INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

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

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

ProQuest Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

Analysis of Binary Longitudinal Data with Dropout and Death

Brenda F. Kurland

A dissertation submitted in partial fulfillment
of the requirements for the degree of

Doctor of Philosophy

University of Washington

2002

Program Authorized to Offer Degree: Public Health and Community Medicine – Biostatistics

UMI Number: 3072104

ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

In presenting this dissertation in partial fulfillment of the requirements for the Doctoral degree at the University of Washington, I agree that the Library shall make its copies freely available for inspection. I further agree that extensive copying of this dissertation is allowable only for scholarly purposes, consistent with "fair use" as prescribed in the U.S. Copyright Law. Requests for copying or reproduction of this dissertation may be referred to Bell and Howell Information and Learning, 300 North Zeeb Road, Ann Arbor, MI 48106-1346, to whom the author has granted "the right to reproduce and sell (a) copies of the manuscript in microform and/or (b) printed copies of the manuscript made from microform."

Signature _____

Date 11/21/02

University of Washington

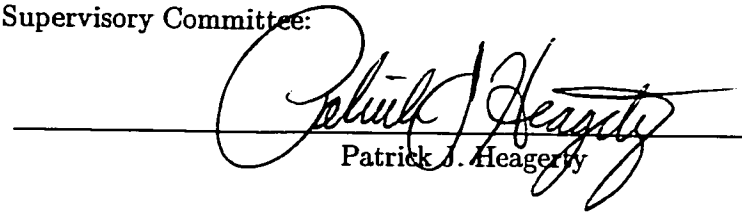
Graduate School

This is to certify that I have examined this copy of a doctoral dissertation by

Brenda F. Kurland

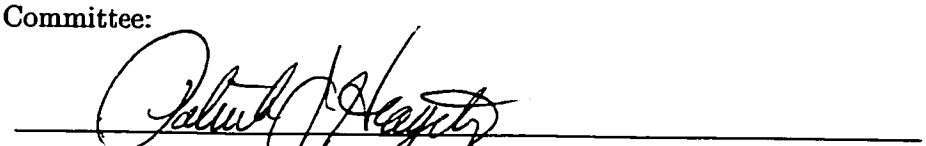
and have found that it is complete and satisfactory in all respects,
and that any and all revisions required by the final
examining committee have been made.

Chair of Supervisory Committee:

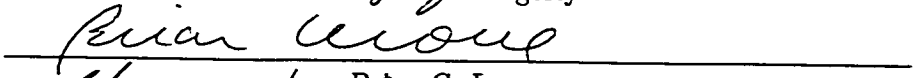


Patrick J. Heagerty

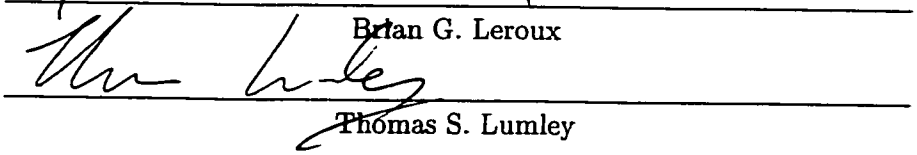
Reading Committee:



Patrick J. Heagerty



Brian G. Leroux



Thomas S. Lumley

Date: November 20, 2002

University of Washington

Abstract

Analysis of Binary Longitudinal Data with Dropout and Death

by Brenda F. Kurland

Chair of Supervisory Committee:

Professor Patrick J. Heagerty
Biostatistics

Dropout (attrition) is a common challenge in analysis of longitudinal data. Additionally, data records may be truncated due to the death of study participants. Response data missing due to death are often modeled by the same process as data missing due to attrition. This dissertation examines dropout and death as separate processes, with different origins and different impact on regression models for longitudinal binary data. We first extend the likelihood-based marginalized transition model (MTM) of Heagerty (2002) to accommodate monotone nonignorable dropout. Simulation studies demonstrate that the MTM displays advantages in misspecification bias and efficiency, compared to inverse probability of censoring weighted generalized estimating equations (Robins et al., 1995), a semiparametric method with comparable regression and selection models. The second section of the dissertation considers Direct Estimation Conditioning on being ALive (DECAL), for longitudinal binary data with both death and monotone dropout. The target of estimation is the mean response value, given that the subject is alive at the response time. Likelihood-based methods are difficult to parameterize directly in terms of this target. Due to implicit imputation beyond time of death, naively employed likelihood-based methods may yield biased parameter estimates. Independence estimating equations, possibly incorporating selection weights, are described for DECAL regression.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I gratefully acknowledge the graduate study financial support provided by a T32 NHLBI Predoctoral Training Grant and the ARCS Foundation. The PEP data were generously provided by Dr. Thomas Gill of Yale University.

My advisor, Patrick Heagerty, is a patient tutor. Many times I would plod toward his office thinking only of my frustration. Within a minute I would be happily immersed in discussion of statistical and technical details. Committee members Thomas Lumley and Brian Leroux have provided sage council at crucial times.

My husband, Mitchell Dernis, has been a wonderful partner for my entire adult life. He provides round-the-clock computing support at reasonable rates, and indulges other erratic requests. Avishai Meir Kurland Dernis has only been around for a small part of my studies, but his fierce determination has been an inspiration during the home stretch: almost anything can be accomplished, though some panting and squealing may be necessary. My parents are unyielding in their confidence in me, and my mother is invaluable as the "Granny Nanny".

I thank office mates Do Peterson, Manisha Desai, Chau Thach, Katherine Guthrie, Laura Lee Johnson, Yingye Zheng, Jon Schildcrout, and Jinbo Chen – as well as many unnamed friends and family – who have lent their support, especially through well-timed wry comments on the nature of scientific innovation, or whatever crisis nagged me.

Chapter 1

INTRODUCTION

This dissertation develops and evaluates methods for the analysis of binary longitudinal data under two common restrictions: monotone dropout, and truncation of followup due to death. A likelihood-based method, the marginalized transition model (Heagerty, 2002) is extended to analyze data with nonignorable monotone dropout. For data with both dropout and death, likelihood-based methods and generalized estimating equations (Liang & Zeger, 1986) with non-independence working correlation may be biased for regression defined as conditional means given subjects' being alive. This chapter introduces two motivating examples: the PANSS Schizophrenia Data, a clinical trial with monotone dropout, and the Precipitating Events Project (PEP), an observational study with both death and monotone dropout.

1.1 Example #1: PANSS Schizophrenia Data

In longitudinal studies, attrition is undesirable but often inevitable. Study participants may move out of town, become disenchanted with a study, or feel too sick (or too healthy) to want to participate. Relevance of the dropout process in regression modeling is apparent in Figure 1.1, which shows fitted proportions of subjects showing clinical improvement in the three arms of a double-blind longitudinal study comparing treatments for schizophrenia (Diggle, 1998). Quadratic longitudinal models for each treatment group are fitted using all available data ("available case") or only subjects who completed all 5 followup assessments ("complete case"). The sample size for the available case analysis starts at 420 for the three groups combined, but decreases to 227 over time, as shown in Figure 1.1. The complete case analysis uses data from the 227 "completers" at all timepoints. Time trends and overall

treatment effects seem similar for the two models. However, the fitted proportion showing improvement is consistently higher for participants who complete the study. The difference in fitted values is greatest for the placebo group, which has the highest rate of attrition.

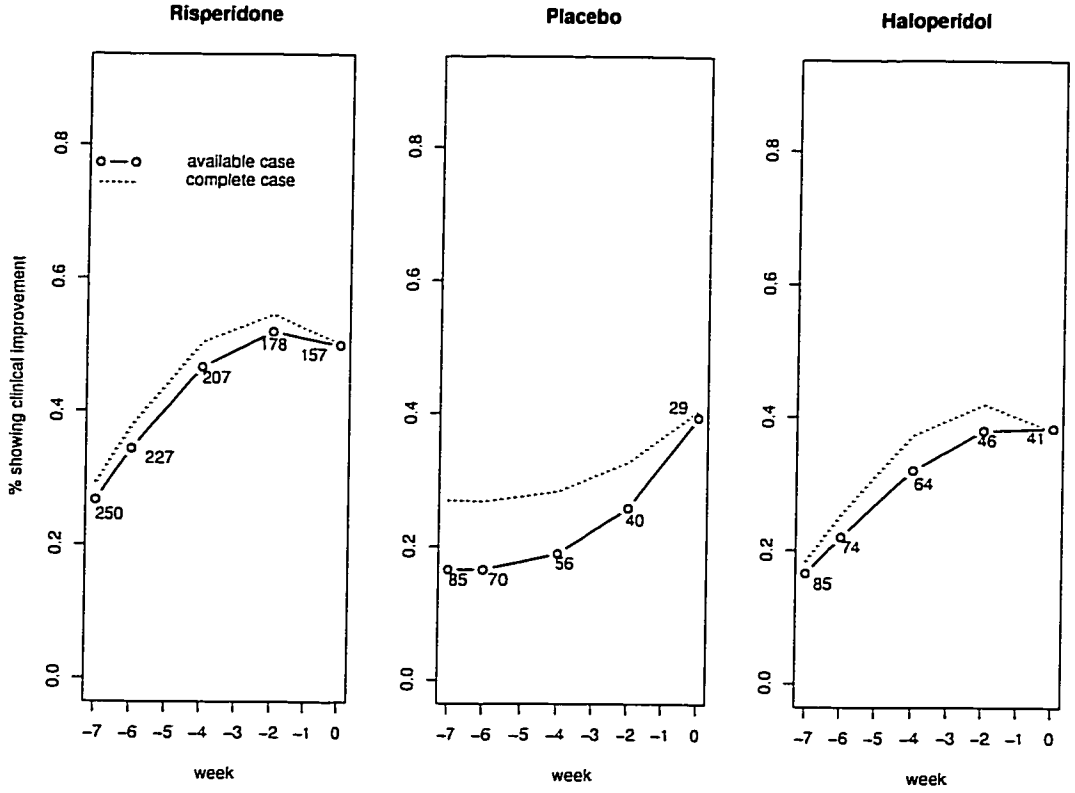


Figure 1.1: Percent improvement over time, by treatment group. PANSS schizophrenia data, quadratic time-by-treatment group interaction. Sample size for each timepoint is labeled for the available case analysis.

Figure 1.2 also illustrates the potential impact of dropout on data analysis, by showing time trends in clinical improvement separately for completers and for participants who drop out after 1, 2, 4, or 6 weeks of treatment. Patients who remain in the study for the entire 8 weeks are more likely to show improvement at almost every timepoint. Furthermore, the proportion of subjects showing improvement appears to decrease in the assessment preceding dropout. Do the panels in Figure 1.1 reflect real treatment differences, or artifacts of selectively removing subjects from the study via attrition? Characterizing the missingness

model and choosing regression methods that are valid under different selection scenarios may be necessary to address the primary scientific question regarding treatment differences in clinical improvement.

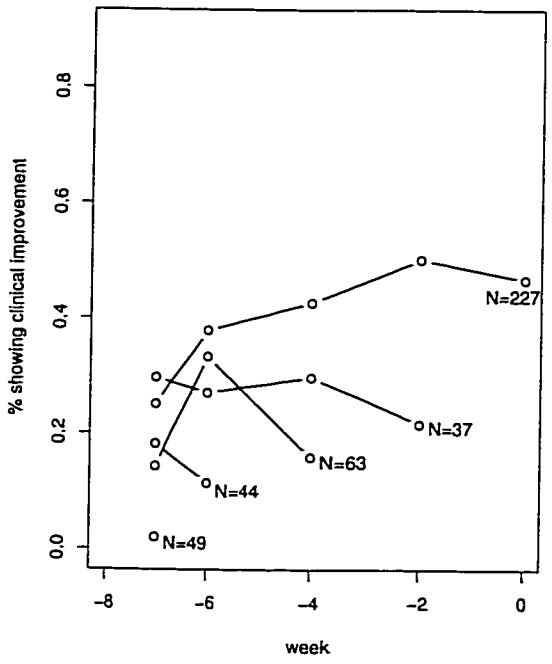


Figure 1.2: Percent showing clinically significant improvement over time, for dropout after 1, 2, 4, 6, or 8 weeks of treatment, PANSS schizophrenia data.

1.1.1 Target of Inference for Data with Dropout

To evaluate the effectiveness of risperidone, haloperidol, and placebo in treatment of schizophrenia symptoms, we introduce a regression model for treatment response over the 8 weeks of the study. Let Y_{it} be a binary (0/1) variable representing whether or not subject i shows clinically significant improvement in symptoms at time t, compared to baseline, according to the Positive and Negative Syndrome Scale (PANSS). Treatment covariates `PLAC` (1 = placebo, 0 otherwise) and `RISP` (1 = risperidone, 0 otherwise) are also binary, and

provide contrasts to the reference group, haloperidol. A logit-linear model, quadratic in time, for PANSS scores could be:

$$\begin{aligned}\text{logit}[P(\mathrm{Y}_{it}=1)] &= \beta_0+\beta_1\cdot\texttt{PLAC}_i+\beta_2\cdot\texttt{RISP}_i+\beta_3\cdot\texttt{week}_{it}+\beta_4\cdot\texttt{week}_{it}^2\\ &\quad+\beta_5\cdot\texttt{PLAC}_i\cdot\texttt{week}_{it}+\beta_6\cdot\texttt{PLAC}_i\cdot\texttt{week}_{it}^2+\beta_7\cdot\texttt{RISP}_i\cdot\texttt{week}_{it}+\beta_8\cdot\texttt{RISP}_i\cdot\texttt{week}_{it}^2.\end{aligned}$$

If week is represented as (-7, -6, -4, -3, -2, 0), parameters β_1 and β_2 compare placebo and risperidone to haloperidol at the end of the study period, and $(\beta_5, \beta_6, \beta_7, \beta_8)$ compare the time trends among the three treatments. The interpretation of treatment contrast β_2 is:

> $\exp(\beta_2)$ = odds ratio for clinically significant improvement after 8 weeks of treatment, for subjects assigned to risperidone compared to those assigned to haloperidol.

Several assumptions underlie this routine interpretation. First, we assume an "intent-to-treat analysis" (Lee et al., 1991) comparing patients *assigned* to the two treatments regardless of whether patients complied with taking the medication. The estimated treatment contrast would probably be different for universal versus imperfect compliance, or if treatment group were made time-varying based on compliance. The intent-to-treat approach maintains randomization, which protects inference in clinical trials from the influence of unmeasured confounders.

Another assumption is that all subjects are included in comparisons at each timepoint. However, some PANSS scores are missing due to dropout. If only complete cases are considered, $\exp(\beta_2)$ is an odds ratio comparing improvement for patients assigned to risperidone and haloperidol *who remain in the study for all 8 weeks.* Many of the dropouts in this trial are due to inadequate response: the assigned treatment did not appear to work, so patients were moved back to their standard treatment. These treatment failures should not be ignored in the PANSS analysis. Chapter 2 describes how analysis methods explicitly or implicitly impute missing responses for subjects who drop out of a longitudinal study. Choice of regression model affects the targets of inference that may be addressed.

1.1.2 Proposed Method: Marginalized Transition Model for Binary Longitudinal Data with Monotone Dropout

The first several chapters of this dissertation extend the marginalized transition model of Heagerty (2002) to accommodate monotone dropout. The model has many features that contribute to ease of modeling and interpretation for clinical trials such as the PANSS study. The regression model allows for treatment comparisons that are not dependent on subject-specific characteristics or dropout pattern. The association among longitudinal binary responses is induced by a "transition model", in which dependence of responses on prior responses is modeled directly. The transition model is "marginalized" so that the regression model used for structuring treatment comparisons does not directly control for past response values. Finally, dropout (like association) is modeled using covariates and prior or current responses as predictors of dropout. Choice of dropout model may influence the estimates of the regression model. In Chapter 3 we formally develop the marginalized transition model (Heagerty, 2002) to accommodate dropout models. Chapter 4 compares the model to inverse probability of censoring weighted GEE (IPCW-GEE, Robins et al., 1995) with respect to efficiency and misspecification bias, and Chapter 5 compares the MTM and IPCW-GEE in analysis of the PANSS data.

1.2 Example #2: PEP Disability Data

The Precipitating Events Project (PEP) is a prospective cohort study of disability in 754 elderly (> 70 years) people in New Haven, CT. Upon entry into the study, participants did not live in a nursing home or other institutionalized care setting, and planned to remain in the New Haven area. Disability, defined as needing assistance in any of 4 key activities of daily living (bathing, dressing, walking within the home, and getting in and out of a chair), was an exclusion criterion at baseline. Enrollment in the study was stratified by low, intermediate, and high risk for activities of daily living (ADL) disability, using a model developed in an earlier study (Gill et al., 1999). Risk group was determined by a rapid gait test, the Folstein Mini-Mental State Examination (MMSE, Folstein et al., 1975), and age.

Disability status is assessed each month by self-report in a telephone interview. For this

analysis, we model the trajectory of disability over 24 months from baseline. Let Y_{it} be a binary random variable taking value 1 if subject i is disabled t months after baseline, and value 0 if not disabled. In principle, we want to model

$$E(Y_{it}) = \text{probability that subject } i \text{ is disabled at } t \text{ months after baseline}$$

as a function of time and, perhaps, other variables. Section 1.1 discusses the complications of modeling longitudinal binary data with dropout. This analysis introduces another element: censoring due to death. Seventy-two months of followup are planned for the PEP study. By 24 months after baseline, 64 of 754 subjects (8.5%) had died, and 19 (2.5%) were "permanent refusers" who had dropped out of the study. Two of the permanent refusers are among the 64 decedents. Survival information for all participants (including permanent refusers) is obtained through family contacts and review of local obituaries (Gill et al., 2002). The second major section of the dissertation, motivated by the PEP data, discusses targets of inference for binary longitudinal data in the presence of dropout and death, and appropriate methods of analysis.

1.2.1 Target of Inference for Data with Deaths

Death and dropout are often treated as comparable sources of missing data:

> For the purposes of this analysis, we follow Robins, Rotnitzky, and Zhao (1995) and regard nonresponse due to death as no different from other types of nonresponse. (Rotnitzky et al., 1998, p. 1322)

However, the accommodation of responses missing due to dropout and death may be quite different. The analysis objective with data missing due to dropout is to recover statistically the contribution of the missing observations. Using the notation introduced above, and R_{it} as a dichotomous (binary) variable for retention (whether or not Y_{it} was observed at month t), the target of inference is $E(Y_{it})$, not $E(Y_{it}|R_{it} = 1)$: the expected value of the response, not the expected value of the response given that the response was observed. The diligence of data collectors should not affect the target of inference.

By contrast, in most applications it is not desirable to "recover" values that are missing due to death. Many outcome variables (such as blood pressure) have no clear meaning when the subject is deceased. Others, such as disability or quality of life measures, are sometimes imputed as the lowest score available for deceased subjects. This approach has been questioned: it can be argued that in some cases death is preferable to being alive and in agony. Also, if the response is an instrument with well-established psychometric properties, such as the SF-36, assigning values to deceased subjects will invalidate the scaling of the instrument (Troxel, 1998; Diehr et al., 2001). Instead of imputing a low value for deceased subjects, some researchers impute data to recover values as if the subject were alive (Revicki et al., 2001). However, many acknowledge the special status of death (Brayne et al., 1999; Miller et al., 2001).

A common approach to analyzing longitudinal data in which deaths occur is to exclude deceased subjects, or to analyze their data separately. This approach may bias results (Diehr et al., 1995; McHorney, 1996). For example, in a randomized clinical trial where one treatment kills weaker participants, the treatment may look successful if only stronger, surviving patients are evaluated. Analyzing deceased and nondeceased subjects separately is especially suspect in prospective research, where survivors and non-survivors cannot be identified from the beginning of the study. Quality of life studies can sometimes modify the outcome variable to accommodate death as an outcome. For example, Diehr et al. (1995) suggest "probability that a person will be healthy in two years" as a health-related outcome that incorporates death in an intuitive, undisputed manner. Whereas death can sometimes be incorporated into definitions of quality of life outcomes, other responses (such as blood pressure or, arguably, disability) experience death as a censoring mechanism, rather than part of the response. Survival and response processes may be modeled jointly (De Gruttola & Tu, 1994; Gray & Brookmeyer, 2000; Ribaudo et al., 2000). However, for the PEP example, censoring by survival status is part of the regression model, but otherwise is a nuisance factor.

Likelihood-based methods implicitly impute missing values, as illustrated by Diggle (1998) for continuous responses. While desirable for data missing due to dropout, this feature of likelihood-based models may be misleading when data are missing due to death.

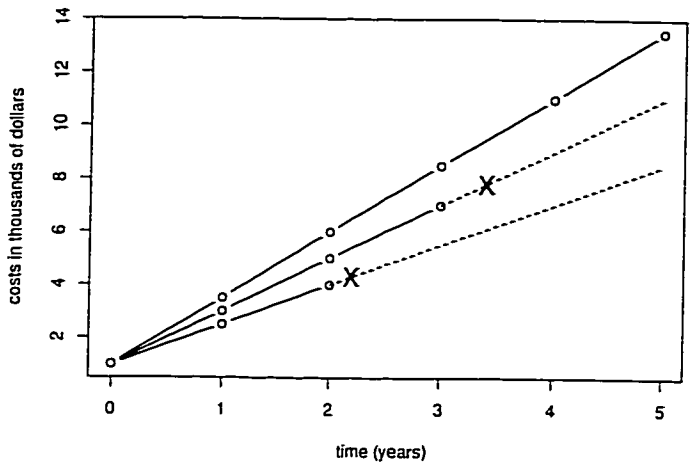


Figure 1.3: Illustration of implicit imputation for deceased subjects in random slopes model.

For example, consider Figure 1.3, a random slopes model of hypothetical annual health-care costs (in thousands of dollars) following a diabetes diagnosis. Solid lines connect observed data points (hollow dots). Dotted lines show implicit imputation beyond deaths (marked by "x"). The random slopes model estimates the average patient cost four years after diagnosis as $E(Y_4) \approx (7+9+11)/3 = \$9000$. An insurance administrator is interested in the total cost to the insurance company: the average cost for a similar group of three patients is $E(Y_4) \approx (0+0+11)/3 \approx \$3667$. Finally, a family planning finances would want to know the average cost, given that the patient survives: $E(Y_4|S > 4) = \$11{,}000$ where S is the time of death. The random slopes expected value does not correspond to the target of inference for either the administrator or the patient.

1.2.2 Direct Methods for Analysis of Longitudinal Binary Data with Deaths

Table 1.1 shows four sample records from the PEP data. The monthly entry is "1" if a participant reports any of 4 ADL disabilities, and "0" if no disability is reported. The symbol "." denotes a month with missing data due to intermittent or permanent dropout. Symbol "x" appears in each month after the study participant has died. None of the four

Table 1.1: Sample data records for PEP study. (x=deceased, .=missing)

ID	ADL months 1-24
2452	000001000000xxxxxxxxxxx
2049	000000000000000001.....
1054	00000000011....xxxxxxxxx
2663	0000000000.1.111101110.0

subjects has 24 months of followup. However, the data for ID 2452 may be considered complete, since values are only missing due to death. The target of inference for the PEP analysis of disability is:

> The probability that subject i is disabled at t months after baseline, given that subject i is still alive at that time.

Or, using notation with Y_{it} as response (as above) and $S_i = s_i$ as the month in which subject i dies:

$$E(Y_{it}|S_i > t).$$

At month 19, 2 subjects in Table 1.1 are deceased, and two are disabled. A simple estimate of $E(Y_{it}|S_i > t)$ for month 19 is the average value for subjects who are living: 100% disability. However, for month 20 the interview for 2049 is missing. Analysis of PEP data must account for missing observations, while preserving a target of inference that remains interpretable by conditioning on survival.

A regression model for direct estimation of $E(Y_{it}|S_i > t)$ must take an appropriate functional form. Regression for $E(Y_{it}|S_i > t)$ may be linear in time, whereas $E(Y_{it}|S_i = s_i)$ or $E(Y_{it})$ may have an entirely different shape. The reverse could also be true, as demonstrated in Figure 1.4. In this example, survival time $S_i = s_i$ indicates that subject i dies at a time shortly following time s_i. (In later chapters, $S_i = s_i$ denotes a subject who dies shortly *before* time s_i.) The linear regression model is

$$E(Y_{it}|S_i = s_i) = s_i \cdot t$$

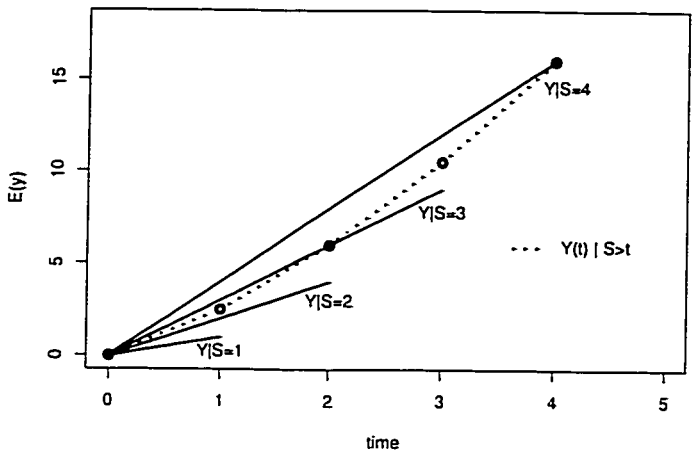


Figure 1.4: Illustration of different functional forms for $E(Y_{it}|S_i = s_i)$ and $E(Y_{it}|S_i > t)$.

for $t = (0, 1, 2, 3, 4)$. $P(S = s) = \frac{1}{4}$ for $S = (1, 2, 3, 4)$. Trajectories are shown for 4 survival times in Figure 1.4, as well as a pointwise sketch of target of inference $E(Y_{it}|S_i > t)$, computed as

$$E(Y_{it}|S_i > t) = \frac{\sum_{s \geq t} E(Y_{it}|S_i = s_i)P(S_i = s_i)}{\sum_{s \geq t} P(S_i = s_i)}.$$

The functional form for $E(Y_{it}|S_i = s_i)$ is linear, while $E(Y_{it}|S_i > t)$ clearly is not. For data with both death and dropout (especially dropout), the observed data may not reveal the functional form of the regression model of interest.

Chapter 6 explores the extent of implicit or explicit imputation of response values for deceased subjects, in likelihood-based methods and in semiparametric estimating equation-based methods (GEE, Liang & Zeger, 1986). Methods are identified in which the target of inference, $E(Y_{it}|S_i > t)$, can only be modeled by stratifying on the survival time, S_i. Chapter 7 derives cases in which the target of inference, $E(Y_{it}|S_i > t)$, can be estimated directly, under monotone dropout, without needing to stratify by or fit the survival distribution. Finally, Chapter 8 illustrates direct methods through analysis of the PEP data, and explores bias in methods that do not distinguish between missing data due to dropout and death. Chapter 9 gives a summary and plan for future research.

1.3 Summary of Dissertation Research

The two sections of this dissertation illustrate how vastly different analysis methods may be preferable under different incomplete data situations. In the case of binary longitudinal data with dropout, such as the PANSS data, a likelihood-based model such as the marginalized transition model compares favorably to inverse probability of censoring weighted GEE (Robins et al., 1995), because of efficiency gains and the robustness of likelihood-based methods under some common dropout patterns. However, for data in which both death and dropout occur (such as the PEP data) we advocate using IPCW-GEE with independence as the working correlation. This approach can be inefficient, but we show that it is the only method examined that will yield direct estimation of the expected value of the response, conditional on survival.

Chapter 2

EXISTING METHODS FOR ANALYSIS OF LONGITUDINAL BINARY DATA WITH DROPOUT

In this chapter we review existing methods for analysis of longitudinal binary data with dropout. After introducing notation, we briefly review approaches to longitudinal binary data and dropout separately. We then survey recent literature for models that could apply to clinical trials such as the PANSS schizophrenia data in Chapter 1.

2.1 Notation

Some notation is necessary to describe methods for analyzing binary longitudinal data with dropout. Models for binary longitudinal data characterize the distribution of values for a response vector, $\mathrm{Y}_i = (\mathrm{Y}_{i1} \ldots \mathrm{Y}_{in_i})$ for individuals $i = 1, \ldots, N$. For the PANSS data, each Y_i is a vector of 0/1 values, with each value reflecting whether or not subject i showed improvement at time t, compared to baseline. The relationship between Y_i and covariates is described by the $(n_i \,\mathrm{x}\, p)$ design matrix $\mathbf{X}_i$ and the regression parameter p-vector β. We restrict ourselves to a logit-linear model:

$$\mathrm{logit}(\mu_{it}^M) = \mathrm{X}_{it}\beta \tag{2.1}$$

where $\mu_{it}^M = \mathrm{E}(\mathrm{Y}_{it}|\mathrm{X}_{it})$ is the probability of positive response for subject i at time t, given covariate vector X_{it}. Serial correlation among members of vector Y_i is parameterized by vector α. The functional form for association (correlation) models and estimation of α differ for the different methods considered, and are discussed in more detail as each method is introduced.

The dropout process is described by a retention vector $\mathrm{R}_i = (\mathrm{R}_{i1} \ldots \mathrm{R}_{in_i})$, where R_{it} is an indicator taking value 1 if observation Y_{it} is observed and 0 if Y_{it} is missing. R_i is modeled by the $(n_i \,\mathrm{x}\, r)$ design matrix for dropout, $\mathbf{D}_i$, and parameter r-vector ϕ. $\mathbf{D}_i$ may include

responses (Y_i), covariates in $\mathbf{X}_i$, and covariates not in $\mathbf{X}_i$. An additional design matrix and parameter vector ($\mathbf{G}_i$ and λ) are introduced below for certain dropout models. Dropout parameter vector ϕ is assumed to be distinct from regression and association parameter vectors β and α. Response vector Y_i can be divided into two vectors, based on whether values are observed or missing: $Y_i = (Y_i^o, Y_i^m)$.

2.2 Binary Longitudinal Data

Three commonly recognized approaches to regression analysis of binary longitudinal data are generalized linear mixed models (Breslow & Clayton, 1993), transition models (Cox & Snell, 1989), and generalized estimating equations (Liang & Zeger, 1986). Generalized linear mixed models specify the full likelihood using regression models with fixed and random effects. Transition models generate association among responses by including prior responses in the regression model. The semiparametric generalized estimating equations (GEE) method incorporates a regression model and working correlation into a quasi-score equation. In this section we briefly review each method, with emphasis on target of inference in the regression model.

The generalized linear mixed model assumes that correlation within a subject's vector of responses, Y_i, is induced via latent variables (random effects). Regression contrasts condition on the values of subject-specific random effects. For example, the regression equation for a random intercept model is:

$$\text{logit}(\mu_{it}) = X_{it}\beta + b_i$$

where for the tth observation on subject i, $\mu_{it} = E(Y_{it}|X_{it}, b_i)$. Random intercepts b_i are typically assumed to have a normal distribution with mean zero.

Transition models introduce correlation among responses by including prior responses in the regression model. Other regression parameters must then be interpreted holding prior response values constant. For example, interpretation of treatment effect β_1 will differ in the following two transition models:

$$\text{logit}(\mu_{it}) = \beta_0 + \beta_1 \texttt{treat}_i + \beta_2 Y_{it-1} \tag{2.2}$$

$$\text{logit}(\mu_{it}) \quad = \quad \beta_0 + \beta_1 \texttt{treat}_i + \beta_2 \text{Y}_{it-1} + \beta_3 \text{Y}_{it-2}. \tag{2.3}$$

In Equation 2.2, $\exp(\beta_1)$ is the odds ratio comparing $\texttt{treat}_i = 1$ and $\texttt{treat}_i = 0$, for subjects with the same value of Y_{it-1}. In Equation 2.3, the contrast is conditional on holding both Y_{it-1} and Y_{it-2} constant. The order of the transition model changes the interpretation of β_1.

For both generalized linear mixed models and transition models, modeling of association among responses affects the regression model. Inference regarding treatment differences in a clinical trial generally desire a "population-averaged" (Zeger et al., 1988) or "marginal" (Heagerty & Zeger, 2000) regression contrast, such as Equation 2.1, where covariates do not include responses or random effects. Generalized estimating equations (Liang & Zeger, 1986) estimate marginal regression parameters. Association among responses is introduced by a working correlation matrix, or by odds ratios for binary response (Lipsitz et al., 1991). If no data are missing, regression parameter estimates are consistent even when the association structure is misspecified. However, estimation can be inefficient compared to likelihood-based methods (Fitzmaurice et al., 1993), and missing data may bias parameter estimates.

2.3 Dropout

Missing data are common in longitudinal studies, and much research has been dedicated to the impact of missing data, and how to incorporate missingness into analysis (Little & Rubin, 1987; Little, 1995; Kenward & Molenberghs, 1999). We restrict our attention to monotone dropout, in which $P(\text{R}_{it} = 1 | \text{R}_{it+1} = 1) = 1$ and $P(\text{R}_{it} = 1 | \text{R}_{it-1} = 0) = 0$. Intermittent missing data are accommodated by some methods discussed, but will not be considered here.

The impact of missing data on regression analysis can be examined via the joint density of responses Y_i and dropout process R_i. Table 2.1 summarizes a hierarchy to assess the influence of missing values on estimation of parameters characterizing the response model. Notation differs slightly from that introduced above, to simplify presentation. Let $\mathbf{Z}_i$ be a design matrix of covariates (but not responses) for subject i. (Responses Y_i are not excluded from design matrix $\boldsymbol{D}_i$ above.) If dropout does not depend on the response Y_i, the

missingness mechanism is described as missing completely at random (MCAR). If dropout depends on observed (Y_i^o), but not unobserved (Y_i^m), responses, data are missing at random (MAR). If the dropout mechanism is dependent on unobserved response values Y_i^m (with or without dependence on Y_i^o), missingness is nonignorable (NI) (Laird, 1988).

Table 2.1: Missing data classification.

$f(R_i\|Y_i^m, Y_i^o, \mathbf{Z}_i) =$	$f(R_i\|\mathbf{Z}_i)$	missing completely at random (**MCAR**)
$=$	$f(R_i\|Y_i^o, \mathbf{Z}_i)$ or $f(R_i\|Y_i^o)$	missing at random (**MAR**)
$=$	$f(R_i\|Y_i^m, Y_i^o, \mathbf{Z}_i), f(R_i\|Y_i^m, Y_i^o)$ or $f(R_i\|Y_i^m)$	nonignorable (**NI**)

R_i: missingness indicator

Y_i^o: observed responses

Y_i^m: unobserved (missing) responses

$\mathbf{Z}_i$: covariates for modeling R_i, other than in Y_i vector

The observed data density used for likelihood estimation is found by integrating the joint density $f(Y_i^o, Y_i^m, R_i|\mathbf{X}_i, \mathbf{Z}_i, \beta, \phi)$ over the density of Y_i^m:

$$f(Y_i^o, R_i|\mathbf{X}_i, \mathbf{Z}_i, \beta, \phi) = \int f(R_i|Y_i^m, Y_i^o, \mathbf{Z}_i, \phi) f(Y_i^o|\mathbf{X}_i, \beta) f(Y_i^m|Y_i^o, \mathbf{X}_i, \beta) dY^m$$

When missingness is MAR (or MCAR), the density $f(R_i)$ can be simplified as shown in Table 2.1, with corresponding full likelihood:

$$\begin{aligned} f(Y_i^o, R_i|\mathbf{X}_i, \mathbf{Z}_i, \beta, \phi) &\overset{\text{MAR}}{=} f(R_i|Y_i^o, \mathbf{Z}_i, \phi) f(Y_i^o|\mathbf{X}_i, \beta) \int f(Y_i^m|Y_i^o, \mathbf{X}_i, \beta) dY^m \\ &= f(R_i|Y_i^o, \mathbf{Z}_i, \phi) f(Y_i^o|\mathbf{X}_i, \beta). \end{aligned}$$

Although Y_i^o and R_i are clearly not independent, the values of β that maximize $f(Y_i^o, R_i|\mathbf{X}_i, \mathbf{Z}_i, \beta, \phi)$ are the same values that maximize $f(Y_i^o|\mathbf{X}_i, \beta)$, when β and ϕ are distinct. Hence, if β is to be found by maximum likelihood and the missingness mechanism is MCAR or MAR, the missingness is "ignorable" and a likelihood-based analysis of available

data will yield consistent estimates for β. MCAR and MAR data have different implications for analyses that are not likelihood-based. This is discussed below.

For the classification of missing data as MCAR, MAR, or NI, the joint distribution $f(Y_i, R_i)$ of response vector Y_i and dropout mechanism R_i is factorized as $f(R_i|Y_i)f(Y_i)$. This factorization results in regression parameters (for Y_i) that do not depend on the dropout mechanism. $f(R_i|Y_i)$ is an outcome-dependent selection model (Heckman, 1979; Little & Rubin, 1987; Diggle & Kenward, 1994; Hogan & Laird, 1997).

An alternate factorization of the joint distribution is $f(R_i)f(Y_i|R_i)$, known as a pattern-mixture model (Little, 1995). In this approach, the distribution of Y_i is stratified by patterns of missingness. This method is generally easy to implement using existing software, by fitting dropout pattern as a covariate. In contrast, nonignorable likelihood-based selection models require explicit computation of the joint likelihood, which can be complicated. Pattern-mixture model regression parameters are conditional on dropout pattern. It is possible to marginalize fitted response values by a weighted average of estimates based on trajectories such as those in Figure 1.2 (Fitzmaurice & Laird, 2000). Thus although marginal summaries are available, direct marginal regression parameters are not estimated.

Because the response distribution $f(Y_i)$ is of primary interest in comparative clinical trials, we will focus on selection model approaches. Both selection and pattern-mixture models may involve random effects for response and/or dropout models (Little, 1995). However, we restrict attention to methods that specify marginal regression parameters directly.

Accommodation of dropout in linear models is reasonably well-developed (Diggle, 1998; Molenberghs et al., 1997a), particularly in the Bayesian framework (Geyer, 1996). Computation of nonlinear (including binary) response models is more challenging than for linear models. Additionally, parameters in random effects models for binary data will require a conditional or "subject-specific" interpretation (Zeger et al., 1988), unless the model is marginalized (Heagerty & Zeger, 2000). For both linear and binary outcomes, nonignorable selection models must address identifiability, as discussed below.

2.4 Analysis of Binary Longitudinal Data with Dropout

2.4.1 Ad Hoc Imputation Methods

Missing values are a common operational nuisance in scientific research. This has led to development of many ad hoc methods for accommodating missing values. Some are appropriate in certain situations, while others can generate large bias in estimation of regression parameters. Most published statistical studies cited below address missing data in linear models, but the principles also apply to a binary outcome.

Complete Case Analysis

One approach to missing data is to delete the records of all subjects for whom any data are missing. This method is simple to describe and to implement, and should yield unbiased point estimates of regression parameters if data are MCAR. However, loss of sample size can be considerable. Little & Rubin (1987) note that for 20 planned observations and a constant 10% chance of missingness, only about 12% of the original subjects are expected to provide complete data. Complete case analyses waste the effort taken to collect incomplete data series, and produce parameter estimates with low precision and power compared to estimates that use all available cases.

In isolated scenarios, complete case analysis may result in unbiased and efficient regression parameter estimation for non-MCAR data (Laird, 1988; Baker & Laird, 1988). In general, though, severe bias may result from complete case analysis when data are MAR or NI (Molenberghs et al., 1997a). For example, if schizophrenics with less severe symptoms are more likely to drop out of the PANSS study, a complete case analysis would overestimate the absolute severity of symptoms, and could distort estimation of treatment effects.

Last Observation Carried Forward

Another strategy is to impute missing values with the last observed value of Y_i. This approach seems sensible if, as for many "longitudinal" studies, the parameter of interest is a mean change score based on a baseline and a final followup measurement. Intervening

observations between baseline and end of followup are ignored if data are complete, but can be proxies for later observations lost to followup.

This approach does allow a randomization-valid test among treatment groups in a randomized parallel group trial (Diggle, 1998). However, if a slope is to be estimated, the bias in carrying observations forward is obvious. Additionally, precision is overestimated by any method where values imputed using observed data are treated on par with genuinely observed data (Rubin, 1987).

Mean Imputation

In "mean imputation", a group mean value is imputed for missing response values. For example, for the binary response PANSS data, 47% of observed subjects show improvement at 8 weeks of followup, so "no improvement" could be substituted for missing 8-week values. Imputation could also be based on treatment-specific means. This "vertical" imputation scheme borrows information from other subjects, whereas "horizontal" imputation such as carrying observations forward takes information from within subjects (Molenberghs et al., 1997a). Using averages from across subjects will cause within-cluster correlation to be underestimated. Vertical imputation may bias results if the imputation model makes strong assumptions about relationships among subjects (Lewis, 1957). As for other imputation methods, precision will be overestimated (Rubin, 1987).

Regression Imputation

Regression imputation may borrow both within- and between-subject information to predict missing responses based on nonmissing response and covariate values. This method is appealing because, depending on the missingness scheme, reasonably good estimates of missing values can be obtained. However, it is difficult to determine whether the model for the missing data is accurate, and an incorrectly specified imputation model can yield great bias in regression estimates. Finally, this single imputation method underestimates variability in the response (Rubin, 1987).

Available Case

Available case methods use all observed data, but do not impute values for missing data. As described below, likelihood-based analyses using available cases may yield valid inference when the nonresponse mechanism is "ignorable". However, this does not mean that available case methods are always unbiased or efficient. If correlations are estimated using only pairwise observed data (as could occur for an unstructured covariance matrix), resulting correlations may show great bias (Little & Rubin, 1987, page 43) or the covariance matrix may not be positive definite (Molenberghs et al., 1997a, page 220).

2.4.2 Likelihood-Based Methods

Section 2.3 demonstrates that missingness is "ignorable" for likelihood-based analysis when data are MCAR or MAR. Analysis of available cases yields consistent estimates for regression parameters β. However, the observed information matrix should be used for computing the asymptotic variance of $\hat{\beta}$ (Little & Rubin, 1987, Section 8.2.2; Molenberghs et al., 1997a, Section 5.8.1), since expected values over $f(Y_i^o)$ are not the same as over $f(Y_i^o, R_i)$.

Even when dropout is ignorable with respect to regression inference, it is often useful to model dropout in a selection model framework. Dropout patterns may be of scientific interest, or models with nonignorable dropout can be compared to those with ignorable dropout for sensitivity analysis.

In addition to ignorability of missingness under MAR, likelihood-based models for longitudinal binary data are also valuable because they are efficient compared to moment-based semiparametric approaches, and allow for likelihood ratio tests to compare models. We focus on methods that adopt marginal regression parameters.

Fitzmaurice & Laird (1993) combine a marginal regression model with an association model based on conditional odds ratios. This method is extended to accommodate nonignorable dropout through a selection model by Fitzmaurice et al. (1996b). The method is likelihood-based: thus, consistency and efficiency of parameter estimates are well-defined. Furthermore, the marginal mean (if correctly specified) is asymptotically robust to misspecification of conditional odds ratio models if no data are missing. However, due to the

dependence on conditional odds ratios, the distribution is not reproducible: if cluster sizes change, the specification and interpretation of the association model changes. Also, the data must be balanced (with observations at the same intervals for all subjects). Finally, the number of parameters needed to model association among responses grows rapidly as the cluster size increases. Models with nonignorable missing data depend on local identifiability, established by a nonsingular Fisher information matrix, for estimation of missingness parameters. Global identifiability is then assessed by an ad hoc method using the available data (Fitzmaurice et al., 1996b).

