Assessing Informative Drop-out in Models for Repeated Binary Data

by

Lee Shean Er

B.Sc, University of Guelph 1998

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

Master of Science

in

THE FACULTY OF GRADUATE STUDIES

(Department of Statistics)

we accept this thesis as conforming
to the required standard

The University of British Columbia

March 2001

Department of _STATISTICS_

The University of British Columbia
Vancouver, Canada

Date _MARCH 5, 2001_

Abstract

Drop-outs are a common problem in longitudinal studies. In terms of statistical models for the data, there are three types of drop-out mechanisms: drop-out occurring completely at random (CRD), drop-out occurring at random (RD) and informative drop-out (ID). The drop-out mechanism is classified as CRD if the drop-out mechanism is independent of the measurements; as RD if the drop-out mechanism depends only on the observed but not the unobserved measurements, and as ID if the drop-out mechanism depends on both the observed and unobserved measurements. CRD and RD are referred to as ignorable because the drop-out mechanism can be ignored for the purpose of making inferences about the observed measurements, while ID is non-ignorable. Analyses based on an assumption of ignorable drop-out, when in reality the drop-out mechanism is non-ignorable, can lead to misleading or biased results. Likelihood-based models for continuous and categorical longitudinal data subject to non-ignorable drop-out have been developed. In this thesis, we focus on exploring likelihood-based models for binary longitudinal data subject to informative drop-out.

The two modelling approaches considered are a selection model proposed by Baker (1995) and a transition model proposed by Liu et al. (1999). We apply these models to a data set from a multiple sclerosis (MS) clinical trial. The aims of the analyses are to investigate whether there is an indication of informative drop-out in this data, and to assess the sentivity of inferences concerning the treatment effects to the underlying drop-out mechanisms. We do not attempt to provide a definitive

analyses of the data set, but rather to explore a variety of models which incorporate informative drop-out.

Contents

List of Tables

List of Figures

Acknowledgements

First and foremost, I would like to thank my supervisor, Professor John Petkau, for his patience, constant support and invaluable guidance throughout the development of this manuscript. I could not have chosen a better person to work with. I am also grateful to Professor Harry Joe for agreeing to review my work on such short notice.

A huge thanks also to Professor Nancy Heckman for her constant encouragement, especially during my first semester in UBC. I would also like to express my gratitude to Ms. Christine Graham for her help. In addition, I would like to thank the entire UBC Statistics Department for making my stay enjoyable.

Most importantly, I am indebted to Ryan Woods, as without his love, encouragement and belief in me, I would not have completed my Master's degree.

<div align="right">LEE SHEAN ER</div>

The University of British Columbia
March 2001

To my beloved late mother, Goh Seok Im

Chapter 1

Introduction

1.1 Background of this Thesis

The defining characteristic of a longitudinal study is a sample design which specifies repeated observations on the same individual (or experimental unit). However, failure to obtain a full set of observations on a given individual (or unit), resulting in incomplete data and/or unbalanced designs, is a common problem in longitudinal studies. The form of missingness in longitudinal studies is typically *drop-outs*, in which sequences of measurements on some individuals terminate prematurely. This drop-out phenomenon is reflected in a data set collected over a 3-year period in a multicenter multiple sclerosis (MS) clinical trial sponsored by Berlex Laboratories of Richmond, California. The work presented in this thesis is motivated by this data set. A detailed description of the clinical trial and the data set to be analyzed can be found in Chapter 2.

Multiple sclerosis is a serious disease of the central nervous system (CNS), the nerves that comprise the brain and spinal cord. The term "multiple sclerosis" refers to multiple areas of patchy scarring, or plaques, that result from the destruction of myelin. Myelin is a white substance which forms a sheath around the spinal cord. When the myelin sheath is destroyed, signals transmitted throughout the CNS are

disrupted which leads to the occurrence of an acute attack, or exacerbation. During these exacerbations patients can suffer from a variety of symptons such as blurred vision, a sensation of numbness or loss of control of the movements in parts of the body. To date, the cause of MS is unknown and no cure exists. A number of treatments examined over the past decade have reduced rates of acute exacerbations and slowed progression in disability. In fact, the Berlex trial, where the treatment investigated was Interferon β-1b, was the first to demonstrate beneficial effects of a treatment for MS patients.

Patients withdrew from the Berlex trial due to reasons such as lack of efficacy, toxicities in excess of prespecified toxicity levels, or other side effects to the treatment. Nevertheless the overall drop-out rate did not exceed that anticipated at the trial's inception. The intent-to-treat analyses of the trial data were performed under the assumption that the drop-out occurred completely at random, as is customary in clinical trials. Methods have been developed for explicitly modelling non-response (not only restricted to drop-outs) under the more general assumption that the non-response may not have occurred completely at random. Our main objective is to investigate the sensitivity of the conclusions concerning the treatment effects to different assumptions about the nature of the drop-out mechanisms. Diggle and Kenward's (1994) classification of drop-out mechanisms, modified from Rubin (1976) and Little and Rubin (1987), is described in Chapter 3. Comparison of different models also allows us to study the nature of the drop-out mechanism in this data set.

1.2   Methods of Analyses

Likelihood-based methods are commonly used for incomplete data, including for the analysis of longitudinal data with drop-outs. According to the Diggle and Kenward (1994) terminology, drop-out mechanisms can be classified as completely random drop-out (CRD), random drop-out (RD) or informative drop-out (ID). For a

2

CRD mechanism, drop-out is independent of the outcome (or measurement) process; for a RD mechanism, drop-out is independent of the *unobserved* outcomes, but depends on the *observed* outcomes. For an ID mechanism, drop-out depends on both the *observed* and *unobserved* outcomes. Likelihood-based methods yield valid results in the presence of CRD or RD provided the model used for the measurement process is valid, and the observed information matrix is used rather than the expected information matrix. If, however, the drop-out mechanism is ID, modelling the drop-out process is necessary to permit valid inferences; see Laird (1988).

Modelling different drop-out mechanisms can provide insight into the nature of the withdrawal process. It can also be used to investigate the sensitivity of inferences to the underlying assumptions. In the past decades, researchers have proposed a number of methods for quantitative longitudinal data (usually normally distributed) and categorical longitudinal data subject to non-random drop-out. Laird (1988) provided an excellent discussion of how the drop-out process can affect the inferences about both continuous and categorical measurement processes.

For continuous longitudinal data, Wu and Carroll (1988) considered ID in a random effects model, with the data for each experimental unit following a linear time trend whose intercept and slope vary between individuals according to a bivariate Gaussian distribution. Their likelihood-based method permits the comparison of non-ID and ID drop-out mechanisms. Schluchter (1992) outlined a new approach based on a log-normal survival model when the primary outcome is the rate of change in a continuous variable subject to informative censoring. More recently, Diggle and Kenward (1994) proposed a general model-based approach for analyzing continuous longitudinal data that combines a multivariate linear model for the response with a logistic regression model for the drop-out process. This is the first paper to develop a modelling strategy that explicitly accommodates CRD and RD as special cases within an ID model.

The issue of how to deal with ID in categorical longitudinal data is not yet

resolved. Further, potential technical difficulties may arise in the likelihood-based methods for correlated categorical data due to the discreteness of the responses. Baker and Laird (1988) developed a log-linear model for categorical response subject to non-ignorable non-response in a sample survey setting and drew attention to the existence of boundary solutions. A number of authors have focused their attention on the multivariate binary data case to better understand some of the potential difficulties for correlated categorical data. Both Baker (1995) and Fitzmaurice *et al.* (1996) used a multivariate binary model where the marginal probabilities for the responses are specified as logistic regressions. However, these authors modelled the associations among the responses differently. These models for the outcomes were combined with logistic models for ignorable and non-ignorable drop-out mechanisms to analyze multivariate binary data. Both papers also highlighted the issue of identifiability of these models. Baker (1995) provided outlines of the proof of model identifiability for certain models he considered. Fitzmaurice *et al.* (1996) gave some suggestions on how to examine the identifiability of non-ignorable drop-out models. More recently, Ten Have *et al.* (1998) presented mixed effects models for longitudinal binary responses with informative drop-out analogous to the Wu and Carroll (1988) models for longitudinal continuous data. Liu *et al.* (1999) adapted the method proposed by Diggle and Kenward (1994) for the analysis of a binary longitudinal outcome.

Most of the likelihood-based models mentioned are formulated within the selection modelling framework (Little and Rubin, 1987). A selection model factors the joint distribution of the measurement and response processes into the marginal measurement distribution and the response distribution, conditional on the measurements. Molenberghs *et al.* (1999) discuss the strengths and limitations of selection models for non-random missingness in the categorical data setting. There are other ways to specify the joint distribution. For categorical responses, a log-linear approach incorporates the measurement and response processes into a sin-

gle log-linear model. A time-ordered approach factors the joint probability into a product of conditional probabilities ordered in time and a pattern-mixture model (Little, 1993) factors the joint distribution into the marginal response distribution and the measurement distribution, given the response distribution. The latter two approaches can be applied to both continuous and categorical repeated measurements data. Ekholm (1998) re-analyzed the children's obesity data set considered by Baker (1995) using a pattern-mixture model. Michiels *et al.* (1999) studied similarities and differences of modelling incomplete data within the selection and pattern-mixture settings assuming a missing at random mechanism.

Pseudo-likelihood and non-parametric approaches have also been proposed for carrying out analyses under different types of drop-out mechanisms. Using Dale's model (1986) for ordinal categorical longitudinal data, Kenward *et al.* (1994) demonstrated that, in the presence of RD, the generalized estimating equations (GEEs) approach proposed by Liang and Zeger (1986) may give misleading results. Robins *et al.* (1995) showed that appropriately weighted GEEs overcome this problem, but not in the presence of ID. More recently, Sun and Song (2000) proposed a non-parametric approach for analyzing the data from a clinical trial of adult schizophrenics with informative censoring.

1.3   Outline of this Thesis

Our attention will be on multivariate binary data. This special form of the data allows us to focus on the aforementioned issues that arise mainly in correlated categorical data. We are also interested in studying the nature of the drop-out process in our data. For this purpose, we choose to work with models within the selection modelling framework.

The remainder of this thesis is outlined as follows: Chapter 2 presents the description of the Berlex trial and the binary responses which will comprise the data set to be analyzed. Chapters 4 and 5 discuss Baker's selection model and the Liu

transition model respectively. These models can be used to examine various types of drop-out mechanisms in our data. The definitions of the drop-out mechanisms is provided in Chapter 3. Non-ignorable (or informative) drop-out (or non-response) models are generally harder to implement due to potential analytical problems such as model identifiability issues. Chapter 6 focuses on the issue of identifiability in models for incomplete binary responses. Proofs of identifiability of some of our models are also included in the chapter. The detailed results of our analyses are reported in Chapter 7. We conclude the thesis with some general discussion in Chapter 8; this includes comments on the two models and suggestions of other possible methods for analyzing the data.

Chapter 2

Data Description

2.1 Description of the Berlex Clinical Trial

The Berlex clinical trial was a phase III trial of the effect of Interferon β-1b on relapsing-remitting multiple sclerosis (MS) patients. The primary outcome measure was the rate of exacerbations. This was a multicenter, randomized, double-blind, placebo-controlled trial with three parallel treatment groups. The study was originally planned with a 2-year treatment period; the trial was later extended to 3 years (because by the end of the second year, some patients had been on the study for almost three years due to different starting dates). The study was carried out in a double-blind fashion for the full three years. The data from the first 2 years of the study established that the Interferon β-1b treatment groups had decreased exacerbation rates and increased proportions of patients remaining exacerbation-free. These beneficial results were also found in the 3 year data. This was the first trial to unequivocally identify an effective treatment for relapsing-remitting MS. Interferon β-1b has emerged as a therapeutic option in MS and has been hailed as a major advance in the management of this disorder.

This trial consisted of 372 patients from 11 centers in the United States and Canada on three parallel treatment groups: placebo (PL), low dose (LD) and high

dose (HD). The dosage for LD and HD were 1.6 and 8.0 million international units (MIU) respectively. All patients were between the ages of 18 and 50 years, had been diagnosed with MS at least 1 year prior to entry to the study, had Krutzke Expanded Disability Status Scale (EDSS) scores of 5.5 or less, and had experienced at least 2 exacerbations in the previous 2 years. Moreover, all had been clinically stable for at least 30 days prior to entry and had received no medications to speed up the recovery from relapse such as ACTH (adrenocorticotrophic hormone) or prednisone during this period.

Patients were randomized to the three treatment groups within each center and divided almost evenly within each center. All patients were blinded to the treatment assignments. Of these 372 patients, 123 received PL, 125 received LD and 124 received HD of Interferon β-1b by injection every other day. Two neurologists were appointed at each center: one who performed the periodic examinations was not aware of the drug side effects, and another who knew about the side effects and injection reactions was responsible for reviewing laboratory findings for toxicity and for overall patient care. Patients were scheduled to be evaluated every 12 weeks except for the first few months of the study, where evaluations were more frequent. In addition, visits were made when symptoms occurred suggesting the possibility of an MS exacerbation. A Scripps Neurological Rating Scale (NRS) score and a Kurtzke EDSS score were determined in each evaluation.

For all patients in the study, the beginning and end dates of all exacerbations as well as the EDSS scores obtained at each visit were recorded. Besides these clinical outcomes, each patient also had a baseline cranial magnetic resonance imaging (MRI) and this was repeated annually. The patients at one of the centers (the University of British Columbia) had cranial MRIs repeated at 6-week intervals for the first 2 years.

2.1.1 Drop-out Rate in this Clinical Trial

Since some beneficial results of Interferon β-1b were found after 3 years of study, patients who remained in the study were offered the high dose treatment for another 2 years. Thus, the entire study continued for over 5 years, but many patients dropped out during this period. Figure 2.1 shows a roughly constant rate of drop-out during the first three and a half years, except for a large number of drop-outs at the end of the second year (the original intended end of the study). The plot also indicates the drop-out rate increased dramatically after the end of the 3-year treatment period. Because of the potential difficulty in interpreting the 5 year data (e.g. how should patients who switched from one treatment to another be treated in the analysis and how should the results obtained be interpreted), we employ the 3-year treatment period data to perform a variety of analyses in this thesis.

We can represent the information shown in Figure 2.1 in another fashion. Figure 2.2 displays Kaplan-Meier survival curves describing the proportion of patients remaining on study by treatment group; the dash, solid and dotted vertical lines indicate the end of the 1-year, 2-year and 3-year periods respectively. The most drop-outs over the 3-year period occurred in the low dose group (approximately 40%). Roughly 20% of patients withdrew from the trial during the first 2 years in all three groups. A number of patients in each treatment arm dropped out around the end of the 2-year period, but the proportion remaining for most of the third year of the study is roughly 70% in both the PL and HD groups and roughly 60% in the LD group. Table 2.1 summarizes the numbers of patients who dropped-out by the end of the first, second and third year of the study.

More details concerning the clinical trial can be found in the published reports of the IFNB Multiple Sclerosis Study Group [27, 35, 36].

9

Figure 2.1: Histogram for Length on Study (3-month bins)
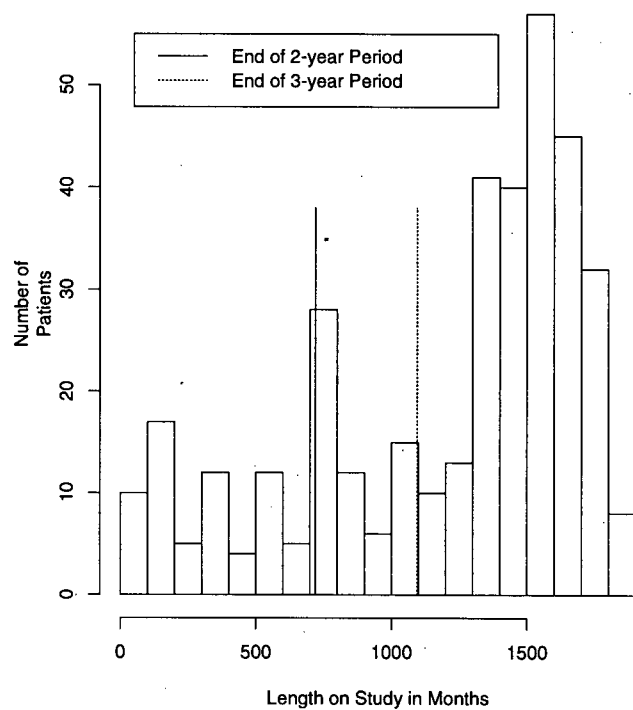
Table 2.1: Cumulative Number of Drop-outs After 1, 2 and 3 Years on Study

	After 1-Year		After 2-Year		After 3-Year	
Group	Number	Proportion	Number	Proportion	Number	Proportion
PL	13	11%	27	22%	41	33%
LD	11	9%	30	24%	49	39%
HD	17	14%	29	23%	35	28%

Figure 2.2: Kaplan-Meier Survival Curves for Time on Study: Over 3-year Treatment Period
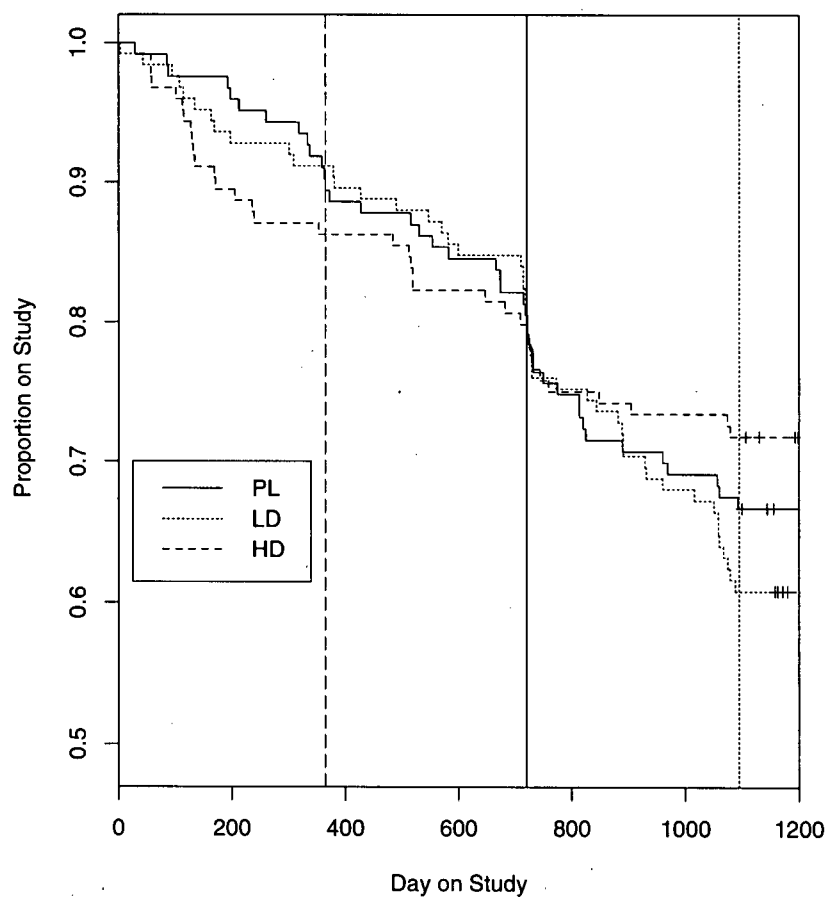
2.2   Description of the Data

The main objective of this thesis is to explore models for longitudinal binary responses incorporating different types of drop-out mechanisms in the context of the Berlex trial. We consider the exacerbation variable as the response variable of interest in our analysis. We choose to represent these data in binary form on an annual basis (whether exacerbations occurred in each 1-year interval) to allow a specific focus on models for binary responses. In other words, the data for each patient will be represented by three binary responses indicating whether they experienced any exacerbations during the 1-year intervals. One of the reasons for proceeding in this way, as opposed to refining the time intervals to 6-month intervals say, is to reduce the number of possible different derived sequences of the binary responses as well as the number of drop-out patterns. This allows a focus on the key ideas for modelling such data. This will become clearer in later chapters.

The rest of this section is structured as follows. In the next subsection, we describe the annual drop-out patterns. We then discuss how the binary responses are derived. We conclude the section with a brief description of the baseline covariates to be included in our analyses.

2.2.1   Drop-out Patterns

The data described in the previous section involve a total of 372 patients. Each patient's termination date from the study was recorded. To derive our annual data set, the data on patients who dropped out are handled as follows:

- **Scenario 1**

    If the patient's termination date was prior to 365 days on study, then we will treat these patients as if they dropped out at the beginning of the study. In other words, these patients have no outcomes in our annual data set.

Table 2.2: Number of Patients in the 4 Drop-out Cases (x=present,o=absent)

0	Year 1	Year 2	Year 3	Number of Patients
1	o	o	o	41
2	x	o	o	45
3	x	x	o	39
4	x	x	x	247

- **Scenario 2**

  If the patient's termination date was after 365 days but prior to 730 days, then we will treat these patients as if they dropped out at the end of the first year of the study. These patients have one outcome in our annual data set.

- **Scenario 3**

  If the patient's termination date was after 730 days and prior to 1095 days, then we will treat these patients as if they dropped out at the end of the second year of the study. That is, these patients are missing only the third year outcome in our annual data set.

- **Scenario 4**

  If the patient's termination date exceeded 1095 days, then we will treat these patients as if they completed the 3-year study and thus all three annual outcomes were observed.

Table 2.2 summarizes the total number of patients according to the four scenarios of available annual outcomes over the 3 year period ("x" denotes present and "o" denotes absent).

Table 2.3 displays the breakdown of the 372 patients in our annual data set by treatment groups and gender according to the total number of patients entering the study, and dropping-out at the beginning, the end of the first year and the end of the second year of the study. Patients were quite evenly distributed across the 3 treatment arms at the beginning of the study, as were the patients who dropped

13

Table 2.3: Summary of the Our Annual Data

	By Treatment Groups			By Gender	
Number of Patients	PL	LD	HD	Males	Females
Entering the Study	123	125	124	113	259
Drop-out At Beginning	13	11	17	16	25
Drop-out At End of Year 1	14	19	12	9	36
Drop-out At End of Year 2	14	19	6	8	31

out at the beginning of the study and at the end of year 1. However, fewer patients in the HD group dropped-out at the end of year 2 than in the PL and LD groups. In summary, the LD group has the highest drop-out rate, followed by the PL group, and both rates increase slightly over time. As expected, the drop-out rate in the HD group is the lowest and it decreases over time. Table 2.3 also shows the drop-out rates for females and males are fairly consistent over time, although the drop-out rates for females are a bit higher than for males.

In the next two sections, we provide a more detailed description of the binary outcome variable and baseline covariates of interest. All the corresponding descriptive statistics presented are based on our annual data.

2.2.2 Binary Outcome Variables

As mentioned earlier in the chapter, the start and end dates of any exacerbations patients experienced during the study were recorded. For our purposes, we do not use the end dates even though they could contain valuable information. All exacerbations are attributed to the annual period in which they began. Recall we divided the time period of the study into three 1-year intervals. Since these intervals are quite wide, some patients experienced multiple exacerbations within these intervals. The number of exacerbations experienced by patients within these annual intervals ranges from 0 to 6; the frequency of these counts by yearly interval is summarized in Table 2.4. Most patients experienced either no exacerbations or a small number of

Table 2.4: Frequency Table of the Exacerbation Counts for Patients with At Least One Outcome

Interval	Group	Number of Exacerbations							Number of Patients
		0	1	2	3	4	5	6	
Year 1	ALL	121	103	63	27	13	2	2	331
	PL	32	36	20	14	8	0	0	110
	LD	39	35	26	7	4	1	2	114
	HD	50	32	17	6	1	1	0	107
Year 2	ALL	122	86	53	15	4	5	1	286
	PL	39	28	16	8	2	3	0	96
	LD	39	28	22	2	2	1	1	95
	HD	44	30	15	5	0	1	0	95
Year 3	ALL	122	74	38	12	0	0	1	247
	PL	37	23	18	4	0	0	0	82
	LD	36	26	8	5	0	0	1	76
	HD	49	25	12	3	0	0	0	89

exacerbations; only a few patients had 4 or more exacerbations within a year. Based on this information and for simplicity of analysis, it seems reasonable to dichotomize these data as no exacerbation or at least 1 exacerbation experienced. Clearly there is some loss of information associated with dichotomizing these data. One way to retain the information is to treat the counts of the total number of exacerbations as if they are Poisson random variables and perform analyses based on the counts. However, we will not explore such analyses in this thesis.

Figure 2.3 shows the proportion of patients experiencing exacerbations over time by treatment group based on these dichotomized annual data. In general, the proportion of patients experiencing exacerbations decreased over the 1-year periods in all groups. Further, the HD group has the lowest proportions among the 3 treatment arms throughout the study. The proportion of patients experiencing exacerbations is slightly higher in the PL group than in the LD group. This plot also suggests a dose-response relationship in these data.

Figure 2.3: Proportion of Patients Experiencing Exacerbations Over Time by Treatment Group Based on Dichotomous Annual Data
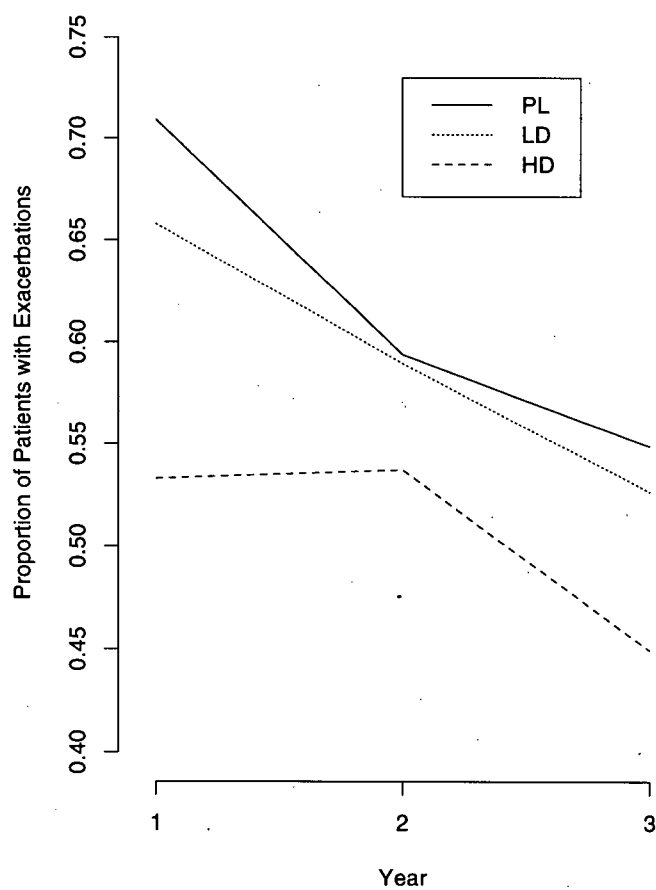
2.2.3 Baseline Covariates

We are primarily interested in the assessment of the treatment effects on the binary outcome variables described in the previous section, but patterns in the data over time and the effects of several baseline covariates are also of interest. The baseline covariates we considered are:

- gender,

- age,

- duration of MS,

- Kurtzke Expanded Disability Status Score (EDSS), and

- burden of disease (BOD).

In general, more females than males suffer from MS. This phenomenon is reflected in this trial; as shown in Table 2.5, the female-to-male ratios are roughly 2.5, 2.1, and 2.3 in the PL, LD and HD groups respectively. Figure 2.4 shows the boxplots of age, duration of MS and EDSS at baseline by treatment group. The ages range between 18 and 50 years. The median age at baseline in the HD group is slightly smaller than in the other groups, but the distribution of the ages is quite similar for the 3 treatment groups. The boxplots also indicate that about 50% of the patients had ages between 30 and 40 years in each treatment group. The duration of MS ranges between 1 and 31 years and the median is slightly higher in the HD group. The boxplots for baseline EDSS indicate a fairly balanced distribution across the three groups, with scores ranging from 0 to 5.5 in each group.

There are two distinct forms of magnetic resonance imaging (MRI) scans of interest in MS studies: T1-weighted scans and T2-weighted scans. A T1-weighted scan uses a small injection of the chemical gadolinium into the patient's bloodstream. The presence of gadolinium will enhance the appearance of active lesions (areas of inflammation on the blood/brain barrier) on the brain stem, and facilitate their

17

Table 2.5: Number of Female and Male Patients in Each Treatment Group

	Treatment Group		
	PL	LD	HD
Female	88	85	86
Male	35	40	38
Total	123	125	124

detection. A T2-weighted scan provides clearer definition of the actual size and shape of each lesion without any gadolinium injection into the bloodstream, which usually blurs the border of the lesions. The MRI measure of interest in this thesis, known as burden of disease (BOD), is a measure of the total volume of all lesions on the T2-weighted scan. In our data set, there are 8 patients who did not have a BOD measurement at baseline; 3 from the PL group, 4 from the LD group, and 1 from the HD group. Excluding these 8 patients, the histogram of BOD at baseline and the boxplots of BOD at baseline by treatment group are shown in Figure 2.5. The distribution of BOD is highly skewed to the right. There are only 2 patients who did not have any lesions at baseline (BOD = 0), but there are 5 patients with BOD greater than 10,000 (mm^2): 2 belong to the PL group and 3 belong to the LD group. This is also reflected in the boxplots in Figure 2.5. Excluding the 3 patients in the LD group who had the largest BOD readings, the general distribution of the BOD measurements is quite similar in each treatment arm.

2.3   Questions of Interest

Having introduced the annual data set to be analyzed, we now describe the study questions we plan to address in this thesis.

Recall that the main focus of this thesis is to explore models for analyzing repeated binary data incorporating different drop-out mechanisms. Although the drop-out rate in our annual data is moderate, we would like to investigate the

Figure 2.4: Boxplots of Age, Duration of MS and EDSS at Baseline by Treatment Group
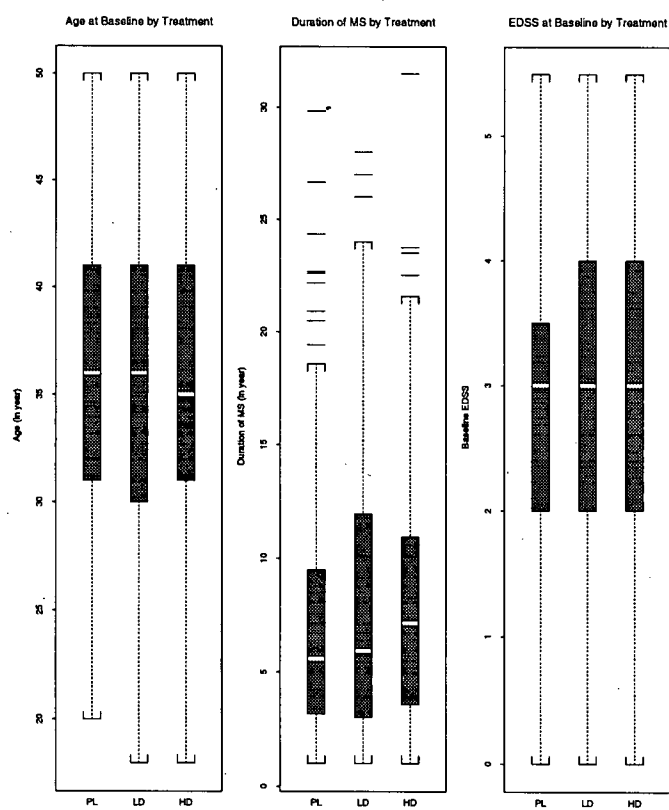
Figure 2.5: Histogram of BOD and Boxplots of BOD by Treatment Group ($n = 364$)
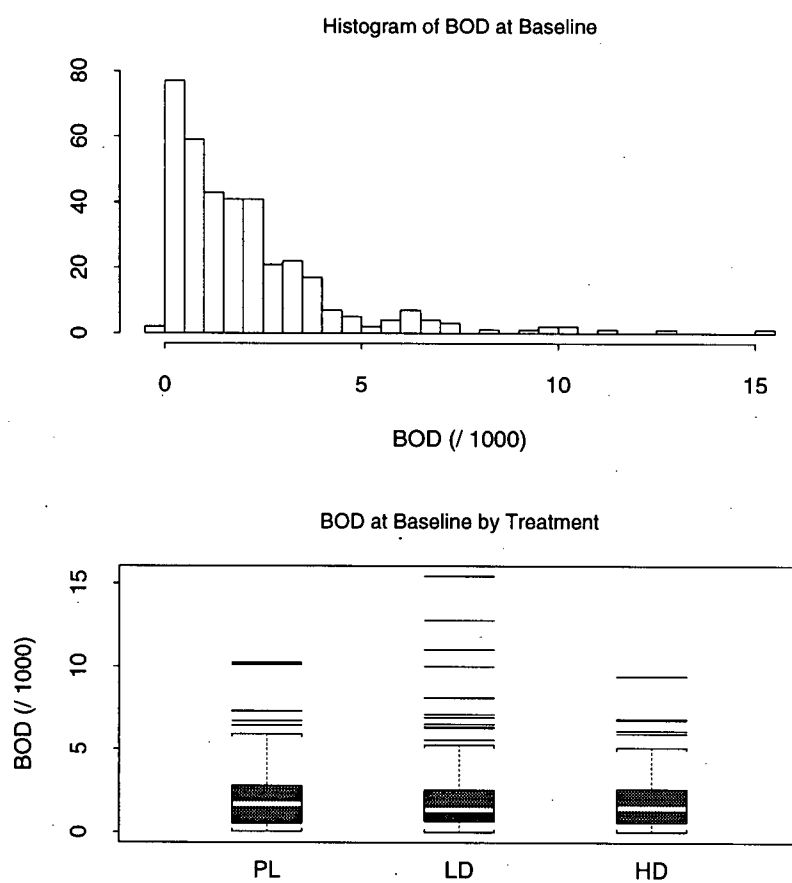
most appropriate form of model for the drop-out process; in particular, to explore whether there is an indication of informative drop-out. It is also of interest to assess the sensitivity of inferences concerning the treatment effects (primarily) to the form of the models for the drop-out mechanism, and to explore the importance of baseline covariates for our annual data.

Chapter 3 provides a discussion of different drop-out mechanisms. We describe general methodology for analyzing incomplete binary data in Chapters 4 and 5. Chapter 6 sheds some light on potential identifiability problems in such models. Chapter 7 contains all the results from the analyses we performed and we conclude the thesis with some discussion.

Chapter 3

Classification of Missing Values in Longitudinal Data

3.1 Introduction

Longitudinal studies are usually characterized by collecting a set of measurements on an individual unit at prespecified points in time; in many cases (typically in clinical trials), the set of prespecified points in time are the same for all units. Missing values arise whenever one or more of the *intended* measurements from units within the study are incomplete. Such missing data are a common problem in longitudinal studies, particularly when the experimental units are human subjects and collecting data involves a visit to a hospital or clinic, or the time between intended measurements is lengthy.

It is important to distinguish between *unbalanced* data and missing values. Unbalanced data result when the set of times of intended measurements is not common to all units; for example, if one chose in advance to take measurements every half hour on one-half of the subjects and every hour on the other half. Such unbalanced data could also be described as incomplete but there are no missing values from the viewpoint of the design of data collection. Missing data also arise in

unbalanced data; however, there are deeper conceptual issues as to *why* the values are missing, and more specifically whether the missingness is related to the questions posed by the study.

