dependent	**CC**	0.842	(0.455, 1.228)	2.320	(1.576, 3.415)
		AC	0.759	(0.415, 1.103)	2.135	(1.514, 3.012)
ADL(4)	Highly dependent	**CC**	1.229	(0.827, 1.631)	3.418	(2.286, 5.110)
		AC	1.227	(0.890, 1.564)	3.412	(2.436, 4.780)
IO(2)	Demented	**CC**	0.674	(0.302, 1.045)	1.961	(1.352, 2.844)
		AC	0.624	(0.313, 0.934)	1.866	(1.367, 2.546)
WT	*10 Pounds*	**CC**	-0.532	(-0.816, -0.249)	*	*
		AC	-0.378	(-0.617, -0.138)	*	*
WT2	*(10 Pounds)*^{2}	**CC**	0.0159	(0.0066, 0.0252)	*	*
		AC	0.0107	(0.0028, 0.0186)	*	*
AGE	*Years*	**CC**	0.0790	(0.0506, 0.1075)	1.082	(1.052, 1.114)
		AC	0.0783	(0.0540, 0.1027)	1.082	(1.056, 1.108)
SEX(2)	Female	**CC**	-0.711	(-1.014, -0.409)	0.491	(0.363, 0.664)
		AC	-0.747	(-1.009, -0.486)	0.474	(0.365, 0.615)

* - Not appropriate due to presence of both linear and quadratic terms

Table 5.4: Comparison of estimates of parameters and hazard ratios from proportional hazards model for complete case (CC) and available case (AC) analyses with an interval-censored approach. Comparison groups: Fully independent (ADL(1)); Non-demented (IO(1)); Male (SEX(1)).

Term	Category/*Units*	Method	PARAMETER Estimate	PARAMETER 95% C.I.	HAZARD RATIO Estimate	HAZARD RATIO 95% C.I.
ADL(2)	Moderately independent	**BM**	0.537	(0.251, 0.822)	1.710	(1.285, 2.276)
		LM	0.565	(0.272, 0.857)	1.759	(1.313, 2.356)
ADL(3)	Moderately dependent	**BM**	0.714	(0.370, 1.058)	2.042	(1.448, 2.879)
		LM	0.768	(0.424, 1.112)	2.156	(1.528, 3.041)
ADL(4)	Highly dependent	**BM**	1.066	(0.740, 1.393)	2.905	(2.096, 4.027)
		LM	1.234	(0.896, 1.572)	3.434	(2.450, 4.815)
IO(2)	Demented	**BM**	0.443	(0.134, 0.752)	1.558	(1.144, 2.122)
		LM	0.648	(0.340, 0.956)	1.912	(1.405, 2.601)
WT	*10 Pounds*	**BM**	-0.377	(-0.622, -0.132)	*	*
		LM	-0.380	(-0.619, -0.141)	*	*
WT2	*(10 Pounds)*^{2}	**BM**	0.0111	(0.0030, 0.0192)	*	*
		LM	0.0110	(0.0031, 0.0189)	*	*
AGE	*Years*	**BM**	0.0775	(0.0543, 0.1007)	1.081	(1.056, 1.106)
		LM	0.0768	(0.0525, 0.1010)	1.080	(1.054, 1.106)
SEX(2)	Female	**BM**	-0.611	(-0.863, -0.359)	0.543	(0.422, 0.699)
		LM	-0.747	(-1.007, -0.486)	0.474	(0.365, 0.615)

* - Not appropriate due to presence of both linear and quadratic terms

Table 5.5: Comparison of estimates of parameters and hazard ratios for available case (AC) analysis using baseline measurements (BM) and last measurements (LM) as covariates in proportional hazards model with interval-censored approach and screen-independent event collection. Comparison groups: Fully independent (ADL(1)); Non-demented (IO(1)); Male (SEX(1)).

Term	Category/*Units*	Method	PARAMETER Estimate	PARAMETER 95% C.I.	HAZARD RATIO Estimate	HAZARD RATIO 95% C.I.
ADL(2)	Moderately independent	**BM**	0.543	(0.235, 0.852)	1.721	(1.265, 2.343)
		LM	0.572	(0.254, 0.889)	1.771	(1.289, 2.433)
ADL(3)	Moderately dependent	**BM**	0.748	(0.386, 1.109)	2.112	(1.471, 3.032)
		LM	0.803	(0.436, 1.171)	2.233	(1.546, 3.224)
ADL(4)	Highly dependent	**BM**	1.106	(0.757, 1.456)	3.023	(2.132, 4.288)
		LM	1.264	(0.895, 1.633)	3.539	(2.447, 5.119)
IO(2)	Demented	**BM**	0.496	(0.176, 0.817)	1.643	(1.192, 2.263)
		LM	0.665	(0.341, 0.988)	1.944	(1.407, 2.686)
WT	*10 Pounds*	**BM**	-0.412	(-0.668, -0.156)	*	*
		LM	-0.404	(-0.659, -0.149)	*	*
WT2	*(10 Pounds)*^{2}	**BM**	0.0119	(0.0034, 0.0205)	*	*
		LM	0.0113	(0.0029, 0.0198)	*	*
AGE	*Years*	**BM**	0.0808	(0.0565, 0.1052)	1.084	(1.058, 1.111)
		LM	0.0754	(0.0497, 0.1010)	1.078	(1.051, 1.106)
SEX(2)	Female	**BM**	-0.679	(-0.948, -0.410)	0.507	(0.388, 0.664)
		LM	-0.815	(-1.097, -0.533)	0.443	(0.334, 0.587)

* - Not appropriate due to presence of both linear and quadratic terms

Table 5.6: Comparison of estimates of parameters and hazard ratios for available case (AC) analysis using baseline measurements (BM) and last measurements (LM) as covariates in proportional hazards model with 'exact' times approach. Comparison groups: Fully independent (ADL(1)); Non-demented (IO(1)); Male (SEX(1)).

5.5 Imputation-based methods

A comparison will now be made between the use of the last available covariate values, corresponding to analysis following *last measurement imputation*, covered extensively in Chapter 4, and a *model-based multiple imputation method*, also investigated in Chapter 4. Each variable considered for inclusion in the model can be classified into one of four types: **time-independent**; **deterministic**; **non-deterministic continuous**; **non-deterministic ordinal**. A time-independent variable is one which, for each individual, is both known exactly and constant for the duration of the study. Such a variable is included in a model as a *fixed* rather than *time-dependent* covariate and, if missing for some cases, can be imputed as its value from another screen. A deterministic variable is time-dependent but only varies in a deterministic manner. This means that if such a variable is missing at a particular screen, its value can, in theory, be predicted **exactly** using its value from another point in the study and, additionally, such a variable is open to consideration for inclusion in a hazard model as a time-dependent covariate. It is assumed that there are no 'measurement errors' in the time-independent and deterministic variables. A non-deterministic variable is one whose value cannot be predicted precisely from the knowledge of its value at other screens and therefore is subject to imputation with uncertainty. Non-deterministic continuous variables will be modelled using standard linear modelling techniques whilst non-deterministic ordinal variables will be modelled using transition proportional odds models, with the special case of logistic models applying if there are only two categories. Missing values at item and unit screen level will be imputed using the same technique. Both last measurement imputation and model-based multiple imputation analyses will firstly be performed under *screen-independent event collection* and then under the assumption that the event collection was *screen-dependent.*

The difference between last measurement analysis considered earlier and last measurement imputation is that only the 2317 binary units available in terms of both all covariates in the model from the screen at the start of the relevant interval and knowledge of the screen-independent outcome at the end of the interval were used in the former whereas all 2458 binary units with a prior measurement of each model covariate and knowledge of the screen-independent outcome were used in the latter. The difference in binary units relates almost entirely to those who refused the second wave screen or failed to have their weights recorded at the second wave screen. In addition there were just seven individuals whose age was not recorded at their first wave screen and one other whose age was not recorded at their second wave screen.

Considering the significant variables SEX, AGE, ADL, IO, WT and WT2, the SEX covariate is *time-independent* and can be imputed as a **known** value from earlier screens and is therefore not subject to different imputation strategies. The AGE variable is *deterministic* as it can, subject to being rounded to the nearest year, be determinined from either the first screen or, more easily and less subject to rounding problems, from the age at posting for the appropriate screen. Age at posting was never missing unless an individual was not offered a screen in that wave. Age is therefore not subjected to the two imputation strategies. This leaves ADL, a non-deterministic ordinal variable with no item non-response, IO, a non-deterministic ordinal (more specifically binary) variable also with no item non-response and WT (and hence WT2), a non-deterministic continuous variable with 23 individuals subject to item non-response with a recorded weight from a screen during the first wave. The two ordinal variables will be subjected to last measurement imputation and multiple imputation based on a transition proportional odds model. In the case of IO, the proportional odds model reduces to the logistic model. The variable weight will be subjected to last measurement imputation and multiple imputation based on a linear model. The models for these three variables will be constucted independently and will consider all variables

from the previous screen using available case analysis.

In the model-based multiple imputation approach the full parameter vector $\underline{\beta}$ is partitioned into four distinct components. These components are $\underline{\beta}_1$, the parameters of the model for WT, $\underline{\beta}_2$, the parameters of the model for IO and $\underline{\beta}_3$, the parameters of the model for ADL and $\underline{\beta}_4$, the parameters of the survival model. Hence, in terms of the survivor function and missing data likelihood contributions for a particular individual, we have:

$$f(\mathrm{WT}_m, \mathrm{IO}_m, \mathrm{ADL}_m, \underline{\beta}_4 \mid \underline{x}_o) = f(\underline{\beta}_4 \mid \mathrm{WT}_m, \mathrm{IO}_m, \mathrm{ADL}_m, \underline{x}_o) \times f(\mathrm{WT}_m, \mathrm{IO}_m, \mathrm{ADL}_m \mid \underline{x}_o)$$

where $\underline{x}_o{}^T = (\mathrm{ADL}_o, \mathrm{IO}_o, \mathrm{IOSCORE}_o, \mathrm{WT}_o, \mathrm{DSPAN}_o, \mathrm{HHCOMP}_o, \mathrm{SH}_o, \mathrm{SYS}_o, \mathrm{DIAS}_o, \mathrm{AGE}, \mathrm{SEX}, \mathrm{PERIOD})$.

Now, assuming independence,

$$\begin{aligned} f(\mathrm{WT}_m, \mathrm{IO}_m, \mathrm{ADL}_m \mid \underline{x}_o) &= f(\mathrm{WT}_m \mid \underline{x}_o) \cdot f(\mathrm{IO}_m \mid \underline{x}_o) \cdot f(\mathrm{ADL}_m \mid \underline{x}_o) \\ &= \int f(\mathrm{WT}_m \mid \underline{\beta}_1, \underline{x}_o) f(\underline{\beta}_1 \mid \underline{x}_o) \, \mathrm{d}\underline{\beta}_1 \\ &\times \int f(\mathrm{IO}_m \mid \underline{\beta}_2, \underline{x}_o) f(\underline{\beta}_2 \mid \underline{x}_o) \, \mathrm{d}\underline{\beta}_2 \\ &\times \int f(\mathrm{ADL}_m \mid \underline{\beta}_3, \underline{x}_o) f(\underline{\beta}_3 \mid \underline{x}_o) \, \mathrm{d}\underline{\beta}_3 \end{aligned}$$

It therefore remains to estimate the parameters of the models for WT, IO and ADL and, *under the assumption of approximate MAR mechanisms causing the missing data*, impute based on $f(\mathrm{WT}_m \mid \underline{\beta}_1, \underline{x}_o)$, $f(\mathrm{IO}_m \mid \underline{\beta}_2, \underline{x}_o)$, $f(\mathrm{ADL}_m \mid \underline{\beta}_3, \underline{x}_o)$ which are assumed to be of the same form as the models for the observed data for these variables. Ten separate data sets are then formed, the imputed values across the data sets being representative of the predictive density

$$f(\mathrm{WT}_m,\ \mathrm{IO}_m, \mathrm{ADL}_m \mid \underline{\boldsymbol{x}}_o)$$

which is generated as

$$f(\mathrm{WT}_m \mid \underline{\boldsymbol{x}}_o) \cdot f(\mathrm{IO}_m \mid \underline{\boldsymbol{x}}_o) \cdot f(\mathrm{ADL}_m \mid \underline{\boldsymbol{x}}_o).$$

The conditional density

$$f(\underline{\boldsymbol{\beta}}_4 \mid \mathrm{WT}_m,\ \mathrm{IO}_m,\ \mathrm{ADL}_m,\ \underline{\boldsymbol{x}}_o)$$

is then averaged over the sample from the predictive density given by the ten data sets completed via imputation, using likelihood methodology to obtain estimates of $\mathrm{E}(\underline{\boldsymbol{\beta}}_4)$ and $\mathrm{Var}(\underline{\boldsymbol{\beta}}_4)$ as described in Section 1.3.1.

The models for WT, IO and ADL were fitted in SAS. PROC REG (SAS, 1996) was used for WT and PROC LOGISTIC (SAS, 1996) was used for both IO and ADL, the model for ADL being a transition proportional odds model (Diggle, Liang and Zeger, 1994) using the ADL values from the previous screen as described in Section 4.5. This model is of the form:

$$\log\left[\frac{\mathrm{P}\,(\mathrm{ADL}_j \leq k \mid \mathrm{ADL}_{j-1} = a_{j-1})}{\mathrm{P}\,(\mathrm{ADL}_j > k \mid \mathrm{ADL}_{j-1} = a_{j-1})}\right] = \theta_k + \underline{\gamma}^T \underline{\boldsymbol{x}}_o + \sum_{m=1}^{3} \tau_m a^*_{j-1\,m}$$

$$j = 2,\ 3; \quad k = 1,\ 2,\ 3\,;$$

where $$a^*_{j-1\,m} = \begin{cases} 1 & \text{if} \quad a_{j-1} \leq m \\ 0 & \text{if} \quad a_{j-1} > m \end{cases}$$

and $\underline{\beta}_3^T = \left(\theta_1, \theta_2, \theta_3, \tau_1, \tau_2, \tau_3, \underline{\gamma}^T\right)$.

For each model, selection of the main effects was via a backward elimination strategy followed by a forward selection of any interactions and quadratic terms based only on the main effects. As the asymptotic multivariate normality of the parameter estimates was to be used for multiple imputation and SAS does not include a function for generating random values from such a distribution, the vector of parameter estimates and its estimated variance-covariance matrix was written to a separate file for each of the three models. These files were then read into short Fortran programs which, using the NAG subroutines G05CBF, G05EAF and G05EZF, generated the ten sets of simulated parameter values for each model. These parameter values were then written to file and read back into SAS where the ten completed data sets for analysis were generated. For each data set the fitted values for the missing wave 2 weights were computed and added to pseudo-random observations generated from a normal distribution with zero mean and an appropriate variance. For each data set, this variance was taken to be the mean square error from the model for WT, scaled by a random value from the χ^2-distribution on the error degrees of freedom (1712) divided by its degrees of freedom. Similarly, for each data set the logistic and proportional odds models were used to obtain the fitted probabilities for the missing wave 2 ADL and IO values based on the appropriate set of parameter values. Random values were then generated from a uniform distribution on the interval [0, 1] and the missing values imputed following appropriate comparison of the fitted probabilites with these random values.

Analyses based on both last measurement imputation and the model-based multiple imputation under screen-independent event collection were performed using SAS PROC LOGISTIC (SAS, 1996) with complementary-log-log link.

Although it is rather artificial to illustrate the use of 'adapted Finkelstein' by

treating the deaths as both interval-censored and collected only at screens, it was decided to pursue this approach purely for illustrative and comparative purposes. A number of actions were taken in order to treat the deaths in this way. Those missing at wave 2 were handled in the same way as those actively or passively refusing at that wave. Individuals 'refusing' at wave 3 were missing for the second interval and, except for those screened in both the first two waves, were also missing for the first interval. Finally, people screened during the first wave, missing or refusers during the second wave and screened or dead for the third wave were treated as follows:

- those dead by the third wave were treated as if they could have died during either the first or second intervals;
- those screened during the third wave were treated as 'alive' during both the first and second intervals.

Given that the deaths are really screen-independent and that those not screened in wave 2 are actually known to be alive at that time, it would be expected that treating those dead by the third wave and not screened during the second in this manner would introduce bias into the estimation. Moreover, as this action will tend to decrease the hazard for the second period and increase that for the first, the omission of a period term, corresponding to the assumption of equal interval-specific cumulative baseline hazard contributions, may not be supported by the data.

Fortran programs were written for maximum likelihood estimation of the parameters using 'adapted Finkelstein' for screen-dependent deaths. Both last measurement imputation and model-based multiple imputation completed data sets were analysed. Additionally, a weight-by-sex interaction was included in each model to enable estimation of sex-specific optimal weights for survival. As these

programs were written merely for illustrative purposes at this stage, it was not deemed time-efficient to consider estimation of the variance-covariance matrix for the parameter estimates.

5.6 Results for imputation-based techniques

In choosing the model for the variable WT, the variables measuring 'weight at the previous screen', 'age at the previous screen' and 'systolic blood pressure at the previous screen' were all highly significant at the end of the backward elimination procedure. Their parameter estimates and standard errors are given in Table 5.7. The R^2 value of 88.25 compared with an R^2 value of 88.07 when the variable measuring 'weight at previous screen' was included alone. There was no evidence of any quadratic or interaction effects of these three variables, nor was there any evidence of heterogeneity of variance or highly influential observations. The effect of the weight at the previous screen was estimated at 0.963 (to 3 s.f.) and is highly significantly different from unity ($t=-4.26$, $p<0.001$). There were, however, a relatively large number of very large standardised residuals, with eight standardised residuals with magnitude greater than four. However, given that there were 1716 observations used to constuct this model, even though eight standardised residuals is some eighty times the number expected, the main result will be that the variability in imputations will be greater leading to a increase in corrected standard errors for the parameters of the hazard model fitted to the multiply imputed data sets. An alternative strategy would be to use hot-deck multiple imputation using the raw residuals, athough it was deemed unnecessary in this case due to the overall adequacy of the model for weight.

The second model related to the imputations for IO, namely for

$f(\text{IO}_m \mid \underline{\beta}_2, \underline{x}_o)$ and for which the distribution of $f(\underline{\beta}_2 \mid \underline{x}_o)$ was estimated using a logistic model for IO based on the covariates from the previous screen using available case analysis. The significant terms following the selection procedure were PERIOD, IOSCORE, SEX, AGE, HHCOMP, SH, WT, WT2 and ADL. It should be noted that including IOSCORE rather than the dementia indicator IO provided a very much better fit. (Twice the negative of the difference in log-likelihoods equalled 73.79 with no difference in model degrees of freedom.) Moreover, there was no evidence of the effect of the IOSCORE value at the previous screen being of non-linear form and, whilst the interaction between ADL and WT was of marginal significance at the 5% level ($\chi^2 = 8.03$, df=3, p=0.045), it was decided to exclude this term on grounds of parsimony. The parameter estimates and their estimated standard errors are given in Table 5.8.

There was a highly significantly greater chance of individuals being classified as suffering from dementia at a wave 3 screen than at a wave 2 screen. In fact, the average score in the IO subtest was greater at wave 2 than at wave 1 amongst

Parameter	Estimate	Standard error
INTERCEPT	26.472	5.333
WT	0.963	0.0087
AGE	-0.198	0.0583
SYS	-0.0415	0.0106

Table 5.7: Parameter estimates and standard errors for model fitted to WT for model-based multiple imputation.

Parameter	Estimate	Standard error
INTERCEPT	-1.922	1.652
PERIOD(2)	0.744	0.128
IOSCORE	-0.471	0.041
WT	-0.037	0.014
WT2	10.0×10^{-5}	4.7×10^{-5}
AGE	0.099	0.014
SEX(2)	-0.496	0.142
HHCOMP(2)	0.008	0.128
HHCOMP(3)	0.493	0.282
SH(2)	0.664	0.132
SH(3)	0.237	0.309
ADL(2)	0.237	0.146
ADL(3)	0.608	0.192
ADL(4)	0.664	0.237

Table 5.8: Parameter estimates and standard errors for model fitted to dementia status (IO) for model-based multiple imputation.

those screened at both the waves. It is believed that this is, at least partially, due to a learning effect but possibly also due to an increased familiarity with the screening process leading to less incorrect answers due to feeling pressurised under questioning. As with the risk of death, those at the extremities of the weight distribution were at greater risk of dementia then those near the centre. Similarly, the risk of dementia was higher for the more elderly and also for males, those living in institutions, those with poor perceived health status and those with moderate or high dependency on others for performing their activities of daily living.