Molenberghs & Lesaffre (1994) adopt a marginal regression model with marginal odds ratios to describe association, based on the global cross-ratio models of Dale (1986). They model ordinal as well as binary response data. A selection model is appended for nonignorable dropout by Molenberghs et al. (1997b). Unlike models specified using conditional odds ratios, the distribution is reproducible. However, regression parameters are not necessarily consistent when the marginal odds ratio model is misspecified. As for all models involving nonignorable missingness, unverifiable assumptions are needed to make parameters identifiable. Like Fitzmaurice et al. (1996b), Molenberghs & Lesaffre (1994) recommend using models with NI assumptions as part of a sensitivity analysis.

The methods described, as well as other likelihood-based methods based on log-linear specification of binary data (Conaway, 1993; Baker, 1995; Lindsey, 2000), are impractical for long series of data because of computational difficulties and interpretation of association parameters. Also, some approaches estimate parameters using the EM algorithm (Dempster et al., 1977), making standard errors difficult to compute. The proposed method addresses these shortcomings.

Bayesian analysis of longitudinal binary data can be extended to accommodate missing values (Geyer, 1996; Clayton, 1996; Park, 1998). For Bayesian fitting of generalized linear mixed models (assuming MAR data), maximum likelihood estimates of (β, α) are produced by imposing either scientifically justified or mathematically convenient prior distributions for all parameters. Evaluation of posterior distributions is computationally intensive, but possible using modern simulation methods.

2.4.3 *Semiparametric Methods*

In this section we discuss methods that adapt generalized estimating equations (GEE, Liang & Zeger, 1986) to accommodate missing data. In GEE, estimates for regression parameter vector β in Equation 2.1 are found as the solution to quasi-score equations summed over subjects $i = 1, \ldots, N$:

$$U(\beta, \hat{\alpha}) = \sum_{i=1}^{N} \frac{\partial \mu_i^M}{\partial \beta} \boldsymbol{V}_i(\hat{\alpha})^{-1} (\mathrm{Y}_i^{\mathrm{o}} - \mu_i^M) = 0$$

where $\frac{\partial \mu_i^M}{\partial \beta}$ is $\mu_i^M(1 - \mu_i^M)\mathrm{X}_i$ for the logit link and $\boldsymbol{V}_i(\hat{\alpha})$ is a working covariance matrix computed with a method of moments estimator for the correlation or odds ratio structure. In the presence of MCAR dropout, estimates of β are consistent, since observed data $\mathrm{Y}_i^{\mathrm{o}}$, where $\mathrm{R}_{it} = 1$, have the same distribution as the full vector Y_i:

$$\begin{aligned}
f(\mathrm{R}_i, \mathrm{Y}_i) &= \int f(\mathrm{R}_i, \mathrm{Y}_i^{\mathrm{o}}, \mathrm{Y}_i^{\mathrm{m}}) d\mathrm{Y}^{\mathrm{m}} \\
&= \int f(\mathrm{R}_i | \mathrm{Y}_i^{\mathrm{o}}, \mathrm{Y}_i^{\mathrm{m}}) f(\mathrm{Y}_i^{\mathrm{m}} | \mathrm{Y}_i^{\mathrm{o}}) f(\mathrm{Y}_i^{\mathrm{o}}) d\mathrm{Y}^{\mathrm{m}} \\
&\overset{\mathrm{MCAR}}{=} \int f(\mathrm{R}_i) f(\mathrm{Y}_i^{\mathrm{o}}) f(\mathrm{Y}_i^{\mathrm{m}} | \mathrm{Y}_i^{\mathrm{o}}) d\mathrm{Y}^{\mathrm{m}} \\
&= f(\mathrm{R}_i) f(\mathrm{Y}_i^{\mathrm{o}}) \\
\rightarrow f(\mathrm{Y}_i) = f(\mathrm{Y}_i^{\mathrm{o}}) & \qquad \text{since } f(\mathrm{R}_i, \mathrm{Y}_i) = f(\mathrm{R}_i) f(\mathrm{Y}_i) \text{ under MCAR.}
\end{aligned}$$

However, under MAR or NI missingness, it is not assured that $\mathrm{E}(\mathrm{Y}_{it} | \mathrm{X}_{it}, \mathrm{R}_{it} = 1) = \mathrm{E}(\mathrm{Y}_{it} | \mathrm{X}_{it}) = \mu_{it}^M$, so the quasi-score equation may not yield consistent estimates of β. Robins et al. (1995) show that consistency is restored by weighting the estimating equations by inverse probability of sampling weights (Horvitz & Thompson, 1952):

$$U(\beta, \hat{\alpha}) = \sum_{i=1}^{N} \frac{\partial \mu_i^M}{\partial \beta} \boldsymbol{V}_i(\hat{\alpha})^{-1} \boldsymbol{W}_i (\mathrm{Y}_i - \mu_i^M) = 0. \tag{2.4}$$

$\boldsymbol{W}_i$ is a diagonal weighting matrix with elements $(\mathrm{R}_{it}/\pi_{it})$ such that

$$\begin{aligned}
\mathrm{R}_{it} &= 1 \text{ if } Y_{it} \text{ observed, } 0 \text{ if } Y_{it} \text{ not observed} \\
\pi_{it} &= P(\mathrm{R}_{it} = 1 | y_{i1}, \ldots y_{it-1}) \\
&= P(\mathrm{R}_{i2} = 1 | y_{i1}) \cdot P(\mathrm{R}_{i3} = 1 | \mathrm{R}_{i2} = 1, y_{i1}, y_{i2}) \\
&\quad \cdots P(\mathrm{R}_{it} = 1 | \mathrm{R}_{it-1} = 1, y_{i1}, \ldots y_{it-1})
\end{aligned}$$

so that the expected value of the quasi-score is zero under the joint density $f(R_i, Y_i)$. Sampling (or censoring) probabilities π_{it} are generally provided through a parametric (logistic regression) selection model. The selection model may accommodate nonignorable missingness in the same manner as for likelihood-based methods above (Rotnitzky et al., 1998). Rather than depending on weak identifiability of parameters in specific models, the authors hold relevant parameters constant in the selection model and examine their impact through sensitivity analysis.

Inverse probability of censoring weighted GEE (IPCW-GEE), like GEE without sampling weights, provides consistent estimation of regression parameters when association is misspecified. However, the weights must be correctly specified. Semiparametric models will also be less efficient than their likelihood-based counterparts, unless the association modeled is very close to the "true" association model (Fitzmaurice et al., 1993). We explore the relative efficiency of IPCW-GEE and the proposed likelihood-based method in Chapter 4. Semiparametric efficient estimation is discussed, but not implemented.

Paik (1997) uses semiparametric sequential imputation to find consistent estimates of mean parameters under MAR and NI dropout. Under monotone MAR dropout, the expected value of Y_{it} given prior elements of Y_i (all of which are observed) is the same whether or not Y_{it} is observed. A value for Y_{it} can therefore be imputed (or multiply imputed) as the sample mean (or selections from a bootstrap sample) of responses with an identical history. Subsequent values are imputed by the same technique, with prior imputed values as part of the response history. For saturated mean and selection models, Paik's mean imputation method yields identical results to the IPCW-GEE of Robins et al. (1995).

2.4.4 Multiple Imputation

Multiple imputation (MI) (Rubin, 1987) is also used to analyze longitudinal data with missing values. Parametric (including Bayesian), semi-parametric, and non-parametric methods can all be used both to impute missing values and to analyze imputed datasets. Standard methods of analysis, such as GEE, can be applied. The only modification required is to combine results for the multiple imputations. Multiple imputation methods do not overestimate

precision, in contrast to the single-imputation strategies listed above. Due to asymptotic equivalence between Bayesian and likelihood-based methods, appropriately specified MI and maximum likelihood methods will yield similar parameter estimates and standard errors, for models with a large sample size and sufficient number of imputations (Rubin, 1987; Collins et al., 2002). A disadvantage of multiple imputation is the stochastic nature of the results: different point estimates can be computed from analysis of the same set of observed responses. This random variation would also occur with repeated sampling from the complete data distribution. The drawback for MI is procedural, not methodological.

2.5 Marginalized Models

The previous sections review existing methods for analysis of binary longitudinal data with ignorable or nonignorable dropout. The methods differ with respect to target of inference, robustness to dropout misspecification, and ease of computation. The next three chapters describe the marginalized transition model, which has useful properties in these three areas.

Heagerty & Zeger (2000) describe a class of "marginalized models" where a conditional model (such as a random effects or transition model) specifies the underlying process of data generation, but marginal mean parameters may be recovered. When regression is via a transition model, interpretation of parameters changes with changes in the association model. The marginalized transition model (Heagerty, 2002), described below, retains an association model using transitions, but permits marginal regression estimation of treatment effects.

The marginalized transition model (MTM), in which association is modeled via conditional dependence on prior responses, and the selection model, in which dropout is related to responses, provide a coherent framework for describing both response and dropout mechanisms. In Chapter 3 we extend the marginalized transition model to accommodate dropout models. Chapter 4 compares the MTM to inverse probability of censoring weighted GEE (IPCW-GEE, Robins et al., 1995) with respect to efficiency and misspecification bias, and Chapter 5 compares the MTM and IPCW-GEE in analysis of the PANSS data introduced in Chapter 1.

Chapter 3

MARGINALIZED TRANSITION MODEL WITH DROPOUT

A likelihood-based model for serial binary data with marginal mean parameters and first-order Markov association is introduced by Azzalini (1994), described in the framework of marginalized models by Heagerty & Zeger (2000), and generalized as a marginalized transition model by Heagerty (2002). In this chapter we describe and model the binary data likelihood function for the first- and second-order marginalized transition model with both ignorable and nonignorable dropout. We first describe the marginalized transition model (MTM), then develop a selection model for the joint likelihood of responses Y_i and missingness R_i. Using a selection model to accommodate monotone dropout leads to a natural separation between dropout and response processes, especially for data missing at random (MAR).

3.1 Marginalized Transition Model

The marginalized transition model (Heagerty, 2002) is constructed around two models: a marginal regression mean, and a conditional mean dependent on prior response values. Association among responses is characterized directly through transitions modeled by vector α. The marginal regression of responses on covariates is parameterized by coefficient vector β. The conditional mean is "marginalized" by exploiting the direct relationship between the conditional and marginal means.

Figure 3.1, adapted from Figure 1 in Heagerty (2002), is a schematic representation of four observations of a second-order marginalized transition model (MTM) with nonignorable (NI) monotone dropout as defined in Chapter 2. Figure 3.1 illustrates the marginal mean, dependence of Y_{it} on past responses, and dependence of R_{it} on past values of R_i, past values of Y_i, and (if dropout is NI) on Y_{it}. Dashed boxes and parameters β represent the marginal mean that is the primary target of inference. Likelihood-based inference requires

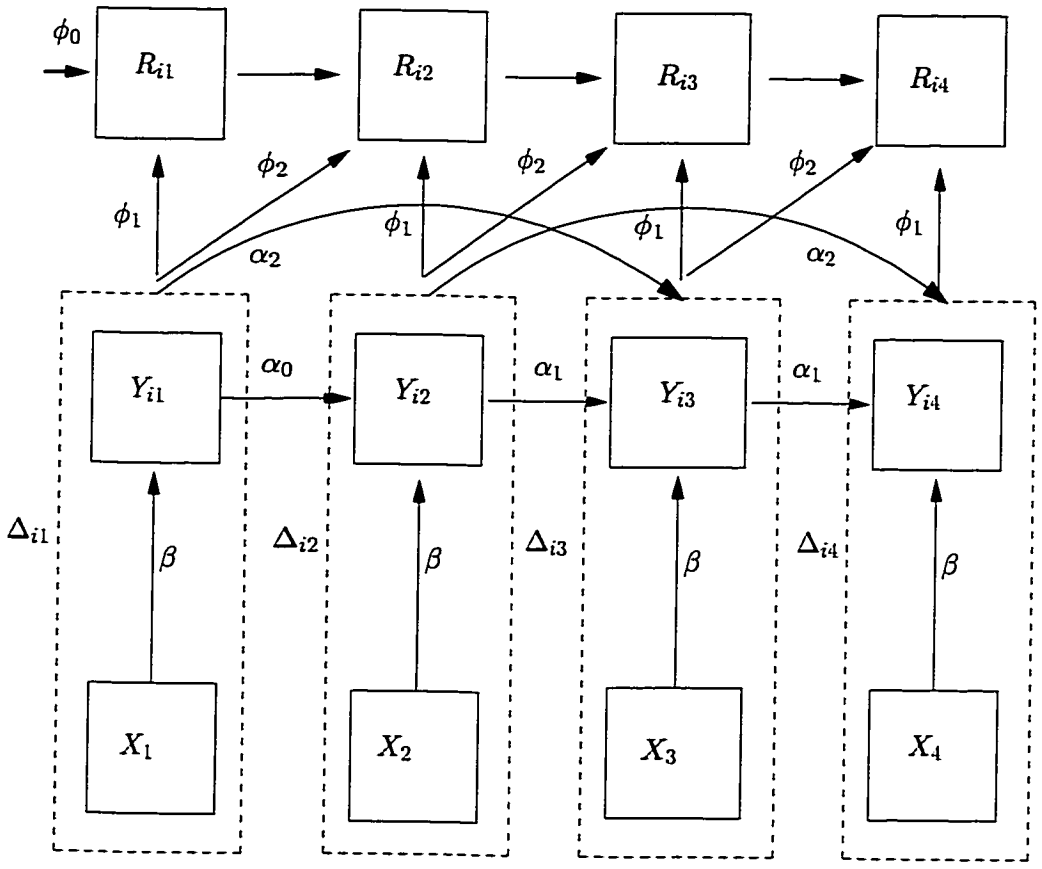


Figure 3.1: Schematic of a second-order marginalized transition model with nonignorable monotone dropout. Illustrates the marginal model (X, β), conditional model (Δ, α), and dropout model (ϕ).

full specification of $f(\mathrm{Y}_i, \mathrm{R}_i)$, though, so further modeling is necessary. The conditional mean, in which response at time t, Y_{it}, depends on covariates and prior responses, is parameterized by Δ and α. Finally, the dropout process $f(\mathrm{R}_i|\mathrm{Y}_i)$ is parameterized by ϕ. In Figure 3.1, whether Y_{it} is observed (whether $\mathrm{R}_{it} = 1$) depends on values of R_{it-1}, Y_{it-1}, and Y_{it}. If $\mathrm{R}_{it-1} = 0$ then $\mathrm{R}_{it} = 0$. Other missingness dependencies are modeled by ϕ. Details of estimation are discussed below.

The marginal mean is a logit-linear model describing the expected value of response vector Y_i, modeled by covariate matrix $\mathbf{X}_i$ through regression coefficients β:

$$\mathrm{logit}(\mu_{it}^M) = \mathrm{X}_{it}\beta$$

where $\mu_{it}^M = \mathrm{E}(\mathrm{Y}_{it}|\mathrm{X}_{it})$ is the marginal mean. For example, time trends and treatment effects for a randomized clinical trial would be modeled through the marginal mean. Association is introduced via the conditional mean, $\mu_{it}^C = \mathrm{E}(\mathrm{Y}_{it}|\mathrm{X}_{it}, \mathrm{Y}_{it-1} \ldots \mathrm{Y}_{i1})$. (The conditional mean model matrix is given as $\mathbf{X}_i$ as for the marginal mean, but may only involve a subset of marginal mean covariates.) The conditional mean is a transition model: current responses depend on values of past responses.

When the dropout process $f(\mathrm{R}_i|\mathrm{Y}_i)$ is ignorable, the likelihood function for marginal mean parameters β and conditional mean (association) parameters α can be defined as follows for response vectors Y_i, for clusters $i = 1, \ldots, N$ at timepoints $t = 1, \ldots, n_i$:

$$\begin{aligned} \boldsymbol{L}_i(\beta, \alpha) &= P(\mathrm{Y}_{i1} = y_{i1})P(\mathrm{Y}_{i2} = y_{i2}|\mathrm{Y}_{i1}) \cdots P(\mathrm{Y}_{in_i} = y_{in_i}|\mathrm{Y}_{i1} \ldots \mathrm{Y}_{in_i-1}) \\ &= \prod_{i=1}^{N} (\mu_{i1}^M)^{y_{i1}} (1 - \mu_{i1}^M)^{(1-y_{i1})} \prod_{i=1}^{N} \prod_{t=2}^{n_i} (\mu_{it}^C)^{y_{it}} (1 - \mu_{it}^C)^{(1-y_{it})}. \end{aligned}$$

For a p^{th}-order marginalized transition model, MTM(p), the logit-linear model for μ_{it}^C is:

$$\begin{aligned} \mathrm{logit}(\mu_{it}^C) &= \Delta_{it} + \sum_{j=1}^{p} \gamma_{it,j} \cdot y_{it-j} \\ \text{for } \gamma_{it,j} &= \mathrm{Z}_{it,j}\alpha_j. \end{aligned} \tag{3.1}$$

The association is modeled through p α-vectors. For example, if each $\mathrm{Z}_{it,j} = 1$, then $\gamma_{it,j} = \alpha_j$, and $(\alpha_1 \ldots \alpha_p)$ are each constants measuring the strength of association between the jth lagged response and the present response. Interactions of lagged response variables are possible in the model for the conditional mean, but they are not included in this presentation. Intercept Δ_{it} describes the conditional mean when the p previous response values are zero. For a given β and α, the value of Δ_{it} may be derived through relationships between marginal and conditional means, as shown in the following convolution equation:

$$\begin{aligned} \mu_{it}^M &= \sum_{y_{it-1}, \cdots y_{it-p}} P(\mathrm{Y}_{it} = 1|\mathrm{Y}_{it-1} \ldots \mathrm{Y}_{it-p}) \cdot P(\mathrm{Y}_{it-1} = y_{it-1}, \ldots, \mathrm{Y}_{it-p} = y_{it-p}) \\ &= \sum_{y_{it-1}, \cdots y_{it-p}} \mu_{it}^C \cdot P(\mathrm{Y}_{it-1} = y_{it-1}, \ldots, \mathrm{Y}_{it-p} = y_{it-p}) \end{aligned}$$

Further details are given by Heagerty (2002). We now derive the likelihood for first- and second-order marginalized transition models.

3.1.1 First-order Marginalized Transition Model

For a first-order marginalized transition model, MTM(1), the conditional mean depends on a single prior response, and no other response values. For $j = t-1$,

$$\mu_{it}^C = \mathrm{E}\left(\mathrm{Y}_{it}|\mathrm{H}_{it}^{(\mathrm{Y})}\right) = \mathrm{E}(\mathrm{Y}_{it}|\mathrm{Y}_{ij}).$$

For the first-order model, transitions can be modeled using odds ratios. Let joint probability $\nu_{itj} = P(\mathrm{Y}_{it}=1, \mathrm{Y}_{ij}=1)$ and odds ratio $\Psi_{itj} = \frac{P(\mathrm{Y}_{it}=1,\mathrm{Y}_{ij}=1)P(\mathrm{Y}_{it}=0,\mathrm{Y}_{ij}=0)}{P(\mathrm{Y}_{it}=1,\mathrm{Y}_{ij}=0)P(\mathrm{Y}_{it}=0,\mathrm{Y}_{ij}=1)}$. We now show how the association model

$$\log(\Psi_{itj}) = \mathrm{Z}_{it}\alpha$$

is incorporated into the likelihood, via the conditional mean μ_{it}^C.

Since

$$\begin{aligned}
\mathrm{P}(\mathrm{Y}_{it}=1|\mathrm{Y}_{ij}=1) = \frac{\mathrm{P}(\mathrm{Y}_{it}=1,\mathrm{Y}_{ij}=1)}{\mathrm{P}(\mathrm{Y}_{ij}=1)} &= \frac{\nu_{itj}}{\mu_{ij}^M} \\
\mathrm{P}(\mathrm{Y}_{it}=0|\mathrm{Y}_{ij}=1) = \frac{\mathrm{P}(\mathrm{Y}_{it}=0,\mathrm{Y}_{ij}=1)}{\mathrm{P}(\mathrm{Y}_{ij}=1)} &= \frac{\mu_{ij}^M - \nu_{itj}}{\mu_{ij}^M} \\
\mathrm{P}(\mathrm{Y}_{it}=1|\mathrm{Y}_{ij}=0) = \frac{\mathrm{P}(\mathrm{Y}_{it}=1,\mathrm{Y}_{ij}=0)}{\mathrm{P}(\mathrm{Y}_{ij}=0)} &= \frac{\mu_{it}^M - \nu_{itj}}{1-\mu_{ij}^M} \\
\mathrm{P}(\mathrm{Y}_{it}=0|\mathrm{Y}_{ij}=0) = \frac{\mathrm{P}(\mathrm{Y}_{it}=0,\mathrm{Y}_{ij}=0)}{\mathrm{P}(\mathrm{Y}_{ij}=0)} &= \frac{1-\mu_{it}^M - \mu_{ij}^M + \nu_{itj}}{1-\mu_{ij}^M},
\end{aligned}$$

probabilities μ_{it}^C are:

$$\begin{aligned}
\mu_{it}^C = \mathrm{P}(\mathrm{Y}_{it}=1|\mathrm{Y}_{ij}=y_{ij}) &= \left(\frac{\nu_{itj}}{\mu_{ij}^M}\right)^{y_{ij}} \cdot \left(\frac{\mu_{it}^M - \nu_{itj}}{1-\mu_{ij}^M}\right)^{1-y_{ij}} \\
1-\mu_{it}^C = \mathrm{P}(\mathrm{Y}_{it}=0|\mathrm{Y}_{ij}=y_{ij}) &= \left(\frac{\mu_{ij}^M - \nu_{itj}}{\mu_{ij}^M}\right)^{y_{ij}} \cdot \left(\frac{1-\mu_{it}^M - \mu_{ij}^M + \nu_{itj}}{1-\mu_{ij}^M}\right)^{1-y_{ij}}.
\end{aligned}$$

We use these identities to give a first-order version of Equation 3.1:

$$
\begin{aligned}
\text{logit}(\mu_{it}^C) &= \underbrace{y_{ij}\cdot(\log(\nu_{itj}) - \log(\mu_{ij}^M)) + (1-y_{ij})\cdot\log(\mu_{it}^M - \nu_{itj}) - (1-y_{ij})\cdot\log(1-\mu_{ij}^M)}_{\log(\mu_{it}^C)} \\
&\quad \underbrace{-y_{ij}\cdot\log(\mu_{ij}^M - \nu_{itj}) + y_{ij}\cdot\log(\mu_{ij}^M)}_{-\log(1-\mu_{it}^C)} \\
&\quad \underbrace{-(1-y_{ij})\cdot\log(1-\mu_{it}^M - \mu_{ij}^M + \nu_{itj}) + (1-y_{ij})\cdot\log(1-\mu_{ij}^M)}_{-\log(1-\mu_{it}^C)\text{ cont.}} \\
&= y_{ik}\cdot(\log(\nu_{itj}) - \log(\mu_{it}^M - \nu_{itj}) - \log(\mu_{ij}^M - \nu_{itj}) + \log(1-\mu_{it}^M - \mu_{ij}^M + \nu_{itj})) \\
&\quad + \log(\mu_{it}^M - \nu_{itj}) - \log(1-\mu_{it}^M - \mu_{ij}^M + \nu_{itj}) \\
&= y_{ij}\cdot\log(\Psi_{itj}) + \log\left(\frac{\mu_{it}^M - \nu_{itj}}{1-\mu_{it}^M - \mu_{ij}^M + \nu_{itj}}\right) \qquad (3.2)
\end{aligned}
$$

Carey et al. (1993) use this identity to compute logistic regression parameters for pairwise odds ratios Ψ_{itj}, holding marginal regression parameter values (and thus μ_{it}^M and μ_{ij}^M) constant. Alternating logistic regressions (Carey et al., 1993) marginalizes correlated binary data by pairing the logistic regression with offset in Equation 3.2 with logistic regression for the marginal regression parameters. For the MTM(1), Equation 3.2 is used as a closed form equation for μ_{it}^C, fully parameterized by α and β.

Returning to the likelihood, and defining $\eta_{it}^M = \text{logit}(\mu_{it}^M) = \mathrm{X}_{it}\beta$ and $\eta_{it}^C = \text{logit}(\mu_{it}^C)$ (using Equation 3.2):

$$
\begin{aligned}
\boldsymbol{L}(\beta,\alpha) &= \prod_{i=1}^{N}(\mu_{i1}^M)^{y_{i1}}\cdot(1-\mu_{i1}^M)^{(1-y_{i1})}\prod_{i=1}^{N}\prod_{t=2}^{n_i}(\mu_{it}^C)^{y_{it}}\cdot(1-\mu_{it}^C)^{(1-y_{it})} \\
\log[\boldsymbol{L}(\beta,\alpha)] &= \sum_{i=1}^{N} y_{i1}\cdot\log(\mu_{i1}^M) + (1-y_{i1})\cdot\log(1-\mu_{i1}^M) \\
&\qquad + \sum_{i=1}^{N}\sum_{t=2}^{n_i} y_{it}\cdot\log(\mu_{it}^C) + (1-y_{it})\cdot\log(1-\mu_{it}^C) \\
&= \sum_{i=1}^{N} y_{i1}\cdot\text{logit}(\mu_{i1}^M) + \log(1-\mu_{i1}^M) + \sum_{i=1}^{N}\sum_{t=2}^{n_i} y_{it}\cdot\text{logit}(\mu_{it}^C) + \log(1-\mu_{it}^C)
\end{aligned}
$$

The score equation for β (where, again, $j = t-1$) is:

$$\frac{\partial \log[\boldsymbol{L}(\beta,\alpha)]}{\partial \beta} = \sum_{i=1}^{N}(y_{i1}-\mu_{i1}^{M})\cdot \mathrm{X}_{i1} + \sum_{i=1}^{N}\sum_{t=2}^{n_i}(y_{it}-\mu_{it}^{C})\cdot\frac{\partial \eta_{it}^{C}}{\partial \mu_{it}^{M}}\frac{\partial \mu_{it}^{M}}{\partial \beta} + \sum_{i=1}^{N}\sum_{t=2}^{n_i}(y_{it}-\mu_{it}^{C})\cdot\frac{\partial \eta_{it}^{C}}{\partial \mu_{ij}^{M}}\frac{\partial \mu_{ij}^{M}}{\partial \beta}$$

and for α:

$$\frac{\partial \log[\boldsymbol{L}(\beta,\alpha)]}{\partial \alpha} = \sum_{i=1}^{N}\sum_{t=2}^{n_i}(y_{it}-\mu_{it}^{C})\cdot\frac{\partial \eta_{it}^{C}}{\partial \Psi_{itj}}\frac{\partial \Psi_{itj}}{\partial \alpha}.$$

Details for solving these score equations are found in Appendix A.

3.1.2 Second-Order Marginalized Transition Model

For the second-order marginalized transition model, MTM(2), the likelihood is set up the same as for MTM(1), but transition probabilities are changed to allow second-order dependence. A first-order conditional probability remains part of the likelihood, when t=2. Its estimation is incorporated into the association model parameterized by α.

Another difference between MTM(1) and MTM(2) is that no closed-form solution such as Equation 3.2 can be used to represent μ_{it}^{C} in terms of marginal regression parameters β and association parameters α. The convolution equation must be applied directly. We introduce new notation to represent the convolution equation of a second-order MTM:

$$\mu_{it}^{M} = \mathrm{E}(\mathrm{Y}_{it}|\mathrm{X}_{it}) = \sum_{j=0}^{1}\sum_{k=0}^{1}\mu_{it(jk)}^{C}\pi_{(jk)}^{t}$$

where

$$\mu_{it(jk)}^{C} = \mathrm{P}(\mathrm{Y}_{it}=1|\mathrm{Y}_{it-1}=j,\mathrm{Y}_{it-2}=k)$$

$$\pi_{(jk)}^{t} = \mathrm{P}(\mathrm{Y}_{it-1}=j,\mathrm{Y}_{it-2}=k).$$

The likelihood for the MTM(2) is:

$$\log[\boldsymbol{L}(\beta,\alpha)] = \sum_{i=1}^{N} y_{i1}\cdot\mathrm{logit}(\mu_{i1}^{M}) + \log(1-\mu_{i1}^{M}) + \sum_{i=1}^{N} y_{i2}\cdot\mathrm{logit}(\mu_{i2}^{C}) + \log(1-\mu_{i2}^{C}) + \sum_{i=1}^{N}\sum_{t=3}^{n_i} y_{it}\cdot\mathrm{logit}(\mu_{it}^{C}) + \log(1-\mu_{it}^{C}).$$

Here, for the second-order marginalized transition model with ignorable (MCAR or MAR) dropout, we focus on solving the convolution equation to find $\frac{\partial \mu_{it}^C}{\partial \beta}$ and $\frac{\partial \mu_{it}^C}{\partial \alpha}$. Details are found in Appendix B.

In summary, first- and second-order marginalized transition models provide likelihood-based parameter estimation simultaneously for transitions and for a marginal regression mean. We discuss estimation below, after incorporating dropout models into the likelihood.

3.2 Selection Model Specification

The likelihood function for regression, association, and dropout parameters (β, α, ϕ) (see Section 2.1 for introduction of notation) is based on observed data: response data for subjects who have not yet dropped out ($\mathrm{Y}_i^{\mathrm{o}}$), and dropout indicator (R_i) and covariate matrix $\mathbf{X}_i$ for all time periods. Data are assumed to be generated under the complete data distribution $f(\mathrm{Y_i}, \mathrm{R}_i)$, where $\mathrm{Y_i} = (\mathrm{Y}_i^{\mathrm{o}}, \mathrm{Y}_i^{\mathrm{m}})$, and $\mathrm{Y}_i^{\mathrm{m}}$ is a vector of responses that are missing. We consider only monotone dropout, not intermittent missing data: if Y_{it} is missing, then Y_{is} are missing for $s > t$. Let $f(\mathrm{Y}_i^{\mathrm{o}})$ be the observed data distribution, and $f(\mathrm{R}_i)$ be the distribution of dropout indicator R_i. The observed data likelihood for marginal regression parameters β, association parameters α, and dropout parameters ϕ is:

$$\begin{aligned} \boldsymbol{L}_i(\beta, \alpha, \phi) \propto f(\mathrm{Y}_i^{\mathrm{o}}, \mathrm{R}_i) &= \int_{\mathrm{Y}^{\mathrm{m}}} f(\mathrm{Y_i}, \mathrm{R}_i) \\ &= \int_{\mathrm{Y}^{\mathrm{m}}} f(\mathrm{R}_i|\mathrm{Y_i}) f(\mathrm{Y_i}) \\ &= \int f(\mathrm{R}_i|\mathrm{Y_i}) f(\mathrm{Y}_i^{\mathrm{o}}) f(\mathrm{Y}_i^{\mathrm{m}}|\mathrm{Y}_i^{\mathrm{o}}) d(\mathrm{Y}_i^{\mathrm{m}}|\mathrm{Y}_i^{\mathrm{o}}). \end{aligned} \tag{3.3}$$

This is a typical selection model factorization (Heckman, 1979; Little & Rubin, 1987).

We now consider modeling of $f(\mathrm{R}_i|\mathrm{Y_i})$ in Equation 3.3 by logistic regression. First, use a telescoping product of conditional distributions to specify vector R_i:

$$\begin{aligned} f(\mathrm{R}_i|\mathrm{Y_i}) &= f(\mathrm{R}_{i1}|\mathrm{Y_i}) f(\mathrm{R}_{i2}|\mathrm{R}_{i1}, \mathrm{Y_i}) \cdots f(\mathrm{R}_{in_i}|\mathrm{H}_{in_i}^{(\mathrm{R})}, \mathrm{Y_i}) \\ &= \prod_{t=2}^{n_i} f(\mathrm{R}_{it}|\mathrm{H}_{it}^{(\mathrm{R})}, \mathrm{H}_{it}^{(\mathrm{Y})}, \mathrm{Y}_{it}) \end{aligned}$$

where $\mathrm{H}_{it}^{(\mathrm{R})}$ is the history for R_i through time $(t-1)$, $\mathrm{H}_{it}^{(\mathrm{R})} = (\mathrm{R}_{i1} \ldots \mathrm{R}_{it-1})$, and $\mathrm{H}_{it}^{(\mathrm{Y})}$ is the history for response Y_{it} in cluster (subject) i. The first observation, Y_{i1}, is assumed never

to be missing: this common assumption (Diggle & Kenward, 1994; Robins et al., 1995) simplifies parameterization of dropout models. Current values R_{it} and Y_{it} are assumed to depend only on past and current values, not the future.

Let $u_{it} = P(R_{it} = 1|H_{it}^{(R)}, H_{it}^{(Y)}, Y_{it})$ and define time d as the first time that response Y is missing. ($R_{id-1} = 1$ and $R_{id} = 0$.) The likelihood contribution for each $R_{it} = r_{it}$ given the response history is:

$$f(R_{it}|H_{it}^{(R)}, H_{it}^{(Y)}, Y_{it}) = u_{it}^{r_{it}}(1-u_{it})^{1-r_{it}}$$

where

$$u_{it} = \begin{cases} p_{it} & : \quad t \leq d \\ 0 & : \quad t > d \end{cases}$$

and

$$p_{it} = P(R_{it} = 1|R_{t-1} = 1, H_{it}^{(Y)}, Y_{it}).$$

Parameterization of p_{it} is described below in Section 3.2.1. The likelihood contribution of $f(R_i|Y_i)$ is:

$$\begin{aligned} f(R_i|Y_i) &= \prod_{t=2}^{d-1} p_{it}^{r_{it}}(1-p_{id})^{1-r_{id}} \prod_{k=d+1}^{n} (1)^{(1-r_{ik})} \\ &= \left(\prod_{t=2}^{d-1} p_{it}\right)(1-p_{id}). \end{aligned} \tag{3.4}$$

Substituting, and rearranging terms from Equations 3.3 and 3.4, the full likelihood may be written as:

$$L_i(\beta, \alpha, \phi) = \underbrace{f(Y_i^o)}_{L1} \underbrace{\left(\prod_{t=2}^{d-1} p_{it}\right)}_{L2} \underbrace{\int (1-p_{id}) f(Y_i^m|Y_i^o) d(Y_i^m|Y_i^o)}_{L3}$$

L1, L2, and L3 correspond to Equations 23-25 of Diggle & Kenward (1994). L1 is the likelihood for the observed responses, the marginalized transition model (MTM) under ignorable dropout. L2 is defined by the dropout model, and L3 is the expected value of $(1-p_{id})$ under the distribution $f(Y_i^m|Y_i^o)$. Derivation of the MTM is described above. We now describe L2 and L3 in detail, then describe maximum likelihood estimation of parameters.

3.2.1 L2: Selection Model

L2 models p_{it}, the probability that Y_{it} is observed given the response history and covariate values. Monotone dropout in discrete time can be analyzed as discrete-time survival analysis using logistic regression (Abbott, 1985; Willett & Singer, 1993). An example of such a dropout model is

$$\text{logit}(p_{it}) = \phi_0 + \phi_1 \cdot \mathrm{Y}_{it} + \phi_2 \cdot \mathrm{Y}_{it-1}.$$

This model describes nonignorable dropout, since it depends on the current observation Y_{it} as well as the first-order response Y_{it-1}. A more general model for dropout is

$$\text{logit}(p_{it}) = \mathrm{D}_{it}\phi + \mathrm{G}_{it}\lambda \cdot \mathrm{Y}_{it}.$$

Here $\mathbf{D}_i$ is a model matrix that may include covariates and lagged observations, with ϕ as the corresponding parameters. $\mathbf{G}_i$ is the model matrix and λ the parameter vector associated with the current observation, Y_{it}. Dropout is MAR if λ is a zero-vector, and MCAR if λ is a zero-vector and $\mathbf{D}_i$ does not contain any response (Y) values. Vector λ is separated from ϕ because of identifiability issues discussed below. Interactions among dropout model variables are possible, but difficult due to identifiability when Y_{it} is involved.

3.2.2 L3: Nonignorable Contribution

L3 is defined in terms of p_{id} and $f(\mathrm{Y}_i^{\mathrm{m}}|\mathrm{Y}_i^{\mathrm{o}})$. Probability p_{id} is determined by the selection model described for L2, and parameterized by ϕ and λ. Because p_{id} only depends on $\mathrm{Y}_i^{\mathrm{m}}$ through Y_{id}, L3 involves the distribution of Y_{id} given the previous response values, $\mathrm{Y}_i^{\mathrm{o}}$. Conveniently, this transition is already modeled by the marginalized transition model:

$$\begin{aligned}
\mathrm{E}_{\mathrm{Y}^{\mathrm{m}}|\mathrm{Y}^{\mathrm{o}}}(1-p_{id}) &= \sum_{\mathrm{Y}_{id}=0,1} (1-p_{id}) P(\mathrm{Y}_{id} = y_{id}|\mathrm{H}_{id}^{(\mathrm{Y})}) \\
&= (1-p_{id|\mathrm{Y}_{id}=1}) P(\mathrm{Y}_{id}=1|\mathrm{H}_{id}^{(\mathrm{Y})}) + (1-p_{id|\mathrm{Y}_{id}=0}) P(\mathrm{Y}_{id}=0|\mathrm{H}_{id}^{(\mathrm{Y})}) \\
&= \frac{1}{1+\exp(\mathrm{D}_{id}\phi + \mathrm{G}_{id}\lambda)}\mu_{id}^{C} + \frac{1}{1+\exp(\mathrm{D}_{id}\phi)}(1-\mu_{id}^{C}).
\end{aligned}$$

Note that L1 only contributes to estimation of (β, α), L2 only contributes to estimation of ϕ, and L3 contributes to all three parameter vectors. However, if the dropout process is

MAR ($\lambda \equiv 0$), then L3 reduces to

$$\frac{1}{1+\exp(\mathrm{D}_{id}\phi)},$$

which depends only on ϕ.

3.3 Estimation

3.3.1 Maximum Likelihood Estimates

The first- and second-order MTM likelihoods are maximized by Newton-Raphson estimation for all parameters. Computations are described in Heagerty (2002), and in detail in Appendices A and B. Newton-Raphson estimation is used instead of Fisher scoring because the expected information matrix is the expectation over $f(\mathrm{Y}_i, \mathrm{R}_i)$, not $f(\mathrm{Y}_i^o)$, and will not yield correct standard errors if data are MAR (Little & Rubin, 1987, Section 8.2.2). Likelihood second derivatives (for the observed information matrix) are computed analytically for MTM(1) and numerically for MTM(2).

Numerical second derivatives are computed as local slopes from small perturbations of parameter estimates (Press et al., 1992). Let $\theta = (\alpha, \beta, \phi)$ be a vector of length $(p+q+r)$ and $\theta^{(i)}$ be a vector of length $(p+q+r-1)$ with the ith element of θ removed. The ith element of θ, θ^i, is perturbed as follows:

$$\delta = .01\left(\frac{\theta^i}{\theta^i + .001}\right)$$

$$\theta_1^i = \begin{cases} \theta^i + \delta & : \quad \delta \geq .001 \\ \theta^i + .001 & : \quad \delta < .001 \end{cases}$$

$$\theta_2^i = \begin{cases} \theta^i - \delta & : \quad \delta \geq .001 \\ \theta^i - .001 & : \quad \delta < .001 \end{cases}$$

Using current estimates of parameters θ, row i of information matrix $-\frac{\partial^2 \log \boldsymbol{L}}{\partial \theta^2}$ is produced:

$$\frac{\partial^2 \log \boldsymbol{L}}{\partial \theta^2}[i,] = \frac{\frac{\partial \log \boldsymbol{L}}{\partial \theta}\Big|_{\theta^{(i)},\theta_2^i} - \frac{\partial \log \boldsymbol{L}}{\partial \theta}\Big|_{\theta^{(i)},\theta_1^i}}{2\delta}$$

3.3.2 Identifiability

Parameters λ are effectively unidentifiable because response Y_{it} is not observed when $R_{it} = 0$. Although parameters may appear estimable, they will always be unidentifiable (Scharfstein et al., 1999). This problem may be overcome by pooling information from a separate source of cross-sectional data into the likelihood (Fitzmaurice et al., 1996a) or by restrictions to the structure of the dropout model (Laird, 1988). A conservative and practical approach, though, is to perform the analysis repeatedly with parameters λ held constant through a series of plausible values. This sensitivity analysis examines the impact of the fixed values on the remainder of the model (Rotnitzky et al., 1998; Kenward & Molenberghs, 1999; Scharfstein et al., 1999).

3.3.3 Standard Errors and Inference

When data are MAR, the marginalized transition model meets the general regularity conditions for maximum likelihood estimation. We assume that response vectors Y_i are independently and identically distributed. Under MAR, or for NI models with λ held at fixed values, the likelihood function is identifiable, as suggested by the derivations above. Finally, the likelihood function should be sufficiently smooth to compute required derivatives with finite expectation. Computations with respect to parameters (β, α, ϕ) are shown above. All these parameters are unbounded in the parameter space, so parameter constraints are not an issue. Although multiple roots may be found for score equations, the likelihood function may be examined to determine which maximizes the likelihood.

Since these regularity conditions are met, correctly specified MTM likelihood estimates will have the expected maximum likelihood properties for sufficiently large samples. Parameter estimates will be consistent: they will converge in probability to the true value of the parameter. The observed information matrix for parameter vector $\theta = (\beta, \alpha, \phi)$ is computed above for use in Newton-Raphson solutions of the score equation:

$$\mathrm{I} = \sum_{i=1}^{N} -\frac{\partial^2 \log \boldsymbol{L}_i(\theta)}{\partial \theta^2}$$

For sufficiently large N $(N \to \infty)$, $\sqrt{N}(\hat{\theta} - \theta)$ converges to a normal distribution with mean

zero and variance $(\lim_{N\to\infty} \frac{1}{N}\mathrm{I})^{-1}$ (the inverse of the information matrix). Finally, the estimates will attain asymptotic efficiency, so that other consistent, asymptotically normal estimates will have asymptotic variance greater or equal to that of the maximum likelihood estimates.

GEE has popularized the use of sandwich standard errors (White, 1982) to provide robust variance estimates in the presence of misspecification. Sandwich standard errors could be valuable if α is misspecified for the MTM (Heagerty, 2002), but only if missing data are MCAR. If data are MAR or NI, the expected value of the standard error may not be the same over $f(\mathrm{Y}_i^o)$ as over $f(\mathrm{Y}_i, \mathrm{R}_i)$. Thus, we compute standard errors using I^{-1}, the inverse of the observed information matrix. Nested marginalized transition models can be compared by likelihood ratio tests or Wald tests.