Little and Rubin (1987) have provided a useful classification of missing value mechanisms. Let $\mathbf{Y}^*$ denote the complete set of measurements for one unit which would have been obtained if there were no missing values. Partition this set into $\mathbf{Y}^* = (\mathbf{Y}^{(o)}, \mathbf{Y}^{(m)})$ with $\mathbf{Y}^{(o)}$ denoting the measurements actually obtained and $\mathbf{Y}^{(m)}$ the measurements which would have been available if they had not been missing, for whatever reason or cause. Let $\mathbf{R}$ denote a set of indicator random variables, denoting which elements of $\mathbf{Y}^*$ fall into $\mathbf{Y}^{(o)}$ and which into $\mathbf{Y}^{(m)}$. We can then specify a probability model for the missing value mechanism as the probability distribution of $\mathbf{R}$ conditional on $\mathbf{Y}^* = (\mathbf{Y}^{(o)}, \mathbf{Y}^{(m)})$. In the terminology used by Little and Rubin, the missing value mechanism is classified as:

1. *completely random* if $\mathbf{R}$ is independent of both $\mathbf{Y}^{(o)}$ and $\mathbf{Y}^{(m)}$;

2. *random* if $\mathbf{R}$ is independent of $\mathbf{Y}^{(m)}$;

3. *informative* if $\mathbf{R}$ is dependent on $\mathbf{Y}^{(m)}$.

We will abuse the notation f to denote a probability density (or mass) function throughout this thesis; the function being referred to will be clear from the context. For likelihood-based inference, the important distinction is between *random* and *informative* missing values. To see this, $f(\mathbf{y}^{(o)}, \mathbf{y}^{(m)}, \mathbf{r})$, the joint probability density function (pdf) of $(\mathbf{Y}^{(o)}, \mathbf{Y}^{(m)}, \mathbf{R})$, can be factored as

$$f(\mathbf{y}^{(o)}, \mathbf{y}^{(m)}, \mathbf{r}) \;=\; f(\mathbf{y}^{(o)}, \mathbf{y}^{(m)}) f(\mathbf{r} \mid \mathbf{y}^{(o)}, \mathbf{y}^{(m)}). \tag{3.1}$$

For a likelihood-based analysis, we need the joint pdf of the observed random variables, $(\mathbf{Y}^{(o)}, \mathbf{R})$, which can be obtained by integrating (3.1) over all possible values for the unobserved random variables

$$f(\mathbf{y}^{(o)}, \mathbf{r}) \;=\; \int f(\mathbf{y}^{(o)}, \mathbf{y}^{(m)}) f(\mathbf{r} \mid \mathbf{y}^{(o)}, \mathbf{y}^{(m)}) d\mathbf{y}^{(m)}. \tag{3.2}$$

If the missing value mechanism is *random*, $f(\mathbf{r} \mid \mathbf{y}^{(o)}, \mathbf{y}^{(m)})$ is independent of $\mathbf{y}^{(m)}$ and (3.2) becomes

$$
\begin{aligned}
f(\mathbf{y}^{(o)}, \mathbf{r}) &= f(\mathbf{r} \mid \mathbf{y}^{(o)}) \int f(\mathbf{y}^{(o)}, \mathbf{y}^{(m)}) d\mathbf{y}^{(m)} \\
&= f(\mathbf{r} \mid \mathbf{y}^{(o)}) f(\mathbf{y}^{(o)}).
\end{aligned}
\tag{3.3}
$$

Taking logarithms in (3.3), the log-likelihood function is

$$
L = \log f(\mathbf{r} \mid \mathbf{y}^{(o)}) + \log f(\mathbf{y}^{(o)}),
\tag{3.4}
$$

which can be maximized by separate maximization of the two terms on the right-hand side provided the parameters appearing in $f(\mathbf{r} \mid \mathbf{y}^{(o)})$ and in $f(\mathbf{y}^{(o)})$ are disjoint. Since the first term contains no information about the distribution of $\mathbf{Y}^{(o)}$, we can ignore it for the purpose of making inferences about $\mathbf{Y}^{(o)}$. Because of the above result, both completely random and random missing value mechanisms are sometimes referred to as *ignorable*. On the other hand, informative missing value mechanisms are referred to as *non-ignorable* because such a missing value mechanism cannot be ignored when making inferences about $\mathbf{Y}^{(o)}$.

3.2   Types of Drop-outs

We have distinguished between unbalanced data and missing values. Now let us focus on different types of missing values. Missing values can occur either *intermittently* or as *drop-outs*. Suppose we intend to obtain a sequence of n measurements, say $Y_1, Y_2, \ldots, Y_n$, on a particular unit. We say that missing values occur as *drop-outs* if whenever measurement Y_j is missing, so are the measurements, Y_k for all $k > j$; otherwise the missing values are *intermittent*.

In this thesis, we are particularly interested in studying drop-out mechanisms. Drop-outs are a common phenomenon in longitudinal studies. They typically arise not as a result of censoring applied to the measurements on the experimental unit, but because some units prematurely terminate their participation in the study.

24

A unit's withdrawal may be for reasons directly or indirectly connected to the measurement process. Thus, it is of interest to investigate whether the drop-out process is related to the measurement process. Following the Little and Rubin (1987) discussion of the classification of missing value mechanisms, Diggle and Kenward (1994) modified the above definitions slightly to describe drop-out processes as:

(a) *Completely Random Drop-out* (CRD): if the drop-out mechanism is independent of the measurement process;

(b) *Random Drop-out* (RD): if the drop-out mechanism is independent of the **unobserved** measurements, but depends on the **observed** measurements;

(c) *Informative Drop-out* (ID): if the drop-out mechanism depends on both the **observed** and **unobserved** measurements.

Both CRD and RD are referred to as *ignorable* drop-outs, while ID is referred to as *non-ignorable* drop-out.

In next two chapters, we give an overview of the selection modelling approach for longitudinal binary data subject to non-ignorable non-response. The basic idea of a selection model is to factor the joint distribution of the measurement variables and the non-response indicator variables, $f(\mathbf{Y}^*, \mathbf{R})$, into $f(\mathbf{R} \mid \mathbf{Y}^*)f(\mathbf{Y}^*)$, where $f(\mathbf{Y}^*)$ is known as the outcome model and $f(\mathbf{R} \mid \mathbf{Y}^*)$ is known as the drop-out model. The only distinction between the next two chapters is in the model for the outcome (or measurement) process.

Chapter 4

Selection Model

4.1 Baker's Selection Model for Binary Longitudinal Data with Informative Non-response

Diggle and Kenward (1994) provided a general methodology for dealing with continuous responses subject to informative, or non-ignorable, drop-outs in a longitudinal study. Baker (1995) provided a discussion of a related model that accounts for non-ignorable non-response. The methodology is connected to that presented by Diggle and Kenward, however Baker's model is for repeated binary data and the non-response is allowed to occur in various patterns, not only as drop-outs.

For simplicity, we limit our discussion to repeated binary data collected at 3 time points, as this coincides with the structure of our data set. Our model is a simplified version of Baker's model as we are only interested in monotonic non-response patterns, i.e. drop-outs.

We first introduce the concepts of incomplete (observed) and complete data. Let t index the time points where measurements are intended to be taken. In this particular context, t represents the three successive 1-year period measurements that are to be taken, coded as $t = 1, 2, 3$. Let $\mathbf{X_t}$ denote a vector of covariates at time t, and denote $\mathbf{X} = (\mathbf{X'_1}, \mathbf{X'_2}, \mathbf{X'_3})$. The vector of random variables for the complete

(possibly unobserved) data is

$$(Y_1^*, Y_2^*, Y_3^*, R_1, R_2, R_3, \mathbf{X}),$$

where Y_t^* is the binary outcome variable at time t which takes on values 0 or 1, and R_t is an indicator variable of non-response at time t with sample space $\{a,p\}$ where "a" denotes absent and "p" denotes present. The vector of random variables for the incomplete (observed) data is

$$(Y_1, Y_2, Y_3, \mathbf{X}),$$

where Y_t has sample space $\{0, 1, a\}$. The complete and incomplete random variables are related as follows:

$$Y_t = \begin{cases} Y_t^* & \text{if } R_t = \text{p}, \\ a & \text{if } R_t = \text{a}. \end{cases}$$

There are several approaches for modelling the joint distribution of the complete data, $\Pr(Y_1^* = y_1^*, Y_2^* = y_2^*, Y_3^* = y_3^*, R_1 = r_1, R_2 = r_2, R_3 = r_3 \mid \mathbf{X})$. Baker chose to use a **selection model** in which the joint distribution is factored into the probability of the outcomes multiplied by the probability of the non-response indicators, given the outcomes; that is,

$$\Pr(R_1 = r_1, R_2 = r_2, R_3 = r_3 \mid Y_1^* = y_1^*, Y_2^* = y_2^*, Y_3^* = y_3^*, \mathbf{X})$$
$$\times \ \Pr(Y_1^* = y_1^*, Y_2^* = y_2^*, Y_3^* = y_3^* \mid \mathbf{X}).$$

Now, denote the outcome model as

$$\Pr(Y_1^* = y_1^*, Y_2^* = y_2^*, Y_3^* = y_3^* \mid \mathbf{X}) \ = \ f^*(y_1^*, y_2^*, y_3^* \mid x; \boldsymbol{\theta}), \qquad (4.1)$$

where $\boldsymbol{\theta}$ is a vector of parameters. Also, denote the non-response model as

$$\Pr(R_1 = r_1, R_2 = r_2, R_3 = r_3 \mid Y_1^* = y_1^*, Y_2^* = y_2^*, Y_3^* = y_3^*, \mathbf{X})$$
$$= q(r_1, r_2, r_3 \mid y_1^*, y_2^*, y_3^*, x; \boldsymbol{\eta}), \qquad (4.2)$$

Table 4.1: All Possible Patterns for Incomplete Data for the Case where 3 Observations were Intended for Every Unit: x = observed, o = missing.

Pattern	y_1	y_2	y_3
1	x	x	x
2	x	x	o
3	x	o	x
4	o	x	x
5	x	o	o
6	o	x	o
7	o	o	x
8	o	o	o

where $\boldsymbol{\eta}$ is a vector of parameters. This construction assumes that the parameters of the outcome and non-response models are distinct (Diggle and Kenward, 1994). This relates back to the idea of ignorable and non-ignorable drop-out mechanisms discussed in Chapter 3 (see p.18). Under RD and CRD, inferences based on the observed data are valid even though the drop-out mechanism is ignored, but this is not true for an informative drop-out mechanism.

Table 4.1 displays all the possible realizations of the incomplete data in this particular scenario, where "x" denotes the measurement is observed and "o" denotes the measurement is unobserved. Using the outcome and non-response models, we can write down the probability of these 8 realizations of incomplete data as follows:

$$f(y_1^*, y_2^*, y_3^* \mid \boldsymbol{x}; \boldsymbol{\theta}, \boldsymbol{\eta}) = f^*(y_1^*, y_2^*, y_3^* \mid \boldsymbol{x}; \boldsymbol{\theta}) q(\mathrm{p}, \mathrm{p}, \mathrm{p} \mid y_1^*, y_2^*, y_3^*, \boldsymbol{x}; \boldsymbol{\eta}), \quad (4.3)$$

$$f(y_1^*, y_2^*, a \mid \boldsymbol{x}; \boldsymbol{\theta}, \boldsymbol{\eta}) = \sum_{y_3^*=0}^{1} \left[f^*(y_1^*, y_2^*, y_3^* \mid \boldsymbol{x}; \boldsymbol{\theta}) \times q(\mathrm{p}, \mathrm{p}, a \mid y_1^*, y_2^*, y_3^*, \boldsymbol{x}; \boldsymbol{\eta}) \right],$$

$$(4.4)$$

$$f(y_1^*, a, y_3^* \mid \boldsymbol{x}; \boldsymbol{\theta}, \boldsymbol{\eta}) = \sum_{y_2^*=0}^{1} \left[f^*(y_1^*, y_2^*, y_3^* \mid \boldsymbol{x}; \boldsymbol{\theta}) \times q(\mathrm{p}, a, \mathrm{p} \mid y_1^*, y_2^*, y_3^*, \boldsymbol{x}; \boldsymbol{\eta}) \right],$$

$$(4.5)$$

$$f(a, y_2^*, y_3^* \mid \boldsymbol{x}; \boldsymbol{\theta}, \boldsymbol{\eta}) = \sum_{y_1^*=0}^{1} \left[f^*(y_1^*, y_2^*, y_3^* \mid \boldsymbol{x}; \boldsymbol{\theta}) \times q(a, p, p \mid y_1^*, y_2^*, y_3^*, \boldsymbol{x}; \boldsymbol{\eta}) \right],$$

(4.6)

$$f(y_1^*, a, a \mid \boldsymbol{x}; \boldsymbol{\theta}, \boldsymbol{\eta}) = \sum_{y_2^*=0}^{1} \sum_{y_3^*=0}^{1} \left[f^*(y_1^*, y_2^*, y_3^* \mid \boldsymbol{x}; \boldsymbol{\theta}) \times q(p, a, a \mid y_1^*, y_2^*, y_3^*, \boldsymbol{x}; \boldsymbol{\eta}) \right],$$

(4.7)

$$f(a, y_2^*, a \mid \boldsymbol{x}; \boldsymbol{\theta}, \boldsymbol{\eta}) = \sum_{y_1^*=0}^{1} \sum_{y_3^*=0}^{1} \left[f^*(y_1^*, y_2^*, y_3^* \mid \boldsymbol{x}; \boldsymbol{\theta}) \times q(a, p, a \mid y_1^*, y_2^*, y_3^*, \boldsymbol{x}; \boldsymbol{\eta}) \right],$$

(4.8)

$$f(a, a, y_3^* \mid \boldsymbol{x}; \boldsymbol{\theta}, \boldsymbol{\eta}) = \sum_{y_1^*=0}^{1} \sum_{y_2^*=0}^{1} \left[f^*(y_1^*, y_2^*, y_3^* \mid \boldsymbol{x}; \boldsymbol{\theta}) \times q(a, a, p \mid y_1^*, y_2^*, y_3^*, \boldsymbol{x}; \boldsymbol{\eta}) \right],$$

(4.9)

$$f(a, a, a \mid \boldsymbol{x}; \boldsymbol{\theta}, \boldsymbol{\eta}) = \sum_{y_1^*=0}^{1} \sum_{y_2^*=0}^{1} \sum_{y_3^*=0}^{1} \left[f^*(y_1^*, y_2^*, y_3^* \mid \boldsymbol{x}; \boldsymbol{\theta}) \times q(a, a, a \mid y_1^*, y_2^*, y_3^*, \boldsymbol{x}; \boldsymbol{\eta}) \right].$$

(4.10)

Baker specified the outcome model, $f^*(y_1^*, y_2^*, y_3^* \mid \boldsymbol{x}; \boldsymbol{\theta})$, in terms of a marginal model which models the marginal probabilities as functions of covariates, and an association model which models the temporal associations, using the idea suggested by Ekholm (1991, 1992). The non-response model, $q(r_1, r_2, r_3 \mid y_1^*, y_2^*, y_3^*, \boldsymbol{x}; \boldsymbol{\eta})$, was modelled by employing a general time-order causal model. With the assumption that drop-out does not depend on future events, the time-ordered causal model for three time points has the form

$$q(r_1, r_2, r_3 \mid y_1^*, y_2^*, y_3^*, \boldsymbol{x}; \boldsymbol{\eta}) =$$

$$\Pr(R_3 = r_3 \mid R_1 = r_1, R_2 = r_2, Y_1^* = y_1^*, Y_2^* = y_2^*, Y_3^* = y_3^*, \boldsymbol{x})$$

$$\times \ \Pr(R_2 = r_2 \mid R_1 = r_1, Y_1^* = y_1^*, Y_2^* = y_2^*, \boldsymbol{x})$$

$$\times \ \Pr(R_1 = r_1 \mid Y_1^* = y_1^*, \boldsymbol{x}).$$

(4.11)

To complete the specification of the non-response model, Baker modelled each of these conditional probabilities as a logistic regression. Under non-ignorable

non-response, these logistic regressions involve the unobserved outcomes as well as the observed outcomes and the covariates.

4.2 Selection Model for Binary Longitudinal Data with Informative Drop-out

Our main focus is to explore models for incomplete binary responses subject to informative drop-out for our annual data set as described in Section 2.2. The approach sketched in the previous section can be modified to serve our purpose.

We consider the case where drop-out occurs either at the first, second or third time point. Our data are then limited to 4 of the 8 possible patterns listed in Table 4.1: patterns 1, 2, 5 and 8. Patterns 2, 5 and 8 form a monotone pattern of non-response, and are also known as drop-outs. Following Baker, the probabilities for these 4 incomplete data patterns are given by equations (4.3), (4.4), (4.7) and (4.10). In other words, our model is a simplified version of Baker's more general selection model.

In the next few subsections, we specify particular forms for the outcome model, $f^*(y_1^*, y_2^*, y_3^* \mid x; \theta)$, and the drop-out model, $q(r_1, r_2, r_3 \mid y_1^*, y_2^*, y_3^*, x; \eta)$. The likelihood function is then assembled according to these models.

4.2.1 Outcome Model

Baker's outcome model $f^*(y_1^*, y_2^*, y_3^* \mid x; \theta)$ is specified in terms of two models: a marginal model (model for the univariate marginal probability) and an association model (model for the multivariate probability). There are several approaches to constructing marginal and association models with binary longitudinal data. Baker used the parameterization introduced by Ekholm (1991, 1992) which expresses $f^*(y_1^*, y_2^*, y_3^* \mid x; \theta)$ as a linear combination of marginal and association models.

Let $\theta = \{\beta, \alpha\}$, where $\beta = \{\beta_1, \beta_2, \beta_3\}$ and $\alpha = \{\alpha_{12}, \alpha_{13}, \alpha_{23}, \alpha_{123}\}$

are vectors of parameters associated with the marginal and association models, respectively. We model the logit of the marginal probability, $\Pr(Y_t^* = 1 \mid x)$ for $t = 1, 2, 3$, as a linear function of the covariates. More precisely, if $g_t(x; \beta_t) = \Pr(Y_t^* = 1 \mid x)$, then the marginal model is given by $\operatorname{logit}\{g_t(x; \beta_t)\} = X_t' \beta_t$. We denote the association model as $g_{st}(x; \alpha_{st}) = \Pr(Y_s^* = 1, Y_t^* = 1 \mid x)$, for $\{s, t\} = \{1, 2\}, \{1, 3\}, \{2, 3\}$ and $g_{123}(x; \alpha_{123}) = \Pr(Y_1^* = 1, Y_2^* = 1, Y_3^* = 1 \mid x)$, where $\operatorname{logit}\{g_{st}(x; \alpha_{st})\} = X_{st}' \alpha_{st}$ and $\operatorname{logit}\{g_{123}(x; \alpha_{123})\} = X_{123}' \alpha_{123}$. The probabilities for the different possible outcomes can then be expressed as follows:

$$
\begin{aligned}
f^*(1,1,1 \mid x; \theta) &= g_{123}(x; \alpha_{123}), \\
f^*(1,1,0 \mid x; \theta) &= g_{12}(x; \alpha_{12}) - g_{123}(x; \alpha_{123}), \\
f^*(1,0,1 \mid x; \theta) &= g_{13}(x; \alpha_{13}) - g_{123}(x; \alpha_{123}), \\
f^*(1,0,0 \mid x; \theta) &= g_1(x; \beta_1) - g_{12}(x; \alpha_{12}) - g_{13}(x; \alpha_{13}) + g_{123}(x; \alpha_{123}), \\
f^*(0,1,1 \mid x; \theta) &= g_{23}(x; \alpha_{23}) - g_{123}(x; \alpha_{123}), \\
f^*(0,1,0 \mid x; \theta) &= g_2(x; \beta_2) - g_{12}(x; \alpha_{12}) - g_{23}(x; \alpha_{23}) + g_{123}(x; \alpha_{123}), \\
f^*(0,0,1 \mid x; \theta) &= g_3(x; \beta_3) - g_{13}(x; \alpha_{13}) - g_{23}(x; \alpha_{23}) + g_{123}(x; \alpha_{123}), \\
f^*(0,0,0 \mid x; \theta) &= 1 - g_1(x; \beta_1) - g_2(x; \beta_2) - g_3(x; \beta_3) + g_{12}(x; \alpha_{12}) \\
&\quad + g_{13}(x; \alpha_{13}) + g_{23}(x; \alpha_{23}) - g_{123}(x; \alpha_{123}).
\end{aligned}
$$

The above probabilities must sum to 1. Further, each of these probabilities must be bounded between 0 and 1, so that there are many constraints on the parameters.

Note that the parameters in the marginal and association models can be interpreted as various types of odds ratios. More detailed interpretation of some of these parameters are provided in Chapter 7 where we discuss the results of the application of this model to our annual data set. But the parameters in the association models do not necessarily have direct interpretation relating to the strength of dependence among the responses. In other words, the magnitude of the parameter estimates may not explicitly reflect whether the responses are positively or negatively associated. Evaluating the correlations among the responses based on this

model are straightforward, although somewhat tedious.

4.2.2 Drop-out Model

We now consider the model for the drop-out process. We adopt Baker's idea for modelling the drop-out process as presented in (4.11). Let $\mathbf{r_{t-1}} = \{r_1, r_2, \ldots, r_{t-1}\}$ denote the previous pattern of non-response indicators up to time $t - 1$ and $\mathbf{y_t^*} = \{y_1^*, y_2^*, \ldots, y_t^*\}$ denote the outcomes up to and including the outcome at time t. Let $\boldsymbol{\eta}_t$ denote vectors of parameters associated with drop-out at time t, where $t = 1, 2, 3$. Further, denote

$$h_t(\mathbf{r_{t-1}}, \mathbf{y_t^*} \mid \boldsymbol{x}; \boldsymbol{\eta}_t) = \Pr(R_t = \mathrm{a} \mid \mathbf{r_{t-1}}, \mathbf{y_t^*}, \boldsymbol{x}) \qquad (4.12)$$

and model $\mathrm{logit}\{h_t(\mathbf{r_{t-1}}, \mathbf{y_t^*} \mid \boldsymbol{x}; \boldsymbol{\eta}_t)\}$ as a linear function of $\mathbf{y_t^*}$ and $\boldsymbol{x}$. The drop-out process is ignorable if $h_t(\mathbf{r_{t-1}}, \mathbf{y_t^*}, \boldsymbol{x}; \boldsymbol{\eta}_t)$ depends only on observed outcomes and covariates. More specifically, if $h_t(\mathbf{r_{t-1}}, \mathbf{y_t^*} \mid \boldsymbol{x}; \boldsymbol{\eta}_t)$ depends only on covariates, the drop-out is completely random; that is, the drop-out mechanism is referred to as CRD. If, on the other hand, $h_t(\mathbf{r_{t-1}}, \mathbf{y_t^*} \mid \boldsymbol{x}; \boldsymbol{\eta}_t)$ depends on the observed outcomes, and perhaps covariates, but not on the unobserved outcomes, the drop-out mechanism is referred to as random drop-out (RD). The drop-out is informative if $h_t(\mathbf{r_{t-1}}, \mathbf{y_t^*} \mid \boldsymbol{x}; \boldsymbol{\eta}_t)$ depends on the unobserved outcomes, and perhaps the observed outcomes and covariates as well.

Various authors such as Baker and Laird (1988), Fitzmaurice, Laird and Zahner (1996), and Glonek (1999), have drawn attention to the issue of identifiability for non-ignorable non-response models. If there are more independent parameters than available degrees of freedom, a model is clearly not identifiable. But with non-ignorable non-response, even some models with fewer independent parameters than available degrees of freedom are not identifiable (Baker and Laird, 1988). Baker (1995) established sufficient conditions for certain non-ignorable non-response models for three repeated binary outcomes to be identifiable. In particular, for $\mathbf{r_{t-1}}$ equal to {p, p}, {p, a}, {a, p}, or {p}, he considered logistic regressions in which

32

the dependence on outcomes is limited to two predictors: y^*_{LOR}, the last observed response (LOR), and y^*_{LUR}, the last unobserved response (LUR). Models that include the predictor y^*_{LUR} are non-ignorable. The values of y^*_{LOR} and y^*_{LUR} depend on the previous patterns of non-response. For $\mathbf{r_{t-1}}$ equal to {a, a}, {a}, and { }, $h_t(\mathbf{r_{t-1}}, \mathbf{y^*_t} \mid x; \boldsymbol{\eta}_t)$ depends only on covariates. More details on Baker's sufficient conditions is provided in Chapter 6.

For our case, $\mathbf{r_{t-1}}$ can take on six patterns: {p, p}, {p, a}, {a, a}, {p}, {a}, and { }. Since the only type of missing responses in our data set corresponds to drop-outs, we only need to model those $h_t(\mathbf{r_{t-1}}, \mathbf{y^*_t} \mid x; \boldsymbol{\eta}_t)$ where $\mathbf{r_{t-1}}$ equals {p, p}, {p}, or { }. More precisely, for the cases of drop-outs:

- consider $\mathbf{r_2}$:

  (i) when $\mathbf{r_2} = \{p, p\} \Rightarrow y^*_{LOR} = y^*_2$, $y^*_{LUR} = y^*_3$, as in Baker (1995);

  (ii) when $\mathbf{r_2} = \{p, a\} \Rightarrow \Pr(R_3 = a \mid \mathbf{r_2}, \mathbf{y^*_3}, x) = 1$;

  (iii) when $\mathbf{r_2} = \{a, a\} \Rightarrow \Pr(R_3 = a \mid \mathbf{r_2}, \mathbf{y^*_3}, x) = 1$.

- consider $\mathbf{r_1}$:

  (i) when $\mathbf{r_1} = \{p\} \Rightarrow y^*_{LOR} = y^*_1$, $y^*_{LUR} = y^*_2$, as in Baker (1995);

  (ii) when $\mathbf{r_1} = \{a\} \Rightarrow \Pr(R_2 = a \mid \mathbf{r_1}, \mathbf{y^*_2}, x) = 1$.

- when $\mathbf{r_0} = \{ \}$, Baker (1995) suggested $h_1(\{ \}, y^*_1 \mid x; \boldsymbol{\eta}_1)$ should depend only on the covariates, not on the observed and unobserved outcomes.

As in Baker (1995), we allow the models for $h_t(\mathbf{r_{t-1}}, \mathbf{y^*_t} \mid x; \boldsymbol{\eta}_t)$ when $\mathbf{r_{t-1}}$ equals {p,p} or {p} to be nested within one of the following:

1. Covariates (COV) * LUR [= COV + LUR + COV × LUR]:

$$
\begin{aligned}
\text{logit}[h_t(\mathbf{r_{t-1}}, \mathbf{y^*_t} \mid x; \boldsymbol{\eta}_t)] &= \eta^{r_{t-1}} + x' \, \eta^{r_{t-1}}_{COV} + \eta^{r_{t-1}}_{LUR} \, y^*_{LUR} \\
&\quad + x' \, \eta^{r_{t-1}}_{COV*LUR} \, y^*_{LUR},
\end{aligned}
\tag{4.13}
$$

33

2. COV + LOR + LUR:

$$\text{logit}[h_t(\mathbf{r_{t-1}}, \mathbf{y}_t^* \mid \boldsymbol{x}; \boldsymbol{\eta}_t)] = \eta^{\mathbf{r_{t-1}}} + \boldsymbol{x}' \, \eta_{COV}^{\mathbf{r_{t-1}}} + \eta_{LOR}^{\mathbf{r_{t-1}}} \, y_{LOR}^*$$
$$+ \, \eta_{LUR}^{\mathbf{r_{t-1}}} \, y_{LUR}^*, \tag{4.14}$$

3. LOR * LUR [= LOR + LUR + LOR × LUR]:

$$\text{logit}[h_t(\mathbf{r_{t-1}}, \mathbf{y}_t^* \mid \boldsymbol{x}; \boldsymbol{\eta}_t)] = \eta^{\mathbf{r_{t-1}}} + \eta_{LOR}^{\mathbf{r_{t-1}}} \, y_{LOR}^* + \eta_{LUR}^{\mathbf{r_{t-1}}} \, y_{LUR}^*$$
$$+ \, \eta_{LOR*LUR}^{\mathbf{r_{t-1}}} \, y_{LOR}^* \, y_{LUR}^*. \tag{4.15}$$

The drop-out model considered in Diggle and Kenward (1994) is a special case of the model COV + LOR + LUR. They assumed the drop-out mechanism only depended on LOR and LUR, and that the effects of these two predictors were the same across different drop-out occasions. Note that the same covariates can appear in both the drop-out and the outcome models.

4.2.3 Likelihood Function

We assemble these models into an explicit expression for the logarithm of the likelihood. Let $n_{y_1, y_2, y_3, x}$ denote the total number of subjects with outcome y_1 at time 1, y_2 at time 2, y_3 at time 3 and categorical covariate at level x. Further, denote $\boldsymbol{\eta} = \{\boldsymbol{\eta}_1, \boldsymbol{\eta}_2, \boldsymbol{\eta}_3\}$. We can then express the log-likelihood as $L(\boldsymbol{\theta}, \boldsymbol{\eta}) = \sum_x L_x(\boldsymbol{\theta}, \boldsymbol{\eta})$, where

$$
\begin{aligned}
L_x(\boldsymbol{\theta}, \boldsymbol{\eta}) = \ & n_{a,a,a,x} \log \, f(a, a, a \mid x; \boldsymbol{\theta}, \boldsymbol{\eta}) \\
& + \sum_{y_1^*=0}^{1} n_{y_1^*, a, a, x} \log \, f(y_1^*, a, a \mid x; \boldsymbol{\theta}, \boldsymbol{\eta}) \\
& + \sum_{y_1^*=0}^{1} \sum_{y_2^*=0}^{1} n_{y_1^*, y_2^*, a, x} \log \, f(y_1^*, y_2^*, a \mid x; \boldsymbol{\theta}, \boldsymbol{\eta}) \\
& + \sum_{y_1^*=0}^{1} \sum_{y_2^*=0}^{1} \sum_{y_3^*=0}^{1} n_{y_1^*, y_2^*, y_3^*, x} \log \, f(y_1^*, y_2^*, y_3^* \mid x; \boldsymbol{\theta}, \boldsymbol{\eta}), \quad (4.16)
\end{aligned}
$$

where the four functions $f(y_1^*, y_2^*, y_3^* \mid x; \boldsymbol{\theta}, \boldsymbol{\eta})$, $f(y_1^*, y_2^*, a \mid x; \boldsymbol{\theta}, \boldsymbol{\eta})$, $f(y_1^*, a, a \mid x; \boldsymbol{\theta}, \boldsymbol{\eta})$, and $f(a, a, a \mid x; \boldsymbol{\theta}, \boldsymbol{\eta})$, are specified in (4.3), (4.4), (4.7) and (4.10) respectively. We obtain the maximum likelihood estimates (MLEs) of the parameters, $\boldsymbol{\theta}$ and $\boldsymbol{\eta}$, by minimizing the negative log-likelihood using a quasi-newton minimization routine [26].

Chapter 5

Transition Model

5.1 Introduction

In this chapter, we model the outcome (or measurement) process using a transition model coupled with several models for the drop-out process as described in Chapter 4. The idea of using a transition model to describe the outcome process is motivated by Liu, Waternaux and Petkova (1999), who investigated the effect of human immunodeficiency virus (HIV) status on neurological impairment on a cohort of HIV positive and negative gay men. These subjects were followed for 5 years and assessed every 6 months. The primary outcome is the presence or absence of neurological impairment which varies over time. Predictors of outcome include fixed and time-varying covariates, such as age at baseline, HIV status, disease progression and time of assessment. Nearly half of the subjects dropped out before the end of the study for reasons that might have been related to the missing neurological data.

Liu *et al.* (1999) adapted the likelihood-based approach proposed by Diggle and Kenward (1994) for the analysis of a Gaussian longitudinal outcome with informative drop-out to analyze these binary longitudinal responses. More precisely, they assumed a first-order Markov chain transition model for the binary longitudinal responses combined with different logit models for the occurrence of drop-out.

Transition models are often used for equally-spaced longitudinal data when the interest is in prediction (Diggle and Kenward, 1994), and Liu *et al.* (1999) proposed such a model for the outcome process as their interest was in predicting neurological impairment.

Our data set consists of yearly observations on the presence or absence of exacerbations in MS patients. According to Liu *et al.* (1999), "In biomedical research, sequences of measurements are often fairly short and, in many cases, a first-order transition model is reasonable". Thus in this thesis, we embrace their idea of modelling the repeated binary responses with first-order transition models.

In the next section, we give an overview of the Liu *et al.* transition model for the outcome process and propose to combine this with Baker's ideas for modelling the drop-out process. We then briefly present the general expression of the log-likelihood under these models to conclude the chapter.

5.2 The Liu *et al.* Transition Model for Binary Longitudinal Data with Informative Drop-out

In this section, we illustrate the general approach of a first-order transition model. To keep the discussion simple and consistent with Chapter 4, we assume each subject is followed at three equally-spaced time points.

As in Chapter 4, $(Y_1^*, Y_2^*, Y_3^*, R_1, R_2, R_3, \mathbf{X})$, is the vector of random variables for the complete data, and $(Y_1, Y_2, Y_3, \mathbf{X})$ is the corresponding set of random variables for the incomplete data. The relationship between them is: $Y_t = Y_t^*$ if $R_t = \text{p}$ and $Y_t = a$ if $R_t = \text{a}$. The joint distribution for the complete data is factored as

$$\Pr(Y_1^* = y_1^*, Y_2^* = y_2^*, Y_3^* = y_3^*, R_1 = r_1, R_2 = r_2, R_3 = r_3 \mid \mathbf{X} = \mathbf{x})$$

$$= \Pr(R_1 = r_1, R_2 = r_2, R_3 = r_3 \mid y_1^*, y_2^*, y_3^*, \mathbf{x})$$

$$\times \Pr(Y_1^* = y_1^*, Y_2^* = y_2^*, Y_3^* = y_3^* \mid \mathbf{x}),$$

where $\Pr(Y_1^* = y_1^*, Y_2^* = y_2^*, Y_3^* = y_3^* \mid \boldsymbol{x})$ is known as the outcome model, and $\Pr(R_1 = r_1, R_2 = r_2, R_3 = r_3 \mid y_1^*, y_2^*, y_3^*, \boldsymbol{x})$ is the drop-out model.

As for Baker's selection model, the basic idea is to construct a model for both the outcome and drop-out processes. These models then specify a model for the incomplete data. The log-likelihood function is expressed in terms of these models and a maximum likelihood is employed to estimate the model parameters. The only difference from the previous chapter is that here we model the outcome process with transition models.