The third model for missing data related to the missing values for ADL, namely for $f(\text{ADL}_m \mid \underline{\beta}_3, \underline{x}_o)$ for which $f(\underline{\beta}_3 \mid \underline{x}_o)$ was estimated using a proportional odds model for ADL based on the covariates from the previous screen, using available case analysis. The proportional odds model chosen following the selection procedure included the variables AGE, WT, DSPAN, HHCOMP, SH, ADL and interactions between AGE and each of SH and HHCOMP. Although this model had 17 parameters and the logistic link led to a better fit than either the probit or complementary-log-log link, there was strong evidence that the proportional odds assumption did not hold. Some investigation of this problem was carried out, involving the reduction of the number of ADL categories from four to three. This was done by combining those in the moderate and high dependence categories, due to the relatively low number of highly dependent individuals, and this action led to a considerably lower degree of lack of proportionality. However as this investigation is designed to provide an illustration of the general applicability of the technique, the time available for further investigation of modelling ADL was severely limited and earlier analyses had indicated that moderate and high dependence for ADLs had different effects on survival, it was decided to proceed with multiple imputation based on the proportional odds model whilst noting that better imputations would arise from a better fitting model for $f(\underline{\beta}_4 \mid \underline{x}_o)$. The parameter estimates and standard errors for the proportional odds model are given in Table 5.9. In the table, the notation ADLXm corresponds to the parameters

τ_m under the model specification given in Section 5.5. It should be noted that, in this model, the effect of the previous ADL value on the log-odds is assumed constant across the levels of present ADL value as there are no interactions between ADLXm and the present ADL category. The strongest effects on present ADL value for a given previous ADL value were found to be those of perceived health status and household composition. In particular, for the younger subjects with good perceived health status or living with others the odds favoured maintaining (or improving) independence amongst ADLs to a greater extent than for those with fair or poor perceived health status or those living alone or in institutions, although these effects tended to decline with age, resulting in a reversal of direction of effect by age 90. Those more independent in performing their activities of daily living and the lighter, taller and younger people were also less likely to deteriorate between successive waves than their more dependent, heavier, shorter and older counterparts.

5.6.1 Screen-independent event collection

The use of imputation rather than simply last measurement analysis meant that information on first wave covariate values could be used to impute for second wave covariate values amongst the 113 refusers who had been screened at wave 1 and were not missing at wave 3, together with weights for those with item missing wave 2 weights. The age of the single individual with age unrecorded at their wave 2 screen was imputed based on their age at posting, as described earlier. Similarly, the ages of the seven individuals with item missing first wave ages were imputed based on their ages at posting. This led to an increase to 2458 units used in the binary response modelling under last measurement imputation. One less observation was available for the model-based multiple imputations due to incomplete wave 1 covariate information for one individual. It was not deemed necessary to take special action for this case.

Parameter	Estimate	Standard error
INTERCEPT 1 (θ_1)	4.158	1.994
INTERCEPT 2 (θ_2)	6.928	1.999
INTERCEPT 3 (θ_3)	9.239	2.001
AGE	-0.135	0.022
WT	-6.71×10^{-3}	2.26×10^{-3}
DSPAN	0.0382	0.0120
HHC(2)	-4.732	2.249
HHC(3)	-8.382	4.345
SH(2)	-5.210	2.384
SH(3)	-11.680	5.826
ADLX1	2.189	0.147
ADLX2	1.826	0.215
ADLX3	1.567	0.310
AGE*SH(2)	0.0581	0.0296
AGE*SH(3)	0.130	0.0709
AGE*HHC(2)	0.0557	0.0280
AGE*HHC(3)	0.0942	0.0516

Table 5.9: Parameter estimates and standard errors for model fitted to ADL dependence level (ADL) for model-based multiple imputation.

When last measurement imputation (Table 5.10) was compared with last measurement analysis (Table 5.6), the magnitudes of the parameter estimates were consistently lower in the former due to the imputation of wave 1 measurements causing a similar effect to that illustrated when baseline analyses were performed. However, due to the increase in the number of binary units available, the standard errors were markedly lower, albeit artificially so due to the imputed values being treated as 'known'. Whilst it is clearly desirable to maximise the number of available cases, this should be done in a manner which does not tend to bias the parameter estimates towards zero, nor give artificially low estimates of their standard errors. The multiple imputation strategy appears, at least partially, to satisfy these requirements. Many of the parameter estimates are of a similar magnitude to those obtained under last measurement analysis rather than under last measurement imputation. However, the effects of 'high dependence' (as opposed to 'full independence') for ADLs, dementia and age are noticeably stronger using multiple imputation than even last measurement analysis, although those for ADL(2) (those 'moderately dependent') and IO(2) ('demented') were slightly smaller. Relatively similar levels of estimated precision are to be expected under the multiple imputation and last measurement analysis approaches as, for 113 of the 140 additional binary units, no new covariate information was available as the non-response was at the unit screen rather than item level. The only additional observed covariates were wave 2 values which had previously been excluded due to missingness in other items for the remaining 27 units. The apparent decrease in precision is likely to be due to the approach to modelling for the imputations. Improvements in the imputation models should lead to an increase in precision and, potentially, less biased estimates of the risk factor effects.

It was shown in the simulations in Chapter 4 that, under screen-indepdendent event collection with data filled-in using last measurements or multiply imputed using a model-based approach, bias tended to be lower for the latter than the

former. The bias in the estimation of the ordinal risk factor parameters tended to be negative whereas that for the continuous risk factor could be positive or negative depending on the proportion of missing data and the mechanism causing the missingness. Therefore, in the absence of real information about the type of missing-data mechanism, it is clearly sensible to use various forms of sensitivity analysis by investigating the effect of different approaches to analysis and, where appropriate, modelling the non-response process and making inferences accordingly.

Considering the estimation of an optimal weight for survival under the two imputation approaches, there was little difference between the estimates of 172.8 pounds under last measurement imputation and 170.8 pounds under the model-based multiple imputation. Moreover, these values are very similar to that of 172.8 pounds obtained under the last measurement analysis. Whilst there was no statistically significant evidence of a weight-by-sex interaction in the model, the inclusion of this interaction was justified mainly on physiological grounds and then investigated for 'exact' survival times in Section 5.4.2. Under last measurement imputation, the optimal weight for survival for males was estimated at 177.1 pounds whilst that for females was estimated to be 163.0 pounds. Similarly, the estimated optimal weights for male and female survival under the model-based multiple imputation were estimated to be 177.2 pounds and 159.0 pounds respectively. When compared with either the corresponding estimates of optimal weight for survival for males and females combined given above, or with the sex-specific optimal weights for survival of 178.1 pounds for males and 168.2 pounds for females under last measurement analysis, it is clear that somewhat different inferences can be drawn depending on the approach taken to data analysis, although the precision of all of these estimates will be extremely low, as illustrated in Section 5.4.2. The differences between point estimates are especially large for females for whom the optimal weight for survival ranges from a value close to 179 pounds (12 stone 11 pounds) under last measurement analysis with

no weight-by-sex interaction to 159 pounds (11 stone 5 pounds) under the multiple imputation approach with a weight-by-sex interaction included in the survival model. These observations are consistent with the theory that the analysis technique will considerably affect estimates, particularly in situations where precision is low or missingness is not *completely at random.*

Term	Category/*Units*	Method	PARAMETER Estimate	PARAMETER 95% C.I.	HAZARD RATIO Estimate	HAZARD RATIO 95% C.I.
ADL(2)	Moderately independent	**LMI**	0.575	(0.296, 0.855)	1.778	(1.345, 2.351)
		MI	0.557	(0.245, 0.869)	1.745	(1.277, 2.385)
ADL(3)	Moderately dependent	**LMI**	0.755	(0.423, 1.086)	2.127	(1.527, 2.963)
		MI	0.798	(0.421, 1.175)	2.222	(1.524, 3.240)
ADL(4)	Highly dependent	**LMI**	1.239	(0.916, 1.562)	3.452	(2.499, 4.768)
		MI	1.376	(0.996, 1.756)	3.960	(2.708, 5.789)
IO(2)	Demented	**LMI**	0.635	(0.341, 0.928)	1.886	(1.407, 2.529)
		MI	0.845	(0.529, 1.160)	2.327	(1.698, 3.189)
WT	*10 Pounds*	**LMI**	-0.330	(-0.564, -0.097)	*	*
		MI	-0.429	(-0.706, -0.152)	*	*
WT2	*(10 Pounds)*^{2}	**LMI**	0.0096	(0.0018, 0.0173)	*	*
		MI	0.0126	(0.0035, 0.0216)	*	*
AGE	*Years*	**LMI**	0.0730	(0.0499, 0.0961)	1.076	(1.051, 1.101)
		MI	0.0926	(0.0652, 0.1200)	1.097	(1.067, 1.128)
SEX(2)	Female	**LMI**	-0.664	(-0.914, -0.414)	0.515	(0.401, 0.661)
		MI	-0.772	(-1.059, -0.484)	0.462	(0.347, 0.616)

* - Not appropriate due to presence of both linear and quadratic terms

Table 5.10: Comparison of estimates of parameters and hazard ratios for available case (AC) analysis using last measurement imputation (LMI) and model-based multiple imputation (MI) in proportional hazards model with interval-censored approach and screen-independent event collection. Comparison groups: Fully independent (ADL(1)); Non-demented (IO(1)); Male (SEX(1)).

5.6.2 Screen-dependent event collection

On treating the collection of death information as screen-dependent, the number of binary units fully available for analysis was reduced from 2458 to 2282 for last measurement imputation and from 2457 to 2254 for the model-based multiple imputation. These reductions were due to missing screen-dependent event information from the second and, in particular, the third waves and were only slightly compensated for by the inclusion of three individuals with a **1 . 1** screen profile plus a single individual with a **1 . 3** profile (see Table 5.2).

The parameter estimates and hazard ratios for both last measurement and model-based multiple imputation are given in Table 5.11. Under both approaches to imputation, the parameter estimates were of similiar magnitude to those obtained under last measurement imputation with screen-independent event collection and are therefore generally weaker than those obtained under multiple imputation with the screen-independent death recording. This would tend to indicate that the loss of information on the habitual refusers who, as described earlier, tend to be at a greater risk of death than those regularly screened, has a weakening effect on the parameter estimates. However, as it was shown in Chapter 4 that bias tended strengthen the apparent effects, especially for the continuous risk factor under both last measurement and multiple imputation techniques, albeit in circumstances where there were more measurements and hence a lesser problem with loss of information due to habitual refusers, it is unclear what the combined effect of these different forms of incompleteness will have on the estimation.

The non-sex-specific estimates of the optimal weight for survival were found to be 173.6 pounds and 172.9 pounds under last measurement imputation and model-based multiple imputation respectively. These values are consistent with estimates obtained under earlier approaches. When differential effects of weight

for males and females were included in the survival model, similar estimates of optimal weight for survival were obtained under the two imputation strategies. These estimates for females were 166.0 pounds and 165.0 pounds and those for males were 177.1 pounds and 176.9 pounds. For both sexes, the lower estimate was when the missing data were filled-in using model-based multiple imputation. These sex-specific estimates of the optimal weight for survival further illustrate the consistency of the values for males and the variability of the values for females under the different approaches to handling the incompleteness of the data.

Term	Category/*Units*	Method	Parameter estimate	Hazard ratio estimate
ADL(2)	Moderately independent	**LMI** **MI**	0.584 0.531	1.794 1.701
ADL(3)	Moderately dependent	**LMI** **MI**	0.793 0.778	2.211 2.178
ADL(4)	Highly dependent	**LMI** **MI**	1.243 1.212	3.467 3.361
IO(2)	Demented	**LMI** **MI**	0.585 0.602	1.796 1.826
WT	*10 Pounds*	**LMI** **MI**	-0.337 -0.334	* *
WT2	*(10 Pounds)*^{2}	**LMI** **MI**	0.0097 0.0097	* *
AGE	*Years*	**LMI** **MI**	0.0679 0.0715	1.070 1.074
SEX(2)	Female	**LMI** **MI**	-0.631 -0.632	0.532 0.532

* - Not appropriate due to presence of both linear and quadratic terms

Table 5.11: Comparison of estimates of parameters and hazard ratios for available case (AC) analysis using last measurement imputation (LMI) and model-based multiple imputation (MI) in proportional hazards model with interval-censored approach and screen-dependent event collection. Comparison groups: Fully independent (ADL(1)); Non-demented (IO(1)); Male (SEX(1)).

5.7 Summary

It has been shown in this chapter that estimates of the parameters of the proportional hazards model can vary quite considerably depending on the approach taken to analysing the event history arising from a screening study. Whilst the results of the simulations in Chapter 4 showed that bias could strengthen the apparent effects in some circumstances, especially when there was screen-dependent event collection, it is likely that the techniques leading to the parameter estimates closer to zero are the more biased and, furthermore, it is believed that the more sophisticated imputation techniques are preferable to those which involve using only complete or available cases, particularly if the missing data mechanism is MAR. As the precise form of the missing data mechanism(s) influencing the incompleteness of the data set analysed here is unknown, but believed to be MAR (or possible weakly non-ignorable), some account should be taken of the missing data to avoid the likely under-representation of cases with extreme survival prognosis and the resultant effects on estimation under a proportional hazards model.

Whilst it was impractical to investigate the form of the non-response mechanism, especially that relating to unit screen non-response, it should be noted that the models fitted and used to multiply impute the data are under the assumption that the missingness is ignorable. If this is not the case, there will be biased estimation of the posterior distribution of the parameters of the models for WT, IO and ADL. This will, in turn, result in the introduction of bias into the parameter estimates for the survival model. An additional stage in this investigation could, therefore, be the use of sensitivity analysis in conjunction with the investigation of likely forms of non-response mechanism.

Differences between parameter estimates under the alternative approaches were, on the whole, relatively small. The effects of weight, dementia and high dependence for performing ADLs tended to be those whose estimates varied to the

greatest degree. The effect of weight on survival showed a greater, but non-significant, estimated degree of curvature in the complete case and model-based multiple imputation methods with the screen-independent event collection than under the other approaches to analysis. However, as with many of the other covariates collected at the screens, whether the relationship is, at least partially, causal is open to question. Likewise, the estimated effect of dementia on mortality is highly dependent on the technique used for analysis. In fact the size of the parameter estimate under the model-based multiple imputation (0.845) is almost twice that when baseline covariate measurements were used under ACA (0.443) with screen-independent death collection and interval-censored times. Moreover, the degree of overlap of the 95% confidence intervals for the parameter is only around one third of each interval, indicating that inferences can be moderately different under alternative approaches to the modelling of survival using this data set. This, of course is despite the existence of only a single update of covariates and a follow-up period of around five years. Similarly, the effect of high dependence for ADLs varies to a reasonable degree under the different analyses. In this case, the 95% confidence intervals under the two situations compared above overlap by a little over one-half of each. As both high dependence for performing ADLs and dementia could be viewed as intermediate states between healthy life and death and those either refusing a screen or not providing measurements of variables such as weight (and demi-span) tend to be amongst the less healthy, it is maybe not surprising that it is *their* effects on mortality for which there is the greatest dependence on the analysis technique. It is worth noting that, whilst it is possible that these effects may be over-estimated by the multiple imputation approach, particularly given the poor overall fit of the transition proportional odds model for imputing for missing values of dependence for performing ADLs, it is believed that the degree of under-estimation is real and reasonably substantial when baseline measurements alone are used as covariates. This re-inforces the recommendation that updated covariate models should be used whenever it is believed that there are time-dependent effects of the risk factors.

Chapter 6

Discussion

Incomplete data can present a serious problem in event history analysis. It is advisable to limit the degree of incompleteness by suitable choice of design. However, removal of all incompleteness likely to introduce bias into the estimation of model parameters is often impractical. In particular, participation in screening studies is usually optional and therefore ensuring complete data are collected presents insurmountable practical difficulties. In addition to the failure to take up the offer of a screen, a common form of incompleteness in screening studies is due to events only being recorded as having occurred between two attended screens. In the case of two-state processes and proportionality of hazards, these can be handled by either existing methods for the analysis of grouped or interval-censored survival times (Prentice and Gloeckler (1978), Finkelstein (1986)) or by imputation of exact times within the interval (Dorey, Little and Schenker (1993), Little (1992)) followed by the use of standard survival analysis techniques for exact times (e.g. Cox (1972)).

An alternative approach discussed briefly in Chapters 2 and 4 is by using stochastic differential equations to model the changes in risk factors over time. This formulation was derived from a biological model of human ageing and was first proposed by Woodbury and Manton (1977). However, as well as being some-

what complex in formulation, this model relies on the assumption of multivariate normality of the risk factors and will therefore not be appropriate unless the screening process consists exclusively of physiological *measurements* (e.g. blood pressure, cholesterol level, heart rate). It has been stated that "where a physical examination or tests form part of an epidemiologic survey, participation is usually lower than *interview only* surveys" and that "those who undergo examination are more likely to be younger" (Arthur, Clarke, Donaldson, Jagger, unpublished manuscript). This provides a justification for limiting or excluding non-question-based assessments as part of a health screening of an elderly population. In fact, in the annual screening of the elderly discussed in Chapter 5, the two variables with the highest levels of unit non-response are the physiological measurements which require physical movement of the subjects, with blood pressure readings having far lower rates of non-response than the measurements of weight or demi-span. As assessment is mainly via subjective responses in the Melton Mowbray screening study (not all of which were coded for the data set used in Chapter 5), the assumption of multivariate normality will not be appropriate and alternative methodologies should be used. Whilst a number of alternative approaches have been proposed to handle categorical data in modelling the health and mortality of elderly populations (Manton, Singer and Suzman, 1993), these are beyond the scope of this thesis.

Missing data in the covariates, caused by failure to take up the offer of a screen or item missingness amongst those screened, is a serious problem, more so than the incompleteness of information on failure times. Likelihood-based methods (e.g. Schluchter and Jackson (1989)) are recommended (Little, 1992b) although such methods apply only when the missing-data mechanism is ignorable and are usually very difficult to implement due to the lack of availability in the present versions of the major software packages and the extensive programming required to implement them otherwise. An alternative to this is imputation of the missing covariate data followed by analysis of the completed data set using standard

methods or adaptations thereof.

To quote Little (1992b) in his conclusions on the analysis of incomplete data in event history analysis "... imputation-based methods also deserve study, particularly if multiple imputation is applied to reflect the uncertainty in the imputation process". A major objective of this thesis has been to investigate methods for imputing missing covariate values in a particular form of event history analysis corresponding to screening studies. A secondary and associated line of investigation relates to the use of time-varying rather than fixed baseline covariates in these studies. Given the widespread use of baseline covariates, even when the real interest is in updated covariate survival models or even in 'strong models' relating the survival process to the evolution of the risk factors, the investigation of the effect of using baseline covariates rather those updated at follow-up was felt to be important. This effect was investigated through simulations in Chapter 4 and then, in Chapter 5, on an on-going screening study of the elderly of Melton Mowbray which provides an addition to the series of studies in to the health and well-being of the elderly population of this Leicestershire market town carried out over the last fifteen years. All previous publications based on these studies and based on survival have used baseline rather than updated covariate models and so it is likely that any estimated effects of risk factors on mortality have been underestimated due to the weakening of effects of baseline covariates as the follow-up time increases.