An approximate score test can also be performed to evaluate whether a higher-order transition model would improve fit. The score test for third-order transitions evaluates the likelihood with third-order transition model:

$$\eta_{it}^C = \mathrm{logit}(\mu_{it}^C) = \Delta_{it} + \sum_{j=1}^{3} \alpha_j \cdot y_{it-j}.$$

The score test statistic for α_3 is:

$$\begin{aligned} \left.\frac{\partial \log\left[\boldsymbol{L}_i(\beta,\alpha,\phi)\right]}{\partial \alpha_3}\right|_{\alpha_3=0} &= \sum_{i=1}^{N}\sum_{t=4}^{n_i}(y_{it}-\mu_{it}^C)\frac{\partial \eta_{it}^C}{\partial \alpha_3} \\ &= \sum_{i=1}^{N}\sum_{t=4}^{n_i} y_{it-3}(y_{it}-\mu_{it}^C) \qquad (3.5) \\ &\quad \text{if } \frac{\partial \Delta_{it}}{\partial \alpha_3} = 0. \end{aligned}$$

Since μ_{it}^C is constructed so that dependence on Y_i should be through α not Δ_{it}, it seems reasonable to dismiss $\frac{\partial \Delta_{it}}{\partial \alpha_3}$. The variance of the approximate score test is approximated by the square of the score. Although Equation 3.5 may only be an approximate score test statistic, it has intuitive appeal since it is related to the correlation between the third-order lagged response Y_{it-3} and the second-order conditional residual (Heagerty, 2002).

3.3.4 Computation

Although the marginalized transition model appears to satisfy regularity conditions for maximum likelihood estimation, asymptotic properties may not be attained for moderate sample sizes. The response for the MTM is merely a series of zeroes and ones: models for binary data can be difficult to characterize completely. In particular, if marginal probabilities are very small or very large, transition parameters α, as well as Δ_{it}, may diverge toward $\pm\infty$ and be difficult to estimate. Numerical adjustments are made in the algorithm for solving MTM score equations, to accommodate numerical difficulties that arise for finite samples. As shown in the previous section, perturbations of score parameters are an absolute size (.01) for large-valued parameters and a percentage (1%) for smaller values. Marginal mean fitted values near the boundary (within 10^{-8} of 1 or 0) are trimmed to 10^{-8} or $1-10^{-8}$, since $\mu_{it}^{M}(1-\mu_{it}^{M})$ appears in the denominator of score and information computations and could cause estimates to "blow up". If transition data are sparse (for example, if few subjects show $Y_{it} = Y_{it-1} = Y_{it-2} = 1$), a ridge of 0.1 is added to the diagonal of the information matrix to stabilize estimation.

Despite the challenges described here, estimation of MTM(1) and MTM(2) parameters is computationally fast, compared to other likelihood-based methods for binary longitudinal data. As described by Heagerty (2002), computational time for the MTM increases linearly with cluster size. For methods described in Chapter 2 (Fitzmaurice et al., 1996b; Molenberghs et al., 1997b), computational time increases exponentially with cluster size.

3.4 Summary

In summary, the marginalized transition model (MTM) models marginal regression parameters, while introducing correlation via a transition model and accommodating dropout by a response-history-dependent selection model. Both the transition model and selection model use observed responses from the past to predict response values and dropout in the present. These features lead to wide applicability for longitudinal binary data with ignorable or non-ignorable dropout. The next chapter examines efficiency and misspecification bias in the MTM, as compared to GEE weighted for nonresponse.

Chapter 4

EFFICIENCY AND MISSPECIFICATION BIAS OF THE MARGINALIZED TRANSITION MODEL AND INVERSE PROBABILITY OF CENSORING WEIGHTED GEE

Based on prior statistical research and known properties of likelihood-based marginalized transition models (MTM) and inverse probability of censoring weighted GEE (IPCW-GEE), we expect:

- Properly specified MTM regression parameter estimates will be more efficient than IPCW-GEE estimates if the association induced by the transition model is closer to the true association model than the IPCW-GEE working correlation.

  Likelihood-based methods are asymptotically more efficient than semiparametric methods. Liang & Zeger (1986) demonstrate efficiency loss in a within-cluster regression parameter for misspecified GEE working correlation, compared to GEE or maximum likelihood methods with the correct correlation structure. Fitzmaurice et al. (1993) find that asymptotic relative efficiency of GEE regression parameters is often quite high, but decreases sharply as marginal correlation increases for binary responses.

- Regression parameter estimates for the MTM are consistent when regression and association are properly specified and data are missing at random (MAR). Missingness must also be modeled properly if missing data are nonignorable (NI).

  As maximum likelihood estimates, MTM parameters will be consistent if the full likelihood for observed data is properly specified. MTM regression parameter estimates may display bias if association is misspecified, or if missingness is NI and not correctly specified. Bias in misspecified likelihood-based models has been demonstrated in simulation studies (Troxel et al., 1997). Asymptotic relative bias for specific models

can be computed by taking the expected value of the fitted score equation over the distribution of the true data-generating model (White, 1982; Fitzmaurice et al., 1993; Rotnitzky & Wypij, 1994; Fitzmaurice et al., 1995; Heagerty & Kurland, 2001). Most studies cited find less than 10% bias in regression models fitted by maximum likelihood methods with misspecified association. Under NI missingness, likelihood-based models assuming MAR show up to 50% bias for within-cluster covariates (Fitzmaurice et al., 1995). For most regression and selection models examined in prior research, cluster-level covariates (such as treatment group comparisons) are relatively robust to misspecification of association and missingness.

- Regression parameter estimates for IPCW-GEE are consistent when regression parameters are properly specified. Missingness must also be modeled properly if missing data are MAR or NI.

  IPCW-GEE regression parameter estimates will be consistent if the missingness is modeled properly (Robins et al., 1995). If conditions for consistent estimation are not met, IPCW-GEE regression will be biased (Fitzmaurice et al., 1993; Fitzmaurice et al., 1995; Troxel et al., 1998; Ziegler & Kastner, 2000). In the misspecification scenarios examined by published studies, cluster-level covariates show less than 10% bias. Bias in published studies is greater for IEE (GEE without correlation or selection weights) compared to GEE without selection weights. When MAR dropout is assumed, IPCW-GEE and likelihood-based models show similar magnitude of bias when missingness is actually NI.

This chapter uses simulation studies to explore relative efficiency and misspecification bias magnitude for the MTM and IPCW-GEE. Regression, association, and missingness parameter values are chosen to approximate possible outcomes of a clinical trial comparing performance of two treatments over time for a binary response.

4.1 *Simulation Details*

Longitudinal binary data with dropout are simulated via a second-order marginalized transition model. The regression model is chosen to represent a common medical application: two treatments compared over time through main effects and an interaction. Response Y_{it} may be interpreted as the presence ($Y_{it} = 1$) or absence ($Y_{it} = 0$) of an undesirable symptom for subject i at time t. The marginal regression model is:

$$\text{logit}\left[\mathrm{E}(Y_{it}|X_{it})\right] = \beta_0 + \beta_1 \cdot \texttt{group}_i + \beta_2 \cdot \texttt{time}_{it} + \beta_3 \cdot \texttt{group}_i \cdot \texttt{time}_{it} \qquad (4.1)$$

where for each subject (cluster) i, $\texttt{group}_i$ is either 0 or 1 (control and treatment, respectively), $\texttt{time}_i$ = (-4, -3, -2, -1, 0), and β = (0.40, -1.25, 0.65, -0.31). The `time` vector ends in 0 to facilitate end-of-study group comparisons. The model is plotted in Figure 4.1. Response levels are nearly equivalent for the two treatments at baseline (`time` = -4), but the response event probability increases at a faster pace for the control group. By the end of followup (`time` = 0), the odds of observing the event for a control group subject are 3.5 times the odds for a subject receiving active treatment.

Association among responses is generated through second-order transitions. The conditional mean $\mu_{it}^C = \mathrm{E}(Y_{it}|X_{it}, Y_{it-1}, Y_{it-2})$ is related to covariates and prior responses:

$$\text{logit}(\mu_{it}^C) = \Delta_{it} + \log(8) \cdot y_{it-1} + \log(2) \cdot y_{it-2}. \qquad (4.2)$$

Holding other covariates constant, the odds of a positive response are 8 times higher for the current response if the previous response is positive, and 2 times higher if the response two timepoints earlier is positive. As described in Chapter 3, intercept Δ_{it} can be derived from the transitions and Equation 4.1 and is not modeled directly. For clusters of size 5, this association structure produces correlations among observations that are similar to an autoregression structure, $\text{corr}(Y_{ij}, Y_{ik}) \approx \rho^{|j-k|}$ with $\rho \approx 0.4$.

Models for monotone dropout are based on a logit-linear selection model with group-by-response interactions (and no main effect for `group`):

$$\begin{aligned}\text{logit}\left[P(R_{it} = 1|R_{it-1} = 1, Y_{it}, H_{it}^{(Y)}, X_{it})\right] &= \phi_0 + \phi_1 \cdot y_{it} + \phi_2 \cdot y_{it-1} + \phi_3 \cdot y_{it-2} \\ &+ \phi_4 \cdot \texttt{group}_i \cdot y_{it} + \phi_5 \cdot \texttt{group}_i \cdot y_{it-1} + \phi_6 \cdot \texttt{group}_i \cdot y_{it-2}. \qquad (4.3)\end{aligned}$$

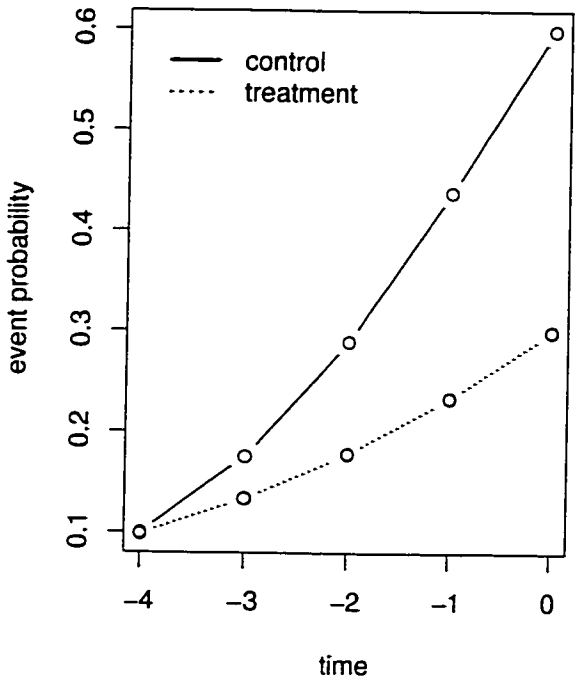


Figure 4.1: Marginal mean regression model used for simulations.

The simulation studies below focus on MAR and NI dropout for two classes of selection models. The first class is a relatively simple missingness model, dependent on main effects of the previous two response values. For these models, $(\phi_1, \phi_4, \phi_5, \phi_6) = 0$ when data are MAR. ϕ_0 is chosen so that overall 20% or 50% of data are missing. Parameters ϕ_2 and ϕ_3 generate three patterns of dropout based on level and increment in the response (Diggle & Kenward, 1994):

1. **level:** Dropout mechanism depends on level of response over previous two time periods, $(Y_{it-1} + Y_{it-2})$. For (ϕ_2, ϕ_3) = (-0.75, -0.75), $(\phi_2 + \phi_3)/2$ = -0.75, and $(\phi_2 - \phi_3)/2 = 0$. Dropout is most likely when $Y_{it-1} = Y_{it-2} = 1$ and least likely when $Y_{it-1} = Y_{it-2} = 0$.

2. **increment:** Dropout mechanism depends on increment (trend) in the response, $(Y_{it-1} - Y_{it-2})$. Here (ϕ_2, ϕ_3) = (-0.75, 0.75), $(\phi_2 + \phi_3)/2 = 0$, and $(\phi_2 - \phi_3)/2$ = -0.75. Dropout is most likely when $Y_{it-2} = 0$ and $Y_{it-1} = 1$, and least likely when $Y_{it-2} = 1$ and $Y_{it-1} = 0$. In our interpretation (Y_{it} as an undesirable symptom), these extremes correspond to deterioration and improvement, respectively.

3. **current:** Dropout depends on Y_{it-1} and not Y_{it-2}: (ϕ_2, ϕ_3) = (-1.5, 0.0). The mechanism is a first-order version of the "level" pattern. Dropout is more likely if the event occurs (Y=1).

Values of ϕ_2 and ϕ_3 are chosen so that an odds ratio of 4.5 is the largest contrast among Y_{it-1} and Y_{it-2} in each pattern.

The second class of models introduces an interaction between lagged responses and group membership in the selection model. For the MAR version of this class, values for ϕ_2 and ϕ_3 in Equation 4.3 are as described above, and $\phi_5 = -\phi_2$ and $\phi_6 = -\phi_3$. For example, the group-by-response MAR selection model for the "level" pattern is:

$$\begin{aligned}\text{logit}\left[P(R_{it} = 1 | R_{it-1} = 1, Y_{it}, H_{it}^{(Y)}, X_{it})\right] &= \phi_0 - 0.75 \cdot y_{it-1} - 0.75 \cdot y_{it-2} \\ &\quad + 0.75 \cdot \text{group}_i \cdot y_{it-1} + 0.75 \cdot \text{group}_i \cdot y_{it-2}\end{aligned}$$

where ϕ_0 is chosen so that 20% or 50% of data are missing. The control group experiences level and increment effects as described above, but the treatment group's missingness is MCAR since the response-dependent terms cancel. This could occur if the treatment has a benign effect unrelated to the response. For example, if a treatment for high blood pressure had mood-enhancing effects, treatment group patients may be equally likely to continue, regardless of whether they experienced angina. In the control group, though, improvement or declining health may affect participation. Nonignorable (NI) missing data models are generated by shifting forward the parameter values: values for $(\phi_2, \phi_3, \phi_5, \phi_6)$ described for MAR models are the values for $(\phi_1, \phi_2, \phi_4, \phi_5)$ in the corresponding NI model, and $\phi_3 = \phi_6 = 0$.

For each of 1000 simulations, a data set with 500 subjects (250 "treatment" and 250 "control") is generated, and several models fitted. Relative efficiency is the ratio of sample

variances of parameter estimates. Relative bias as a percentage of true value β is computed based on the average value of the 1000 sets of parameter estimates $\hat{\beta}_j$:

$$\hat{\beta} = \sum_{j=1}^{1000} \hat{\beta}_j/1000$$

$$\text{percent bias} = \frac{100 \cdot (\hat{\beta} - \beta)}{\beta}.$$

For models with MAR data, correctly specified models are fitted to evaluate efficiency. For NI missingness, correctly specified models are also fitted to validate the simulation characteristics and the fitting algorithms. In most cases, finite sample bias in the "true" regression model parameters based on 1000 simulations is under 5 percentage points when data are MAR, for both IPCW-GEE and the MTM. Bias is eliminated almost entirely by doubling the number of observations to 1000 in each simulated dataset.

Of 18 correctly specified NI models, 4 show greater than 10% bias in a regression parameter (intercept β_0). The largest percentage bias in a regression parameter is -30%, for the intercept of a NI main-effects current-response selection model with 50% missing data (C50NI-main). The bias has little effect on predicted probability of response for the simulation control group at `time` $= 0$: the true probability is 0.599, fitted as 0.570 by the biased estimator.

Bias in regression intercepts estimated for NI models is not reduced by increasing sample size or number of simulations. Rather than small sample size, bias appears to be due to numerical problems in the estimation algorithm. Association parameters α converge quickly to highly attenuated values: first-order parameters are underestimated as closer to log(5.5) than to log(8), and second-order parameter estimates are close to the null value of log(1). Bias in the C50NI-main regression intercept is lowered to -7% by increasing cluster size (without missing data) from 5 to 20, while retaining 50% missing data overall. A model with regression parameters $\beta = (1.2, -2.5, 0.65, -0.31)$, for which marginal probabilities are higher than for the regression model in Figure 4.1, only shows -11% bias in the intercept, compared to -30% for the model with lower prevalence. Bias in β_0 is only -4% in model C20NI-main, with only 20% of data missing. These adjustments suggest that data sparseness is the source of numerical difficulty in model C50NI-main. Proportion of positive response in

observed Y values at `time` $= 0$ is lower in scenarios showing bias in β_0. Comparable models without bias have similar patterns of missing data (i.e. number of complete cases, and number of subjects with only one response), but are more likely to show positive responses at later timepoints. Sparseness in contingency tables of responses and lagged responses could lead to poor estimation of α, and thus of β. However, increasing the sample size to 2000 (from 500) does not reduce bias for the models described.

4.1.1 Efficiency

Asymptotic relative efficiency of IPCW-GEE regression parameter estimates is compared to the second-order MTM. Data are generated using the regression, association, and MAR missingness models described above. Regression and missingness models (and association for the MTM) are correctly specified, so that regression estimates are consistent. IPCW-GEE models are fitted using independence, exchangeable, and autoregressive working correlations.

We expect IPCW-GEE to be less efficient than the MTM. IPCW-GEE efficiency should be highest for the model with autoregressive working correlation, since it is closest to the true model. Efficiency is also compared for cluster-level (`intercept` and `group`) versus within-cluster covariates (`time` and `group:time`), percent missing data (20% versus 50%), different missingness patterns ("level", "current", and "increment" as described above), and correctly specified versus overfitted selection weights in the IPCW-GEE. Efficiency gain with overspecification of IPCW-GEE selection models is described by Robins et al. (1995) and demonstrated by Troxel et al. (1997), Preisser et al. (2000), and Collins et al. (2002).

Finally, IPCW-GEE efficiency for a richer selection model (with group-by-response interactions) is compared to efficiency where the selection model includes only main effects. Efficiency may be poorer when dropout is based on a more complicated selection model, since IPCW-GEE must estimate more dropout parameters. We examine whether the selection model with group-by-response interaction has a greater impact on efficiency for terms involving `group` in the regression model, compared to main-effects selection.

4.1.2 Bias under MAR Dropout

We evaluate bias under MAR dropout for the MTM (with association misspecified) and IPCW-GEE (with missingness misspecified). Data are generated under the regression, association, and MAR missingness models described above. In all cases the regression mean (Equation 4.1) is correctly specified, but there is misspecification elsewhere for the following fitted models:

1. **IEE:** GEE with independent correlation structure (a generalized linear model) and no weights to adjust for missing data.

2. **IPW:** IPCW-GEE with autoregressive (AR1) working correlation and underfitted missingness. Three models are fitted, described using the notation of Equation 4.3:

   (a) **IPW1:** Selection depends on Y_{t-1}. $(\phi_1, \phi_3, \phi_4, \phi_5, \phi_6)$=0:

   $$\text{logit}\left[P(R_{it} = 1 | R_{it-1} = 1, Y_{it}, H_{it}^{(Y)}, X_{it})\right] = \phi_0 + \phi_2 \cdot y_{it-1}$$

   (b) **IPW2:** Selection depends on main effects of Y_{t-1} and Y_{t-2}. $(\phi_1, \phi_4, \phi_5, \phi_6)$=0:

   $$\text{logit}\left[P(R_{it} = 1 | R_{it-1} = 1, Y_{it}, H_{it}^{(Y)}, X_{it})\right] = \phi_0 + \phi_2 \cdot y_{it-1} + \phi_3 \cdot y_{it-2}$$

   (c) **IPW3:** Selection depends on a group-by-response interaction for Y_{t-1} only. $(\phi_1, \phi_3, \phi_4, \phi_6)$=0:

   $$\text{logit}\left[P(R_{it} = 1 | R_{it-1} = 1, Y_{it}, H_{it}^{(Y)}, X_{it})\right] = \phi_0 + \phi_2 \cdot y_{it-1} + \phi_5 \cdot \texttt{group}_i \cdot y_{it-1}$$

   Note that IPW2 is correctly specified for all patterns when there is no group-by-response interaction in the selection model. All IPW models are correctly specified (or overspecified) for "current" patterns with response main effects, and the IPW3 selection model is correct for the "current" pattern with group-by-response interaction.

3. **MTM(1):** First-order MTM. The true association is second-order, so the likelihood is not correctly specified.

As described above, twelve scenarios are used to generate MAR data: "level", "current", and "increment" patterns for two selection models (main-effects and group-by-response) and 20% or 50% missing data. We examine how percent bias varies for different covariates, for different missingness models, and for different amounts of missing data. We examine the direction of the bias, and the impact of bias on the fitted trajectories.

Additional data are simulated with 50% missing data and the "level" selection pattern to explore the impact of association strength and cluster length on bias. We explore bias in a model with a second-order association of (α_1=log(16), α_2=log(8)). Corresponding parameter values in Equation 4.2 are (log(8), log(2)). The underspecified MTM(1) is expected to show more bias for the model with larger α values. Strength of association may also affect estimation of weights for IPCW-GEE. Since the true association is close to an autoregression structure, bias may be attenuated for larger cluster sizes (n=10 instead of n=5), since missing data will not be correlated strongly with other responses.

4.1.3 Bias under NI Dropout

Under nonignorable (NI) dropout, IPCW-GEE with censoring weights based on MAR missingness does not provide consistent estimates of regression parameters. MTM regression parameters are expected to be biased as well if the dropout model is not correctly specified. We explore the extent of bias when data are generated according to nonignorable dropout based on level, increment, or current response as above, with parameter values shifted so that missingness is nonignorable. Bias is estimated under the following fitted models:

1. **IEE:** GEE with independent correlation structure (a generalized linear model).

2. **IPW:** IPCW-GEE with autoregressive working correlation and underspecified missingness:

   (a) **IPW2:** Selection depends on Y_{it-1} and Y_{it-2}. Dropout parameters fitted are (ϕ_0, ϕ_2, ϕ_3). Missingness actually depends on (ϕ_0, ϕ_1, ϕ_2) from Equation 4.3. Especially for the "increment" model for dropout, perhaps the trend can be

estimated sufficiently fitting parameters to Y_{it-1} and Y_{it-2}, rather than Y_{it} and Y_{it-1}.

(b) **IPW4:** IPCW-GEE with autoregressive working correlation and selection dependent on Y_{it-1} and Y_{it-2}, plus interactions with treatment group. This model is a MAR version of group-by-response selection. We compare models where $(\phi_0, \phi_2, \phi_3, \phi_5, \phi_6)$ are fitted to models where selection is generated based on $(\phi_0, \phi_1, \phi_2, \phi_4, \phi_5)$.

3. **MAR:** Second-order MTM (assumes MAR data).

4. **NI:** Second-order MTM with misspecified selection model for nonignorable missingness:

   (a) **NI1:** Selection model estimates ϕ_0 and ϕ_2 (and ϕ_5 for group-by-response selection), and holds $\lambda = (\phi_1, \phi_4)$ constant as $(-\frac{\phi_1}{2}, -\frac{\phi_4}{2})$. Recall that λ is effectively not identifiable from the likelihood. Will bias be worse when NI missingness is assumed to be in the incorrect direction?

   (b) **NI2:** Second-order MTM for nonignorable data, fitting (ϕ_0, ϕ_2, ϕ_5) and holding λ constant as $(\frac{\phi_1}{2}, \frac{\phi_4}{2})$. Perhaps partial modeling of NI missingness (in the correct direction) results in less bias than a MAR model.

In addition to the 12 basic scenarios (generating data with NI versions of the MAR scenarios in the previous subsection), additional models are fitted with larger cluster size or stronger association, again as described for MAR models. We also conduct simulations to examine whether bias is relative or absolute. Fitzmaurice et al. (1995) examine asymptotic bias of a range of regression parameter values, holding association and percent missing constant. They find evidence that bias is absolute, not relative. To explore this issue further, we fit additional models under the "level" missingness pattern, with 50% of data missing, for both MAR and NI missingness and both main-effects and group-by-response selection models. Models are fitted with $\beta = (1.2, -2.5, 0.65, -0.31)$. This regression model is chosen

to double the **group** parameter value (the primary parameter of interest for treatment differences) while maintaining trajectories shaped similar to those in Figure 4.1.

4.2 Results: Efficiency

Asymptotic relative efficiency (based on 1000 simulations of datasets with 500 subjects) of IPCW-GEE with correctly specified selection models is compared to correctly specified second-order MTMs. When 20% of data are missing, efficiency of IPCW-GEE is 83% or greater for all regression covariates, and is generally over 90%. However, efficiency loss for IPCW-GEE is substantial when 50% of data are missing. Efficiency is as low as 40% (for `group:time` under autoregressive working correlation, and main-effects selection model), and no higher than 89% (for the `intercept` in models based on increment of responses, with exchangeable working correlation).

Efficiency is shown for models with 50% missing data in Figure 4.2 (for the simpler missingness model) and Figure 4.3 (for group-by-response selection). Percent efficiencies for different covariates in the same model are connected by solid lines (for autoregressive working correlation) or dashed lines (for independence working correlation). These results are displayed in tabular form in Tables C.2 and C.3 in Appendix C. These tables and Table C.1 also report efficiency for overfitted selection models, exchangeable working correlation, and for 20% missing data.

Unlike prior research (Fitzmaurice et al., 1993), we do not find GEE efficiency greater for cluster-level covariates. For models with a main-effects dropout model (Figure 4.2), `intercept` has slightly greater efficiency than other covariates, while models with a group-by-response dropout model (Figure 4.3) show greater efficiency for covariates **group** and `group:time`.

Robins & Rotnitzky (1995) and Rotnitzky & Robins (1995) show that overfitting the IPCW-GEE missingness model leads to improved efficiency. This is supported in Tables C.1 and C.2 by comparing columns fitting main-effects ("TRUE") and group-by-response ("OVER") selection with the same working correlation. Efficiency gains of up to 9 percentage points are observed for overfitted IPCW-GEE models compared to cor-

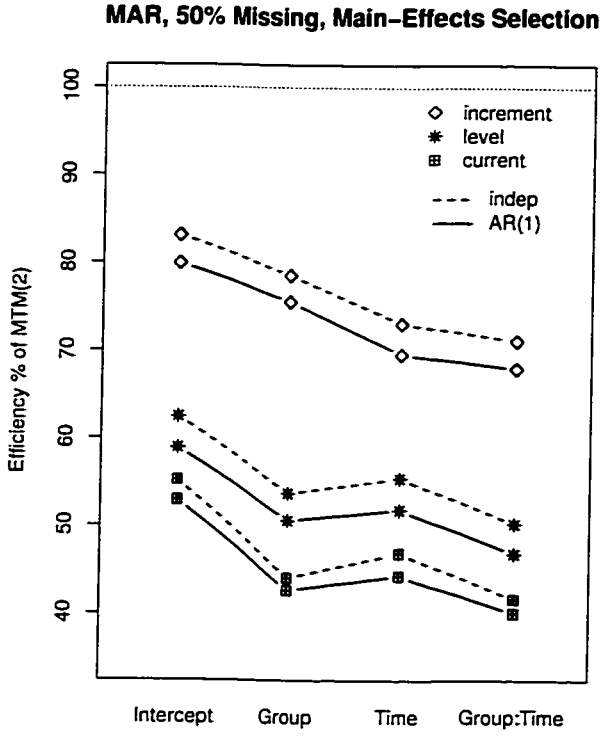


Figure 4.2: Efficiency of IPCW-GEE compared to MTM(2) for correctly specified models with 50% of data missing at random (MAR). Selection model has main effects of first- and second-order lagged responses.

rectly specified IPCW-GEE. The `group` covariate efficiency appears to benefit most from overfitting.

Figures 4.2 and 4.3 suggest that correctly specified (not overfitted) missingness models may result in greater regression parameter efficiency when the missingness model is more complicated. Efficiency for `group` and `group:time` under group-by-response selection models is greater than for the same covariates in the simpler main-effects selection model. A similar counterintuitive trend is seen for missingness patterns. The "current" selection model has fewer parameters to fit, and lower efficiency in all regression covariates compared to "level" and "increment".

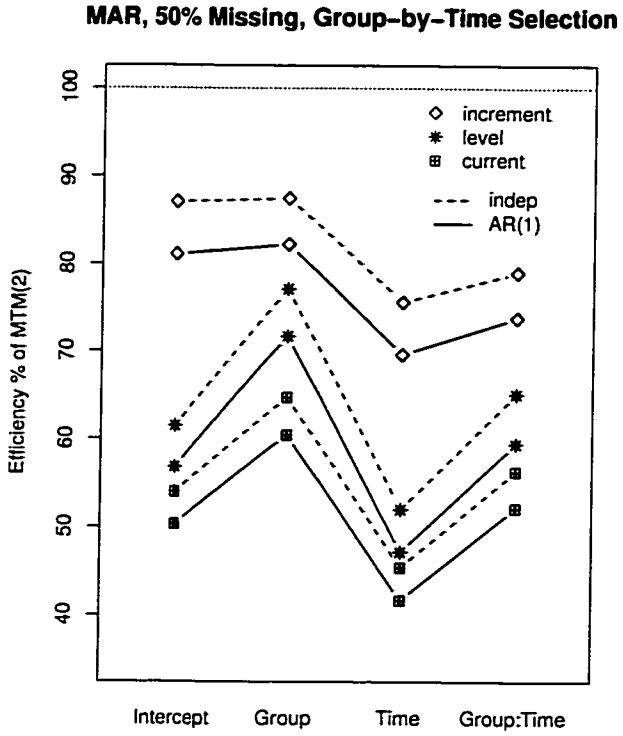


Figure 4.3: Efficiency of IPCW-GEE compared to MTM(2) for correctly specified models with 50% of data missing at random (MAR). Selection model has group-by-response interaction for first- and second-order lagged responses.

Figures 4.2 and 4.3 show a small efficiency advantage for independence working correlation over autoregression. This was unexpected. Previous research shows GEE or IPCW-GEE to be somewhat more efficient with working correlation closer to the true association (Liang & Zeger, 1986; Fitzmaurice et al., 1993; Mancl & Leroux, 1996; Heagerty, 2002). We note first that the efficiency differences are slight, usually less than 5 percentage points compared to the MTM. Also, the cluster size for complete data (5) is rather small, so large efficiency differences are not expected for different working correlations.

Estimation of working correlation parameter α could be related both to efficiency being higher for more complicated missingness models (although the trend is also seen under

working independence), and to the efficiency advantage of working independence. Simulation data are generated to have association close to an autoregression structure with parameter approximately 0.40. However, the largest average fitted value for α (with 50% missing data) is 0.345, under the increment-based scenario with group-by-response missingness. The smallest average value is 0.22, for the current-response scenario with main-effects missingness. It may not be a coincidence that extremes in estimation of working correlation correspond with extremes in efficiency. In the algorithm used to solve IPCW-GEE quasi-scores, working correlation is modeled for observed responses only, not for $\mathrm{R}_{ij}(\mathrm{Y}_{ij} - \mu_{ij}^{M})$ over the joint distribution $f(\mathrm{Y}_i, \mathrm{R}_i)$. Robins et al. (1995) do not specify how to compute working correlation, saying only that it is arbitrary.

In summary, efficiency for IPCW-GEE compared to MTM(2) is as low as 40% for both the main-effects and group-by-response missingness models, when 50% of data are missing. The selection model differs from MCAR only when the response event occurs, which has only 10% outcome probability at early timepoints in the simulations. Estimates of the selection weights are therefore quite variable, leading to efficiency loss compared to the MTM(2). Specification of working correlation close to the true association model does not improve efficiency of IPCW-GEE, but overspecification of the selection model results in modest efficiency gains.

4.3 Results: Bias

4.3.1 MAR

We first consider the simpler dropout model, in which selection is based on main effects of lagged responses. The selection model used to generate data is:

$$\text{logit}\left[P(\mathrm{R}_{it} = 1 | \mathrm{R}_{it-1} = 1, \mathrm{Y}_{it}, \mathrm{H}_{it}^{(\mathrm{Y})})\right] = \phi_0 + \phi_2 \cdot y_{it-1} + \phi_3 \cdot y_{it-2}$$

and the association model (Equation 4.2) is

$$\text{logit}(\mu_{it}^{C}) = \text{logit}\left[P(\mathrm{Y}_{it} = 1 | \mathrm{H}_{it}^{(\mathrm{Y})})\right] = \Delta_{it} + \log(8) \cdot y_{it-1} + \log(2) \cdot y_{it-2}.$$

Percent bias in regression parameter estimates fitted to second-order MTMs with response-dependent selection are shown in Table C.4 in Appendix C. IPCW-GEE model IPW1 fits

selection model

$$\text{logit}\left[P(\mathrm{R}_{it}=1|\mathrm{R}_{it-1}=1,\mathrm{Y}_{it},\mathrm{H}_{it}^{(\mathrm{Y})})\right]=\phi_0+\phi_2\cdot y_{it-1}$$

and first-order marginalized transition model MTM(1) fits association model

$$\text{logit}(\mu_{it}^{C})=\Delta_{it}+\alpha_1\cdot y_{it-1}.$$

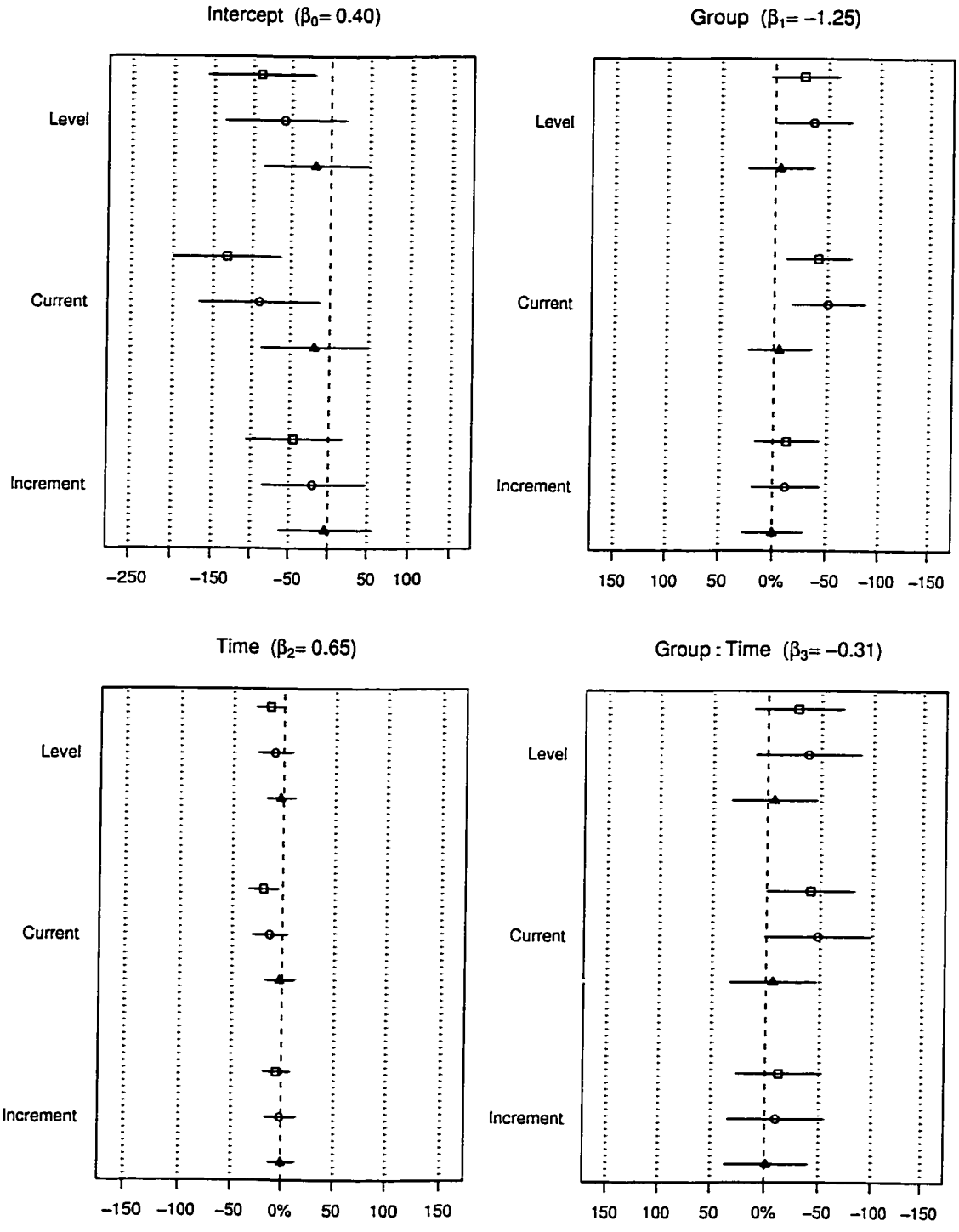
For covariates `intercept` and `time`, a naive IEE model assuming MCAR and independent responses shows up to 54% (when 20% of data are missing) or 125% bias (50% of data missing) toward the null regression parameter. Bias in both the IPW1 model (with missingness underspecified) and MTM(1) (with association underspecified) is not as great: 22% or less for IPW1 and 19% or less for MTM(1) for these two covariates. Bias in `group` and `group:time` is negligible, even for the unweighted GEE model. For 20% and 50% missingness, under selection based on level, increment, and current patterns, bias in these two parameter estimates is no more than 5% for IEE, no more than 3% for IPCW-GEE, and no more than 4% for the MTM. Under the main effects selection model, fitted models show little bias in the treatment group main effects and interactions that are of primary interest in clinical trials. However, bias can be considerable when group-by-response terms appear in the selection model.

Figure 4.4 shows bias of fitted regression parameters for data generated by the group-by-response selection model with 50% missing. Trends in bias for 20% missing data are similar, with about half the magnitude in percent bias. Details are given for both 20% and 50% missing data in Table C.5. The four panels display results for the four covariates. In each panel, three missingness patterns (level, increment, and current response) are displayed. For each pattern, two semiparametric models are fitted, and the average parameter estimate in simulations is shown with a hollow plotting character. MTM average parameter estimates are shown with solid plotting characters. Error bars reflect one standard deviation in each direction for fitted parameters in the simulations. A vertical dashed line represents the true value of each covariate, and vertical dotted lines show percent bias in 50% increments.

For all parameters, MTM(1) bias is relatively small. IPW1 bias is larger (12-53%) for parameters involving `group`. Bias in `intercept` and `time` is somewhat smaller for Model

Figure 4.4: Average regression parameter estimates $\pm$ one standard deviation based on 1000 simulations, for the regression model in Equation 4.1. 50% of data are missing at random (MAR), with selection dependent on a group-by-response interaction.

□	IEE	unweighted logistic regression
○	IPW1	underspecified IPCW-GEE dropout
▲	MTM(1)	underspecified MTM


Intercept (β_0= 0.40)
Group (β_1= −1.25)
Time (β_2= 0.65)
Group : Time (β_3= −0.31)
Level
Current
Increment
−250
−150
−50
50
100
150
0%

IPW1 than for IEE. This suggests that underspecified selection weights are better than no selection weights. IPW2, with second-order response main effects in the missingness model, does not appear to mitigate bias much more than IPW1 (Table C.5), but a first-order selection model with group-by-response interaction (IPW3) shows greatly reduced bias in `group` and `group:time` compared to IPW1. The two cluster-level covariates (`intercept` and `group`) in Figure 4.4 appear to show bias in opposite directions. However, this is an artifact of plotting fitted values while labeling axes by percent bias. The true value for `intercept` is positive, and the true value for `group` is negative. Bias is negative (toward the null value) in each case.

Fitted trajectories for selected models are shown in Figure 4.5. The rows of panels show the three missingness schemes: "level", "current value" (single lagged response), and "increment". The first column is the fitted trajectory for the control group, and the second column is the fitted trajectory for the treatment group. The true regression model (Equation 4.1) is plotted as a solid line, but mostly overlaps with the fitted MTM(1) trajectory. As described above, missingness in the treatment group is MCAR, and this is reflected in fitted models. The misspecified association in MTM(1) yields little bias in estimates for either treatment group. For IEE, MCAR is assumed and is the correct model for the treatment group. Control group estimates are biased toward lower probability of response. People who experience the event are more likely to drop out in all the missingness schemes, yielding underestimation of response values in later time periods. Note that the large percent bias in the intercept ($\beta_0 = 0.4$; up to -132% for the "current" model) yields only a moderate difference in fitted probabilities. At `time` $= 0$, the correct control group event probability is 0.60 (second row, first column of Figure 4.5), but is fitted as 0.47 under IEE. A difference of 13 percentage points can be enormous on a public health scale, but is not as great as the percent bias might suggest. A larger parameter value may show a smaller percent bias: relative versus absolute bias is discussed in Section 4.3.2.

IPCW-GEE model IPW1, with underspecified selection, is closer to the truth than IEE for the control group fitted values, but overestimates response probabilities in the treatment group. Dropout is modeled as related to response, but the model does not include the group-by-response interaction. Thus fitted values are biased downward for the control group and

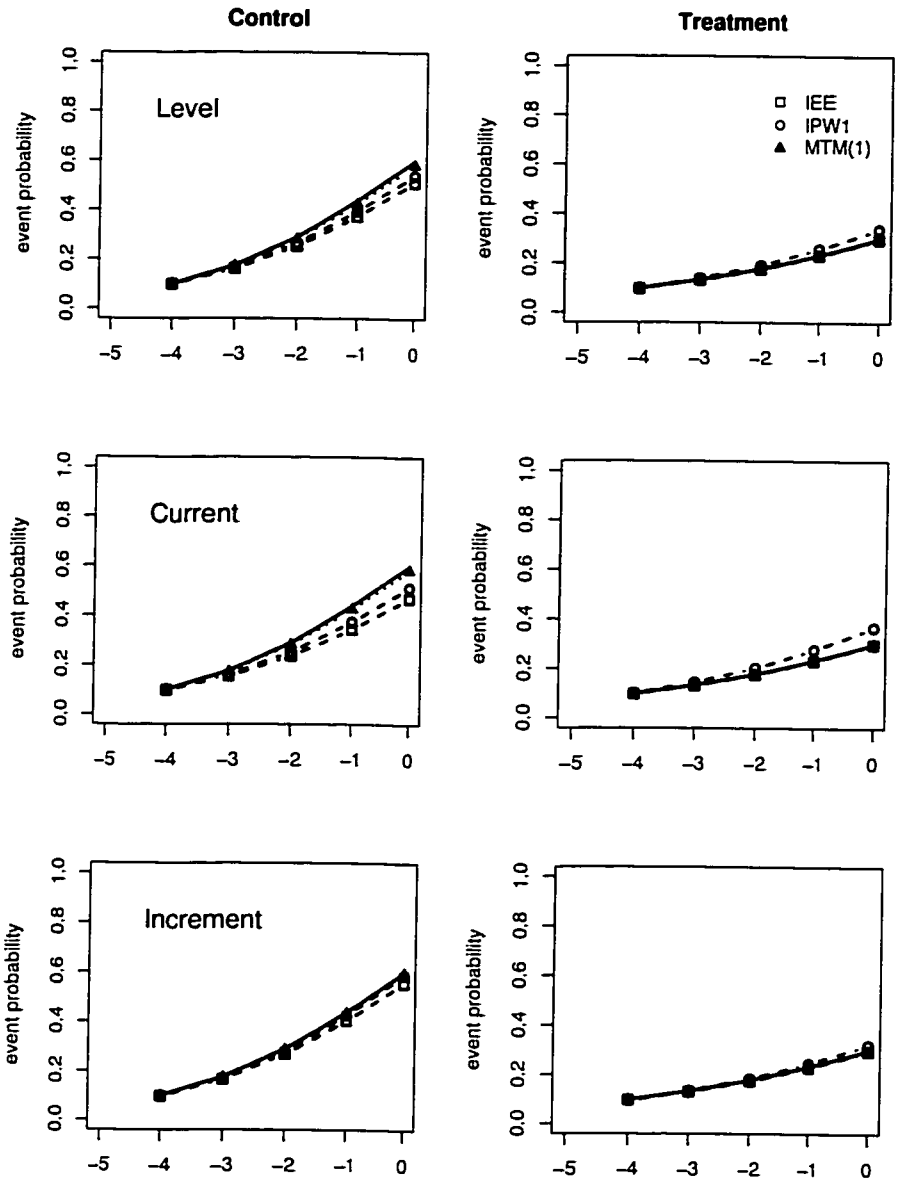


Figure 4.5: Trajectories of average regression parameter estimates fitted by misspecified models under 50% data missing at random (MAR) and selection dependent on a group-by-response interaction.

upward for the treatment group. This dilutes group differences over time, as shown by the bias toward the null in Figure 4.4 for `group` and `group:time`. There is no bias for values

at baseline (time $= -4$). No data are missing at that timepoint, so naive methods (IEE) yield the correct fitted value.