5.2.1 Outcome Model

Denote $H_t = \{y_1^*, \ldots, y_{t-1}^*\}$ as the responses up to but not including time t. The joint distribution of the equally-spaced outcome variables given the covariates, i.e. $\Pr(Y_1^* = y_1^*, \ldots, Y_t^* = y_t^* \mid \boldsymbol{x})$, can be decomposed as

$$
\begin{aligned}
\Pr(Y_1^* = y_1^*, \ldots, Y_t^* = y_t^* \mid \boldsymbol{x}) \; = \; & \Pr(Y_t^* = y_t^* \mid H_t, \boldsymbol{x}) \\
& \times \Pr(Y_{t-1}^* = y_{t-1}^* \mid H_{t-1}, \boldsymbol{x}) \\
& \times \ldots \times \Pr(Y_1^* = y_1^* \mid \boldsymbol{x}).
\end{aligned}
\tag{5.1}
$$

A transition model of order $q > 0$ postulates that the conditional distribution of y_t^*, given the history H_t, depends only on the observations: $y_{t-q}^*, \ldots, y_{t-1}^*$. A first-order $(q = 1)$ transition model for the case of three repeated responses is of form

$$
\begin{aligned}
\Pr(Y_1^* = y_1^*, Y_2^* = y_2^*, Y_3^* = y_3^* \mid \boldsymbol{x}) \; = \; & \Pr(Y_3^* = y_3^* \mid y_2^*, \boldsymbol{x}) \\
& \times \; \Pr(Y_2^* = y_2^* \mid y_1^*, \boldsymbol{x}) \\
& \times \; \Pr(Y_1^* = y_1^* \mid \boldsymbol{x}).
\end{aligned}
\tag{5.2}
$$

Liu *et al.* (1999) proposed using a first-order transition model for specifying the joint distribution of an equally-spaced binary outcome process. They employed a specific model for the conditional probabilities of the binary elements in the complete outcome vector $\mathbf{y}^*$ in which the conditional probabilities are assumed to depend only

on the covariates observed at the immediately previous time point, denoted $\mathbf{x_{t-1}}$. The form of the model is

$$\text{logit}\{\Pr(Y_t^* = 1 \mid y_{t-1}^*, \boldsymbol{x})\} = \mathbf{x_{t-1}'}\boldsymbol{\beta_1} + \beta_2 y_{t-1}^*, \qquad (5.3)$$

where $\boldsymbol{\beta_1}$ and β_2 are parameters to be estimated. The parameter β_2 represents the log odds ratio for presence at time t given presence at time $t-1$, against presence at time t given absence at time $t-1$.

At first glance, the assumed form of dependence on the covariates seems a bit peculiar since the covariates measured at time t should have a stronger influence on the response y_t^* than the covariates measured at time $t-1$. However, they noted that in most biomedical studies, there will be no information available after a subject drops out of the study; that is, if y_t^* is not observed, then $\mathbf{x_t}$ would not be observed either. They chose to overcome this (potential) data limitation problem by the aforementioned approach.

In summary, the outcome model is specified in terms of conditional distribution of y_t^* with the assumption that it depends only on y_{t-1}^* and $\mathbf{x_{t-1}}$. The structure of the associations among the responses is more restricted than in Baker's selection model in that this model assumes the association between Y_1^* and Y_2^* to be the same as the association between Y_2^* and Y_3^*.

5.2.2 Drop-out Model

Similarly, the drop-out model $\Pr(R_1 = r_1, R_2 = r_2, R_3 = r_3 \mid y_1^*, y_2^*, y_3^*, \boldsymbol{x})$ is specified in terms of R_t given $(\mathbf{r_{t-1}}, H_t, y_t^*, \boldsymbol{x})$ for $t = 2, 3$ as in Chapter 4. Liu *et al.* (1999) modelled these conditional probabilities as:

$$\text{logit}\{h_t(\mathbf{r_{t-1}}, H_t, y_t^* \mid \boldsymbol{x}; \eta_t)\} = \eta_0 + \eta_1 y_{t-1}^* + \eta_2 y_t^*. \qquad (5.4)$$

This is a special case of (4.13) and (4.14) in which the drop-out mechanism is assumed to be independent of the covariates and the parameters in the logistic regressions are the same regardless of $\mathbf{r_{t-1}}$. In their data set, the first observation

was always observed, so they did not need to consider drop-out models for the case where r_{t-1} equals $\{a\}$ or $\{\ \}$. But for our purposes, the drop-out probability with r_{t-1} equals $\{a\}$ is always 1 . For the case where r_{t-1} equals $\{\ \}$, we model $h_t(\{\ \}, y_1^* \mid x, \eta_t)$ according to Baker's (1995) suggestion; that is, this probability should depend only on the baseline covariates, not on the outcome measurements.

5.2.3 Likelihood Function

The general expression of the log-likelihood is the same as (4.15). Similarly, the four models for the incomplete data, $f(y_1^*, y_2^*, y_3^*)$, $f(y_1^*, y_2^*, a)$, $f(y_1^*, a, a)$ and $f(a, a, a)$ have the forms (4.3), (4.4), (4.7) and (4.10) respectively, which are specified in terms of the models described in the previous two subsections. Liu et $al.$ (1999) used the S-PLUS function ms to obtain the maximum likelihood estimates (MLE) for the parameters in their problem. As in Chapter 4, we use a quasi-newton minimization routine to obtain the MLEs for β and η.

In the next chapter, we discuss potential identifiability problems in non-ignorable non-response models for incomplete binary data before proceeding to use the Baker's selection model and the Liu et $al.$ transition model to analyze our annual data in Chapter 7.

Chapter 6

Identifiability in Models for Incomplete Binary Data

6.1 Introduction

Analyses based on an assumption of ignorable non-response when the non-response mechanism is informative (or non-ignorable) can lead to misleading or biased results. Thus in the past decade, various authors have developed models for continuous and categorical response data subject to non-ignorable non-response. In particular, likelihood-based analyses have been widely employed since there is a choice of whether or not to introdure an explicit model for the non-response mechanism. Little and Rubin (1987) noted that, by incorporating a model for non-response in a likelihood-based approach, valid inferences can be obtained when the non-response mechanism is non-ignorable provided the non-response model correctly represents the non-response mechanism. Most of these papers have emphasized the formulation and implementation of those models. However, it has been observed that such models present certain analytical difficulties. In particular, it can happen that the parameters of the non-ignorable models are not identifiable or the maximum likelihood solutions can lie on the boundary of the parameter space.

Baker and Laird (1988) drew attention to the issue of boundary solutions to the maximum likelihood equations in a non-longitudinal setting. They illustrated this issue with the pre-election data from four successive Roper polls carried out to predict the proportion of voters preferring Truman in the 1948 presidential election. The four variables used in their analyses were time of survey (X_T = July, August, September, October), economic class of voter (X_E = A, B, C, D), voter preference (Y = Truman, Dewey, other), and expression of preference (R = yes, no). They employed two different log-linear models to describe the related regressions: the marginal outcome model for the $X_T X_E Y$ margin (a $4 \times 4 \times 3$ array) which describes the regression of Y on X_T and X_E, and the non-response model for the full contingency table $X_T X_E Y R$, which describes the regression of R on X_T, X_E, and Y. For this framework, they showed that with non-ignorable non-response models, over-parameterized and saturated models may not yield a perfect fit and the likelihood equations can be satisfied by boundary values even when all observed counts are strictly positive.

As discussed in Chapter 4, Baker (1995) used a selection model to analyze data from the Muscatine Risk Factor Study to investigate the effects of gender and age on obesity in schoolchildren who ranged between ages of 5 and 13 years. In these data, each child was intended to have three binary responses at 2-year intervals indicating whether or not they were obese at that point in time. However, there was a substantial amount of non-response due to no consent from the parents or the child not being in school on the day of the examination. In this special setting, Baker (1995) obtained sufficient conditions for non-ignorable non-response models to be identifiable. Following Baker's ideas, we establish sufficient conditions for non-ignorable drop-out models to be identifiable in the last section of this chapter.

For the context of models for incomplete multivariate binary data, Fitzmaurice *et al.* (1996) suggested some simple procedures for examining local and global identifiability in models with non-ignorable non-response. A summary of this por-

tion of that paper is given in the following section. More recently, Glonek (1999) formulated the specific application considered in Section 3 of Fitzmaurice *et al.* (1996) in a somewhat more general fashion to discuss the identifiability issue for models for incomplete binary data. He derived necessary and sufficient conditions for certain simple non-ignorable non-response models (including some of the models considered by Fitzmaurice *et al.* for their application) to be identifiable. His results show that these models are identifiable except at a set of special parameter values where the conditions fail to hold.

The consideration of model identifiability is an issue that should be resolved prior to estimation, because it does not make sense to attempt interpretation of an estimate of a parameter that is not statistically identifiable. In Section 6.2, we describe the procedures suggested by Fitzmaurice *et al.* (1996) for checking model identifiability. We also describe the necessary and sufficient conditions obtained by Glonek (1999) and the implications of these results for Fitzmaurice *et al.*'s suggested approaches to examining the identifiability of non-ignorable non-response models. Baker's (1995) development of sufficient conditions for the identifiability of certain non-ignorable non-response models for the case where the data consist of three repeated binary responses with all possible patterns of non-response is briefly summarized in Section 6.3. We conclude this chapter by applying Baker's ideas to the special situation of interest here of models corresponding to monotone non-response patterns.

6.2 Discussion in Fitzmaurice *et al.* (1996) and Glonek (1999)

Fitzmaurice *et al.* (1996) proposed a likelihood-based regression model for analyzing incomplete multivariate binary responses based on the multivariate binary model proposed by Fitzmaurice and Laird (1993). The latter model is extended to

accommodate incomplete data by assuming a logistic model for the non-response mechanism which depends on covariates and on both the observed and unobserved responses. This idea is motivated by Diggle and Kenward (1994) and Molenberghs, Kenward, and Lesaffre (1997). Throughout Fitzmaurice *et al.* (1996), monotone non-response is assumed.

Various authors have pointed out that the identifiability is an important yet unresolved issue in non-ignorable non-response models. As Fitzmaurice *et al.* (1996) stated, "So far, no general and practically useful necessary and sufficient conditions for identifiability are available". Fitzmaurice *et al.* (1996) suggested some simple procedures for examining the identifiability status of non-ignorable models for the case of discrete response variables; these are described in the next subsection. The following subsection describes Glonek's results and the implications of those results for the procedures suggested by Fitzmaurice *et al.* (1996).

6.2.1 Fitzmaurice *et al.*'s Suggested Procedures

Fitzmaurice *et al.* (1996) indicate what they mean by a non-identifiable model. Consider a non-ignorable model with parameters $(\boldsymbol{\theta}, \boldsymbol{\eta})$, where $\boldsymbol{\theta}$ and $\boldsymbol{\eta}$ are the vectors of parameters associated with the outcome model and the non-response model respectively. If it is the case that there are distinct parameter vectors $(\boldsymbol{\theta}_0, \boldsymbol{\eta}_0) \neq (\boldsymbol{\theta}_1, \boldsymbol{\eta}_1)$ such that

$$f(\mathbf{y}_{\mathbf{o}i}, \mathbf{r}_i \mid \boldsymbol{\theta}_0, \boldsymbol{\eta}_0) = f(\mathbf{y}_{\mathbf{o}i}, \mathbf{r}_i \mid \boldsymbol{\theta}_1, \boldsymbol{\eta}_1)$$

for all $\mathbf{y}_{\mathbf{o}i}$ (the vector of observed responses for the *i-th* subject) and $\mathbf{r}_i$ (the vector of response indicators for the *i-th* subject), then $L(\boldsymbol{\theta}_0, \boldsymbol{\eta}_0) = L(\boldsymbol{\theta}_1, \boldsymbol{\eta}_1)$ and the model is not statistically identifiable. Showing algebraically that all of the parameters in non-ignorable models are identifiable is not trivial (Fitzmaurice *et al.*, 1996). If there are more parameters to be estimated than available degrees of freedom in the data, the model is clearly not identifiable. But having no more parameters to

44

be estimated than the available degrees of freedom is not sufficient to guarantee identifiability for non-ignorable non-response models (Baker and Laird, 1988).

Fitzmaurice et al. (1996) suggested some simple procedures for examining the identifiability of non-ignorable non-response models. Since local identifiability (the model is identifiable in a subspace of the entire parameter space) is a necessary condition for a model to be globally identifiable (the model is identifiable throughout the entire parameter space), a first step is to examine the local identifiability status of the model by checking that the Fisher information matrix is nonsingular. Rothenberg (1971) has shown that, subject to certain regularity conditions, if the Fisher information matrix is nonsingular, then the model is locally identifiable. This idea of using the Fisher information matrix to determine the identifiability status of a model was described in the context of latent class models by Goodman (1974).

- *Checking for Local Identifiability*

Fitzmaurice et al. (1996) suggested selecting a reasonable set of parameter values for (θ, η) and evaluating the Fisher information matrix at this particular set of parameter values. This can be accomplished by taking the expectation of the outer-product of the score equations, summing all the possible realizations weighted by their respective probabilities. In other words, for each possible realization of $(\mathbf{Y_i}, \mathbf{R_i}, \mathbf{X_i})$, calculate the sample covariance matrix of the scores and weight these contributions by their respective joint probabilities. By summing over all possible realizations, the Fisher information matrix is obtained. The information matrix can then be checked to see whether it is nonsingular at this set of parameter values.

- *Checking for Global Identifiability*

Having established local identifiability, Fitzmaurice et al. (1996) recommend assessing global identifiability with the following procedure:

1. Select a set of reasonable values for the parameters (θ, η) (e.g. the estimated values) and use them to generate an artificial sample comprising one observa-

tion for each possible realization of $(\mathbf{Y_i}, \mathbf{R_i}, \mathbf{X_i})$.

2. Solve for $(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\eta}})$ from the likelihood equations obtained by weighting the contribution for each possible realization by its respective probability.

If the resulting estimate $(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\eta}})$ does not equal $(\boldsymbol{\theta}, \boldsymbol{\eta})$, then the model is not globally identifiable, and those parameters that give a different value are not statistical identifiable. If the estimate $(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\eta}})$ equals $(\boldsymbol{\theta}, \boldsymbol{\eta})$ for a whole grid of reasonable values for $(\boldsymbol{\theta}, \boldsymbol{\eta})$, then the model is *most likely* identifiable (Fitzmaurice *et al.*, 1996).

Fitzmaurice *et al.* (1996) provide a simple example intended to show that the model identifiability problem exists even when the number of parameters is no more than the available degrees of freedom from the data. For the *i-th* patient there are two binary responses, Y_{i1} and Y_{i2}, and a dichotomous covariate, X_i. Y_{i1} is always observed but Y_{i2} is subject to missingness. Thus for each value of X_i, there are 6 possible outcomes for (Y_{i1}, Y_{i2}): $(0,0), (0,1), (0,a), (1,0), (1,1), (1,a)$. Consequently, the observed data have 10 degrees of freedom, 5 for each of the two possible values of X_i.

The outcome model they considered is not fully saturated and they also considered several non-ignorable non-response models. More specifically, the outcome model consists of two parts: a marginal model (for the means of the responses) and an association model. The (unrestricted) marginal model is parametrized as

$$\text{logit}\{E(Y_{ij})\} = \beta_{0j} + \beta_{1j}X_i;$$

for $j = 1, 2$, but the association between Y_{i1} and Y_{i2} is assumed to be constant across X_i, i.e. the conditional log odds ratios are assumed to be constant across X_i. Thus, the outcome model involves 5 parameters. This outcome model is coupled with 8 non-ignorable non-response models having at most 5 parameters. With R_{i2}

denoting the response indicator for Y_{i2}, these models are:

1. $\quad \text{logit}\{\Pr(R_{i2} = \text{p})\} \quad = \quad \eta_0 + \eta_1 Y_{i2}$

2. $\quad \text{logit}\{\Pr(R_{i2} = \text{p})\} \quad = \quad \eta_0 + \eta_1 X_i + \eta_2 Y_{i2}$

3. $\quad \text{logit}\{\Pr(R_{i2} = \text{p})\} \quad = \quad \eta_0 + \eta_1 Y_{i1} + \eta_2 Y_{i2}$

4. $\quad \text{logit}\{\Pr(R_{i2} = \text{p})\} \quad = \quad \eta_0 + \eta_1 X_i + \eta_2 Y_{i2} + \eta_3 X_i \times Y_{i2}$

5. $\quad \text{logit}\{\Pr(R_{i2} = \text{p})\} \quad = \quad \eta_0 + \eta_1 X_i + \eta_2 Y_{i1} + \eta_3 Y_{i2}$

6. $\quad \text{logit}\{\Pr(R_{i2} = \text{p})\} \quad = \quad \eta_0 + \eta_1 Y_{i1} + \eta_2 Y_{i2} + \eta_3 Y_{i1} \times Y_{i2}$

7. $\quad \text{logit}\{\Pr(R_{i2} = \text{p})\} \quad = \quad \eta_0 + \eta_1 X_i + \eta_2 Y_{i1} + \eta_3 Y_{i2} + \eta_4 X_i \times Y_{i1}$

8. $\quad \text{logit}\{\Pr(R_{i2} = \text{p})\} \quad = \quad \eta_0 + \eta_1 X_i + \eta_2 Y_{i1} + \eta_3 Y_{i2} + \eta_4 X_i \times Y_{i2}.$

Based on the use of their suggested procedures, Fitzmaurice *et al.* (1996) claimed that only three of these eight non-response models (Models 1, 2, and 4) are statistically identifiable. However, they do not indicate how they selected reasonable sets of values for the parameters and how many sets they checked to reach their conclusions.

6.2.2 Glonek's Necessary and Sufficient Conditions

Glonek (1999) attacks the model identifiability problem from a different point of view. He formulated the problem considered in Fitzmaurice *et al.* (1996) in a more general fashion to address the issue of identifiability. Two binary responses, Y_1 and Y_2, and a categorical covariate X with I levels are considered. Only Y_2 is subject to non-response and R_2 is the response indicator for Y_2 ($R_2 = \text{p}$ if Y_2 is observed and $R_2 = \text{a}$ otherwise).

The outcome model is denoted as

$$\pi_{ijk} \quad = \quad \Pr(Y_1 = j, Y_2 = k \mid X = i)$$

for $j, k = 0, 1$. The non-response model is denoted as

$$\rho_{ijk} \quad = \quad \Pr(R_2 = \text{p} \mid Y_1 = j, Y_2 = k, X = i).$$

Thus, the observations corresponding to the *i-th* level of the covariate X are multinomial across six cells with probabilities

$$\theta_{ijk} = \pi_{ijk}\rho_{ijk} = \Pr(Y_1 = j, Y_2 = k, \text{both responses observed} \mid X = i)$$

$$\theta_{ij*} = \pi_{ij0}(1 - \rho_{ij0}) + \pi_{ij1}(1 - \rho_{ij1}) = \Pr(Y_1 = j, Y_2 \text{ unobserved} \mid X = i).$$

The simple example used by Fitzmaurice *et al.* (1996) to illustrate their suggested procedures for checking model identifiability is of this form. As described in the previous subsection, for the case of a binary covariate ($I = 2$), they considered a restricted model for π_{ijk} involving no three-factor interaction and eight different models for ρ_{ijk}.

Combined with an unrestricted model for π_{ijk}, Glonek (1999) considered homogeneous non-response models of two forms:

$$\rho_{ijk} = \rho_{jk}, \tag{6.1}$$

and

$$\rho_{ijk} = \rho_{ik}. \tag{6.2}$$

In the first of these models, the probability of response is independent of the covariate, while in the second, the probability of response does not depend on the first response variable. Non-response models 1, 3 and 6 of Fitzmaurice *et al.* (1996) are of the first form, whereas models 1, 2, and 4 are of the second form; models 5, 7, and 8 are of more general forms.

For the case $I = 2$ with non-response model (6.1), Glonek showed that the condition

$$\Pr(Y_2 = 1 \mid Y_1 = j, X = 1) \neq \Pr(Y_2 = 1 \mid Y_1 = j, X = 2) \tag{6.3}$$

for $j = 0, 1$, is necessary and sufficient for the parameters of the model to be identified. The condition (6.3) would generally be satisfied, even under the restriction of no three-factor interaction incorporated into the Fitzmaurice *et al.* (1996) outcome

model. However, the restriction does not imply the condition (6.3); the condition could fail to hold for specific values of the parameters. Hence, their outcome model combined with any of their non-ignorable non-response models 1, 3, and 6 is identifiable except at those special values of the parameters where (6.3) fails to hold.

Similarly for the case $I = 2$ with non-response model (6.2), a necessary and sufficient condition for the parameters of the model to be identified is

$$\Pr(Y_2 = 1 \mid Y_1 = 0, X = i) \quad \neq \quad \Pr(Y_2 = 1 \mid Y_1 = 1, X = i) \qquad (6.4)$$

for $i = 1, 2$. The proof is provided in Appendix A. Again, even under the restriction of no three-factor interaction in the outcome model, the condition (6.4) would generally be satisfied. But the restriction does not imply the condition. Hence, the Fitzmaurice et al. (1996) outcome model combined with any of their non-ignorable non-response models 1, 2 and 4 is identifiable except at those special values of the parameters where (6.4) fails to hold.

Contrary to the conclusions of Fitzmaurice et al. (1996), Glonek was able to establish that with these homogeneous non-response models, the Fitzmaurice et al. models 1, 2, 3, 4 and 6 for this simple example are identifiable except at a set of special values of the parameters. (He did not address the issue for the Fitzmaurice et al. non-response models 5, 7 and 8.) Thus, Glonek established that the identifiability status of these models depends on the particular values of the parameters.

Glonek also provided a simple example with a non-homogeneous non-response model where this phenomenon occurs. This is problematic for inference since it may happen for a particular set of data that the maximum likelihood estimates are well-defined in the sense that the parameters are identified while, in fact, the true values of the parameters that generated the data are not. In such cases, it is clear that local calculations performed at the MLE will not bring to light this underlying non-identifiability. This phenomenon is different from the structural type of non-identifiability that would lead to rank deficiency in the Fisher information matrix, as considered by Fitzmaurice et al. (1996). Hence, the procedures suggested by

Fitzmaurice *et al.* (1996) are not adequate to resolve the issue of identifiability.

Our annual data setting is slightly different from the problem Glonek considered. We have three binary responses, Y_1, Y_2, Y_3, and all are subject to non-response. The derivation of the necessary and sufficient conditions for the identifiability of non-ignorable non-response models following Glonek's ideas appears to be much more complicated in our setting.

However, we were able to establish sufficient conditions for certain non-ignorable models to be identified in our setting, by following the ideas illustrated in Baker (1995). We briefly describe Baker's ideas in the next section and conclude this chapter with a description of the sufficient conditions we established.

6.3 Discussion of Model Identifiability for Incomplete Binary Responses in Baker (1995)

In Chapter 4, we described Baker's selection model for three repeated binary responses. He pointed out that all models with ignorable non-response are identifiable, but identifiability becomes a concern with non-ignorable non-response models. To restrict his models to those that are identifiable, he introduced two predictors for the non-response model: y^*_{LOR}, the last observed response (LOR), and y^*_{LUR}, the last unobserved response (LUR). Non-response models that include the predictor y^*_{LUR} are non-ignorable.

Recall that Baker modelled $\Pr(R_1 = r_1, R_2 = r_2, R_3 = r_3 \mid y_1^*, y_2^*, y_3^*, x)$ in terms of conditional probabilities assuming the non-response does not depend on future events; that is,

$$
\begin{aligned}
\Pr(R_1 = r_1, R_2 = r_2, R_3 = r_3 \mid y_1^*, y_2^*, y_3^*, x) =& \\
P(R_3 = r_3 \mid R_1 = r_1, R_2 = r_2, Y_1^* = y_1^*, Y_2^* = y_2^*, Y_3^* = y_3^*, x) & \\
\times\, P(R_2 = r_2 \mid R_1 = r_1, Y_1^* = y_1^*, Y_2^* = y_2^*, x) & \\
\times\, P(R_1 = r_1 \mid Y_1^* = y_1^*, x). &
\end{aligned}
\tag{6.5}
$$

Each of these conditional probabilities is modelled as a logistic regression that depends on y^*_{LOR} and y^*_{LUR}. The values of these predictors are determined by the previous observation pattern, $r_{t-1} = \{r_1, r_2, \ldots, r_{t-1}\}$. He claimed that the non-ignorable non-response models are identifiable if the following conditions are satisfied:

A. When r_{t-1} equals $\{a, a\}$, $\{a\}$, or $\{\ \}$, the corresponding conditional non-response probabilities should depend only on covariates.

B. When r_{t-1} equals $\{p, p\}$, $\{p, a\}$, $\{a, p\}$, or $\{p\}$, the non-response models should be nested within one of the following three types:

   (a) COV ∗ LUR;

   (b) COV + LOR + LUR;

   (c) LOR ∗ LUR.

Baker allowed the model parameters to differ for each of the previous observation patterns. Some of the details of the verification of identifiability are presented in the appendix of his paper.

Our situation is slightly different from that Baker considered. He had 7 non-response history patterns to consider, i.e. $\{p, p\}$, $\{p, a\}$, $\{a, p\}$, $\{a, a\}$, $\{p\}$, $\{a\}$, and $\{\ \}$. Since the non-response in our data set is monotonic, we need to consider only three different non-response history patterns: $\{p, p\}$, $\{p\}$, and $\{\ \}$. In the following section, we present verifications of the identifiability of the non-ignorable non-response models considered in our context.

6.4 Discussion of Model Identifiability

Our data set is a special case of Baker's general data structure as we have only monotone non-responses, i.e. drop-outs. In particular, we have four monotone non-response patterns to consider: $\{p, p, p\}$, $\{p, p, a\}$, $\{p, a, a\}$, and $\{a, a, a\}$. Recall that

$\boldsymbol{\theta} = \{\beta, \boldsymbol{\alpha}\}$ and $\boldsymbol{\eta} = \{\boldsymbol{\eta}_1, \boldsymbol{\eta}_2, \boldsymbol{\eta}_3\}$. As in Chapter 4, we model the incomplete data in terms of the product of the outcome model,

$$\Pr(Y_1^* = y_1^*, Y_2^* = y_2^*, Y_3^* = y_3^* \mid \mathbf{X}) = f^*(y_1^*, y_2^*, y_3^* \mid \boldsymbol{x}; \boldsymbol{\theta}),$$

and the drop-out model,

$$\Pr(R_1 = r_1, R_2 = r_2, R_2 = r_2 \mid Y_1^* = y_1^*, Y_2^* = y_2^*, Y_3^* = y_3^*, \mathbf{X})$$
$$= q(r_1, r_2, r_3 \mid y_1^*, y_2^*, y_3^*, \boldsymbol{x}; \boldsymbol{\eta}),$$

where $q(r_1, r_2, r_3 \mid y_1^*, y_2^*, y_3^*, \boldsymbol{x}; \boldsymbol{\eta})$ is specified as in (6.5). Recall also that $\mathbf{y}_t^*$ denotes the outcomes up to and including occasion t and $\mathbf{r}_{t-1}$ denotes the non-response history prior to time t. The three conditional non-response probabilities are denoted as follows:

$$\Pr(R_3 = \mathrm{a} \mid \mathbf{r_2} = \{\mathrm{p,p}\}, \mathbf{y}_3^*, \boldsymbol{x}) = h_3(\{\mathrm{p,p}\}, \mathbf{y}_3^* \mid \boldsymbol{x}; \boldsymbol{\eta}_3)$$
$$\Pr(R_2 = \mathrm{a} \mid \mathbf{r_1} = \{\mathrm{p}\}, \mathbf{y}_2^*, \boldsymbol{x}) = h_2(\{\mathrm{p}\}, \mathbf{y}_2^* \mid \boldsymbol{x}; \boldsymbol{\eta}_2)$$
$$\Pr(R_1 = \mathrm{a} \mid \mathbf{r_0} = \{\ \}, \mathbf{y}_3^*, \boldsymbol{x}) = h_1(\{\ \}, \mathbf{y}_1^* \mid \boldsymbol{x}; \boldsymbol{\eta}_1).$$

Consequently, the drop-out models for the four monotone non-response patterns are

$$q(\mathrm{p,p,p} \mid \mathbf{y}_3^*, \boldsymbol{x}; \boldsymbol{\eta}) = [1 - h_3(\{\mathrm{p,p}\}, \mathbf{y}_3^* \mid \boldsymbol{x}; \boldsymbol{\eta}_3)][1 - h_2(\{\mathrm{p}\}, \mathbf{y}_2^* \mid \boldsymbol{x}; \boldsymbol{\eta}_2)]$$
$$\times [1 - h_1(\{\ \}, \mathbf{y}_1^* \mid \boldsymbol{x}; \boldsymbol{\eta}_1)]$$
$$q(\mathrm{p,p,a} \mid \mathbf{y}_2^*, \boldsymbol{x}; \boldsymbol{\eta}) = h_3(\{\mathrm{p,p}\}, \mathbf{y}_3^* \mid \boldsymbol{x}; \boldsymbol{\eta}_3)[1 - h_2(\{\mathrm{p}\}, \mathbf{y}_2^* \mid \boldsymbol{x}; \boldsymbol{\eta}_2)]$$
$$\times [1 - h_1(\{\ \}, \mathbf{y}_1^* \mid \boldsymbol{x}; \boldsymbol{\eta}_1)]$$
$$q(\mathrm{p,a,a} \mid \mathbf{y}_1^*, \boldsymbol{x}; \boldsymbol{\eta}) = h_2(\{\mathrm{p}\}, \mathbf{y}_2^* \mid \boldsymbol{x}; \boldsymbol{\eta}_2)[1 - h_1(\{\ \}, \mathbf{y}_1^* \mid \boldsymbol{x}; \boldsymbol{\eta}_1)]$$
$$q(\mathrm{a,a,a} \mid \boldsymbol{x}; \boldsymbol{\eta}) = h_1(\{\ \}, \mathbf{y}_1^* \mid \boldsymbol{x}; \boldsymbol{\eta}_1).$$

For the case of categorical covariates, the kernel of the log-likelihood function is $L(\boldsymbol{\theta},\boldsymbol{\eta}) = \sum_x L_x(\boldsymbol{\theta},\boldsymbol{\eta})$, where

$$
\begin{aligned}
L_x(\boldsymbol{\theta},\boldsymbol{\eta}) &= \sum_{y_1^*=0}^{1}\sum_{y_2^*=0}^{1}\sum_{y_3^*=0}^{1} n_{y_1^*,y_2^*,y_3^*,x} \log\Big\{ f^*(y_1^*,y_2^*,y_3^* \mid x,\boldsymbol{\theta})[1 - h_3(\{p,p\},\mathbf{y_3^*} \mid x;\boldsymbol{\eta}_3)] \\
&\qquad \times [1 - h_2(\{p\},\mathbf{y_2^*} \mid x;\boldsymbol{\eta}_2)][1 - h_1(\{\ \},\mathbf{y_1^*} \mid x;\boldsymbol{\eta}_1)]\Big\} \\
&+ \sum_{y_1^*=0}^{1}\sum_{y_2^*=0}^{1} n_{y_1^*,y_2^*,a,x} \log\Big\{ \sum_{y_3^*=0}^{1} f^*(y_1^*,y_2^*,y_3^* \mid x,\boldsymbol{\theta}) h_3(\{p,p\},\mathbf{y_3^*} \mid x;\boldsymbol{\eta}_3) \\
&\qquad \times [1 - h_2(\{p\},\mathbf{y_2^*} \mid x;\boldsymbol{\eta}_2)][1 - h_1(\{\ \},\mathbf{y_1^*} \mid x;\boldsymbol{\eta}_1)]\Big\} \\
&+ \sum_{y_1^*=0}^{1} n_{y_1^*,a,a,x} \log\Big\{ \sum_{y_2^*=0}^{1}\sum_{y_3^*=0}^{1} f^*(y_1^*,y_2^*,y_3^* \mid x,\boldsymbol{\theta}) h_2(\{p\},\mathbf{y_2^*} \mid x;\boldsymbol{\eta}_2) \\
&\qquad \times [1 - h_1(\{\ \},\mathbf{y_1^*} \mid x;\boldsymbol{\eta}_1)]\Big\} \\
&+ n_{a,a,a,x} \log\Big\{ \sum_{y_1^*=0}^{1}\sum_{y_2^*=0}^{1}\sum_{y_3^*=0}^{1} f^*(y_1^*,y_2^*,y_3^* \mid x,\boldsymbol{\theta}) h_1(\{\ \},\mathbf{y_1^*} \mid x;\boldsymbol{\eta}_1)\Big\}.
\end{aligned}
$$

(6.6)

In the following subsections, we discuss the identifiability of the drop-out models by verifying whether the conditional probabilities $h_3(\{p,p\},\mathbf{y_3^*} \mid x;\boldsymbol{\eta}_3)$, $h_2(\{p\},\mathbf{y_2^*} \mid x;\boldsymbol{\eta}_2)$ and $h_1(\{\ \},\mathbf{y_1^*} \mid x;\boldsymbol{\eta}_1)$ are identifiable under the conditions described in the previous section. That is, $h_3(\{p,p\},\mathbf{y_3^*} \mid x;\boldsymbol{\eta}_3)$ could depend on y_3^*, y_2^* and x, while $h_2(\{p\},\mathbf{y_2^*} \mid x;\boldsymbol{\eta}_2)$ could depend on y_1^*, y_2^* as well as x. However, $h_1(\{\ \},\mathbf{y_1^*} \mid x;\boldsymbol{\eta}_1)$ is allowed to depend only on x.

6.4.1 Identifiability of $h_3(\{p,p\},\mathbf{y_3^*} \mid x;\boldsymbol{\eta}_3)$

The contribution of $h_3(\{p,p\},\mathbf{y_3^*} \mid x;\boldsymbol{\eta}_3)$ to $L_x(\boldsymbol{\theta},\boldsymbol{\eta})$ is given by

$$\sum_{y_1^*=0}^{1}\sum_{y_2^*=0}^{1}\sum_{y_3^*=0}^{1} n_{y_1^*,y_2^*,y_3^*,x} \log\Big\{ f^*(y_1^*,y_2^*,y_3^* \mid x,\boldsymbol{\theta})[1-h_3(\{p,p\},\mathbf{y_3^*} \mid x;\boldsymbol{\eta}_3)]$$

$$\times\, [1-h_2(\{p\},\mathbf{y_2^*} \mid x;\boldsymbol{\eta}_2)][1-h_1(\{\ \},\mathbf{y_1^*} \mid x;\boldsymbol{\eta}_1)]\Big\}$$

$$+\sum_{y_1^*=0}^{1}\sum_{y_2^*=0}^{1} n_{y_1^*,y_2^*,a,x} \log\Big\{ \sum_{y_3^*=0}^{1} f^*(y_1^*,y_2^*,y_3^* \mid x,\boldsymbol{\theta})h_3(\{p,p\},\mathbf{y_3^*} \mid x;\boldsymbol{\eta}_3)$$

$$\times\, [1-h_2(\{p\},\mathbf{y_2^*} \mid x;\boldsymbol{\eta}_2)][1-h_1(\{\ \},\mathbf{y_1^*} \mid x;\boldsymbol{\eta}_1)]\Big\} \quad (6.7)$$

To simplify the notation, substitute i for y_1^*, j for y_2^* and k for y_3^* and denote $n_{y_1^*,y_2^*,y_3^*,x} = m_{xijk}$ and $n_{y_1^*,y_2^*,a,x} = w_{xij}$. We further define

$$\rho_{xijk} = f^*(y_1^*,y_2^*,y_3^* \mid x,\boldsymbol{\theta})[1-h_3(\{p,p\},\mathbf{y_3^*} \mid x;\boldsymbol{\eta}_3)]$$

$$\times\, [1-h_2(\{p\},\mathbf{y_2^*} \mid x;\boldsymbol{\eta}_2)]\,[1-h_1(\{\ \},\mathbf{y_1^*} \mid x;\boldsymbol{\eta}_1)]$$

$$\phi_{xjk} = \frac{h_3(\{p,p\},\mathbf{y_3^*} \mid x;\boldsymbol{\eta}_3)}{1-h_3(\{p,p\},\mathbf{y_3^*} \mid x;\boldsymbol{\eta}_3)},$$

where the notation reflects that $h_3(\{p,p\},\mathbf{y_3^*} \mid x;\boldsymbol{\eta}_3)$ does not depend upon y_1^*. Then (6.7) can be re-expressed as:

$$\sum_{i=0}^{1}\sum_{j=0}^{1}\sum_{k=0}^{1} m_{xijk} \log \rho_{xijk} + \sum_{i=0}^{1}\sum_{j=0}^{1} w_{xij} \log\Big\{ \sum_{k=0}^{1} \rho_{xijk}\phi_{xjk}\Big\}.$$

This is identical to the log-likelihood for a contingency table $\{m_{xijk}\}$ with a supplementary margin $\{w_{xij}\}$ corresponding to cases where k was not observed. Therefore, the expected cell counts for m_{xijk} and w_{xij} are $\mu_{xijk} = \rho_{xijk}(m_{x+++} + w_{x++})$ and $\sum_{k=0}^{1} \mu_{xijk}\phi_{xjk}$, respectively.