Chapter 3 investigated random coefficient models for growth curve modelling and concentrated on selection models, providing a useful insight into these models, especially that proposed by Wu and Carroll (1988). However, these models showed little potential for extension to event history analysis. Wu and Carroll's pseudo-maximum likelihood estimator based on a probit dropout model (PPMLE) proved computationally complex with a large number of parameters which required joint estimation in the maximisation process. The estimation of the parameters relat-

ing to informative dropout processes with a range of intercept and gradient effects was relatively unbiased, with bias ranging from 0.9% to 4.0% for that relating to the gradient parameter and from 0.5% to 10.0% for that relating to the intercept parameter, although in only one out of six sets of simulations was the percentage bias for the latter parameter over 5.5%. Although this illustrates that PPMLE performs adequately in the estimation of parameters of the dropout or survival process, these parameters are not the usual log-hazard ratio parameters but relate this process to the parameters of the repeated measures model. Moreover, these parameters may have a varied interpretation depending on their nature (and the nature of the covariate if the model were extended) and would prove extremely difficult to explain to non-statisticians. There would also be problems in extending the technique to cover additional covariates, particularly those of non-continuous form, and to dropout processes not of the probit form required for an analytic evaluation of the marginal likelihood function. However, on the positive side, the results of the simulations given in Chapter 3 and those of Wu and Bailey (1988) indicated that the performance of PPMLE may not be adversely affected by deviations from a probit form of dropout, especially those of Cox (or complementary-log-log) form. Due to the problems associated with this model, it was decided that further work into potential extensions of random coefficient selection models to apply to typical forms of event history data was likely to prove unfruitful.

Several methods presently available for the analysis of grouped or interval-censored survival data were discussed in Chapter 4. In particular, the methods of Prentice and Gloeckler (1978) and Finkelstein (1986) have the advantages of a reasonable degree of simplicity, a natural interpretation of the parameters of interest in terms of (log-)hazard ratios and, in the case of the former, ready availability in generalised linear model routines in a number of statistical software packages. Moreover, it was found that the desirable feature of time-varying covariates which was present in Prentice and Gloeckler's model and that of irregularly-spaced follow-up times for survival present in Finkelstein's model could be incorporated

into a single model formulation which was termed *'adapted Finkelstein'*. However, missing data in the covariates still presented a problems and, to allow the 'adapted Finkelstein' to be applied to such data, model-based multiple imputation techniques were proposed. With incomplete repeated measures data which incorporate both continuous and categorical variables, a very limited range of multivariate techniques is available for modelling and these methods apply only under the assumption of a MAR process. One technique which, in practice, could be applied to the imputation of data of this form is that introduced by Little and Schluchter (1985) and, in fact, this is one application of the model suggested by the authors.

Although it is recognised that efficient and unbiased imputation methods are the ideal, there is a strong argument for using simpler imputation methods than that described above, providing any deterioration in performance is minor. The imputation methods considered in Chapters 4 and 5 of this thesis are relatively simple univariate models. First-order autoregressive models were used for modelling repeated realisations of continuous variables and Markov chain regression models were used for modelling ordinal variables. One advantage of all the models discussed here is their potential for extension to non-ignorable missing data situations by suitable selection from the predictive density via the chosen missing data mechanism based on the value to be imputed (or some function thereof). This is an extension of the approach used in Chapters 4 and 5 and would merit further investigation although, for the Melton data set used for illustration in Chapter 5, it is *believed* that the non-response mechanism or mechanisms are at least approximately MAR.

The results of the simulations described in Chapter 4 show that the performance of the first-order autoregressive process used for imputation of missing values of a repeatedly measured continuous variable was acceptable in terms of bias, particularly with low percentages (up to 10%) of data missing. It was also

preferable to the cold-deck imputation method using the last recorded value for various forms of missing data mechanism (MCAR, covariate-based MAR, value-based non-ignorable and difference-based non-ignorable). However, the performance of the multiple imputation approach deteriorated considerably when an ordinal repeatedly-measured variable was also included, the estimation of the parameters relating to the effect of the ordinal variable being especially susceptible to bias. In all these cases, both screen-independent and screen-dependent forms of event collection were investigated. The relative deterioration of the performance of the model-based multiple imputation technique with increasing percentages of missing data was generally greater with screen-independent event collection, particularly when compared with the corresponding deterioration of the imputation of the previously recorded value.

When the methods derived and applied through simulations in Chapter 4 were then applied to survival data arising from the Melton Mowbray screening study of the elderly in Chapter 5, it was shown that, whilst some estimates of effects were of similar magnitude when traditional complete or available case methods and model-based multiple imputation were used, others were considerably different. Whilst both item and unit screen non-response are likely to be approximately missing at random (MAR), ignoring information on recorded covariates, which happens when there is either unit screen non-response at the end of an interval or item missing data, will tend to force the non-response into a non-ignorable state. This means that bias is introduced into the likelihood-based estimation process and the use of appropriate imputation methods will tend to reduce this bias because they preserve a greater proportion of the cases. In particular, those highly dependent on others for performing their activities of daily living (ADLs) or classified as demented at the previous screen were more likely to be item missing, particularly on the weight variable which was included in all the survival models. Ignoring these cases entirely will tend to lead to a relative under-representation of those likely to die *in a later inteval between screens* but with *present* serious

disability or dementia. Even a reasonable imputation stategy will tend to recover most of this information, especially given that the scope for improvement from clincially-diagnosed dementia and from a highly dependent state for ADLs will be extremely limited. This is borne out in the results from Chapter 5 where the size of the estimated effects of dementia and high dependence for ADLs are both amongst those noticeably greater under model-based multiple imputation than complete or available case methods.

As emphasised be Diggle, Liang and Zeger (1994), it is highly important to correctly specify the form of the model when considering repeated outcomes of dichotomous or ordinal responses. A transition rather than random effects model would seem appropriate for the Melton study as parameters at a population rather than individual level are required for the imputation model, particularly given that the imputations made in Chapter 5 were always based on a single previous measurement for that individual, this often being at the only screen they underwent. However, the proportional odds model used did not appear to be the most appropriate model for transitions in states of dependence for performing the activities of daily living assessed given the lack of proportionality found in the odds. Further work would be required either to investigate other forms of model for the ordinal responses or, alternatively, multivariate models such as that of Little and Schluchter (1985).

In terms of the use of baseline covariate rather than updated covariate models for survival data where the true change of state process relates to recent realisations of risk factors, the simulations support the conclusions of Altman and de Stavola (1994) that, if the covariate effects are constant over the follow-up period (as is expected to be true for the Melton Mowbray data), the effects estimated from an updated covariate model will be larger in absolute value 'because of the time delay of the effect of entry values'. It is likely, therefore, that the effects of baseline covariates will decline over time and that the estimates of hazard

ratios for the risk factors for mortality reported in earlier publications (Jagger and Clarke (1988), Jagger and Sutton (1991)) relating to the survival of the 1981 cohort of the population of Melton Mowbray aged 75 years and over (based solely on baseline values of the risk factors) will be biased towards unity.

It would be sensible, if practical in the future, to check that the true form of examination scheme for the Melton Mowbray screening study corresponds to a non-informative process. In Section 2.1.2, four simple models for examination scheme (Grüger et al., 1991) are outlined, of which 'patient self-selection' is clearly acting here. This form of process will tend to be non-ignorable although it may, as suspected here, be approximately MAR. Whilst this may be a reasonable assumption to make, it would clearly be desirable to validate it by following-up those missing offered screens. Alternatively, a proposal of one or more plausible patient self-selection mechanisms (plus item non-response mechanisms if deemed potentially non-ignorable) would enable further sensitivity analysis to be performed.

Whilst the imputation techniques employed in Chapter 5 have reduced the amount of missing data, they have not led to a fully completed data set. For this to be attained, additional imputation strategies need to be investigated and then employed. Although the left-truncation, due mainly to those aged under 75 when the first wave population was defined, is not a problem, imputation of covariates for a missed first-offered screen is a major methodological hurdle to overcome in order to complete the data set. Choice of an appropriate strategy for this will depend greatly on the determination of the true form of the patient self-selection mechanism. Otherwise, potential bias of an indeterministic amount will remain in the estimates of the hazard ratios for the risk factors. Also, treating each wave as a single interval is somewhat simplistic. An alternative view would be to consider the twelve areas of Melton separately, define twelve separate points of follow-up for each wave corresponding to the first post-dates for each area and use 'adapted Finkelstein' for analysis. This would treat removals within the town

in a more appropriate manner than the approach used in Chapter 5 but, unless there is variation in the ordering of the areas screened across waves, the limited number of removals is likely to have little impact on inferences. Moreover, there would then be the potential for thirty-five (rather than two) nuisance parameters for the time periods based on these three waves and, unless some parametric assumption is made about the baseline hazard function, the total number of model parameters requiring estimation quickly becomes unmanageable as the number of waves increases. An additional problem is the asymptotic results relating to likelihood estimation which rely on the number of parameters not increasing with the number of observations. Whilst this is a minor problem in the form of analysis performed in Chapter 5 for which there would be one additional parameter per screen, the problem will be more substantial for this alternative suggestion as there would be twelve additional parameters per screen. It is therefore recommended that, whilst 'adapted Finkelstein' offers the potential for this approach to be used, it should not be adopted for the modelling of the Melton Mowbray study of the elderly.

To summarise, it has been shown that the potential exists for extending the method of Finkelstein (1985) to include time-dependent covariates of the form described by Prentice and Gloeckler (1978) in which updates are performed at the follow-up times. However, unless these follow-up measurements are at regular times for all individuals, the potential exists for inconsistent assessment and interpretation of the parameters. In particular, in screening studies with a regular schedule of screens in which the real interest is in the effect of the values of risk factors at a particular screen time on survival to the next scheduled screen, missed screens will lead to failure to update the risk factor values. As shown in Chapter 4, this will lead to a weakening of the true effects, of the same form as caused by the inappropriate use of baseline covariates. One solution for such studies is the multiple imputation of unrecorded updates of risk factors. Whilst some investigation have been performed here into model-based methods for imputing risk

factors of mixed continuous and ordinal form, further work is required to widen the choice of imputation models and to select the most appropriate model for the repeated measurements from the Melton Mowbray screening study of the elderly.

Further work motivated by the research presented in this thesis includes the extension to cover non-ignorable missing data mechanisms. For the Melton Mowbray screening study, this will be particularly important if the collection of information on a random subset of refusers at a particular screen were to show that unit screen non-response is not a MAR process. Additionally, as outlined earlier, it would be of interest to consider methods for imputation of risk factor values for individuals refusing a screen prior to their first acceptance of a screen in association with screen-independent event collection, especially as the time to event may be linked to unit screen non-response. Finally, the extension of the scope of these techniques to cover multi-state processes with a single absorbing state, for example the three-state process for an elderly population in which the three states are:

- alive and non-demented;
- alive and demented;
- dead;

would be of considerable interest. Whilst such methods exist in cases of complete data (Hsieh, Crowley and Tormey (1983), Collett (1994)), further work would be required to consider the extension to interval-censored state transition times (or mixed 'exact' and interval-censored times) in the presence of incomplete updates of risk factors.

To close, whilst the Melton Mowbray study has provided much of the motivation for the research into the analytic techniques described here, resource constraints

in performing the screens have led to the third wave only recently being completed. There has therefore been, at most, one *update* of risk factors available for use in conjunction with interval-censored survival times. The limited amount of imputation required for these updates using techniques proposed in Chapter 4 has been a problem in the comparison of different approaches to modelling the state transition data from this study. The real test of these and alternative imputation-based models will come when two or three more screens have been completed.

Appendix A

Derivation of the likelihood function for PPMLE

An individual's contribution to the marginal likelihood is given by

$$L_i = \frac{D}{(2\pi)^2 \mid \boldsymbol{C}_{1i} \mid^{1/2} \mid \boldsymbol{\Sigma}_\beta \mid^{1/2}} \int \exp\left[-\frac{1}{2}\left(\underline{\hat{\boldsymbol{\beta}}}_i - \underline{\boldsymbol{\beta}}_i\right)^T \boldsymbol{C}_{1i}{}^{-1}\left(\underline{\hat{\boldsymbol{\beta}}}_i - \underline{\boldsymbol{\beta}}_i\right)\right.$$

$$\left. -\frac{1}{2}\left(\underline{\boldsymbol{\beta}}_i - \underline{\boldsymbol{B}}_k\right)^T \boldsymbol{\Sigma}_\beta{}^{-1}\left(\underline{\boldsymbol{\beta}}_i - \underline{\boldsymbol{B}}_k\right)\right]$$

$$\times \prod_{j=2}^{J}\left[\left(\Phi\left(\underline{\boldsymbol{\alpha}}^T\underline{\boldsymbol{\beta}}_i + \alpha_{0j}\right) - \Phi\left(\underline{\boldsymbol{\alpha}}^T\underline{\boldsymbol{\beta}}_i + \alpha_{0j-1}\right)\right)^{Z(i,j-1)}\right]$$

$$\times \left(1 - \Phi\left(\underline{\boldsymbol{\alpha}}^T\underline{\boldsymbol{\beta}}_i + \alpha_{0J}\right)\right)^{(1-\sum_j Z(i,j))} d\underline{\boldsymbol{\beta}}_i \qquad i\,\varepsilon\,k \quad \text{and} \quad k = 1, 2\,.$$

Now

$$\exp\left[-\frac{1}{2}\left(\hat{\underline{\boldsymbol{\beta}}}_i - \underline{\boldsymbol{\beta}}_i\right)^T \boldsymbol{C}_{1i}{}^{-1}\left(\hat{\underline{\boldsymbol{\beta}}}_i - \underline{\boldsymbol{\beta}}_i\right) - \frac{1}{2}\left(\underline{\boldsymbol{\beta}}_i - \underline{\boldsymbol{B}}_k\right)^T \boldsymbol{\Sigma}_\beta{}^{-1}\left(\underline{\boldsymbol{\beta}}_i - \underline{\boldsymbol{B}}_k\right)\right]$$

$$= \exp\left[-\frac{1}{2}\left(\underline{\boldsymbol{\beta}}_i{}^T \boldsymbol{C}_{1i}{}^{-1}\underline{\boldsymbol{\beta}}_i - \underline{\boldsymbol{\beta}}_i{}^T \boldsymbol{C}_{1i}{}^{-1}\hat{\underline{\boldsymbol{\beta}}}_i - \hat{\underline{\boldsymbol{\beta}}}_i{}^T \boldsymbol{C}_{1i}{}^{-1}\underline{\boldsymbol{\beta}}_i + \hat{\underline{\boldsymbol{\beta}}}_i{}^T \boldsymbol{C}_{1i}{}^{-1}\hat{\underline{\boldsymbol{\beta}}}_i\right.\right.$$

$$\left.\left.+\underline{\boldsymbol{\beta}}_i{}^T \boldsymbol{\Sigma}_\beta{}^{-1}\underline{\boldsymbol{\beta}}_i - \underline{\boldsymbol{B}}_k{}^T \boldsymbol{\Sigma}_\beta{}^{-1}\underline{\boldsymbol{\beta}}_i - \underline{\boldsymbol{\beta}}_i{}^T \boldsymbol{\Sigma}_\beta{}^{-1}\underline{\boldsymbol{B}}_k + \underline{\boldsymbol{B}}_k{}^T \boldsymbol{\Sigma}_\beta{}^{-1}\underline{\boldsymbol{B}}_k\right)\right]$$

$$= \exp\left[-\frac{1}{2}\left(\underline{\boldsymbol{\beta}}_i{}^T\left(\boldsymbol{C}_{1i}{}^{-1} + \boldsymbol{\Sigma}_\beta{}^{-1}\right)\underline{\boldsymbol{\beta}}_i - \underline{\boldsymbol{\beta}}_i{}^T\left(\boldsymbol{C}_{1i}{}^{-1}\hat{\underline{\boldsymbol{\beta}}}_i + \boldsymbol{\Sigma}_\beta{}^{-1}\underline{\boldsymbol{B}}_k\right)\right.\right.$$

$$\left.\left.-\left(\hat{\underline{\boldsymbol{\beta}}}_i{}^T \boldsymbol{C}_{1i}{}^{-1} + \underline{\boldsymbol{B}}_k{}^T \boldsymbol{\Sigma}_\beta{}^{-1}\right)\underline{\boldsymbol{\beta}}_i + \left(\hat{\underline{\boldsymbol{\beta}}}_i{}^T \boldsymbol{C}_{1i}{}^{-1}\hat{\underline{\boldsymbol{\beta}}}_i + \underline{\boldsymbol{B}}_k{}^T \boldsymbol{\Sigma}_\beta{}^{-1}\underline{\boldsymbol{B}}_k\right)\right)\right]$$

$$= \exp\left[-\frac{1}{2}\left(\underline{\boldsymbol{\beta}}_i{}^T \boldsymbol{C}_{3i}{}^{-1}\underline{\boldsymbol{\beta}}_i - \underline{\boldsymbol{\beta}}_i{}^T \underline{\boldsymbol{d}}_{ik} - \underline{\boldsymbol{d}}_{ik}{}^T \underline{\boldsymbol{\beta}}_i + \left(\hat{\underline{\boldsymbol{\beta}}}_i{}^T \boldsymbol{C}_{1i}{}^{-1}\hat{\underline{\boldsymbol{\beta}}}_i + \underline{\boldsymbol{B}}_k{}^T \boldsymbol{\Sigma}_\beta{}^{-1}\underline{\boldsymbol{B}}_k\right)\right)\right] ,$$

where

$$\underline{\boldsymbol{d}}_{ik} = \boldsymbol{C}_{1i}{}^{-1}\hat{\underline{\boldsymbol{\beta}}}_i + \boldsymbol{\Sigma}_\beta{}^{-1}\underline{\boldsymbol{B}}_k \quad \text{and} \quad \boldsymbol{C}_{3i} = (\boldsymbol{C}_{1i}{}^{-1} + \boldsymbol{\Sigma}_\beta{}^{-1})^{-1} ,$$

$$= \exp\left[-\frac{1}{2}\left(\left(\underline{\boldsymbol{\beta}}_i - \boldsymbol{C}_{3i}\underline{\boldsymbol{d}}_{ik}\right)^T \boldsymbol{C}_{3i}{}^{-1}\left(\underline{\boldsymbol{\beta}}_i - \boldsymbol{C}_{3i}\underline{\boldsymbol{d}}_{ik}\right) - \underline{\boldsymbol{d}}_{ik}{}^T \boldsymbol{C}_{3i}\underline{\boldsymbol{d}}_{ik}\right.\right.$$

$$\left.\left.+\hat{\underline{\boldsymbol{\beta}}}_i{}^T \boldsymbol{C}_{1i}{}^{-1}\hat{\underline{\boldsymbol{\beta}}}_i + \underline{\boldsymbol{B}}_k{}^T \boldsymbol{\Sigma}_\beta{}^{-1}\underline{\boldsymbol{B}}_k\right)\right]$$

$$= \exp\left[-\frac{1}{2}\left(\left(\underline{\boldsymbol{\beta}}_i - \boldsymbol{C}_{3i}\underline{\boldsymbol{d}}_{ik}\right)^T \boldsymbol{C}_{3i}{}^{-1}\left(\underline{\boldsymbol{\beta}}_i - \boldsymbol{C}_{3i}\underline{\boldsymbol{d}}_{ik}\right)\right.\right.$$

$$- \left(\boldsymbol{C}_{1i}{}^{-1}\hat{\underline{\boldsymbol{\beta}}}_i + \boldsymbol{\Sigma}_\beta{}^{-1}\underline{\boldsymbol{B}}_k\right)^T \left(\boldsymbol{C}_{1i}{}^{-1} + \boldsymbol{\Sigma}_\beta{}^{-1}\right)^{-1}\left(\boldsymbol{C}_{1i}{}^{-1}\hat{\underline{\boldsymbol{\beta}}}_i + \boldsymbol{\Sigma}_\beta{}^{-1}\underline{\boldsymbol{B}}_k\right)$$

$$\left.\left. + \hat{\underline{\boldsymbol{\beta}}}_i{}^T \boldsymbol{C}_{1i}{}^{-1}\hat{\underline{\boldsymbol{\beta}}}_i + \underline{\boldsymbol{B}}_k{}^T \boldsymbol{\Sigma}_\beta{}^{-1}\underline{\boldsymbol{B}}_k\right)\right]$$

$$= \exp\left[-\frac{1}{2}\left(\left(\underline{\boldsymbol{\beta}}_i - \boldsymbol{C}_{3i}\underline{\boldsymbol{d}}_{ik}\right)^T \boldsymbol{C}_{3i}{}^{-1}\left(\underline{\boldsymbol{\beta}}_i - \boldsymbol{C}_{3i}\underline{\boldsymbol{d}}_{ik}\right)\right.\right.$$