Figures 4.4 and 4.5 show similar bias under "level" and "current" selection models, although the bias is larger for "current", which depends on fewer lagged responses. Bias is smaller overall for "increment", though in the same direction as for other selection patterns. The reasons are not clear except that the increment model appears to deviate less from MCAR than the others. (Compared to level or current patterns, dropout intercept ϕ_0 is larger for increment models, to maintain 20% or 50% missing response.)

The "level" selection pattern with group-by-response interaction and 50% missing data is examined for additional models in Figure 4.6. (Percent bias is shown in tabular form for this and main-effects selection models in Table C.6.) Again, the four panels show results for the four regression parameters. The "Level50" group of estimates are the same as in Figure 4.4. For the "Big Serial" estimates, data are generated with stronger serial correlation. The MTM(1) is missing a parameter of value log(8) instead of log(2): bias is almost double that of the original "Level50" scenario. However, MTM bias is still small compared to bias in semiparametric models. Stronger association also results in greater bias for the IEE and IPCW-GEE models.

The lower set of models in Figure 4.6 generates data with the same regression, association, missingness, and percent missing overall, but the complete data cluster size is 10 instead of 5. Bias is smaller than for the cluster size of 5, but is still as great as 50%. Since the association degrades over time, the MTM(1) likelihood is reasonably correct for association among distant responses. Both MTM and IPCW-GEE may have less bias compared to "Level50" because there is more information (more data per person) even though 50% of data are still missing.

In summary, bias in regression parameters can be substantial for models based on semiparametric methods. Bias only emerges in parameters involving treatment group when a treatment group-by-response interaction is included in the selection model used to generate data. Fitted trajectories (Figure 4.5) illustrate that IEE bias in group is due to overestimation of control group response probabilities. For IPCW-GEE with group-by-response-dependent selection, bias in group occurs when the IPC weights do not distinguish between

MAR missingness in the control group and MCAR missingness in the treatment group. Underspecification of association in the MTM results in smaller bias than underspecification of missingness in the IPCW-GEE model.

4.3.2 NI

Patterns of bias when missingness is nonignorable (NI) are similar compared to the MAR models: IEE models show the most bias, and MTMs assuming MAR show less bias than the IPCW-GEE. However, magnitude of bias is generally larger for NI models. Table C.7 shows results for the simulations with main effects selection described in Section 4.1.3. Bias in the intercept is as great as -407%, but even IEE models show little bias in parameters `group` and `group:time`. Percent bias for models with NI missingness and a group-by-response interaction in the selection model are shown in Table C.8 and Figure 4.7.

Model IPW4 is an IPCW-GEE model with missingness an MAR version (modeling on Y_{it-1} and Y_{it-2}) of the true selection (based on Y_{it} and Y_{it-1}). Bias in IPW4 is somewhat smaller than for marginalized transition model assuming MAR (MTM-MAR) for all missingness patterns. Both generally show less bias than IEE estimates, with the "increment" model as an exception. Perhaps IEE is closer to the truth for increment-based models because these models are closer to MCAR: intercept parameter ϕ_0 in the missingness model is largest in increment-based models to achieve 20% or 50% missing data.

MTM model NI1, with parameters ϕ_1 and ϕ_4 from Equation 4.3 fixed in the wrong direction, does lead to greater bias than for MTM-MAR. Fitting a MTM with NI missingness and (ϕ_1, ϕ_4) at half their true values (in the correct direction) leads to some mitigation of bias. For most models and parameters, percent bias in Model NI2 is about half that of Model MTM-MAR.

Figure 4.8 shows fitted trajectories for the models displayed in Figure 4.7, except for model NI1. The true regression model is the solid line without symbols. All models fit the (MCAR) treatment group values well, even IPW4 with weights from an incorrect group-by-response selection model. For "level" and "current" models, IEE fitted control group trajectories greatly underestimate the response probability at later timepoints. IPW4 and

Figure 4.6: Average regression parameter estimates ± one standard deviation based on 1000 simulations, for additional models. 50% of data are missing at random (MAR), through a selection model dependent on a group-by-response interaction.

□	IEE	unweighted logistic regression
○	IPW1	underspecified IPCW-GEE dropout
▲	MTM(1)	underspecified MTM

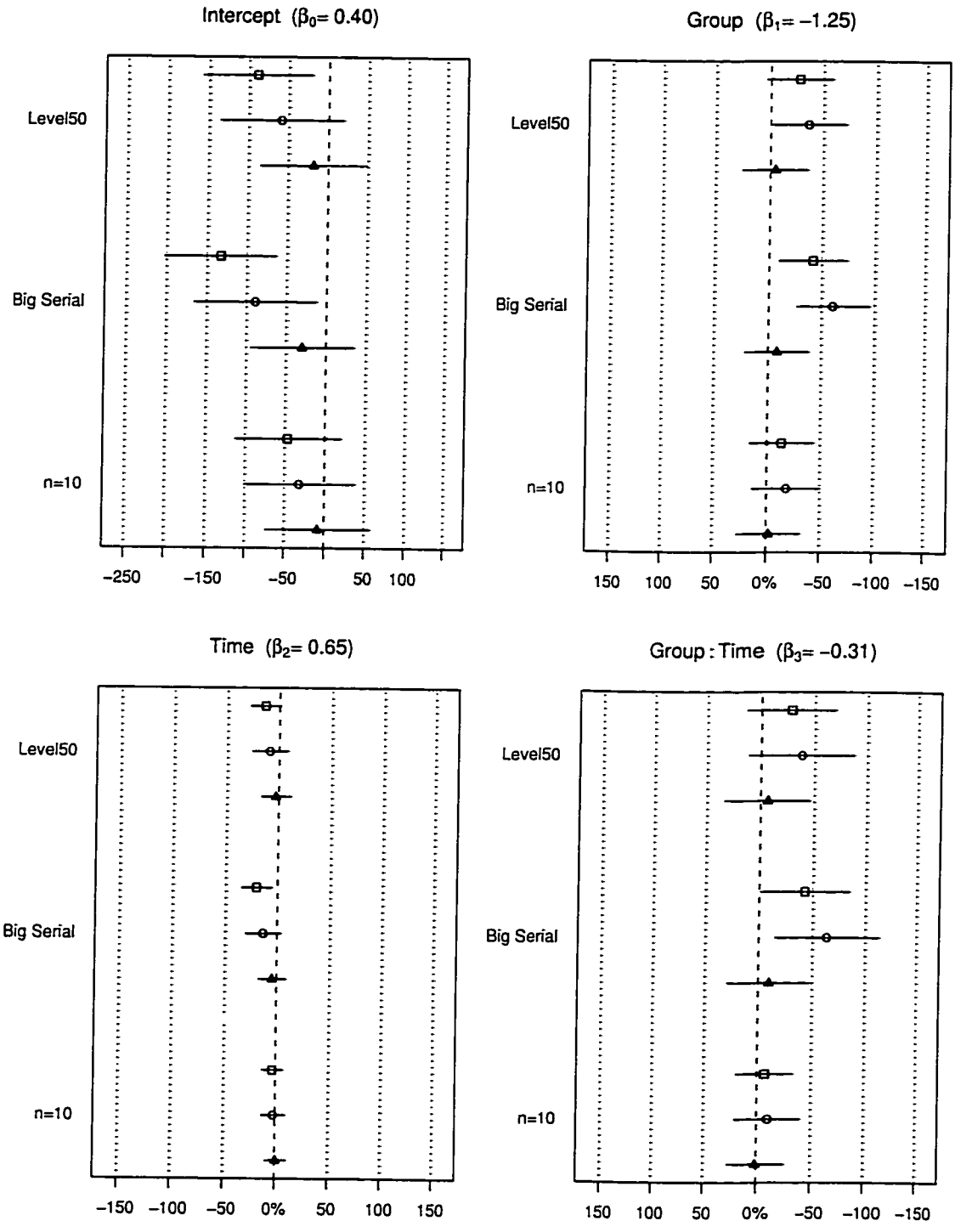
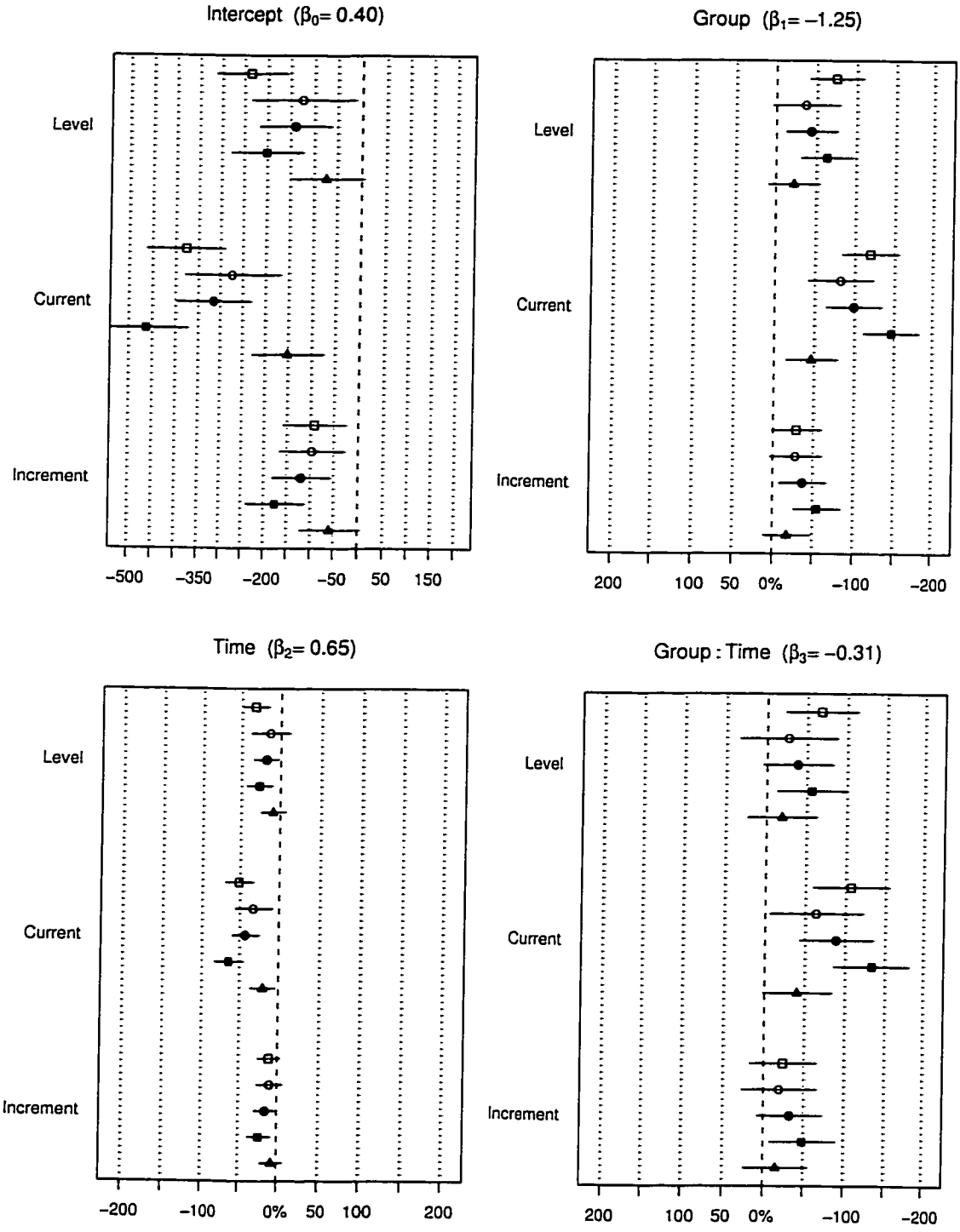

Intercept (β_0= 0.40)
Group (β_1= −1.25)
Time (β_2= 0.65)
Group : Time (β_3= −0.31)
Level50
Big Serial
n=10
−250 −150 −50 50 100
150 100 50 0% −50 −100 −150
−150 −100 −50 0% 50 100 150

Figure 4.7: Average regression parameter estimates ± one standard deviation based on 1000 simulations, for the regression model in Equation 4.1. Missingness (50%) is dependent on current response (NI) through a group-by-response interaction.

□	IEE	unweighted logistic regression
○	IPW4	misspecified IPCW-GEE (MAR)
●	MTM(2)	MTM assuming MAR
■	NI1	MTM with -1/2 true λ
▲	NI2	MTM with 1/2 true λ


Intercept (β_0= 0.40)
Group (β_1= −1.25)
Time (β_2= 0.65)
Group : Time (β_3= −0.31)
Level
Current
Increment
−500 −350 −200 −50 50 150
200 100 50 0% −100 −200
−200 −100 0% 50 100 200
200 100 50 0% −100 −200

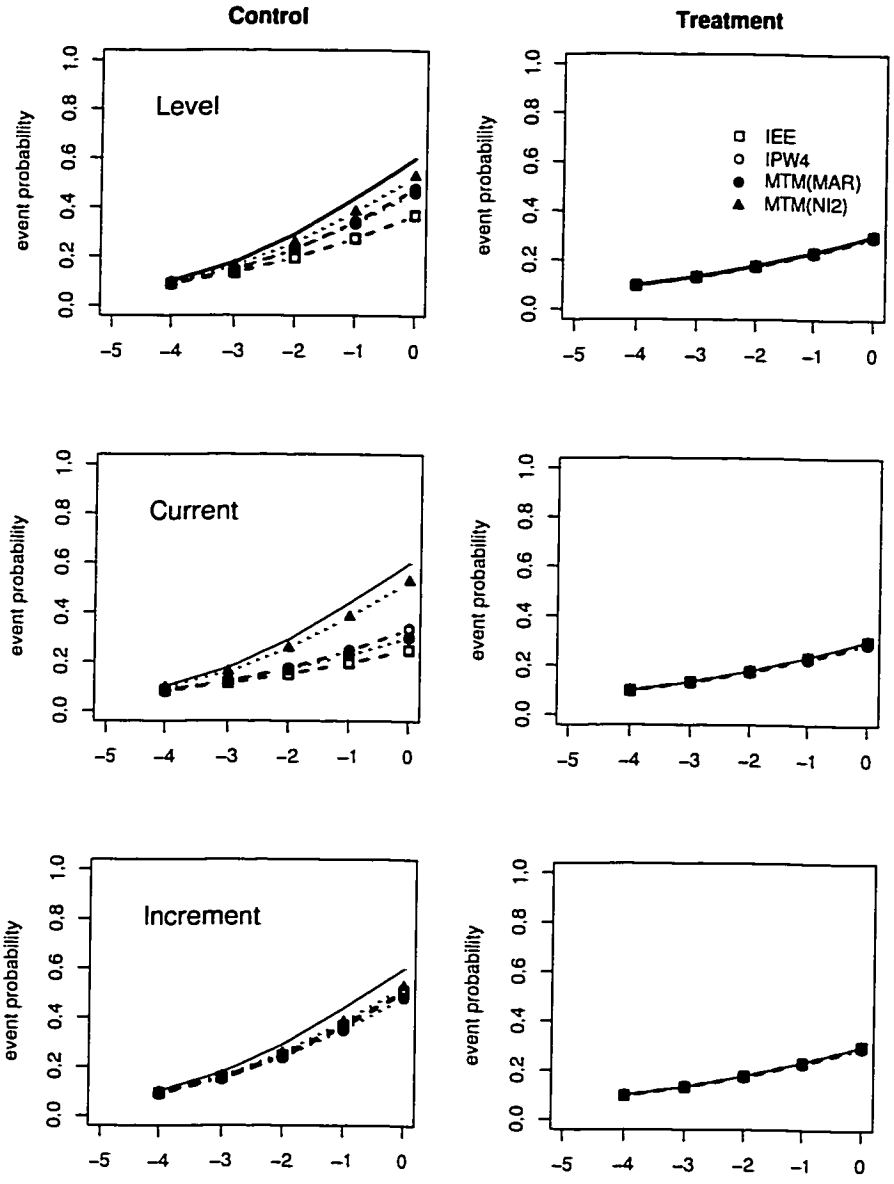


Figure 4.8: Trajectories of average regression parameter estimates fitted by misspecified models under 50% nonignorable (NI) missing data. Selection is dependent on a group-by-response interaction.

MTM-MAR yield values somewhat closer to the truth than IEE, and the NI2 MTM fitted values are quite close to the true regression model. For the "increment" pattern, fitted

values for misspecified models all underestimate control group response rates, but not as badly as for other patterns.

Sensitivity to strength of association and cluster length are examined in Table C.9. Comparisons to the reference model parallel those for the MAR models. When association is stronger, percent bias is slightly greater for most parameters, but the same order of magnitude as for the reference model. Percent bias is reduced for larger cluster sizes, for the same percent of missing data.

Finally, we examine additional simulations to explore whether bias is relative or absolute. Absolute bias in regression parameters for the two regression models are contrasted in Figure 4.9. Relative (percent) bias is shown in Figure 4.10 for the same models. The four columns contain values for the four covariates, and the rows correspond to different selection models to generate missing data. The y-axis range for both figures is from -350% to 50% bias in parameter estimates. Bias is shown for IEE, IPCW-GEE, and MTM models. Percent bias values are shown in tabular form in Tables C.6 and C.9.

Lines connecting bias values for the last two columns are almost all flat, reflecting that the regression parameters (and bias) are similar for both regression models used to generate data. However, they are not identical: differences in `intercept` and `group` values do affect estimation of `time` and `group:time`. Bias for the two values of parameter `group` is similar on both the absolute and relative scale for most models. However, for the group-by-response selection model with nonignorable missing data (bottom row NIint), percent bias in `group` appears smaller for the larger value of β_1. This trend is seen for most models for the intercept (the first column). Absolute bias is relatively constant between the two models, even though the true value for regression parameter on the right is three times the size of that on the left. Percent bias is smaller for larger parameter value. Like Fitzmaurice et al. (1995), we conclude that bias is probably of less concern the farther a parameter is from a null value. Absolute bias appears to be the underlying bias mechanism. However, we continue to report relative bias, because of greater interpretability for parameters in the same regression model, and for different regression models.

Figure 4.9: Absolute bias for parameters in two regression models, for four missingness scenarios with 50% of data missing. "orig" regression parameters on left side of each plot are the fitted coefficients for (`intercept`, `group`, `time`, `group:time`) used in previous simulations: $\beta = (0.4, -1.25, 0.65, -0.31)$. "new" parameters are: $\beta = (1.2, -2.5, 0.65, -0.31)$.

MARmain	MAR, main-effects selection
MARint	MAR, group-by-response selection
NImain	NI, main-effects selection
NIint	NI, group-by-response selection

Missing at Random (MAR) Missingness:

□	IEE	unweighted logistic regression
○	IPW1	underspecified IPCW-GEE dropout
▲	MTM(1)	underspecified MTM

Nonignorable (NI) Missingness:

□	IEE	unweighted logistic regression
○	IPW4	misspecified IPCW-GEE (MAR)
●	MTM(2)	MTM assuming MAR
■	NI1	MTM with -1/2 true λ
▲	NI2	MTM with 1/2 true λ

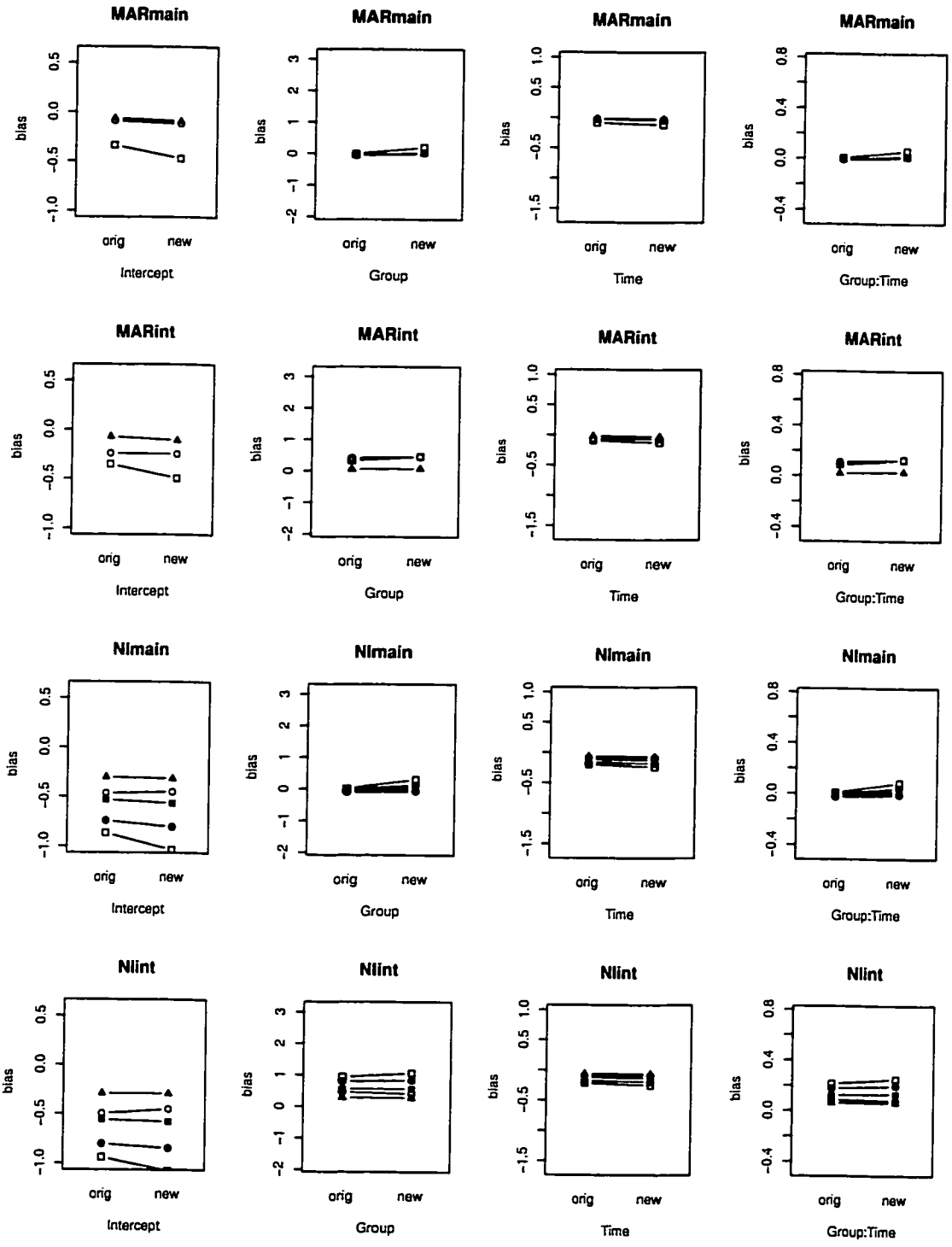

MARmain
MARmain
MARmain
MARmain
MARint
MARint
MARint
MARint
Nlmain
Nlmain
Nlmain
Nlmain
Nlint
Nlint
Nlint
Nlint
bias
orig
new
Intercept
Group
Time
Group:Time

Figure 4.10: Percent bias for parameters in two regression models, for four missingness scenarios with 50% of data missing. "orig" regression parameters on left side of each plot are the fitted coefficients for (`intercept`, `group`, `time`, `group:time`) used in previous simulations: $\beta = (0.4, -1.25, 0.65, -0.31)$. "new" parameters are: $\beta = (1.2, -2.5, 0.65, -0.31)$.

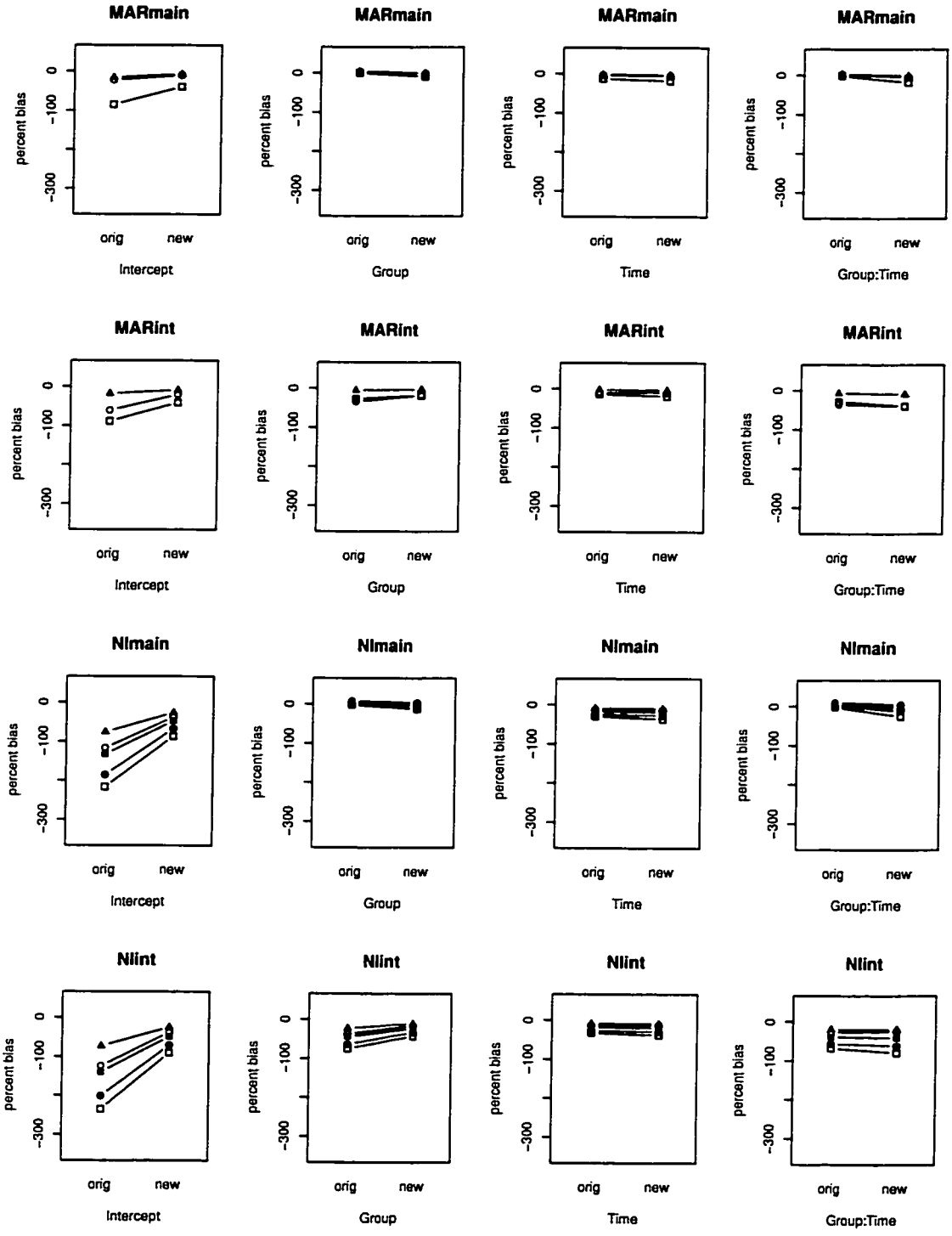
MARmain	MAR, main-effects selection
MARint	MAR, group-by-response selection
NImain	NI, main-effects selection
NIint	NI, group-by-response selection

Missing at Random (MAR) Missingness:

□	IEE	unweighted logistic regression
○	IPW1	underspecified IPCW-GEE dropout
▲	MTM(1)	underspecified MTM

Nonignorable (NI) Missingness:

□	IEE	unweighted logistic regression
○	IPW4	misspecified IPCW-GEE (MAR)
●	MTM(2)	MTM assuming MAR
■	NI1	MTM with -1/2 true λ
▲	NI2	MTM with 1/2 true λ


MARmain
MARmain
MARmain
MARmain
MARint
MARint
MARint
MARint
NImain
NImain
NImain
NImain
NIint
NIint
NIint
NIint
percent bias
0
-100
-300
orig
new
Intercept
Group
Time
Group:Time

4.4 *Discussion (Efficiency)*

Enhancing GEE with IPC weights yields consistent estimation of regression parameters β, but IPCW-GEE as described here is not the most efficient semiparametric estimation method for β (Robins & Rotnitzky, 1995). The semiparametric efficiency bound is the supremum of the Cramér-Rao variance bounds for $\hat{\beta}$ over all parametric submodels for the semiparametric model (Rotnitzky, 1998). Semiparametric efficient estimation theory (Bickel et al., 1993) defines an efficient score function for β under which the semiparametric efficiency bound will be attained.

Semiparametric estimating equations yielding unbiased, regular, asymptotically linear estimators for regression parameter β are of the form:

$$U(\beta) = \sum_{i=1}^{N} \mathbf{D}_i \boldsymbol{W}_i (\mathrm{Y}_i - \mu_i^M) = 0.$$

For IPCW-GEE as an extension of logistic regression for correlated data, $\mathbf{D}_i = \frac{\partial \mu_i^M}{\partial \beta} \boldsymbol{V}_i(\hat{\alpha})^{-1}$ and $\boldsymbol{W}_i$ contains IPC weights as described in Chapter 2. The semiparametric efficient score for longitudinal data with missing values is:

$$U_{\text{eff}}(\beta) = \sum_{i=1}^{N} \mathbf{D}_i^* \boldsymbol{W}_i^* (\mathrm{Y}_i - \mu_i^M) = 0$$

where

$$\begin{aligned}
\mathbf{D}_i^* &= \frac{\partial \mu_i^M}{\partial \beta} \left(\mathrm{E}\left[(\mathrm{U}_i^* - \mathrm{P}_i^*)^{\otimes 2} \middle| \mathrm{X}_i \right] \right)^{-1} \\
\boldsymbol{W}_i^* &= \text{diagonal matrix with elements } \frac{\mathrm{R}_{it}}{\pi_{it}} \\
\pi_{it} &= \prod_{j=1}^{t} \lambda_{ij}^* \\
\lambda_{ij}^* &= \text{weighted sum of fitted selection model probabilities } \lambda_{ij} \text{ and } \mathbf{D}_i^* \mathrm{Q}_{ij}^* \\
\mathrm{U}_i^* &= \boldsymbol{W}_i^* (\mathrm{Y}_i - \mu_i^M) \\
\mathrm{P}_i^* &= \sum_{t=1}^{n_i} (\mathrm{R}_{it} - \lambda_{it} \mathrm{R}_{it-1}) \mathrm{Q}_{it}^* \\
\mathrm{Q}_{it}^* &= \left(\underbrace{0, \ldots, 0}_{1, \ldots, t-1} \underbrace{\frac{\mathrm{G}_{itj}}{\pi_{it}}, \ldots, \frac{\mathrm{G}_{itn_i}}{\pi_{it}}}_{j=t, \ldots, j=n_i} \right)
\end{aligned}$$

$$G_{itj} = E(Y_{ij} - \mu_{ij}^{M} | R_{it-1} = 1, H_{it}^{(Y)}).$$

The form of G_{itj} and $E\left[(U_i^* - P_i^*)^{\otimes 2} \middle| X_i\right]$ are not specified by the model for the marginal mean or missingness model for IPCW-GEE. (There is no need to specify a working correlation among elements of Y_i for the semiparametric efficient score, since $E\left[(U_i^* - P_i^*)^{\otimes 2} \middle| X_i\right]$ provides analogous weights.) Robins & Rotnitzky (1995) note that the optimal estimator is not available for analysis of observed data, and propose a two-stage adaptive estimator that is asymptotically equivalent to the estimator based on the semiparametric efficient score.

The adaptive estimator requires three additional finite-parameter regression models:

$$P\left(R_{ij} = 1 | R_{it-1} = 1, X_{it}, H_{it}^{(Y)}, \chi\right) \text{ for } j \geq t \tag{4.4}$$

$$E\left(\frac{\pi_{it-1}}{\pi_{ij}}(Y_{ij} - \mu_{ij}^{M}) | R_{ij} = 1, H_{it}^{(Y)}, \tau\right) \text{ for } j \geq t \tag{4.5}$$

$$E\left[(U_i^* - P_i^*)^{\otimes 2} \middle| X_i, \Psi\right]. \tag{4.6}$$

Equations 4.4 - 4.6 correspond to Equations (11), (12), and (13) of Robins & Rotnitzky (1995). Equations 4.4 and 4.5 model the dropout indicator R_{ij} and response Y_{ij} (the latter weighted by IPC weights for time j and t) conditioning on the partial history of R_i and Y_i until time t, where $j \geq t$. Equation 4.6 models the (p x p) covariance matrix for a penalized quasi-score used in estimation of the semiparametric efficient score. These three models (estimating parameter vectors χ, τ, and Ψ) add considerable overhead to the modeling process. IPCW-GEE requires only the marginal regression model (to estimate parameters of interest), working correlation (including independence), and a dropout model that adds complexity linearly by cluster length and model order. By contrast, the models required for adaptive estimates of efficient IPCW-GEE scores add $\binom{n_i}{2}$ models in addition to estimating (β, ϕ): one model for each combination of t and $j \geq t$. The number of additional parameters is large, and the semiparametric efficiency bound is not attained if these models are misspecified. Although consistent estimation of regression parameters β is robust to misspecification of the models in Equations 4.4 - 4.6, the modeling burden is equivalent to that of *both* weighting methods such as IPCW-GEE and imputation (Equation 4.5) such as the sequential imputation of Paik (1997).

Efficiency gain of adaptive estimates compared to IPCW-GEE does not appear to be

substantial in finite samples (Robins & Rotnitzky, 1995, Table 1). For simulated linear models with 4 timepoints, the fitted mean at the fourth timepoint is about 90% efficient for IPCW-GEE compared to semiparametric efficient weighted estimating equations, with or without adaptive estimation. Robins & Wang (1998) showed more substantial efficiency gain in a logistic regression model for a simulated two-phase study. The effective sample size increased more than five-fold by use of locally efficient estimators. The simulations in this chapter show that IPCW-GEE estimates can be much less efficient than maximum likelihood regression parameter estimates. However, for longitudinal studies with many observations per subject, or with continuous covariates, the potential benefit of semiparametric efficient estimation is less clear due to modeling overhead.

4.5 Summary

In conclusion, efficiency of IPCW-GEE regression parameters is as low as 40% compared to the marginalized transition model (MTM). Because we do not know the semiparametric efficiency bound for the models examined, we do not know how much of this efficiency loss is due to use of a semiparametric method, and how much is due to the choice of working correlation and dropout model for the semiparametric method. Independence estimating equations (IEE) generally show the largest relative bias of models considered. When data are missing at random (MAR), IPCW-GEE regression parameters with misspecified selection show more bias than a MTM with misspecified association, although there is no scale on which to make misspecifications comparable. When dropout is nonignorable (NI) and IPCW-GEE and MTM both assume data are MAR, the MTM generally (but not always) shows greater bias in regression parameters. Fitting a NI MTM model with current response parameters fixed in the "right direction" shows less relative bias compared to a MTM assuming MAR. In a regression model with treatment group effects and group-by-treatment interactions, covariates involving treatment group show little bias when the selection model generating dropout depends only on main effects of prior responses. However, when a group-by-response interaction is introduced in the selection model, bias toward the null value in treatment group parameters can be substantial under both MAR and NI missingness.

Chapter 5

ANALYSIS OF PANSS SCHIZOPHRENIA DATA BY MARGINALIZED TRANSITION MODELS (MTM) AND INVERSE PROBABILITY OF CENSORING WEIGHTED GEE (IPCW-GEE)

This chapter applies the marginalized transition model (MTM) for binary data with monotone dropout in a double-blind phase III longitudinal study comparing treatments for schizophrenia. MTM results are compared to models fitted using independence estimating equations (IEE) and inverse probability of censoring weighted generalized estimating equations (IPCW-GEE). We also examine sensitivity of MTM parameter estimates to assumptions about missing at random (MAR) and nonignorable (NI) dropout processes.

The schizophrenia study protocol randomly assigned four hundred thirty-one patients to one of six treatments: placebo, 20 mg/day of haloperidol, or one of four doses (2 mg, 6 mg, 10 mg, 16 mg/day) of risperidone. Randomization and treatment commencement occurred after a one-week washout period where prior treatment ceased and participants took placebo. Treatment continued for 8 weeks. Details about eligibility criteria and clinical results of the study are reported elsewhere (Chouinard et al., 1993; Marder & Meibach, 1994).

The Positive and Negative Syndrome Scale (PANSS: Kay et al., 1987), an established instrument for measuring severity of schizophrenia, was administered at selection into the trial, after the washout period, and after 1,2,4,6, and 8 weeks of treatment. Diggle (1998) models these 7 total PANSS scores as an illustration of models for dropout in data with a continuous outcome. Clinical studies (Chouinard et al., 1993; Marder & Meibach, 1994) use a binary primary outcome: clinically significant improvement, measured as a 20% reduction in PANSS compared to baseline. Analyses published in the clinical literature compare treatment effectiveness after 8 weeks, carrying the last measured score forward if the 8-week score is missing. On average, subjects taking the three highest doses of risperidone are more likely to show clinical improvement after 8 weeks of treatment than members of the

haloperidol and the placebo groups. We consider clinically significant improvement as the study response, and examine three treatment groups: placebo, haloperidol, and risperidone (6-16 mg combined), over the sequence of 5 post-baseline scores.

Treatment assignments for the 431 people selected for the study are: 87 placebo, 87 haloperidol, and 257 in the combined risperidone group (6 mg, 10 mg, or 16 mg/day). Two subjects are excluded from the analysis for missing baseline scores, and nine are excluded for missing all post-baseline scores. Additionally, two observations of two subjects are deleted in order to make the dropout pattern monotone. For the purpose of analysis, then, the number of participants is 420, with a total of 1609 response measurements.

5.1 Modeling Considerations

Figure 5.1 describes dropout patterns by group. The risperidone group clearly suffers less dropout than the other groups: almost twice as many subjects in the risperidone group (63%) complete all 5 measurements as in the placebo group (34%). Exploratory analysis also shows group differences in clinical improvement. Figure 1.1 in Chapter 1 models the percentage of patients at each time point who had a clinically significant PANSS score reduction. Each treatment group demonstrates a time trend: a greater percentage show improvement at later times. Individual profiles of improvement over time (not shown) reveal variability in patterns of improvement. The most common pattern is for improvement never to occur in up to 5 measurements ($Y_{it} = 0$ for $t = 1, \ldots, 5$ where Y_{it} is the dichotomous response for subject i at time t): 59 patients (69%) in the placebo group, 53 (62%) in the haloperidol group, and 107 (43%) in the risperidone group never show improvement. Of those who do attain a clinically significant PANSS reduction, some subjects show the reduction followed by a relapse, while others maintain improvement over time, and a few show improvement immediately following baseline and maintain it. Since improvement is not always maintained, we model the entire response profile, rather than only the 8-week response or time until first improvement. Inference about treatment differences is based on the estimated probability of clinically significant improvement at the end of followup (8 weeks).

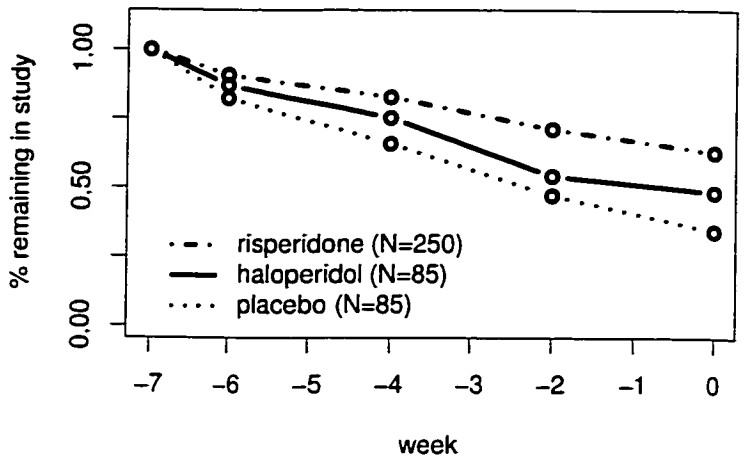


Figure 5.1: Dropout profiles for three treatment groups (with baseline sample size) in schizophrenia trial.

The treatment groups differ in both improvement and dropout rates. By modeling dropout, we can investigate the influence of dropout patterns on parameters relating treatment group to response. Logistic regression (ignoring dependence among subjects' measurements) suggests a selection model dependent on subjects' prior response values. We also explore the relationship between dropout and treatment group.

The marginalized transition model (MTM) with dropout and IPCW-GEE both require specification of three components: a marginal regression model, an association model, and a dropout model. Table 5.1 summarizes the models used for analysis of the schizophrenia data. The marginal regression model reflects the primary research question: how do patients respond to risperidone, placebo, and haloperidol in the treatment of schizophrenia? Level of improvement and both linear and quadratic time trends are allowed to vary by treatment group. Covariate `week` takes values (-7, -6, -4, -2, 0) so that variables `PLAC` and `RISP` are interpretable as comparisons to haloperidol after 8 weeks of treatment. Thus, the model allows inference on the 8-week response rates.

Serial dependence for MTMs is in terms of first- and second-order lagged responses.

Table 5.1: Summary of models considered for schizophrenia data.