We address the identifiability of $h_3(\{p,p\},\mathbf{y_3^*} \mid x;\boldsymbol{\eta}_3)$ for the three specific forms of non-response models introduced by Baker (1995). In each case, a saturated outcome model is assumed.

- **COV * LUR**

This model has $\phi_{xjk} = \phi_{xk}$, implying two distinct parameters for each level of x.

A perfect fit requires $\mu_{xijk} = m_{xijk}$ and $w_{xij} = \sum_{k=0}^{1} \mu_{xijk} \ \phi_{xk}$. Hence, we require $w_{xij} = \sum_{k=0}^{1} m_{xijk}\phi_{xk}$. Thus, for each level of x, we have four equations in the two unknowns, ϕ_{x0} and ϕ_{x1}. The parameters are overdetermined even if x has only one level. Hence, $h_3(\{p, p\}, \mathbf{y}_3^* \mid x; \boldsymbol{\eta}_3)$ is identifiable under this specification.

- **COV + LOR + LUR**

In this model, we can represent $\phi_{xjk} = \phi_x \ \phi_j \ \phi_k$. If we denote

$$\phi_{111} = \phi, \ \ \phi_{110} = \phi \ \phi_K, \ \ \phi_{101} = \phi \ \phi_J, \ \ \phi_{100} = \phi \ \phi_J \ \phi_K,$$

then if x has only 2 levels, we can write

$$\phi_{211} = \phi \ \phi_X, \ \ \phi_{210} = \phi \ \phi_X \ \phi_K, \ \ \phi_{201} = \phi \ \phi_X \ \phi_J, \ \ \phi_{200} = \phi \ \phi_X \ \phi_J \ \phi_K.$$

A perfect fit requires $w_{xij} = \sum_{k=0}^{1} m_{xijk}\phi_x\phi_j\phi_k$. For $x = 1$ (level 1), we have the following equations:

$$w_{111} \ = \ m_{1111} \ \phi + m_{1110} \ \phi \ \phi_K \qquad (6.8)$$

$$w_{110} \ = \ m_{1101} \ \phi \ \phi_J + m_{1100} \ \phi \ \phi_J \ \phi_K \qquad (6.9)$$

$$w_{101} \ = \ m_{1011} \ \phi + m_{1010} \ \phi \ \phi_K \qquad (6.10)$$

$$w_{100} \ = \ m_{1001} \ \phi \ \phi_J + m_{1000} \ \phi \ \phi_J \ \phi_K \qquad (6.11)$$

We can solve the two linear equations (6.8) and (6.10) for the two unknowns ϕ and ϕ_K. Substituting these solutions into (6.9) and (6.11) yields two equations for ϕ_J and thus ϕ_J is overdetermined. The equation:

$$w_{211} \ = \ m_{2111} \ \phi \ \phi_X + m_{2110} \ \phi \ \phi_X \ \phi_K,$$

then yields a value for ϕ_X. Indeed, each of the w_{2ij} equations yields an equation for ϕ_X.

If x has more than 2 levels, we would write ϕ as ϕ_1^X for the first level, ϕ_X as ϕ_2^X for the second level, etc. In other words, there is one parameter for each

level of x and these parameters can all be identified. Thus, this specification for $h_3(\{p,p\}, \mathbf{y_3^*} \mid x; \boldsymbol{\eta_3})$ is identifiable.

- **LOR * LUR**

This model has $\phi_{xjk} = \phi_{jk}$ as there is no dependence on the covariate x. Thus, there are only four parameters for all the levels of x. As before, a perfect fit requires $w_{xij} = \sum_{k=0}^{1} m_{xijk}\phi_{jk}$ which represents four linear equations in the same four unknowns for each level of x. Hence, $h_3(\{p,p\}, \mathbf{y_3^*} \mid x; \boldsymbol{\eta_3})$ is also identifiable under this parameterization.

In summary, $h_3(\{p,p\}, \mathbf{y_3^*} \mid x; \boldsymbol{\eta_3})$ can be identified if its form is one of the three types considered above.

6.4.2 Identifiability of $h_2(\{p\}, \mathbf{y_2^*} \mid x; \boldsymbol{\eta_2})$

The verification of the identifiability of $h_2(\{p\}, \mathbf{y_2^*} \mid x; \boldsymbol{\eta_2})$ is similar to that for $h_3(\{p,p\}, \mathbf{y_3^*} \mid x; \boldsymbol{\eta_3})$. In addition to the notation from the previous subsection, denote $\dot{v}_{xi} = n_{y_1^*, a, a, x}$ and

$$\gamma_{xij} = \frac{h_2(\{p\}, \mathbf{y_2^*} \mid x; \boldsymbol{\eta_2})}{1 - h_2(\{p\}, \mathbf{y_2^*} \mid x; \boldsymbol{\eta_2})}.$$

The contribution of $h_2(\{p\}, \mathbf{y_2^*} \mid x; \boldsymbol{\eta_2})$ to $L_x(\boldsymbol{\theta}, \boldsymbol{\eta})$ in (6.6) can be expressed as:

$$\sum_{i=0}^{1}\sum_{j=0}^{1}\sum_{k=0}^{1} m_{xijk} \log \rho_{xijk} + \sum_{i=0}^{1}\sum_{j=0}^{1} w_{xij} \log\Big\{\sum_{k=0}^{1} \rho_{xijk}\phi_{xjk}\Big\}$$

$$+ \sum_{i=0}^{1} v_{xi} \log\Big\{\sum_{j=0}^{1}\sum_{k=0}^{1} \rho_{xijk}[1 + \phi_{xjk}]\gamma_{xij}\Big\}. \qquad (6.12)$$

This is identical to the log-likelihood function for a contingency table $\{m_{xijk}\}$ with two supplementary margins, namely $\{w_{xij}\}$ (where k was not observed) and $\{v_{xi}\}$ (where neither of j and k were observed). Therefore, the expected cell counts for m_{xijk}, w_{xij} and v_{xi} are $\mu_{xijk} = \rho_{xijk}(m_{x+++} + w_{x++} + v_{x+})$, $\sum_{k=0}^{1} \mu_{xijk}\phi_{xjk}$ and $\sum_{j=0}^{1}\sum_{k=0}^{1} \mu_{xijk}(1 + \phi_{xjk})\gamma_{xij}$, respectively.

• COV * LUR

This model has $\phi_{xjk} = \phi_{xk}$ and $\gamma_{xij} = \gamma_{xj}$. A perfect fit requires $\mu_{xijk} = m_{xijk}$, $w_{xij} = \sum_{k=0}^{1} \mu_{xijk}\phi_{xk}$ and $v_{xi} = \sum_{j=0}^{1}\sum_{k=0}^{1} \mu_{xijk}(1 + \phi_{xk})\gamma_{xj}$. Hence, we require

$$w_{xij} = \sum_{k=0}^{1} m_{xijk}\phi_{xk}, \tag{6.13}$$

and

$$v_{xi} = \sum_{j=0}^{1}\sum_{k=0}^{1} m_{xijk}(1 + \phi_{xk})\gamma_{xj}. \tag{6.14}$$

For a fixed level of x, (6.13) represents four linear equations in the two unknowns, ϕ_{x0} and ϕ_{x1}, indicating these are overdetermined. With solutions for ϕ_{x0} and ϕ_{x1}, (6.14) represents two linear equations in the two unknowns, γ_{x0} and γ_{x1}. Thus, $h_2(\{p\}, \mathbf{y}_2^* \mid x; \boldsymbol{\eta}_2)$ is identifiable under this model.

• COV + LOR + LUR

In this model, we can represent $\phi_{xjk} = \phi_x \ \phi_j \ \phi_k$ and $\gamma_{xij} = \gamma_x \ \gamma_i \ \gamma_j$. The equations for ϕ_{xjk} are identical to the earlier case for this model and so are identifiable provided the covariate takes on at least two levels. It remains to show that the parameters γ_{xij} can also be identified.

The equations for v_{xi} are

$$v_{xi} = \sum_{j=0}^{1}\sum_{k=0}^{1} m_{xijk}(1 + \phi_{xjk})\gamma_x \ \gamma_i \ \gamma_j$$

$$= \sum_{j=0}^{1} \gamma_x \ \gamma_i \ \gamma_j \ M_{xij}, \tag{6.15}$$

where $M_{xij} = \sum_{k=0}^{1} m_{xijk}(1 + \phi_{xjk})$ is treated as known since solutions for the ϕ's exist. Suppose x has 2 levels. Using the same representation for γ_{xij} as was used

for ϕ_{xjk} earlier, these equations become

$$v_{10} = M_{100} \, \gamma \, \gamma_I \, \gamma_J + M_{101} \, \gamma \, \gamma_I, \tag{6.16}$$

$$v_{11} = M_{110} \, \gamma \, \gamma_J + M_{111} \, \gamma, \tag{6.17}$$

$$v_{20} = M_{200} \, \gamma \, \gamma_X \, \gamma_I \, \gamma_J + M_{201} \, \gamma \, \gamma_X \, \gamma_I, \tag{6.18}$$

$$v_{21} = M_{210} \, \gamma \, \gamma_X \, \gamma_J + M_{211} \, \gamma \, \gamma_X. \tag{6.19}$$

Taking the ratio of (6.18) to (6.16) to eliminate $\gamma \, \gamma_I$ and of (6.19) to (6.17) to eliminate γ leads to two equations in γ_X and γ_J from which γ_X is easily eliminated. This leads to a quadratic equation in γ_J; that is, $A \, \gamma_J^2 + B \, \gamma_J + C = 0$, where

$$A = M_{100}M_{210} - v_{OR}M_{110}M_{200}$$

$$B = (M_{101}M_{210} + M_{211}M_{100}) - v_{OR}(M_{111}M_{200} + M_{201}M_{110})$$

$$C = M_{101}M_{211} - v_{OR}M_{201}M_{111}$$

$$v_{OR} = \frac{v_{10}/v_{11}}{v_{20}/v_{21}}.$$

A perfect fit requires real roots, or $B^2 - 4AC \geq 0$. Thus, $h_2(\{p\}, \mathbf{y}_2^* \mid x; \boldsymbol{\eta}_2)$ is identifiable under this model provided the covariate takes on at least two levels and the equation $B^2 - 4AC \geq 0$ is satisfied.

• LOR * LUR

This model has $\phi_{xjk} = \phi_{jk}$ and $\gamma_{xij} = \gamma_{ij}$. Thus, there are 4 distinct parameters of each type for all the levels of x. These 8 parameters can be identified from the equations for a perfect fit:

$$w_{xij} = \sum_{k=0}^{1} m_{xijk}\phi_{jk}, \tag{6.20}$$

and

$$v_{xi} = \sum_{j=0}^{1}\sum_{k=0}^{1} m_{xijk}(1 + \phi_{jk})\gamma_{ij}. \tag{6.21}$$

58

For each x, (6.20) corresponds to 4 linear equations in the same 4 unknowns as in the verification for $h_3(\{p, p\}, \mathbf{y}_3^* \mid x; \boldsymbol{\eta}_3)$. Substituting these solutions for the ϕ's into (6.21) leads to 2 linear equations in the same 4 unknowns for each x. The 4 γ_{ij} parameters are determined as long as x has 2 or more levels. Hence, $h_2(\{p\}, \mathbf{y}_2^* \mid x; \boldsymbol{\eta}_2)$ is identifiable provided the covariate x has 2 or more levels.

6.4.3 Identifiability of $h_1(\{\ \}, \mathbf{y}_1^* \mid x; \boldsymbol{\eta}_1)$

In addition to the notation from the previous subsection, denote $z_x = n_{a,a,a,x}$ and

$$\delta_x = \frac{h_1(\{\ \}, \mathbf{y}_1^* \mid x; \boldsymbol{\eta}_1)}{1 - h_1(\{\ \}, \mathbf{y}_1^* \mid x; \boldsymbol{\eta}_1)}. \tag{6.22}$$

The contribution of $h_1(\{\ \}, \mathbf{y}_1^* \mid x; \boldsymbol{\eta}_1)$ to $L_x(\boldsymbol{\theta}, \boldsymbol{\eta})$ can then be expressed as:

$$\sum_{y_1^*=0}^{1} \sum_{y_2^*=0}^{1} \sum_{y_3^*=0}^{1} m_{xijk} \log \rho_{xijk} + \sum_{y_1^*=0}^{1} \sum_{y_2^*=0}^{1} w_{xij} \log \left\{ \sum_{k=0}^{1} \rho_{xijk} \phi_{xjk} \right\}$$

$$+ \sum_{i} v_{xi} \log \left\{ \sum_{j=0}^{1} \sum_{k=0}^{1} \rho_{xijk} (1 + \phi_{xjk}) \gamma_{xij} \right\}$$

$$+ z_x \log \left\{ \sum_{y_1^*=0}^{1} \sum_{y_2^*=0}^{1} \sum_{y_3^*=0}^{1} \rho_{xijk} (1 + \phi_{xjk})(1 + \gamma_{xij}) \delta_x \right\}. \tag{6.23}$$

A perfect fit requires

$$w_{xij} = \sum_{k=0}^{1} m_{xijk} \phi_{xjk}, \tag{6.24}$$

$$v_{xi} = \sum_{j=0}^{1} \sum_{k=0}^{1} m_{xijk} (1 + \phi_{xjk}) \gamma_{xij} \tag{6.25}$$

and

$$z_x = \sum_{y_1^*=0}^{1} \sum_{y_2^*=0}^{1} \sum_{y_3^*=0}^{1} m_{xijk} (1 + \phi_{xjk})(1 + \gamma_{xij}) \delta_x. \tag{6.26}$$

● **COV * LUR**

This implies $\phi_{xjk} = \phi_{xk}$ and $\gamma_{xij} = \gamma_{xj}$, while the equations (6.19) for w_{xij} and

(6.20) for v_{xi} are the same as before. The argument in the previous subsections shows that the ϕ_{xk} and γ_{xj} are identified. With these solutions, (6.21) becomes a single equation in one unknown, namely δ_x. Thus, δ_x is also identified. In other words, $h_1(\{\ \}, \mathbf{y_1^*} \mid x; \boldsymbol{\eta}_1)$ is identifiable.

• COV + LOR + LUR

In this model, we can represent $\phi_{xjk} = \phi_x\ \phi_j\ \phi_k$ and $\gamma_{xij} = \gamma_x\ \gamma_i\ \gamma_j$. The argument for the identifiability of the ϕ and γ parameters is identical to that in the previous subsection. Additionally, we have a δ_x parameter for each level of x in (6.21). In other words, there exists a solution for δ_x provided the solutions for the ϕ and γ parameters exist. Hence, $h_1(\{\ \}, \mathbf{y_1^*} \mid x; \boldsymbol{\eta}_1)$ is identifiable.

• LOR * LUR

This model implies $\phi_{xjk} = \phi_{jk}$ (4 parameters for all levels of x), $\gamma_{xij} = \gamma_{ij}$ (4 parameters for all levels of x) and $\delta_x = \delta$ (1 parameter for all levels of x). The argument for the identifiability of the ϕ and γ parameters is again identical to that in the previous subsection. The additional parameter, δ, can be determined from (6.21) provided solutions exist for the ϕ and γ parameters. Hence, $h_1(\{\ \}, \mathbf{y_1^*} \mid x; \boldsymbol{\eta}_1)$ is identifiable.

Thus, we have shown that, when coupled with a saturated outcome model, the parameters in the drop-out models of the three forms suggested by Baker (1995) are identifiable. Notice that we only consider the case where the covariates are categorical. In the next chapter, we analyze our annual data set with the models mentioned in the previous chapters.

Chapter 7

Application to the Data

7.1 Introduction

In this chapter, we implement the selection model approach for our annual MS data as described in Chapter 2. Recall our study questions of interest are:

- to investigate the most appropriate form of drop-out model for our annual data (in particular, to explore whether the data provide evidence of informative drop-out);

- to assess the sensitivity of inferences concerning the treatment effects (and other covariate effects) to the form of drop-out model employed;

- to explore the influence of baseline covariates.

Recall that the basic idea of a selection model is to factor the joint distribution for the response variables ($\mathbf{Y}$) and the indicator variables corresponding to whether or not the response variables are observed ($\mathbf{R}$) as follows:

$$f(\mathbf{Y}, \mathbf{R}) \;\; = \cdot\, f(\mathbf{R} \mid \mathbf{Y}) f(\mathbf{Y}). \tag{7.1}$$

Thus, the selection model approach involves the specification of a model for the outcomes, $f(\mathbf{Y})$, and for the drop-out pattern conditional on the outcomes, $f(\mathbf{R} \mid \mathbf{Y})$.

The outline of this chapter is as follows: Section 7.2 considers a simple structure for Baker's selection model where only treatment group and time are included as covariates in the outcome model. This outcome model is coupled with a LOR+LUR type of drop-out model. In Section 7.3, we consider three more general model specifications for the drop-out process in conjunction with the same outcome model: COV * LUR, COV + LOR + LUR, and LOR * LUR. We extend this simple model by incorporating other baseline covariates described in Section 2.2.3 into the outcome model in Section 7.4. The latter two sections can be viewed as further explorations of Baker's selection model. We conclude the chapter with a brief discussion of the use of the Liu *et al.* transition model for the outcome model.

7.2 Baker's Selection Model: With Only Treatment Groups and Time as Covariates

As described in Chapter 4, Baker (1995) suggested specifying the outcome model in terms of marginal and association models. The drop-out process is modelled using a time-dependent causal model assuming the non-response does not depend on future events.

♦ Repeated Binary Outcomes with Informative Drop-out ♦

⋄ *Outcome Model*

The outcome model $f^*(y_1^*, y_2^*, y_3^* \mid x; \boldsymbol{\theta})$ is expressed in terms of marginal and association models. As is apparent from Figure 2.3, the proportion of patients with exacerbations seems to vary across the treatment groups and with time, so the marginal model employed is

$$\text{logit}\{g_t(\boldsymbol{x}; \boldsymbol{\beta})\} = \beta_0 + \beta_1 LD + \beta_2 HD + \beta_3 t, \tag{7.2}$$

where $t = 1, 2, 3$, and LD and HD are indicator variables to represent the treatment groups. For patients in the LD group, $LD = 1$ and $HD = 0$. Similarly, $LD = 0$ and

$HD = 1$ if patients belong to the HD group. For patients in the PL group, both LD and HD take on value 0.

We propose modelling the 2-way and 3-way associations with different intercept parameters to describe different degrees of association. We further assume the association among the responses is related to the treatment arms. For simplicity, these treatment effects are taken to be the same for all associations.

- **Models for 2-way Association:**

$$\text{logit}\{g_{st}(\boldsymbol{x}; \boldsymbol{\alpha}_{st})\} = \alpha_{st} + \alpha_1 LD + \alpha_2 HD \qquad (7.3)$$

where $st = \{12, 13, 23\}$.

- **Model for 3-way Association:**

$$\text{logit}\{g_{123}(\boldsymbol{x}; \boldsymbol{\alpha}_{123})\} = \alpha_{123} + \alpha_1 LD + \alpha_2 HD. \qquad (7.4)$$

Both the marginal and association models remain the same throughout the analyses in this section regardless of the assumption on the drop-out mechanism.

The adequacy of this non-saturated outcome model for our data has been confirmed by comparing it to various more general models. This information is presented in the next subsection.

◇ **Drop-out Model**

We model the drop-out process using time-dependent causal models assuming the non-response does not depend on future events. We allow different regression parameters for the logistic regressions specifying the different conditional probabilities of absence, $h_t(\mathbf{r}_{t-1}, \mathbf{y}_t^* \mid \boldsymbol{x}, \boldsymbol{\eta}_t)$; see (4.12). To simplify the notation, we introduce two subscripts for these regression parameters:

$$\text{logit}\{h_3(\mathbf{r}_2 = \{\text{p}, \text{p}\}, \mathbf{y}_3^* \mid \boldsymbol{x}, \boldsymbol{\eta}_3)\} = \eta_{03} + \eta_{13} y_2^* + \eta_{23} y_3^*$$
$$\text{logit}\{h_2(\mathbf{r}_1 = \{\text{p}\}, \mathbf{y}_2^* \mid \boldsymbol{x}, \boldsymbol{\eta}_2)\} = \eta_{02} + \eta_{12} y_1^* + \eta_{22} y_2^* \qquad (7.5)$$

63

Table 7.1: Drop-out Models under Different Drop-out Mechanisms: $\sqrt{}$ denotes inclusion of a parameter and η_i denotes parameters which are restricted to be equal

Drop-out Mechanism	Model	Parameter						
		η_{03}	η_{13}	η_{23}	η_{02}	η_{12}	η_{22}	η_{01}
ID	1	$\sqrt{}$	$\sqrt{}$	$\sqrt{}$	$\sqrt{}$	$\sqrt{}$	$\sqrt{}$	$\sqrt{}$
	2	$\sqrt{}$	η_1	η_2	$\sqrt{}$	η_1	η_2	$\sqrt{}$
	3	η_0	η_1	η_2	η_0	η_1	η_2	η_0
	4	$\sqrt{}$	-	$\sqrt{}$	$\sqrt{}$	-	$\sqrt{}$	$\sqrt{}$
	5	$\sqrt{}$	-	η_2	$\sqrt{}$	-	η_2	$\sqrt{}$
	6	η_0	-	η_2	η_0	-	η_2	η_0
RD	1	$\sqrt{}$	$\sqrt{}$	-	$\sqrt{}$	$\sqrt{}$	-	$\sqrt{}$
	2	$\sqrt{}$	η_1	-	$\sqrt{}$	η_1	-	$\sqrt{}$
	3	η_0	η_1	-	η_0	η_1	-	η_0
CRD	1	$\sqrt{}$	-	-	$\sqrt{}$	-	-	$\sqrt{}$
	2	η_0	-	-	η_0	-	-	η_0

where the first subscript indexes the specific parameter in the model, while the second subscript indexes the year the drop-out occurred. According to Baker (1995), if the conditional non-response probability in the first year, $\Pr(R_1 = a \mid y_1^*, x) = h_1(\mathbf{r_1} = \{\ \}, y_1^* \mid x, \boldsymbol{\eta}_1)$, depends only on the covariates, then the non-ignorable non-response models under consideration will be identifiable. In our case, the model for $h_1(\mathbf{r_1} = \{\ \}, y_1^* \mid x, \boldsymbol{\eta}_1)$ becomes:

$$\text{logit}\{h_1(\mathbf{r_1} = \{\ \}, y_1^* \mid x, \boldsymbol{\eta}_1)\} = \eta_{01}. \tag{7.6}$$

These drop-out models belong to Baker's LOR + LUR class of models. For simplicity, we have taken the drop-out mechanism to be independent of the available covariates. We relax this assumption in Section 7.3.

To explore the adequacy of simpler models, we consider five other model specifications which are obtained by letting certain parameters be equal or be equal to zero. The ID models to be considered are summarized in the first six rows of Table 7.1.

♦ Repeated Binary Outcomes with Ignorable Drop-out ♦

To investigate the types of drop-out in our annual data, we also fit the data to models under ignorable drop-out assumptions, i.e. with RD and CRD models.

◇ *Random Drop-out*

We consider the three RD models summarized in Table 7.1. Modifying an ID model by setting the parameters associated with the unobserved response to zero leads to an RD model. For instance, RD1 (Model 1 under RD) is obtained by setting $\eta_{23} = \eta_{22} = 0$ in ID1 (Model 1 under ID). RD2 and RD3 are similarly obtained from ID2 and ID3.

◇ *Completely Random Drop-out*

The two CRD models considered in Table 7.1 are obtained by simplifying the RD models. Under CRD, the drop-out mechanism is independent of the measurement process. Thus, CRD1 (obtained by setting $\eta_{13} = \eta_{12} = 0$ in RD1, or $\eta_1 = 0$ in RD2) and CRD2 (obtained by setting $\eta_1 = 0$ in RD3) each consist of only intercept parameters.

For both the RD and CRD cases, we also have the opportunity to examine the sensitivity of the covariate effects (treatment and time) under different forms of the RD or CRD models.

These outcome and drop-out models can be assembled into explicit expressions for the logarithm of the likelihood (see (4.16)). The maximun likelihood estimates of the parameters in these models are obtained by minimizing the negative log-likelihood function using a quasi-newton (QN) minimization procedure. This procedure is briefly described in the following subsection. The corresponding results are summarized in next subsection.

7.2.1   The Quasi-Newton (QN) Algorithm

The QN algorithm used to maximize the log-likelihood is a variable metric algorithm. All variable metric methods seek to minimize a certain function $S(\boldsymbol{\theta})$ (in our case, $S(\boldsymbol{\theta})$ is the negative log-likelihood function) of p parameters by means of a sequence of basic iterative steps

$$\boldsymbol{\theta}' = \boldsymbol{\theta} - k\mathbf{B}\boldsymbol{g} \tag{7.7}$$

where $\boldsymbol{g}$ is the gradient of the function S, $\mathbf{B}$ is a matrix defining a transformation of the gradient and k is a step length.

Consider the set of nonlinear equations formed by the gradient at a minimum

$$\boldsymbol{g}(\boldsymbol{\theta}') = 0. \tag{7.8}$$

As in the one-dimensional root-finding problem, one can use a linear approximation from the current $\boldsymbol{\theta}$, that is

$$\boldsymbol{g}(\boldsymbol{\theta}') \approx \boldsymbol{g}(\boldsymbol{\theta}) + \mathbf{H}(\boldsymbol{\theta})(\boldsymbol{\theta}' - \boldsymbol{\theta}) \tag{7.9}$$

where $\mathbf{H}(\boldsymbol{\theta})$ is the *Hessian* matrix (the matrix of second derivatives of the function S). For convex functions, $\mathbf{H}$ will be positive definite. From (7.8), (7.9) becomes

$$\boldsymbol{\theta}' \approx \boldsymbol{\theta} - \mathbf{H}^{-1}(\boldsymbol{\theta})\boldsymbol{g}(\boldsymbol{\theta}) \tag{7.10}$$

which is Newton's method for a function of p parameters. This is equivalent to (7.7) with $\mathbf{B} = \mathbf{H}^{-1}$ and $k = 1$.

Newton's method is generally preferable if second derivatives can be analytically computed. But the implementation of Newton's method may induce errors when closed form expressions for the second derivatives do not exist as it involves composing subroutines for evaluating p first derivatives, p^2 second derivatives and a matrix inversion. For these reasons, Newton's method does not recommend itself for some problems.

If $\mathbf{H}^{-1}$ could be approximated directly from the first derivative information available at each step of the iteration, this would save a great deal of work in computing both the matrix $\mathbf{H}$ and its inverse. This is precisely the role of the matrix $\mathbf{B}$ in the iteration defined by (7.7). The transformed gradients in the matrix $\mathbf{B}$ are used to generate linearly independent search directions; equivalently, these search directions are conjugate to each other with respect to $\mathbf{H}$. Further, the step parameter k is rarely fixed; its value is usually determined by some form of a linear search. In particular, the role of k is to allow a search for values of $\boldsymbol{\theta}'$ at which the function value is reduced, i.e. $S(\boldsymbol{\theta}') < S(\boldsymbol{\theta})$. Since the second derivatives required in Newton's method are approximated in the iteration (7.7), this algorithm is known as a *quasi*-Newton method.

We employ the QN algorithm suggested in Nash (1979). It involves specific choices of the formula for updating the matrix $\mathbf{B}$ and of the linear search procedure for obtaining the updated values of $\boldsymbol{\theta}'$. An 'acceptable point' search procedure suggested by Fletcher (1970) and a matrix-updating formula for $\mathbf{B}$ due to Broyden (1970a, 1970b), Fletcher (1970) and Shanno (1970) are employed. Generally speaking, the algorithm first goes through a linear search to find one value for $\boldsymbol{\theta}$ which gives a smaller function value than that at the previous value for $\boldsymbol{\theta}$. The approximation to the Hessian matrix is then updated accordingly. The algorithm stops when all the parameter values on consecutive iterations are sufficiently close. For our purposes, the absolute difference between the parameter values of consecutive iterations must be smaller than 10^{-7}. A detailed outline of this algorithm can be found in Chapter 15 of Nash (1979).

Note that in this version of the QN algorithm, the matrix $\mathbf{B}$ is initialized as a unit matrix. This simple choice nevertheless has the advantage of generating the steepest descent direction (Nash, 1970). To ensure rounding errors which occur in updating the matrix $\mathbf{B}$ and forming the search directions, t, through the equation

$$t = \boldsymbol{\theta}' - \boldsymbol{\theta} = -k\mathbf{B}g$$

have not accidentally given a direction in which the function S cannot be reduced, a reset of **B** to a unit matrix is suggested in any of the following cases:

(i) $t^{\mathrm{T}}g \geq 0$; that is, the direction of the search is 'uphill';

(ii) $\theta' = \theta$; that is, no change is made in the parameters by the linear search along t;

(iii) $t^{\mathrm{T}}\{g(\theta') - g(\theta)\} \leq 0$; that is, an updating contrary to the objective of the method to reduce S along t ($t^{\mathrm{T}}g(\theta')$ is expected to be greater (less negative) than $t^{\mathrm{T}}g(\theta)$), indicating a danger that matrix **B** may no longer be positive definite.

If either (i) or (ii) occurs during the first step after **B** has been set to the unit matrix, the algorithm is taken to have converged.

All results described in this thesis are obtained using this QN algorithm implemented in C. The results for the models described in the beginning of this section are discussed in the next subsection.

7.2.2 Results

• **Adequacy of the Outcome Model**

To verify the adequacy of our reduced (non-saturated) outcome model, we consider four more general outcome model specifications. These outcome models are

1. **Saturated**: a saturated marginal model (9 distinct parameters) and a saturated association model of the same form as (7.3) and (7.4) but with regression parameters that differ for each of the 2-way and 3-way association models (12 distinct parameters);

2. **Semi-saturated I**: a saturated marginal model and a reduced association model with common treatment effects in the 2-way associations (8 distinct parameters);

Table 7.2: Negative Log-likelihood Values for Five Outcome Model Specifications

Outcome Model	Negative Log-likelihood	Number of Parameters
Saturated	928.923	28
Semi-saturated I	930.450	24
Semi-saturated II	931.680	22
Semi-saturated III	930.304	23
Reduced	933.407	17

3. **Semi-saturated II**: a saturated marginal model and a reduced association model with common treatment effects for all associations (6 distinct parameters). Note that this reduced association model is exactly (7.3) and (7.4);

4. **Semi-saturated III**: a reduced marginal model assuming linearity in time (4 parameters) and a saturated association model (12 distinct parameters). Note that this reduced marginal model is exactly (7.2).

The negative log-likelihood values presented in Table 7.2 correspond to these outcome models coupled with the drop-out model (7.5).

The likelihood ratio test (LRT) indicates the reduction from the fully saturated outcome model to semi-saturated I is reasonable (LR statistic $= 3.05$ on degrees of freedom (df) $= 4$; p-$value = 0.55$). To examine whether the treatment effects in the association model can be taken to be common across all associations, we compare semi-saturated I to semi-saturated II. The LR statistic of 2.46 (df $= 2$; p-$value = 0.29$), indicates the reduction is permissible. The result based on a direct comparison between the saturated and semi-saturated II models also agrees (LR statistic $= 5.51$ on df $= 6$; p-$value = 0.48$). This indicates that an association model with common treatment effects for all associations is reasonable for our data set. The further reduction to our reduced outcome model is also allowed (LR statistic $= 3.45$ on df $= 5$; p-$value = 0.63$).

As our primary focus is on the marginal model, a more interesting comparison is between the semi-saturated III and saturated outcome models. In the context of a saturated association model, this provides an assessment of whether the reduced marginal model (7.2), which incorporates additive treatment effects and a linear pattern over time for the log odds of having exacerbations, is reasonable. The LRT allows this reduction ($p\text{-}value = 0.74$). As should be expected from the earlier comparisons, the semi-saturated III model can be further reduced to our non-saturated model ($p\text{-}value = 0.40$).

Both sequences of model reductions lead to the same conclusion: the reduction to the model presented in the beginning of this section is permitted. This reduced model also provides an adequate fit to our data. The usual goodness-of-fit statistics based on the 15 different possible patterns of binary responses for each treatment arm lead to $G^2 = 24.65$ and $X^2 = 22.80$ on 25 degrees of freedom ($p\text{-}values = 0.48$ and 0.59 respectively). Thus, we can proceed confidently with further work using this reduced model as a starting point in the investigations.

- **Informative Drop-out (ID)**

The detailed results corresponding to the six ID models described in Table 7.1 can be found in Appendix B: Tables B.1 to B.6. These tables include the sets of starting values (SV) used, and the maximum likelihood estimates for the parameters (Est), the corresponding standard errors (SE), and the negative log-likelihood computed at the MLE which are all provided as part of the output from the QN minimization procedure. The number of iterations needed to achieve convergence is also cited in the tables.

In each of these tables, regardless of the starting values in the QN procedure, the corresponding negative log-likelihoods computed at the parameter estimates (at convergence) are the same (at least up to the 4 significant decimal digits displayed). However, in Tables B.1, B.2, B.4 and B.5, not all the reported MLEs are the same

(see especially for parameters η_{03} and η_{23} in Tables B.1 and B.4, and parameters η_{03}, η_2 and η_{02} in Tables B.2 and B.5). Also, in these four tables, the SEs for the estimates vary quite a bit across different sets of starting values. This phenomenon might be due to how the Hessian matrix is approximated in the minimization procedure. As mentioned earlier, the Hessian matrix is approximated based on the search directions for the parameter estimates obtained in each successive iteration. To illustrate, consider starting value Sets #1 and #4 in Table B.1. For Set #1, the estimated Hessian matrix was reset to a unit matrix at the 56th iteration due to it not being a positive definite matrix. The final SEs as displayed thus depend on both the parameter estimates at convergence and the corresponding search directions at the subsequent iterations, i.e. the 57th iteration until convergence was achieved (at the 71st iteration). The estimated Hessian matrix for Set #4, however, was reset to a unit matrix three times during the process of minimization (at the 4th, 9th, and 75th iterations), with convergence established at the 91st iteration. Since the process of minimization for the two sets was quite different, this might be the reason why the estimated SEs differ considerably from one set of starting values to another.