$$- \hat{\underline{\boldsymbol{\beta}}}_i{}^T \boldsymbol{C}_{1i}{}^{-1}\left(\boldsymbol{C}_{1i}{}^{-1} + \boldsymbol{\Sigma}_\beta{}^{-1}\right)^{-1}\left(\boldsymbol{C}_{1i}{}^{-1}\hat{\underline{\boldsymbol{\beta}}}_i + \boldsymbol{\Sigma}_\beta{}^{-1}\underline{\boldsymbol{B}}_k\right) + \hat{\underline{\boldsymbol{\beta}}}_i{}^T \boldsymbol{C}_{1i}{}^{-1}\hat{\underline{\boldsymbol{\beta}}}_i$$

$$\left.\left. + \underline{\boldsymbol{B}}_k{}^T \boldsymbol{\Sigma}_\beta{}^{-1}\underline{\boldsymbol{B}}_k - \underline{\boldsymbol{B}}_k{}^T \boldsymbol{\Sigma}_\beta{}^{-1}\left(\boldsymbol{C}_{1i}{}^{-1} + \boldsymbol{\Sigma}_\beta{}^{-1}\right)^{-1}\left(\boldsymbol{C}_{1i}{}^{-1}\hat{\underline{\boldsymbol{\beta}}}_i + \boldsymbol{\Sigma}_\beta{}^{-1}\underline{\boldsymbol{B}}_k\right)\right)\right]$$

$$= \exp\left[-\frac{1}{2}\left(\left(\underline{\boldsymbol{\beta}}_i - \boldsymbol{C}_{3i}\underline{\boldsymbol{d}}_{ik}\right)^T \boldsymbol{C}_{3i}{}^{-1}\left(\underline{\boldsymbol{\beta}}_i - \boldsymbol{C}_{3i}\underline{\boldsymbol{d}}_{ik}\right)\right.\right.$$

$$- \hat{\underline{\boldsymbol{\beta}}}_i{}^T \left(\boldsymbol{I} + \boldsymbol{\Sigma}_\beta{}^{-1}\boldsymbol{C}_{1i}\right)^{-1}\left(\boldsymbol{C}_{1i}{}^{-1}\hat{\underline{\boldsymbol{\beta}}}_i + \boldsymbol{\Sigma}_\beta{}^{-1}\underline{\boldsymbol{B}}_k\right) + \hat{\underline{\boldsymbol{\beta}}}_i \boldsymbol{C}_{1i}{}^{-1}\hat{\underline{\boldsymbol{\beta}}}_i$$

$$\left.\left. + \underline{\boldsymbol{B}}_k{}^T \boldsymbol{\Sigma}_\beta{}^{-1}\underline{\boldsymbol{B}}_k - \underline{\boldsymbol{B}}_k{}^T \left(\boldsymbol{C}_{1i}{}^{-1}\boldsymbol{\Sigma}_\beta + \boldsymbol{I}\right)^{-1}\left(\boldsymbol{C}_{1i}{}^{-1}\hat{\underline{\boldsymbol{\beta}}}_i + \boldsymbol{\Sigma}_\beta{}^{-1}\underline{\boldsymbol{B}}_k\right)\right)\right]$$

$$= \exp\left[-\frac{1}{2}\left(\left(\underline{\boldsymbol{\beta}}_i - \boldsymbol{C}_{3i}\underline{\boldsymbol{d}}_{ik}\right)^T \boldsymbol{C}_{3i}{}^{-1}\left(\underline{\boldsymbol{\beta}}_i - \boldsymbol{C}_{3i}\underline{\boldsymbol{d}}_{ik}\right)\right.\right.$$

$$- \hat{\underline{\boldsymbol{\beta}}}_i{}^T\left(\boldsymbol{I} + \boldsymbol{\Sigma}_\beta{}^{-1}\boldsymbol{C}_{1i}\right)^{-1}\boldsymbol{C}_{1i}{}^{-1}\hat{\underline{\boldsymbol{\beta}}}_i + \hat{\underline{\boldsymbol{\beta}}}_i{}^T\boldsymbol{C}_{1i}{}^{-1}\hat{\underline{\boldsymbol{\beta}}}_i + \underline{\boldsymbol{B}}_k{}^T\boldsymbol{\Sigma}_\beta{}^{-1}\underline{\boldsymbol{B}}_k$$

$$- \hat{\underline{\boldsymbol{\beta}}}_i{}^T\left(\boldsymbol{I} + \boldsymbol{\Sigma}_\beta{}^{-1}\boldsymbol{C}_{1i}\right)^{-1}\boldsymbol{\Sigma}_\beta{}^{-1}\underline{\boldsymbol{B}}_k - \underline{\boldsymbol{B}}_k{}^T\left(\boldsymbol{C}_{1i}{}^{-1}\boldsymbol{\Sigma}_\beta + \boldsymbol{I}\right)^{-1}\boldsymbol{C}_{1i}{}^{-1}\hat{\underline{\boldsymbol{\beta}}}_i$$

$$\left.\left. - \underline{\boldsymbol{B}}_k{}^T\left(\boldsymbol{C}_{1i}{}^{-1}\boldsymbol{\Sigma}_\beta + \boldsymbol{I}\right)^{-1}\boldsymbol{\Sigma}_\beta{}^{-1}\underline{\boldsymbol{B}}_k\right)\right]$$

$$= \exp\left[-\frac{1}{2}\left(\left(\underline{\boldsymbol{\beta}}_i - \boldsymbol{C}_{3i}\underline{\boldsymbol{d}}_{ik}\right)^T \boldsymbol{C}_{3i}{}^{-1}\left(\underline{\boldsymbol{\beta}}_i - \boldsymbol{C}_{3i}\underline{\boldsymbol{d}}_{ik}\right) - \hat{\underline{\boldsymbol{\beta}}}_i{}^T\left(\boldsymbol{C}_{1i} + \boldsymbol{C}_{1i}\boldsymbol{\Sigma}_\beta{}^{-1}\boldsymbol{C}_{1i}\right)^{-1}\hat{\underline{\boldsymbol{\beta}}}_i\right.\right.$$

$$+ \hat{\underline{\boldsymbol{\beta}}}_i{}^T\boldsymbol{C}_{1i}{}^{-1}\hat{\underline{\boldsymbol{\beta}}}_i + \underline{\boldsymbol{B}}_k{}^T\boldsymbol{\Sigma}_\beta{}^{-1}\underline{\boldsymbol{B}}_k - \hat{\underline{\boldsymbol{\beta}}}_i{}^T\left(\boldsymbol{\Sigma}_\beta + \boldsymbol{C}_{1i}\right)^{-1}\underline{\boldsymbol{B}}_k$$

$$\left.\left. - \underline{\boldsymbol{B}}_k{}^T\left(\boldsymbol{\Sigma}_\beta + \boldsymbol{C}_{1i}\right)^{-1}\hat{\underline{\boldsymbol{\beta}}}_i - \underline{\boldsymbol{B}}_k{}^T\left(\boldsymbol{\Sigma}_\beta\boldsymbol{C}_{1i}{}^{-1}\boldsymbol{\Sigma}_\beta + \boldsymbol{\Sigma}_\beta\right)^{-1}\underline{\boldsymbol{B}}_k\right)\right]$$

$$= \exp\left[-\frac{1}{2}\left(\left(\underline{\boldsymbol{\beta}}_i - \boldsymbol{C}_{3i}\underline{\boldsymbol{d}}_{ik}\right)^T \boldsymbol{C}_{3i}{}^{-1}\left(\underline{\boldsymbol{\beta}}_i - \boldsymbol{C}_{3i}\underline{\boldsymbol{d}}_{ik}\right) - \hat{\underline{\boldsymbol{\beta}}}_i{}^T\left(\boldsymbol{\Sigma}_\beta + \boldsymbol{C}_{1i}\right)^{-1}\underline{\boldsymbol{B}}_k\right.\right.$$

$$- \hat{\underline{\boldsymbol{\beta}}}_i{}^T\left\{\left(\boldsymbol{C}_{1i} + \boldsymbol{C}_{1i}\boldsymbol{\Sigma}_\beta{}^{-1}\boldsymbol{C}_{1i}\right)^{-1} - \boldsymbol{C}_{1i}{}^{-1}\right\}\hat{\underline{\boldsymbol{\beta}}}_i$$

$$\left.\left. - \underline{\boldsymbol{B}}_k{}^T\left\{\left(\boldsymbol{\Sigma}_\beta\boldsymbol{C}_{1i}{}^{-1}\boldsymbol{\Sigma}_\beta + \boldsymbol{\Sigma}_\beta\right)^{-1} - \boldsymbol{\Sigma}_\beta{}^{-1}\right\}\underline{\boldsymbol{B}}_k - \underline{\boldsymbol{B}}_k{}^T\left(\boldsymbol{\Sigma}_\beta + \boldsymbol{C}_{1i}\right)^{-1}\hat{\underline{\boldsymbol{\beta}}}_i\right)\right]$$

$$= \exp\left[-\frac{1}{2}\left(\left(\underline{\boldsymbol{\beta}}_i - \boldsymbol{C}_{3i}\underline{\boldsymbol{d}}_{ik}\right)^T \boldsymbol{C}_{3i}{}^{-1}\left(\underline{\boldsymbol{\beta}}_i - \boldsymbol{C}_{3i}\underline{\boldsymbol{d}}_{ik}\right) - \hat{\underline{\boldsymbol{\beta}}}_i{}^T \boldsymbol{C}_{2i}{}^{-1}\underline{\boldsymbol{B}}_k\right.\right.$$

$$- \underline{\boldsymbol{B}}_k{}^T \boldsymbol{C}_{2i}{}^{-1}\hat{\underline{\boldsymbol{\beta}}}_i - \hat{\underline{\boldsymbol{\beta}}}_i{}^T \left\{\boldsymbol{C}_{1i}{}^{-1}\left(\mathbf{I} + \boldsymbol{C}_{1i}\boldsymbol{\Sigma}_\beta{}^{-1}\right)^{-1} - \boldsymbol{C}_{1i}{}^{-1}\right\}\hat{\underline{\boldsymbol{\beta}}}_i$$

$$\left.\left. - \underline{\boldsymbol{B}}_k{}^T \left\{\boldsymbol{\Sigma}_\beta{}^{-1}\left(\boldsymbol{I} + \boldsymbol{\Sigma}_\beta \boldsymbol{C}_{1i}{}^{-1}\right)^{-1} - \boldsymbol{\Sigma}_\beta{}^{-1}\right\}\underline{\boldsymbol{B}}_k\right)\right] ,$$

$$\text{where} \qquad \boldsymbol{C}_{2i} = \boldsymbol{C}_{1i} + \boldsymbol{\Sigma}_\beta ,$$

$$= \exp\left[-\frac{1}{2}\left(\left(\underline{\boldsymbol{\beta}}_i - \boldsymbol{C}_{3i}\underline{\boldsymbol{d}}_{ik}\right)^T \boldsymbol{C}_{3i}{}^{-1}\left(\underline{\boldsymbol{\beta}}_i - \boldsymbol{C}_{3i}\underline{\boldsymbol{d}}_{ik}\right) - \hat{\underline{\boldsymbol{\beta}}}_i{}^T \boldsymbol{C}_{2i}{}^{-1}\underline{\boldsymbol{B}}_k\right.\right.$$

$$- \underline{\boldsymbol{B}}_k{}^T \boldsymbol{C}_{2i}{}^{-1}\hat{\underline{\boldsymbol{\beta}}}_i - \hat{\underline{\boldsymbol{\beta}}}_i{}^T \left\{\boldsymbol{C}_{1i}{}^{-1}\boldsymbol{\Sigma}_\beta\left(\boldsymbol{\Sigma}_\beta + \boldsymbol{C}_{1i}\right)^{-1} - \boldsymbol{C}_{1i}{}^{-1}\right\}\hat{\underline{\boldsymbol{\beta}}}_i$$

$$\left.\left. - \underline{\boldsymbol{B}}_k{}^T \left\{\boldsymbol{\Sigma}_\beta{}^{-1}\boldsymbol{C}_{1i}\left(\boldsymbol{\Sigma}_\beta + \boldsymbol{C}_{1i}\right)^{-1} - \boldsymbol{\Sigma}_\beta{}^{-1}\right\}\underline{\boldsymbol{B}}_k\right)\right]$$

$$= \exp\left[-\frac{1}{2}\left(\left(\underline{\boldsymbol{\beta}}_i - \boldsymbol{C}_{3i}\underline{\boldsymbol{d}}_{ik}\right)^T \boldsymbol{C}_{3i}{}^{-1}\left(\underline{\boldsymbol{\beta}}_i - \boldsymbol{C}_{3i}\underline{\boldsymbol{d}}_{ik}\right) - \hat{\underline{\boldsymbol{\beta}}}_i{}^T \boldsymbol{C}_{2i}{}^{-1}\underline{\boldsymbol{B}}_k - \underline{\boldsymbol{B}}_k{}^T \boldsymbol{C}_{2i}{}^{-1}\hat{\underline{\boldsymbol{\beta}}}_i\right.\right.$$

$$\left.\left. - \hat{\underline{\boldsymbol{\beta}}}_i{}^T \left(\boldsymbol{C}_{1i}{}^{-1}\boldsymbol{\Sigma}_\beta \boldsymbol{C}_{2i}{}^{-1} - \boldsymbol{C}_{1i}{}^{-1}\right)\hat{\underline{\boldsymbol{\beta}}}_i - \underline{\boldsymbol{B}}_k{}^T \left(\boldsymbol{\Sigma}_\beta{}^{-1}\boldsymbol{C}_{1i}\boldsymbol{C}_{2i}{}^{-1} - \boldsymbol{\Sigma}_\beta{}^{-1}\right)\underline{\boldsymbol{B}}_k\right)\right]$$

$$= \exp\left[-\frac{1}{2}\left(\left(\underline{\boldsymbol{\beta}}_i - \boldsymbol{C}_{3i}\underline{\boldsymbol{d}}_{ik}\right)^T \boldsymbol{C}_{3i}{}^{-1}\left(\underline{\boldsymbol{\beta}}_i - \boldsymbol{C}_{3i}\underline{\boldsymbol{d}}_{ik}\right) - \hat{\underline{\boldsymbol{\beta}}}_i{}^T \boldsymbol{C}_{2i}{}^{-1}\underline{\boldsymbol{B}}_k - \underline{\boldsymbol{B}}_k{}^T \boldsymbol{C}_{2i}{}^{-1}\hat{\underline{\boldsymbol{\beta}}}_i\right.\right.$$

$$-\hat{\underline{\boldsymbol{\beta}}}_i{}^T\left\{\boldsymbol{C}_{1i}{}^{-1}\left[\boldsymbol{\Sigma}_\beta - (\boldsymbol{\Sigma}_\beta + \boldsymbol{C}_{1i})\right]\boldsymbol{C}_{2i}{}^{-1}\right\}\hat{\underline{\boldsymbol{\beta}}}_i$$

$$\left.\left. - \underline{\boldsymbol{B}}_k{}^T\left\{\boldsymbol{\Sigma}_\beta{}^{-1}\left[\boldsymbol{C}_{1i} - (\boldsymbol{\Sigma}_\beta + \boldsymbol{C}_{1i})\right]\boldsymbol{C}_{2i}{}^{-1}\right\}\underline{\boldsymbol{B}}_k\right)\right]$$

$$= \exp\left[-\frac{1}{2}\left(\left(\underline{\boldsymbol{\beta}}_i - \boldsymbol{C}_{3i}\underline{\boldsymbol{d}}_{ik}\right)^T \boldsymbol{C}_{3i}{}^{-1}\left(\underline{\boldsymbol{\beta}}_i - \boldsymbol{C}_{3i}\underline{\boldsymbol{d}}_{ik}\right) - \hat{\underline{\boldsymbol{\beta}}}_i{}^T \boldsymbol{C}_{2i}{}^{-1}\underline{\boldsymbol{B}}_k\right.\right.$$

$$\left.\left. - \underline{\boldsymbol{B}}_k{}^T \boldsymbol{C}_{2i}{}^{-1}\hat{\underline{\boldsymbol{\beta}}}_i + \hat{\underline{\boldsymbol{\beta}}}_i{}^T \boldsymbol{C}_{2i}{}^{-1}\hat{\underline{\boldsymbol{\beta}}}_i + \underline{\boldsymbol{B}}_k{}^T \boldsymbol{C}_{2i}{}^{-1}\underline{\boldsymbol{B}}_k\right)\right]$$

$$= \exp\left[-\frac{1}{2}\left(\left(\underline{\boldsymbol{\beta}}_i - \boldsymbol{C}_{3i}\underline{\boldsymbol{d}}_{ik}\right)^T \boldsymbol{C}_{3i}{}^{-1}\left(\underline{\boldsymbol{\beta}}_i - \boldsymbol{C}_{3i}\underline{\boldsymbol{d}}_{ik}\right)\right.\right.$$

$$\left.\left. + \left(\hat{\underline{\boldsymbol{\beta}}}_i - \underline{\boldsymbol{B}}_k\right)^T \boldsymbol{C}_{2i}{}^{-1}\left(\hat{\underline{\boldsymbol{\beta}}}_i - \underline{\boldsymbol{B}}_k\right)\right)\right] .$$

So,

$$L_i = \frac{D}{(2\pi)^2 \mid \boldsymbol{C}_{1i} \mid^{1/2} \mid \boldsymbol{\Sigma}_\beta \mid^{1/2}} \int \exp\left[-\frac{1}{2}\left(\underline{\boldsymbol{\beta}}_i - \boldsymbol{C}_{3i}\underline{\boldsymbol{d}}_{ik}\right)^T \boldsymbol{C}_{3i}{}^{-1}\left(\underline{\boldsymbol{\beta}}_i - \boldsymbol{C}_{3i}\underline{\boldsymbol{d}}_{ik}\right)\right]$$

$$\times \quad \exp\left[-\frac{1}{2}\left(\hat{\underline{\boldsymbol{\beta}}}_i - \underline{\boldsymbol{B}}_k\right)^T \boldsymbol{C}_{2i}{}^{-1}\left(\hat{\underline{\boldsymbol{\beta}}}_i - \underline{\boldsymbol{B}}_k\right)\right]$$

$$\times \quad \prod_{j=2}^{J}\left[\left(\Phi\left(\underline{\boldsymbol{\alpha}}^T\underline{\boldsymbol{\beta}}_i + \alpha_{0j}\right) - \Phi\left(\underline{\boldsymbol{\alpha}}^T\underline{\boldsymbol{\beta}}_i + \alpha_{0j-1}\right)\right)^{Z(i,j-1)}\right]$$

$$\times \quad \left(1 - \Phi\left(\underline{\boldsymbol{\alpha}}^T\underline{\boldsymbol{\beta}}_i + \alpha_{0J}\right)\right)^{(1-\sum_j Z(i,j))} d\underline{\boldsymbol{\beta}}_i \ ,$$

i.e.

$$L_i = \frac{D \exp\left[-\frac{1}{2}\left(\hat{\underline{\beta}}_i - \underline{\boldsymbol{B}}_k\right)^T \boldsymbol{C}_{2i}{}^{-1}\left(\hat{\underline{\beta}}_i - \underline{\boldsymbol{B}}_k\right)\right]}{(2\pi)^2 \mid \boldsymbol{C}_{1i} \mid^{1/2} \mid \boldsymbol{\Sigma}_\beta \mid^{1/2}}$$

$$\times \int \exp\left[-\frac{1}{2}\left(\underline{\boldsymbol{\beta}}_i - \boldsymbol{C}_{3i}\underline{\boldsymbol{d}}_{ik}\right)^T \boldsymbol{C}_{3i}{}^{-1}\left(\underline{\boldsymbol{\beta}}_i - \boldsymbol{C}_{3i}\underline{\boldsymbol{d}}_{ik}\right)\right]$$

$$\times \prod_{j=2}^{J}\left[\left(\Phi\left(\underline{\boldsymbol{\alpha}}^T\underline{\boldsymbol{\beta}}_i + \alpha_{0j}\right) - \Phi\left(\underline{\boldsymbol{\alpha}}^T\underline{\boldsymbol{\beta}}_i + \alpha_{0j-1}\right)\right)^{Z(i,j-1)}\right]$$

$$\times \left(1 - \Phi\left(\underline{\boldsymbol{\alpha}}^T\underline{\boldsymbol{\beta}}_i + \alpha_{0J}\right)\right)^{(1-\sum_j Z(i,j))} d\underline{\boldsymbol{\beta}}_i \ ;$$

i.e.