Marginal Mean (all models)

$$\text{logit}\left[P(\mathrm{Y}_{it}=1|\mathrm{X}_{it})\right]=\beta_0+\beta_1\cdot\mathtt{PLAC}_i+\beta_2\cdot\mathtt{RISP}_i+\beta_3\cdot\mathtt{week}_{it}+\beta_4\cdot\mathtt{week}_{it}^2$$
$$+\beta_5\cdot\mathtt{PLAC}_i\cdot\mathtt{week}_{it}+\beta_6\cdot\mathtt{PLAC}_i\cdot\mathtt{week}_{it}^2+\beta_7\cdot\mathtt{RISP}_i\cdot\mathtt{week}_{it}+\beta_8\cdot\mathtt{RISP}_i\cdot\mathtt{week}_{it}^2$$

Serial Dependence

Marginalized Transition Models:

$$\text{logit}\left[P(\mathrm{Y}_{it}=1|\mathrm{X}_{it},\mathrm{Y}_{it-1},\mathrm{Y}_{it-2})\right]=$$

$t=2$	$\Delta_{i2}^{(2)}+\alpha_0\cdot Y_{it-1}$
$t>2$	$\Delta_{it}+\alpha_1\cdot Y_{it-1}+\alpha_2\cdot Y_{it-2}$

IPCW-GEE: Autoregressive working correlation

Dropout

$$\text{logit}\left[P(\mathrm{R}_{it}=1|\mathrm{R}_{it-1}=1,\mathrm{Y}_{it},\mathrm{H}_{it}^{(\mathrm{Y})},\mathrm{X}_{it})\right]=$$

$t=2$	ϕ_0
$t>2$	
(IPW1)	$\phi_1+\phi_2\cdot\mathtt{PLAC}_i+\phi_3\cdot\mathtt{RISP}_i+\phi_4\cdot y_{it-1}+\phi_5\cdot y_{it-2}$
(IPW2)	$\phi_1+\phi_2\cdot\mathtt{PLAC}_i+\phi_3\cdot\mathtt{RISP}_i+\phi_4\cdot y_{it-1}+\phi_5\cdot y_{it-2}$ $+\phi_6\cdot y_{it-1}\cdot\mathtt{PLAC}_i+\phi_7\cdot y_{it-1}\cdot\mathtt{RISP}_i$
(NI1: $\lambda_0=1.5$)	$\phi_1+\phi_2\cdot\mathtt{PLAC}_i+\phi_3\cdot\mathtt{RISP}_i+\phi_4\cdot y_{it-1}+\lambda_0\cdot y_{it}$
(NI2: $\lambda_0=-1.5$)	
(NI3: $\lambda=(1.5,0,1)$)	$\phi_1+\phi_2\cdot\mathtt{PLAC}_i+\phi_3\cdot\mathtt{RISP}_i+\phi_4\cdot y_{it-1}$ $+\lambda_0\cdot y_{it}+\lambda_1\cdot y_{it}\cdot\mathtt{PLAC}_i+\lambda_2\cdot y_{it}\cdot\mathtt{RISP}_i$
(NI4: $\lambda=(0.75,0,-1)$)	

For the first-order MTM, all observations have the conditional mean specification listed in Table 5.1 as "$t=2$". As described in Chapter 3, second-order marginalized transition model intercept Δ_{it} is implicitly defined and does not need to be fitted separately. The IPCW-GEE working correlation is chosen as autoregressive, based on exploratory analysis.

The interval between measurements is not constant, so we use visit number rather than week for establishing transitions and characterizing serial correlation.

Finally, the selection model for dropout is similar to that used by Diggle & Kenward (1994) and Diggle (1998): a logit-linear model. For both missing at random (MAR) and nonignorable (NI) conditions, dropout is modeled through group effects, and lagged and/or current responses. Observed data suggest that subjects who have shown improvement are more likely to stay in the study. Also, dropout is preceded by a drop in the percentage showing improvement (Figure 1.2). We therefore expect both "level" and "increment" effects (as discussed in Chapter 4) in retention. Fitting two lagged terms (ϕ_4 and ϕ_5 in Table 5.1) allows for separate estimation of level and increment. Group effects are not statistically significant in the selection models, but are retained since overspecification can result in modest efficiency gains, while underspecification may lead to severe bias (Robins et al., 1995). Two IPCW-GEE censoring weight models are computed, with and without interactions between treatment group and the first-order lagged response. MAR selection model parameters are orthogonal to regression and serial association parameters specified for the MTM, a likelihood-based model.

Four MTMs with NI dropout are fitted, to address sensitivity of MTM parameter estimates to hypothesized missing data mechanisms. For the first two NI models, the parameter λ_0 in Table 5.1 is set to 1.5 and -1.5 respectively. Holding other dropout parameters constant, when $\lambda_0 = 1.5$ the odds of remaining in the study when clinical improvement has occurred are $\exp(1.5) \approx 4.5$ times the odds if clinical improvement has not occurred. For $\lambda_0 = -1.5$ the effect is reversed. The final two NI selection models explore a hypothetical interaction between treatment group and response. In Model NI3, $(\lambda_0, \lambda_1, \lambda_2) = (0.75, 0, 1.0)$: the odds ratio for retention comparing $Y_{it} = 1$ to $Y_{it} = 0$ is about 2.1 for haloperidol and placebo, and about 5.8 for the risperidone group. This could occur if risperidone had additional desirable effects in addition to easing schizophrenia symptoms as measured by the PANSS. In model NI4, $(\lambda_0, \lambda_1, \lambda_2) = (1.5, 0, -1.0)$: people showing improvement are still always more likely to be retained, but the odds ratio for risperidone (1.6) is lower than for the other two groups (4.5). There is no evidence in the data for these models, but they represent alternate scenarios to explore for the impact of missingness model assumptions

on estimates of regression parameters.

A final methodical sensitivity analysis for nonignorable dropout models holds λ_0 constant in a series of values between -1.5 and 2.0 (with $\lambda_1 = \lambda_2 = 0$), and examines trends in all fitted parameters across the range of λ_0 values.

5.2 Results

Figure 5.2 shows marginal regression parameter estimates (converted to odds ratios) and one standard error in either direction, for the MAR models proposed in Table 5.1. A vertical reference line is given for no effect (an odds ratio of 1.0). Model IEE is fitted as a generalized linear model, estimated by GEE (Liang & Zeger, 1986) with independence working correlation. The two IPW models are fitted by GEE with autoregressive working correlation, weighted by inverse probability of censoring (Robins et al., 1995). Censoring weights for IPW1 are modeled by first- and second-order lagged responses and treatment group. An interaction between Y_{it-1} and treatment group is included in model IPW2. Model MTM(1) is a first-order and Model MTM(2) a second-order marginalized transition model: both assume MAR data, and have no explicitly specified dropout model. A likelihood ratio test indicates that the second-order terms α_0 and α_2 improve fit compared to a first-order model ($\chi^2_2 = 19.2, p < .001$). An approximate score test ($\chi^2_1 = 0.0, p = 0.994$), as described in Chapter 3, indicates that a third-order lagged term would not be statistically significant. Numerical details for these fitted models are shown in Table D.1, except for IPW1, which Figure 5.2 shows is similar to IPW2. Fitted trajectories for the three treatment groups are shown in Figure 5.3.

Parameter estimates do not differ greatly among the various MAR models (Figure 5.2). Fitted trajectories for the MTM and IPCW-GEE models are also very similar (Figure 5.3). Models IPW2 and MTM(2) both estimate the probability of clinical improvement after 8 weeks of treatment to be 33% for patients taking haloperidol, about 30% for placebo, and 46% for patients taking risperidone. The IEE model predicted probabilities are about 5 percentage points higher for haloperidol and risperidone groups, and 9 percentage points higher for the placebo group. The odds ratio for risperidone compared to haloperidol is

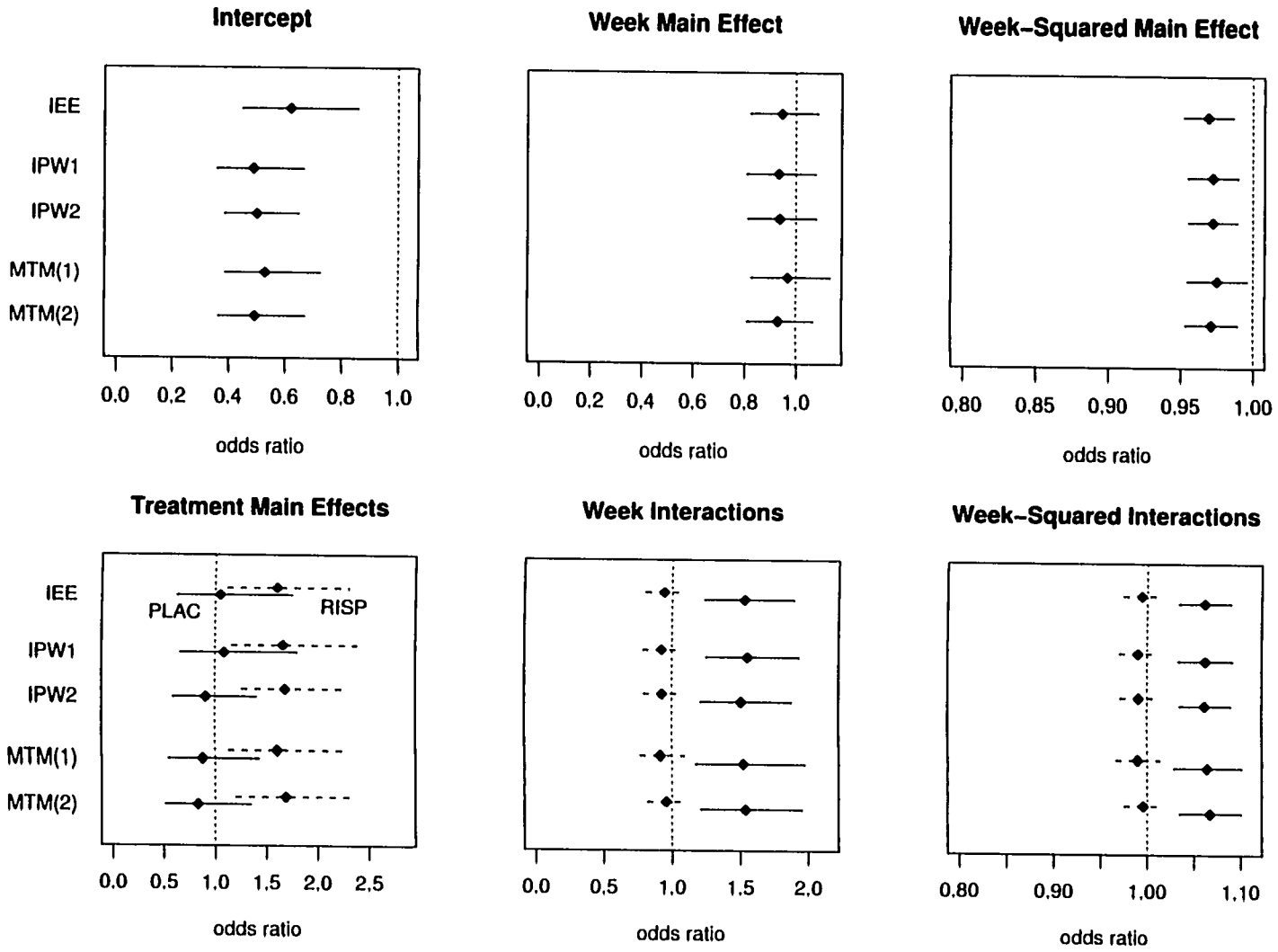


Figure 5.2: Fitted regression parameters ± one sandwich standard error (IEE, IPW) or model-based standard error (MTM) for MAR models for schizophrenia data (N=420). Model descriptions are given in Table 5.1.

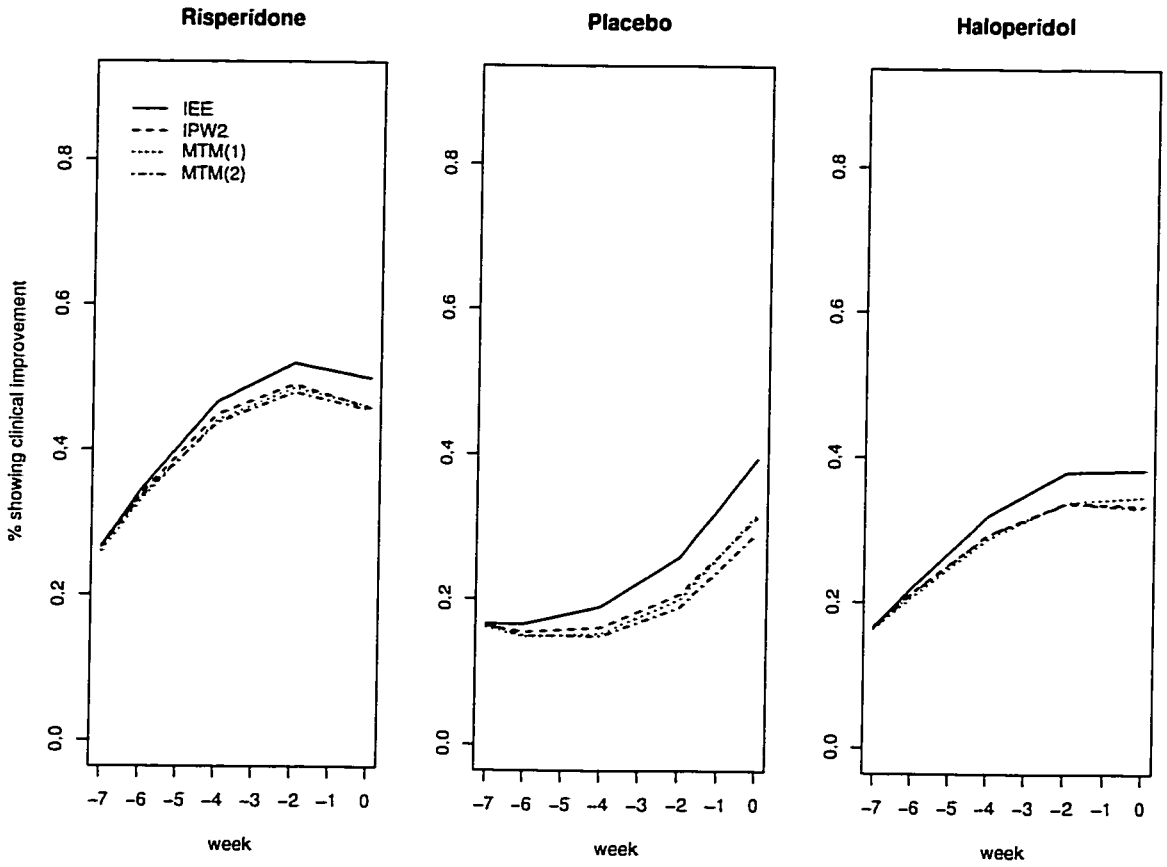


Figure 5.3: Fitted trajectories for schizophrenia data (N=420): Models assuming that data are missing at random (MAR).

estimated as 1.69 by IPW2 and MTM(2), and 1.60 by the IEE model. Risperidone is also estimated to be superior to placebo at 8 weeks, with odds ratios of 2.03, 1.85, and 1.53 estimated for MTM(2), IPW2, and IEE respectively. However, treatment differences are not statistically significant (for .05 Type I error) in any of the fitted models.

Time trends for risperidone and haloperidol have similar shapes, while the placebo group response probability does not climb as steeply in the early weeks of treatment (Figure 5.3). Probability of improvement is overestimated for IEE for timepoints after baseline, since patients who do not show improvement are more likely to leave the study. By contrast, likelihood-based MTMs implicitly impute the time trend for these cases (Laird, 1988), and

IPCW-GEE weights observations based on selection probabilities. However, overestimation of IEE response probabilities appears to affect the treatment groups equally: treatment group differences at time $= 0$ are similar for IEE and the other MAR models. It is notable that the placebo group shows improvement, on average, over the course of the trial. Perhaps this reflects the natural history of schizophrenia episodes.

5.3 Sensitivity Analysis

As background for a sensitivity analysis with nonignorable (NI) dropout models for the second-order MTM, we consider the MAR dropout model used for Model IPW1. Parameter estimates and model-based standard errors are shown in the first column of Table D.2. The dropout process depends on the previous two response values and treatment group. Patients who show clinical improvement in two consecutive measurements have odds of staying in the study for the next measurement $\exp(1.609 - 0.706) \approx 2.5$ times the odds for patients in the same treatment group who had not shown clinical improvement. This "level" effect suggests that people whose symptoms improve compared to baseline are more likely to stay in the study. The "increment" effect is measured as odds of retention at time t for a subject who does not show improvement at time $t-2$ but does at time $t-1$: the odds of retention are $\exp(1.609) \approx 5$ times higher than the odds of retention for someone who does not show improvement in PANSS at either time. People may be more likely to stay as the treatment seems to start working.

Fitted values for MTMs with the NI selection described in Section 5.1 are shown in Table D.2. Dropout model parameters involving Y_{it} are held constant, and other selection model parameters are estimated by maximum likelihood. We suspect that the fitted value $\hat{\phi}_4$ might change relative to λ_0 to maintain the level and increment interpretations for the MAR model. Table 5.2, which shows level and increment values based on values of λ_0 and $\hat{\phi}_4$, shows that this is not the case. Model NI1 has moderate level and increment effects. Although the fixed λ_0 is close to the fitted $\hat{\phi}_4$ in the MAR model, $\hat{\phi}_4$ in NI1 is not close to $\hat{\phi}_5$ in MTM(MAR). Model NI2 has a positive level but a negative increment effect. (In NI2, retention is most likely if the subject shows clinically significant improvement at time $t-1$

but not at time t.) In NI3 and NI4, level and increment effects are both positive (except for the model NI4 risperidone group), but level effects are larger, unlike in the MAR model.

Table 5.2: Level and Increment log-odds ratios for sensitivity analysis selection models for schizophrenia data (N=420).

Model	Level	Increment
MAR*	0.45	1.16
NI1	0.99	0.51
NI2	0.29	-1.79
NI3 (placebo,haloperidol)	0.65	0.10
NI3 (risperidone)	1.15	0.60
NI4 (placebo,haloperidol)	1.16	0.34
NI4 (risperidone)	0.66	-0.16

*MAR effects computed using $\hat{\phi}_4$ and $\hat{\phi}_5$ (Table 5.1).

Level: $(\lambda_0 + \lambda_2 \cdot \texttt{RISP} + \hat{\phi}_4)/2$

Increment: $(\lambda_0 + \lambda_2 \cdot \texttt{RISP} - \hat{\phi}_4)/2$

Of course the primary interest in sensitivity analysis is not how the dropout parameters ϕ are affected, but in the resulting marginal regression parameter estimates. Figure 5.4 shows fitted values $\hat{\beta}$ and one standard error in each direction, converted to odds ratios. Vertical reference lines are shown for no effect (odds ratio of 1). For interactions, haloperidol is the reference group, solid lines represent placebo terms, and dashed lines represent the risperidone group. Of the four NI models, NI2 parameters differ the most in comparison to the MAR (top) model. Predicted improvement probabilities are higher overall for Model NI2, due to implicit imputation of healthier subjects. Fitted values for the sensitivity analysis models are shown in Figure 5.5. Model NI2 trajectories resemble those for IEE in Figure 5.3. The remaining NI models show somewhat lower predicted probability of improvement for all treatment groups at all time points, compared to the MAR model. An

induced interaction between current response Y_{it} and the treatment group in the selection model (NI3 and NI4) does *not* greatly alter the treatment group effects in the regression model. Parameter estimates for models NI3 and NI4 do differ slightly from MTM(2) (Figure 5.4), but the fitted values over the range of the data are almost identical (Figure 5.5). NI1 and NI4 fitted values are extremely close to each other, and somewhat lower than for MTM(2). Fitting different NI models does affect inference: NI4 is the only model to find a statistically significant difference between risperidone and haloperidol after 8 weeks of treatment (odds ratio 2.03, Z=2.09, p=0.02).

The final sensitivity analysis shows changes in all parameters over a range of values for λ_0, where $\lambda_1 = \lambda_2 = 0$. Figure 5.6 shows parameter estimates (converted to odds ratios) plotted against values for λ_0. Odds ratios are inverted if parameter values are less than zero for most or all values of λ_0. For no parameter are fitted values symmetric about the MAR value ($\lambda_0 = 0$): deviations from MAR depend on the value of λ_0, not only distance from MAR. Risperidone and placebo estimated contrasts with haloperidol response at `time` $= 0$ are somewhat larger for positive values of λ_0. Serial dependence is smaller in models with $\lambda_0 < 0$, compared to $\lambda_0 > 0$. The difference between placebo and haloperidol linear time trends is greater for $\lambda_0 < 0$. The first-order lagged response term in the selection model (ϕ_4) is very large when $\lambda_0 < 0$ and approaches 1 for large positive λ_0.

5.4 Summary

The fitted marginalized transition models appropriately addresses the primary research question: there is a nonsignificant trend for risperidone to be superior to both placebo and haloperidol in the treatment of chronic schizophrenia. MTM and IPCW-GEE models under different assumptions about missingness provide information about the regression, association, and missingness structures of the PANSS data:

- Regression

  According to the second-order marginalized transition model under MAR missingness (Table D.1), the odds of showing clinical improvement for patients taking risperidone

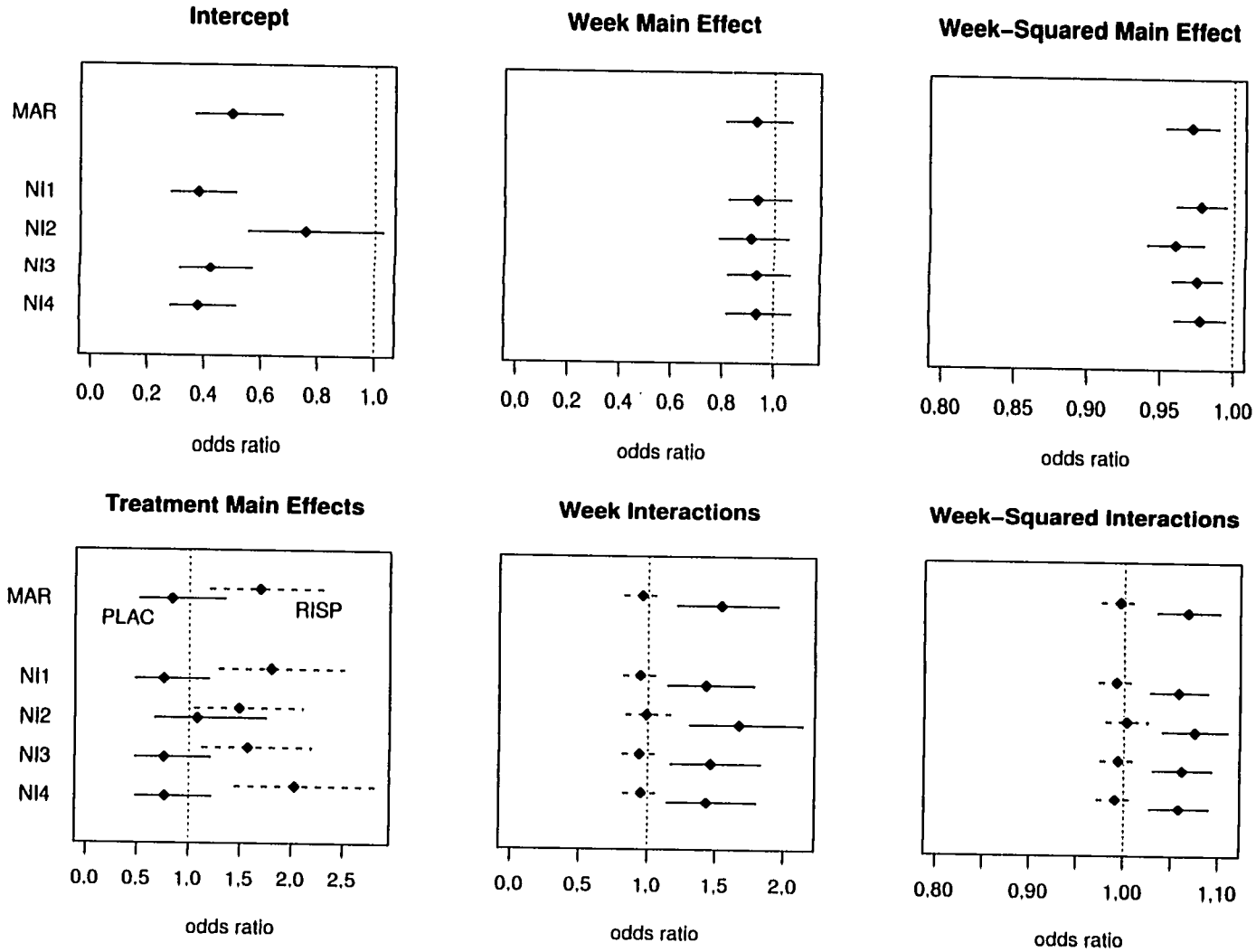


Figure 5.4: Fitted regression parameters ± one model-based standard error: PANSS sensitivity analysis (N=420): Second-order marginalized transition model (MAR) and 4 MTMs with nonignorable missing data, described in Table 5.1.

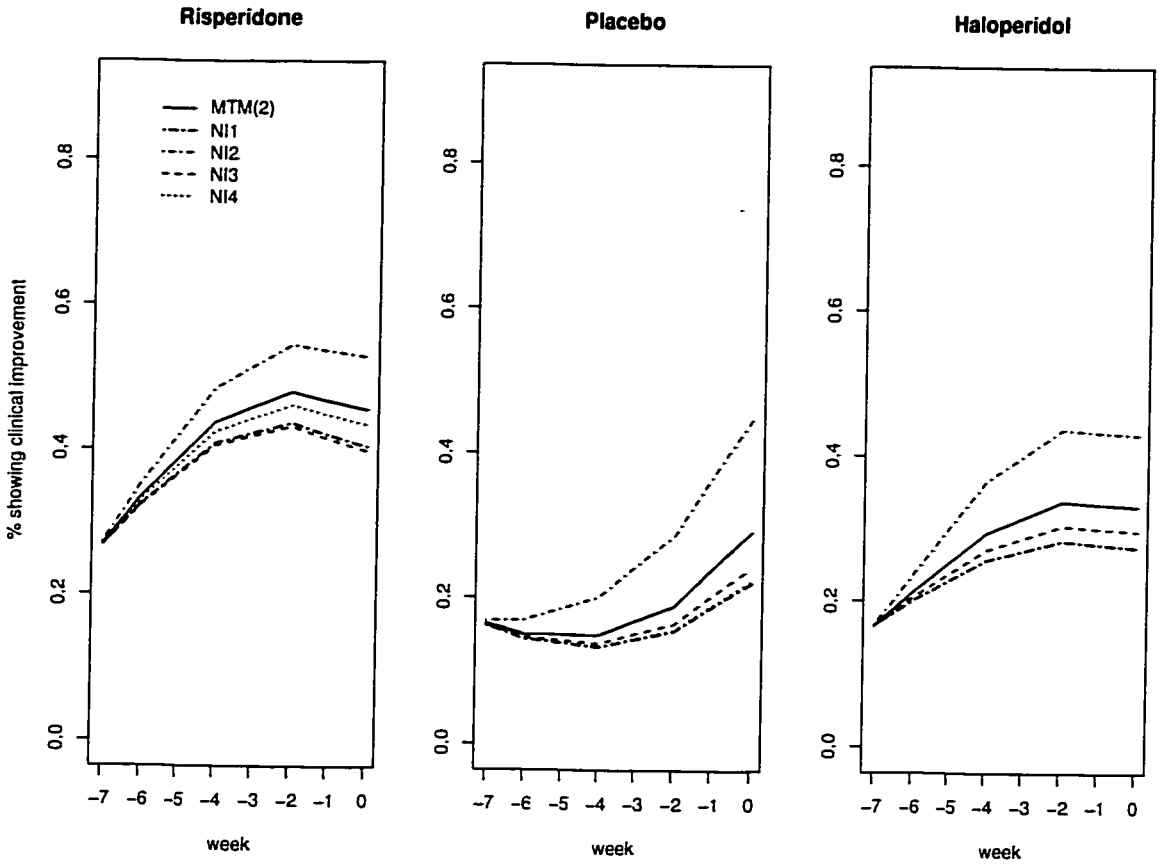


Figure 5.5: Fitted trajectories for schizophrenia data (N=420): Sensitivity analysis for MTMs with data assumed to be missing at random (MAR) or nonignorable (NI).

are $\exp(.524+.184) \approx 2.0$ times as great as for those taking placebo, and $\exp(.524) \approx 1.7$ times as great as for the haloperidol group at week 8. If only week 8 values are compared to baseline, with the last observation carried forward for patients who drop out, an odds ratio of 3.9 (Z=4.05, $p < .001$) is fitted comparing risperidone to placebo, and 1.9 (Z=2.33, $p < .01$) for risperidone compared to haloperidol.

- Association

The MTM(2) fitted association model suggests that the odds of improvement given improvement at the previous timepoint are over 10 times the odds of improvement

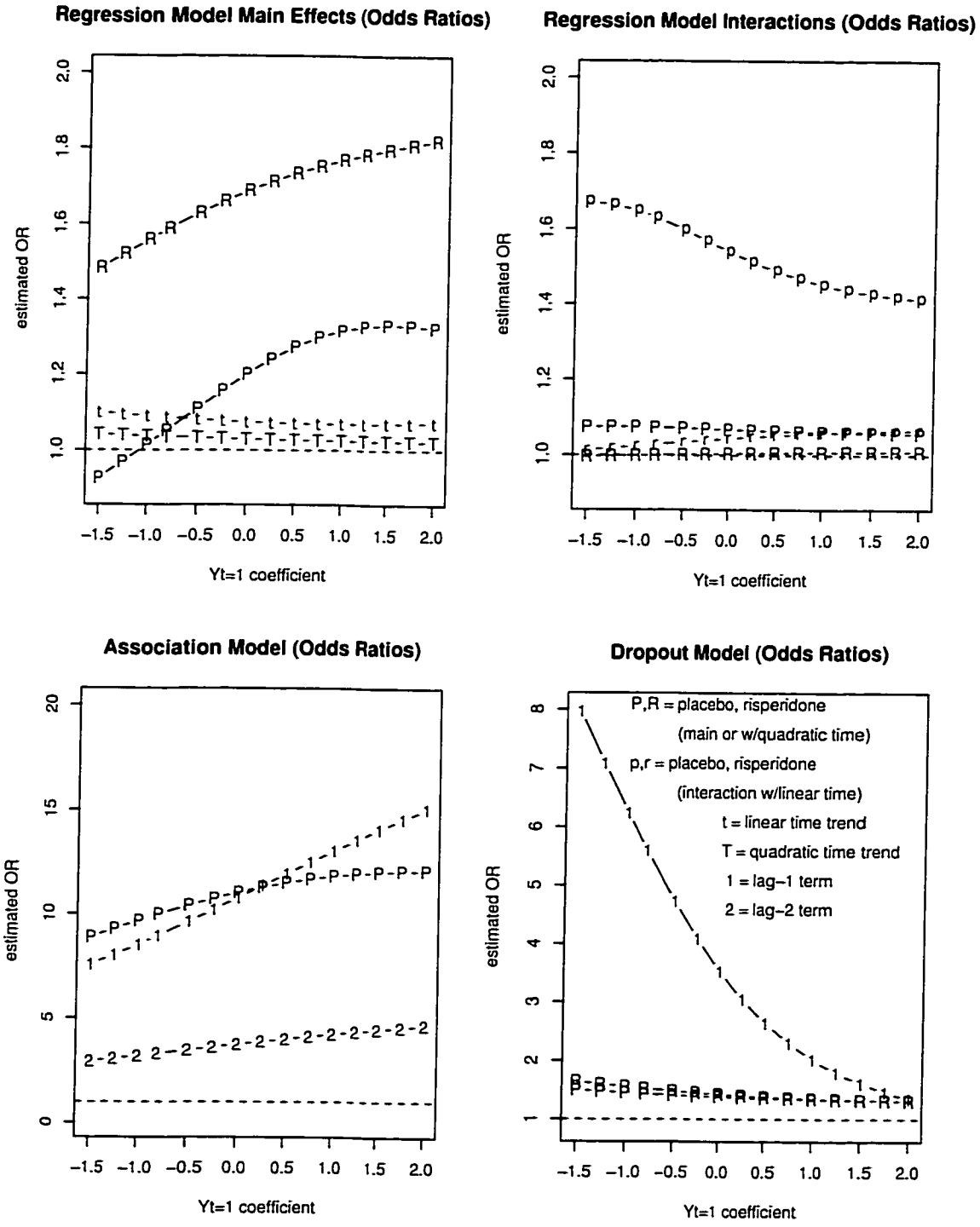


Figure 5.6: Influence of selection model current response parameter (λ_0) on fitted parameter values of regression, association, and selection model for schizophrenia data.

having not shown improvement at the previous assessment. Additionally, the odds of improvement are 3.75 times as great if the pre-previous response shows improvement compared to baseline.

- Missingness

  The dropout model shows that patients who have recently shown improvement tend to remain in the study.

In summary, the MTM accounting for dropout is a rich model that finds maximum likelihood estimates for important, interpretable parameters. For the PANSS data, in which dropout appears to vary by both treatment group and response values, treatment effects are attenuated compared to semiparametric models that do not accommodate MAR dropout. Sensitivity analysis with four hypothesized models for nonignorable dropout changes inference for regression model parameters. Fitted values for MTM models with nonignorable dropout vary somewhat more than for MTM and IPCW-GEE models assuming data are MAR.

Chapter 6

LONGITUDINAL BINARY DATA WITH DROPOUT AND DEATH: EXISTING METHODS

The PEP study, introduced in Chapter 1, is an example of longitudinal binary data with both dropout and death. The target of inference is the probability of disability in a specific month, among subjects that remain alive during that month. The analysis method should implicitly or explicitly impute values for responses that are missing due to dropout, but should not impute or otherwise account for responses missing due to death. This chapter examines the compatibility of such an analysis for likelihood-based methods and for inverse probability of censoring weighted GEE (IPCW-GEE, Robins et al., 1995). We consider the type of dropout (MCAR, MAR, or survival-dependent missingness), ignorability of the survival distribution, and direct estimation of parameters of interest.

6.1 Notation and Terminology

Notation for this chapter, summarized in Table 6.1, is similar to that for Chapter 2, with additional terms added for survival status. Note that for the PEP data, Y_{it} is described as disability status t months after baseline. Table 6.1 shows a more general representation, Y_{ij} at time t_j. Time is still measured as discrete, but does not have to be equally spaced values such as 1, 2, 3, etc.

Extension of missing data methods to data with both dropout and death also requires expanding definitions of MCAR, MAR, and NI missingness, discussed in Chapter 2 for data with dropout but no deaths. The MCAR/MAR/NI distinction is useful for dropout models because MCAR dropout is ignorable for most analysis methods, MAR dropout is ignorable for likelihood-based methods but not GEE, and analysis under NI dropout requires sensitivity analysis or dependence on weakly identified parameters (Laird, 1988). When

Table 6.1: Notation for analysis of binary longitudinal data with dropout and death.

n_i	scalar	number of responses for subject i until death or the planned end of the study
R_i	(n_i x 1)	missingness indicators (1=observed) (Monotone dropout: t_d is first timepoint in which R=0)
S_i	scalar	survival time (When $S_i = s_i$, t_s is the first timepoint in which subject i is deceased)
A_i	(n_i x 1)	survival indicators (1=alive) (Discrete-time vector version of S_i)
C_i	(n_i x 1)	censoring of survival information indicators (1=*not* censored) (introduced in Section 7.6)
$\mathrm{Y}_i^{\mathrm{o}}$	(m_i x 1)	observed responses
$\mathrm{Y}_i^{\mathrm{m}}$	($n_i - m_i$ x 1)	responses unobserved due to dropout
$\mathrm{H}_{ij}^{(\mathrm{Y})}$	($j-1$ x 1)	"history" vector for response
$\mathrm{H}_{ij}^{(\mathrm{R})}$	($j-1$ x 1)	"history" vector for missingness indicator
β	(p x 1)	regression parameters
α	(q x 1)	association parameters
ϕ	(r x 1)	dropout parameters
η	(l x 1)	survival parameters
N	scalar	number of clusters (subjects)
t_j	scalar	time at which Y_{ij} observed (or not observed)

the dropout process is *ignorable*, consistent estimates for β (regression parameters for the response Y_i) can be computed without explicit modeling of the dropout process, $f(\mathrm{R}_i|\phi)$. Table 6.2 describes a discrete-time selection model taxonomy that incorporates dependence on survival times. For missing completely at random (MCAR) data, missingness vector R_i does not depend on either responses Y_i or the specific survival time S_i. Each R_{ij} depends on S_i in that if subject i is deceased at time t_j, no Y_{ij} will be observed. However, given that

$S_i > t_j$, R_{ij} does not depend on the value of S_i for MCAR data. Under survival-dependent MCAR (MCAR-S), R_{ij} depends on S_i but not any element of Y_i.

Table 6.2: Taxonomy for missingness due to dropout and death, with comparison to dropout-only taxonomy.

Dropout		**Dropout and Death**	
$P(\mathrm{R}_{ij} = 1 \mid \mathrm{Y}_i^{\mathrm{o}}, \mathrm{Y}_i^{\mathrm{m}})$		$P(\mathrm{R}_{ij} = 1 \mid \mathrm{Y}_i^{\mathrm{o}}, \mathrm{Y}_i^{\mathrm{m}}, S_i = s_i, S_i > t_j)$	
$P(\mathrm{R}_{ij} = 1)$	**MCAR**	$P(\mathrm{R}_{ij} = 1 \mid S_i > t_j)$	**MCAR**
		$P(\mathrm{R}_{ij} = 1 \mid S_i = s_i, S_i > t_j)$	**MCAR-S**
$P(\mathrm{R}_{ij} = 1 \mid \mathrm{Y}_i^{\mathrm{o}})$	**MAR**	$P(\mathrm{R}_{ij} = 1 \mid \mathrm{Y}_i^{\mathrm{o}}, S_i > t_j)$	**MAR**
		$P(\mathrm{R}_{ij} = 1 \mid \mathrm{Y}_i^{\mathrm{o}}, S_i = s_i, S_i > t_j)$	**MAR-S**
$P(\mathrm{R}_{ij} = 1 \mid \mathrm{Y}_i^{\mathrm{o}}, \mathrm{Y}_i^{\mathrm{m}})$	**NI**	$P(\mathrm{R}_{ij} = 1 \mid \mathrm{Y}_i^{\mathrm{o}}, \mathrm{Y}_i^{\mathrm{m}}, S_i > t_j)$	**NI**
or $P(\mathrm{R}_{ij} = 1 \mid \mathrm{Y}_i^{\mathrm{m}})$		or $P(\mathrm{R}_{ij} = 1 \mid \mathrm{Y}_i^{\mathrm{m}}, S_i > t_j)$	
		$P(\mathrm{R}_{ij} = 1 \mid \mathrm{Y}_i^{\mathrm{o}}, \mathrm{Y}_i^{\mathrm{m}}, S_i = s_i, S_i > t_j)$	**NI-S**
		or $P(\mathrm{R}_{ij} = 1 \mid \mathrm{Y}_i^{\mathrm{m}}, S_i = s_i, S_i > t_j)$	

Data missing at random (MAR) and nonignorable missingness (NI) are defined as for data with dropout, given that the subject is living. Survival-dependent MAR (MAR-S) and survival-dependent NI (NI-S) data occur when the missingness process R_i depends on S_i as well as $\mathrm{Y}_i^{\mathrm{o}}$ (MAR-S) or $\mathrm{Y}_i^{\mathrm{m}}$ (NI-S). We concentrate here on methods for MCAR, MCAR-S, MAR, and MAR-S data, and leave discussion of NI dropout for future work.

The remainder of this chapter shows how ignorability of the survival process differs from ignorability of dropout under the dropout taxonomy in Table 6.2. Again a distribution P is ignorable with respect to response distribution $f(\mathrm{Y}_i|\beta)$ if modeling of P is not required to find consistent estimates for β. We also distinguish between direct and indirect methods for examining the target of inference. For *direct* methods, $\mu_{ij}^A = \mathrm{E}(\mathrm{Y}_{ij}|\mathrm{X}_{ij}, S_i > t_j)$ is parameterized by β. For *indirect* methods, target of inference μ_{ij}^A must be recovered from the model $\mu_{ij}^S = \mathrm{E}(\mathrm{Y}_{ij}|\mathrm{X}_{ij}, S_i = s_i)$, parameterized by β^s. The focus of this portion of the dissertation is Direct Estimation Conditioning on being ALive (DECAL).

6.2 *Likelihood-Based Analysis*

Before considering likelihood-based analysis of binary longitudinal data with both dropout and death, recall that in the presence of dropout (without death), the first step in a likelihood-based analysis can be to factorize the joint distribution function $f(\mathrm{Y}_i, \mathrm{R}_i)$. Chapter 2 focuses on selection models $f(\mathrm{Y}_i)f(\mathrm{R}_i|\mathrm{Y}_i)$ rather than the pattern-mixture factorization $f(\mathrm{R}_i)f(\mathrm{Y}_i|\mathrm{R}_i)$ because $\mathrm{E}(\mathrm{Y}_{ij})$, not $\mathrm{E}(\mathrm{Y}_{ij}|\mathrm{R}_i)$ or $\mathrm{E}(\mathrm{Y}_{ij}|\mathrm{R}_{ij} = 1)$, is the target of inference. The target of inference is again considered when factorizing the joint distribution $f(\mathrm{Y}_i, \mathrm{R}_i, S_i)$:

$$L(\beta^s, \alpha^s, \phi, \eta|\mathrm{Y}_i, \mathrm{R}_i, S_i) \propto f(S_i|\eta)f(\mathrm{Y}_i|S_i, \beta^s, \alpha^s)f(\mathrm{R}_i|\mathrm{Y}_i, S_i, \phi).$$

A full likelihood approach requires parameterization of the *entire* joint distribution, $f(\mathrm{Y}_i, \mathrm{R}_i, S_i|\theta)$. Accordingly, $f(\mathrm{Y}_i|S_i, \beta^s, \alpha^s)$ is the distribution of Y_i conditioning on the specific survival time, rather than partly conditioning on $S_i > t_j$. The remaining likelihood terms parallel the selection model for dropout and response.

If data are MCAR, MCAR-S, MAR, or MAR-S and (β^s, α^s), ϕ, and η are distinct, then dropout process $f(\mathrm{R}_i|\phi)$ is ignorable when considering the response process Y_i. The derivation is essentially the same as for ignorable dropout without death. For $\theta = (\beta^s, \alpha^s, \phi, \eta)$:

$$\begin{aligned} L(\theta|\mathrm{Y}_i^{\mathrm{o}}, \mathrm{R}_i, S_i) \;\propto\; & \int f(S_i|\eta)f(\mathrm{Y}_i^{\mathrm{m}}|\mathrm{Y}_i^{\mathrm{o}}, S_i, \beta^s, \alpha^s)f(\mathrm{Y}_i^{\mathrm{o}}|S_i, \beta^s, \alpha^s)f(\mathrm{R}_i|\mathrm{Y}_i^{\mathrm{m}}, \mathrm{Y}_i^{\mathrm{o}}, S_i, \phi)d\mathrm{Y}_i^{\mathrm{m}} \\ \overset{\text{MAR-S}}{=} \; & f(S_i|\eta)f(\mathrm{Y}_i^{\mathrm{o}}|S_i, \beta^s, \alpha^s)f(\mathrm{R}_i|\mathrm{Y}_i^{\mathrm{o}}, S_i, \phi) \int f(\mathrm{Y}_i^{\mathrm{m}}|\mathrm{Y}_i^{\mathrm{o}}, S_i, \beta^s, \alpha^s)d\mathrm{Y}_i^{\mathrm{m}} \\ = \; & f(S_i|\eta)f(\mathrm{Y}_i^{\mathrm{o}}|S_i, \beta^s, \alpha^s)f(\mathrm{R}_i|\mathrm{Y}_i^{\mathrm{o}}, S_i, \phi) \end{aligned}$$

and the three functions can be maximized separately. Dropout process $f(\mathrm{R}_i|\phi)$ is therefore ignorable under survival-dependent and non-survival-dependent MCAR and MAR missingness.