The substantially different values of the estimates obtained with different sets of starting values for some of the parameters in models ID1, ID2, ID4 and ID5 indicates a more fundamental difficulty. Consider the results for model ID1, for example. Table B.1 shows the parameter η_{03} is always estimated as being large negative, while η_{23} is always estimated as large positive. Furthermore, for all four sets of starting values, the sum of these two parameter estimates equals a constant value, -1.548. This suggests the maximum likelihood estimates for this data set satisfy the constraint $\hat{\eta}_{03} + \hat{\eta}_{23} = -1.548$, with the MLE occurring on the boundary of the parameter space ($\hat{\eta}_{03} = -\infty$ or $\hat{\eta}_{23} = \infty$). Recall that the non-response probability for the third observation is modelled as a logistic regression on last observed outcome (y_2^*) and last unobserved outcome (y_3^*); see (7.5). When $\hat{\eta}_{03} = -\infty$, $\hat{\eta}_{03} + \hat{\eta}_{23} = -1.548$ and $\hat{\eta}_{13}$ is finite, the probability that the third observation

71

is missing is estimated to be zero if the history is either $\{y_2^* = 0, y_3^* = 0\}$ or $\{y_2^* = 1, y_3^* = 0\}$, but non-zero for the remaining two histories.

The same phenomenon is observed for model ID4 in Table B.4, but with $\hat{\eta}_{03} + \hat{\eta}_{23} = -1.165$. This phenomenon is also apparent for models ID2 (Table B.2) and ID5 (Table B.5), but manifests itself in a slightly different fashion. Here, the parameters η_{03} and η_{02} are always estimated as being large negative, while the parameter η_2 is always estimated as large positive. However, the sum of $\hat{\eta}_{03}$ and $\hat{\eta}_2$ always equals a constant, and the sum of $\hat{\eta}_{02}$ and $\hat{\eta}_2$ equals another constant. The pair of constants differ from model ID2 to model ID5. Thus under models ID2 and ID5, the probabilities for the second and third observations to be missing are estimated to be zero when the past observations are either $\{y_1^* = 0, y_2^* = 0\}$ or $\{y_1^* = 1, y_2^* = 0\}$, and when the history is either $\{y_2^* = 0, y_3^* = 0\}$ or $\{y_2^* = 1, y_3^* = 0\}$, respectively.

In the next few paragraphs, we discuss the issue of boundary solutions for model ID1 in greater detail. The corresponding discussion for models ID2, ID4, and ID5 is omitted as the details are essentially identical to model ID1. But the results for these three models evaluated at the boundary solutions are also presented.

- **Discussion of Boundary Solutions**

Consider model ID1. The estimates obtained for η_{03} and η_{23} displayed in Table B.1 vary across different starting values, but in each case $\hat{\eta}_{03} + \hat{\eta}_{23} = -1.548$. Further the negative log-likelihood remains the same up to the four decimal digits displayed. We believe that the MLE is located on the boundary of the parameter space. To confirm this conjecture, we first use a graphical visualization of the negative log-likelihood function incorporating the special feature (e.g. η_{03} is estimated with large negative value, while η_{23} is estimated with large positive values, and the sum of the two is always the same) observed in Table B.1.

Figure 7.1: A Two-Dimensional Profile Log-likelihood Surface for Model ID1
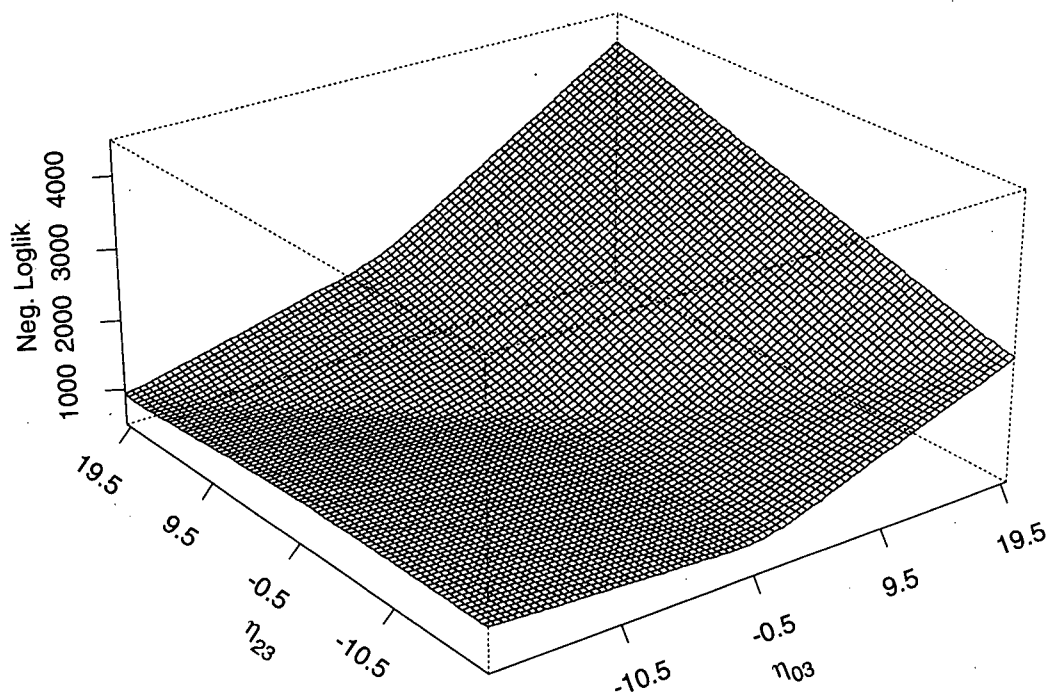
Figure 7.1 is a graphical representation of the profile log-likelihood surface for the parameters η_{03} and η_{23} in model ID1. This three-dimensional plot is produced by maximizing the log-likelihood over all parameters except η_{03} and η_{23}. For fixed values of η_{03} and η_{23}, we apply the QN minimization procedure to the negative log-likelihood function. This log-likelihood value is then plotted against these values for η_{03} and η_{23} using the S-PLUS function "persp". We chose the values for η_{03} and η_{23} to be a sequence of numbers between -20 and 20 with increment size of 0.5. This yields a 81 by 81 grid of log-likelihood values. Notice that there seems to be a steady, but very shallow, decrease in this surface along a line (where $\eta_{03} + \eta_{23} = -1.548$) in the grid where η_{03} and η_{23} take on values ranging from -20.0 to -0.5, and from 0.0 to 20.0, respectively. This seems to agree with the results presented in Table B.1.

We also computed the log-likelihood on the boundary of the parameter space to check that the log-likelihood values obtained in Table B.1 are what one would obtain at the suggested point on the boundary. Because the parameter estimates appear to satisfy the constraint $\hat{\eta}_{03} + \hat{\eta}_{23} = -1.548$, it is useful to re-parameterize in terms of η_{03} and $\eta_{23} = -\eta_{03} + \Delta$, where Δ is a finite-valued parameter. As η_{03} approaches $-\infty$, the log-likelihood is a function of the remaining parameters and Δ. For the probability of non-response, $\Pr(R_3 = a \mid \{p, p\}, \mathbf{y}_3^*, x)$, we substitute the values presented in Table 7.3 to obtain the reduced log-likelihood function.

Applying the QN minimization routine to this reduced negative log-likelihood function yields the results summarized in Table 7.4. The estimates for the model parameters are essentially the same as those presented in Table B.1 and the log-likelihood value also agrees. Thus, both Figure 7.1 and this computation of the log-likelihood at the indicated boundary point seem to support our conjecture that the parameter estimates for model ID1 occur on the boundary of the parameter space. The same values of the estimates reported in Table 7.4 were obtained with different choices of starting values and these minimizations required many fewer iterations than those presented in Table B.1. Further, the estimated Hessian matrix

74

was never reset to a unit matrix during these minimizations.

Notice that the standard errors for $\hat{\eta}_{02}$ and $\hat{\eta}_{22}$ in Table 7.4 are relatively large. One might suspect this reflects a potential boundary solution phenomenon for the reduced log-likelihood even though these estimates did not vary with the sets of starting values chosen (see also Table B.1). Perhaps these large standard errors are simply indicating that our data set does not contain sufficient information to obtain precise estimates for these parameters. We explored this further graphically.

Figure 7.2 shows the profile log-likelihood surface for the parameters η_{02} and η_{22} of the reduced model ID1. The values for η_{02} and η_{22} were chosen to be a sequence of numbers between -20 and 20 with increment size of 0.5. The plot is not very informative in terms of revealing the existence of optimal solutions. The rotating option in "persp" allowed us to view Figure 7.2 from different directions and convinced us of the existence of optimal solutions in the interior of the parameter space for this reduced log-likelihood function. For further assurance, we also calculated the log-likelihood values at various points in the neighbourhood of the suggested estimates for η_{02} and η_{22}; these values are all larger than 933.407. Thus we are certain that this situation does not indicate a boundary solution, but simply indicates a lack of information in the data to precisely estimate these parameters.

One can easily show, in a similar fashion, that the parameter estimates for models ID2, ID4 and ID5 also occur on the boundary of the parameter space. The corresponding results for these three models computed at the suggested boundary points are presented in Tables 7.5, 7.6, and 7.7. Note that the parameter estimates in the outcome models for ID2 and ID5 are the same. With the imposed boundary constraints, the log-likelihood functions can be expressed as the sum of a function of the parameters in the outcome model and a function of the parameters in the drop-out model. Hence, the parameters in the outcome and drop-out models can be maximized separately. Compared to the minimizations summarized in Tables B.2, B.4 and B.5, the convergence for these three cases is achieved with many fewer

Figure 7.2: A Two-Dimensional Profile Log-likelihood Surface for Model ID1 with Boundary Constraint $\eta_{03} \to -\infty$ and $\eta_{03} + \eta_{23} = \Delta$
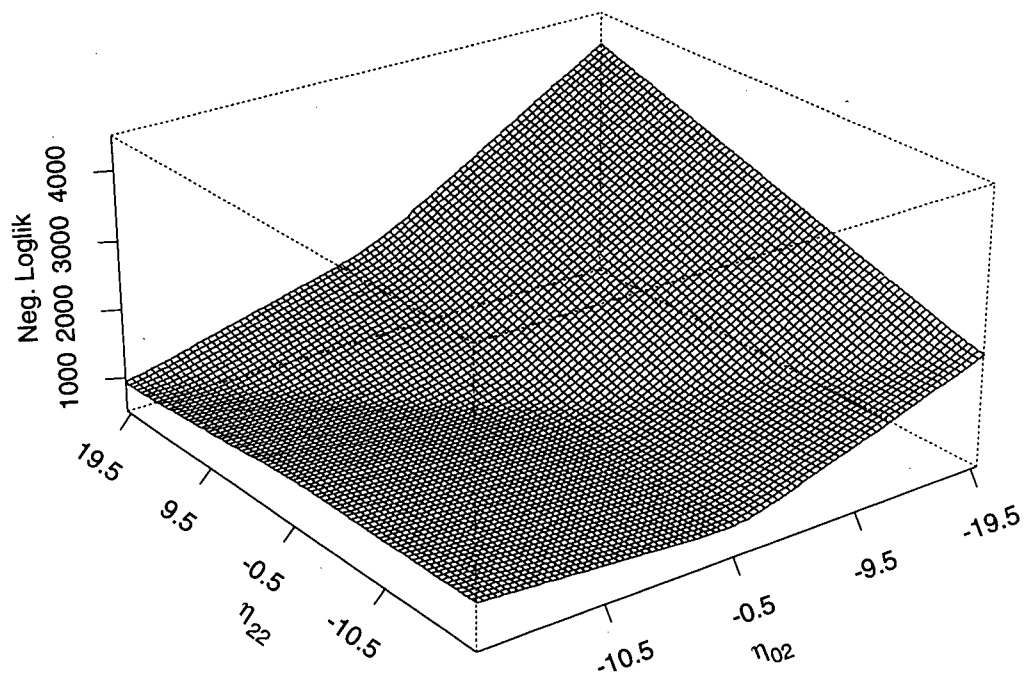
Table 7.3: Non-response Probability for the Third Response Using Model ID1 with $\eta_{03} \to -\infty$ and $\eta_{03} + \eta_{23} = \Delta$

y_2^*	y_3^*	$\mathrm{logit}\{\Pr(R_3 = \mathrm{a} \mid \{\mathrm{p},\mathrm{p}\}, \mathbf{y}_3^*, x)\}$
0	0	$-\infty$
0	1	Δ
1	0	$-\infty$
1	1	$\eta_{13} + \Delta$

iterations. Further, the estimated Hessian matrices were never reset to a unit matrix during the course of minimization. As expected, the standard errors for η_{02} and η_{22} in Table 7.6 behave similarly as in Table 7.4. This is again verified (by the same approach) not to reflect a boundary solution. On the other hand, the standard errors for all the estimates in Tables 7.5 and 7.7 look quite reasonable.

This feature of boundary solutions does not appear in models ID3 and ID6. For both models, the solutions obtained by the QN minimization are located in the interior of the parameter space. Different sets of starting values lead to the same parameter estimates and similar standard errors for the estimates, as shown in Tables B.3 and B.6. Even though the Hessian matrix was never reset to unity during the minimization process, the small discrepancy in the estimated SEs is expected due to the way the Hessian matrix is approximated. For these two models, the convergence is achieved between 17 and 21 iterations, which is much faster than for the models where the solutions are located on the boundary of the parameter space. This concludes the discussion concerning the existence of boundary solutions.

- **Results for the ID Models**

Now we examine if the treatment effects are sensitive to the form of the informative drop-out model based on the results presented in Tables 7.4, 7.5, B.3, 7.6, 7.7 and B.6. Our primary focus is on the treatment effects in the marginal model for the exacerbation rates even though treatment effects are also incorporated in

Table 7.4: Results for Model ID1 Evaluated on the Boundary: $\eta_{03} \to -\infty$ and $\eta_{03} + \eta_{23} = \Delta$

Parameter	Estimate	SE
β_0	0.876	0.206
β_1 (LD)	-0.028	0.200
β_2 (HD)	-0.489	0.195
β_3 (time)	-0.122	0.074
α_{12}	-0.020	0.170
α_{13}	-0.031	0.168
α_{23}	-0.136	0.183
α_{123}	-0.534	0.187
α_1	-0.113	0.213
α_2	-0.657	0.221
η_{13}	0.558	0.409
Δ	-1.548	0.347
η_{02}	-3.360	2.218
η_{12}	0.140	0.417
η_{22}	1.860	2.615
η_{01}	-2.089	0.167
Neg. Loglik	933.407 (# Iter = 25)	

Table 7.5: Results for Model ID2 Evaluated on the Boundary: $\eta_{03} \rightarrow -\infty$, $\eta_{02} \rightarrow -\infty$, $\eta_{03} + \eta_2 = \Delta_1$ and $\eta_{02} + \eta_2 = \Delta_2$

Parameter	Estimate	SE
β_0	0.886	0.204
β_1 (LD)	−0.017	0.195
β_2 (HD)	−0.484	0.194
β_3 (time)	−0.118	0.074
α_{12}	−0.004	0.163
α_{13}	−0.010	0.161
α_{23}	−0.111	0.173
α_{123}	−0.511	0.177
α_1	−0.103	0.208
α_2	−0.649	0.217
η_1	0.286	0.275
Δ_1	−1.356	0.264
Δ_2	−1.499	0.258
η_{01}	−2.089	0.164
Neg. Loglik	933.922 (# Iter = 20)	

Table 7.6: Results for Model ID4 Evaluated on the Boundary: $\eta_{03} \rightarrow -\infty$ and $\eta_{03} + \eta_{23} = \Delta$

Parameter	Estimate	SE
β_0	0.880	0.189
β_1 (LD)	−0.024	0.190
β_2 (HD)	−0.487	0.187
β_3 (time)	−0.120	0.071
α_{12}	−0.013	0.145
α_{13}	−0.022	0.137
α_{23}	−0.126	0.151
α_{123}	−0.524	0.152
α_1	−0.109	0.202
α_2	−0.654	0.214
Δ	−1.165	0.181
η_{02}	−3.819	3.002
η_{22}	2.464	3.217
η_{01}	−2.089	0.165
Neg. Loglik	934.432 (# Iter = 27)	

Table 7.7: Results for Model ID5 Evaluated on the Boundary: $\eta_{03} \to -\infty$, $\eta_{02} \to -\infty$, $\eta_{03} + \eta_2 = \Delta_1$ and $\eta_{02} + \eta_2 = \Delta_2$

Parameter	Estimate	SE
β_0	0.886	0.202
β_1 (LD)	-0.017	0.198
β_2 (HD)	-0.484	0.192
β_3 (time)	-0.118	0.073
α_{12}	-0.004	0.162
α_{13}	-0.010	0.160
α_{23}	-0.111	0.172
α_{123}	-0.511	0.176
α_1	-0.103	0.211
α_2	-0.649	0.217
Δ_1	-1.165	0.182
Δ_2	-1.293	0.168
η_{01}	-2.089	0.165
Neg. Loglik	934.473 (# Iter = 21)	

the association model. The structure of the ID drop-out model does not change the conclusions about the treatment effects in the marginal model. All six models conclude that the exacerbation rates in the LD and PL groups at any given time are not significantly different (approximate two-sided $p\text{-}value > 0.62$ based on $\hat{\beta}_1$ in each case). On the other hand, the exacerbation rate in the HD group is estimated to be significantly lower than in the PL group at all time points (two-sided $p\text{-}value < 0.02$ based on $\hat{\beta}_2$ in each case). The odds of experiencing exacerbations in the PL group are roughly 1.6 times higher than in the HD group. There is a weak suggestion of a linear decrease with time in the log odds of experiencing exacerbations under models ID1, ID2, ID4 and ID5 (two-sided $p\text{-}value \approx 0.10$ in each model), but the estimates of β_3 in both ID3 and ID6 provide a strong indication of a linear decrease over time (two-sided $p\text{-}values < 0.008$).

The conclusions regarding the treatment effects in the association model are similar. All six models indicate that the odds of having exacerbations at two

occasions or at all three occasions in the study are not significantly different between the LD and PL groups (two-sided p-values ≥ 0.38 based on $\hat{\alpha}_1$). But the models suggest that the odds in the HD group are significantly smaller than in the PL group (two-sided p-values ≤ 0.004).

Under models ID1, ID2, ID4 and ID5, the estimates of the intercept parameters, α_{12} and α_{13}, are fairly similar while $\hat{\alpha}_{23}$ is slightly more negative. As would be expected, the estimate for the intercept in the 3-way association model is most negative. The situation is similar for models ID3 and ID6, although the estimates are slightly more negative. Note that the estimates for α_{12}, α_{13} and α_{23} are not very different, suggesting a possibility of a common intercept parameter for all the 2-way association models. However, the reduction to a model with the same intercept parameter for all 2- and 3-way association models may not seem reasonable since the estimate for α_{123} is always quite different from the others. Further, we could explore explicitly whether the responses are positively or negatively associated by comparing the joint probabilities of the responses with those obtained under the independence assumption. If the joint probabilities are larger than the product of the marginal probabilities, then there is some positive dependence among the responses; otherwise, the responses are negatively correlated. See Chapter 8 for more details.

We now consider selecting a parsimonious ID model to describe our data. Table 7.8 summarizes the negative log-likelihood and available degrees of freedom for all models listed in Table 7.1. Based on the LRT, the reduction from model ID1 to ID2 is permissible (p-value $= 0.60$), indicating the dependence on the previous and current observations is similar at time points 2 and 3. Using model ID2 as the base model and comparing to model ID3 examines whether the odds of dropping out (for the same history) change over time; that is, the hypothesis is $\eta_{03} = \eta_{02} = \eta_{01} = \eta_0$. But the LRT statistic indicates this reduction is not reasonable (p-value $= 0.03$). Note that one can also assess the reduction from model ID1 directly to ID3, although this assessment is not as sensitive as the comparison between models ID2 and ID3.

Table 7.8: Negative Log-likelihood Values for Models in Table 7.1

Drop-out Mechanism	Model	Negative Log-likelihood	Degrees of Freedom (df)
ID	1	933.407	25
	2	933.922	27
	3	937.349	29
	4	934.432	27
	5	934.473	28
	6	938.464	30
RD	1	936.833	27
	2	937.250	28
	3	937.457	30
CRD	1	940.422	29
	2	941.040	31

The associated *p-value* is 0.096, indicating only fairly weak evidence against reducing from model ID1 to ID3. Thus, based on the more sensitive assessment, we conclude that model ID2 is the simplest permissible ID model among these three.

To consider further model reductions, we next compare model ID2 to ID5. The LRT statistic suggests this reduction is reasonable (*p-value* = 0.29). The overall reduction from model ID1 to ID5 also agrees (*p-value* = 0.54). In model ID5, the drop-out probabilities do not depend on the last observed response, only on the last unobserved response. The further reduction from model ID5 to ID6 is not allowed (*p-value* = 0.02).

We conclude that model ID5 is the simplest of these six informative drop-out models that can be used to describe our annual data set. The two reduced models, ID2 and ID5, both fit the data adequately. For model ID2, $G^2 = 25.94$ and $X^2 = 23.81$ on 27 degrees of freedom (*p-values* = 0.52 and 0.64 respectively). For model ID5, $G^2 = 26.53$ and $X^2 = 24.09$ on 28 degrees of freedom (*p-value* = 0.54 and 0.68 respectively). Note that all parameter estimates in the outcome model are the same for drop-out models ID2 and ID5. This phenomenon is induced by the

imposed boundary constraints mentioned earlier which allow separate maximizations for the parameters in the outcome and drop-out models.

- **Ignorable Drop-out**

Under the assumption of ignorable drop-out (either RD or CRD), the maximum likelihood estimates obtained by the QN minimization are in the interior of the parameter space. The results are summarized in Tables B.7 to B.11.

As expected, the parameter estimates in the measurement process are the same in all the RD and CRD models. Hence, the conclusions about the treatment effects in the marginal model for the exacerbation rates do not differ across the different specifications of these drop-out models. Only the HD group has a different effect on the exacerbation rates compared to the PL group (two-sided p-$value \approx 0.01$ based on $\hat{\beta}_2$); the odds of having exacerbations in the PL group are about 1.6 times the odds in the HD group. There is a strong indication of a linear decrease over time in the log odds of having exacerbations (two-sided p-$value \approx 0.001$ based on $\hat{\beta}_3$).

The treatment effects express themselves similarly in the association model. There are no apparent differences between the LD and PL groups in the odds of having exacerbations at two and three occasions (two-sided p-$value \approx 0.40$ based on $\hat{\alpha}_1$), but the HD and PL groups differ (two-sided p-$value \approx 0.004$ based on $\hat{\alpha}_2$). The intercept parameter estimates are quite similar, although slightly more negative, to those obtained under models ID3 and ID6. Again, the estimated values for α_{12}, α_{13} and α_{23} are reasonably similar, and the estimate for α_{123} is somewhat more negative. This indicates a model which assumes a common intercept parameter for all the 2-way association models and a separate intercept parameter for the 3-way association may be reasonable for our data.

We next consider selecting a simpler model among the three RD models. Based on the LRT, the model reduction from RD1 to RD2 is permissible (p-$value =$

0.36). One can also reduce model RD2 to RD3 (*p-value* = 0.81). The LRT statistic comparing model RD1 to RD3 also indicates the reduction to model RD3 is reasonable (*p-value* = 0.74). Thus, model RD3 is the simplest permissble model under the RD assumption. Similarly, if a CRD mechanism is assumed, model CRD2 can be used instead of CRD1 to describe our annual data (*p-value* = 0.54).

- **Types of Drop-out in the Data**

In the earlier part of this section, we determined that reductions from model ID1 to models ID2 and ID5 are permissible, with model ID5 being the simplest possible model among the six ID models considered. These three models can be used to examine whether the drop-out mechanisms in our data is ID, RD or CRD according to the classification by Little and Rubin (1987).

To assess whether the drop-out occurred at random (RD), we can compare model ID1 to RD1. This comparison examines $\eta_{23} = \eta_{22} = 0$. The LR statistic of 6.85 (df = 2; *p-value* = 0.03) provides evidence against this reduction. As already established, it is reasonable to have common regression parameters describing drop-out at the different time points (reduce from ID1 to ID2). Hence, the comparison between model ID2 and RD2 should provide a more sensitive assessment of our question. In this case, we investigate whether $\eta_2 = 0$ and the result agrees with the previous assessment (LR statistic = 6.66, df = 1; *p-value* = 0.01). The less sensitive comparison of model ID1 to RD2 also sugguests one should not reduce to the simpler model (LR statistic = 7.69, df = 3; *p-value* = 0.05). Thus, the data indicate that the drop-out did not occur at random.

As reduction to an RD model is not allowed, presumably reduction to a CRD model will also not be allowed. For the sake of completeness, we perform various assessments to examine this. Model CRD1 can be compared to model ID1, ID2 and ID5 to examine the dependence between the drop-out and the outcome processes. The LR test comparing models ID1 and CRD1 clearly indicates the reduction is

84

not permissible (LR statistic = 14.03, df = 4; *p-value* = 0.007). The LR statistics for examining the reduction from model ID2 and ID5 to CRD1 are 13.00 (df = 2; *p-value* = 0.002) and 11.90 (df = 1; *p-value* < 0.001), respectively. As expected, the comparison to ID5 provides the strongest evidence. Thus, the data provide strong evidence against the hypothesis that the drop-out process is independent of the outcome process.

According to these comparisons, one cannot reduce from the ID models to any of these RD and CRD models. We can thus confidently conclude that the drop-out process in our data is informative.

7.2.3 Summary

We fitted six ID models and the maximum likelihood solutions for four of these models lie on the boundary of the parameter space. This phenomenon does not occur in the case where the drop-out mechanism is assumed to be ignorable.

Based on LR tests, we conclude that the drop-out mechanism in our data is informative and model ID5 is determined to be the simplest possible model for our data. The treatment effects appear in both the marginal and association models. However, we focus primarily on the treatment effects in the marginal model. Under model ID5, the HD group has a lower rate of exacerbations compared to the PL group. The odds ratio of having exacerbations in the HD group relative to the PL group is estimated to be 0.62 and the corresponding approximate 95% confidence interval (CI) is (0.42, 0.90). The indication of a linear decrease in the odds of having exacerbations over time is quite weak; the approximate 95% CI for β_3 is $(-0.26, 0.03)$. The treatment effects in the association model convey a similar story: the odds of experiencing exacerbations at two occasions and at all three occasions in the LD group are not significantly different from the PL group, but these odds are clearly lower in the HD group.

Interestingly, these conclusions are not very sensitive to the underlying drop-

out mechanisms for this data set. In particular, the parameter estimates (and standard errors) in the outcome model obtained with the ID assumption are fairly similar to those obtained with the ignorable drop-out assumptions.

7.3   Baker's Selection Model: Extensions of the Drop-out Model

In this section, we are interested in investigating the impact of different specifictions of the drop-out model on inferences concerning the treatment effects. The outcome model remains the same as in the previous section, and is coupled with the drop-out models considered in Baker (1995); that is, COV + LOR + LUR, COV * LUR and LOR * LUR. Since the only covariates to be used are the treatment groups indicators, we replace COV with TRT throughout this section.

We have established that models ID1, ID2 and ID5 can be used to describe our annual data but no reduction to the RD and CRD models is allowed. Model ID1 is of form LOR + LUR, with different parameters associated with each time of occurrence of the drop-outs. Models ID2 is obtained from model ID1 by assuming the regression parameters to be common at each time of occurrence of the drop-outs, while model ID5 corresponds to the further assumption that the drop-out probabilities do not depend on LOR.

In this section, we retain the feature of common regression parameters in all drop-out models considered. The three non-nested ID models considered for $h_t(\mathbf{r_{t-1}}, \mathbf{y}_t^* \mid \boldsymbol{x}; \boldsymbol{\eta}_t)$ are:

1. TRT * LUR:

For $t = 2, 3$ ($\mathbf{r_{t-1}}$ equal to $\{p, p\}$ or $\{p\}$), we have

$$
\begin{aligned}
\mathrm{logit}[h_t(\mathbf{r_{t-1}}, \mathbf{y}_t^* \mid \boldsymbol{x}; \boldsymbol{\eta}_t)] &= \eta_{0t} + \eta_1 LD + \eta_2 HD + \eta_3 y_t^* \\
&\quad + \eta_4 LD y_t^* + \eta_5 HD y_t^*,
\end{aligned} \tag{7.11}
$$

and for $t = 1$ ($\mathbf{r_{t-1}}$ equal to $\{\ \}$), the model is

$$\text{logit}[h_1(\{\ \}, y_1^* \mid \boldsymbol{x}; \boldsymbol{\eta}_1)] \quad = \quad \eta_{01} + \eta_1 LD + \eta_2 HD; \qquad (7.12)$$

2. TRT + LOR + LUR:

For $t = 2, 3$, the model is

$$\text{logit}[h_t(\mathbf{r_{t-1}}, \mathbf{y}_t^* \mid \boldsymbol{x}; \boldsymbol{\eta}_t)] \quad = \quad \eta_{0t} + \eta_1 LD + \eta_2 HD$$
$$+ \ \eta_3 y_{t-1}^* + \eta_4 y_t^*, \qquad (7.13)$$

and for $t = 1$, we have

$$\text{logit}[h_1(\{\ \}, y_1^* \mid \boldsymbol{x}; \boldsymbol{\eta}_1)] \quad = \quad \eta_{01} + \eta_1 LD + \eta_2 HD; \qquad (7.14)$$

3. LOR * LUR:

For $t = 2, 3$, the model is

$$\text{logit}[h_t(\mathbf{r_{t-1}}, \mathbf{y}_t^* \mid \boldsymbol{x}; \boldsymbol{\eta}_t)] \quad = \quad \eta_{0t} + \eta_1 y_{t-1}^* + \eta_2 y_t^* + \eta_3 y_{t-1}^* y_t^*, \quad (7.15)$$

and for $t = 1$, we have

$$\text{logit}[h_1(\{\ \}, y_1^* \mid \boldsymbol{x}; \boldsymbol{\eta}_1)] \quad = \quad \eta_{01}. \qquad (7.16)$$

One can view these models as expansions of models ID2 and ID5. More specifically, all three drop-out models are expansions of model ID5. Further, models TRT * LUR and TRT + LOR + LUR can also be considered as expansions of model ID2. Hence we can compare these models to models ID2 or ID5 for examining the improvement of the fit with these more general models. The results are presented in next subsection.

7.3.1 Results

Tables C.1 to C.3 in Appendix C display detailed summaries of the results corresponding to the three extended drop-out models. For each drop-out model, we

report the starting values used to obtain the parameter estimates, the estimated standard errors, negative log-likelihood values, and the number of iterations required to achieve convergence. The phenomenon observed in models ID2 and ID5 can also be seen in these drop-out models. For the drop-out model TRT * LUR (see Table C.1), the same parameter estimates are obtained regardless of the starting values used except for the intercept parameters, η_{03} and η_{02}, and the parameter associated with LUR (η_3). Parameters η_{03} and η_{02} are estimated as large negative values, and η_3 is estimated as large positive. Further the estimates of η_{03} and η_3 always sum to -1.226, and $\hat{\eta}_{02} + \hat{\eta}_3 = -1.350$. Similarly, for model TRT+LOR+LUR (see Table C.2), the intercept parameters, η_{03} and η_{02}, are estimated as being large negative, and the estimated value for η_4 (the regression parameter corresponding to LUR) is large positive, but $\hat{\eta}_{03} + \hat{\eta}_4 = -1.430$ and $\hat{\eta}_{02} + \hat{\eta}_4 = -1.573$.

The situation for model LOR * LUR is more complicated. Here we have the same phenomenon described for both models TRT * LUR and TRT + LOR + LUR, but the estimates of η_1 (the parameter corresponding to LOR) and η_3 (the parameter associated with the interaction term, LOR × LUR) also appear to satisfy the constraint, $\hat{\eta}_1 + \hat{\eta}_3 = 0.286$. The parameter estimates obtained from the fourth set of the starting values, in particular, indicate that the maximum likelihood solution corresponds to $\hat{\eta}_1 \to -\infty$ with $\hat{\eta}_1 + \hat{\eta}_3 = 0.286$.

To make comparison to models ID2 or ID5, we need to verify that the maximum likelihood solutions for these extended models occur at the suggested points on the boundary of the parameter space. Re-parameterizing in a similar fashion as previously, the conditional drop-out probabilities at years 2 and 3 can be expressed as in Table 7.9. We then substitute these expressions into the log-likelihood functions for the three models. To obtain the MLEs, we minimize the negative log-likelihood functions using the QN procedure. The results are reported in Tables 7.10 to 7.12.

The minimizations reported in Tables C.1 to C.3 required a large number of iterations for convergence and, in each case, the estimated Hessian matrix was

Table 7.9: Non-response Probability for the Second and Third Responses

Model: TRT * LUR			
With $\eta_{02} + \eta_3 = \Delta_1$, $\eta_{03} + \eta_3 = \Delta_2$			
LOR	LUR	logit$\{\Pr(R_2 = \text{a} \mid \{\text{p}\}, \mathbf{y_2^*}, x)\}$	logit$\{\Pr(R_3 = \text{a} \mid \{\text{p,p}\}, \mathbf{y_3^*}, x)\}$
0/1	0	$-\infty$	$-\infty$
0/1	1	$\Delta_1 + (\eta_1 + \eta_4)LD + (\eta_2 + \eta_5)HD$	$\Delta_2 + (\eta_1 + \eta_4)LD + (\eta_2 + \eta_5)HD$

Model: TRT + LOR + LUR			
With $\eta_{02} + \eta_4 = \Delta_1$, $\eta_{03} + \eta_4 = \Delta_2$			
LOR	LUR	logit$\{\Pr(R_2 = \text{a} \mid \{\text{p}\}, \mathbf{y_2^*}, x)\}$	logit$\{\Pr(R_3 = \text{a} \mid \{\text{p,p}\}, \mathbf{y_3^*}, x)\}$
0/1	0	$-\infty$	$-\infty$
0	1	$\Delta_1 + \eta_1 LD + \eta_2 HD$	$\Delta_2 + \eta_1 LD + \eta_2 HD$
1	1	$\Delta_1 + \eta_1 LD + \eta_2 HD + \eta_3$	$\Delta_2 + \eta_1 LD + \eta_2 HD + \eta_3$

Model: LOR * LUR			
With $\eta_{02} + \eta_2 = \Delta_1$, $\eta_{03} + \eta_2 = \Delta_2$, $\eta_1 + \eta_3 = \Delta_3$			
LOR	LUR	logit$\{\Pr(R_2 = \text{a} \mid \{\text{p}\}, \mathbf{y_2^*}, x)\}$	logit$\{\Pr(R_3 = \text{a} \mid \{\text{p,p}\}, \mathbf{y_3^*}, x)\}$
0/1	0	$-\infty$	$-\infty$
0	1	Δ_1	Δ_2
1	1	$\Delta_1 + \Delta_3$	$\Delta_2 + \Delta_3$

reset to a unit matrix in the course of the computations. These features were not found for the minimizations reported in Tables 7.10 to 7.12. In particular, the number of iterations needed in Tables 7.10 to 7.12 is, on average, only one-third the number required in Tables C.1 to C.3. Furthermore, the estimated Hessian matrix in Tables 7.10, 7.11 and 7.12 was never reset to unity throughout the minimization process. The parameter estimates and the log-likelihood values in the corresponding tables in these two sets are identical to the number of digits displayed, but the log-likelihood is always slightly larger at the boundary point than at the interior points located by the original minimizations. Hence, we have shown that the maximum likelihood solutions for these extended models are indeed located at the suggested points on the boundary.