$$L_i = \frac{D \exp\left[-\frac{1}{2}\left(\hat{\underline{\beta}}_i - \underline{\boldsymbol{B}}_k\right)^T \boldsymbol{C}_{2i}{}^{-1}\left(\hat{\underline{\beta}}_i - \underline{\boldsymbol{B}}_k\right)\right]}{(2\pi)^2 \mid \boldsymbol{C}_{1i} \mid^{1/2} \mid \boldsymbol{\Sigma}_\beta \mid^{1/2}}$$

$$\times \int \left(\sqrt{2\,\pi}\right)^2 \mid \boldsymbol{C}_{3i} \mid^{1/2} \phi_2(\underline{\boldsymbol{\beta}}_i, \boldsymbol{C}_{3i}\underline{\boldsymbol{d}}_{ik}, \boldsymbol{C}_{3i})$$

$$\times \prod_{j=2}^{J}\left[\left(\Phi\left(\underline{\boldsymbol{\alpha}}^T\underline{\boldsymbol{\beta}}_i + \alpha_{0j}\right) - \Phi\left(\underline{\boldsymbol{\alpha}}^T\underline{\boldsymbol{\beta}}_i + \alpha_{0j-1}\right)\right)^{Z(i,j-1)}\right]$$

$$\times \left(1 - \Phi\left(\underline{\boldsymbol{\alpha}}^T\underline{\boldsymbol{\beta}}_i + \alpha_{0J}\right)\right)^{(1-\sum_j Z(i,j))} d\underline{\boldsymbol{\beta}}_i \ ;$$

i.e.

$$L_i = \frac{D \exp\left[-\frac{1}{2}\left(\underline{\hat{\beta}}_i - \underline{\boldsymbol{B}}_k\right)^T \boldsymbol{C}_{2i}{}^{-1}\left(\underline{\hat{\beta}}_i - \underline{\boldsymbol{B}}_k\right)\right] \mid \boldsymbol{C}_{3i} \mid^{1/2}}{2\pi \mid \boldsymbol{C}_{1i} \mid^{1/2} \mid \boldsymbol{\Sigma}_\beta \mid^{1/2}}$$

$$\times \int \phi_2(\underline{\boldsymbol{\beta}}_i, \boldsymbol{C}_{3i}\underline{\boldsymbol{d}}_{ik}, \boldsymbol{C}_{3i}) \prod_{j=2}^{J} \left[\left(\Phi\left(\underline{\boldsymbol{\alpha}}^T\underline{\boldsymbol{\beta}}_i + \alpha_{0j}\right) - \Phi\left(\underline{\boldsymbol{\alpha}}^T\underline{\boldsymbol{\beta}}_i + \alpha_{0j-1}\right)\right)^{Z(i,j-1)}\right]$$

$$\times \left(1 - \Phi\left(\underline{\boldsymbol{\alpha}}^T\underline{\boldsymbol{\beta}}_i + \alpha_{0J}\right)\right)^{(1-\sum_j Z(i,j))} d\underline{\boldsymbol{\beta}}_i \quad .$$

But it can be shown that:

$$\int \Phi(\underline{\boldsymbol{\alpha}}^T\underline{\boldsymbol{\beta}}_i + c)\, \phi_2(\underline{\boldsymbol{\beta}}_i, \underline{\boldsymbol{\mu}}, \boldsymbol{V})\, d\underline{\boldsymbol{\beta}}_i = \Phi\left[\frac{\underline{\boldsymbol{\alpha}}^T\underline{\boldsymbol{\mu}} + c}{\sqrt{1 + \underline{\boldsymbol{\alpha}}^T\boldsymbol{V}\underline{\boldsymbol{\alpha}}}}\right] \quad ,$$

where c is a constant.

It easily follows that:

$$\int \left(1 - \Phi(\underline{\boldsymbol{\alpha}}^T\underline{\boldsymbol{\beta}}_i + c)\right) \phi_2(\underline{\boldsymbol{\beta}}_i, \underline{\boldsymbol{\mu}}, \boldsymbol{V})\, d\underline{\boldsymbol{\beta}}_i = 1 - \Phi\left[\frac{\underline{\boldsymbol{\alpha}}^T\underline{\boldsymbol{\mu}} + c}{\sqrt{1 + \underline{\boldsymbol{\alpha}}^T\boldsymbol{V}\underline{\boldsymbol{\alpha}}}}\right] \quad .$$

So,

$$L_i = \frac{D \exp\left[-\frac{1}{2}\left(\hat{\underline{\beta}}_i - \underline{B}_k\right)^T {C_{2i}}^{-1}\left(\hat{\underline{\beta}}_i - \underline{B}_k\right)\right] \mid ({C_{1i}}^{-1} + {\Sigma_\beta}^{-1})^{-1} \mid^{1/2}}{2\pi \mid C_{1i} \mid^{1/2} \mid \Sigma_\beta \mid^{1/2}}$$

$$\times \left[\prod_{j=2}^{J}\left\{\Phi\left[\left((C_{3i}\underline{d}_{ik})^T \underline{\alpha} + \alpha_{0j}\right)\left(1 + \underline{\alpha}^T C_{3i}\underline{\alpha}\right)^{-\frac{1}{2}}\right]\right.\right.$$

$$\left.\left. - \Phi\left[\left((C_{3i}\underline{d}_{ik})^T \underline{\alpha} + \alpha_{0j-1}\right)\left(1 + \underline{\alpha}^T C_{3i}\underline{\alpha}\right)^{-\frac{1}{2}}\right]\right\}^{Z(i,j-1)}\right]$$

$$\times \left\{1 - \Phi\left[\left((C_{3i}\underline{d}_{ik})^T \underline{\alpha} + \alpha_{0J}\right)\left(1 + \underline{\alpha}^T C_{3i}\underline{\alpha}\right)^{-\frac{1}{2}}\right]\right\}^{(1-\sum_j Z(i,j))}$$

as exactly one of the last J terms in the product is other than unity.

Therefore

$$L_i = \frac{D \exp\left[-\frac{1}{2}\left(\hat{\underline{\beta}}_i - \underline{B}_k\right)^T {C_{2i}}^{-1}\left(\hat{\underline{\beta}}_i - \underline{B}_k\right)\right]}{2\pi \mid C_{1i} \mid^{1/2} \mid \Sigma_\beta \mid^{1/2} \mid ({C_{1i}}^{-1} + {\Sigma_\beta}^{-1}) \mid^{1/2}}$$

$$\times \left[\prod_{j=2}^{J}\left\{\Phi\left[\left(\underline{d}_{ik}{}^T {C_{3i}}^T \underline{\alpha} + \alpha_{0j}\right)\left(1 + \underline{\alpha}^T C_{3i}\underline{\alpha}\right)^{-\frac{1}{2}}\right]\right.\right.$$

$$\left.\left. - \Phi\left[\left(\underline{d}_{ik}{}^T {C_{3i}}^T \underline{\alpha} + \alpha_{0j-1}\right)\left(1 + \underline{\alpha}^T C_{3i}\underline{\alpha}\right)^{-\frac{1}{2}}\right]\right\}^{Z(i,j-1)}\right]$$

$$\times \left\{1 - \Phi\left[\left(\underline{d}_{ik}{}^T {C_{3i}}^T \underline{\alpha} + \alpha_{0J}\right)\left(1 + \underline{\alpha}^T C_{3i}\underline{\alpha}\right)^{-\frac{1}{2}}\right]\right\}^{(1-\sum_j Z(i,j))}$$

$$= \frac{D \exp\left[-\frac{1}{2}\left(\hat{\underline{\beta}}_i - \underline{\boldsymbol{B}}_k\right)^T \boldsymbol{C}_{2i}{}^{-1}\left(\hat{\underline{\beta}}_i - \underline{\boldsymbol{B}}_k\right)\right]}{2\pi \mid \boldsymbol{C}_{1i} \mid^{1/2} \mid \boldsymbol{\Sigma}_\beta \mid^{1/2} \mid \boldsymbol{\Sigma}_\beta{}^{-1}(\boldsymbol{\Sigma}_\beta + \boldsymbol{C}_{1i})\boldsymbol{C}_{1i}{}^{-1} \mid^{1/2}}$$

$$\times \left[\prod_{j=2}^{J} \left\{\Phi\left[\left(\underline{\boldsymbol{d}}_{ik}{}^T \boldsymbol{C}_{3i}{}^T \underline{\boldsymbol{\alpha}} + \alpha_{0j}\right)\left(1 + \underline{\boldsymbol{\alpha}}^T \boldsymbol{C}_{3i} \underline{\boldsymbol{\alpha}}\right)^{-\frac{1}{2}}\right]\right.\right.$$

$$\left.\left. - \Phi\left[\left(\underline{\boldsymbol{d}}_{ik}{}^T \boldsymbol{C}_{3i}{}^T \underline{\boldsymbol{\alpha}} + \alpha_{0j-1}\right)\left(1 + \underline{\boldsymbol{\alpha}}^T \boldsymbol{C}_{3i} \underline{\boldsymbol{\alpha}}\right)^{-\frac{1}{2}}\right]\right\}^{Z(i,j-1)}\right]$$

$$\times \left\{1 - \Phi\left[\left(\underline{\boldsymbol{d}}_{ik}{}^T \boldsymbol{C}_{3i}{}^T \underline{\boldsymbol{\alpha}} + \alpha_{0J}\right)\left(1 + \underline{\boldsymbol{\alpha}}^T \boldsymbol{C}_{3i} \underline{\boldsymbol{\alpha}}\right)^{-\frac{1}{2}}\right]\right\}^{(1-\sum_j Z(i,j))} .$$

So

$$L_i = \frac{D \exp\left[-\frac{1}{2}\left(\hat{\underline{\beta}}_i - \underline{\boldsymbol{B}}_k\right)^T \boldsymbol{C}_{2i}{}^{-1}\left(\hat{\underline{\beta}}_i - \underline{\boldsymbol{B}}_k\right)\right]}{2\pi \mid \boldsymbol{C}_{1i} \mid^{1/2} \mid \boldsymbol{\Sigma}_\beta \mid^{1/2} \mid \boldsymbol{\Sigma}_\beta \mid^{-1/2} \mid \boldsymbol{C}_{2i} \mid^{1/2} \mid \boldsymbol{C}_{1i} \mid^{-1/2}}$$

$$\times \left[\prod_{j=2}^{J} \left\{\Phi\left[\left(\underline{\boldsymbol{d}}_{ik}{}^T \boldsymbol{C}_{3i}{}^T \underline{\boldsymbol{\alpha}} + \alpha_{0j}\right)\left(1 + \underline{\boldsymbol{\alpha}}^T \boldsymbol{C}_{3i} \underline{\boldsymbol{\alpha}}\right)^{-\frac{1}{2}}\right]\right.\right.$$

$$\left.\left. - \Phi\left[\left(\underline{\boldsymbol{d}}_{ik}{}^T \boldsymbol{C}_{3i}{}^T \underline{\boldsymbol{\alpha}} + \alpha_{0j-1}\right)\left(1 + \underline{\boldsymbol{\alpha}}^T \boldsymbol{C}_{3i} \underline{\boldsymbol{\alpha}}\right)^{-\frac{1}{2}}\right]\right\}^{Z(i,j-1)}\right]$$

$$\times \left\{1 - \Phi\left[\left(\underline{\boldsymbol{d}}_{ik}{}^T \boldsymbol{C}_{3i}{}^T \underline{\boldsymbol{\alpha}} + \alpha_{0J}\right)\left(1 + \underline{\boldsymbol{\alpha}}^T \boldsymbol{C}_{3i} \underline{\boldsymbol{\alpha}}\right)^{-\frac{1}{2}}\right]\right\}^{(1-\sum_j Z(i,j))} .$$

But

$$\boldsymbol{C}_{3i}{}^{T} = \left[\left(\boldsymbol{C}_{1i}{}^{-1} + \boldsymbol{\Sigma}_{\beta}{}^{-1}\right)^{-1}\right]^{T} = \left[\left(\boldsymbol{C}_{1i}{}^{-1} + \boldsymbol{\Sigma}_{\beta}{}^{-1}\right)^{T}\right]^{-1}$$

$$= \left[\left(\boldsymbol{C}_{1i}{}^{-1}\right)^{T} + \left(\boldsymbol{\Sigma}_{\beta}{}^{-1}\right)^{T}\right]^{-1}$$

$$= \left[\left(\boldsymbol{C}_{1i}{}^{T}\right)^{-1} + \left(\boldsymbol{\Sigma}_{\beta}{}^{T}\right)^{-1}\right]^{-1}$$

$$= \left(\boldsymbol{C}_{1i}{}^{-1} + \boldsymbol{\Sigma}_{\beta}{}^{-1}\right)^{-1} = \boldsymbol{C}_{3i}$$

as $\boldsymbol{C}_{1i}$ and $\boldsymbol{\Sigma}_{\beta}$ are symmetric.

So,

$$L_i = \frac{D \exp\left[-\frac{1}{2}\left(\hat{\underline{\boldsymbol{\beta}}}_i - \underline{\boldsymbol{B}}_k\right)^{T} \boldsymbol{C}_{2i}{}^{-1}\left(\hat{\underline{\boldsymbol{\beta}}}_i - \underline{\boldsymbol{B}}_k\right)\right]}{2\pi \mid \boldsymbol{C}_{2i} \mid^{1/2}}$$

$$\times \left[\prod_{j=2}^{J}\left\{\Phi\left[\left(\underline{\boldsymbol{d}}_{ik}{}^{T}\boldsymbol{C}_{3i}\underline{\boldsymbol{\alpha}} + \alpha_{0j}\right)\left(1 + \underline{\boldsymbol{\alpha}}^{T}\boldsymbol{C}_{3i}\underline{\boldsymbol{\alpha}}\right)^{-\frac{1}{2}}\right]\right.\right.$$

$$\left.\left. - \Phi\left[\left(\underline{\boldsymbol{d}}_{ik}{}^{T}\boldsymbol{C}_{3i}\underline{\boldsymbol{\alpha}} + \alpha_{0j-1}\right)\left(1 + \underline{\boldsymbol{\alpha}}^{T}\boldsymbol{C}_{3i}\underline{\boldsymbol{\alpha}}\right)^{-\frac{1}{2}}\right]\right\}^{Z(i,j-1)}\right]$$

$$\times \left\{1 - \Phi\left[\left(\underline{\boldsymbol{d}}_{ik}{}^{T}\boldsymbol{C}_{3i}\underline{\boldsymbol{\alpha}} + \alpha_{0J}\right)\left(1 + \underline{\boldsymbol{\alpha}}^{T}\boldsymbol{C}_{3i}\underline{\boldsymbol{\alpha}}\right)^{-\frac{1}{2}}\right]\right\}^{(1-\sum_j Z(i,j))} .$$

Appendix B

Detailed tables for parameter estimation via 'adapted Finkelstein' (including nuisance parameters)

Parameter	True Value	Bias	95% C.I. for bias
β	0.07	0.148×10^{-3}	$(-0.344\times10^{-3}, 0.640\times10^{-3})$
γ_1	-9.2000	-0.0255	(-0.0693, 0.0183)
γ_2	-9.0019	-0.0222	(-0.0661, 0.0217)
γ_3	-8.9237	-0.0303	(-0.0739, 0.0134)
γ_4	-8.8728	-0.0191	(-0.0644, 0.0263)
γ_5	-8.8349	-0.0194	(-0.0652, 0.0264)
γ_6	-8.8047	-0.0213	(-0.0678, 0.0252)
γ_7	-8.7796	-0.0255	(-0.0711, 0.0201)
γ_8	-8.7581	-0.0268	(-0.0724, 0.0188)
γ_9	-8.7393	-0.0232	(-0.0697, 0.0233)
γ_{10}	-8.7226	-0.0311	(-0.0782, 0.0159)

Parameter	True Value	MSE	Standard Error of MSE
β	0.07	3.149×10^{-5}	1.96×10^{-6}
γ_1	-9.2000	0.2496	0.0168
γ_2	-9.0019	0.2511	0.0158
γ_3	-8.9237	0.2484	0.0160
γ_4	-8.8728	0.2676	0.0167
γ_5	-8.8349	0.2732	0.0168
γ_6	-8.8047	0.2817	0.0174
γ_7	-8.7796	0.2713	0.0168
γ_8	-8.7581	0.2705	0.0165
γ_9	-8.7393	0.2816	0.0181
γ_{10}	-8.7226	0.2880	0.0184

Table B.1: Comparison of bias and MSE from simulations of a Weibull mortality process with shape parameter 1.15 and $\sigma_W = 1.5$ and $\sigma_B = 8.0$ using 'adapted Finkelstein' with $\sigma_B = 8.0$ and a single continuous repeated measurement with no missing data.

Parameter	True Value	EVENT INDEPENDENTLY COLLECTED Bias	95% C.I. for bias	EVENT DEPENDENTLY COLLECTED Bias	95% C.I. for bias
β	0.07	0.155×10^{-3}	$(-0.338\times10^{-3}, 0.647\times10^{-3})$	0.471×10^{-3}	$(-0.023\times10^{-3}, 0.965\times10^{-3})$
γ_1	-9.2000	-0.0260	(-0.0699, 0.0178)	-0.0220	(-0.0659, 0.0219)
γ_2	-9.0019	-0.0194	(-0.0633, 0.0245)	-0.0411	(-0.0851, 0.0029)
γ_3	-8.9237	-0.0274	(-0.0711, 0.0163)	-0.0508	(-0.0947, -0.0068)
γ_4	-8.8728	-0.0161	(-0.0615, 0.0292)	-0.0430	(-0.0886, 0.0026)
γ_5	-8.8349	-0.0164	(-0.0623, 0.0294)	-0.0403	(-0.0862, 0.0056)
γ_6	-8.8047	-0.0183	(-0.0648, 0.0282)	-0.0445	(-0.0911, 0.0021)
γ_7	-8.7796	-0.0224	(-0.0680, 0.0232)	-0.0483	(-0.0942, -0.0024)
γ_8	-8.7581	-0.0237	(-0.0693, 0.0219)	-0.0486	(-0.0943, -0.0029)
γ_9	-8.7393	-0.0201	(-0.0666, 0.0264)	-0.0451	(-0.0916, 0.0015)
γ_{10}	-8.7226	-0.0279	(-0.0749, 0.0192)	-0.0354	(-0.0828, 0.0121)

Parameter	True Value	EVENT INDEPENDENTLY COLLECTED MSE	Standard Error of MSE	EVENT DEPENDENTLY COLLECTED MSE	Standard Error of MSE
β	0.07	3.150×10^{-5}	1.96×10^{-6}	3.192×10^{-5}	2.00×10^{-6}
γ_1	-9.2000	0.2502	0.0168	0.2507	0.0167
γ_2	-9.0019	0.2508	0.0157	0.2531	0.0161
γ_3	-8.9237	0.2482	0.0159	0.2532	0.0162
γ_4	-8.8728	0.2673	0.0167	0.2721	0.0171
γ_5	-8.8349	0.2729	0.0167	0.2754	0.0169
γ_6	-8.8047	0.2809	0.0173	0.2840	0.0177
γ_7	-8.7796	0.2709	0.0168	0.2758	0.0172
γ_8	-8.7581	0.2704	0.0166	0.2735	0.0167
γ_9	-8.7393	0.2813	0.0181	0.2836	0.0184
γ_{10}	-8.7226	0.2881	0.0183	0.2934	0.0188

Table B.2: Comparison of bias and MSE from simulations of a Weibull mortality process with shape parameter 1.15 and $\sigma_W = 1.5$ and $\sigma_B = 8.0$ using 'adapted Finkelstein' with a single continuous repeated measurement with 5% data missing via a MCAR process under two different situations for the assessment of the interval containing the event of interest using imputation of the last measurement.