Unlike dropout, survival process $f(S_i|\eta)$ is not ignorable. The likelihood function specifies $\mu_{ij}^S = \mathrm{E}(\mathrm{Y}_{ij}|S_i = s_i)$. Target of inference $\mu_{ij}^A = \mathrm{E}(\mathrm{Y}_{ij}|S_i > t_j)$ is recovered from the pattern-mixture model $f(\mathrm{Y}_i^{\mathrm{o}}|S_i, \beta^s, \alpha^s)f(S_i|\eta)$, by summing over the survival distribution. In discrete time:

$$\mathrm{E}(\mathrm{Y}_{ij}|S_i > t_j) = \mu_{ij}^A = \frac{\sum_{k>j} \mathrm{E}(\mathrm{Y}_{ij}|S_i = t_k, \beta^s, \alpha^s)P(S_i = t_k|\eta)}{\sum_{k>j} P(S_i = t_k|\eta)}. \tag{6.1}$$

Section 1.2.2 describes how likelihood-based analyses implicitly impute data according to the fitted regression model. Equation 6.1 prevents this implicit imputation for likelihood-based DECAL analysis. The expected value μ_{ij}^{A} is computed by modeling survival time explicitly and summing expected values over the support restricted to when the subject is living. This summation yields unbiased estimates of $\mathrm{E}(\mathrm{Y}_{ij}|S_i > t_j)$, but only if $f(\mathrm{Y}_i|\beta^s, \alpha^s)$ and $f(S_i|\eta)$ are both specified correctly. The survival function $f(S_i|\eta)$ is therefore not ignorable for likelihood-based regression estimation of μ_{ij}^{A} – even if the dropout process is non-survival-dependent, and even if no data are missing due to dropout.

Equation 6.1 shows that, in addition to $f(S_i|\eta)$ being nonignorable, likelihood-based methods as described are not *direct*. The regression models $\mathrm{E}(\mathrm{Y}_{ij}|S_i = s_i)$ for survival time S_i (parameterized by β^s), not $\mathrm{E}(\mathrm{Y}_{ij}|S_i > t_j)$, modeled by β. Regression coefficients β cannot be recovered from the mixing equation, unless strong assumptions are made about modeling relationships among β and $(\beta^s, \alpha^s, \eta)$. This situation is similar to the marginalization of pattern-mixture models for dropout by Fitzmaurice & Laird (2000). A confidence interval can be constructed for μ_{ij}^{A} at each t_j using the delta method, but comparisons such as treatment main effects are not directly parameterized.

Misspecification bias in modeling $f(S_i|\eta)$ may be avoided by incorporating survival times via an empirical distribution. Brayne et al. (1999) model cognitive status, measured by the Mini-Mental State Examination (MMSE, Folstein et al., 1975), at four timepoints over nine years, and state the target of inference as "the mean MMSE ... had there been no attrition apart from death" (p. 1284). Random lines models are fitted to MMSE scores, using Markov Chain Monte Carlo to incorporate both MAR and NI dropout for a sensitivity analysis. Results are reported separately for each timepoint. Brayne et al. (1999) appear to avoid implicit imputation by truncating predicted values at time of death and reporting separate results at each timepoint. Each subject's S_i value determines where to stop the fitted response trajectory, so $f(S_i|\eta)$ is not modeled. However, $\mathrm{E}(\mathrm{Y}_{ij}|S_i > t_j)$ is still not directly parameterized, and empirical determination of $f(S_i|\eta)$ is impractical for models with continuous covariates.

In summary, a likelihood-based analysis can estimate $\mu_{ij}^{A} = \mathrm{E}(\mathrm{Y}_{ij}|S_i > t_j)$, but in the standard factorization (a pattern-mixture model for Y_i and S_i) the survival distribution is

not ignorable, and parameterization of μ_{ij}^{A} is not direct. Expected values must be recovered using a mixing distribution, and confidence intervals computed using the delta method.

6.3 *Generalized Estimating Equations (GEE)*

In this section we consider semiparametric methods of analysis, as an alternative to likelihood-based methods, for modeling $\mu_{ij}^{A} = \mathrm{E}(\mathrm{Y}_{ij}|S_i > t_j)$. To specify the full likelihood, dependence of response Y_{ij} on survival time S_i is modeled in terms of specific survival time. Using a moment-based approach, generalized estimating equations (GEE, Liang & Zeger, 1986), μ_{ij}^{A} may be modeled directly.

6.3.1 *Independence Estimating Equations (IEE)*

First consider the case where no data are missing due to dropout. For independence estimating equations (IEE), a logit-linear model with independence working correlation, (p x 1) covariate vector X_{ij} at time t_j and marginal mean $\mu_{ij}^{M} = \mathrm{E}(\mathrm{Y}_{ij}|\mathrm{X}_{ij}, \beta)$, an unbiased, linear quasi-score equation for β is:

$$U(\beta) = \sum_{i=1}^{N}\sum_{j=1}^{n_i} \mathrm{X}_{ij}(\mathrm{Y}_{ij} - \mu_{ij}^{M}).$$

To change the target of inference to $\mu_{ij}^{A} = \mathrm{E}(\mathrm{Y}_{ij}|S_i > t_j)$, we restrict quasi-score contributions to time periods when the subject is alive:

$$U(\beta) = \sum_{i=1}^{N}\sum_{j=1}^{n_i} \mathrm{X}_{ij}\mathrm{A}_{ij}(\mathrm{Y}_{ij} - \mu_{ij}^{A})$$

where indicator A_{ij} equals 1 when $S_i > t_j$ and 0 otherwise. Taking the expectation under $(\mathrm{Y}|S > t_j)$:

$$\begin{aligned} \mathrm{E}_{\mathrm{Y}|\mathrm{A}}\left[U(\beta)\right] &= \sum_{i=1}^{N}\sum_{j=1}^{n_i} \mathrm{X}_{ij}\mathrm{A}_{ij}\mathrm{E}_{\mathrm{Y}|\mathrm{A}}(\mathrm{Y}_{ij} - \mu_{ij}^{A}) \\ &= 0 \text{ if } \mu_{ij}^{A} = \mathrm{E}(\mathrm{Y}_{ij}|\mathrm{X}_{ij}, S_i > t_j). \end{aligned}$$

The only requirement for unbiased estimation of μ_{ij}^{A} is that the regression model is correctly specified. The DECAL (Direct Estimation Conditioning on being ALive) IEE approach does not require modeling of the survival distribution: $f(S_i|\eta)$ is ignorable.

6.3.2 *GEE with nonindependence working correlation*

The previous section illustrates direct estimation of β when no data are missing due to dropout, for GEE with independence working correlation. We next show that estimation is *not* direct and that survival distribution $f(S_i|\eta)$ is *not* ignorable when generalized estimating equations (GEE) with non-independence working correlation is applied.

For analysis with available cases, survival time determines the dimension of the working correlation matrix if death occurs before the planned end of followup. Under working independence (IEE), terms in the quasi-score sum over cluster values with variance weights that are a function of the mean but not the cluster size. When the variance matrix is non-diagonal, though, the variance inverse is different for different cluster sizes. This is apparent in IPCW-GEE without deaths: in Section 2.4.3, the vector R_i reflects that only observed values contribute to the quasi-score, but the cluster size is the length of Y_i, not observed response vector Y_i^o. Missing data due to dropout do not affect cluster size n_i, but when subjects die n_i will be truncated.

The quasi-score for GEE reflects dependence on survival time $S_i = s_i$. Let $w_{ijk}^{[n_i(s_i)]}$ be weights for the logistic-linear model corresponding to $\mathbf{D}_i^T \boldsymbol{V}_i(\hat{\alpha})^{-1}$ in Equation 2.4. The weights depend on X_i, μ_{ij}^S, and the variance inverse, which depends on μ_{ij}^S, α, and cluster size $n_i(s_i)$ (which depends on survival time s_i). For example, under exchangeable working correlation, estimated values of $\mathrm{corr}(\mathrm{Y}_{ij}, \mathrm{Y}_{ik})$ do not depend on value of $(j, k \in 1, \ldots, n_i)$, but values of the correlation matrix inverse depend on $n_i(s_i)$:

$$\begin{pmatrix} 1 & 0.4 & 0.4 & 0.4 \\ 0.4 & 1 & 0.4 & 0.4 \\ 0.4 & 0.4 & 1 & 0.4 \\ 0.4 & 0.4 & 0.4 & 1 \end{pmatrix}^{-1} \approx \begin{pmatrix} 1.36 & -0.30 & -0.30 & -0.30 \\ -0.30 & 1.36 & -0.30 & -0.30 \\ -0.30 & -0.30 & 1.36 & -0.30 \\ -0.30 & -0.30 & -0.30 & 1.36 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 0.4 & 0.4 \\ 0.4 & 1 & 0.4 \\ 0.4 & 0.4 & 1 \end{pmatrix}^{-1} \approx \begin{pmatrix} 1.30 & -0.37 & -0.37 \\ -0.37 & 1.30 & -0.37 \\ -0.37 & -0.37 & 1.30 \end{pmatrix}.$$

The quasi-score for β is:

$$U(\beta) = \sum_{i=1}^{N} \sum_{j=1}^{n_i(s_i)} \sum_{k=1}^{n_i(s_i)} w_{ijk}^{[n_i(s_i)]} I(S_i > t_j)(\mathrm{Y}_{ij} - \mu_{ij}^S).$$

Unlike the IEE quasi-score, it is necessary to sum over all n_i^2 pairwise combinations. This is due to the presence of the variance *inverse* in $w_{ijk}^{[n_i(s_i)]}$. Since $w_{ijk}^{[n_i(s_i)]}$ depends on the specific value of survival time S_i, the expectation of the score under $(\mathrm{Y}|S > t_j)$ will not result in unbiased estimation of μ_{ij}^A. We can take the expectation under $(\mathrm{Y}|S = s_i)$:

$$\begin{aligned} \mathrm{E}_{\mathrm{Y}|S}\left[U(\beta)\right] &= \sum_{i=1}^{N} \sum_{j=1}^{n_i(s_i)} \sum_{k=1}^{n_i(s_i)} w_{ijk}^{[n_i(s_i)]} \mathrm{A}_{ij} \mathrm{E}_{\mathrm{Y}|S}(\mathrm{Y}_{ij} - \mu_{ij}^S) \\ &= 0 \text{ if } \mu_{ij}^S = \mathrm{E}(\mathrm{Y}_{ij}|\mathrm{X}_{ij}, S_i = s_i) \end{aligned}$$

Unlike the IEE case above, $f(S_i|\eta)$ is not ignorable and parameterization of β is not direct. Once consistent estimates of (β^s, α^s) are found that parameterize $\mu_{ij}^S = \mathrm{E}(\mathrm{Y}_{ij}|\mathrm{X}_{ij}, S = s_i)$, the target of inference $\mathrm{E}(\mathrm{Y}_{ij}|\mathrm{X}_{ij}, S_i > t_j)$ can be estimated using a mixing distribution with $f(S_i|\eta)$, as for likelihood-based analyses described above. Ignorability of $f(\mathrm{R}_i|\phi)$ is not discussed for GEE because we focus on direct methods for estimating μ_{ij}^A, such as IEE.

6.4 Summary and Discussion

This chapter has explored direct parameterization for $\mu_{ij}^A = \mathrm{E}(\mathrm{Y}_{ij}|S_i > t_j)$ using existing methods for analyzing binary longitudinal data. Ignorability of dropout and survival processes are also of interest. Direct estimation is generally not feasible for likelihood-based methods or for GEE with non-independence working correlation. The next chapter describes in detail Direct Estimation Conditioning on being ALive (DECAL), using independence estimating equations (IEE).

Chapter 7

DIRECT ESTIMATION IN LONGITUDINAL BINARY DATA WITH DROPOUT AND DEATH

This chapter describes Direct Estimation Conditioning on being ALive (DECAL) for longitudinal binary data with both dropout and death. Under direct estimation, regression models parameterize $E(Y_{ij}|S_i > t_j)$. Independence estimating equations (IEE) models are established for estimation under MCAR, MCAR-S, MAR, and MAR-S dropout. We develop inverse probability of censoring weights when needed for consistent estimation of DECAL regression parameters. The impact of censored survival time on DECAL regression is considered.

7.1 IPC Weights under Dropout and Death

Section 1.2.2 shows that for data records truncated by death but not dropout, a DECAL model can be fitted using IEE. No weights or special programming are needed. The only requirement to distinguish death from dropout is to choose a functional form for the regression model so that $\mu_{ij}^{A} = E(Y_{ij}|S_i > t_j)$. This chapter adapts inverse probability of censoring weighted GEE (IPCW-GEE) to accommodate DECAL regression for data with both dropout and death.

For data with MAR dropout and *no deaths*, IPCW-GEE, described in Section 2.4.3, weights observed values Y_{ij} by the inverse of $P(R_{ij} = 1|H_{ij}^{(Y)})$, the probability of being observed. Response Y_{ij} is given more weight in estimating equations if responses with similar characteristics are likely to be missing due to dropout. Just as implicit imputation is a concern in likelihood-based methods (Section 1.2.1), we wish to avoid assigning inverse probability of censoring (IPC) weights based on potential responses of *deceased* subjects.

For data with both MAR dropout and deaths, IPC weights π_{ij} should be modeled so

Table 7.1: Data records to illustrate weighting schemes under dropout and death. ($\surd$ = observed, $\cdot$ = missing, $\times$ = deceased)

id	time 1	time 2	time 3	time 4	time 5
1	$\surd$	$\surd$	$\surd$	$\surd$	$\times$
2	$\surd$	$\cdot$	$\times$	$\times$	$\times$
3	$\surd$	$\surd$	$\cdot$	$\cdot$	$\times$
4	$\surd$	$\times$	$\times$	$\times$	$\times$
5	$\surd$	$\surd$	$\surd$	$\cdot$	$\times$

that

$$\mathrm{E}\left(\frac{\mathrm{R}_{ij}\mathrm{Y}_{ij}}{\pi_{ij}}|\mathrm{X}_{ij}, S_i > t_j\right) = \mathrm{E}(\mathrm{Y}_{ij}|\mathrm{X}_{ij}, S_i > t_j).$$

This section illustrates that the IPC weights must condition on survival status. Sections 7.2-7.5 explore the form of weights for specific dropout patterns. For survival-dependent dropout patterns (MCAR-S and MAR-S) finer conditioning is needed, since dropout at time t_j depends on $S_i = s_i$ rather than $S_i > t_j$.

Table 7.1 shows hypothetical missingness information for 5 subjects in a study with 5 timepoints. Data for subjects 1 and 4 are complete because all responses are observed ($\surd$) until the subjects die ($\times$). For subject 2, the response at timepoint 2 is missing due to dropout ($\cdot$), but the subject dies before timepoint 3. All five subjects are deceased by timepoint 5. Assuming that the subjects are similar with respect to treatment group and other covariates, IPC weights for the fourth response of subject 1 should represent all three living subjects (subjects 1,3, and 5).

Assume that R_j is a binary indicator of nonmissingness at time j, and that $P(\mathrm{R}_1 = 1) = 1$. Since dropout is monotone, $(\mathrm{R}_j = 1)$ implies that $(\mathrm{R}_k = 1)$ for $k < j$, and $P(\mathrm{R}_j = 1)$ can be represented as a telescoping product, with each element of R conditioning on past values. We now consider how to condition on the survival time so that the IPC weights are computed correctly. First, if dropout is modeled conditioning on being alive:

$$\begin{aligned} P(\mathrm{R}_4{=}1|S>4) &= P(\mathrm{R}_4=1|\mathrm{R}_3=1,S>4)\cdot P(\mathrm{R}_3=1|\mathrm{R}_2=1,S>4)\cdot P(\mathrm{R}_2=1|S>4) \\ &= \frac{1}{2}\cdot\frac{2}{3}\cdot\frac{3}{3} \\ &= \frac{1}{3} \text{ the correct weight.} \end{aligned}$$

Second, if the dropout model ignores survival, and assigns $\mathrm{R}_j = 0$ for subjects who are deceased at time j:

$$\begin{aligned} P(\mathrm{R}_4{=}1) &= P(\mathrm{R}_4=1|\mathrm{R}_3=1)\cdot P(\mathrm{R}_3=1|\mathrm{R}_2=1)\cdot P(\mathrm{R}_2=1) \\ &= \frac{1}{2}\cdot\frac{2}{3}\cdot\frac{3}{5} \\ &= \frac{1}{5} \text{ (incorrect).} \end{aligned}$$

Finally, if the conditioning on survival changes as the dropout model telescopes:

$$\begin{aligned} P(\mathrm{R}_4{=}1|S>4) &= P(\mathrm{R}_4=1|\mathrm{R}_3=1,S>4)\cdot P(\mathrm{R}_3=1|\mathrm{R}_2=1,S>3)\cdot P(\mathrm{R}_2=1|S>2) \\ &= \frac{1}{2}\cdot\frac{2}{3}\cdot\frac{3}{4} \\ &= \frac{1}{4} \text{ (incorrect).} \end{aligned}$$

The proper target for IPC weights for non-survival-dependent dropout in DECAL models is $P(\mathrm{R}_{ij}=1|S_i>t_j)$. Survival-dependent dropout models $P(\mathrm{R}_{ij}=1|S_i=s_i)$ for MCAR-S and MAR-S data are discussed below. Furthermore, under modeling assumptions for some dropout patterns, the second or third telescoping examples above may have the same expected value as the first (Section 7.3.1).

7.2 MCAR

Working independence estimation of β using available data (all observed binary responses) can be represented by the following quasi-score:

$$U(\beta) = \sum_{i=1}^{N}\sum_{j=1}^{n_i} \mathrm{X}_{ij}\mathrm{A}_{ij}\mathrm{R}_{ij}(\mathrm{Y}_{ij}-\mu_{ij}^{A}),$$

where indicator R_{ij} reflects missingness status ($R_{ij} = 0$ if Y_{ij} is missing), and A_{ij} indicates vital status ($A_{ij} = 1$ if subject i is alive at time t_j). Under MCAR, $P(R_{ij} = 1|Y_{ij}, H_{ij}^{(Y)}, S_i) = P(R_{ij} = 1|S_i > t_j)$, so

$$\begin{aligned} E_{Y,R|A}[U(\beta)] &= \sum_{i=1}^{N}\sum_{j=1}^{n_i} X_{ij}A_{ij}E_{Y|A}\left[(Y_{ij} - \mu_{ij}^{A})E_{R|Y,A}(R_{ij})\right] \\ &\overset{\text{MCAR}}{=} \sum X_{ij}A_{ij}E_{Y|A}\left(Y_{ij} - \mu_{ij}^{A}\right)P(R_{ij} = 1|S_i > t_j) \\ &= 0 \text{ if } \mu_{ij}^{A} = E(Y_{ij}|X_{ij}, S_i > t_j). \end{aligned}$$

As for data with no dropout, estimation of β using IEE is consistent under MCAR if the regression model is properly specified. Weights based on MCAR dropout may affect efficiency, but not consistency, of regression estimators β. Both $f(R_i|\phi)$ and $f(S_i|\eta)$ are ignorable, as long as regression parameter vector β is distinct from dropout and survival parameters ϕ and η.

7.3 MAR

Non-survival-dependent MAR is defined in Table 6.2 as:

$$P(R_{ij} = 1|Y_i^o, Y_i^m, S_i) = P(R_{ij} = 1|Y_i^o, S_i > t_j).$$

As long as subject i is alive at time t_j, R_{ij} does not depend on specific survival time. We assume that R_{ij} depends on observed responses occurring *before* time t_j, not in the future. Thus, the MAR assumption is more precisely stated as:

$$P(R_{ij} = 1|Y_i, S_i) = P(R_{ij} = 1|H_{ij}^{(Y)}, S_i > t_j). \tag{7.1}$$

This section describes DECAL models under MAR dropout, using inverse probability of censoring weighted independence estimating equations (IPCW-IEE).

For inverse probability of censoring weights $\pi_{ij} = P(R_{ij} = 1|H_{ij}^{(Y)}, S_i > t_j)$, the quasi-score with independence working correlation is:

$$U(\beta) = \sum_{i=1}^{N}\sum_{j=1}^{n_i} X_{ij}A_{ij}\frac{R_{ij}}{\pi_{ij}}(Y_{ij} - \mu_{ij}^{A}).$$

Taking the expectation over the observed data distribution:

$$\begin{aligned} \mathrm{E}_{\mathrm{Y,R|A}}\left[U(\beta)\right] &= \sum_{i=1}^{N}\sum_{j=1}^{n_i} \mathrm{X}_{ij}\mathrm{A}_{ij}\mathrm{E}_{\mathrm{Y|A}}\left[\mathrm{E}_{\mathrm{R|Y,A}}\left(\frac{\mathrm{R}_{ij}}{\pi_{ij}}\right)(\mathrm{Y}_{ij}-\mu_{ij}^{A})\right] \\ &\overset{\mathrm{MAR}}{=} \sum \mathrm{X}_{ij}\mathrm{A}_{ij}\mathrm{E}_{\mathrm{Y|A}}\left[\frac{\pi_{ij}}{\pi_{ij}}(\mathrm{Y}_{ij}-\mu_{ij}^{A})\right] \\ &= 0 \text{ if } \mu_{ij}^{A} = \mathrm{E}(\mathrm{Y}_{ij}|\mathrm{X}_{ij}, S_i > t_j). \end{aligned}$$

Dropout process $f(\mathrm{R}_i|\phi)$ is not ignorable, since dropout must be modeled correctly for R_i not to affect inference for Y_i. However, survival distribution $f(S_i|\eta)$ is ignorable under IEE for MAR data, as it was under MCAR and MCAR-S.

7.3.1 Modeling IPC Weights

We now consider details for estimation of the dropout model that produces IPC weights $\pi_{ij} = P(\mathrm{R}_{ij} = 1|\mathrm{Y}_i, S_i > t_j)$. Notation is reviewed in Table 6.1 on page 87. Again, the quasi-score for IPCW-IEE estimation of β under data missing at random is:

$$U(\beta) = \sum_{i=1}^{N}\sum_{j=1}^{n_i} \mathrm{X}_{ij}\mathrm{A}_{ij}\frac{\mathrm{R}_{ij}}{\pi_{ij}}(\mathrm{Y}_{ij}-\mu_{ij}^{A}).$$

Inverse probability of censoring (dropout) weights may be estimated by a logit-linear model. As for most work involving dropout, we assume that the first observation in a cluster is not missing (Diggle & Kenward, 1994; Robins et al., 1995). We first examine how dropout models are altered by the presence of both dropout and death, then simplify the dropout model under MAR assumptions. We parameterize dropout (ϕ) through logistic regression as follows. First consider $\pi_{ij} = P(\mathrm{R}_{ij} = 1|\mathrm{H}_{ij}^{(\mathrm{Y})}, S_i > t_j)$:

$$\begin{aligned} &P(\mathrm{R}_{ij} = 1|\mathrm{H}_{ij}^{(\mathrm{Y})}, S_i > t_j) \\ &= P(\mathrm{R}_{i1} = 1, \ldots, \mathrm{R}_{ij} = 1|\mathrm{H}_{ij}^{(\mathrm{Y})}, S_i > t_j) \\ &= P(\mathrm{R}_{i2} = 1|\mathrm{R}_{i1} = 1, \mathrm{Y}_{i1}, S_i > t_j) \cdots P(\mathrm{R}_{ij} = 1|\mathrm{H}_{ij}^{(\mathrm{R})} \equiv 1, \mathrm{H}_{ij}^{(\mathrm{Y})}, S_i > t_j) \\ &= \prod_{k=2}^{j} P(\mathrm{R}_{ik} = 1|\mathrm{H}_{ik}^{(\mathrm{R})} \equiv 1, \mathrm{H}_{ik}^{(\mathrm{Y})}, S_i > t_j). \end{aligned}$$

Let $u_{ik(j)} = P(\mathrm{R}_{ik} = 1|\mathrm{H}_{ik}^{(\mathrm{R})}, \mathrm{H}_{ik}^{(\mathrm{Y})}, S_i > t_j)$ where $2 \leq k \leq j$. When dropout time is t_d and

survival time t_s:

$$P(\mathrm{R}_{ik} = r_{ik} | \mathrm{H}_{ik}^{(\mathrm{R})} \equiv 1, \mathrm{H}_{ik}^{(\mathrm{Y})}, S_i > t_j) = u_{ik(j)}^{r_{ik}} (1 - u_{ik(j)})^{1 - r_{ik}}$$

and

$$u_{ik(j)} = \begin{cases} p_{ik(j)} & : \quad k \leq d \text{ and } k < s \\ 0 & : \quad k > d \text{ and } k < s \\ \text{undefined} & : \quad k \geq s \text{ (deceased)} \end{cases}$$

where $p_{ik(j)} = P(\mathrm{R}_{ik} = 1 | \mathrm{H}_{ik}^{(\mathrm{R})} \equiv 1, \mathrm{H}_{ik}^{(\mathrm{Y})}, S_i > t_j), = P(\mathrm{R}_{ik} = 1 | \mathrm{R}_{ik-1} = 1, \mathrm{H}_{ik}^{(\mathrm{Y})}, S_i > t_j)$ under monotone dropout. So

$$\begin{aligned} & P(\mathrm{R}_{ij} = r_{ij} | \mathrm{H}_{ij}^{(\mathrm{Y})}, S_i > t_j) \\ &= \prod_{k=2}^{(d-1) \wedge (s-1) \wedge j} p_{ik(j)}^{r_{ik}} \left[(1 - p_{id(j)})^{I(d \leq j < s)} \right] \prod_{k=(d+1)}^{j \wedge (s-1)} (1)^{(1 - r_{ik}) I((d+1) < s)} \\ &= \left(\prod_{k=2}^{(d-1) \wedge (s-1) \wedge j} p_{ik(j)} \right) (1 - p_{id(j)})^{I(d \leq j < s)}. \end{aligned}$$

This is a discrete-time survival function, modeling the hazard for the event $\mathrm{R}_{ij} = 0$ for times ($j = 1, \ldots, n_i$). Death is a censoring mechanism. Using a logistic regression model with model matrix $\mathbf{D}_i$ that can include elements of $\mathrm{H}_{ik}^{(\mathrm{Y})}$ and other covariates,

$$\mathrm{logit}(p_{ik(j)}) = \mathrm{D}_{ik} \phi_j.$$

7.3.2 Simplified Weights Under MAR

We now show that under non-survival-dependent MAR, $p_{ik(j)} = p_{ik}$. It follows that $\phi_j = \phi$. The dropout model for R_{ij} does not need to stratify by timepoint.

The MAR assumption in Equation 7.1 describes marginal, not conditional probabilities. For monotone missingness, dropout models are built through the telescoping property of the dropout vector:

$$\begin{aligned} P(\mathrm{R}_{ij} = 1) &= P(\mathrm{R}_{i1} = 1) \cdot P(\mathrm{R}_{i2} = 1 | \mathrm{R}_{i1} = 1) \cdots P(\mathrm{R}_{ij} = 1 | \mathrm{H}_{ij}^{(\mathrm{R})} \equiv 1) \\ &= P(\mathrm{R}_{i1} = 1) \prod_{k=2}^{j} P(\mathrm{R}_{ik} = 1 | \mathrm{H}_{ik}^{(\mathrm{R})} \equiv 1) \end{aligned}$$

$$= P(\mathrm{R}_{i1}=1)\prod_{k=2}^{j} P(\mathrm{R}_{ik}=1|\mathrm{R}_{ik-1}=1) \qquad \text{(since dropout is monotone).}$$

Appendix E shows that under the MAR assumption for monotone dropout, Equation 7.1, the conditional dropout probabilities used for telescoping do not depend on survival time, as long as the subject is alive:

$$P(\mathrm{R}_{ij}=1|\mathrm{R}_{ij-1}=1,\mathrm{Y}_{ij},\mathrm{H}_{ij}^{(\mathrm{Y})},S_i) \overset{\mathrm{MAR}}{=} P(\mathrm{R}_{ij}=1|\mathrm{R}_{ij-1}=1,\mathrm{H}_{ij}^{(\mathrm{Y})},S_i>t_j). \qquad (7.2)$$

We now show that $p_{ik(j)}=p_{ik}$. Consider $p_{ik(j)}$, where $j \geq k$:

$$\begin{aligned}
p_{ik(j)} &= P(\mathrm{R}_{ik}=1|\mathrm{R}_{ik-1}=1,\mathrm{H}_{ik}^{(\mathrm{Y})},S_i>t_j) \\
&= \frac{P(\mathrm{R}_{ik}=1,S_i>t_j|\mathrm{R}_{ik-1}=1,\mathrm{H}_{ik}^{(\mathrm{Y})})}{P(S_i>t_j|\mathrm{R}_{ik-1}=1,\mathrm{H}_{ik}^{(\mathrm{Y})})} \\
&= \frac{\sum_{r>j}P(\mathrm{R}_{ik}=1|\mathrm{R}_{ik-1}=1,\mathrm{H}_{ik}^{(\mathrm{Y})},S_i=t_r)P(S_i=t_r|\mathrm{R}_{ik-1}=1,\mathrm{H}_{ik}^{(\mathrm{Y})})}{\sum_{r>j}P(S_i=t_r|\mathrm{R}_{ik-1}=1,\mathrm{H}_{ik}^{(\mathrm{Y})})} \\
&\overset{\mathrm{MAR}}{=} \frac{P(\mathrm{R}_{ik}=1|\mathrm{R}_{ik-1}=1,\mathrm{H}_{ik}^{(\mathrm{Y})},S_i>t_k)\sum_{r>j}P(S_i=t_r|\mathrm{R}_{ik-1}=1,\mathrm{H}_{ik}^{(\mathrm{Y})})}{\sum_{r>j}P(S_i=t_r|\mathrm{R}_{ik-1}=1,\mathrm{H}_{ik}^{(\mathrm{Y})})} \\
&\qquad\qquad \text{(Equation 7.2. } r>j \text{ and } j\geq k \text{ so } t_k<t_r\text{)} \\
&= P(\mathrm{R}_{ik}=1|\mathrm{R}_{ik-1}=1,\mathrm{H}_{ik}^{(\mathrm{Y})},S_i>t_k) \\
&= p_{ik}.
\end{aligned}$$

The implication of this section is a simple adjustment of IPC weights for IPCW-IEE estimating equations when deaths occur in addition to dropout. Dropout parameters are not stratified by time ($\phi_j \equiv \phi$), but death and MAR dropout must be differentiated when modeling ϕ. The distribution for the dropout process used for IPC weights is thus:

$$P(\mathrm{R}_{ij}=r_{ij}|\mathrm{H}_{ij}^{(\mathrm{Y})},S_i>t_j) = \left(\prod_{k=2}^{(d-1)\wedge(s-1)\wedge j} p_{ik}\right)(1-p_{id})^{I(d\leq j<s)} \qquad (7.3)$$

with p_{ik} modeled using logistic regression:

$$\mathrm{logit}(p_{ik}) = \mathrm{D}_{ik}\phi.$$

In summary, comparing IPCW-GEE for data with dropout only and IPCW-IEE for MAR dropout and death, the only differences are:

1. Choose a functional form for the regression model for μ_{ij}^{A} (DECAL)

2. Use independence working correlation to avoid dependence on specific survival time

3. Use survival time as a censoring time for dropout

7.4 MCAR-S

For survival-dependent dropout completely at random (MCAR-S), dropout depends on the specific survival time, but not observed elements of Y_i:

$$P(\mathrm{R}_{ij} = 1|\mathrm{Y}_{ij}, \mathrm{H}_{ij}^{(\mathrm{Y})}, S_i) = P(\mathrm{R}_{ij} = 1|S_i = s_i, S_i > t_j).$$

For estimation using IEE, the quasi-score for regression parameter vector β is:

$$\begin{aligned} \mathrm{E}_{\mathrm{Y,R},S|\mathrm{A}}\left[U(\beta)\right] &= \sum_{i=1}^{N}\sum_{j=1}^{n_i} \mathrm{X}_{ij}\mathrm{A}_{ij}\mathrm{E}_{\mathrm{Y}|\mathrm{A}}\left(\mathrm{E}_{S|\mathrm{Y,A}}\left[\mathrm{E}_{\mathrm{R}|S,\mathrm{Y,A}}(\mathrm{R}_{ij})(\mathrm{Y}_{ij} - \mu_{ij}^{A})\right]\right) \\ &\overset{\text{MCAR-S}}{=} \sum \mathrm{X}_{ij}\mathrm{A}_{ij}\mathrm{E}_{\mathrm{Y}|\mathrm{A}}\left[P(\mathrm{R}_{ij} = 1|S_i = s_i)(\mathrm{Y}_{ij} - \mu_{ij}^{A})\right]. \end{aligned}$$

The MCAR-S dropout process $f(\mathrm{R}_i|\phi)$ is not ignorable: it depends on the specific value of an observed random variable, S_i.

MCAR-S missing data can be accommodated by including IPC weights in the IPCW-IEE quasi-score:

$$U(\beta) = \sum_{i=1}^{N}\sum_{j=1}^{n_i} \mathrm{X}_{ij}\mathrm{A}_{ij}\frac{\mathrm{R}_{ij}}{\pi_{ij}^{s}}(\mathrm{Y}_{ij} - \mu_{ij}^{A})$$

where $\pi_{ij}^{s} = P(\mathrm{R}_{ij} = 1|S_i = s_i)$. Models for $f(\mathrm{R}_i|\phi)$ may be more complicated than for $f(\mathrm{Y}_{ij}|S_i > t_j)$, and may include covariates that are related to Y_{ij} but unmodeled (such as S_i, or past responses $\mathrm{H}_{ij}^{(\mathrm{Y})}$ for MAR data). Robins et al. (1995) emphasize that covariates related both to R_i and Y_i may be modeled in $f(\mathrm{R}_i|\phi)$ but not in $f(\mathrm{Y}_i|\beta)$. An inclusive model for π_{ij}^{s} is necessary to ensure consistent estimation of β, while the regression model

must be of substantive interest (and may exclude known confounders). In short, any model for R_i using observed data makes direct estimation of μ_{ij}^A possible, as long as

$$E\left(\frac{R_{ij}Y_{ij}}{\pi_{ij}^s}\middle| X_{ij}, S_i > t_j\right) = E(Y_{ij}|X_{ij}, S_i > t_j).$$

Details of estimation for IPCW weights are given in Section 7.3. Having modeled π^s, β will be estimated consistently:

$$\begin{aligned} E_{Y,R,S|A}[U(\beta)] &= \sum_{i=1}^{N}\sum_{j=1}^{n_i} X_{ij}A_{ij}E_{Y|A}\left(E_{S|Y,A}\left[E_{R|Y,S}\left(\frac{R_{ij}}{\pi_{ij}^s}\right)(Y_{ij}-\mu_{ij}^A)\right]\right) \\ &\overset{\text{MCAR-S}}{=} \sum X_{ij}A_{ij}E_{Y|A}\left(E_{S|Y,A}\left[\frac{\pi_{ij}^s}{\pi_{ij}^s}(Y_{ij}-\mu_{ij}^A)\right]\right) \\ &= \sum X_{ij}A_{ij}E_{Y|A}(Y_{ij}-\mu_{ij}^A). \end{aligned}$$

Direct estimation of μ_{ij}^A is possible, and the survival distribution $f(S_i|\eta)$ is ignorable, since it is not involved in the moment-based estimating equation. However, unlike non-survival-dependent MCAR dropout (for data with or without deaths), survival-dependent MCAR (MCAR-S) dropout is not ignorable for estimation of regression parameters β by independence estimating equations (IEE). Ignorability of $f(R_i|\phi)$ is lost because dropout depends on observed random variable S_i. By analogy, MAR dropout depends on observed random variables in $H_{ij}^{(Y)}$ and is also not ignorable for analysis using GEE. Augmenting IEE with properly specified inverse probability of censoring weights can yield consistent estimates for DECAL parameters β.

7.5 MAR-S

Section 7.4 shows that even when dropout is survival-dependent MCAR (MCAR-S), $f(S_i|\eta)$ may still be ignorable and direct estimation of β is possible. Moment-based methods (as opposed to likelihood-based) can estimate $E(R_{ij}|S_i = s_i)$ as part of the quasi-score defining $E(Y_{ij}|S_i > t_j)$. MAR-S estimation is similar to MAR and MCAR-S. IPC weights are the inverse of

$$\pi_{ij}^s = \prod_{k=2}^{j} p_{ik}^s$$

where $p^s_{ik} = P(\mathrm{R}_{ik} = 1 | \mathrm{R}_{ik-1} = 1, S_i = s_i, S_i > t_j)$ and

$$\mathrm{logit}(p^s_{ik}) = \mathrm{D}_{ik}\phi,$$

where model matrix row D_{ik} includes S_i, relevant elements of $\mathrm{H}^{(\mathrm{Y})}_{ik}$, and other covariates.

A practical problem with MAR-S is that if dropout and death rarely both occur, estimation of ϕ may not be possible without placing restrictions on dropout parameters for S_i. A reasonable way to reduce the number of parameters to estimate under MAR-S is to model survival time S_i as distance from t_j. For example, p^s_{ij} could be modeled as "probability of dropping out at time t_j, given that Y_{ij-1} was observed and that subject i will die in $(t_s - t_j)$ months." This approach is illustrated using the sample data in Table 7.1:

$P(\mathrm{R}_4 = 1|$ die within one time period)

$$\begin{aligned} &= P(\mathrm{R}_4 = 1 | R_3 = 1, \text{ die in 1}) \cdot P(\mathrm{R}_3 = 1 | R_2 = 1, \text{ die in 2}) \cdot P(\mathrm{R}_2 = 1 | \text{ die in 3}) \\ &= \frac{1}{2} \cdot \frac{2}{3} \cdot \frac{3}{3} \\ &= \frac{1}{3} \quad \text{the correct weight.} \end{aligned}$$

That the correct weight is the same as for the previous example is a coincidence due to the simplicity of the sample records. The example shows how the telescoping works and shows that the dropout process may still be modeled by discrete-time survival methods.

7.6 Survival Censored for Some Subjects

If a subject has dropped out or if the study period has ended, survival status may be unknown for some subjects. Censored survival differs from monotone dropout, both in mechanisms and in character of missing information. *Dropout* may occur when a subject refuses to participate, or when the subject cannot be contacted but a relative confirms that the subject is alive. After the time of dropout, values of dropout indicator R_{ij} are 0, and response values (Y_{ij}) exist but are unknown. Monotone dropout does not affect n_i, the planned number of responses for subject i. *Censored survival* may occur if followup ceases due to lack of funds, or if all available contacts are exhausted without locating a subject. When survival status is unknown for subject i at time t_j, R_{ij} is 0 (if the subject is alive) or

Table 7.2: Data records to illustrate weighting schemes under dropout and death, with censored survival time. ($\surd$ = observed, $\cdot$ = response missing, $\times$ = deceased, $\odot$ = survival status unknown, and response missing)

id	time 1	time 2	time 3	time 4
1	$\surd$	$\surd$	$\surd$	$\surd$
2	$\surd$	$\cdot$	$\times$	$\times$
3	$\surd$	$\surd$	$\odot$	$\odot$
4	$\surd$	$\times$	$\times$	$\times$
5	$\surd$	$\surd$	$\surd$	$\odot$

undefined (if the subject is dead), Y_{ij} is missing or undefined, and n_i may or may not be truncated due to death. This section explores the implications of censored survival status on IPCW-IEE.

To illustrate the problem of censored survival times in the presence of dropout and death, consider Table 7.2, which is Table 7.1 with "time 5" omitted so that survival time is no longer known for subjects 1, 3, and 5. Symbol "$\cdot$" represents data missing due to dropout. The subject is known to be alive. Symbol "$\odot$" appears when no response is observed and survival status is unknown. Should the time 4 observed data for subject 1 be weighted to represent subjects 3 and 5? Weighting for a DECAL IPCW-IEE model is based on the inverse probability of censoring *conditioning on being alive*, and we do not know if subjects 3 and 5 are alive at time 4.

7.6.1 Censored Survival without Dropout

We first consider data for which survival status is unknown whenever a value for Y_{ij} is not recorded. "Dropout" as defined above does not occur, but survival status (and response) may be censored. An example of this situation is a dataset with subjects 1, 3, 4, and 5 of Table 7.2, but not subject 2.

To examine the impact of censored survival on DECAL by IEE, we introduce random variable C_{ij}, which takes the value 1 if survival status is known at time t_j, and value 0 if

survival is unknown. The IEE quasi-score for β is:

$$U(\beta) = \sum_{i=1}^{N} \sum_{j=1}^{n_i} \mathrm{X}_{ij}\mathrm{A}_{ij}\mathrm{C}_{ij}(\mathrm{Y}_{ij} - \mu_{ij}^{\mathrm{A}}).$$

When $\mathrm{C}_{ij} = 0$, we do not know whether A_{ij} is 0 or 1, or whether Y_{ij} is missing (if $\mathrm{A}_{ij} = 1$, the subject is alive) or undefined (if the subject is dead). The true value of n_i is also unknown, since the trajectory may or may not be truncated by death.

Censoring Completely at Random

Since it is unclear whether censored observations are even missing, we address ignorable censoring before considering recovery of information lost due to censoring. Under what conditions is censoring ignorable? Analogous to MCAR in the dropout taxonomy, censored survival need not be modeled for IEE estimation if

$$P(\mathrm{C}_{ij} = 1|\mathrm{D}_{ij}^{\mathrm{c}}, \mathrm{Y}_{ij}, \mathrm{H}_{ij}^{(\mathrm{Y})}, S_i > t_j) = P(\mathrm{C}_{ij} = 1|\mathrm{D}_{ij}^{\mathrm{c}}, S_i > t_j) \tag{7.4}$$

where $\mathbf{D}_i^{\mathrm{c}}$ is the model matrix for censoring survival. Equation 7.4 implies that

$$\mathrm{E}(\mathrm{Y}_{ij}|\mathrm{X}_{ij}, S_i > t_j, \mathrm{C}_{ij} = 1) = \mathrm{E}(\mathrm{Y}_{ij}|\mathrm{X}_{ij}, S_i > t_j) \tag{7.5}$$

where X_{ij} is a subset of $\mathrm{D}_{ij}^{\mathrm{c}}$. This can be described as a form of independent censoring (Fleming & Harrington, 1991, p. 27), for which expected values of Y are estimated instead of a hazard function. Extending the analogy to MCAR, we refer to Equation 7.4 as *censoring completely at random* (CCAR). Administrative censoring, where followup ends because of the end of the study, is an example where CCAR is a reasonable assumption.

In contrast to MCAR, CCAR is not a testable assumption. Hypotheses about MCAR versus MAR dropout models can be tested using observed data. Because S_i is unobserved when $\mathrm{C}_{ij} = 0$, we cannot test whether Equation 7.4 is true.

Censoring not Completely at Random

What if the identity in Equation 7.4 cannot be assumed to be true? Can we recover consistent estimation of β by IEE through inverse probability of censoring weights? The

sample data records in Table 7.2 show that IPCW-GEE may not be straightforward to apply. For data with dropout, observed data are weighted to represent unobserved data. In Table 7.1 the IPC weight for subject 1 at time 4 is clearly 3. For data censored by death, a DECAL approach requires weighting only for subjects who are alive. In Table 7.2 the correct weight for subject 1 at time 4 cannot be determined, since survival status is unknown for subjects 3 and 5.