Table 7.10: Results for Model TRT * LUR Evaluated on the Boundary: $\eta_{03} \to -\infty$, $\eta_{02} \to -\infty$, $\eta_{02} + \eta_3 = \Delta_1$ and $\eta_{03} + \eta_3 = \Delta_2$

Parameter	Estimate	SE
β_0	0.886	0.204
$\beta_1 \ (LD)$	-0.017	0.201
$\beta_2 \ (HD)$	-0.484	0.198
$\beta_3 \ (time)$	-0.118	0.075
α_{12}	-0.004	0.165
α_{13}	-0.010	0.165
α_{23}	-0.111	0.180
α_{123}	-0.511	0.182
α_1	-0.103	0.213
α_2	-0.649	0.221
η_{01}	-2.136	0.293
$\eta_1(LD)$	-0.203	0.433
$\eta_2(HD)$	0.296	0.394
$\eta_4(LD \times LUR)$	0.571	0.521
$\eta_5(HD \times LUR)$	-0.620	0.518
Δ_1	-1.350	0.244
Δ_2	-1.226	0.251
Neg. Loglik	931.223 (# Iter = 23)	

90

Table 7.11: Results for Model TRT + LOR + LUR Evaluated on the Boundary: $\eta_{03} \to -\infty$, $\eta_{02} \to -\infty$, $\eta_{02} + \eta_4 = \Delta_1$ and $\eta_{03} + \eta_4 = \Delta_2$

Parameter	Estimate	SE
β_0	0.886	0.198
$\beta_1\ (LD)$	-0.017	0.195
$\beta_2\ (HD)$	-0.484	0.190
$\beta_3\ (time)$	-0.118	0.074
α_{12}	-0.004	0.157
α_{13}	-0.010	0.155
α_{23}	-0.111	0.169
α_{123}	-0.511	0.173
α_1	-0.103	0.207
α_2	-0.649	0.215
η_{01}	-2.156	0.223
$\eta_1(LD)$	-0.209	0.238
$\eta_2(HD)$	-0.023	0.249
$\eta_3(LOR)$	0.290	0.202
Δ_1	-1.573	0.283
Δ_2	-1.430	0.242
Neg. Loglik	933.350 (# Iter = 25)	

Table 7.12: Results for Model LOR * LUR Evaluated on the Boundary: With $\eta_{03} \rightarrow -\infty, \eta_{02} \rightarrow -\infty, \eta_1 \rightarrow -\infty, \eta_{02} + \eta_2 = \Delta_1, \eta_{03} + \eta_2 = \Delta_2$ and $\eta_1 + \eta_3 = \Delta_3$

Parameter	Estimate	SE
β_0	0.886	0.206
β_1 (LD)	-0.017	0.196
β_2 (HD)	-0.484	0.194
β_3 $(time)$	-0.118	0.074
α_{12}	-0.004	0.163
α_{13}	-0.010	0.160
α_{23}	-0.111	0.172
α_{123}	-0.511	0.177
α_1	-0.103	0.208
α_2	-0.649	0.217
η_{01}	-2.089	0.167
Δ_1	-1.499	0.265
Δ_2	-1.356	0.265
Δ_3	0.286	0.277
Neg. Loglik	933.922 (# Iter = 21)	

There is an interesting point to note before moving on to the comparisons between these models and the models described in the previous section. Tables 7.10 to 7.12 (see also Tables C.1 to C.3) display identical estimates for all the parameters in the outcome model. In fact, these parameter estimates are identical to those reported in Tables 7.5 and 7.7 (see also Tables B.2 and B.5) for models ID2 and ID5, respectively. The explanation for this is simple: for these drop-out models, the conditional probabilities that the second and third observations are missing are estimated to be zero when LUR = 0 (for both values of LOR). This simplifies the log-likelihood functions and allows the parameters in the outcome model and in the drop-out model to be maximized separately. As the 5 models share the same specification for the outcome process, it is then no surprise that the estimates of the parameters in the outcome model are identical even though the model specifications for the drop-out process differ.

Table 7.13: Results for Model TRT + LUR Evaluated on the Boundary: $\eta_{03} \to -\infty$, $\eta_{02} \to -\infty$, $\eta_{02} + \eta_4 = \Delta_1$ and $\eta_{03} + \eta_4 = \Delta_2$

Parameter	Estimate	SE
β_0	0.886	0.207
β_1 (LD)	-0.017	0.193
β_2 (HD)	-0.484	0.194
β_3 $(time)$	-0.118	0.076
α_{12}	-0.004	0.163
α_{13}	-0.010	0.159
α_{23}	-0.111	0.171
α_{123}	-0.511	0.175
α_1	-0.103	0.206
α_2	-0.649	0.218
η_{01}	-2.136	0.214
$\eta_1(LD)$	0.191	0.232
$\eta_2(HD)$	-0.051	0.248
Δ_1	-1.349	0.211
Δ_2	-1.222	0.226
Neg. Loglik	933.910 (# Iter = 27)	

To examine if one of these more complicated models should be employed for the drop-out process, we compare models TRT * LUR, TRT + LOR + LUR and LOR * LUR to models ID2 and ID5. By comparing model TRT * LUR to model ID5, we are examining whether the additional treatment effects (η_1, η_2) and the interaction between the treatment effects and the last unobserved response (η_4, η_5) provide a significant improvement on the fit of model ID5. The LR statistic (6.50 on df = 4; *p-value* = 0.16) indicates that there is not strong evidence that we should employ model TRT * LUR instead of model ID5.

Table 7.10 suggests the two interaction terms contribute the major improvement in expanding the model from ID5 to TRT * LUR. Further, the comparisons of model ID5 with models TRT + LOR + LUR and LOR * LUR seem to agree with this observation (LR statistics = 2.25 and 1.10, df = 3 and 2; *p-values* = 0.52 and 0.58, respectively). That is, neither the terms LOR and TRT nor the terms LOR and LOR × LUR contribute significant improvement to the fit of model ID5. Thus model TRT + LUR (obtained by setting $\eta_4 = \eta_5 = 0$ in model TRT * LUR) is an interesting intermediate model between models ID5 and TRT * LUR. The detailed results for model TRT + LUR are provided in Table C.4, while the maximum likelihood estimates evaluated on the suggested point on the boundary of the parameter space is presented in Table 7.13. Comparing model TRT * LUR to TRT + LUR examines the contribution of the interaction terms, TRT × LUR. The corresponding LRT statistic is 5.37 on 2 degrees of freedom (*p-value* = 0.07), indicating fairly weak evidence against the hypothesis that the interaction terms are negligible. The cautious approach in this situation might be to retain the more general model, i.e. TRT * LUR, rather than reducing to the simpler TRT + LUR. But the evidence is not compelling, so we choose to reduce to the simpler TRT + LUR as the drop-out model. We then further examine whether the reduction from model TRT + LUR to ID5 is reasonable. Not surprisingly, in view of the earlier comparisons of model TRT * LUR to ID5, the LRT shows that the data provide no evidence to conclude

that the additional TRT covariates improve the fit of model ID5 (*p-value* = 0.57).

We have already identified that models TRT + LOR + LUR and LOR * LUR do not improve the fit of model ID5. We can also examine whether these extended drop-out models provide improvements to model ID2 (LOR + LUR). The LR statistics are 1.15 and 0.00 (due to possible round-off error) on 2 and 1 degrees of freedom, respectively, indicating insufficient evidence to conclude that these extended drop-out models improve upon the fit of ID2 to our data set. Thus, neither the addition of TRT nor of LOR × LUR, provides a meaningful improvement in fit to ID2 (LOR + LUR). Hence, the simpler models ID2 or ID5 can be used to describe the drop-out process in our annual data set.

7.3.2 Summary

We explored various ways of modelling the drop-out process in our data. More specifically, the three models considered can be viewed as extensions of ID2 and ID5, two of the permissible drop-out models described in the previous section. We introduce treatment effects and interactions terms into the drop-out model with a view to examining whether there is any impact on the conclusions about the treatment effects. Because some of the conditional drop-out probabilities at years 2 and 3 are estimated to be zero for each of these three drop-out models (see Table 7.9), the estimates of the parameters in the outcome model from these three drop-out model specifications are identical to those obtained under models ID2 and ID5. It is also of interest to investigate whether a more general model specification for the drop-out process improves the fit. The results indicate that the simpler drop-out models ID2 or ID5 are adequate for our annual MS data. Thus, models ID2 and ID5 would be used throughout the next section.

7.4 Baker's Selection Model: Extension of the Outcome Model

In this section, we explore extensions of the outcome model considered in the two previous sections based on including other baseline covariates such as gender, age, duration of MS, EDSS and BOD, in addition to the treatment arms and time. The main purpose of this section is to investigate whether or not inclusion of other baseline covariates in the model has any impact on the conclusions about the treatment effects identified in Section 7.2.

For simplicity, we only consider the five baseline covariates described in Section 2.2.3, and these are introduced only into the marginal model for the exacerbation rates. The structure of the associations among the measurements is assumed to remain as previously described. This is thought reasonable as our primary interest focuses on the impact of additional covariates on the conclusions about the treatment effects in the marginal model for the exacerbation rates.

The baseline covariates are included one at a time into the marginal component of the outcome model. The forward stepwise procedure for inclusion of the baseline covariates in addition to the treatment and time effects is carried out in the following fashion:

(1). Consider each covariate for inclusion in the marginal model and examine if it has a significant effect;

(2). If any covariates have significant effects, include the most significant covariate in the marginal model and repeat (1). Stop when no remaining covariates are found to be significant;

(3). If no covariates have significant effects, terminate the procedure.

Even though EDSS score is an ordinal variable, for simplicity, we treat it as a continuous variable in our analysis. The BOD at baseline is skewed to the right

as is evident in Figure 2.5. Further, this covariate has a much larger scale than the other covariates. To avoid potential difficulties these features could induce in the estimating procedure, we use a logarithm transformation of the baseline BOD. Baseline BOD and its logarithm are highly associated (the correlation between them is roughly 0.7 based on the 362 patients who had baseline BOD greater than zero).

Among the ID models considered with the original form of the outcome model, we found that the reduced models ID2 and ID5 were adequate. The extensions considered in Section 7.3 did not improve the fit significantly, so these same drop-out models will be considered here. The inclusion of additional covariates in the outcome model contemplated here could improve the overall fit, in which case it would again be of interest to examine whether the drop-out process is ID, RD or CRD. As noted in Section 7.2, model ID2 is more suitable for this purpose. Hence, model ID2 is used to describe the drop-out process throughout this section.

7.4.1   Results

The results of the forward stepwise procedure to examine the role of each baseline covariate are summarized in Table 7.14. These log-likelihood values correspond to maximum likelihood estimates on the boundary of the parameter space as in the earlier fitting with models ID2 and ID5. Detailed summaries for the several cases reported in Table 7.14 appear in Tables D.1 to D.5 of Appendix D. The minimization process for obtaining the estimates reported in Tables D.1 to D.5 are similar to those described earlier. These maximum likelihood estimates were, on average, obtained at the 24th iteration and the Hessian matrix was never reset to a unit matrix in any of the minimizations.

The first baseline covariate in addition to the treatment group included in the model is gender of the patients (*Gender*). The LRT indicates gender is not an important covariate when estimating the exacerbation rate. This agrees with Wald test (see Table D.1: *z-score* = 1.17, *p-value* = 0.24). The effects of baseline EDSS

Table 7.14: The LRT Statistics in the Forward Stepwise Procedure

	Neg. Loglik for Model with $LD + HD + time$: 933.922				
Case	Additional COV	Neg. Loglik	LRT	*p-value*	Comment
1	*Gender*	933.244	1.357	0.24	
2	*EDSS*	933.901	0.043	0.84	
3	*Dur*	933.768	0.154	0.69	
4	*Age*	933.354	1.137	0.29	
5	log(BOD)	933.088	1.668	0.20	Based on Imputed Set 1
	log(BOD)	933.215	1.414	0.23	Based on Imputed Set 2
	log(BOD)	933.083	1.677	0.20	Based on Imputed Set 3
	log(BOD)	933.211	1.421	0.23	Based on Imputed Set 4

($EDSS$), duration of MS at baseline (Dur), and age at baseline (Age), are similarly not significant; see Table 7.14.

As mentioned before, there are 8 patients with missing BOD at baseline. In addition, 2 patients did not have any lesions at baseline, i.e. their baseline BOD value is zero. This creates a minor difficulty for converting baseline BOD to the log scale. Since the smallest non-zero baseline BOD value is 9, we impute a value between 0 and 9 for these 2 patients and perform a sensitivity analysis to determine whether the specific value chosen has any impact on the conclusion of our analysis. The arbitrary values chosen are 1.0 and 4.5. For the 8 patients who did not have any reading on BOD at baseline, one way to impute values for them is with the expectation-maximization (EM) algorithm, utilizing the other baseline covariates. For our purposes, it is sufficient to use the following values to fill in the 8 missing values and perform a sensitivity assessment:

- the average of the log of the baseline BOD from 362 patients (excluding the 10 patients mentioned earlier), i.e. 7.085 (BOD = 1194.516);

- the average of the log of the BOD from 364 patients (2 patients with zero

Table 7.15: Data sets used for assessing the sensitivity of the results when considering log(BOD) in addition to treatment group and gender as a covariate

Data Set	In terms of BOD	
	The 8 Patients	The 2 Patients
Imputed Set 1	1194.516	1.0
Imputed Set 2	1194.516	4.5
Imputed Set 3	1148.905	1.0
Imputed Set 4	1158.439	4.5

baseline BOD imputed to have a value 1.0), i.e. 7.047 (BOD = 1148.905);

- the average of the log of the BOD from 364 patients data (2 patients with zero baseline BOD imputed to have a value 4.5), i.e. 7.055 (BOD = 1158.439).

The four different combinations of values for imputing the 8 missing values and the 2 zero baseline BOD values are listed in Table 7.15. All four imputed data sets lead to a similar conclusion: $\log(BOD)$ is not a statistically important factor; see Table D.5 for the detailed results.

Since the other baseline covariates are demonstrated to be not important for estimating the rate of exacerbations, we can also perform an alternative assessment for the significance of $\log(BOD)$. In particular, the 8 patients with missing baseline BOD are withheld from the analysis and the 2 patients with zero baseline BOD are imputed to have values of 1.0 and 4.5. The results evaluated on the boundary of the parameter space are displayed in Table D.6. To perform a LRT, we re-fit model ID2 with this reduced data set; see Table D.7. The conclusion from this assessment remains the same as in the previous analyses. The LR statistics corresponding to the data sets with zero baseline BOD imputed as 1.0 and 4.5 are 1.83 and 1.56 on 1 degree of freedom ($p\text{-}values = 0.17$ and 0.21, respectively). The Wald-test for β_4 also leads to the same conclusion ($z\text{-}scores = 1.30$ and 1.20, with $p\text{-}values = 0.19$ and 0.23, respectively).

As expected, the parameter estimates associated with the drop-out process

are identical in Tables D.1 to D.5. The reason is exactly as in the previous section. Because the conditional drop-out probabilities at years 2 and 3 are estimated to be zero, the log-likelihood functions in all five cases can be expressed as the sum of a function of the parameters for the outcome model and a function of the parameters for the drop-out model. Hence, the MLEs for the parameters in the two processes can be obtained separately. Since we employ the ID2 drop-out model in all five cases, the parameter estimates are expected to be identical.

7.4.2 Summary

In the previous sections, the outcome model includes only the treatment groups and time as covariates. Here we consider also including the five baseline covariates, gender of the patients, EDSS, duration of disease, age and BOD, into the marginal model for the exacerbation rates. Model ID2 is used to described the drop-out process throughout the section. We found that none of these five baseline covariates contribute significantly to the fit in estimating the exacerbation rates.

7.5 Overall Summary for Baker's Selection Model

We have used Baker's selection modelling approach to address various questions, and we provided a brief summary of our findings at the end of Sections 7.2, 7.3 and 7.4. In this section, we briefly describe what we have learned about the data according to the results obtained with the simplest acceptable model.

In Section 7.2, we first determined that the non-saturated outcome model described in (7.2) − (7.4) is sufficient for our data by comparing it to various more general outcome models. This outcome model was then used throughout the section, coupled with drop-out models of the type LOR + LUR, to address questions of interest. We discovered that the maximum likelihood solutions for four (models ID1, ID2, ID4 and ID5) of the six informative drop-out models are located on the boundary of the parameter space. This results in identical parameter estimates for

the outcome model associated with drop-out models ID2 and ID5 for our data set. This boundary phenomenon does not arise in any of the ignorable drop-out models, i.e. the RD and CRD models.

Based on likelihood ratio tests, we concluded the drop-out mechanism in our data set is informative. Models ID1, ID2 and ID5 are permissible and adequate models for modelling the drop-out process in our data. Model ID5, the simplest permissible informative drop-out model, indicates that the drop-out process in our data depends on the outcome process only through the last unobserved measurement (LUR).

In Section 7.3, we explored several drop-out models that can be viewed as generalizations of models ID2 and ID5. In particular, we allowed the drop-out process to depend on the treatment groups. We found that these general drop-out models do not provide significant improvement to the fit of models ID2 or ID5. Thus, our drop-out process can be described by the simpler models ID2 and ID5.

In Section 7.4, we addressed the question of the significance of other baseline covariates such as gender, EDSS, duration of MS, age and BOD in estimating the rate of exacerbations. These covariates were considered for inclusion only in the marginal component of the outcome model. Based on the forward stepwise procedure, none of these covariates were found to contribute significantly to the fit.

Consequently, the simplest Baker's selection model consists of an outcome model composed of (7.2) $-$ (7.4), and the drop-out process described by model ID5; see Table 7.7. This model fits the data quite adequately (p-$value \geq 0.54$). The observed and expected counts for the 15 (observation patterns) by 3 (treatment groups) contingency table are presented in Table 7.16. None of the expected cell counts are zero even though this model estimates some of the conditional probabilities of drop-out to be zero. The discrepencies between the observed and the expected counts are generally small, indicating the data are well-described by the model. Thus, we make inferences based on our data using this model in Chapter 8.

Table 7.16: The Observed and Expected Cell Counts for Baker's Selection Model with Drop-Out Model ID5 ("$*$" denotes missing)

Pattern	PL		LD		HD	
$(0,0,0)$	14	(13.5)	9	(9.1)	15	(15.3)
$(0,0,1)$	3	(3.0)	5	(5.0)	11	(7.7)
$(0,1,0)$	6	(5.2)	7	(7.4)	12	(10.3)
$(0,1,1)$	5	(6.4)	7	(6.3)	7	(5.3)
$(1,0,0)$	9	(6.7)	9	(9.5)	9	(13.9)
$(1,0,1)$	12	(10.2)	10	(10.2)	6	(8.6)
$(1,1,0)$	8	(10.7)	11	(10.7)	13	(9.0)
$(1,1,1)$	25	(24.5)	18	(23.4)	16	(15.8)
$(0,0,*)$	0	(0.9)	1	(1.6)	1	(2.4)
$(0,1,*)$	2	(2.0)	3	(2.0)	0	(1.6)
$(1,0,*)$	1	(3.2)	5	(3.2)	2	(2.7)
$(1,1,*)$	11	(7.7)	10	(7.3)	3	(4.9)
$(0,*,*)$	2	(3.7)	7	(4.3)	4	(4.7)
$(1,*,*)$	12	(11.8)	12	(11.3)	8	(8.1)
$(*,*,*)$	13	(13.6)	11	(13.8)	17	(13.7)

Goodness-of-fit Tests
$G^2 = 26.53$ on 28 degrees of freedom; $p\text{-}value = 0.54$
$X^2 = 24.09$ on 28 degrees of freedom; $p\text{-}value = 0.68$

7.6   The Liu *et al.* Transition Model

In this section, we apply the Liu transition model to our annual data. Recall that Liu *et al.* (1999) employ a first-order transition model to model the outcome process. Further they assume that each of the conditional probabilities, $\Pr(Y_t^* = y_t^* \mid Y_{t-1}^* = y_{t-1}^*, x)$, does not depend on the covariates measured at time t which seems somewhat unusual (see Chapter 5 for details). In our case, we can proceed with their idea without making such an assumption about the dependence on the covariates measured at time t as we consider only covariates measured at baseline.

For the drop-out process, we consider three models: ID1, ID2 and ID3 as described in Table 7.1. Based on the LRT, we can select the simplest permissible model among the three. The basic idea of these models is similar to those considered in Liu *et al.* (1999) in the sense that the drop-out probabilities are assumed to depend only on the response observed prior the drop-out (LOR) and the response which would be observed if drop-out had not occurred (LUR). But in their data set, the first observation is always observed. Thus their models are slightly different than ours as they do not need a model for the case where the response pattern r_{t-1} is equal to { }.

♦ Repeated Binary Outcomes with Informative Drop-out ♦

◇ *Outcome Model*

A first-order transition model is assumed for the binary longitudinal data. This means that the current measurement, y_t^*, is related only to the previous measurement, y_{t-1}^*, for $t = 2, 3$, as well as to the baseline covariates of interest. Here only the treatment assignment and time are considered in the analysis since the results in the previous section indicate that gender of the patients, baseline EDSS, age at baseline, duration of MS, and baseline BOD were not important covariates in estimating the

rates of exacerbation. Thus, the outcome model employed can be expressed as:

$$\text{logit}\{\Pr(Y_t^* = 1 \mid Y_{t-1}^* = y_{t-1}^*, \boldsymbol{x}_t)\} = \beta_0 + \beta_1 LD + \beta_2 HD + \beta_3 t + \beta_4 y_{t-1}^*$$

$$(7.17)$$

◇ *Drop-out Model*

Models similar to ID3 and ID6 from Table 7.1 were considered in Liu *et al.* (1999). Here we propose to model the drop-out process using models ID1, ID2, ID3 and ID5. We choose to focus on these three ID models out of the six listed in Table 7.1 because they allow straightforward investigation for the form of the drop-out mechanisms according to the terminology by Little and Rubin (1987). Furthermore, it will be interesting to determine if this leads to the same choice of the ID models for the drop-out process, namely ID2 and ID5, as the Baker selection model approach.

♦ Repeated Binary Outcomes with Ignorable Drop-out ♦

To investigate the impact of different drop-out mechanisms on the treatment effects, we also consider drop-out models assuming the drop-out occurred at random (RD) and completely at random (CRD). The RD and CRD models are the same in Table 7.1.

Likelihood ratio tests can be performed to examine the type of drop-out in our annual data based on these models. The results for the parameter estimates under different drop-out mechanisms are presented in the subsequent subsection. We conclude the section with a brief summary.

7.6.1 Results

● **Informative Drop-out**

The maximum likelihood solutions for models ID1 and ID2 lie on the boundary of the parameter space, while those for model ID3 exist in the interior. The detailed

results for these models are summarized in Tables E.1 to E.6 of Appendix E. The boundary solutions for models ID1 and ID2 occur in a similar fashion as in Baker's selection model (see Tables E.1 and E.2). We present the MLEs computed on the boundary for drop-out models ID1 and ID2 in Tables 7.17 and 7.18, respectively. These reported estimates are obtained with many fewer iterations than those in Tables E.1 and E.2. Moreover, the estimated Hessian matrix in both cases was never reset to unity throughout the minimization process.

Notice that the MLEs for the parameters in the outcome model are identical for drop-out models ID1 and ID2. This is again because some of the conditional probabilities of drop-out at years 2 and 3 are estimated to be zero and hence the parameters in the outcome and drop-out models can be estimated separately. The G^2 and X^2 goodness-of-fit statistics shown in Table 7.20 provide some evidence of lack-of-fit in each case. Although the evidence is not compelling, the fit of these models for our data are somewhat questionable; perhaps a more complicated association structure or a more general drop-out model should be employed. However, our objective is not to perform a definitive analysis on our annual data, but rather to explore different approaches for modelling incomplete longitudinal binary data with informative drop-outs. Hence, despite their somewhat questionable fit, we do not elaborate on these models but rather go on to consider the best choices within this collection of models.

All three ID models lead to similar conclusions about the treatment effects. In particular, the chance that an exacerbation would be experienced, given the past history (whether or not an exacerbation occurred at the previous time point), is not significantly different between the LD and PL groups (all $p\text{-}values \geq 0.47$). Nevertheless, the LD effect is estimated to be much stronger in model ID3 than in models ID1 and ID2. All three models conclude that the HD group has a lower chance than the PL group to experience an exacerbation, given the past history ($p\text{-}values \leq 0.01$). Further, there is a strong suggestion of a linear decrease over

Table 7.17: Results for Liu Transition Model with Drop-out Model ID1 Evaluated on the Boundary: $\eta_{03} \to -\infty, \eta_{02} \to -\infty$, $\eta_{03} + \eta_{23} = \Delta_1$ and $\eta_{02} + \eta_{22} = \Delta_2$

Parameter	Estimate	SE
β_0	1.007	0.206
β_1 (LD)	-0.040	0.168
β_2 (HD)	-0.462	0.167
β_3 (time)	-0.324	0.095
β_4	0.692	0.161
η_{01}	-2.089	0.165
η_{13}	0.558	0.413
η_{12}	0.048	0.374
Δ_1	-1.548	0.348
Δ_2	-1.327	0.314
Neg. Loglik	942.259 (# Iter = 16)	

Table 7.18: Results for Liu Transition Model with Drop-out Model ID2 Evaluated on the Boundary: $\eta_{03}, \eta_{02} \to -\infty$, $\Delta_1 = \eta_{03} + \eta_2$ and $\Delta_2 = \eta_{02} + \eta_2$

Parameter	Estimate	SE
β_0	1.007	0.199
β_1 (LD)	-0.040	0.167
β_2 (HD)	-0.462	0.167
β_3 (time)	-0.324	0.094
β_4	0.692	0.161
η_{01}	-2.089	0.169
η_1	0.286	0.296
Δ_1	-1.356	0.279
Δ_2	-1.499	0.273
Neg. Loglik	942.687 (# Iter = 15)	

Table 7.19: Results for Liu Transition Model with Drop-out Model ID5 Evaluated on the Boundary: $\eta_{03}, \eta_{02} \rightarrow -\infty$, $\Delta_1 = \eta_{03} + \eta_2$ and $\Delta_2 = \eta_{02} + \eta_2$

Parameter	Estimate	SE
β_0	1.007	0.206
β_1 (LD)	-0.040	0.169
β_2 (HD)	-0.462	0.168
β_3 (time)	-0.324	0.095
β_4	0.692	0.161
η_{01}	-2.089	0.166
Δ_1	-1.356	0.185
Δ_2	-1.499	0.168
Neg. Loglik	943.239 (# Iter = 14)	

Table 7.20: Goodness-of-fit Statistics for Liu Transition Model with Drop-out Models ID1, ID2, ID3 and ID5

Model	Degrees of Freedom	G^2	p-value	X^2	p-value
ID1	30	42.56	0.06	40.91	0.09
ID2	32	42.77	0.10	41.15	0.13
ID3	34	48.09	0.06	46.06	0.08
ID5	33	46.10	0.06	45.76	0.07

time in the log odds of having exacerbations given the past history ($p\text{-}value < 0.001$ based on $\hat{\beta}_3$ in each model). The association parameter β_4 is highly significant (all $p\text{-}values < 0.001$). Under models ID1 and ID2, the odds of having an exacerbation given there was an exacerbation at the previous visit are 2.0 times the odds of having an exacerbation given there was no exacerbation at the previous visit; the corresponding approximate 95% CI for the odds ratio is $(1.46, 2.74)$. Under model ID3, the odds ratio is estimated as 1.8 and the approximate 95% CI is $(1.27, 2.48)$.

The LR statistic for the reduction from model ID1 to model ID2 is 0.86 on 2 degrees of freedom ($p\text{-}value = 0.65$) and hence is permissible. However, we cannot further reduce model ID2 to model ID3 (LR statistic $= 6.43$, df $= 2$; $p\text{-}value = 0.04$). Thus, the simplest ID model among these three is ID2, which is the same conclusion obtained with Baker's selection model.

Recall that with Baker's selection model, drop-out model ID5 is a reasonable reduction of model ID2. Thus, it is of interest to perform this assessment with the Liu transition model. The parameter estimates obtained from the QN minimization with drop-out model ID5 are summarized in Table E.4. The results indicate a similar feature of boundary solutions as in model ID2. Table 7.7 presents the maximum likelihood estimates obtained at the suggested boundary points for model ID5. The LR statistic indicates that the term corresponding to the last observed response included in ID2 does not provide an important improvement to the fit (LR statistic $= 1.10$, df $= 1$; $p\text{-}value = 0.29$). Further, while the goodness-of-fit of model ID5 is slightly less satisfactory than for ID2 (see Table 7.20), the evidence against the adequacy of model ID5 is not overly compelling. These conclusions are qualitatively similar to those obtained with Baker's selection model.

- **Ignorable Drop-out**

The results for the RD and CRD models are displayed in Tables E.5 and E.6, respectively. As expected, the parameter estimates for the outcome model are identical

under both drop-out mechanisms. All parameter estimates are located in the interior of the parameter space.

Under the assumption that the drop-out process is ignorable, the Wald tests suggest that the chance a patient would have an exacerbation given the past history is similar in the LD and PL group (p-$value \approx 0.50$). But the risk is significantly lower in the HD group than in the PL group (p-$value \approx 0.01$). As in the ID case, the suggestion of a linear decrease over time in the log odds of having exacerbations given the past history is quite strong (z-$score \approx -4.4$ based on $\hat{\beta}_3$; p-$value < 0.001$). The odds of having an exacerbation given there was an exacerbation in the previous period are about 1.8 times the odds of having an exacerbation given there was no exacerbation in the previous period; the approximate 95% CI for the odds ratio is (1.32, 2.50).

We perform LRTs for selecting the simplest RD and CRD models. The reduction from model RD1 to RD2 is permissible (p-$value = 0.36$), but the further reduction from model RD2 to RD3 is not allowed (p-$value = 0.007$). Under the CRD assumption, CRD1 is identified as the simplest possible model, as the reduction from CRD1 to CRD2 is not permissible (p-$value = 0.004$). These choices differ from those for Baker's selection model; see Section 7.2.

- **Types of Drop-out**

We established that models ID1 and ID2 are reasonable for describing our data. To investigate the types of drop-out in our annual data set, we compare these models with some RD and CRD models.

For assessing if the drop-out mechanism is of type RD, model ID1 can be compared to model RD1 and similarly, model ID2 can be compared to model RD2. The reduction from ID1 to RD1 is permissible (LR statistic = 3.68, df = 2; p-$value = 0.16$). However, the more sensitive assessment comparing model RD2 to ID2 (since the reduction from ID1 to ID2 is reasonable) provides a less definite

conclusion; the LR statistic equals 3.66 on 1 degree of freedom ($p\text{-}value = 0.06$). With a 5% level of significance, we would not reject the hypothesis that $\eta_2 = 0$, but with only a slightly larger acceptable type I error, we would reject the hypothesis. Thus further investigation is required.

The LR test indicates one cannot reduce from model RD2 to CRD1 (LR statistic = 6.14, df = 1; $p\text{-}value = 0.01$). The reduction from model ID2 to CRD1 is also not permitted (LR statistic = 9.80, df = 2; $p\text{-}value = 0.007$). Thus we need to make a decision based on the comparison between model ID2 and RD2. In such an ambiguous situation, one would usually prefer not to reduce from ID2 to RD2 because the simpler model may be more susceptible to potential bias in the results. As mentioned earlier, model ID2 can be further reduced to ID5. The comparison of model ID5 to CRD1 confirms that model CRD1 is not appropriate for our data (LR statistic = 8.70, df = 1; $p\text{-}value = 0.003$). Thus we conclude that the drop-out process in our data appears to be informative.

7.6.2 Summary

We considered a first-order transition model for modelling the outcome process, coupled with the same drop-out models considered in Section 7.2. Based on the likelihood ratio tests, it appears that the drop-out process in our model cannot be ignored. Model ID5 is identified as the simplest drop-out model that is acceptable for our data.

Based on this model, we computed the expected cell counts for the 15 (observation patterns) by 3 (treatment arms) contingency table; see Table 7.21. Despite some of conditional drop-out probabilities being estimated as zero, the expected counts are all nonzero. Notice that the differences between the observed and expected counts in some cells are quite large. For instance, the differences in cells $(0, 0, 0)$ and $(1, 1, 0)$ for the PL group and in cell $(1, 1, 1)$ for the LD group are larger than 5.0 in magnitude. This is also reflected in the values of G^2 and X^2, both

Table 7.21: The Observed and Expected Cell Counts for the Liu Transition Model with Drop-Out Model ID5 ("*" denotes missing)

Pattern	PL		LD		HD	
$(0,0,0)$	14	(7.4)	9	(8.1)	15	(15.6)
$(0,0,1)$	3	(6.1)	5	(6.4)	11	(8.1)
$(0,1,0)$	6	(5.8)	7	(6.1)	12	(8.3)
$(0,1,1)$	5	(9.5)	7	(9.6)	7	(8.6)
$(1,0,0)$	9	(9.3)	9	(9.7)	9	(13.2)
$(1,0,1)$	12	(7.6)	10	(7.7)	6	(6.9)
$(1,1,0)$	8	(14.4)	11	(14.6)	13	(14.0)
$(1,1,1)$	25	(23.6)	18	(23.1)	16	(14.5)
$(0,0,*)$	0	(1.6)	1	(1.6)	1	(2.1)
$(0,1,*)$	2	(2.4)	3	(2.5)	0	(2.2)
$(1,0,*)$	1	(2.0)	5	(2.0)	2	(1.8)
$(1,1,*)$	11	(6.1)	10	(6.0)	3	(3.7)
$(0,*,*)$	2	(3.9)	7	(4.1)	4	(4.3)
$(1,*,*)$	12	(9.8)	12	(9.8)	8	(7.2)
$(*,*,*)$	13	(13.6)	11	(13.8)	17	(13.7)

Goodness-of-fit Tests
$G^2 = 46.10$ on 33 degrees of freedom; $p\text{-}value = 0.06$
$X^2 = 45.76$ on 33 degrees of freedom; $p\text{-}value = 0.07$

indicating a potential lack-of-fit of this model. One could explore more complicated drop-out models or association structures to improve the fit of the model, but such extensions are not our main interest. Thus, we go on to make inferences based on our data using this model in the concluding chapter.

Chapter 8

Conclusions

8.1 Conclusions

The main focus of this thesis has been on exploring likelihood-based methods for analyzing longitudinal binary responses under informative (or non-ignorable) drop-out. The two modelling approaches considered were Baker's selection model and the Liu *et al.* transition model. Both models belong to a general class of models known as selection models. A selection model factors the joint distribution for the response variables ($\mathbf{Y}$) and the indicator variables denoting whether the response variables were observed ($\mathbf{R}$) as

$$f(\mathbf{Y}, \mathbf{R}) = f(\mathbf{R} \mid \mathbf{Y}) f(\mathbf{Y}), \qquad (8.1)$$

where $f(\mathbf{R} \mid \mathbf{Y})$ is the model for the drop-out process and $f(\mathbf{Y})$ corresponds to the model for the measurement (or outcome) process.