Parameter	True Value	EVENT INDEPENDENTLY COLLECTED Bias	95% C.I. for bias	EVENT DEPENDENTLY COLLECTED Bias	95% C.I. for bias
β	0.07	0.144×10^{-3}	$(-0.346\times10^{-3}, 0.635\times10^{-3})$	0.808×10^{-3}	$(0.314\times10^{-3}, 1.302\times10^{-3})$
γ_1	-9.2000	-0.0251	(-0.0689, 0.0186)	-0.0184	(-0.0622, 0.0254)
γ_2	-9.0019	-0.0153	(-0.0591, 0.0284)	-0.0610	(-0.1050, -0.0169)
γ_3	-8.9237	-0.0228	(-0.0663, 0.0206)	-0.0728	(-0.1167, -0.0289)
γ_4	-8.8728	-0.0114	(-0.0566, 0.0337)	-0.0626	(-0.1082, -0.0169)
γ_5	-8.8349	-0.0116	(-0.0572, 0.0341)	-0.0649	(-0.1108, -0.0190)
γ_6	-8.8047	-0.0133	(-0.0596, 0.0331)	-0.0680	(-0.1145, -0.0215)
γ_7	-8.7796	-0.0174	(-0.0628, 0.0281)	-0.0701	(-0.1161, -0.0241)
γ_8	-8.7581	-0.0187	(-0.0640, 0.0267)	-0.0720	(-0.1177, -0.0264)
γ_9	-8.7393	-0.0151	(-0.0614, 0.0312)	-0.0688	(-0.1155, -0.0221)
γ_{10}	-8.7226	-0.0228	(-0.0696, 0.0241)	-0.0356	(-0.0831, 0.0118)

Parameter	True Value	EVENT INDEPENDENTLY COLLECTED MSE	Standard Error of MSE	EVENT DEPENDENTLY COLLECTED MSE	Standard Error of MSE
β	0.07	3.129×10^{-5}	1.93×10^{-6}	3.231×10^{-5}	2.01×10^{-6}
γ_1	-9.2000	0.2493	0.0165	0.2497	0.0162
γ_2	-9.0019	0.2490	0.0155	0.2560	0.0162
γ_3	-8.9237	0.2458	0.0157	0.2561	0.0163
γ_4	-8.8728	0.2652	0.0165	0.2747	0.0171
γ_5	-8.8349	0.2709	0.0163	0.2778	0.0170
γ_6	-8.8047	0.2790	0.0172	0.2858	0.0178
γ_7	-8.7796	0.2684	0.0166	0.2794	0.0175
γ_8	-8.7581	0.2677	0.0163	0.2755	0.0168
γ_9	-8.7393	0.2792	0.0178	0.2883	0.0187
γ_{10}	-8.7226	0.2854	0.0180	0.2937	0.0186

Table B.3: Comparison of bias and MSE from simulations of a Weibull mortality process with shape parameter 1.15 and $\sigma_W = 1.5$ and $\sigma_B = 8.0$ using 'adapted Finkelstein' with a single continuous repeated measurement with 10% data missing via a MCAR process under two different situations for the assessment of the interval containing the event of interest using imputation of the last measurement.

Parameter	True Value	EVENT INDEPENDENTLY COLLECTED Bias	95% C.I. for bias	EVENT DEPENDENTLY COLLECTED Bias	95% C.I. for bias
β	0.07	0.170×10^{-3}	$(-0.322\times10^{-3}, 0.663\times10^{-3})$	1.612×10^{-3}	$(1.113\times10^{-3}, 2.112\times10^{-3})$
γ_1	-9.2000	-0.0274	(-0.0713, 0.0166)	-0.0145	(-0.0588, 0.0297)
γ_2	-9.0019	-0.0110	(-0.0549, 0.0329)	-0.1078	(-0.1525, -0.0630)
γ_3	-8.9237	-0.0165	(-0.0601, 0.0270)	-0.1227	(-0.1670, -0.0783)
γ_4	-8.8728	-0.0047	(-0.0500, 0.0405)	-0.1140	(-0.1602, -0.0678)
γ_5	-8.8349	-0.0045	(-0.0502, 0.0412)	-0.1253	(-0.1720, -0.0786)
γ_6	-8.8047	-0.0063	(-0.0527, 0.0402)	-0.1221	(-0.1689, -0.0753)
γ_7	-8.7796	-0.0103	(-0.0558, 0.0353)	-0.1273	(-0.1738, -0.0808)
γ_8	-8.7581	-0.0114	(-0.0568, 0.0341)	-0.1252	(-0.1713, -0.0791)
γ_9	-8.7393	-0.0077	(-0.0541, 0.0388)	-0.1144	(-0.1620, -0.0668)
γ_{10}	-8.7226	-0.0155	(-0.0625, 0.0315)	-0.0449	(-0.0929, 0.0312)

Parameter	True Value	EVENT INDEPENDENTLY COLLECTED MSE	Standard Error of MSE	EVENT DEPENDENTLY COLLECTED MSE	Standard Error of MSE
β	0.07	3.157×10^{-5}	1.94×10^{-6}	3.500×10^{-5}	2.18×10^{-6}
γ_1	-9.2000	0.2517	0.0167	0.2547	0.0166
γ_2	-9.0019	0.2508	0.0156	0.2719	0.0172
γ_3	-8.9237	0.2462	0.0156	0.2703	0.0173
γ_4	-8.8728	0.2664	0.0164	0.2902	0.0177
γ_5	-8.8349	0.2714	0.0163	0.2985	0.0180
γ_6	-8.8047	0.2802	0.0172	0.2995	0.0184
γ_7	-8.7796	0.2696	0.0166	0.2969	0.0185
γ_8	-8.7581	0.2684	0.0162	0.2917	0.0178
γ_9	-8.7393	0.2803	0.0178	0.3072	0.0201
γ_{10}	-8.7226	0.2872	0.0179	0.3013	0.0191

Table B.4: Comparison of bias and MSE from simulations of a Weibull mortality process with shape parameter 1.15 and $\sigma_W = 1.5$ and $\sigma_B = 8.0$ using 'adapted Finkelstein' with a single continuous repeated measurement with 20% data missing via a MCAR process under two different situations for the assessment of the interval containing the event of interest using imputation of the last measurement.

Parameter	True Value	EVENT INDEPENDENTLY COLLECTED Bias	95% C.I. for bias	EVENT DEPENDENTLY COLLECTED Bias	95% C.I. for bias
β	0.07	-0.126×10^{-3}	$(-0.622\times10^{-3}, 0.369\times10^{-3})$	4.536×10^{-3}	$(4.023\times10^{-3}, 5.047\times10^{-3})$
γ_1	-9.2000	-0.0024	(-0.0464, 0.0416)	0.0553	(0.0102, 0.1004)
γ_2	-9.0019	0.0340	(-0.0100, 0.0779)	-0.2527	(-0.2992, -0.2061)
γ_3	-8.9237	0.0422	(-0.0012, 0.0856)	-0.2817	(-0.3269, -0.2364)
γ_4	-8.8728	0.0617	(0.0166, 0.1068)	-0.2768	(-0.3244, -0.2292)
γ_5	-8.8349	0.0659	(0.0205, 0.1114)	-0.3017	(-0.3497, -0.2537)
γ_6	-8.8047	0.0660	(0.0196, 0.1124)	-0.2923	(-0.3408, -0.2437)
γ_7	-8.7796	0.0637	(0.0184, 0.1090)	-0.2936	(-0.3413, -0.2458)
γ_8	-8.7581	0.0632	(0.0178, 0.1085)	-0.2597	(-0.3072, -0.2122)
γ_9	-8.7393	0.0683	(0.0218, 0.1147)	-0.2070	(-0.2562, -0.1578)
γ_{10}	-8.7226	0.0610	(0.0142, 0.1077)	-0.0603	(-0.1103, -0.0102)

Parameter	True Value	EVENT INDEPENDENTLY COLLECTED MSE	Standard Error of MSE	EVENT DEPENDENTLY COLLECTED MSE	Standard Error of MSE
β	0.07	3.193×10^{-5}	1.97×10^{-6}	5.461×10^{-5}	3.24×10^{-6}
γ_1	-9.2000	0.2516	0.0164	0.2671	0.0172
γ_2	-9.0019	0.2522	0.0157	0.3456	0.0207
γ_3	-8.9237	0.2461	0.0154	0.3454	0.0218
γ_4	-8.8728	0.2681	0.0165	0.3715	0.0227
γ_5	-8.8349	0.2725	0.0160	0.3907	0.0242
γ_6	-8.8047	0.2842	0.0177	0.3920	0.0244
γ_7	-8.7796	0.2708	0.0168	0.3824	0.0241
γ_8	-8.7581	0.2714	0.0164	0.3606	0.0226
γ_9	-8.7393	0.2852	0.0180	0.3574	0.0232
γ_{10}	-8.7226	0.2877	0.0177	0.3291	0.0208

Table B.5: Comparison of bias and MSE from simulations of a Weibull mortality process with shape parameter 1.15 and $\sigma_W = 1.5$ and $\sigma_B = 8.0$ using 'adapted Finkelstein' with a single continuous repeated measurement with 50% data missing via a MCAR process under two different situations for the assessment of the interval containing the event of interest using imputation of the last measurement.

Parameter	True Value	EVENT INDEPENDENTLY COLLECTED Bias	95% C.I. for bias	EVENT DEPENDENTLY COLLECTED Bias	95% C.I. for bias
β	0.07	0.079×10^{-3}	$(-0.413 \times 10^{-3}, 0.572 \times 10^{-3})$	0.139×10^{-3}	$(-0.354 \times 10^{-3}, 0.633 \times 10^{-3})$
γ_1	-9.2000	-0.0197	(-0.0635 , 0.0241)	0.0061	(-0.0377, 0.0498)
γ_2	-9.0019	-0.0161	(-0.0600 , 0.0279)	-0.0169	(-0.0609, 0.0270)
γ_3	-8.9237	-0.0242	(-0.0679 , 0.0195)	-0.0257	(-0.0696, 0.0182)
γ_4	-8.8728	-0.0130	(-0.0584 , 0.0324)	-0.0175	(-0.0631, 0.0280)
γ_5	-8.8349	-0.0132	(-0.0591 , 0.0326)	-0.0146	(-0.0605, 0.0313)
γ_6	-8.8047	-0.0152	(-0.0617 , 0.0314)	-0.0186	(-0.0652, 0.0279)
γ_7	-8.7796	-0.0193	(-0.0650 , 0.0264)	-0.0223	(-0.0682, 0.0235)
γ_8	-8.7581	-0.0206	(-0.0662 , 0.0250)	-0.0223	(-0.0680, 0.0233)
γ_9	-8.7393	-0.0169	(-0.0635 , 0.0297)	-0.0188	(-0.0653, 0.0277)
γ_{10}	-8.7226	-0.0249	(-0.0720 , 0.0222)	-0.0079	(-0.0553, 0.0395)

Parameter	True Value	EVENT INDEPENDENTLY COLLECTED MSE	Standard Error of MSE	EVENT DEPENDENTLY COLLECTED MSE	Standard Error of MSE
β	0.07	3.151×10^{-5}	1.96×10^{-6}	3.163×10^{-5}	1.97×10^{-6}
γ_1	-9.2000	0.2495	0.0168	0.2491	0.0166
γ_2	-9.0019	0.2511	0.0158	0.2514	0.0160
γ_3	-8.9237	0.2486	0.0159	0.2509	0.0160
γ_4	-8.8728	0.2675	0.0166	0.2698	0.0168
γ_5	-8.8349	0.2731	0.0167	0.2735	0.0167
γ_6	-8.8047	0.2812	0.0173	0.2818	0.0175
γ_7	-8.7796	0.2714	0.0169	0.2737	0.0170
γ_8	-8.7581	0.2707	0.0166	0.2712	0.0165
γ_9	-8.7393	0.2821	0.0181	0.2816	0.0182
γ_{10}	-8.7226	0.2888	0.0183	0.2920	0.0185

Table B.6: Comparison of bias and MSE from simulations of a Weibull mortality process with shape parameter 1.15 and $\sigma_W = 1.5$ and $\sigma_B = 8.0$ using 'adapted Finkelstein' with a single continuous repeated measurement with 5% data missing via a MCAR process under two different situations for the assessment of the interval containing the event of interest using linear interpolation imputation.

Parameter	True Value	EVENT INDEPENDENTLY COLLECTED Bias	95% C.I. for bias	EVENT DEPENDENTLY COLLECTED Bias	95% C.I. for bias
β	0.07	-0.009 $\times 10^{-3}$	(-0.500$\times 10^{-3}$, 0.483 $\times$ 10^{-3})	0.123$\times 10^{-3}$	(-0.370$\times 10^{-3}$, 0.616 $\times$ 10^{-3})
γ_1	-9.2000	-0.0122	(-0.0559 , 0.0315)	0.0396	(-0.0040, 0.0833)
γ_2	-9.0019	-0.0084	(-0.0523 , 0.0355)	-0.0109	(-0.0550, 0.0331)
γ_3	-8.9237	-0.0166	(-0.0602 , 0.0270)	-0.0212	(-0.0651, 0.0227)
γ_4	-8.8728	-0.0053	(-0.0505 , 0.0400)	-0.0103	(-0.0559, 0.0353)
γ_5	-8.8349	-0.0055	(-0.0513 , 0.0403)	-0.0122	(-0.0580, 0.0337)
γ_6	-8.8047	-0.0073	(-0.0537 , 0.0392)	-0.0149	(-0.0613, 0.0316)
γ_7	-8.7796	-0.0113	(-0.0568 , 0.0343)	-0.0168	(-0.0627, 0.0292)
γ_8	-8.7581	-0.0126	(-0.0581 , 0.0329)	-0.0181	(-0.0637, 0.0274)
γ_9	-8.7393	-0.0089	(-0.0555 , 0.0376)	-0.0143	(-0.0611, 0.0324)
γ_{10}	-8.7226	-0.0169	(-0.0639 , 0.0302)	0.0210	(-0.0264, 0.0684)

Parameter	True Value	EVENT INDEPENDENTLY COLLECTED MSE	Standard Error of MSE	EVENT DEPENDENTLY COLLECTED MSE	Standard Error of MSE
β	0.07	3.138$\times 10^{-5}$	1.93$\times 10^{-6}$	3.153$\times 10^{-5}$	1.94$\times 10^{-6}$
γ_1	-9.2000	0.2485	0.0166	0.2487	0.0161
γ_2	-9.0019	0.2501	0.0156	0.2519	0.0159
γ_3	-8.9237	0.2470	0.0157	0.2507	0.0158
γ_4	-8.8728	0.2662	0.0164	0.2700	0.0166
γ_5	-8.8349	0.2724	0.0164	0.2733	0.0165
γ_6	-8.8047	0.2802	0.0172	0.2806	0.0174
γ_7	-8.7796	0.2699	0.0166	0.2743	0.0171
γ_8	-8.7581	0.2688	0.0163	0.2699	0.0164
γ_9	-8.7393	0.2816	0.0179	0.2837	0.0182
γ_{10}	-8.7226	0.2876	0.0181	0.2923	0.0182

Table B.7: Comparison of bias and MSE from simulations of a Weibull mortality process with shape parameter 1.15 and $\sigma_W = 1.5$ and $\sigma_B = 8.0$ using 'adapted Finkelstein' with a single continuous repeated measurement with 10% data missing via a MCAR process under two different situations for the assessment of the interval containing the event of interest using linear interpolation imputation.

Parameter	True Value	EVENT INDEPENDENTLY COLLECTED Bias	95% C.I. for bias	EVENT DEPENDENTLY COLLECTED Bias	95% C.I. for bias
β	0.07	-0.277 $\times 10^{-3}$	(-0.766$\times 10^{-3}$, 0.213 $\times$ 10^{-3})	0.066$\times 10^{-3}$	(-0.429$\times 10^{-3}$, 0.561 $\times$ 10^{-3})
γ_1	-9.2000	0.0104	(-0.0332 , 0.0540)	0.1171	(0.0734, 0.1608)
γ_2	-9.0019	0.0151	(-0.0286 , 0.0588)	0.0067	(-0.0377, 0.0511)
γ_3	-8.9237	0.0068	(-0.0366 , 0.0502)	-0.0066	(-0.0507, 0.0375)
γ_4	-8.8728	0.0182	(-0.0269 , 0.0632)	0.0036	(-0.0422, 0.0495)
γ_5	-8.8349	0.0181	(-0.0274 , 0.0637)	-0.0064	(-0.0527, 0.0400)
γ_6	-8.8047	0.0164	(-0.0298 , 0.0626)	-0.0022	(-0.0487, 0.0443)
γ_7	-8.7796	0.0128	(-0.0326 , 0.0581)	-0.0066	(-0.0528, 0.0396)
γ_8	-8.7581	0.0116	(-0.0338 , 0.0569)	-0.0035	(-0.0493, 0.0423)
γ_9	-8.7393	0.0153	(-0.0310 , 0.0617)	0.0088	(-0.0385, 0.0561)
γ_{10}	-8.7226	0.0070	(-0.0398 , 0.0539)	0.0831	(0.0355, 0.1307)

Parameter	True Value	EVENT INDEPENDENTLY COLLECTED MSE	Standard Error of MSE	EVENT DEPENDENTLY COLLECTED MSE	Standard Error of MSE
β	0.07	3.119$\times 10^{-5}$	1.92$\times 10^{-6}$	3.178$\times 10^{-5}$	1.94$\times 10^{-6}$
γ_1	-9.2000	0.2465	0.0165	0.2619	0.0169
γ_2	-9.0019	0.2484	0.0157	0.2563	0.0162
γ_3	-8.9237	0.2447	0.0156	0.2525	0.0159
γ_4	-8.8728	0.2642	0.0163	0.2731	0.0163
γ_5	-8.8349	0.2696	0.0163	0.2792	0.0165
γ_6	-8.8047	0.2777	0.0170	0.2804	0.0172
γ_7	-8.7796	0.2675	0.0166	0.2772	0.0172
γ_8	-8.7581	0.2674	0.0163	0.2727	0.0166
γ_9	-8.7393	0.2794	0.0179	0.2909	0.0187
γ_{10}	-8.7226	0.2852	0.0177	0.3014	0.0184

Table B.8: Comparison of bias and MSE from simulations of a Weibull mortality process with shape parameter 1.15 and $\sigma_W = 1.5$ and $\sigma_B = 8.0$ using 'adapted Finkelstein' with a single continuous repeated measurement with 20% data missing via a MCAR process under two different situations for the assessment of the interval containing the event of interest using linear interpolation imputation.