Consider *censoring at random*, where

$$P(\mathrm{C}_{ij} = 1|\mathrm{D}^{c}_{ij}, \mathrm{Y}_{ij}, \mathrm{H}^{(\mathrm{Y})}_{ij}, S_i > t_j) = P(\mathrm{C}_{ij} = 1|\mathrm{D}^{c}_{ij}, \mathrm{H}^{(\mathrm{Y})}_{ij}, S_i > t_j).$$

Under censored survival and no dropout, IPC weights π^{c}_{ij} should be modeled so that

$$\mathrm{E}\left(\frac{\mathrm{C}_{ij}\mathrm{Y}_{ij}}{\pi^{c}_{ij}} \middle| \mathrm{X}_{ij}, S_i > t_j \right) = \mathrm{E}(\mathrm{Y}_{ij}|X_{ij}, S_i > t_j).$$

The correct weights are the inverse of $P(\mathrm{C}_{ij} = 1|\mathrm{D}^{c}_{ij}, \mathrm{H}^{(\mathrm{Y})}_{ij}, S_i > t_j)$ (see Section 7.1): the probability of survival not being censored, given that the subject is alive. However, no data records exist where $\mathrm{C}_{ij} = 0$, and $S_i > t_j$ is known to be true. Without further assumptions, the model for IPC weights π^{c}_{ij} is not estimable, due to missing covariate information about survival status. One such assumption could be that $P(\mathrm{C}_{ij} = 1|S_i > t_j) = P(\mathrm{C}_{ij} = 1|S_i > t_{j-1})$.

7.6.2 Censored Survival and Dropout (MCAR, MAR)

We next address data for which both dropout and censored survival status occur. Dropout and censored survival occur for different reasons and have different implications for DECAL models. Therefore, we model the processes separately, by random vectors R_i and C_i. We consider the impact of both R_i and C_i on DECAL using IEE or IPCW-IEE. What assumptions or extra modeling are required for consistent estimation of regression parameter vector β?

Impact on Regression Model (β)

The IPCW-IEE quasi-score for β must reflect the presence of dropout and censored survival data. Again, R_{ij} is an indicator for whether Y_{ij} is not missing due to dropout. Indicator

C_{ij} equals 1 if survival status is known at time t_j, and is 0 if survival is unknown. Survival status also appears in the quasi-score through indicator random variable A_{ij}. ($A_{ij} = 1$ if subject i is alive at time t_j.) The IEE quasi-score is:

$$U(\beta) = \sum_{i=1}^{N} \sum_{j=1}^{n_i} X_{ij} A_{ij} C_{ij} R_{ij} (Y_{ij} - \mu_{ij}^A).$$

We first note that if survival status is unknown at time t_j ($C_{ij} = 0$), we do not know whether Y_{ij} is missing (if subject i is alive at time t_j) or undefined (if subject i is dead). The presence of C_{ij} in the quasi-score does not have any direct effect on *estimation* of β: if $C_{ij} = 0$ and subject i is alive, then $R_{ij} = 0$ as well. If $C_{ij} = 0$ and subject i is not alive, then $A_{ij} = 0$. The contribution to the score is zero when $C_{ij} = 0$, whether or not subject i is alive. However, we must consider whether the presence of C_{ij} affects the *target of inference* for the regression model, or for IPC weights (below).

If dropout is MCAR (Table 6.2) and survival is censored completely at random (Equation 7.4) then, for X_{ij} including relevant elements of D_{ij}^c:

$$\begin{aligned}
&P(Y_{ij} = y_{ij} | X_{ij}, S_i > t_j, C_{ij} = 1, R_{ij} = 1) \qquad (X_{ij} \text{ suppressed on next 2 lines}) \\
&= \frac{P(R_{ij} = 1 | Y_{ij}, S_i > t_j, C_{ij} = 1) \cdot P(C_{ij} = 1 | Y_{ij}, S_i > t_j) \cdot P(Y_{ij} | S_i > t_j) \cdot P(S_i > t_j)}{P(R_{ij} = 1 | S_i > t_j, C_{ij} = 1) \cdot P(C_{ij} = 1 | S_i > t_j) \cdot P(S_i > t_j)} \\
&\underline{\underline{\text{MCAR}}} \quad \frac{P(R_{ij} = 1 | S_i > t_j, C_{ij} = 1) \cdot P(C_{ij} = 1 | Y_{ij}, S_i > t_j) \cdot P(Y_{ij} = y_{ij} | S_i > t_j)}{P(R_{ij} = 1 | S_i > t_j, C_{ij} = 1) \cdot P(C_{ij} = 1 | S_i > t_j)} \\
&\underline{\underline{\text{CCAR}}} \quad \frac{P(C_{ij} = 1 | X_{ij}, S_i > t_j) \cdot P(Y_{ij} = y_{ij} | X_{ij}, S_i > t_j)}{P(C_{ij} = 1 | X_{ij}, S_i > t_j)} \\
&= \quad P(Y_{ij} = y_{ij} | X_{ij}, S_i > t_j).
\end{aligned}$$

Both dropout and censoring processes are ignorable. IEE using available data results in consistent estimation of β.

Impact on Dropout Model (ϕ)

What if data are CCAR but not MCAR? IPCW-IEE must be adapted to acknowledge MAR dropout and censored survival. (Random vectors R_i and C_i are not independent or conditionally independent, since $R_{ij} \neq 1$ when $C_{ij} = 0$.) Consider IPC weights $\pi_{ij}^r =$

$P(\mathrm{R}_{ij} = 1|\mathrm{D}_{ij}, S_i > t_j, \mathrm{C}_{ij} = 1)$, where matrix $\mathbf{D}_i$ includes covariates and elements of response history $\mathrm{H}_{ij}^{(\mathrm{Y})}$. Weights π_{ij}^r can be modeled using observed data, since

$$\begin{aligned}
\pi_{ij}^r &= P(\mathrm{R}_{ij} = 1|\mathrm{D}_{ij}, S_i > t_j, \mathrm{C}_{ij} = 1) \\
&= \frac{[P(\mathrm{C}_{ij} = 1|\mathrm{D}_{ij}, S_i > t_j, \mathrm{R}_{ij} = 1) = 1] \cdot P(\mathrm{R}_{ij} = 1|\mathrm{D}_{ij}, S_i > t_j) \cdot P(S_i > t_j|\mathrm{D}_{ij})}{P(\mathrm{C}_{ij} = 1|\mathrm{D}_{ij}, S_i > t_j) \cdot P(S_i > t_j|\mathrm{D}_{ij})} \\
&= \frac{P(\mathrm{R}_{ij} = 1|\mathrm{D}_{ij}, S_i > t_j)}{P(\mathrm{C}_{ij} = 1|\mathrm{D}_{ij}, S_i > t_j)}. \qquad (7.6)
\end{aligned}$$

The numerator in Equation 7.6 is known not to depend on Y_{ij} (MAR), and the denominator is independent of both Y_{ij} and $\mathrm{H}_{ij}^{(\mathrm{Y})}$ (since survival is CCAR). Therefore, we can proceed using weights π_{ij}^r. Telescoping properties under MAR (Appendix E) apply in this case as well.

Assuming π_{ij}^r is modeled correctly,

$$\begin{aligned}
\mathrm{E}\left(\frac{\mathrm{C}_{ij}\mathrm{R}_{ij}\mathrm{Y}_{ij}}{\pi_{ij}^r}\middle| \mathrm{X}_{ij}, S_i > t_j, \mathrm{C}_{ij} = 1\right) &= \mathrm{E}(\mathrm{Y}_{ij}|\mathrm{X}_{ij}, S_i > t_j, \mathrm{C}_{ij} = 1) \\
&\overset{\text{CCAR}}{=} \mathrm{E}(\mathrm{Y}_{ij}|\mathrm{X}_{ij}, S_i > t_j). \quad \text{(See Eq. 7.5, page 105.)}
\end{aligned}$$

If data are MAR (rather than MCAR) and CCAR, IPCW-IEE can be applied using available data.

Censored survival truncates R_i for the estimating equation $U(\phi)$, just as death does. As shown above (Section 7.3.1), estimation of dropout parameter vector ϕ is through discrete-time survival analysis. We can write the estimating equation for ϕ as follows:

$$U(\phi) = \sum_{i=1}^{N} \sum_{j=1}^{n_i} \mathrm{D}_{ij}\mathrm{A}_{ij}\mathrm{R}_{ij-1}\mathrm{C}_{ij}(\mathrm{R}_{ij} - p_{ij}^r).$$

Let $u_{ik} = P(\mathrm{R}_{ik} = 1|\mathrm{H}_{ik}^{(\mathrm{R})}, \mathrm{H}_{ik}^{(\mathrm{Y})}, S_i > t_k, \mathrm{C}_{ik} = 1)$ and $p_{ik}^r = P(\mathrm{R}_{ik} = 1|\mathrm{H}_{ik}^{(\mathrm{R})} \equiv 1, \mathrm{H}_{ik}^{(\mathrm{Y})}, S_i > t_k, \mathrm{C}_{ik} = 1)$. Dropout time is t_d, survival time is t_s, and time of censoring is t_c. Censoring of survival status is considered to be monotone. If death is confirmed after a period when survival is censored, the specific survival time will be recorded. The alternative would be interval censoring of survival times. Then:

$$P(\mathrm{R}_{ik} = r_{ik}|\mathrm{H}_{ik}^{(\mathrm{R})} \equiv 1, \mathrm{H}_{ik}^{(\mathrm{Y})}, S_i > t_k, \mathrm{C}_{ij} = 1) = u_{ik}^{r_{ik}}(1 - u_{ik})^{1-r_{ik}}$$

and

$$u_{ik} = \begin{cases} p^r_{ik} & : \ k \leq d, k < s, k < c \\ 0 & : \ k > d, k < s, k < c \\ \text{undefined} & : \ k \geq s \text{ (deceased)} \\ \text{undefined} & : \ k \geq c \text{ (censored survival)} \end{cases}$$

The logistic regression model for dropout parameters ϕ:

$$\text{logit}(p^r_{ik}) = \mathrm{D}_{ik}\phi$$

is solved using the likelihood:

$$P(\mathrm{R}_{ij} = r_{ij} | \mathrm{H}^{(\mathrm{Y})}_{ij}, S_i > t_j, \mathrm{C}_{ij} = 1) = \left(\prod_{k=2}^{(d-1)\wedge(s-1)\wedge(c-1)\wedge j} p^r_{ik} \right) (1 - p^r_{id})^{I[d \leq j < (s \wedge c)]}. \quad (7.7)$$

Equation 7.7 differs from Equation 7.3 on page 100 in that death and censored survival both censor the dropout process. Solving Equation 7.7 for ϕ yields probabilities $p^r_{ij} = P(\mathrm{R}_{ij} = 1 | \mathrm{R}_{ij-1} = 1, S_i > t_j, \mathrm{C}_{ij} = 1)$, and IPC weights are:

$$\pi^r_{ij} = P(\mathrm{R}_{ij} = 1 | S_i > t_j, \mathrm{C}_{ij} = 1) = \prod_{k=2}^{j} p^r_{ij}.$$

Sensitivity Analysis When Not CCAR

Section 7.6.1 (page 104) notes that if data are not CCAR, IPC weights $\pi^c_{ij} = P(\mathrm{C}_{ij} = 1 | S_i > t_j)$ are not generally estimable from observed data, since covariate A_{ij} is missing whenever $\mathrm{C}_{ij} = 0$. However, sensitivity analysis can be conducted with values of $\pi^c_{ij} = P(\mathrm{C}_{ij} = 1 | S_i > t_j)$ held constant, in the same manner as for data with NI dropout (see Section 5.3). In the presence of both missing data and censored survival, weights π^*_{ij} are needed such that

$$\mathrm{E}\left(\frac{\mathrm{C}_{ij}\mathrm{R}_{ij}\mathrm{Y}_{ij}}{\pi^*_{ij}} \middle| \mathrm{X}_{ij}, S_i > t_j \right) = \mathrm{E}(\mathrm{Y}_{ij} | \mathrm{X}_{ij}, S_i > t_j).$$

The weights π^*_{ij} are $\mathrm{E}[(\mathrm{C}_{ij} \wedge \mathrm{R}_{ij}) | \mathrm{D}^*_{ij}, S_i > t_j]$ for model matrix $\mathbf{D}^*_i$ that may include past response values ($\mathrm{H}^{(\mathrm{Y})}_{ij}$). Suppressing dependence on $\mathbf{D}^*_i$:

$$\begin{aligned} \pi^*_{ij} = P(\mathrm{C}_{ij} = 1, \mathrm{R}_{ij} = 1 | S_i > t_j) &= P(\mathrm{R}_{ij} = 1 | \mathrm{C}_{ij} = 1, S_i > t_j) \cdot P(\mathrm{C}_{ij} = 1 | S_i > t_j) \\ &= \pi^r_{ij} \cdot \pi^c_{ij}. \end{aligned}$$

For estimation using fixed weights π^c_{ij} when data are not CCAR, dropout weights π^r_{ij} and censoring weights π^c_{ij} are simply multiplied to give appropriate weights for IPCW-IEE DECAL models. If dropout is MCAR (not MAR), π^r_{ij} does not depend on response values and does need to be estimated.

7.6.3 Censored Survival and Dropout (MCAR-S, MAR-S)

The previous section shows how MCAR or MAR data may be analyzed using IEE or IPCW-IEE if the survival is censored completely at random (CCAR). However, if the missingness mechanism depends on specific survival time (MCAR-S or MAR-S), a covariate (survival time) is missing from the estimation of IPC weights needed for consistent estimation of DECAL model parameters.

Finally, we return to Table 7.2 (page 104) to consider how assumptions about dropout and censored survival affect DECAL models for this example. If data are MAR and CCAR then weights π^r_{ij} (page 108) are needed for consistent estimation of β, and are estimable using observed data:

$$P(\mathrm{R}_4 = 1 | S > 4, \mathrm{C}_4 = 1) = \frac{1}{1}.$$

The two cases with censored survival (subjects 3 and 5) are not considered when computing weights π^r_{ij}. When survival is censored completely at random (CCAR), symbol $\odot$ truncates the data records without affecting consistency of DECAL parameters. Although it seems that censored survival is treated as death (since both truncate the data record), a more appropriate comparison is that censored survival is akin to truncation due to the end of followup, a form of "independent censoring" that should not affect parameter estimation.

7.7 Summary and Discussion

Table 7.3 summarizes methods for analysis of longitudinal binary data with dropout and death, described in this chapter and in Chapter 6. Some methods permit DECAL analysis (Direct Estimation Conditioning on being ALive), while others are not direct. As shown in Chapter 6, likelihood-based models do not directly parameterize $\mu^A_{ij} = \mathrm{E}(\mathrm{Y}_{ij} | S_i > t_j)$ and so must recover expected values using a mixing distribution with $f(S_i|\eta)$. Estimation

using GEE or IPCW-GEE with non-independence working correlation is not direct because covariance weights depend on cluster size n_i, which is truncated when subjects die. Independence estimating equations (IEE) and IPCW-IEE are examined in this chapter. Estimation of parameter vector β for regression modeling of μ_{ij}^A can be direct, although $f(\mathrm{R}_i|\phi)$ is not always ignorable. The taxonomy for DECAL methods differs from data missing due to monotone dropout in that $f(\mathrm{R}_i|\phi)$ is not ignorable for survival-dependent MCAR data (MCAR-S). MCAR data do not depend on any random variables, either S_i or Y_i. MCAR-S, MAR, and MAR-S depend on observed random variables: S_i, $\mathrm{H}_{ij}^{(\mathrm{Y})}$, and both (respectively).

For methods examined in Table 7.3, survival distribution is always ignorable whenever estimation of β is direct. This is not necessarily true for all methods of analysis. Miller et al. (2001) present a method in which response Y_{ij} is modeled conditioning on covariates, prior responses ($\mathrm{H}_{ij}^{(\mathrm{Y})}$), and survival until timepoint t_j ($S_i \geq t_j$). Survival and responses are jointly modeled under a factorization in which estimation of Y_{ij} is at least partially direct, but $f(S_i|\eta)$ is not ignorable. The method also includes IPC weights for monotone dropout, modeled as $P(\mathrm{R}_{it} = 1|\mathrm{R}_{it-1} = 1, \mathrm{H}_{ij}^{(\mathrm{Y})}, S_i \geq t_j)$.

When survival is censored for some subjects, DECAL estimation is relatively unchanged if censoring is independent of the response, conditioning on the subject's being alive. If censoring is not independent of the response, censoring weights cannot be estimated because weights would condition on unknown survival status.

Examination of R_i and C_i above suggests that dropout and censoring of survival are similar processes. When dropout occurs, response data (Y_i) are neither completely observed nor completely missing. When survival is censored, time of death (S_i) is known for some subjects, but not all. Under dropout and censored survival, Y_i and S_i are examples of *coarse data*. Ignorability of coarsening mechanisms (such as R_i and C_i) are examined as a general framework by Heitjan & Rubin (1991). MAR and CCAR are both special cases of "coarsening at random".

IPCW-IEE is a straightforward method that allows for direct parameterization of $\mathrm{E}(\mathrm{Y}_{ij}|S_i > t_j)$ without a great deal of additional modeling. However, even if censoring of survival is independent of response (CCAR), if survival status is always (or almost always) unknown when subjects drop out, then parameters ϕ may not be estimable, since data

records with $C_{ij} = R_{ij} = 0$ do not contribute to the score equations for ϕ. One potential solution is multiple imputation (Rubin, 1987) of survival times. Multiple imputation is a natural approach to solving many of the complicated missing data problems described here and in the previous chapter, and merits further investigation.

This chapter has identified situations in which DECAL estimation is possible. The next chapter includes an example of DECAL analysis (using the PEP data from Chapter 1), and demonstrates bias when an incorrect target of inference is assumed for methods that do not permit DECAL regression, such as likelihood-based methods and GEE with non-independence working correlation.

Table 7.3: Summary: Direct Estimation Conditioning on Being ALive (DECAL).

	Ignorable $f(\mathrm{R}_i\|\phi)$	Ignorable $f(S_i\|\eta)*$	Comments
Likelihood			
MCAR,MCAR-S	√	x	Mixture of μ_{ij}^S, $f(S_i\|\eta)$
MAR, MAR-S	√	x	Mixture of μ_{ij}^S, $f(S_i\|\eta)$
GEE, IPCW-GEE (non-IEE)			
MCAR	√	x	Dimensions depend on $S_i = s_i$.
MCAR-S	x	x	Mixture of μ_{ij}^S, $f(S_i\|\eta)$
MAR,MAR-S	x	x	
IEE, IPCW-IEE			
MCAR	√	√	
MCAR-S	x	√	Model $P(\mathrm{R}_{ij}=1\|S_i = s_i, S_i > t_j)$
MAR	x	√	Model $P(\mathrm{R}_{ij}=1\|\mathrm{H}_{ij}^{(\mathrm{Y})}, S_i > t_j)$
MAR-S	x	√	Model $P(\mathrm{R}_{ij}=1\|\mathrm{H}_{ij}^{(\mathrm{Y})}, S_i = s_i, S_i > t_j)$

* Also, direct β for μ_{ij}^A

MAR non-survival-dependent MAR data

$P(\mathrm{R}_{ij} = 1|\mathrm{Y}_i, S_i = s_i) = P(\mathrm{R}_{ij} = 1|\mathrm{H}_{ij}^{(\mathrm{Y})}, S_i > t_j)$

MAR-S survival-dependent MAR data

$P(\mathrm{R}_{ij} = 1|\mathrm{H}_{ij}^{(\mathrm{Y})}, S_i = s_i) \neq P(\mathrm{R}_{ij} = 1|\mathrm{H}_{ij}^{(\mathrm{Y})}, S_i > t_j)$

μ_{ij}^A $\mathrm{E}(\mathrm{Y}_{ij}|S_i > t_j, \beta)$

μ_{ij}^S $\mathrm{E}(\mathrm{Y}_{ij}|S_i = s_i, \beta^s)$

Chapter 8

BIAS OF STANDARD METHODS FOR DIRECT ESTIMATION CONDITIONING ON BEING ALIVE (DECAL)

Chapters 1 and 6 introduce Direct Estimation Conditioning on being ALive (DECAL), in which $\mu_{ij}^A = \mathrm{E}(\mathrm{Y}_{ij}|S_i > t_j)$ is the target of inference for response vector Y_i in regression models for longitudinal data with survival time S_i. Chapter 7 examines methods that can parameterize μ_{ij}^A directly. In this chapter, simulation studies demonstrate bias in parameter estimates when regression models are fitted that do not account for survival status. To illustrate that target of inference issues are not limited to longitudinal data with binary outcome, we include simulation results for both linear and binary response data. Data from the Precipitating Events Project (PEP), introduced in Chapter 1, are analyzed to show that inverse probability of censoring weighted generalized estimating equations (IPCW-GEE) and the likelihood-based marginalized transition model (MTM) do not yield DECAL regression models when fitted to available data.

8.1 Simulation Overview

Data generated for simulation studies parameterize regression mean $\mu_{ij}^A = \mathrm{E}(\mathrm{Y}_{ij}|S_i > t_j)$. However, specification of μ_{ij}^A alone does not provide information about the association among responses, or between response values and the survival distribution. For linear responses, the bivariate normal distribution $f(\mathrm{Y}_{ij}, S_i)$ is used to specify the regression model for $\mathrm{E}(\mathrm{Y}_{ij}|S_i > t_j)$ (Johnson & Kotz, 1972; Heckman, 1979). To simulate binary data, the DECAL model μ_{ij}^A, marginal survival mean, and association between S_i and Y_i are used to derive a marginal mean $\mu_{ij}^M = \mathrm{E}(\mathrm{Y}_{ij})$ for responses Y_i. Notation for models with dropout and death is described in Table 6.1. Additional notation is introduced below, as needed.

8.2 Simulation Study: Linear Response

8.2.1 Simulation Details

To generate linear responses correlated to survival time, we adopt a linear mixed model in which survival distribution $f(S_i)$ and response distribution $f(Y_i)$ share random effects vector b_i.

The regression model for response values conditioning on survival is:

$$E(Y_{ij}|S_i > t_j) = \beta_0 + \beta_1 \cdot \texttt{group}_i + \beta_2 \cdot \texttt{time}_{ij} + \beta_3 \cdot \texttt{group}_i \cdot \texttt{time}_{ij}.$$

For subject i, $\texttt{group}_i$ is a binary cluster-level covariate. Interpreted as sex, $\texttt{group}_i = 0$ for women and $\texttt{group}_i = 1$ for men. Covariate vector $\texttt{time}_i$ is planned to be $(0, 5, 10, 15, 20)$ for all subjects, but is truncated for subjects who die before the end of followup. Regression parameter vector $\beta = (28, -2, -0.25, 0.1)$ produces the trajectories plotted in panel (a) of Figure 8.1. The plot represents hypothetical Mini-Mental State Examination (MMSE, Folstein et al., 1975) scores for ages 65-85 ($\texttt{time}+65$) in a hypothetical observational study. At age 65, surviving women enjoy an advantage in cognitive functioning, but predicted MMSE for men and women are equal by age 85.

Survival times are generated according to a normal distribution. This is generally not realistic, but has been applied in data analysis as well as simulation studies (De Gruttola & Tu, 1994), and is adopted here for illustrative purposes. The average survival time is 85 years for women ($\texttt{group} = 0$) and 80 years for men. The standard deviation for both is 2.5 years.

Between-subject variation is introduced through a random lines linear mixed model. Response and survival models share random intercept b_{i0} and random slope b_{i1} where

$$\begin{pmatrix} b_{i0} \\ b_{i1} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \mathbf{D}_i = \begin{pmatrix} 1 & -.02 \\ -.02 & .06 \end{pmatrix} \right).$$

The model matrix for the response random effects is $\mathbf{Z}_i$ (intercept and $\texttt{time}$ vector). Survival depends only on the random intercept. (Vector $\mathbf{V}_i = (1.5, 0.0)$.) Mean-zero normally distributed random error is added to responses (standard deviation 1) and survival time (standard deviation 2.5 years).

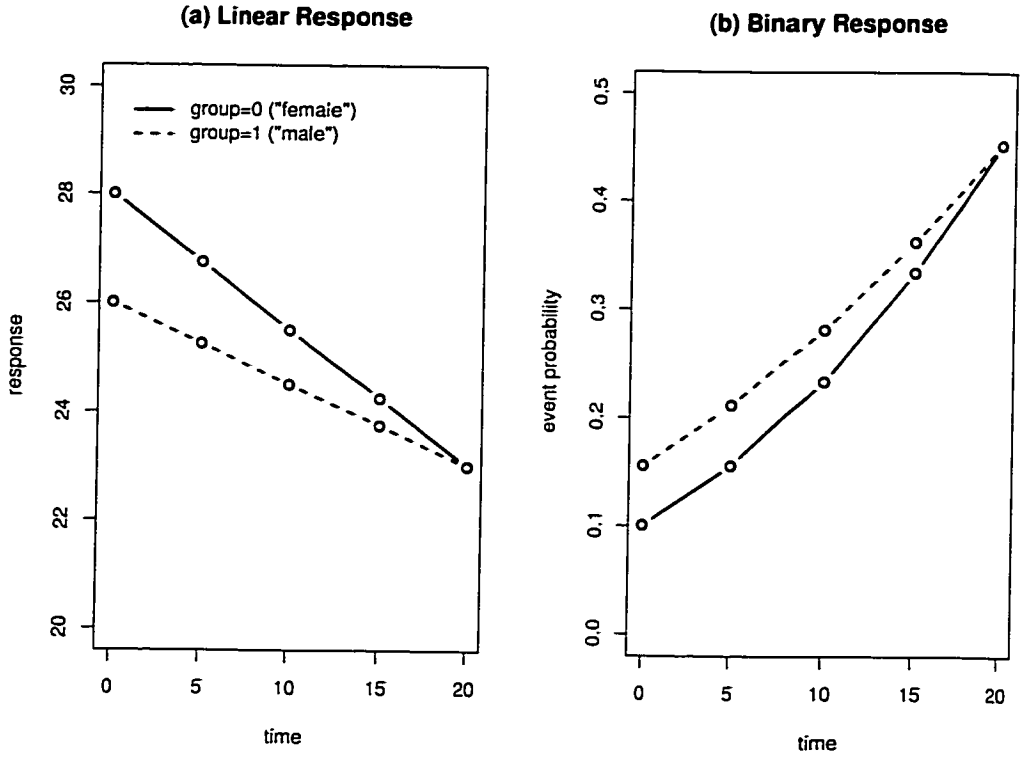


Figure 8.1: DECAL regression model used for simulations.

The regression model for response and survival values is:

$$\begin{aligned} \mathrm{Y}_i &= \mu_i^M + \mathbf{Z}_i b_i + \epsilon_i, & \epsilon_{ij} \sim \mathrm{N}(0, \sigma_\epsilon^2 = 1) \\ S_i &= \mathrm{X}_i^* \eta + \mathbf{V}_i b_i + e_i, & e_i \sim \mathrm{N}(0, \sigma_e^2 = 2.5^2) \end{aligned}$$

where X_i^* is a model matrix for intercept and group. We assume that e_i and ϵ_i are independent. Covariance of Y_i and S_i is specified through matrix Σ_i, where $\mathbf{I}_i$ is an identity matrix of dimension 5:

$$\Sigma_i = \begin{pmatrix} \mathbf{Z}_i \mathbf{D}_i \mathbf{Z}_i^T + \sigma_\epsilon^2 \cdot \mathbf{I}_i & \mathbf{Z}_i \mathbf{D}_i \mathbf{V}_i^T \\ \mathbf{V}_i \mathbf{D}_i \mathbf{Z}_i^T & \mathbf{V}_i \mathbf{D}_i \mathbf{V}_i + \sigma_e^2 \end{pmatrix}.$$

The joint multivariate normal distribution $f(\mathrm{Y}_i, S_i)$ is *not* directly parameterized by β, which models μ_{ij}^A rather than μ_{ij}^M. Heckman (1979) notes a result of Johnson & Kotz (1972),

that for a multivariate normal distribution, the expectation $\mathrm{E}(\mathrm{Y}_{ij}|S_i > t_j)$ can be expressed in relation to the joint distribution of Y_i and S_i using Mill's ratio. For each time t_j,

$$\mu_{ij}^A = \mathrm{E}(\mathrm{Y}_{ij}|S_i > t_j) = \mu_{ij}^M + \frac{\sigma_{SY}}{\sqrt{\sigma_{SS}}}\lambda_{ij} \tag{8.1}$$

where $\sigma_{SY} = \mathrm{Cov}(S_i, \mathrm{Y}_{ij})$ and $\sigma_{SS} = \mathrm{Var}(S_i)$ are elements of Σ_i defined above, and

$$\lambda_{ij} = \frac{\phi(z_{ij})}{1 - \Phi(z_{ij})}$$

is the inverse of Mill's ratio, where ϕ is the density function and Φ the cumulative density function of a standard normal random variable, and

$$z_{ij} = \frac{t_j - \mathrm{E}(S_i)}{\sqrt{\sigma_{SS}}}.$$

Equation 8.1 is solved to find μ_{ij}^M.

One thousand datasets with 1000 clusters (500 for each group) are generated according to the linear mixed model. Response values Y_{ij} after survival time S_i are defined by μ_{ij}^M but are deleted from simulated datasets. No data are missing due to dropout. About 39% of simulated subjects survive to the end of followup. 24% complete four responses before dying, 20% contribute three responses, 12% contribute 2 responses, and 5% record one response before dying. About 1% of subjects die before responses are measured ($\texttt{time} = 0$) and do not contribute any data. Three models are fitted to each dataset:

1. **IEE:** GEE with independent correlation structure (equivalent to ordinary least squares regression). This model should be unbiased for β, as shown in Chapter 6.

2. **GEE:** GEE with exchangeable correlation structure. Chapter 6 shows that this model is not consistent for estimation of DECAL parameter vector β.

3. **LMM:** Linear mixed model fitted to available data. The target of estimation for this maximum likelihood method is μ_{ij}^M, not μ_{ij}^A, so we expect bias in regression parameter estimates.

Percent relative bias in β is computed as for the simulations in Chapter 4.

Table 8.1: Percent bias for generalized estimating equations (IEE and GEE), inverse probability of censoring-weighted generalized estimating equations (IPCW-IEE and IPCW-GEE), linear mixed models (LMM), and marginalized transition models (MTM). Percent bias is based on 1000 simulated data sets with 1000 subjects. For linear data $\beta = (28, -2, -0.25, 0.10)$. For binary data $\beta = (-2.19, 0.50, 0.10, -0.025)$.

	INT. (β_0)	GROUP (β_1)	TIME (β_2)	GROUP:TIME (β_3)
(a) Linear No Dropout				
IEE	0	0	0	0
GEE	0	0	10	-11
LMM	0	0	12	-9
(b) Binary No Dropout				
IEE	0	0	0	-1
GEE	2	4	8	-2
MTM	3	5	12	-5
(c) Binary MCAR-S Dropout				
IPCW-IEE	0	1	1	0
IEE	2	2	-11	-12
IPCW-GEE	2	4	9	-1
MTM	2	4	0	-16

8.2.2 Results: Linear Response

Percent relative bias in estimation of β is summarized for the fitted models in section (a) of Table 8.1. For linear data with death but no dropout, GEE and maximum likelihood regression models fit the correct fixed effects, but do not condition on survival (being alive at the time of measurement). This misspecification yields negligible bias percentages for the intercept and cluster-level covariate `group`. Bias in `time` and `group:time` is of similar magnitude (about 10%) for GEE and linear mixed models. The `group:time` coefficient shows negative bias, toward no difference in time trends for the two groups. Bias in `time`

is toward a more extreme slope for the reference group, women.

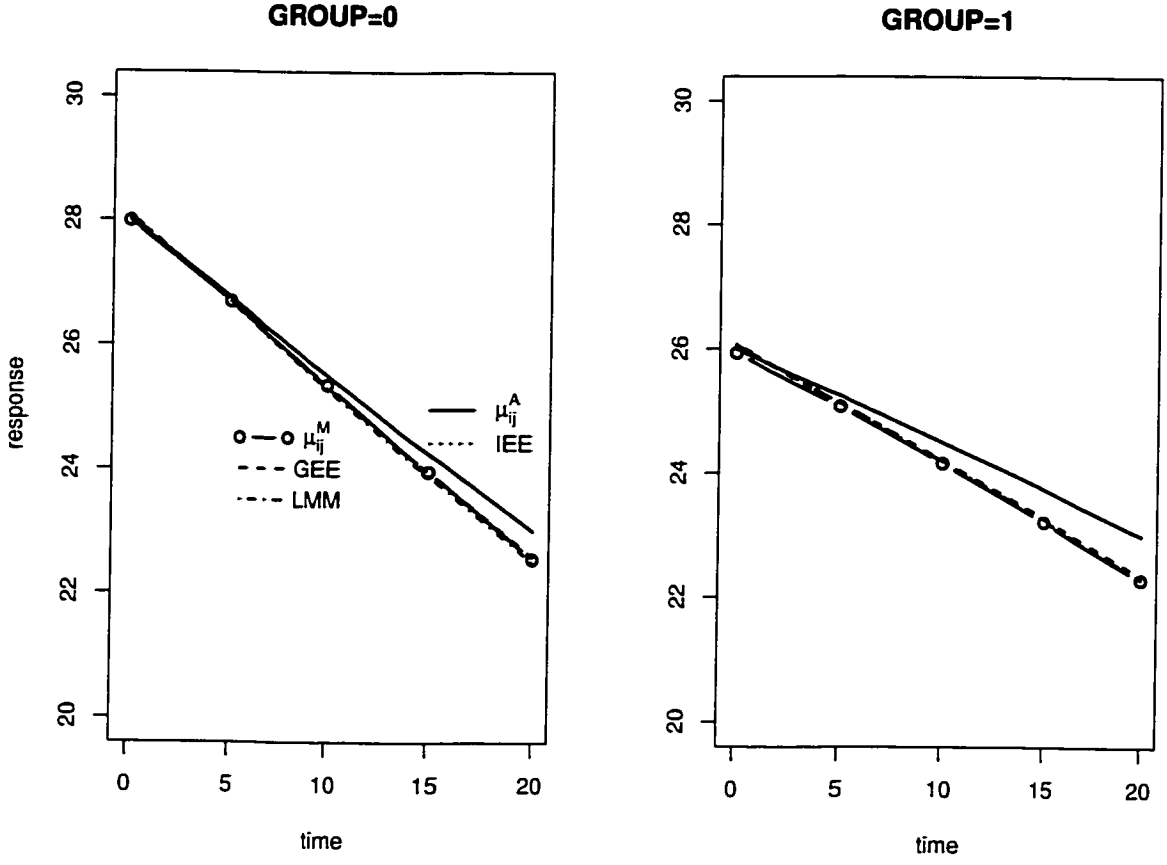


Figure 8.2: Fitted trajectories: Bias in linear regression model conditioning on survival.

The impact of bias in fitted GEE and LMM models is illustrated in Figure 8.2. Regression model trajectories are shown using average β values for 1000 fitted models. IEE fits the DECAL trajectory closely, and is indistinguishable from the plotted μ_{ij}^{A} trajectory. The GEE and LMM trajectories fit marginal model μ_{ij}^{M} instead of DECAL model μ_{ij}^{A}. Mill's ratio in Equation 8.1 is not linear in either `group` or `time`, but μ_{ij}^{M} is approximately linear in `time` and is fitted well by GEE and LMM.

Marginal response values μ_{ij}^{M} are attenuated compared to DECAL model μ_{ij}^{A}, since

data for deceased subjects (who are expected to have lower response values) are implicitly imputed by the marginal model. For men (group = 1), the slope for marginal models is somewhat farther from the true slope than for women. Men are more likely to die in the simulated data, so more implicit imputation lowers the slope.

8.3 Simulation Study: Binary Response

8.3.1 Simulation Details

Binary response data are generated for a hypothetical observational study of disability in activities of daily living for men and women aged 65-85. The DECAL regression model for binary response simulations includes an interaction between group and time:

$$\text{logit}\left[\text{E}(\text{Y}_{ij}|S_i > t_j)\right] = \beta_0 + \beta_1 \cdot \texttt{group}_i + \beta_2 \cdot \texttt{time}_{ij} + \beta_3 \cdot \texttt{group}_i \cdot \texttt{time}_{ij}$$

where $\beta = (-2.19, 0.5, 0.1, -0.025)$. The response trajectory is shown in panel (b) of Figure 8.1. Here women (group = 0) are less likely to show the binary response (disability) among survivors at age 65. By age 85 (time = 20), both men and women are equally likely to be disabled, given that they are alive.

Data are generated using the DECAL regression model and a probit model for the joint distribution of survival time and responses. The binary response Y_{ij} is an indicator for $(\text{Y}^*_{ij} > 0)$ where (Y^*_i, S_i) follow a multivariate normal distribution:

$$\begin{pmatrix} \text{Y}^*_i \\ S_i \end{pmatrix} \sim N\left(\begin{pmatrix} \mu^M_i \\ \text{E}(S_i) \end{pmatrix}, \Sigma_i = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \right) \tag{8.2}$$

for

$a =$ autoregressive correlation matrix for $\text{Y}^*_i, \rho = 0.7$

$b = c = \sigma_{SY}$

$d = 1$ (corresponds to $\text{Var}(S_i) = 5$)

For these simulations, σ_{SY} is -0.4 for women (group = 0), and -0.3 for men (group = 1). The mean survival time is 85 years for women and 80 for men. The standard deviation is 2.2 ($\sqrt{5}$) for both groups.

Marginal mean $\mu_{ij}^M = \mathrm{E}(\mathrm{Y}_{ij}^*)$ is found through the following correspondence:

$$\begin{aligned}\mu_{ij}^A = \mathrm{E}(\mathrm{Y}_{ij}|S_i > t_j) &= P(\mathrm{Y}_{ij}^* > 0|S_i > t_j) \\ &= \frac{P(\mathrm{Y}_{ij}^* > 0, S_i > t_j)}{P(S_i > t_j)}.\end{aligned}$$

Response conditioning on survival (μ_{ij}^A) and the marginal survival distribution for S_i are known, as are all parameters of the joint cumulative density (Equation 8.2) except for marginal mean μ_{ij}^M. We therefore can find μ_{ij}^M by bisection, to an arbitrary tolerance (0.0001). Once the joint distribution $f(\mathrm{Y}_i^*, S_i)$ is specified, S_i and Y_i^* are generated by the function `rmvnorm()` in package `mvtnorm` of R version 1.5.1. Y_i^* is dichotomized to give Y_i. Again, 1000 datasets are simulated with 1000 clusters each. For the first set of simulations, no dropout occurs other than truncation due to simulated death. For the second set of simulations, monotone MCAR-S dropout is simulated according to the following model:

$$\mathrm{logit}\left[P(\mathrm{R}_{ij} = 1|\mathrm{R}_{ij-1} = 1, \mathrm{X}_{it}, S_i)\right] = \phi_0 + \phi_1 \cdot (s_i - \mathtt{time}_{ij})$$

where missingness parameter vector $\phi = (-0.5, 0.15)$. The odds of not dropping out 15 years before death are $\exp(10 \cdot \phi_1) \approx 4.5$ times the odds of not dropping out 5 years before death. No data are missing due to dropout at $\mathtt{time} = 0$, but about 33% of responses (Y_{ij} at times when the subject is alive) are missing due to dropout overall. For the simulation with no dropout, 33% of simulated subjects survive to have 5 responses. 34% have 4 responses, 24% have 3, 8% contribute 2, and 1% survive to give only one response.

For the simulation with no data missing, three models are fitted:

1. **IEE:** GEE with independent correlation structure (a generalized linear model). This model should be unbiased for β, as shown in Chapter 6.

2. **GEE:** GEE with autoregressive correlation structure. Section 6.3.2 shows that this model is not consistent for estimation of DECAL parameter vector β.

3. **MTM:** Second-order marginalized transition model fitted to available data. The target of estimation for this maximum likelihood method is μ_{ij}^M, not μ_{ij}^A, so we expect bias in regression parameter estimates.

For the simulation with MCAR-S dropout, four models are fitted:

1. **IPCW-IEE:** IEE with correctly specified inverse probability of censoring weights. This model should yield consistent estimates of β.

2. **IEE:** GEE with independent correlation structure (a generalized linear model). Although GEE parameter estimates are usually consistent for MCAR dropout, MCAR-S dropout depends on random variable S_i and bias is expected in β (Section 7.4).

3. **IPCW-GEE:** GEE with autoregressive correlation structure and correctly specified inverse probability of censoring weights. Estimates of DECAL parameter vector β are not expected to be consistent.

4. **MTM:** Second-order marginalized transition model fitted to available data.

Percent relative bias in β is computed as for the simulations in Chapter 4.

8.3.2 Results: Binary Response

Section (b) of Table 8.1 shows that when data records are truncated due to death but not dropout, GEE and the MTM show little bias in cluster-level variables (`intercept` and `group`), and some bias (up to 12%) for within-cluster covariates. Figure 8.3 shows trajectories using μ_{ij}^A, μ_{ij}^M, and average β values from fitted models. The IEE model fits the DECAL regression very closely. (The trajectories cannot be differentiated in the figure.) GEE and MTM slopes overestimate the probability of being disabled compared to the slope for regression conditioning on being alive. Unlike the linear model, the GEE and MTM parameters do not follow the marginal (μ_{ij}^M) trajectory.