The main difference between Baker's selection model and the Liu transition model resides in the model specification for the measurement process. Baker's selection model uses a parameterization proposed by Ekholm (1991, 1992) to acco-modate longitudinal binary measurements. That is, the outcome model is expressed in terms of a model for the (univariate marginal) probabilities of the responses and

113

an association model for the temporal associations among the responses. The Liu transition model, however, employs a first-order Markov chain transition model for the measurement process. The conditional distribution of response at time t (y_t) given the history of the responses up to time $t - 1$ is assumed to depend only on the response at the previous time point (y_{t-1}). These outcome models are coupled with a drop-out model specified as a time-ordered causal model incorporating the assumption that the drop-out does not depend on future events.

Given that the two approaches model the outcome process differently, this raises the question of the advantages and disadvantages of the two approaches. If the objective of the study is to study the effects of covariates on the marginal probabilities of the responses, marginal models provide a direct answer to this question. However, transition models should be used when the interest is in prediction (Diggle et al., 1994). Baker's selection model incorporates a more general structure for the strength of association among the responses than the Liu transition model. The structure for the associations among the responses in the Liu transition model is completely specified in terms of a single lagged effect. (Additional lagged effects could be added to the model but the nature of the association structure is limited by this parameterization.) For Baker's selection model, the expression for the outcome model for a sequence with more than three responses becomes more complicated, and the number of parameters increases rapidly. This is particularly so for the association model if no assumptions are made regarding the nature of the association structure among the responses. Unlike Baker's selection model, the number of parameters in the Liu transition outcome model need not change with the length of the response sequence.

Both models were applied to our annual version of the Berlex exacerbation data described in Chapter 2 to examine the sensitivity of the estimated effects of Interferon β-1b on the exacerbation rates in relapsing-remitting MS patients to various assumed forms for the drop-out mechanisms. More fundamentally, we were

114

Table 8.1: Estimated Chance of Exacerbations Based on Baker's Selection Model

Treatment Group	Year 1	Year 2	Year 3
PL	0.68	0.66	0.63
LD	0.68	0.65	0.63
HD	0.57	0.54	0.51

interested in studying the nature of the drop-out process in this clinical trial.

Using Baker's selection modelling approach, we verified that the relationships expressed in (7.2) − (7.4) are sufficient for describing the outcome process in our data. This outcome model coupled with drop-out model ID5 is determined to be the most parsimonious yet adequate model among other more general models considered. In other words, the drop-out process in our data is informative and it depends on the last unobserved response, but not on the last observed response. Based on this model, we conclude that the low dose effect is not significant. The odds of having exacerbations in the LD group are reduced only by 1.7% relative to the odds of having exacerbations in the PL group. The corresponding approximate 95% CI for the precent reduction in the odds is $(-44.9\%, 33.3\%)$. The high dose effect, however, is evidently different from the placebo effect. The odds of having exacerbations in the HD group are roughly 38.4% lower than the odds in the PL group (95% CI: 10.1%, 57.7%). Under the model assumption that the log odds of having exacerbations changes linearly over time, the odds are estimated to decrease by 11.1% per year in each group. The approximate 95% CI for the relative reduction in odds over time is $(-2.6\%, 23.0\%)$, indicating the reduction is not statistically significant. The estimated chances of having exacerbations at each occasion presented in Table 8.1 also reflect these conclusions. The chances of experiencing exacerbations are almost the same in the LD and PL groups, but are much smaller in the HD group. In each group, these chances decrease only slightly over time.

As for the association models, the LD and PL groups seem to have similar

Table 8.2: Estimated Chances of Exacerbations Based on the Liu *et al.* Transition Model

Exacerbation Experienced in Previous Period			
Treatment Group	Year 1	Year 2	Year 3
PL	0.80	0.74	0.67
LD	0.79	0.73	0.67
HD	0.71	0.64	0.57

No Exacerbation Experienced in Previous Period			
Treatment Group	Year 1	Year 2	Year 3
PL	0.66	0.59	0.51
LD	0.66	0.58	0.50
HD	0.56	0.47	0.39

chances of having exacerbations at exactly two or all three time points during the study, but these chances are lower in the HD group. The odds ratios for the LD and PL groups are estimated as 0.90, reflecting a 9.8% reduction in the odds in the LD group. The corresponding approximate 95% CI is $(-36.4\%, 40.3\%)$, implying the LD effect is not statistically significant. On the other hand, the odds in the HD group are only about half the odds in the PL group. The approximate 95% CI for the decrease in the odds in the HD group is $(20.1\%, 65.8\%)$. The estimates for the intercept parameters, α_{12}, α_{13} and α_{23}, are all quite small. This suggests a possible reduction to a model with all the 2-way associations in each treatment group being the same, i.e. $\alpha_{12} = \alpha_{13} = \alpha_{23}$. On the other hand, a separate intercept parameter for the 3-way association appears to be useful as $\hat{\alpha}_{123}$ is considerably larger in magnitude. Notice that, the estimated joint probabilities of the responses obtained from our model are slightly larger than those obtained under the independence assumption, indicating that there is some positive dependence among the responses; see Table 8.3.

With the Liu *et al.* transition approach, the simplest acceptable drop-out model is also identified to be ID5, again indicating the drop-out mechanism in our

116

data is informative. Even though the outcome model, and hence the parameters being estimated, are different than in Baker's selection model, the conclusion regarding the treatment effects remain quite similar. For fixed t and previous response y^*_{t-1}, the odds of having exacerbations are reduced by 3.9% (95% CI: $-33.8\%, 31.0\%$) in the LD group and by 37.0% (95% CI: 12.5%, 54.6%) in the HD group relative to the PL group. This indicates that only the high dosage of Interferon β-1b effectively reduces the odds of experiencing exacerbations in MS patients.

Similarly, the parameters β_3 and β_4 can also be interpretated as log odds ratios. In particular, $\exp(\beta_3)$ represents the ratio of the odds of having exacerbations at time $t+1$ as relative to time t for a patient with the same history at times $t-1$ and t ($y^*_{t-1} = y^*_t$). This odds ratio is estimated as 0.72 with approximate 95% CI (0.60, 0.87). The odds of having exacerbations given exacerbations in the previous period are 2.00 ($= \exp(\hat{\beta}_4)$) times the odds given no exacerbations in the previous period; the corresponding 95% CI for the odds ratio is (1.46, 2.74).

The estimated chances of experiencing exacerbations given the previous history presented in Table 8.2 also indicate similar conclusions regarding the treatment effects: the risks are much smaller in the HD group than in the LD and PL groups. Given that exacerbations were observed in the previous period (i.e. $y^*_{t-1} = 1$), the relative differences in the chances between the HD and PL groups are 11%, 14% and 15% at years 1, 2, and 3, respectively. For the case where no exacerbations were detected in the previous period (i.e. $y^*_{t-1} = 0$), the relative differences are slightly larger: 15%, 20% and 24% at years 1, 2, and 3, correspondingly.

Table 8.3 displays the values of $\hat{\Pr}(Y^*_s = 1, Y^*_t = 1)$ where $\{s,t\} = \{1,2\}$, $\{1,3\}$, $\{2,3\}$ and $\hat{\Pr}(Y^*_1 = 1, Y^*_2 = 1, Y^*_3 = 1)$ obtained from Baker's selection model and the Liu *et al.* transition model. The estimates are generally similar for the two approaches except for the estimated probability of exacerbations at visits 1 and 3 and at all three visits. The differences are more substantial for the former estimated probabilities; the magnitudes of the (absolute) differences are 0.08, 0.06, 0.12 in the

Table 8.3: Estimated $\Pr(Y_s^* = 1, Y_t^* = 1)$ and $\Pr(Y_1^* = 1, Y_2^* = 1, Y_3^* = 1)$ by Treatment Groups

Baker's Selection Model			
	PL	LD	HD
$\Pr(Y_1^* = 1, Y_2^* = 1)$	0.50	0.47	0.34
$\Pr(Y_1^* = 1, Y_3^* = 1)$	0.50	0.47	0.34
$\Pr(Y_2^* = 1, Y_3^* = 1)$	0.47	0.45	0.32
$\Pr(Y_1^* = 1, Y_2^* = 1, Y_3^* = 1)$	0.38	0.35	0.24
Assuming Independent Responses			
	PL	LD	HD
$\Pr(Y_1^* = 1, Y_2^* = 1)$	0.45	0.44	0.31
$\Pr(Y_1^* = 1, Y_3^* = 1)$	0.43	0.43	0.29
$\Pr(Y_2^* = 1, Y_3^* = 1)$	0.42	0.41	0.28
$\Pr(Y_1^* = 1, Y_2^* = 1, Y_3^* = 1)$	0.28	0.28	0.15

Liu et $al.$ Transition Model			
	PL	LD	HD
$\Pr(Y_1^* = 1, Y_2^* = 1)$	0.49	0.48	0.36
$\Pr(Y_1^* = 1, Y_3^* = 1)$	0.42	0.41	0.28
$\Pr(Y_2^* = 1, Y_3^* = 1)$	0.47	0.45	0.32
$\Pr(Y_1^* = 1, Y_2^* = 1, Y_3^* = 1)$	0.33	0.32	0.20

PL, LD and HD groups respectively.

In the intent-to-treat analyses reported in [35] (which assumed the drop-out occurred completely at random), the exacerbation rate was defined as the number of exacerbations experienced in one year. This is different from the exacerbation rate referred to throughout this thesis (the chance of having one or more exacerbations in a year). Nevertheless, it is of interest to compare the two sets of estimated treatment effects in terms of the relative change in the exacerbation rates. From the intent-to-treat analyses, the exacerbation rates in the PL, LD and HD group were 1.21, 1.05 and 0.84, respectively. Thus, the rates were 13% and 31% lower for the LD and HD groups relative to the PL group. Based on Baker's selection

model, the odds of having exacerbations are reduced by 1.7% and 38.4% in the LD and HD groups, respectively. Similarly, they are reduced by 3.9% and 37.0% under the Liu *et al.* transition model. The relative changes for the low dose effect are quite different between our approaches and the intent-to-treat analyses, but the variation is not as large for the high dose effect. Even though the magnitudes of the relative changes are quite different, the results convey a similar conclusion; that is, the effect of the high dosage of Interferon β-1b is much more evident than that of the low dosage. We also found that there is a weak positive association over time in the presence/absence of exacerbations, and that the influence of the association is present over more than 1 time period.

In the previous chapter, we provided the results from goodness-of-fit tests for both Baker's selection model and the Liu *et al.* transition model. The tests provided no evidence to suggest any lack-of-fit of Baker's selection model for our data. However, the adequacy of the Liu *et al.* transition model (*p-values* = 0.06 and 0.07 for G^2 and X^2 respectively) is questionable. The discrepency between some of the observed and expected counts obtained from the Liu *et al.* model is quite large (see Table 7.21). This seems to suggest the restrictive assumption on the form of the associations among the responses in the Liu transition model may not be adequate for our data; that is, a higher-ordered transition model could possible be used instead. Alternatively, this may suggest a more general model for the drop-out process should be employed. Between Baker's selection model and the Liu transition model, Baker's selection model seems much more satisfactory as it fits the data quite well (see Table 7.20).

In summary, analyses based on an assumption of ignorable non-response when the non-response mechanism is informative could lead to misleading results. By incorporating a non-response model in a likelihood-based approach, valid inferences can be obtained when the non-response mechanism is non-ignorable provided the non-response model correctly describes the non-response mechanism (Little and

Rubin, 1987). However, this approach is not without analytical difficulties. The parameters of the non-ignorable models may not be identifiable or the solutions to the likelihood equations (which may not be the maximum) may lie on the boundary of the parameter space. In Chapter 6, we showed that, with a saturated outcome model, the informative models of types COV * LUR, COV + LOR + LUR and LOR * LUR where COV represents categorical covariates, are identifiable. In the course of our analyses in Chapter 7, we demonstrated that the maximum likelihood solutions for some of our non-ignorable models were located on the boundary of the parameter space. This boundary phenomenon did not occur in any of the ignorable non-response models considered.

8.2   Further Work

- **Other Approaches of Interest**

There are approaches other than selection models that can be used for analyzing incomplete data. In particular, the pattern-mixture modelling framework proposed by Little (1993) has become an area of active research. The pattern-mixture approach specifies the joint distribution of the measurement and response processes in terms of the marginal distribution of the responses multiplied by the distribution of measurements, conditional on the response patterns. Pattern-mixture models are natural when the interest is in population strata defined by missing data patterns, but these models are typically underidentified (Little, 1993). Thus the models require restrictions or prior information to identify the parameters. Unlike selection models, with the pattern-mixture approach one can avoid specifying the form of the missing data mechanism as it is incorporated indirectly via parameter restrictions (Little, 1993). This is a possible attractive feature over the selection model approach as the latter is vulnerable to misspecification of the form of the missing-data mechanism. Further, pattern-mixture models are closer to the form of the data and sometimes simpler to fit. Thus, it would be of interest to re-analyze our annual data

with this approach and compare the results to those reported here.

- **Generalizations of the Data**

We chose to express the exacerbation data in terms of annual binary outcome variables. One could perform similar analyses on the binary data with more refined time intervals; for instance, semi-annual intervals. This semi-annual data may contain more information and may provide more precise estimates for the parameters.

As mentioned at the outset, there is a loss of information associated with dichotomizing the data. To retain all the information, one could analyze the count data presented in Table 2.4 treating these as realizations of Poisson random variables [18, 19]. One could also use this approach with finer time-intervals, semi-annual intervals say. The conclusions obtained from these annual and semi-annual count data might be more informative than those based on the dichotomized data.

Bibliography

[1] Baker, S.G. and Laird, N.M. (1988). Regression analysis for categorical variable with outcome subject to nonignorable nonresponse. *Journal of the American Statistical Association* **83**, 62-69.

[2] Baker, S.G. (1995). Marginal regression for repeated binary data with outcome subject to non-ignorable non-response. *Biometrics* **51**, 1042-1052.

[3] Broyden, C.G. (1970a). The convergence of a class of double-rank minimization algorithms, pt 1. *Journal of the Institute of Mathematics and Its Applications* **6**, 76-90.

[4] Broyden, C.G. (1970b). The convergence of a class of double-rank minimization algorithms, pt 2. *Journal of the Institute of Mathematics and Its Applications* **6**, 222-231.

[5] Dale, J. (1986). Global cross-ratio models for bivariate discrete ordered responses. *Biometrics* **42**, 909-917.

[6] Diggle, P.J. and Kenward, M.G. (1994). Informative drop-out in longitudinal data analysis. *Applied Statistics* **43**, 49-93.

[7] Ekholm, A. (1991). Fitting regression models to a multivariate binary response. In: *A Spectrum of Statistical Thought: Essays in Statistical Theory, Economics, and Population Genetics in Honour of Johan Fellman*, G. Rosenqvist,

K. Juselius, K. Nordström, J. Palmgren (eds), 19-32. Helsinki: Swedish School of Economics and Business Administration.

[8] Ekholm, A. (1992). Discussion of: Multivariate regression analysis for categorical data by K. Liang, S.L. Zeger, and B. Qaqish. *Journal of the American Statistical Association* **81**, 354-365.

[9] Ekholm, A. (1998). The muscatine children's obesity data reanalysed using pattern mixture models. *Applied Statistics* **47**, 251-263.

[10] Fitzmaurice, G.M. and Laird, N.M. (1993). A likelihood-based method for analysing longitudinal binary responses. *Biometrika* **80**, 141-151.

[11] Fitzmaurice, G.M., Laird, N.M. and Zahner, E.P. (1996). Multivariate logistic models for incomplete binary responses. *Journal of the American Statistical Association* **91**, 99-108.

[12] Fletcher, R. (1970). A new approach to variable metric algorithms. *The Computer Journal* **13**, 317-322.

[13] Glonek, G.F.V. (1999). On identifiability in models for incomplete binary data. *Statistics & Probability Letters* **41**, 191-197.

[14] Goodman, L.A. (1974). Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika* **61**, 215-231.

[15] Kenward, M.G., Lesaffre, E. and Molenberghs, G. (1994). An application of maximum likelihood and generalized estimating equations to the analysis of ordinal data from a longitudinal study with cases missing at random. *Biometrics* **50**, 945-953.

[16] Laird, N.M. (1988). Missing data in longitudinal studies. *Statistics in Medicine* **7**, 305-315.

[17] Liang, K.Y. and Zeger, S.L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73**, 13-22.

[18] Lindsey, J.K. (1997). *Applying Generalized Linear Models*. Springer-Verlag, New York.

[19] Lindsey, J.K. (1999). *Models for Repeated Measurements*. Oxford University Press, New York.

[20] Little, R.J.A. and Rubin, D.B. (1987). *Statistical Analysis with Missing Data*. John Wiley, New York.

[21] Little, R.J.A. (1993). Pattern-mixture models for multivariate incomplete data. *Journal of the American Statistical Association* **88**, 125-134.

[22] Liu, X., Waternaux, C. and Petkova, E. (1999). Influence of human immunodeficiency virus infection on neurological impairment: an analysis of longtudinal binary data with informative drop-out. *Applied Statistics* **48**, 103-115.

[23] Michiels, B., Molenberghs, G. and Lipsitz, S.R. (1999). Selection models and pattern-mixture models for incomplete data with covariates. *Biometrics* **55**, 978-983.

[24] Molenberghs, G., Kenward, M.G. and Lesaffre, E. (1997). The analysis of longitudinal ordinal data with non-random dropout. *Biometrika* **84**, 33-44.

[25] Molenberghs, G., Goetghebeur, E.J.T., Lipsitz, S.R. and Kenward, M.G. (1999). Nonrandom missingness in categorical data: strengths and limitations. *The American Statistician* **53**, 110-118.

[26] Nash, J.C. (1979). *Compact Numerical Methods for Computers: Linear Algebra and Function Minimisation*. Adam Hilger Ltd, Bristol.

[27] Paty, D.W., Li, D.K.B., The UBC MS/MRI Study Group, and The IFNB Multiple Sclerosis Study Group (1993). Interferon β-1b is effective in relapsing-remitting multiple sclerosis: II. MRI analysis results of a multicenter, randomized, double-blind, placebo-controlled trial. *Neurology* **43**, 662-668.

[28] Robins, J.M. and Rotnitzky, A. (1995) Semiparametric efficiency in multivariate regression models with missing data. *Journal of the American Statistical Association* **90**, 122-129.

[29] Rothenberg, T.J. (1971). Identification in parametric models. *Econometrica* **39**, 577-591.

[30] Rubin, D.B. (1976). Inference and missing data. *Biometrika* **63**, 581-592.

[31] Schluchter, M.D. (1992). Methods for the analysis of informatively censored longitudinal data. *Statistics in Medicine* **11**, 1861-1870.

[32] Shanno, D.F. (1970). Conditioning of quasi-Newton methods for function minimization. *Mathematics of Computation* **24**, 647-656.

[33] Sun, W. and Song, P. (2000). Statistical analysis of repeated measurements with informative cersoring times. *Statistics in Medicine*. To appear.

[34] Ten Have, T.R., Kunselman, A.R., Pulkstenis, E.P. and Landis, J.R. (1998). Mixed effects logistic regression models for longitudinal binary response data with informative drop-out. *Biometrics* **54**, 367-383.

[35] The IFNB Multiple Sclerosis Study Group (1993). Interferon β-1b is effective in relapsing-remitting multiple sclerosis: I. Clinical results of a multicenter, randomized, double-blind, placebo-controlled trial. *Neurology* **43**, 655-661.

[36] The IFNB Multiple Sclerosis Study Group (1995). Interferon β-1b in the treatment of multiple sclerosis: final outcome of the randomized controlled trial. *Neurology* **45**, 1277-1285.

[37] Wu, M.C. and Carroll, R.J. (1988). Estimation and comparison of changes in the presence of informative right censoring by modelling the censoring process. *Biometrics* **44**, 175-188.

Appendix A

Proof for Condition (6.4)

As in Section 6.2.2, there are two binary responses, Y_1 and Y_2, with only Y_2 subject to non-response. The outcome model is $\Pr(Y_1 = j, Y_2 = k \mid X = i) = \pi_{ijk}$, for $j, k = 0, 1$. The non-response model, $\Pr(R_2 = \mathrm{p} \mid Y_1 = j, Y_2 = k, X = i) = \rho_{ijk}$, is assumed to be homogeneous in Y_1; that is, $\rho_{ijk} = \rho_{ik}$. Thus, the joint probabilities for the observed data are

$$\Pr(Y_1 = j, Y_2 = k, \text{both observed} \mid X = i) = \theta_{ijk} = \pi_{ijk}\rho_{ik}$$

$$\Pr(Y_1 = j, Y_2 \text{ unobserved} \mid X = i) = \theta_{ij*} = \pi_{ij0}(1 - \rho_{i0}) + \pi_{ij1}(1 - \rho_{i1}),$$

and the marginal probabilities for Y_1 are

$$\pi_{ij\cdot} = \pi_{ij0} + \pi_{ij1} = \theta_{ij*} + \theta_{ij0} + \theta_{ij1}.$$

Let $\phi_{ik} = 1/\rho_{ik}$ and assume $I = 2$. The ϕ_{ik} must satisfy the following system of equations:

$$
\begin{pmatrix}
\theta_{100} & \theta_{101} & 0 & 0 \\
\theta_{110} & \theta_{111} & 0 & 0 \\
0 & 0 & \theta_{200} & \theta_{201} \\
0 & 0 & \theta_{210} & \theta_{211}
\end{pmatrix}
\begin{pmatrix}
\phi_{10} \\
\phi_{11} \\
\phi_{20} \\
\phi_{21}
\end{pmatrix}
=
\begin{pmatrix}
\pi_{10\cdot} \\
\pi_{11\cdot} \\
\pi_{20\cdot} \\
\pi_{21\cdot}
\end{pmatrix}. \tag{A.1}
$$

Given the multinomial probabilities $\boldsymbol{\theta}$, there is a unique solution for the ϕ_{ik} provided the coefficient matrix is non-singular; that is, provided the determinant of the coefficient matrix does not equal to 0.

The determinant of the coefficient matrix, $(\theta_{111}\theta_{100} - \theta_{101}\theta_{110})(\theta_{211}\theta_{200} - \theta_{201}\theta_{210})$, will be non-zero provided

$$\theta_{111}\theta_{100} - \theta_{101}\theta_{110} \;\neq\; 0 \qquad\qquad (A.2)$$

and

$$\theta_{211}\theta_{200} - \theta_{201}\theta_{210} \;\neq\; 0. \qquad\qquad (A.3)$$

To satisfy (A.2), we require

$$\theta_{111}\theta_{100} \;\neq\; \theta_{101}\theta_{110}$$
$$\Leftrightarrow \quad \pi_{111}\rho_{11}\pi_{100}\rho_{10} \;\neq\; \pi_{101}\rho_{11}\pi_{110}\rho_{10}$$
$$\Leftrightarrow \quad \pi_{111}(\pi_{10\cdot} - \pi_{101}) \;\neq\; \pi_{101}(\pi_{11\cdot} - \pi_{111})$$
$$\Leftrightarrow \quad \pi_{111}/\pi_{11\cdot} \;\neq\; \pi_{101}/\pi_{10\cdot}$$
$$\Leftrightarrow \quad \Pr(Y_2 = 1 \mid Y_1 = 1, X = 1) \;\neq\; \Pr(Y_2 = 1 \mid Y_1 = 0, X = 1).$$

Similarly, to satisfy (A.3) requires

$$\Pr(Y_2 = 1 \mid Y_1 = 1, X = 2) \neq \Pr(Y_2 = 1 \mid Y_1 = 0, X = 2).$$

Thus the necessary and sufficient condition for the coefficient matrix to be non-singular is

$$\Pr(Y_2 = 1 \mid Y_1 = 1, X = i) \neq \Pr(Y_2 = 1 \mid Y_1 = 0, X = i)$$

for $i = 1, 2$. Thus, the ϕ_{ik} are identifiable unless this condition fails to hold. Note that, in contrast to the argument leading to condition (6.3), the argument leading to this condition remains the same if the number of levels of the categorical covariate X is greater than 2 ($I > 2$).

Appendix B

Detailed Results for the Selection Models Described in Section 7.2

Table B.1: Results for Model ID1

Parameter	Set 1		Set 2	
	SV	Estimate (SE)	SV	Estimate (SE)
β_0	0.82	0.876 (0.919)	0.90	0.876 (0.816)
β_1	0.00	-0.028 (1.036)	-0.02	-0.028 (0.584)
β_2	0.00	-0.489 (0.896)	-0.50	-0.489 (0.357)
β_3	-0.26	-0.122 (0.568)	-0.12	-0.122 (0.388)
α_{12}	-0.60	-0.020 (0.579)	-0.02	-0.020 (0.404)
α_{13}	-0.63	-0.031 (0.840)	-0.03	-0.031 (0.378)
α_{23}	-0.77	-0.136 (0.959)	-0.14	-0.136 (0.486)
α_{123}	-1.15	-0.534 (0.702)	-0.50	-0.534 (0.446)
α_1	0.00	-0.113 (1.188)	-0.11	-0.113 (0.656)
α_2	0.00	-0.657 (0.938)	-0.66	-0.657 (0.391)
η_{03}	-1.95	-14.421 (1.025)	-1.00	-15.848 (0.784)
η_{13}	0.00	0.558 (1.003)	0.50	0.558 (0.769)
η_{23}	0.00	12.874 (1.015)	1.00	14.301 (0.775)
η_{02}	-1.95	-3.360 (1.042)	-2.00	-3.360 (0.690)
η_{12}	0.00	0.140 (1.001)	0.14	0.140 (0.540)
η_{22}	0.00	1.860 (1.030)	2.00	1.860 (0.787)
η_{01}	-1.95	-2.089 (1.057)	-2.00	-2.089 (0.563)
Neg. Loglik	933.407 (# Iter = 71)		933.407 (# Iter = 70)	
Parameter	Set 3		Set 4	
	SV	Estimate (SE)	SV	Estimate (SE)
β_0	0.90	0.876 (0.799)	0.876	0.876 (0.330)
β_1	-0.03	-0.028 (0.266)	-0.028	-0.028 (0.338)
β_2	0.00	-0.489 (0.286)	-0.489	-0.489 (0.341)
β_3	-0.12	-0.122 (0.394)	-0.122	-0.122 (0.079)
α_{12}	-0.02	-0.020 (0.287)	-0.020	-0.020 (0.302)
α_{13}	-0.04	-0.031 (0.251)	-0.031	-0.031 (0.311)
α_{23}	-0.15	-0.136 (0.434)	-0.136	-0.136 (0.321)
α_{123}	-0.50	-0.534 (0.292)	-0.534	-0.534 (0.334)
α_1	0.00	-0.113 (0.287)	-0.113	-0.113 (0.355)
α_2	0.00	-0.657 (0.310)	-0.657	-0.657 (0.371)
η_{03}	-2.00	-15.400 (0.799)	-20.000	-15.171 (0.737)
η_{13}	0.56	0.558 (0.865)	0.558	0.558 (0.503)
η_{23}	0.00	13.853 (0.794)	1.000	13.624 (0.739)
η_{02}	-2.40	-3.360 (0.806)	-3.360	-3.360 (0.800)
η_{12}	0.15	0.140 (0.783)	0.140	0.140 (0.828)
η_{22}	2.00	1.860 (0.813)	1.860	1.860 (0.813)
η_{01}	0.00	-2.089 (0.266)	-2.089	-2.089 (0.171)
Neg. Loglik	933.407 (# Iter = 59)		933.407 (# Iter = 91)	

Table B.2: Results for Model ID2

| Parameter | Set 1 | | Set 2 | |
	SV	Estimate (SE)	SV	Estimate (SE)
β_0	0.82	0.886 (0.553)	0.88	0.886 (0.738)
β_1	0.00	$-$ 0.017 (0.538)	-0.02	$-$ 0.017 (0.603)
β_2	0.00	$-$ 0.484 (0.538)	-0.48	$-$ 0.484 (0.700)
β_3	-0.26	$-$ 0.118 (0.248)	-0.12	$-$ 0.118 (0.379)
α_{12}	-0.60	$-$ 0.004 (0.341)	0.00	$-$ 0.004 (0.487)
α_{13}	-0.63	$-$ 0.010 (0.325)	-0.01	$-$ 0.010 (0.683)
α_{23}	-0.77	$-$ 0.111 (0.373)	-0.11	$-$ 0.111 (0.848)
α_{123}	-1.15	$-$ 0.511 (0.368)	-0.51	$-$ 0.511 (0.607)
α_1	0.00	$-$ 0.103 (0.592)	-0.10	$-$ 0.103 (0.713)
α_2	0.00	$-$ 0.649 (0.647)	-0.65	$-$ 0.649 (0.818)
η_{03}	-1.95	-14.421 (0.888)	-1.95	-14.760 (1.075)
η_1	0.00	0.286 (0.815)	0.00	0.286 (0.844)
η_2	0.00	13.065 (0.708)	0.00	13.404 (0.875)
η_{02}	-1.95	-14.564 (0.889)	-1.95	-14.903 (0.982)
η_{01}	-1.95	$-$ 2.089 (0.990)	-2.00	$-$ 2.089 (0.995)
Neg. Loglik	933.922 (# Iter = 67)		933.922 (# Iter = 64)	

| Parameter | Set 3 | | Set 4 | |
	SV	Estimate (SE)	SV	Estimate (SE)
β_0	0.90	0.886 (0.788)	0.886	0.886 (0.530)
β_1	-0.02	$-$ 0.017 (0.734)	-0.017	-0.017 (0.389)
β_2	-0.50	$-$ 0.484 (0.751)	-0.484	-0.484 (0.377)
β_3	-0.12	$-$ 0.118 (0.324)	-0.118	-0.118 (0.286)
α_{12}	0.00	$-$ 0.004 (0.562)	-0.004	-0.004 (0.370)
α_{13}	-0.01	$-$ 0.010 (0.694)	-0.010	-0.010 (0.420)
α_{23}	-0.11	$-$ 0.111 (0.593)	-0.111	-0.111 (0.546)
α_{123}	-0.50	$-$ 0.511 (0.599)	-0.511	-0.511 (0.489)
α_1	-0.10	$-$ 0.103 (0.792)	-0.103	-0.103 (0.441)
α_2	-0.60	$-$ 0.649 (0.847)	-0.649	-0.649 (0.426)
η_{03}	-6.00	-14.384 (1.033)	-14.384	-13.732 (0.616)
η_1	-0.30	0.286 (0.957)	0.000	0.286 (0.318)
η_2	6.00	13.028 (0.990)	0.000	12.376 (0.584)
η_{02}	-4.00	-14.527 (0.904)	0.000	-13.875 (0.642)
η_{01}	-2.00	$-$ 2.089 (0.958)	0.000	-2.089 (0.838)
Neg. Loglik	933.922 (# Iter = 56)		933.922 (# Iter = 72)	

Table B.3: Results for Model ID3

Parameter	Set 1		Set 2	
	SV	Estimate (SE)	SV	Estimate (SE)
β_0	0.88	0.986 (0.210)	1.00	0.986 (0.210)
β_1	−0.02	−0.097 (0.199)	−0.10	−0.097 (0.200)
β_2	−0.48	−0.475 (0.196)	−0.50	−0.475 (0.195)
β_3	−0.12	−0.230 (0.083)	−0.20	−0.230 (0.083)
α_{12}	0.00	−0.082 (0.169)	−0.08	−0.082 (0.171)
α_{13}	−0.01	−0.189 (0.177)	−0.20	−0.189 (0.178)
α_{23}	−0.11	−0.345 (0.195)	−0.30	−0.345 (0.198)
α_{123}	−0.51	−0.706 (0.200)	−0.70	−0.706 (0.202)
α_1	−0.10	−0.191 (0.219)	−0.20	−0.191 (0.219)
α_2	−0.60	−0.648 (0.226)	−0.60	−0.648 (0.224)
η_0	−1.95	−2.195 (0.159)	−2.00	−2.195 (0.161)
η_1	0.00	0.416 (0.286)	0.40	0.416 (0.290)
η_2	0.00	0.222 (0.449)	0.20	0.222 (0.459)
Neg. Loglik	937.349 (# Iter = 20)		937.349 (# Iter = 21)	

Table B.4: Results for Model ID4

Parameter	Set 1		Set 2	
	SV	Estimate (SE)	SV	Estimate (SE)
β_0	0.82	0.880 (0.731)	0.88	0.880 (0.217)
β_1	0.00	−0.024 (0.252)	0.00	−0.024 (0.197)
β_2	0.00	−0.487 (0.241)	−0.50	−0.487 (0.202)
β_3	−0.26	−0.120 (0.386)	−0.12	−0.120 (0.075)
α_{12}	−0.60	−0.013 (0.216)	0.00	−0.013 (0.174)
α_{13}	−0.63	−0.022 (0.225)	−0.02	−0.022 (0.168)
α_{23}	−0.77	−0.126 (0.422)	−0.13	−0.126 (0.179)
α_{123}	−1.15	−0.524 (0.257)	−0.52	−0.524 (0.189)
α_1	0.00	−0.109 (0.298)	−0.10	−0.109 (0.208)
α_2	0.00	−0.654 (0.274)	−0.65	−0.654 (0.230)
η_{03}	−1.95	−14.818 (0.730)	−4.00	−16.284 (1.015)
η_{23}	0.00	13.654 (0.751)	2.00	15.119 (1.008)
η_{02}	−1.95	−3.819 (0.728)	−3.80	−3.819 (2.606)
η_{22}	0.00	2.464 (0.761)	0.00	2.464 (2.754)
η_{01}	−1.95	−2.089 (0.209)	−2.08	−2.089 (0.142)
Neg. Loglik	934.432 (# Iter = 67)		934.432 (# Iter = 63)	
Parameter	Set 3		Set 4	
	SV	Estimate (SE)	SV	Estimate (SE)
β_0	0.880	0.880 (0.880)	0.880	0.880 (0.170)
β_1	−0.024	−0.024 (0.720)	−0.024	−0.024 (0.165)
β_2	−0.486	−0.487 (0.703)	−0.487	−0.487 (0.171)
β_3	−0.120	−0.120 (0.434)	−0.120	−0.120 (0.074)
α_{12}	−0.010	−0.013 (0.438)	−0.013	−0.013 (0.118)
α_{13}	−0.022	−0.022 (0.771)	−0.022	−0.022 (0.121)
α_{23}	−0.125	−0.126 (0.845)	−0.126	−0.126 (0.139)
α_{123}	−0.524	−0.524 (0.641)	−0.524	−0.524 (0.131)
α_1	−0.109	−0.109 (0.776)	−0.109	−0.109 (0.174)
α_2	−0.650	−0.654 (0.820)	−0.654	−0.654 (0.195)
η_{03}	0.000	−15.432 (0.986)	−15.432	−15.432 (1.363)
η_{23}	0.000	14.268 (0.988)	14.268	14.268 (1.368)
η_{02}	−3.820	−3.819 (0.725)	−3.819	−3.819 (1.973)
η_{22}	2.460	2.464 (0.766)	2.464	2.464 (2.093)
η_{01}	−2.090	−2.089 (1.002)	−2.089	−2.089 (0.156)
Neg. Loglik	934.432 (# Iter = 61)		934.432 (# Iter = 24)	