Parameter	True Value	EVENT INDEPENDENTLY COLLECTED Bias	95% C.I. for bias	EVENT DEPENDENTLY COLLECTED Bias	95% C.I. for bias
β	0.07	-3.053 $\times 10^{-3}$	(-3.545$\times 10^{-3}$, -2.562 $\times 10^{-3}$)	-1.146$\times 10^{-3}$	(-1.648$\times 10^{-3}$, -0.644 $\times 10^{-3}$)
γ_1	-9.2000	0.2443	(0.2011 , 0.2876)	0.5457	(0.5020, 0.5894)
γ_2	-9.0019	0.2544	(0.2105 , 0.2982)	0.1807	(0.1345, 0.2269)
γ_3	-8.9237	0.2490	(0.2055 , 0.2925)	0.1459	(0.1008, 0.1909)
γ_4	-8.8728	0.2619	(0.2168 , 0.3071)	0.1525	(0.1053, 0.1998)
γ_5	-8.8349	0.2629	(0.2173 , 0.3084)	0.1307	(0.0830, 0.1783)
γ_6	-8.8047	0.2620	(0.2155 , 0.3085)	0.1434	(0.0951, 0.1916)
γ_7	-8.7796	0.2596	(0.2140 , 0.3051)	0.1453	(0.0977, 0.1929)
γ_8	-8.7581	0.2592	(0.2137 , 0.3046)	0.1852	(0.1382, 0.2322)
γ_9	-8.7393	0.2638	(0.2174 , 0.3103)	0.2451	(0.1964, 0.2937)
γ_{10}	-8.7226	0.2556	(0.2089 , 0.3023)	0.4062	(0.3569, 0.4555)

Parameter	True Value	EVENT INDEPENDENTLY COLLECTED MSE	Standard Error of MSE	EVENT DEPENDENTLY COLLECTED MSE	Standard Error of MSE
β	0.07	4.065$\times 10^{-5}$	2.44$\times 10^{-6}$	3.405$\times 10^{-5}$	2.03$\times 10^{-6}$
γ_1	-9.2000	0.3029	0.0197	0.5461	0.0288
γ_2	-9.0019	0.3144	0.0199	0.3095	0.0184
γ_3	-8.9237	0.3077	0.0187	0.2847	0.0173
γ_4	-8.8728	0.3332	0.0201	0.3132	0.0190
γ_5	-8.8349	0.3389	0.0197	0.3119	0.0179
γ_6	-8.8047	0.3493	0.0214	0.3232	0.0196
γ_7	-8.7796	0.3367	0.0207	0.3154	0.0198
γ_8	-8.7581	0.3356	0.0201	0.3213	0.0189
γ_9	-8.7393	0.3501	0.0214	0.3680	0.0218
γ_{10}	-8.7226	0.3485	0.0203	0.4806	0.0261

Table B.9: Comparison of bias and MSE from simulations of a Weibull mortality process with shape parameter 1.15 and $\sigma_W = 1.5$ and $\sigma_B = 8.0$ using 'adapted Finkelstein' with a single continuous repeated measurement with 50% data missing via a MCAR process under two different situations for the assessment of the interval containing the event of interest using linear interpolation imputation.

Parameter	True Value	EVENT INDEPENDENTLY COLLECTED Bias	95% C.I. for bias	EVENT DEPENDENTLY COLLECTED Bias	95% C.I. for bias
β	0.07	0.255×10^{-3}	$(-0.232\times10^{-3}, 0.742\times10^{-3})$	0.133×10^{-3}	$(-0.335\times10^{-3}, 0.602\times10^{-3})$
γ_1	-9.2000	-0.0272	(-0.0711, 0.0167)	0.0011	(-0.0401, 0.0422)
γ_2	-9.0019	-0.0297	(-0.0734, 0.0140)	-0.0133	(-0.0553, 0.0287)
γ_3	-8.9237	-0.0309	(-0.0749, 0.0132)	-0.0210	(-0.0643, 0.0222)
γ_4	-8.8728	-0.0350	(-0.0802, 0.0102)	-0.0138	(-0.0575, 0.0300)
γ_5	-8.8349	-0.0232	(-0.0688, 0.0223)	-0.0129	(-0.0553, 0.0295)
γ_6	-8.8047	-0.0389	(-0.0850, 0.0072)	-0.0207	(-0.0644, 0.0231)
γ_7	-8.7796	-0.0353	(-0.0804, 0.0098)	-0.0261	(-0.0706, 0.0184)
γ_8	-8.7581	-0.0308	(-0.0752, 0.0137)	-0.0306	(-0.0749, 0.0138)
γ_9	-8.7393	-0.0344	(-0.0806, 0.0119)	-0.0115	(-0.0569, 0.0339)
γ_{10}	-8.7226	-0.0432	(-0.0895, 0.0030)	-0.0048	(-0.0494, 0.0398)

Parameter	True Value	EVENT INDEPENDENTLY COLLECTED MSE	Standard Error of MSE	EVENT DEPENDENTLY COLLECTED MSE	Standard Error of MSE
β	0.07	3.084×10^{-5}	1.78×10^{-6}	2.850×10^{-5}	1.66×10^{-6}
γ_1	-9.2000	0.2514	0.0150	0.2198	0.0127
γ_2	-9.0019	0.2488	0.0147	0.2293	0.0137
γ_3	-8.9237	0.2533	0.0153	0.2430	0.0140
γ_4	-8.8728	0.2664	0.0152	0.2489	0.0152
γ_5	-8.8349	0.2698	0.0159	0.2337	0.0133
γ_6	-8.8047	0.2774	0.0172	0.2488	0.0150
γ_7	-8.7796	0.2656	0.0157	0.2579	0.0151
γ_8	-8.7581	0.2574	0.0146	0.2562	0.0151
γ_9	-8.7393	0.2790	0.0163	0.2682	0.0154
γ_{10}	-8.7226	0.2800	0.0174	0.2586	0.0152

Table B.10: Comparison of bias and MSE from simulations of a Weibull mortality process with shape parameter 1.15 and $\sigma_W = 1.5$ and $\sigma_B = 8.0$ using 'adapted Finkelstein' with a single continuous repeated measurement with 5% data missing via a MCAR process under two different situations for the assessment of the interval containing the event of interest using multiple imputation with five imputations for each missing observation.

Parameter	True Value	EVENT INDEPENDENTLY COLLECTED Bias	95% C.I. for bias	EVENT DEPENDENTLY COLLECTED Bias	95% C.I. for bias
β	0.07	0.053×10^{-3}	$(-0.541\times10^{-3}, 0.436\times10^{-3})$	0.350×10^{-3}	$(-0.113\times10^{-3}, 0.814\times10^{-3})$
γ_1	-9.2000	0.0075	(-0.0369, 0.0519)	0.0056	(-0.0359, 0.0470)
γ_2	-9.0019	-0.0024	(-0.0465, 0.0418)	-0.0266	(-0.0699, 0.0167)
γ_3	-8.9237	-0.0062	(-0.0512, 0.0388)	-0.0274	(-0.0693, 0.0144)
γ_4	-8.8728	-0.0011	(-0.0455, 0.0433)	-0.0319	(-0.0752, 0.0115)
γ_5	-8.8349	-0.0040	(-0.0495, 0.0415)	-0.0321	(-0.0749, 0.0108)
γ_6	-8.8047	0.0023	(-0.0440, 0.0486)	-0.0376	(-0.0820, 0.0068)
γ_7	-8.7796	-0.0023	(-0.0484, 0.0438)	-0.0301	(-0.0735, 0.0133)
γ_8	-8.7581	-0.0051	(-0.0505, 0.0403)	-0.0404	(-0.0835, 0.0027)
γ_9	-8.7393	-0.0102	(-0.0575, 0.0371)	-0.0285	(-0.0721, 0.0150)
γ_{10}	-8.7226	-0.0174	(-0.0634, 0.0286)	-0.0026	(-0.0470, 0.0418)

Parameter	True Value	EVENT INDEPENDENTLY COLLECTED MSE	Standard Error of MSE	EVENT DEPENDENTLY COLLECTED MSE	Standard Error of MSE
β	0.07	3.099×10^{-5}	1.93×10^{-6}	2.803×10^{-5}	1.78×10^{-6}
γ_1	-9.2000	0.2562	0.0160	0.2234	0.0145
γ_2	-9.0019	0.2531	0.0157	0.2444	0.0153
γ_3	-8.9237	0.2627	0.0172	0.2279	0.0140
γ_4	-8.8728	0.2563	0.0153	0.2451	0.0171
γ_5	-8.8349	0.2693	0.0164	0.2393	0.0144
γ_6	-8.8047	0.2788	0.0177	0.2576	0.0166
γ_7	-8.7796	0.2757	0.0167	0.2455	0.0160
γ_8	-8.7581	0.2679	0.0168	0.2431	0.0145
γ_9	-8.7393	0.2903	0.0185	0.2475	0.0160
γ_{10}	-8.7226	0.2754	0.0167	0.2566	0.0159

Table B.11: Comparison of bias and MSE from simulations of a Weibull mortality process with shape parameter 1.15 and $\sigma_W = 1.5$ and $\sigma_B = 8.0$ using 'adapted Finkelstein' with a single continuous repeated measurement with 10% data missing via a MCAR process under two different situations for the assessment of the interval containing the event of interest using multiple imputation with five imputations for each missing observation.

Parameter	True Value	EVENT INDEPENDENTLY COLLECTED Bias	95% C.I. for bias	EVENT DEPENDENTLY COLLECTED Bias	95% C.I. for bias
β	0.07	-0.492×10^{-3}	$(-0.944\times10^{-3}, -0.039\times10^{-3})$	0.149×10^{-3}	$(-0.357\times10^{-3}, 0.656\times10^{-3})$
γ_1	-9.2000	0.0273	(-0.0147, 0.0693)	0.1067	(0.0623, 0.1510)
γ_2	-9.0019	0.0244	(-0.0164, 0.0652)	0.0005	(-0.0452, 0.0463)
γ_3	-8.9237	0.0308	(-0.0097, 0.0712)	-0.0073	(-0.0525, 0.0380)
γ_4	-8.8728	0.0335	(-0.0094, 0.0764)	-0.0161	(-0.0627, 0.0306)
γ_5	-8.8349	0.0331	(-0.0099, 0.0761)	-0.0066	(-0.0548, 0.0416)
γ_6	-8.8047	0.0340	(-0.0096, 0.0776)	-0.0191	(-0.0671, 0.0288)
γ_7	-8.7796	0.0278	(-0.0146, 0.0702)	-0.0160	(-0.0639, 0.0318)
γ_8	-8.7581	0.0279	(-0.0149, 0.0706)	-0.0204	(-0.0669, 0.0261)
γ_9	-8.7393	0.0349	(-0.0083, 0.0780)	-0.0011	(-0.0491, 0.0469)
γ_{10}	-8.7226	0.0283	(-0.0152, 0.0718)	0.0741	(0.0259, 0.1224)

Parameter	True Value	EVENT INDEPENDENTLY COLLECTED MSE	Standard Error of MSE	EVENT DEPENDENTLY COLLECTED MSE	Standard Error of MSE
β	0.07	2.684×10^{-5}	1.64×10^{-6}	3.336×10^{-5}	1.88×10^{-6}
γ_1	-9.2000	0.2301	0.0145	0.2670	0.0161
γ_2	-9.0019	0.2168	0.0132	0.2723	0.0156
γ_3	-8.9237	0.2135	0.0133	0.2660	0.0156
γ_4	-8.8728	0.2402	0.0147	0.2829	0.0165
γ_5	-8.8349	0.2412	0.0151	0.3023	0.0177
γ_6	-8.8047	0.2479	0.0155	0.2991	0.0172
γ_7	-8.7796	0.2343	0.0146	0.2973	0.0178
γ_8	-8.7581	0.2378	0.0145	0.2816	0.0158
γ_9	-8.7393	0.2432	0.0154	0.2988	0.0170
γ_{10}	-8.7226	0.2467	0.0160	0.3085	0.0182

Table B.12: Comparison of bias and MSE from simulations of a Weibull mortality process with shape parameter 1.15 and $\sigma_W = 1.5$ and $\sigma_B = 8.0$ using 'adapted Finkelstein' with a single continuous repeated measurement with 20% data missing via a MCAR process under two different situations for the assessment of the interval containing the event of interest using multiple imputation with five imputations for each missing observation.

Parameter	True Value	EVENT INDEPENDENTLY COLLECTED Bias	95% C.I. for bias	EVENT DEPENDENTLY COLLECTED Bias	95% C.I. for bias
β	0.07	-1.665×10^{-3}	$(-2.149\times10^{-3}, -1.180\times10^{-3})$	-0.882×10^{-3}	$(-1.367\times10^{-3}, -0.397\times10^{-3})$
γ_1	-9.2000	0.1260	(0.0836, 0.1684)	0.5073	(0.4641, 0.5504)
γ_2	-9.0019	0.1437	(0.1004, 0.1871)	0.1727	(0.1276, 0.2177)
γ_3	-8.9237	0.1207	(0.0767, 0.1648)	0.1290	(0.0845, 0.1736)
γ_4	-8.8728	0.1332	(0.0884, 0.1780)	0.1311	(0.0863, 0.1760)
γ_5	-8.8349	0.1386	(0.0937, 0.1835)	0.1070	(0.0628, 0.1511)
γ_6	-8.8047	0.1373	(0.0922, 0.1824)	0.1199	(0.0735, 0.1663)
γ_7	-8.7796	0.1314	(0.0860, 0.1769)	0.1206	(0.0740, 0.1672)
γ_8	-8.7581	0.1343	(0.0897, 0.1788)	0.1503	(0.1053, 0.1953)
γ_9	-8.7393	0.1326	(0.0871, 0.1781)	0.2135	(0.1655, 0.2614)
γ_{10}	-8.7226	0.1268	(0.0812, 0.1724)	0.3670	(0.3198, 0.4142)

Parameter	True Value	EVENT INDEPENDENTLY COLLECTED MSE	Standard Error of MSE	EVENT DEPENDENTLY COLLECTED MSE	Standard Error of MSE
β	0.07	3.326×10^{-5}	2.12×10^{-6}	3.132×10^{-5}	1.91×10^{-6}
γ_1	-9.2000	0.2492	0.0164	0.4990	0.0271
γ_2	-9.0019	0.2651	0.0167	0.2937	0.0180
γ_3	-8.9237	0.2664	0.0169	0.2744	0.0182
γ_4	-8.8728	0.2784	0.0181	0.2785	0.0168
γ_5	-8.8349	0.2806	0.0175	0.2647	0.0166
γ_6	-8.8047	0.2832	0.0183	0.2940	0.0185
γ_7	-8.7796	0.2856	0.0177	0.2966	0.0179
γ_8	-8.7581	0.2761	0.0178	0.2856	0.0184
γ_9	-8.7393	0.2863	0.0179	0.3437	0.0208
γ_{10}	-8.7226	0.2863	0.0179	0.4238	0.0237

Table B.13: Comparison of bias and MSE from simulations of a Weibull mortality process with shape parameter 1.15 and $\sigma_W = 1.5$ and $\sigma_B = 8.0$ using 'adapted Finkelstein' with a single continuous repeated measurement with 50% data missing via a MCAR process under two different situations for the assessment of the interval containing the event of interest using multiple imputation with five imputations for each missing observation.

Appendix C

The log-likelihood and its derivatives for use in the ordinal variable multiple imputation procedure

The multiple imputation for the four category ordinal variable was based on a transition proportional odds regression model (Diggle, Liang and Zeger, 1994) which is a special case of the proportional odds model (McCullagh, 1980).

The form of the transition proportional odds model is:

$$\eta_{ijk} = \log\left[\frac{P(Z_{ij} \leq k \mid Z_{ij-1} = z_{ij-1})}{P(Z_{ij} > k \mid Z_{ij-1} = z_{ij-1})}\right] = \theta_k + \sum_{m=1}^{3} \tau_{mk} z^*_{ij-1\,m} \qquad \text{(C.1)}$$

where $$z^*_{ijm} = \begin{cases} 1 & \text{if } z_{ij} \leq m \\ 0 & \text{if } z_{ij} > m \end{cases}$$

$$\underline{\boldsymbol{\theta}}^T = (\theta_1, \theta_2, \theta_3, \tau_{11}, \tau_{12}, \tau_{13}, \tau_{21}, \tau_{22}, \tau_{23}, \tau_{31}, \tau_{32}, \tau_{33})$$

and Z_{ij} is the ordinal variable for the i^{th} individual at the start of the $j+1^{\text{th}}$ year.

Using an extension of the notation used in McCullagh (1980), it is convenient to let

$$P(Z_{ij} = z_{ij} \mid Z_{ij-1} = z_{ij-1}) = \pi_k(z_{ij-1}), \qquad k = 1, \ldots, 4$$

represent the probability that the ordinal variable for the i^{th} individual takes the value k at the start of the $j+1^{\text{th}}$ year, given that it takes the value z_{ij-1} at the start of the j^{th} year and to let $\gamma_k(z_{ij-1}) = \sum_{l=1}^{k} \pi_l(z_{ij-1})$, $k = 1, \ldots, 4$, represent the corresponding cumulative probabilities. Equation C.1 can then be written:

$$\log\left(\frac{\gamma_k(z_{ij-1})}{1 - \gamma_k(z_{ij-1})}\right) = \theta_k + \sum_{m=1}^{3} \tau_{mk} z^*_{ij-1\,m} \quad .$$

The contribution from observation on the i^{th} individual at the start of the $j+1^{\text{th}}$ year to this likelihood function, as given in McCullagh (1980) and based on the product of three components is, on supressing the conditioning on z_{ij-1} from the notation,

$$\left\{\left(\frac{\gamma_1}{\gamma_2}\right)^{z^*_{ij1}} \left(\frac{\gamma_2 - \gamma_1}{\gamma_2}\right)^{z^*_{ij2} - z^*_{ij1}}\right\} \left\{\left(\frac{\gamma_2}{\gamma_3}\right)^{z^*_{ij2}} \left(\frac{\gamma_3 - \gamma_2}{\gamma_3}\right)^{z^*_{ij3} - z^*_{ij2}}\right\}$$

$$\times \left\{(\gamma_3)^{z^*_{ij3}} (1 - \gamma_3)^{1 - z^*_{ij3}}\right\} \quad .$$

The first component represents the probability, given z^*_{ij-12}, that the first two cells divide in the ratio $z^*_{ij2} - z^*_{ij1} : z^*_{ij1}$. In the general proportional odds model as proposed by McCullagh (1980), this partitioning is more appropriate as each

'observation' is a multinomial one and so the random variables representing the cumulative observations are, unlike the z^*_{ijm} in this special case, not restricted to the values 0 and 1. The second and third components represent, respectively, the probability, given z^*_{ij-13}, that the third cell and the first two cells taken together divide in the ratio $z^*_{ij3} - z^*_{ij2} : z^*_{ij2}$ and the probability that the fourth cell and the first three cells taken together divide in the ratio $z^*_{ij4} - z^*_{ij3} : z^*_{ij3}$.

Defining

$$\phi_k = \log\left\{\frac{\gamma_k}{(\gamma_{k+1} - \gamma_k)}\right\} = \text{logit}\left(\frac{\gamma_k}{\gamma_{k+1}}\right)$$

and

$$g(\phi_k) = \log\{1 + \exp(\phi_k)\} = \log\left\{\frac{\gamma_{k+1}}{(\gamma_{k+1} - \gamma_k)}\right\},$$

the contribution from the observation on the i^{th} individual at the start of the $j+1^{\text{th}}$ year to the **log**-likelihood function is

$$l_{ij} = \left(z^*_{ij1}\phi_1 - z^*_{ij2}g(\phi_1)\right) + \left(z^*_{ij2}\phi_2 - z^*_{ij3}g(\phi_2)\right) + \left(z^*_{ij3}\phi_3 - g(\phi_3)\right).$$

Differentiating this log-likelihood contribution with repect to $\underline{\theta}$ gives

$$\frac{\partial l_{ij}}{\partial \theta_r} = \sum_{m=1}^{3} \frac{\partial l_{ij}}{\partial \phi_m}\left(\frac{\partial \phi_m}{\partial \gamma_m} \cdot \frac{\partial \gamma_m}{\partial \eta_{ijm}} \cdot \frac{\partial \eta_{ijm}}{\partial \theta_r} + \frac{\partial \phi_m}{\partial \gamma_{m+1}} \cdot \frac{\partial \gamma_{m+1}}{\partial \eta_{ij\,m+1}} \cdot \frac{\partial \eta_{ij\,m+1}}{\partial \theta_r}\right) \qquad \text{(C.2)}.$$

For the transition proportional odds model used here,

$$\frac{\partial l_{ij}}{\partial \phi_m} = z^*_{ijm} - z^*_{ijm+1}\frac{\partial g(\phi_m)}{\partial \phi_m} \qquad m = 1, \ldots, 3,$$

where

$$\begin{aligned}\frac{\partial g(\phi_m)}{\partial \phi_m} &= \frac{\partial}{\partial \phi_m}\left[\log\left\{1 + e^{\phi_m}\right\}\right] \\ &= \frac{e^{\phi_m}}{1 + e^{\phi_m}} \quad .\end{aligned}$$

So

$$\frac{\partial l_{ij}}{\partial \phi_m} = z^*_{ijm} - z^*_{ijm+1}\frac{e^{\phi_m}}{1 + e^{\phi_m}} \qquad m = 1, \ldots, 3\,.$$

Now, as $\phi_m = \text{logit}\left(\frac{\gamma_m}{\gamma_{m+1}}\right)$, $\frac{e^{\phi_m}}{1 + e^{\phi_m}} = \frac{\gamma_m}{\gamma_{m+1}} \qquad m = 1, \ldots, 3,$

this can be written in the alternative form

$$\frac{\partial l_{ij}}{\partial \phi_m} = z^*_{ijm} - z^*_{ijm+1}\frac{\gamma_m}{\gamma_{m+1}} \qquad m = 1, \ldots, 3\,.$$

It is this last form that was used in the Fortran subroutine written for implementing the likelihood estimation of the parameters of the transition proportional odds model in Chapter 4.