Table B.5: Results for Model ID5

Parameter	Set 1		Set 2	
	SV	Estimate (SE)	SV	Estimate (SE)
β_0	0.82	0.886 (0.926)	0.890	0.886 (0.169)
β_1	0.00	-0.017 (0.898)	-0.017	-0.017 (0.178)
β_2	0.00	-0.484 (0.874)	-0.480	-0.484 (0.187)
β_3	-0.26	-0.118 (0.543)	-0.120	-0.118 (0.070)
α_{12}	-0.60	-0.004 (0.971)	0.000	-0.004 (0.134)
α_{13}	-0.63	-0.010 (0.965)	-0.010	-0.010 (0.140)
α_{23}	-0.77	-0.111 (0.990)	-0.110	-0.111 (0.153)
α_{123}	-1.15	-0.511 (0.891)	-0.510	-0.511 (0.151)
α_1	0.00	-0.103 (0.901)	-0.100	-0.103 (0.195)
α_2	0.00	-0.649 (0.940)	-0.650	-0.649 (0.204)
η_{03}	-1.95	-13.737 (1.000)	-4.000	-15.608 (301.091)
η_2	0.00	12.573 (1.002)	0.000	14.443 (301.092)
η_{02}	-1.95	-13.866 (1.000)	-2.000	-15.737 (301.091)
η_{01}	-1.95	-2.089 (1.000)	-2.080	-2.089 (0.166)
Neg. Loglik	934.473 (# Iter = 55)		934.473 (# Iter = 60)	
	Set 3		Set 4	
Parameter	SV	Estimate (SE)	SV	Estimate (SE)
β_0	0.886	0.886 (0.213)	0.900	0.886 (0.400)
β_1	-0.017	-0.017 (0.206)	-0.017	-0.017 (0.573)
β_2	-0.484	-0.484 (0.192)	-0.480	-0.484 (0.482)
β_3	-0.118	-0.118 (0.074)	-0.120	-0.118 (0.214)
α_{12}	-0.004	-0.004 (0.171)	0.000	-0.004 (0.305)
α_{13}	-0.010	-0.010 (0.167)	-0.010	-0.010 (0.306)
α_{23}	-0.111	-0.111 (0.178)	-0.110	-0.111 (0.456)
α_{123}	-0.511	-0.511 (0.185)	-0.510	-0.511 (0.460)
α_1	-0.103	-0.103 (0.220)	-0.100	-0.103 (0.619)
α_2	-0.649	-0.649 (0.214)	-0.650	-0.649 (0.534)
η_{03}	-15.608	-15.608 (0.582)	0.000	-13.737 (0.942)
η_2	14.443	14.443 (0.568)	0.000	12.573 (0.587)
η_{02}	-15.737	-15.737 (0.569)	0.000	-13.866 (0.896)
η_{01}	-2.089	-2.089 (0.166)	-2.000	-2.089 (0.577)
Neg. Loglik	934.473 (# Iter = 20)		934.473 (# Iter = 71)	

Table B.6: Results for Model ID6

Parameter	Set 1		Set 2	
	SV	Estimate (SE)	SV	Estimate (SE)
β_0	0.90	0.962 (0.209)	1.00	0.962 (0.208)
β_1	−0.10	−0.080 (0.197)	−0.08	−0.080 (0.198)
β_2	−0.50	−0.483 (0.198)	−0.50	−0.483 (0.195)
β_3	−0.20	−0.201 (0.077)	−0.20	−0.201 (0.078)
α_{12}	−0.05	−0.057 (0.171)	−0.06	−0.057 (0.168)
α_{13}	−0.10	−0.137 (0.172)	−0.10	−0.137 (0.170)
α_{23}	−0.20	−0.279 (0.187)	−0.30	−0.279 (0.186)
α_{123}	−0.50	−0.646 (0.193)	−0.60	−0.646 (0.190)
α_1	−0.10	−0.172 (0.213)	−0.20	−0.172 (0.216)
α_2	−0.60	−0.655 (0.223)	−0.70	−0.655 (0.224)
η_0	−1.95	−2.206 (0.151)	−2.00	−2.206 (0.161)
η_2	0.00	0.661 (0.264)	0.70	0.661 (0.269)
Neg. Loglik	938.464 (# Iter = 19)		938.464 (# Iter = 17)	

Table B.7: Results for Model RD1

Parameter	Set 1		Set 2	
	SV	Estimate (SE)	SV	Estimate (SE)
β_0	0.82	0.999 (0.211)	1.00	0.999 (0.216)
β_1	0.00	−0.106 (0.196)	−0.12	−0.106 (0.194)
β_2	0.00	−0.470 (0.196)	−0.47	−0.470 (0.194)
β_3	−0.26	−0.246 (0.080)	−0.20	−0.246 (0.082)
α_{12}	−0.60	−0.097 (0.166)	−0.10	−0.097 (0.173)
α_{13}	−0.63	−0.219 (0.169)	−0.22	−0.219 (0.172)
α_{23}	−0.77	−0.384 (0.182)	−0.38	−0.384 (0.188)
α_{123}	−1.15	−0.742 (0.188)	−0.74	−0.742 (0.195)
α_1	0.00	−0.201 (0.216)	−0.20	−0.201 (0.217)
α_2	0.00	−0.643 (0.229)	−0.64	−0.643 (0.227)
η_{03}	−1.95	−2.416 (0.335)	−2.41	−2.416 (0.336)
η_{13}	0.00	0.878 (0.396)	0.90	0.878 (0.396)
η_{02}	−1.95	−2.117 (0.300)	−2.11	−2.117 (0.293)
η_{12}	0.00	0.401 (0.360)	0.40	0.401 (0.350)
η_{01}	−1.95	−2.089 (0.165)	−2.08	−2.089 (0.164)
Neg. Loglik	936.833 (# Iter = 25)		936.833 (# Iter = 23)	

Table B.8: Results for Model RD2

Parameter	Set 1		Set 2	
	SV	Estimate (SE)	SV	Estimate (SE)
β_0	0.82	0.999 (0.204)	0.999	0.999 (0.208)
β_1	0.00	-0.106 (0.193)	-0.106	-0.106 (0.197)
β_2	0.00	-0.470 (0.187)	-0.470	-0.470 (0.196)
β_3	-0.26	-0.246 (0.080)	-0.246	-0.246 (0.077)
α_{12}	-0.60	-0.097 (0.166)	-0.097	-0.097 (0.168)
α_{13}	-0.63	-0.219 (0.162)	-0.219	-0.219 (0.170)
α_{23}	-0.77	-0.384 (0.174)	-0.384	-0.384 (0.186)
α_{123}	-1.15	-0.742 (0.183)	-0.742	-0.742 (0.193)
α_1	0.00	-0.201 (0.217)	-0.201	-0.201 (0.217)
α_2	0.00	-0.643 (0.229)	-0.643	-0.643 (0.227)
η_{03}	-1.95	-2.239 (0.258)	-2.239	-2.239 (0.247)
η_1	0.00	0.625 (0.261)	0.625	0.625 (0.264)
η_{02}	-1.95	-2.278 (0.250)	-2.278	-2.278 (0.253)
η_{01}	-2.08	-2.089 (0.169)	-2.089	-2.089 (0.167)
Neg. Loglik	937.250 (# Iter = 26)		937.250 (# Iter = 17)	

Table B.9: Results for Model RD3

Parameter	Set 1		Set 2	
	SV	Estimate (SE)	SV	Estimate (SE)
β_0	0.82	0.999 (0.206)	0.999	0.999 (0.208)
β_1	0.00	-0.106 (0.197)	-0.106	-0.106 (0.195)
β_2	0.00	-0.470 (0.179)	-0.470	-0.470 (0.196)
β_3	-0.26	-0.246 (0.078)	-0.246	-0.246 (0.073)
α_{12}	-0.60	-0.097 (0.167)	-0.097	-0.097 (0.168)
α_{13}	-0.63	-0.219 (0.168)	-0.219	-0.219 (0.163)
α_{23}	-0.77	-0.384 (0.180)	-0.384	-0.384 (0.181)
α_{123}	-1.15	-0.742 (0.187)	-0.742	-0.742 (0.191)
α_1	0.00	-0.201 (0.227)	-0.201	-0.201 (0.217)
α_2	0.00	-0.643 (0.219)	-0.643	-0.643 (0.225)
η_0	-1.95	-2.153 (0.133)	0.000	-2.153 (0.121)
η_1	0.00	0.518 (0.196)	0.000	0.518 (0.188)
Neg. Loglik	937.457 (# Iter = 22)		937.457 (# Iter = 24)	

Table B.10: Results for Model CRD1

Parameter	Set 1		Set 2	
	SV	Estimate (SE)	SV	Estimate (SE)
β_0	0.82	0.999 (0.203)	1.00	0.999 (0.212)
β_1	0.00	-0.106 (0.173)	-0.10	-0.106 (0.199)
β_2	0.00	-0.470 (0.189)	-0.40	-0.470 (0.197)
β_3	-0.26	-0.246 (0.078)	-0.20	-0.246 (0.080)
α_{12}	-0.60	-0.097 (0.172)	0.00	-0.097 (0.169)
α_{13}	-0.63	-0.219 (0.169)	-0.20	-0.219 (0.171)
α_{23}	-0.77	-0.384 (0.186)	-0.40	-0.384 (0.186)
α_{123}	-1.15	-0.742 (0.192)	-0.70	-0.742 (0.193)
α_1	0.00	-0.201 (0.196)	-0.20	-0.201 (0.221)
α_2	0.00	-0.643 (0.231)	-0.60	-0.643 (0.229)
η_{03}	0.00	-1.846 (0.161)	-2.00	-1.846 (0.170)
η_{02}	0.00	-1.849 (0.152)	-2.00	-1.849 (0.159)
η_{01}	0.00	-2.089 (0.154)	-2.10	-2.089 (0.165)
Neg. Loglik	940.322 (# Iter = 29)		940.322 (# Iter = 21)	

Table B.11: Results for Model CRD2

Parameter	Set 1		Set 2	
	SV	Estimate (SE)	SV	Estimate (SE)
β_0	0.82	0.999 (0.281)	1.00	0.999 (0.210)
β_1	0.00	-0.106 (0.211)	-0.10	-0.106 (0.192)
β_2	0.00	-0.470 (0.228)	-0.40	-0.470 (0.193)
β_3	-0.26	-0.246 (0.078)	-0.20	-0.246 (0.079)
α_{12}	-0.60	-0.097 (0.232)	0.00	-0.097 (0.169)
α_{13}	-0.63	-0.219 (0.221)	-0.20	-0.219 (0.172)
α_{23}	-0.77	-0.384 (0.227)	-0.40	-0.384 (0.185)
α_{123}	-1.15	-0.742 (0.250)	-0.70	-0.742 (0.190)
α_1	0.00	-0.201 (0.228)	-0.20	-0.201 (0.217)
α_2	0.00	-0.643 (0.239)	-0.60	-0.643 (0.224)
η_0	-1.95	-1.933 (0.106)	-2.00	-1.933 (0.095)
Neg. Loglik	941.040 (# Iter = 21)		941.040 (# Iter = 19)	

Appendix C

Detailed Results for the Selection Models Described in Section 7.3

Table C.1: Results for Drop-out Model: TRT * LUR

Parameter	Set 1			Set 2		
	SV	Estimate (SE)		SV	Estimate (SE)	
β_0	0.80	0.886 (1.178)		0.90	0.886 (0.926)	
$\beta_1(LD)$	-0.10	-0.017 (0.779)		-0.02	-0.017 (0.988)	
$\beta_2(HD)$	-0.50	-0.484 (0.876)		-0.50	-0.484 (1.095)	
$\beta_3(time)$	-0.20	-0.118 (0.422)		-0.10	-0.118 (0.911)	
α_{12}	-0.08	-0.004 (1.143)		0.00	-0.004 (1.365)	
α_{13}	-0.20	-0.010 (1.025)		-0.01	-0.010 (1.198)	
α_{23}	-0.30	-0.111 (0.877)		-0.10	-0.111 (1.265)	
α_{123}	-0.70	-0.511 (0.617)		-0.50	-0.511 (1.193)	
α_1	-0.20	-0.103 (1.096)		-0.10	-0.103 (0.987)	
α_2	-0.70	-0.649 (0.885)		-0.60	-0.649 (1.143)	
η_{03}	-1.95	-13.608 (1.053)		-1.00	-15.118 (1.047)	
η_{02}	-1.95	-13.732 (1.069)		-1.00	-15.242 (1.044)	
η_{01}	-1.95	-2.136 (1.235)		-2.10	-2.136 (1.026)	
$\eta_1(LD)$	0.00	-0.203 (1.232)		-0.20	-0.203 (1.020)	
$\eta_2(HD)$	0.00	0.296 (1.044)		0.30	0.296 (1.133)	
$\eta_3(LUR)$	0.00	12.382 (1.122)		1.00	13.892 (1.348)	
$\eta_4(LD \times LUR)$	0.00	0.571 (1.001)		0.60	0.571 (1.025)	
$\eta_5(HD \times LUR)$	0.00	-0.620 (1.101)		-0.60	-0.620 (1.100)	
Neg. Loglik	931.223 (# Iter = 60)			931.223 (# Iter = 65)		

139

Table C.2: Results for Drop-out Model: TRT + LOR + LUR

Parameter	Set 1		Set 2	
	SV	Estimate (SE)	SV	Estimate (SE)
β_0	0.80	0.886 (0.602)	0.90	0.886 (1.280)
$\beta_1(LD)$	−0.10	−0.017 (0.687)	−0.02	−0.017 (1.112)
$\beta_2(HD)$	−0.50	−0.484 (0.650)	−0.50	−0.484 (1.012)
$\beta_3(time)$	−0.20	−0.118 (0.267)	−0.10	−0.118 (0.683)
α_{12}	−0.08	−0.004 (0.561)	0.00	−0.004 (1.067)
α_{13}	−0.20	−0.010 (0.899)	−0.01	−0.010 (1.112)
α_{23}	−0.30	−0.111 (0.715)	−0.10	−0.111 (1.089)
α_{123}	−0.70	−0.511 (0.572)	−0.50	−0.511 (1.104)
α_1	−0.20	−0.103 (0.732)	−0.10	−0.103 (0.978)
α_2	−0.70	−0.649 (0.774)	−0.60	−0.649 (1.188)
η_{03}	−1.95	−13.920 (1.010)	−2.00	−14.006 (1.019)
η_{02}	−1.95	−14.063 (0.998)	−1.00	−14.149 (1.368)
η_{01}	−1.95	−2.156 (0.923)	−2.10	−2.156 (1.102)
$\eta_1(LD)$	0.00	0.209 (0.934)	0.20	−0.209 (1.287)
$\eta_2(HD)$	0.00	−0.023 (0.974)	−0.20	−0.023 (1.381)
$\eta_3(LOR)$	0.00	0.290 (0.873)	0.30	0.290 (1.145)
$\eta_4(LUR)$	0.00	12.490 (0.777)	1.00	12.576 (2.942)
Neg. Loglik	933.350 (# Iter = 65)		933.350 (# Iter = 60)	

Table C.3: Results for Drop-out Model: LOR * LUR

Parameter	Set 1		Set 2	
	SV	Estimate (SE)	SV	Estimate (SE)
β_0	0.80	0.886 (0.664)	0.90	0.886 (0.650)
$\beta_1(LD)$	−0.10	−0.017 (0.411)	−0.02	−0.017 (0.733)
$\beta_2(HD)$	−0.50	−0.484 (0.639)	−0.50	−0.484 (0.800)
$\beta_3(time)$	−0.20	−0.118 (0.361)	−0.10	−0.118 (0.188)
α_{12}	−0.08	−0.004 (0.448)	0.00	−0.004 (0.487)
α_{13}	−0.20	−0.010 (0.539)	−0.01	−0.010 (0.766)
α_{23}	−0.30	−0.111 (0.711)	−0.10	−0.111 (0.586)
α_{123}	−0.70	−0.511 (0.643)	−0.50	−0.511 (0.610)
α_1	−0.20	−0.103 (0.456)	−0.10	−0.103 (0.788)
α_2	−0.70	−0.649 (0.751)	−0.60	−0.649 (0.847)
η_{03}	−1.95	−12.674 (0.953)	−1.00	−14.340 (0.992)
η_{02}	−1.95	−12.817 (0.936)	−1.00	−14.483 (1.001)
η_{01}	−1.95	−2.089 (0.979)	−2.10	−2.089 (0.998)
$\eta_1(LOR)$	−0.10	−0.711 (0.850)	0.10	0.152 (1.002)
$\eta_2(LUR)$	0.00	11.318 (0.785)	1.00	12.984 (0.995)
$\eta_3(LOR \times LUR)$	0.20	0.996 (0.852)	0.20	0.134 (0.998)
Neg. Loglik	933.922 (# Iter = 61)		933.922 (# Iter = 66)	

Parameter	Set 3		Set 4	
	SV	Estimate (SE)	SV	Estimate (SE)
β_0	0.80	0.886 (0.908)	0.90	0.886 (0.704)
$\beta_1(LD)$	−0.10	−0.017 (0.882)	−0.02	−0.017 (0.722)
$\beta_2(HD)$	−0.50	−0.484 (0.963)	−0.50	−0.484 (0.628)
$\beta_3(time)$	−0.20	−0.118 (0.593)	−0.10	−0.118 (0.345)
α_{12}	−0.08	−0.004 (0.956)	0.00	−0.004 (0.710)
α_{13}	−0.20	−0.010 (0.945)	−0.01	−0.010 (0.806)
α_{23}	−0.30	−0.111 (0.887)	−0.10	−0.111 (0.812)
α_{123}	−0.70	−0.511 (0.884)	−0.50	−0.511 (0.705)
α_1	−0.20	−0.103 (0.918)	−0.10	−0.103 (0.787)
α_2	−0.70	−0.649 (0.985)	−0.60	−0.649 (0.793)
η_{03}	−1.95	−14.338 (1.002)	−4.00	−13.092 (0.922)
η_{02}	−1.95	−14.481 (1.004)	−3.00	−13.235 (1.002)
η_{01}	−1.95	−2.089 (1.002)	−2.10	−2.089 (0.992)
$\eta_1(LOR)$	−0.50	−0.496 (1.001)	0.10	−3.387 (0.916)
$\eta_2(LUR)$	−0.10	12.982 (1.009)	2.00	11.736 (0.861)
$\eta_3(LOR \times LUR)$	0.10	0.782 (1.001)	−0.20	3.672 (0.929)
Neg. Loglik	933.922 (# Iter = 60)		933.922 (# Iter = 59)	

Table C.4: Results for Drop-out Model: TRT + LUR

Parameter	Set 1		Set 2	
	SV	Estimate (SE)	SV	Estimate (SE)
β_0	0.80	0.886 (0.202)	0.80	0.886 (0.927)
$\beta_1(LD)$	−0.10	−0.017 (0.195)	−0.02	−0.017 (0.995)
$\beta_2(HD)$	−0.50	−0.484 (0.192)	−0.50	−0.484 (0.995)
$\beta_3(time)$	−0.20	−0.118 (0.074)	−0.10	−0.118 (0.600)
α_{12}	−0.08	−0.004 (0.157)	0.00	−0.004 (0.988)
α_{13}	−0.20	−0.010 (0.156)	−0.01	−0.010 (0.981)
α_{23}	−0.30	−0.111 (0.168)	−0.10	−0.111 (0.981)
α_{123}	−0.70	−0.511 (0.171)	−0.50	−0.511 (0.976)
α_1	−0.20	−0.103 (0.207)	−0.10	−0.103 (1.000)
α_2	−0.70	−0.649 (0.215)	−0.60	−0.649 (0.997)
η_{01}	−1.95	−2.140 (0.151)	−2.10	−2.140 (1.001)
η_{02}	−1.95	−15.364 (0.862)	−2.00	−13.697 (1.001)
η_{03}	−1.95	−15.236 (0.807)	−3.00	−13.569 (1.001)
$\eta_1(LD)$	0.00	0.191 (0.205)	0.20	0.191 (1.002)
$\eta_2(HD)$	0.00	−0.051 (0.219)	−0.05	−0.051 (1.002)
$\eta_4(LUR)$	0.00	14.014 (0.876)	1.00	12.347 (1.021)
Neg. Loglik	933.910 (# Iter = 67)		933.910 (# Iter = 54)	

Appendix D

Detailed Results for the Selection Models Described in Section 7.4

Table D.1: Results for Case 1 in Table 7.14 Evaluated at the Boundary: $\eta_{03} \to -\infty$, $\eta_{02} \to -\infty$, $\eta_{03} + \eta_2 = \Delta_1$, $\eta_{02} + \eta_2 = \Delta_2$

Parameter	Estimate	SE
β_0	0.861	0.206
$\beta_1\ (LD)$	-0.007	0.196
$\beta_2\ (HD)$	-0.495	0.193
$\beta_3\ (time)$	-0.122	0.073
$\beta_4\ (gender)$	0.052	0.045
α_{12}	-0.002	0.164
α_{13}	-0.011	0.159
α_{23}	-0.111	0.172
α_{123}	-0.511	0.176
α_1	-0.094	0.208
α_2	-0.661	0.218
$\eta_1(LOR)$	0.286	0.275
Δ_1	-1.356	0.264
Δ_2	-1.499	0.262
η_{01}	-2.089	0.167
Neg. Loglik	933.244 (# Iter = 24)	

Table D.2: Results for Case 2 in Table 7.14 Evaluated at the Boundary: $\eta_{03} \to -\infty$, $\eta_{02} \to -\infty$, $\eta_{03} + \eta_2 = \Delta_1$, $\eta_{02} + \eta_2 = \Delta_2$

Parameter	Estimate	SE
β_0	0.894	0.205
β_1 (LD)	-0.016	0.193
β_2 (HD)	-0.483	0.190
β_3 $(time)$	-0.117	0.073
β_4 $(EDSS)$	-0.004	0.017
α_{12}	-0.004	0.157
α_{13}	-0.011	0.155
α_{23}	-0.111	0.168
α_{123}	-0.511	0.170
α_1	-0.101	0.207
α_2	-0.649	0.215
$\eta_1(LOR)$	0.286	0.251
Δ_1	-1.356	0.209
Δ_2	-1.499	0.237
η_{01}	-2.089	0.164
Neg. Loglik	933.901 (# Iter = 24)	

Table D.3: Results for Case 3 in Table 7.14 Evaluated at the Boundary: $\eta_{03} \to -\infty$, $\eta_{02} \to -\infty$, $\eta_{03} + \eta_2 = \Delta_1$, $\eta_{02} + \eta_2 = \Delta_2$

Parameter	Estimate	SE
β_0	0.877	0.205
β_1 (LD)	-0.025	0.196
β_2 (HD)	-0.484	0.194
β_3 ($time$)	-0.119	0.073
β_4 ($duration$)	0.002	0.003
α_{12}	0.000	0.163
α_{13}	-0.008	0.161
α_{23}	-0.108	0.173
α_{123}	-0.508	0.177
α_1	-0.110	0.208
α_2	-0.651	0.216
η_1(LOR)	0.286	0.279
Δ_1	-1.356	0.263
Δ_2	-1.499	0.265
η_{01}	-2.089	0.165
Neg. Loglik	933.768 (# Iter = 24)	

Table D.4: Results for Case 4 in Table 7.14 Evaluated at the Boundary: $\eta_{03} \to -\infty$, $\eta_{02} \to -\infty$, $\eta_{03} + \eta_2 = \Delta_1$, $\eta_{02} + \eta_2 = \Delta_2$

Parameter	Estimate	SE
β_0	0.745	0.241
β_1 (LD)	−0.025	0.195
β_2 (HD)	−0.486	0.185
β_3 (time)	−0.118	0.073
β_4 (age)	0.004	0.004
α_{12}	−0.001	0.162
α_{13}	−0.005	0.162
α_{23}	−0.110	0.174
α_{123}	−0.507	0.179
α_1	−0.108	0.208
α_2	−0.654	0.209
η_1(LOR)	0.286	0.241
Δ_1	−1.356	0.247
Δ_2	−1.499	0.232
η_{01}	−2.089	0.138
Neg. Loglik	933.354 (# Iter = 26)	

Table D.5: Results for Case 5 in Table 7.14 Evaluated at the Boundary: $\eta_{03} \to -\infty$, $\eta_{02} \to -\infty$, $\eta_{03} + \eta_2 = \Delta_1$, $\eta_{02} + \eta_2 = \Delta_2$

Parameter	Imputed Set 1		Imputed Set 2	
	Estimate	SE	Estimate	SE
β_0	0.718	0.253	0.726	0.251
$\beta_1(LD)$	0.008	0.193	0.006	0.195
$\beta_2(HD)$	-0.474	0.190	-0.475	0.192
$\beta_3(time)$	-0.108	0.075	-0.109	0.073
$\beta_4(log(BOD))$	0.020	0.016	0.019	0.016
α_{12}	-0.022	0.164	-0.021	0.162
α_{13}	-0.031	0.161	-0.029	0.159
α_{23}	-0.131	0.172	-0.130	0.171
α_{123}	-0.530	0.178	-0.529	0.175
α_1	-0.071	0.207	-0.074	0.210
α_2	-0.624	0.216	-0.626	0.218
$\eta_1(LOR)$	0.286	0.274	0.286	0.277
Δ_1	-1.356	0.264	-1.356	0.266
Δ_2	-1.499	0.263	-1.499	0.261
η_{01}	-2.089	0.164	-2.089	0.166
Neg. Loglik	933.088 (# Iter = 23)		933.215 (# Iter = 23)	

Parameter	Imputed Set 3		Imputed Set 4	
	Estimate	SE	Estimate	SE
β_0	0.717	0.251	0.725	0.249
$\beta_1(LD)$	0.009	0.190	0.006	0.195
$\beta_2(HD)$	-0.474	0.189	-0.475	0.189
$\beta_3(time)$	-0.108	0.075	-0.109	0.073
$\beta_4(log(BOD))$	0.020	0.016	0.019	0.016
α_{12}	-0.022	0.160	-0.021	0.160
α_{13}	-0.031	0.158	-0.029	0.155
α_{23}	-0.131	0.172	-0.129	0.167
α_{123}	-0.530	0.173	-0.529	0.171
α_1	-0.071	0.204	-0.073	0.210
α_2	-0.624	0.215	-0.625	0.214
$\eta_1(LOR)$	0.286	0.278	0.286	0.283
Δ_1	-1.356	0.266	-1.356	0.268
Δ_2	-1.499	0.264	-1.499	0.267
η_{01}	-2.089	0.166	-2.089	0.166
Neg. Loglik	933.083 (# Iter = 23)		933.211 (# Iter = 23)	

Table D.6: Results for Case 5 in Table 7.14 Evaluated at the Boundary (364 patients): $\eta_{03} \to -\infty$, $\eta_{02} \to -\infty$, $\eta_{03} + \eta_2 = \Delta_1$, $\eta_{02} + \eta_2 = \Delta_2$

Parameter	Imputed with 1.0		Imputed with 4.5	
	Estimate	SE	Estimate	SE
β_0	0.686	0.248	0.694	0.249
$\beta_1(LD)$	0.063	0.193	0.061	0.193
$\beta_2(HD)$	−0.446	0.196	−0.447	0.192
$\beta_3(time)$	−0.108	0.074	−0.108	0.074
$\beta_4(log(BOD))$	0.020	0.016	0.020	0.016
α_{12}	−0.061	0.166	−0.060	0.160
α_{13}	−0.059	0.162	−0.058	0.159
α_{23}	−0.161	0.175	−0.160	0.172
α_{123}	−0.567	0.180	−0.565	0.176
α_1	−0.022	0.208	−0.025	0.206
α_2	−0.587	0.220	−0.589	0.212
$\eta_1(LOR)$	0.301	0.274	0.301	0.276
Δ_1	−1.342	0.264	−1.342	0.264
Δ_2	−1.491	0.261	−1.491	0.263
η_{01}	−2.149	0.168	−2.149	0.171
Neg. Loglik	914.912 (# Iter = 24)		915.047 (# Iter = 26)	

Table D.7: Results for Model ID2 Evaluated at the Boundary (364 patients): $\eta_{03} \to -\infty$, $\eta_{02} \to -\infty$, $\eta_{03} + \eta_2 = \Delta_1$, $\eta_{02} + \eta_2 = \Delta_2$

Parameter	Estimate	SE
β_0	0.859	0.206
$\beta_1(LD)$	0.035	0.193
$\beta_2(HD)$	−0.457	0.201
$\beta_3(time)$	−0.118	0.074
α_{12}	−0.041	0.171
α_{13}	−0.037	0.161
α_{23}	−0.140	0.174
α_{123}	−0.546	0.179
α_1	−0.056	0.207
α_2	−0.614	0.224
$\eta_1(LOR)$	0.301	0.276
Δ_1	−1.342	0.250
Δ_2	−1.491	0.266
η_{01}	−2.149	0.170
Neg. Loglik	915.825 (# Iter = 21)	

Appendix E

Detailed Results for the Liu *et al.* Transition Models Described in Section 7.6

Table E.1: Results for Liu Transition Model with Drop-out Model ID1

Parameter	SV	Estimate (SE)	SV	Estimate (SE)
β_0	0.89	1.007 (0.558)	1.00	1.007 (0.447)
β_1	-0.12	-0.040 (0.876)	-0.04	-0.040 (0.475)
β_2	-0.50	-0.462 (1.336)	-0.50	-0.462 (0.372)
β_3	-0.42	-0.324 (0.369)	-0.30	-0.324 (0.404)
β_4	0.90	0.692 (0.707)	0.70	0.692 (0.361)
η_{03}	-1.95	-12.083 (2.727)	-2.00	-10.451 (1.197)
η_{13}	0.00	0.558 (1.013)	0.60	0.558 (0.767)
η_{23}	0.00	10.535 (0.888)	2.00	8.903 (0.970)
η_{02}	-1.95	-18.645 (1.144)	-1.00	-22.807 (0.746)
η_{12}	0.00	0.048 (2.246)	0.05	0.048 (0.646)
η_{22}	0.00	17.318 (0.913)	2.00	21.480 (0.746)
η_{01}	-1.95	-2.089 (0.828)	-2.00	-2.089 (0.363)
Neg. Loglik	942.259 (# Iter = 137)		942.259 (# Iter = 97)	

Table E.2: Results for Liu Transition Model with Drop-out Model ID2

Parameter	SV	Estimate (SE)	SV	Estimate (SE)
β_0	0.89	1.007 (0.817)	1.00	1.007 (0.873)
β_1	-0.12	-0.040 (0.748)	-0.04	-0.040 (0.710)
β_2	-0.50	-0.462 (0.604)	-0.50	-0.462 (0.625)
β_3	-0.42	-0.324 (0.483)	-0.30	-0.324 (0.481)
β_4	0.90	0.692 (0.631)	0.70	0.692 (0.697)
η_{03}	-1.95	-14.182 (0.991)	-2.00	-14.175 (0.908)
η_{02}	-1.95	-14.324 (1.388)	-1.00	-14.318 (0.909)
η_{01}	-1.95	-2.089 (0.924)	-2.00	-2.089 (0.886)
η_1	0.00	0.286 (0.710)	0.30	0.286 (0.415)
η_2	0.00	12.826 (0.812)	3.00	12.819 (0.594)
Neg. Loglik	942.687 (# Iter = 58)		942.687 (# Iter = 61)	

Table E.3: Results for Liu Transition Model with Drop-out Model ID3

Parameter	SV	Estimate (SE)	SV	Estimate (SE)
β_0	0.89	1.123 (0.211)	1.10	1.123 (0.212)
β_1	−0.12	−0.128 (0.178)	−0.13	−0.128 (0.176)
β_2	−0.50	−0.437 (0.174)	−0.40	−0.437 (0.173)
β_3	−0.42	−0.443 (0.102)	−0.50	−0.443 (0.103)
β_4	0.90	0.573 (0.172)	0.60	0.573 (0.177)
η_0	−1.95	−2.023 (0.163)	−2.00	−2.023 (0.161)
η_1	0.00	0.542 (0.311)	0.50	0.542 (0.313)
η_2	0.00	−0.262 (0.610)	−0.30	−0.262 (0.609)
Neg. Loglik	939.471 (# Iter = 17)		939.471 (# Iter = 15)	

Table E.4: Results for Liu Transition Model with Drop-out Model ID5

Parameter	SV	Estimate (SE)	SV	Estimate (SE)
β_0	0.89	1.007 (0.931)	1.10	1.007 (0.929)
β_1	−0.12	−0.040 (1.026)	−0.03	−0.040 (0.994)
β_2	−0.50	−0.462 (0.984)	−0.40	−0.462 (0.992)
β_3	−0.42	−0.324 (0.463)	−0.30	−0.324 (0.485)
β_4	0.90	0.692 (0.975)	0.60	0.692 (0.971)
η_{03}	−1.95	−14.704 (0.863)	−1.00	−13.967 (0.970)
η_{02}	−1.95	−14.832 (0.972)	−0.00	−14.096 (0.894)
η_{01}	−1.95	−2.089 (1.028)	−1.00	−2.089 (0.992)
η_2	0.00	13.539 (0.742)	0.00	12.803 (0.570)
Neg. Loglik	943.239 (# Iter = 55)		943.239 (# Iter = 57)	

Table E.5: Results for Liu Transition Model with Random Drop-out (RD)

Parameter	RD1		RD2		RD3	
	Estimate	SE	Estimate	SE	Estimate	SE
β_0	1.113	0.206	1.113	0.209	1.113	0.212
β_1	-0.118	0.170	-0.118	0.172	-0.118	0.175
β_2	-0.445	0.165	-0.445	0.172	-0.445	0.173
β_3	-0.431	0.096	-0.431	0.099	-0.431	0.099
β_4	0.596	0.160	0.596	0.164	0.596	0.169
η_{03}	-2.416	0.327	-2.239	0.248	–	–
η_{02}	-2.117	0.297	-2.278	0.261	–	–
η_{01}	-2.089	0.108	-2.089	0.163	–	–
η_0	–	–	–	–	-2.068	0.132
η_{13}	0.878	0.367	–	–	–	–
η_{12}	0.401	0.337	–	–	–	–
η_1	–	–	0.625	0.261	0.432	0.195
Neg. Loglik	944.101 (# Iter = 21)		944.518 (# Iter = 17)		939.578 (# Iter = 14)	

Table E.6: Results for Liu Transition Model with Drop-out Completely At Random (CRD)

Parameter	CRD1		CRD2	
	Estimate	SE	Estimate	SE
β_0	1.113	0.209	1.113	0.209
β_1	-0.118	0.174	-0.118	0.172
β_2	-0.445	0.172	-0.445	0.172
β_3	-0.431	0.099	-0.431	0.099
β_4	0.596	0.169	0.596	0.164
η_{03}	-1.846	0.172	–	–
η_{02}	-1.849	0.160	–	–
η_{01}	-2.089	0.166	–	–
η_0	–	–	-1.880	0.096
Neg. Loglik	947.589 (# Iter = 13)		942.074 (# Iter = 11)	