Differentiating ϕ_m partially with respect to γ_m gives

$$\frac{\partial \phi_m}{\partial \gamma_m} = \frac{\gamma_{m+1}}{\gamma_m\left(\gamma_{m+1} - \gamma_m\right)},$$

and similarly with respect to γ_{m+1} gives

$$\frac{\partial \phi_m}{\partial \gamma_{m+1}} = -\frac{1}{\gamma_{m+1} - \gamma_m} \quad .$$

Differentiating γ_m with respect to η_{ijm} gives

$$\frac{\partial \gamma_m}{\partial \eta_{ijm}} = \frac{e^{\eta_{ijm}}}{(1 + e^{\eta_{ijm}})^2} ,$$

and differentiating γ_{m+1} with respect to $\eta_{ij\,m+1}$ gives

$$\frac{\partial \gamma_{m+1}}{\partial \eta_{ij\,m+1}} = \frac{e^{\eta_{ij\,m+1}}}{(1 + e^{\eta_{ij\,m+1}})^2} \quad .$$

From (C.1),

$$\eta_{ijk} = \theta_k + \sum_{m=1}^{3} \tau_{mk} z^*_{i\,j-1\,m} ,$$

so

$$\frac{\partial \eta_{ijm}}{\partial \theta_r} = \begin{cases} 1 & \text{if } r = m \\ 0 & \text{otherwise} \end{cases} ,$$

and

$$\frac{\partial \eta_{ijm}}{\partial \tau_{pr}} = \begin{cases} z^*_{i\,j-1\,p} & \text{if } r = m \\ 0 & \text{otherwise} \end{cases} .$$

Similarly,

$$\frac{\partial \eta_{ij\,m+1}}{\partial \theta_r} = \begin{cases} 1 & \text{if } r = m+1 \\ 0 & \text{otherwise} \end{cases} ,$$

and

$$\frac{\partial \eta_{ij\,m+1}}{\partial \tau_{pr}} = \begin{cases} z^*_{i\,j-1\,p} & \text{if } r = m+1 \\ 0 & \text{if otherwise} \end{cases} .$$

The log-likelihood function and its first partial derivatives are then obtained by summing these contributions over all observed screens for all the individuals. The NAG subroutine E04KCF used in the maximisation process used the value of the log-likelihood function and the values of its first, but not its second, partial derivatives.

It has been shown (McCullagh (1980)) that the derivative of the log-likelihood with respect to the parameter vector $\underline{\theta}$ given in (C.2) can be written as

$$\frac{\partial l_{ij}}{\partial \theta_r} = \sum_{m=1}^{3} \frac{\partial l_{ij}}{\partial \phi_m} V_m^{-1} q_{ijmr} ,$$

where $$V_m = \frac{\partial \gamma_m}{\partial \phi_m} , \quad \frac{\partial \phi_m}{\partial \gamma_{m+1}} = \left(-\frac{\gamma_m}{\gamma_{m+1}}\right) V_m^{-1}$$

and $$q_{ijmr} = \left\{ \frac{\partial \gamma_m}{\partial \eta_{ijm}} \frac{\partial \eta_{ijm}}{\partial \theta_r} - \frac{\gamma_m}{\gamma_{m+1}} \frac{\partial \gamma_{m+1}}{\partial \eta_{ij\,m+1}} \frac{\partial \eta_{ij\,m+1}}{\partial \theta_r} \right\} .$$

Similarly, the contribution to the $(rs)^{\text{th}}$ element of the Fisher information matrix from the i^{th} individual at the start of the $j+1^{\text{th}}$ year was shown to be given by

$$A_{ijrs} = -\mathrm{E}\left(\frac{\partial^2 l_{ij}}{\partial\beta_r\partial\beta_s}\right) = \sum_{m=1}^{3} V_m^{-1} q_{ijmr} q_{ijms} \quad .$$

Whilst the Fisher information was not used in the NAG subroutine to maximise the log-likelihood, its inverse was used as an approximation to the variance-covariance matrix of the vector of parameter estimates.

Bibliography

Afifi, A.A. and Elashoff, R.M. 'Missing observations in multivariate statistics III: large sample analysis of simple linear regression', *Journal of the American Statistical Association*, **64**, 337-358 (1969a).

Afifi, A.A. and Elashoff, R.M. 'Missing observations in multivariate statistics IV: a note on simple linear regression', *Journal of the American Statistical Association*, **64**, 359-365 (1969b).

Altman, D.G and De Stavola, B.L. 'Practical problems in fitting a proportional hazards model to data with updated measurements of the covariates', *Statistics in Medicine*, **13**, 301-341 (1994).

Andersen, P.K. 'Repeated assessment of risk factors in survival analysis', *Statistical Methods in Medical Research*, **1**, 297-315 (1992).

Arthur, A.J., Clarke, M., Donaldson, M.M.K. and Jagger, C. 'Community surveys of older people: how to minimise non-response bias', unpublished manuscript.

BMDP Statistical Software, Inc., *BMDP Software Manual, Volume 2*, University of California Press, Berkeley, California, 1990.

Buck, S.F. 'A method of estimation of missing values in multivariate data suitable for use with an electronic computer' *Journal of the Royal Statistical Society, B*, **37**, 302-306 (1960).

Carter, R.L., Resnick, M.B., Ariet, M. and Shieh, G. 'A random coefficient growth curve analysis of mental development in low-birthweight infants', *Statistics in Medicine*, **11**, 243-256 (1992).

Clarke, M., Clarke, S.J., Odell, A. and Jagger, C. 'The elderly at home: health and social status', *Health Trends*, **16**, 3-7 (1984).

Clarke, M., Jagger, C., Anderson, J., Battcock, T., Kelly, F. and Stern, M.C., 'The prevalence of dementia in a total population - a comparison of two screening instruments', *Age and Ageing*, **20**, 396-403 (1991).

Collett, D. *Modelling Survival Data in Medical Research*, Chapman and Hall, London, 1994.

Cox, D.R. 'Regression models and life-tables', *Journal of the Royal Statistical Society, B*, **74**, 187-220 (1972).

Cox, D.R. 'Partial likelihood', *Biometrika*, **62**, 269-276 (1975).

Crépeau, H., Koziol, J., Reid, N. and Yuh, Y.S. 'Analysis of incomplete multivariate data from repeated measurement experiments', *Biometrics*, **41**, 505-514 (1985).

Crowder, M.J. and Hand, D.J. *Analysis of Repeated Measures*, Chapman and Hall, London, 1990.

Cupples, L.A., D'Agostino, R.B., Anderson, K. and Kannel, W.B. 'Comparison of baseline and repeated measure covariate techniques in the Framingham Heart Study', *Statistics in Medicine*, **7**, 205-218 (1988).

Dempster, A.P., Little, R.J.A. and Rubin, D.B. 'Maxiumum likelihood from incomplete data via the EM algorithm' *Journal of the Royal Statistical Society, B*, **39**, 1-38 (1977).

Diggle, P. and Kenward, M.G. 'Informative drop-out in longitudinal data analysis', *Applied Statistics*, **43**, 49-93 (1994).

Diggle, P.J. 'An approach to the analysis of repeated measurements', *Biometrics*, **44**, 959-971 (1988).

Diggle, P.J. 'Testing for random dropouts in repeated measurement data', *Biometrics*, **45**, 1255-1258 (1989).

Diggle, P.J., Liang, K.-Y. and Zeger, S.L. *Analysis of Longitudinal Data*, Oxford University Press, Oxford, 1994.

Dorey, F.J., Little, R.J.A. and Schenker, N. 'Multiple imputation for threshold-crossing data with interval censoring', *Statistics in Medicine*, **12**, 1589-1603 (1993).

Dwyer, J.H., 'Differential equation models for longitudinal data', in Dwyer, J.H., Feinleib, M., Lippert, P. and Hoffmeister, H. (eds.), *Statistical Models for Longitudinal Studies of Health*, Oxford University Press, Oxford, 1992, pp. 71-98.

Finkelstein, D.M. 'A proportional hazards model for interval-censored failure time data', *Biometrics*, **42**, 845-854 (1986).

Goldstein, H. *Multilevel Statistical Models*, Arnold, London, 1995.

Grüger, J., Kay, R. and Schumacher, M. 'The validity of inferences based on incomplete observations in disease state models', *Biometrics*, **47**, 595-605 (1991).

Hsieh, F.Y, Crowley, J. and Tormey, D.C. 'Some test statistics for use in multistate survival analysis', *Biometrika*, **70**, 111-119 (1983).

Hughes, M.D. and Pocock, S.J. 'Within-subject diastolic blood pressure variablility: implications for risk assessment and screening', *Journal of Clinical Epidemiology*, **45**, 985-998 (1992).

Jagger, C. and Clarke, M. 'Mortality risks in the elderly: five-year follow-up of a total population', *International Journal of Epidemiology*, **17**, 111-114.

Jagger, C., Clarke, M. and Clarke, S.J. 'Getting older - feeling younger: the changing health profile of the elderly', *International Journal of Epidemiology*, **20**, 234-238 (1991).

Jagger, C., Clarke, M. and Cooke, A.J. 'Mental and physical health of elderly people: five-year follow up of a total population', *Age and Ageing*, **18**, 77-82 (1989).

Jagger, C., Clarke, M. and Davies, R.A. 'The elderly at home: indices of disability', *Journal of Epidemiology and Community Health*, **40**, 139-143 (1986).

Jagger, C., Clarke, M, O'Shea, C. and Gannon, M. 'Annual visits to patients over the age of 75 - who is missed?', *Family Practice*, **13**, 22-27 (1996).

Jagger, C., Spiers, N.A. and Clarke, M. 'Factors associated with decline in function, institutionalisation and mortality in the elderly', *Age and Ageing*, **22**, 190-197 (1993).

Jagger, C. and Sutton, C.J. 'Death after marital bereavement - is the risk increased?', *Statistics in Medicine*, **10**, 395-404 (1991).

Jennrich, R.I. and Schluchter, M.D. 'Unbalanced repeated measures models with structured covariance matrices', *Biometrics*, **42**, 805-820 (1986).

Kalbfleisch, J.D. and Prentice, R.L. 'Marginal likelihoods based on Cox's regression and life model', *Biometrika*, **60**, 267-278 (1973).

Kaplan, E.L. and Meier, P. 'Nonparametric estimation from incomplete observations', *Journal of the American Statistical Society*, **53**, 457-481 (1958).

Kuk, A.Y.C. and Chen, C.-H. 'A mixture model combining logistic regression with proportional hazards regression', *Biometrika*, **79**, 531-541 (1992).

Laird, N.M. 'Missing data in longitudinal studies', *Statistics in Medicine*, **7**, 305-315 (1988).

Laird, N.M., Lange, N. and Stram, D. 'Maximum likelihood computations with repeated measures: application of the EM algorithm', *Journal of the American Statistical Association*, **82**, 97-105 (1987).

Laird, N.M., and Ware, J.H. 'Random-effects models for longitudinal data', *Biometrics*, **38**, 963-974 (1982).

Lan, K.K.G. and DeMets, D.L. 'Discrete sequential boundaries for clinical trials', *Biometrika*, **70**, 659-663 (1983).

Liang, K.-Y. and Zeger, S.L. 'Longitudinal data analysis using generalized linear models', *Biometrika*, **73**, 13-22 (1986).

Liang, K.-Y., Zeger, S.L. and Qaqish B. 'Multivariate regression analysis for categorical data', *Journal of the Royal Statistical Society B*, **54**, 3-40 (1992).

Lindsey, J.C. and Ryan, L.M. 'A three-state multiplicative model for rodent tumorigenicity experiments', *Applied Statistics*, **42**, 283-300 (1993).

Lindsey, J.K. *Models for Repeated Measurements*, Oxford Science Publications, Oxford, 1993.

Little, R.J.A. 'Regression with missing X's : a review', *Journal of the American Statistical Association*, **87**, 1227-1237 (1992a).

Little, R.J.A. 'Incomplete data in event history analysis', in Trussell, J, Hankinson, R. and Tilton, J. (eds.), *Demographic Applications of Event History Analysis*, Oxford University Press, Oxford, 1992b, pp. 209-230.

Little, R.J.A. 'Modelling the drop-out mechanism in repeated-measures studies', *Journal of the American Statistical Association*, **90**, 1112-1121 (1995).

Little, R.J.A., and Rubin, D.B. *Statistical Analysis with Missing Data*, John Wiley, New York, 1987.

Little, R.J.A. and Schluchter, M.D. 'Maximum likelihood estimation for mixed continuous and categorical data with missing values', *Biometrika*, **72**, 497-512 (1985).

McCullagh, P. 'Regression models for ordinal data', *Journal of the Royal Statistical Society, B*, **42**, 109-142 (1980).

Manton, K.G. 'Health forecasting and models of aging', in Manton, K.G., Singer, B.H., Suzman, R.M. (eds.), *Forecasting the Health of Elderly Populations*, Springer-Verlag, New York, 1993, pp. 79-106.

Manton, K.G. Dowd, J.E. and Stallard, E. 'The effects of risk factors on male and female cardiovascular risks in middle and late age', in Manton, K.G., Singer, B.H., Suzman, R.M. (eds.), *Forecasting the Health of Elderly Populations*, Springer-Verlag, New York, 1993, pp. 137-158.

Manton, K.G., Stallard, E. and Liu, K. 'Frailty and forecasts of active life expectancy in the United States', in Manton, K.G., Singer, B.H., Suzman, R.M. (eds.), *Forecasting the Health of Elderly Populations*, Springer-Verlag, New York, 1993, 159-181.

Manton, K.G., Stallard, E. and Woodbury, M.A. 'Chronic disease evolution and human aging: a general model for assessing the impact of chronic disease, in human populations', *Mathematical Modelling*, **7**, 1155-1171 (1986).

Manton, K.G., Woodbury, M.A. and Stallard E. 'Models of the interaction of mortality and the evolution of risk factor distribution : a general stochastic process formulation', *Statistics in Medicine*, **7**, 239-256 (1988).

Matthews, J.N.S. 'A refinement to the analysis of serial data using summary measures', *Statistics in Medicine*, **12**, 27-37 (1993).

NAG *Fortran Subroutine Workstation Library*, Numerical Algorithms Group, Oxford, 1991.

Orchard, T. and Woodbury, M.A. 'A missing information principle: theory and applications', *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley, University of California Press, 1972, pp. 697-715.

Pattie, A.H. and Gilleard, C.J. *Manual of the Clifton Assessment Procedure of the Elderly (CAPE)*, Sevenoaks, Hodder and Stoughton, 1979.

Peto, R. 'Contribution to the discussion of a paper by D.R.Cox', *Journal of the Royal Statistical Society, B*, **34**, 205-207 (1972).

Prentice, R.L. and Gloeckler L.A. 'Regression analysis of grouped survival data with application to breast cancer data', *Biometrics*, **34**, 57-67 (1978).

Reilly, M. 'Data analysis using hot deck multiple imputation', *The Statistician*, **42**, 307-313 (1993).

Rubin, D.B. 'Inference and missing data', *Biometrics*, **48**, 970-972 (1976).

Rubin, D.B. *Multiple Imputation for Nonresponse in Surveys*, John Wiley, New York, 1987.

SAS Institute Inc., *SAS/STAT Software: Changes and Enhancements, through Release 6.11*, SAS Institute Inc., Cary, North Carolina, U.S.A., 1996.

Schluchter, M.D. 'Analysis of incomplete multivariate data using linear models with structured covariance matrices', *Statistics in Medicine*, **7**, 317-324 (1988).

Schluchter, M.D. 'Methods for the analysis of informatively censored longitudinal data', *Statistics in Medicine*, **11**, 1861-1870 (1992).

Schluchter, M.D. and Jackson K.L. 'Log-linear analysis of censored survival data with partially observed covariates', *Journal of the American Statistical Association*, **84**, 42-52 (1989).

Shih, W.J. 'On informative and random dropouts in longitudinal studies', Letter to the editor of *Biometrics*, **48**, 970-972 (1992).

Sinha, D., Tanner, M.A. and Hall, W.J. 'Maximization of the marginal likelihood of grouped survival data', *Biometrika*, **81**, 53-60 (1994).

Tanner, M.A. and Wong, W.H. 'Data-based non-parametric estimation of the hazard function with applications to model diagnostics and exploratory analysis', *Journal of the American Statistical Association*, **79**, 174-182 (1984).

Tanner, M.A. and Wong, W.H. 'The calculation of posterior distributions by data augmentation', *Journal of the American Statistical Association*, **82**, 528-550 (1987a).

Tanner, M.A. and Wong, W.H. 'An application of imputation to an estimation problem in grouped lifetime analysis', *Technometrics*, **29**, 23-32 (1987b).

Thomas, D. 'Use of biologic models to analyse epidemiologic data', in Dwyer, J.H., Feinleib, M., Lippert, P. and Hoffmeister, H. (eds.), *Statistical Models for Longitudinal Studies of Health*, Oxford University Press, Oxford.

Thompson, R. 'Survival data and GLIM', Letter to the editor of *Applied Statistics*, **30**, 310.

Tu, X.M., Meng, X.-L. and Pagano, M. 'The AIDS epidemic: estimating survival after AIDS diagnosis from surveillance data', *Journal of the American Statistical Association*, **88**, 421, 26-36 (1993).

Vacek, P.M., Mickey, R.M. and Bell, D.Y. 'Application of a two-stage model to longitudinal pulmonary function data from sarcoidosis patients', *Statistics in Medicine*, 8, 189-200 (1989).

Vonesh, E.F. and Carter, R.L. 'Efficient inference for random-coefficient growth curve models with unbalanced data', *Biometrics*, **43**, 617-628 (1987).

Wang, R., Sedransk, J. and Jinn, J.H. 'Secondary data analysis when there are missing observations', *Journal of the American Statistical Association*, **87**, 952-961 (1992).

Whitehead, J. 'The analysis of relapse trials with application to a comparison of two ulcer treatments', *Statistics in Medicine*, 8, 1439-1454 (1989).

Woodbury, M.A. and Manton, K.G. 'A random walk model of human mortality and aging', *Theoretical Population Biology*, **11**, 37-48 (1977).

Woodbury, M.A., Manton, K.G. and Stallard, E. 'Longitudinal analysis of the dynamics and risk of coronary heart disease in the Framingham Study', *Biometrics*, **35**, 575-585 (1979).

Wu, M., Ware, J.H. 'On the use of repeated measurements in regression analysis with dichotomous responses', *Biometrics*, **35**, 513-521 (1979).

Wu, M.C. and Bailey, K. 'Analysing changes in the presence of informative right-censoring caused by death and withdrawal', *Statistics in Medicine*, **7**, 337-346 (1988).

Wu, M.C. and Bailey, K.R. 'Estimation and comparison of changes in the presence of informative right-censoring: conditional linear model', *Biometrics*, **45**, 939-955 (1989).

Wu, M.C. and Carroll, R.J. 'Estimation and comparison of changes in the presence of informative right-censoring by modeling the censoring process', *Biometrics*, **44**, 175-188 (1988).

Wu M.C. and Lan K.K.G. 'Sequential monitoring for comparison of change in a response variable in clinical studies', *Biometrics*, **48**, 765-779 (1992).

Yashin, A.I., Manton, K.G. and Stallard E. 'Evaluating the effects of observed and unobserved diffusion processes in survival analysis of longitudinal data', *Mathematical Modelling*, **7**, 1353-1363 (1986).

Zeger, S.L. and Liang, K.-Y. 'Longitudinal data analysis for discrete and continuous outcomes', *Biometrics*, **42**, 121-130 (1986).

Zhao, L.P. and Prentice, R.L. 'Correlated binary regression using a quadratic exponential model', *Biometrika*, **77**, 642-648 (1990).