Section (c) of Table 8.1 shows percent bias in regression parameters for models fitted to binary data with deaths and with dropout that is survival-dependent MCAR (MCAR-S). Figure 8.4 shows regression trajectories using average fitted regression parameters. The IEE model without inverse probability of censoring weights underestimates event probabilities for both groups. People closer to death are more likely to drop out, so by not giving weight to those dropouts, we underestimate the probability of disability in the target population of

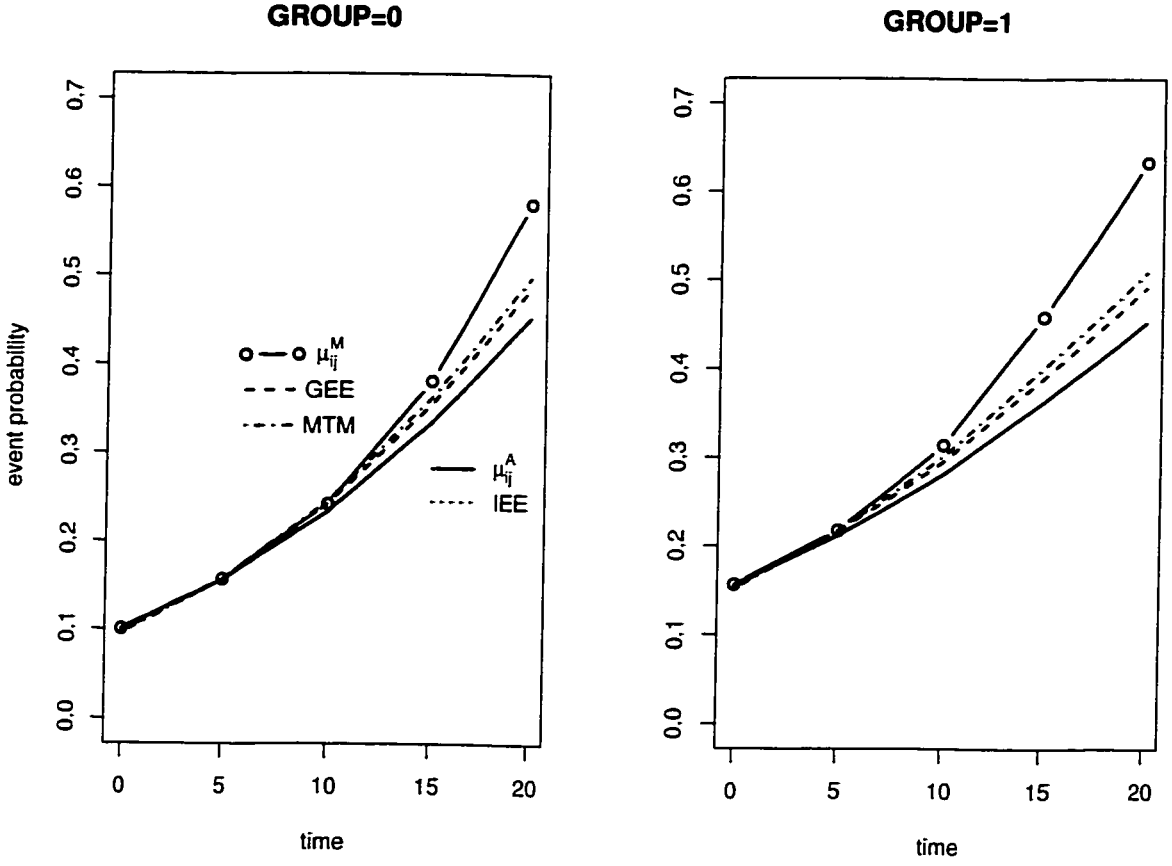


Figure 8.3: Fitted trajectories: Bias in binary regression model conditioning on survival.

survivors (both dropouts and non-dropouts). In contrast, because of implicit imputation, GEE with censoring weights overestimates event probabilities, since undue weight is given to decedents. IPW-GEE fitted values are in between the marginal model μ_{ij}^{M} and the desired target, μ_{ij}^{A}. The MTM, like GEE without IPC weights (not shown), displays 16% negative bias in `group:time` but small ($\leq 5\%$) bias in `time` and other regression parameters. Predicted response probabilities are slightly smaller for women and slightly larger for men, compared to μ_{ij}^{A}.

These three simulation studies demonstrate that the target of estimation can be incor-

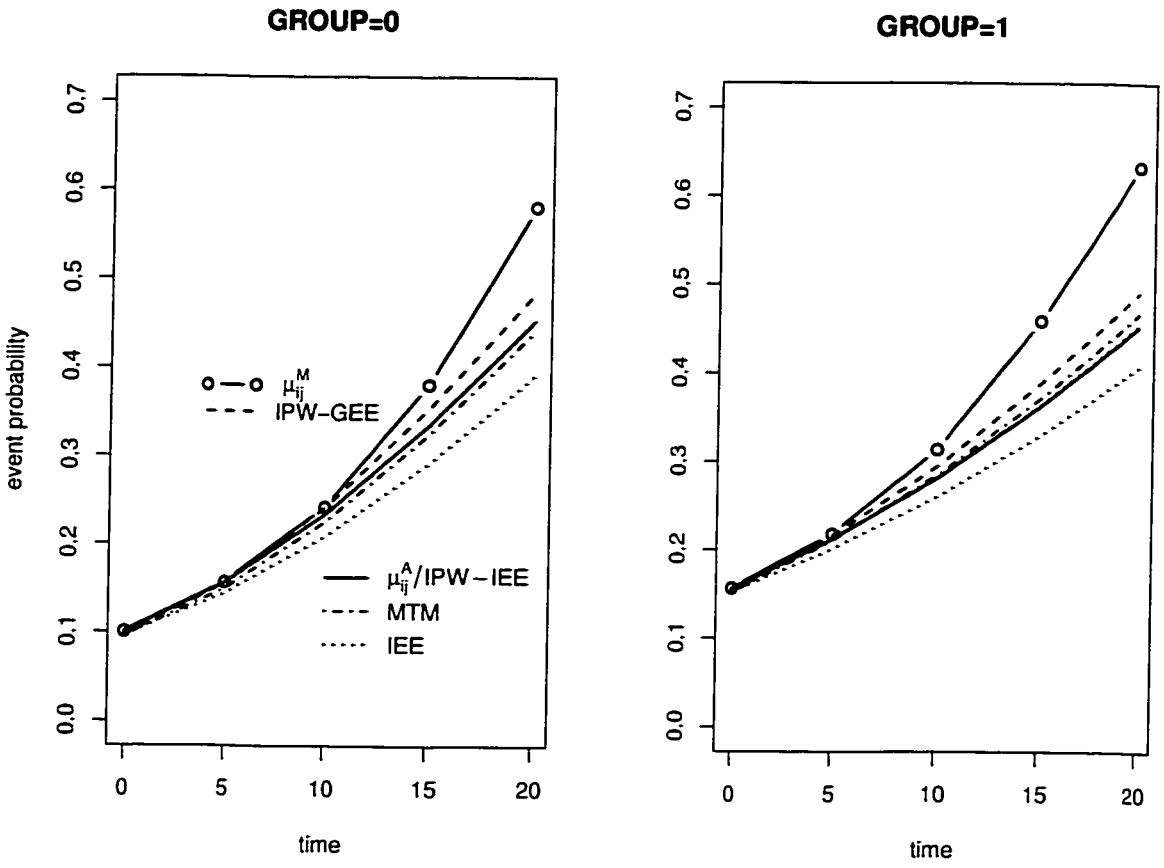


Figure 8.4: Fitted trajectories: Bias in binary regression model with MCAR-S dropout, conditioning on survival.

rect, even when the proper covariates are selected for the regression model. Fitted parameter estimates and fitted values for non-DECAL models differ noticeably from the DECAL model used to generate data. The next section illustrates DECAL and non-DECAL models for the Precipitating Events Project data.

8.4 Illustration: PEP Data

Introduced in Chapter 1, the Precipitating Events Project (PEP) explores monthly patterns of disability in senior citizens. Risk for disability (low, medium, or high) is established by physical and mental status at a baseline interview, and disability in activities of daily living (ADLs) is reported monthly by telephone interviews. This section illustrates the impact of analysis method on targets of inference for regression models. Using 24 months of PEP data, we model a quadratic time trend, with linear effects differing by risk group.

Absence or presence of ADL disability for subject i at month $(j-12)$ is denoted by Y_{ij}. Risk of disability is modeled over time for the three risk groups:

$$\begin{aligned}\text{logit}[P(Y_{ij}=1|X_{ij},S_i>t_j)] &= \beta_0+\beta_1\cdot \texttt{MED}_i+\beta_2\cdot \texttt{HIGH}_i+\beta_3\cdot \texttt{month}_{ij}+\beta_4\cdot \texttt{month}_{ij}^2\\ &\quad +\beta_5\cdot \texttt{MED}_i\cdot \texttt{month}_{ij}+\beta_6\cdot \texttt{HIGH}_i\cdot \texttt{month}_{ij}.\end{aligned}$$

Four approaches are used to fit this regression model:

1. **IEE:** GEE with independent correlation structure. If data are not MCAR, regression may be biased.

2. **IPW-GEE:** Inverse probability of censoring weighted GEE with autoregressive correlation structure. Section 6.3 shows that this model is not expected to be consistent for estimation of DECAL parameter vector β.

3. **MTM:** Second-order marginalized transition model (MTM). Fitted to available data, assumes data are MAR. The target of estimation for this maximum likelihood method is μ_{ij}^M, not μ_{ij}^A, so we expect bias in regression parameter estimates.

4. **IPW-IEE:**Inverse probability of censoring weighted IEE. If the dropout weights and regression model are specified correctly, we expect this model to be unbiased for β.

A small number responses are missing. Of 17,401 anticipated months of response data, 219 are missing due to intermittent dropout, and 224 are missing due to the monotone

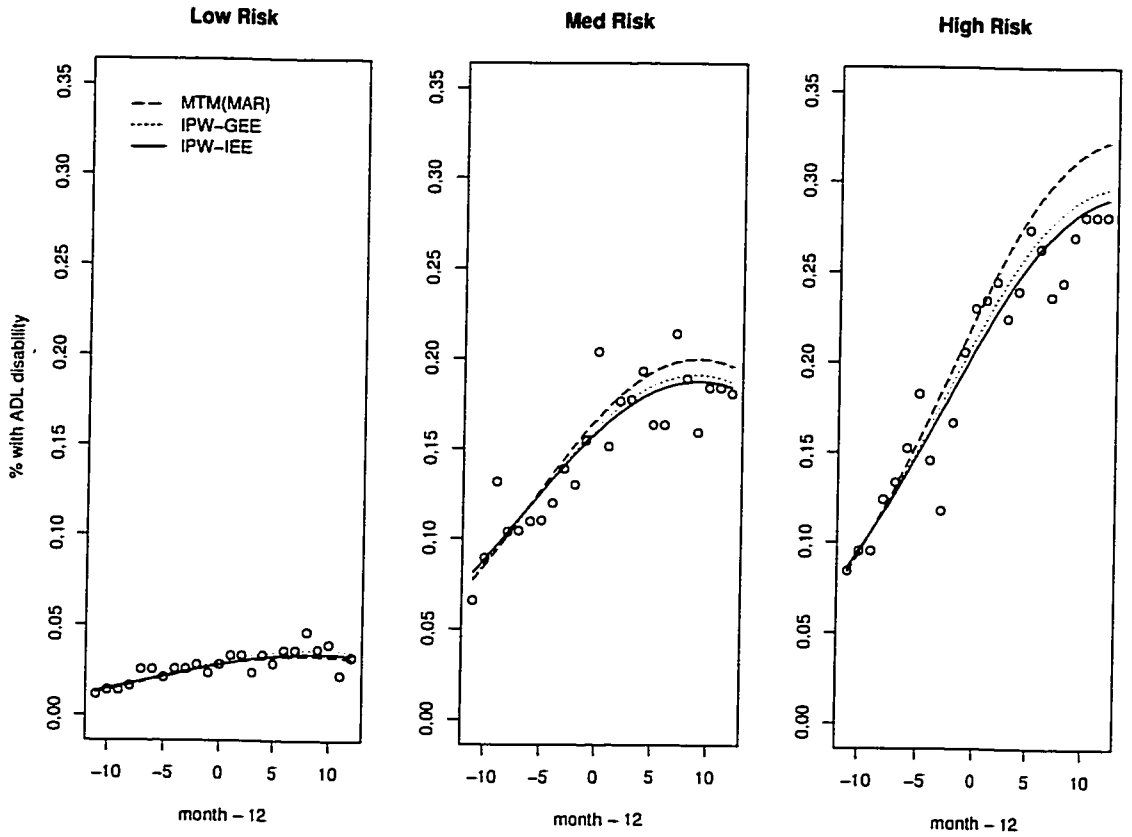


Figure 8.5: Fitted trajectories and cross-sectional means for PEP data (N=754): Different analysis methods yield different targets of inference.

dropout of 19 subjects. Intermittent missing responses are imputed using a nonparametric approximate Bayesian bootstrap (Rubin, 1987). Missing data is matched to nonmissing data by decedent status, risk group, sex, mean number of months disabled, 6-month interval within the followup period, and ADL function in adjacent months (Gill & Kurland, 2003). The selection model at time t_j for the two IPCW-GEE models is MAR with the covariates from the regression model, plus additional main effects: sex, ADL disability status at time t_{j-1}, and an indicator for depression at baseline, from the short version of the CES-D (Kohout et al., 1993). The dropout model is very dense for describing a small amount of missing data: the fitted probability of retention is at least 98% for all data points. However, an inclusive strategy in dropout modeling is advisable since an underspecified dropout model

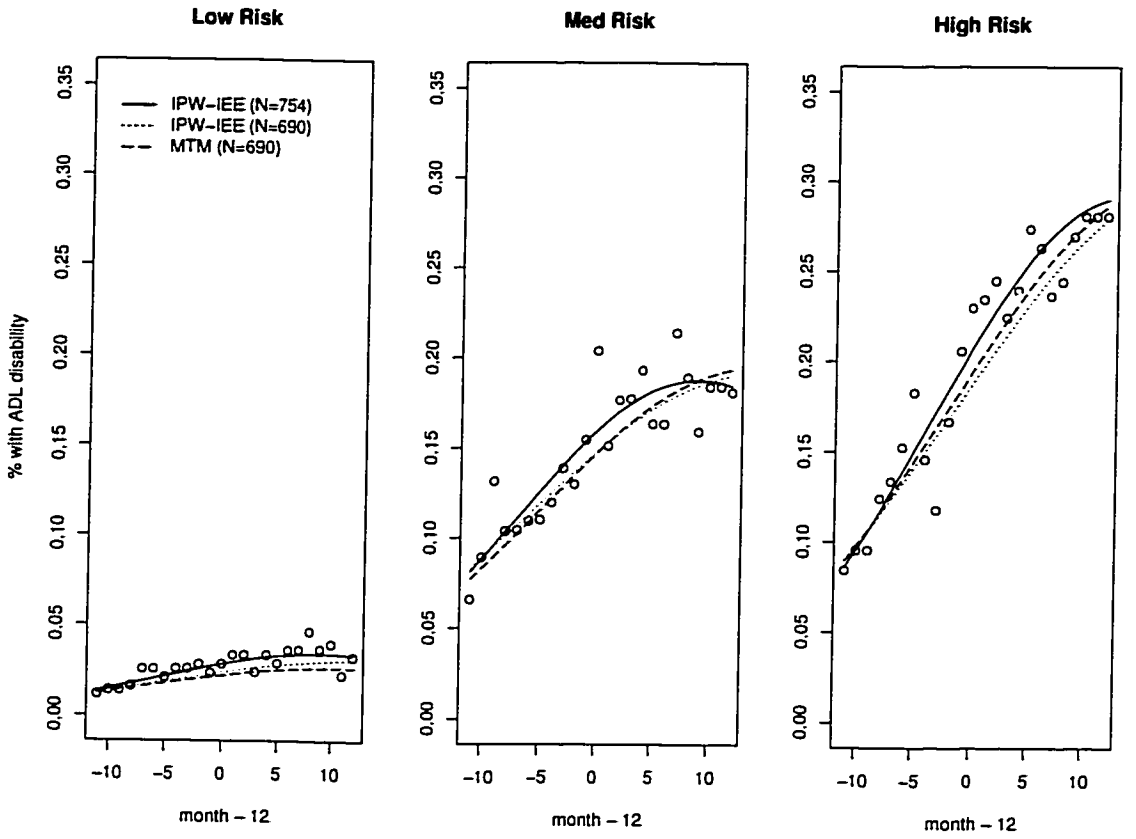


Figure 8.6: Fitted trajectories for PEP data: IPW-IEE and cross-sectional means (N=754) contrasted with analysis only of non-decedents (N=690).

may result in regression model bias (Robins et al., 1995).

Figure 8.5 shows fitted regression trajectories for three models fitted to the PEP data. The IEE model differs only slightly from IPCW-IEE since very little data are missing due to dropout. Cross-sectional means for living subjects are shown for a dataset in which both intermittent missing data and data missing due to dropout are imputed using the single imputation method described above. The IPW-IEE trajectory (solid line) is closer to the raw data than IPW-GEE (dotted line) or the MTM (dashed line). Implicit imputation of data missing due to dropout and death causes the MTM fitted probabilities to be higher than cross-sectional means, especially for subjects at high risk of disability. This implicit imputation is desirable if data are missing due to dropout (see Figure 9.6 in Diggle (1998)),

but show that an incorrect target of inference is modeled by the MTM. Trajectories are extended for deceased subjects, who are presumably more likely to experience ADL disability compared to subjects whose death is not imminent. Therefore, the probability of ADL disability is overestimated by both MTM and IPW-GEE models.

Table 8.2: Fitted models for PEP data (N = 754). Parameter estimates, with sandwich standard errors (IEE, IPCW) or model-based standard errors (MTM) in parentheses.

Model	IEE	IPCW-GEE	MTM(MAR)	IPCW-IEE
Regression Mean				
Intercept (β_0)	-3.535 (0.184)	-3.492 (0.179)	-3.544 (0.125)	-3.536 (0.186)
MED (β_1)	1.864 (0.226)	1.848 (0.221)	1.936 (0.154)	1.874 (0.228)
HIGH (β_2)	2.183 (0.245)	2.183 (0.239)	2.303 (0.181)	2.192 (0.247)
Month (β_3)	0.042 (0.014)	0.046 (0.014)	0.043 (0.015)	0.043 (0.014)
Month^2 (β_4)	-0.002 (0.001)	-0.003 (0.001)	-0.003 (0.001)	-0.002 (0.001)
MED*Month (β_5)	0.000 (0.016)	0.000 (0.016)	0.006 (0.018)	0.001 (0.016)
HIGH*Month (β_6)	0.023 (0.019)	0.022 (0.018)	0.032 (0.021)	0.024 (0.019)
Serial Dependence				
Y_{it-1}			2.967 (0.090)	
Y_{it-2}			1.972 (0.097)	
AR correlation		0.53		

Fitted regression models are shown in tabular form in Table 8.2. The MTM model-based standard errors for risk group parameters (`MED` and `HIGH`) are markedly smaller than the sandwich standard errors for the semiparametric models. This suggests misspecification of the association among responses. The score test for third-order dependence does not suggest that a third-order marginalized transition model is needed ($\chi^2_1 = 1.03, p = 0.31$). The second-order dependence could be misspecified. For example, serial dependence could differ by risk group. However, since the regression mean is biased for DECAL estimation,

the standard errors of the MTM are not of great concern for this analysis.

Model fitting when all decedents are excluded is illustrated in Figure 8.6. The IPW-IEE model and cross-sectional means are the same as in Figure 8.5, but the other two models exclude the 64 subjects who die during the first 24 months of the study. For medium- and high-risk subjects, the trajectories excluding decedents appear closer to the target of inference conditioning on being alive than the trajectories in Figure 8.5. However, since sicker people have presumably been selected out, $P(Y_{ij} = 1)$ is attenuated for all risk groups, most notably the low-risk group.

The impact of dropout and death is unlikely to be large in the PEP dataset, since very few subjects drop out (n=19, 2.5%) and only 64 (8.5%) die in the first two years of the study. Thirty of 432 low-risk group subjects die (7%), and 15/214 (7%) and 19/108 (18%) die in the medium- and high-risk disability groups, respectively. Although inference about parameter estimates does not differ greatly between models, choice of model does have a noticeable impact on regression parameters and fitted values. For example, the fitted odds ratio comparing odds of disability in the high- and low-risk groups (ignoring the nonsignificant group-by-time interaction) is 8.95 under the IPW-IEE model, and 10.00 for the second-order MTM fitted to all available data. In summary, this modest example illustrates that independence estimating equations (IEE) is a promising approach when attempting Direct Estimation Conditioning on being ALive (DECAL).

8.5 Summary

For linear and binary longitudinal data with moderate correlation among responses and survival time, fitted values for $\mu_{ij}^A = \mathrm{E}(Y_{ij}|S_i > t_j)$ may be quite different from the marginal mean $\mu_{ij}^M = \mathrm{E}(Y_{ij})$, which implicitly imputes responses after death. Fitting GEE or maximum likelihood models without accounting for μ_{ij}^A as the target of inference results in both positive and negative bias for within-cluster regression parameters of magnitude in the low teens as a percentage of true β values. The impact of targets of inference is evident in regression models fitted to the PEP data, despite low rates of dropout and death.

Chapter 9

CONCLUSIONS AND FUTURE WORK

The primary contributions of this dissertation are extension of the marginalized transition model (MTM, Heagerty, 2002) to accommodate nonignorable (NI) dropout, and development of a taxonomy to describe direct regression estimation conditioning on being alive. The likelihood-based marginalized transition model has several valuable features for analysis of longitudinal binary data with dropout and no death. The marginal model, conditional (transition) model, and selection (dropout) model are all logit-linear models, so models fitted to observed data can be compared – separately, or as part of the full likelihood. For nonignorable dropout, dependence of the dropout model on unobserved responses can be imposed directly for sensitivity analysis.

For data with missing at random (MAR) dropout, the modeling burden of the marginalized transition model is two logit-linear models: the regression mean and the conditional (transition) mean. Inverse probability of censoring weighted generalized estimating equations, a semiparametric method, also involves two logit-linear models: the regression mean and a dropout model. Simulation studies demonstrate efficiency advantages for using the likelihood-based method, and comparable misspecification bias (or an advantage for the marginalized transition model) when association or dropout models are underspecified.

For binary longitudinal data with both dropout and death, this dissertation identifies expected value of the response conditioning on being alive as a target of inference. Likelihood-based methods implicitly impute beyond the time data records are truncated. This implicit imputation is the reason that MAR dropout is ignorable for likelihood-based methods, but limits the ability of likelihood-based methods to parameterize regression models conditioning on current survival status. Generalized estimating equations (GEE) with nonindependence working correlation also implicitly imputes data beyond time of death, so only independence estimating equations (IEE) is considered for Direct Estimation Condi-

tioning on being ALive (DECAL).

When the regression target is average response for people surviving to report a response, the survival distribution itself is not of interest. Although responses and survival time are assumed to be related, DECAL regression models can generally be fitted using IEE without needing to model the survival distribution. The dropout process is also ignorable if dropout does not depend on responses or the survival distribution. When dropout is related to these observed random variables, inverse probability of censoring weights can be computed for IEE and DECAL regression parameters can be estimated consistently.

If survival time is censored completely at random (if the censoring process is unrelated to responses, conditioning on covariates and survival to the time of censoring), censoring of survival times will not cause DECAL regression to be inconsistent. However, if censoring is not completely at random, the censoring process cannot be modeled for weighted IEE without simplifying assumptions, because the censoring model depends on a missing covariate (survival status at the time of censoring). Finally, the dissertation demonstrated bias in DECAL regression parameters when likelihood-based methods or GEE are fitted without conditioning on being alive.

Future work for the marginalized transition model could include extensions to ordinal responses or transitions in continuous time. Also, the dissertation compares asymptotic bias and efficiency of the MTM and weighted GEE. The small-sample properties of the MTM would also be of interest.

DECAL regression using multiple imputation (MI) is mentioned above as a solution when survival status is censored for dropouts. Multiple imputation could also result in efficiency gains compared to weighted IEE, since with enough imputations the efficiency of MI is comparable to that of likelihood-based methods modeling observed data (Rubin, 1987). Since MI involves modeling the response process, misspecification bias for MI should be considered in balance of efficiency gains over IEE.

When the survival process can be modeled well and inference on point estimates is desirable (such as treatment comparisons at the end of a study), likelihood-based methods may be explored to provide efficiency gains compared to IEE or weighted IEE. However, censoring of survival times is problematic for methods reliant on mixing distributions, even

if censoring is completely at random as defined in Chapter 7.

In conclusion, this dissertation has provided a new method for modeling binary longitudinal data with nonignorable dropout, and has explored target of inference issues surrounding data with both dropout and death.

BIBLIOGRAPHY

ABBOTT, R. D. (1985). Logistic regression in survival analysis. *American Journal of Epidemiology* **121**, 465–471.

AZZALINI, A. (1994). Logistic regression for autocorrelated data with application to repeated measures. *Biometrika* **81**, 767–775.

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

BAKER, S. G. & LAIRD, N. M. (1988). Regression analysis for categorical variables with outcome subject to nonignorable nonresponse. *Journal of the American Statistical Association* **83**, 62–69.

BICKEL, P. J., KLAASSEN, C. A. J., RITOV, Y., & WELLNER, J. (1993). *Efficient and adaptive estimation for semiparametric models.* Johns Hopkins University Press, Baltimore, MD.

BRAYNE, C., SPIEGELHALTER, D. J., DUFOUIL, C., CHI, L. Y., DENING, T. R., PAYKEL, E. S., O'CONNOR, D. W., AHMED, A., MCGEE, M. A., & HUPPERT, F. A. (1999). Estimating the true extent of cognitive decline in the old old. *Journal of the American Geriatrics Society* **47**, 1283–1288.

BRESLOW, N. E. & CLAYTON, D. G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association* **88**, 9–25.

CAREY, V., ZEGER, S. L., & DIGGLE, P. (1993). Modelling multivariate binary data wtih alternating logistic regressions. *Biometrika* **80**, 517–526.

CHOUINARD, G., JONES, B., REMINGTON, G., BLOOM, D., ADDINGTON, D., MACEWAN, G., LABELLE, A., BEAUCLAIR, L., & ARNOTT, W. (1993). A Canadian multicenter placebo-controlled study of fixed doses of risperidone and haloperidol in the treatment of chronic schizophrenia patients. *Journal of Clinical Psychopharmacology* **13**, 25–40.

CLAYTON, D. G. (1996). Generalized linear mixed models. In Gilks, W. R., Richardson, S., & Spiegelhalter, D. J., editors, *Markov Chain Monte Carlo in Practice*, pages 275–301. Chapman and Hall.

COLLINS, L. M., SCHAFER, J. L., & KAM, C.-M. (2002). A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychological Methods* **6**, 330–351.

CONAWAY, M. R. (1993). Non-ignorable non-response models for time-ordered categorical variables. *Applied Statistics* **42**, 105–115.

COX, D. R. & SNELL, E. J. (1989). *Analysis of Binary Data.* Chapman and Hall, London, second edition.

DALE, J. R. (1986). Global cross-ratio models for bivariate, discrete, ordered responses. *Biometrics* **42**, 909–917.

DE GRUTTOLA, V. & TU, X. M. (1994). Modelling progression of CD4-lymphocyte count and its relationship to survival time. *Biometrics* **50**, 1003–1014.

DEMPSTER, A. P., LAIRD, N. M., & RUBIN, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B* **39**, 1–22.

DIEHR, P., PATRICK, D., HEDRICK, S., ROTHMAN, M., GREMBOWSKI, D., RAGHUNATHAN, T. E., & BERESFORD, S. (1995). Including deaths when measuring health status over time. *Medical Care* **33**, AS164–AS172.

DIEHR, P., PATRICK, D. L., SPERTUS, J., KIEFE, C. I., MCDONELL, M., & FIHN, S. D. (2001). Transforming self-rated health and the SF-36 scales to include death and improve interpretability. *Medical Care* **39**, 670–680.

DIGGLE, P. & KENWARD, M. G. (1994). Informative drop-out in longitudinal data analysis (with discussion). *Applied Statistics* **43**, 49–93.

DIGGLE, P. J. (1998). Dealing with missing values in longitudinal studies. In Everitt, B. & Dunn, G., editors, *Statistical Analysis of Medical Data*, pages 203–228. Arnold.

FITZMAURICE, G. M., HEATH, A. F., & CLIFFORD, P. (1996a). Logistic regression models for binary panel data with attrition. *Journal of the Royal Statistical Society, Series A* **159**, 249–263.

FITZMAURICE, G. M. & LAIRD, N. M. (1993). A likelihood-based method for analyzing longitudinal binary responses. *Biometrika* **80**, 141–151.

FITZMAURICE, G. M. & LAIRD, N. M. (2000). Generalized linear mixture models for handling nonignorable dropouts in longitudinal studies. *Biostatistics* **1**, 141–156.

FITZMAURICE, G. M., LAIRD, N. M., & ROTNITZKY, A. G. (1993). Regression models for discrete longitudinal response (with discussion). *Statisical Science* **8**, 284–309.

FITZMAURICE, G. M., LAIRD, N. M., & ZAHNER, G. E. P. (1996b). Multivariate logistic models for incomplete binary responses. *Journal of the American Statistical Association* **91**, 99–108.

FITZMAURICE, G. M., MOLENBERGHS, G., & LIPSITZ, S. R. (1995). Regression models for longitudinal binary responses with informative drop-outs. *Journal of the Royal Statistical Society, Series B* **57**, 691–704.

FLEMING, T. R. & HARRINGTON, D. P. (1991). *Counting processes and survival analysis.* John Wiley & Sons, New York.

FOLSTEIN, M. F., FOLSTEIN, S. E., & MCHUGH, P. R. (1975). "Mini-Mental State": A practical method for grading the cogntive state of patients for the clinician. *Journal of Psychiatric Research* **12**, 189–198.

GEYER, C. J. (1996). Estimation and optimization of functions. In Gilks, W. R., Richardson, S., & Spiegelhalter, D. J., editors, *Markov Chain Monte Carlo in Practice*, pages 241–258. Chapman and Hall.

GILL, T. M., HARDY, S. E., & WILLIAMS, C. S. (2002). Underestimation of disability in community-living older persons. *Journal of the American Geriatrics Society* in press.

GILL, T. M. & KURLAND, B. F. (2003). The burden and patterns of ADL disability among community-living older persons. In press, *Archives of Internal Medicine*.

GILL, T. M., WILLIAMS, C. S., & TINETTI, M. E. (1999). The combined effects of baseline vulnerability and acute hospital events on the development of functional dependence among community-living older persons. *The Journals of Gerontology. Series A, Biological Sciences and Medical Sciences* **54**, M377–M383.

GRAY, S. M. & BROOKMEYER, R. (2000). Multidimensional longitudinal data: Estimating a treatment effect from continuous, discrete, or time-to-event response variables. *Journal of the American Statistical Association* **95**, 396–406.

HEAGERTY, P. J. (2002). Marginalized transition models and likelihood inference for longitudinal categorical data. *Biometrics* **58**, 342–351.

HEAGERTY, P. J. & KURLAND, B. F. (2001). Misspecified maximum likelihood estimates and generalised linear mixed models. *Biometrika* **88**, 973–985.

HEAGERTY, P. J. & ZEGER, S. L. (2000). Marginalized multilevel models and likelihood inference (with discussion). *Statisical Science* **15**, 1–26.

HECKMAN, J. J. (1979). Sample selection bias as a specification error. *Econometrica* **47**, 153–162.

HEITJAN, D. F. & RUBIN, D. B. (1991). Ignorability and coarse data. *Annals of Statistics* **19**, 2244–2253.

HOGAN, J. W. & LAIRD, N. M. (1997). Model-based approaches to analysing incomplete longitudinal and failure time data. *Statistics in Medicine* **16**, 259–272.

HORVITZ, D. G. & THOMPSON, D. J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association* **47**, 663–685.

JOHNSON, N. & KOTZ, S. (1972). *Distributions in Statistics: Continuous Multivariate Distributions.* John Wiley & Sons, New York.

KAY, S. R., FISZBEIN, A., & OPLER, L. A. (1987). The positive and negative syndrome scale (PANSS) for schizophrenia. *Schizophrenia Bulletin* **13**, 261–276.

KENWARD, M. G. & MOLENBERGHS, G. (1999). Parametric models for incomplete continuous and categorical longitudinal data. *Statistical Methods in Medical Research* **8**, 51–83.

KOHOUT, F. J., BERKMAN, L. F., EVANS, D. A., & CORNONI-HUNTLEY, J. (1993). Two shorter forms of the CES-D depression symptoms index. *Journal of Aging and Health* **5**, 179–193.

LAIRD, N. M. (1988). Missing data in longitudinal studies. *Statistics in Medicine* **7**, 305–315.

LEE, Y. J., ELLENBERG, J. H., HIRTZ, D. G., & NELSON, K. B. (1991). Analysis of clinical trials by treatment actually received: Is it really an option? *Statistics in Medicine* **10**, 1595–1605.

LEWIS, H. R. (1957). The data enrichment method. *Operations Research* **5**, 551–554.

LIANG, K.-Y. & ZEGER, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73**, 13–22.

LINDSEY, J. K. (2000). Obtaining marginal estimates from conditional categorical repeated measurements models with missing data. *Statistics in Medicine* **19**, 801–809.

LIPSITZ, S. R., LAIRD, N. M., & HARRINGTON, D. P. (1991). Generalized estimating equations for correlated binary data: Using the odds ratio as a measure of association. *Biometrika* **78**, 153–160.

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

LITTLE, R. J. A. (1995). Modeling the drop-out mechanism in repeated-measures studies. *Journal of the American Statistical Association* **90**, 1112–1121.

LITTLE, R. J. A. & RUBIN, D. B. (1987). *Statistical analysis with missing data.* John Wiley & Sons, New York.

MANCL, L. A. & LEROUX, B. G. (1996). Efficiency of regression estimates for clustered data. *Biometrics* **52**, 500–511.

MARDER, S. R. & MEIBACH, R. C. (1994). Risperidone in the treatment of schizophrenia. *The American Journal of Psychiatry* **151**, 825–835.

MCHORNEY, C. A. (1996). Measuring and monitoring general health status in elderly persons: practical and methodological issues in using the SF-36 health survey. *The Gerontologist* **36**, 571–583.

MILLER, M. E., TEN HAVE, T. R., REBOUSSIN, B. A., LOHMAN, K. K., & REJESKI, W. J. (2001). A marginal model for analyzing discrete outcomes from longitudinal surveys

with outcomes subject to multiple-cause nonresponse. *Journal of the American Statistical Association* **96**, 844–857.

MOLENBERGHS, G., BIJNENS, L., & SHAW, D. (1997a). Linear mixed models and missing data. In Verbeke, G. & Molenberghs, G., editors, *Linear Mixed Models in Practice*, pages 191–274. Springer-Verlag, New York.

MOLENBERGHS, G., KENWARD, M. G., & LESAFFRE, E. (1997b). The analysis of longitudinal ordinal data with nonrandom drop-out. *Biometrika* **84**, 33–44.

MOLENBERGHS, G. & LESAFFRE, E. (1994). Marginal modeling of correlated ordinal data using a multivariate Plackett distribution. *Journal of the American Statistical Association* **89**, 633–644.

PAIK, M. C. (1997). The generalized estimating equation approach when data are not missing completely at random. *Journal of the American Statistical Association* **92**, 1320–1329.

PARK, T. (1998). An approach to categorical data with nonignorable nonresponse. *Biometrics* **54**, 1579–1590.

PREISSER, J. S., GALECKI, A. T., LOHMAN, K. K., & WAGENKNECHT, L. E. (2000). Analysis of smoking trends with incomplete longitudinal binary responses. *Journal of the American Statistical Association* **95**, 1021–1031.

PRESS, W. H., TEUKOLSKY, S. A., VETTERLING, W. T., & FLANNERY, B. P. (1992). *Numerical Recipes in C: The Art of Scientific Computing.* Cambridge University Press, Cambridge, UK, second edition.

REVICKI, D. A., GOLD, K., BUCKMAN, D., CHAN, K., KALLICH, J. D., & WOOLLEY, J. M. (2001). Imputing physical health status scores missing owing to mortality. *Medical Care* **39**, 61–71.

RIBAUDO, H. J., THOMPSON, S. G., & ALLEN-MERSH, T. G. (2000). A joint analysis of quality of life and survival using a random effect selection model. *Statistics in Medicine* **19**, 3237–3250.

ROBINS, J. M. & ROTNITZKY, A. (1995). Semiparametric efficiency in multivariate regression models with missing data. *Journal of the American Statistical Association* **90**, 122–129.

ROBINS, J. M., ROTNITZKY, A., & ZHAO, L. P. (1995). Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *Journal of the American Statistical Association* **90**, 106–121.

ROBINS, J. M. & WANG, N. (1998). Discussion on the papers by Forster and Smith and Clayton *et al.* *Journal of the Royal Statistical Society, Series B* **60**, 91–93.

ROTNITZKY, A. (1998). Efficiency and efficient estimators. In Armitage, P. & Colton, T., editors, *Encyclopedia of Biostatistics*, pages 1286–1292. John Wiley & Sons, New York.

ROTNITZKY, A. & ROBINS, J. M. (1995). Semiparametric regression estimation in the presence of dependent censoring. *Biometrika* **82**, 805–820.

ROTNITZKY, A., ROBINS, J. M., & SCHARFSTEIN, D. O. (1998). Semiparametric regression for repeated outcomes with nonignorable nonresponse. *Journal of the American Statistical Association* **93**, 1321–1339.

ROTNITZKY, A. & WYPIJ, D. (1994). A note on the bias of estimators with missing data. *Biometrics* **50**, 1163–1170.

RUBIN, D. B. (1987). *Multiple imputation for nonresponse in surveys.* John Wiley & Sons, New York.

SCHARFSTEIN, D. O., ROTNITZKY, A., & ROBINS, J. M. (1999). Adjusting for nonignorable drop-out using semiparametric nonresponse models (with discussion). *Journal of the American Statistical Association* **94**, 1096–1146.

TROXEL, A. B. (1998). A comparative analysis of quality of life data from a Southwest Oncology Group randomized trial of advanced colorectal cancer. *Statistics in Medicine* **17**, 767–779.

TROXEL, A. B., FAIRCLOUGH, D. L., CURRAN, D., & HAHN, E. A. (1998). Statistical analysis of quality of life with missing data in cancer clinical trials. *Statistics in Medicine* **17**, 653–666.

TROXEL, A. B., LIPSITZ, S. R., & BRENNAN, T. A. (1997). Weighted estimating equations with nonignorably missing response data. *Biometrics* **53**, 857–869.

WHITE, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica* **50**, 1–26.

WILLETT, J. B. & SINGER, J. D. (1993). Investigating onset, cessation, relapse, and recovery: Why you should, and how you can, use discrete-time survival analysis to examine event occurrence. *Journal of Consulting and Clinical Psychology* **61**, 952–965.

ZEGER, S. L., LIANG, K.-Y., & ALBERT, P. S. (1988). Models for longitudinal data: A generalized estimating equations approach. *Biometrics* **44**, 1049–1060.

ZIEGLER, A. & KASTNER, C. (2000). The effect of misspecified response probabilities on parameter estimates from weighted estimating equations. In *München, SFB 386*, number 200.

Appendix A

DERIVATION OF SCORE EQUATIONS FOR FIRST-ORDER MARGINALIZED TRANSITION MODEL WITH IGNORABLE AND NONIGNORABLE MONOTONE DROPOUT

Setting: binary outcome, clustered data

Notation and Model:

i=1,...,N (clusters)

t=1,...,n_i (cluster size)

$j = t - 1$

A.1 Likelihood under MAR Missingness

$$\begin{aligned} \boldsymbol{L}_i(\beta,\alpha) &= P(\mathrm{Y}_{i1} = y_{i1}, \mathrm{Y}_{i2} = y_{i2}, \cdots, \mathrm{Y}_{in_i} = y_{in_i}) \\ &= P(\mathrm{Y}_{i1} = y_{i1})P(\mathrm{Y}_{i2} = y_{i2}|\mathrm{Y}_{i1} = y_{i1}) \cdots P(\mathrm{Y}_{in_i} = y_{in_i}|\mathrm{Y}_{in_i-1} = y_{in_i-1}) \\ &\quad \text{(since dependence is first-order)} \end{aligned}$$

where:

$P(\mathrm{Y}_{i1} = y_{i1}) = (\mu_{i1}^M)^{y_{i1}}(1 - \mu_{i1}^M)^{(1-y_{i1})}$

$\mu_{it}^M = P(\mathrm{Y}_{it} = 1)$

$\mathrm{logit}(\mu_{it}^M) = \eta_{it}^M = \mathrm{X}_{it}\beta$

$P(\mathrm{Y}_{it} = y_{it}|\mathrm{Y}_{ij} = y_{ij}) = (\mu_{it}^C)^{y_{it}}(1 - \mu_{it}^C)^{(1-y_{it})}$

$\mu_{it}^C = P(\mathrm{Y}_{it} = 1|\mathrm{Y}_{ij} = y_{ij})$

$\mathrm{logit}(\mu_{it}^C) = \eta_{it}^C = y_{ij}\log(\Psi_t) + \log\frac{\mu_{it}^M - \nu_{itj}}{1-\mu_{it}^M-\mu_{ij}^M+\nu_{itj}}$ (Carey, Zeger, Diggle '93, *Biometrika*, p. 520)

$$\begin{aligned}\nu_{itj} &= P(\mathrm{Y}_{it}=1,\mathrm{Y}_{ij}=1)\\ &= \frac{1-(\mu_{it}^M+\mu_{ij}^M)(1-\Psi_t)-\left[(1-(1-\Psi_t)(\mu_{it}^M+\mu_{ij}^M))^2+4\Psi_t\mu_{it}^M\mu_{ij}^M(1-\Psi_t)\right]^{\frac{1}{2}}}{2(\Psi_t-1)}\end{aligned}$$

(Diggle, Liang, & Zeger '94, p. 150)

$$\Psi_t = \Psi_{itj} = \frac{P(\mathrm{Y}_{it}=1,\mathrm{Y}_{ij}=1)P(\mathrm{Y}_{it}=0,\mathrm{Y}_{ij}=0)}{P(\mathrm{Y}_{it}=1,\mathrm{Y}_{ij}=0)P(\mathrm{Y}_{it}=0,\mathrm{Y}_{ij}=1)}$$

$$\log(\Psi_t) = \mathrm{Z}_{it}\alpha$$

Returning to the likelihood:

$$\begin{aligned}\boldsymbol{L}(\beta,\alpha) &= \prod_{i=1}^{N}(\mu_{i1}^M)^{y_{i1}}(1-\mu_{i1}^M)^{(1-y_{i1})}\prod_{i=1}^{N}\prod_{t=2}^{n_i}(\mu_{it}^C)^{y_{it}}(1-\mu_{it}^C)^{(1-y_{it})}\\ \log(\boldsymbol{L}(\beta,\alpha)) &= \sum_{i=1}^{N} y_{i1}\log(\mu_{i1}^M)+(1-y_{i1})\log(1-\mu_{i1}^M)\\ &\qquad +\sum_{i=1}^{N}\sum_{t=2}^{n_i} y_{it}\log(\mu_{it}^C)+(1-y_{it})\log(1-\mu_{it}^C)\\ &= \sum_{i=1}^{N}\underbrace{y_{i1}\mathrm{logit}(\mu_{i1}^M)+\log(1-\mu_{i1}^M)}_{\log\boldsymbol{L}1i}+\sum_{i=1}^{N}\sum_{t=2}^{n_i}\underbrace{y_{it}\mathrm{logit}(\mu_{it}^C)+\log(1-\mu_{it}^C)}_{\log\boldsymbol{L}2i}\end{aligned}$$

$$\begin{aligned}\frac{\partial\log\boldsymbol{L}1i}{\partial\beta} &= \frac{\partial\log\boldsymbol{L}1i}{\partial\eta_{i1}}\frac{\partial\eta_{i1}}{\partial\beta}\\ &= (y_{i1}-\mu_{i1}^M)\mathrm{X}_{i1}\end{aligned}$$

$$\frac{\partial\log\boldsymbol{L}2i}{\partial\beta} = \frac{\partial\log\boldsymbol{L}2i}{\partial\mu_{it}^C}\frac{\partial\mu_{it}^C}{\partial\eta_{it}^C}\frac{\partial\eta_{it}^C}{\partial\mu_{it}^M}\frac{\partial\mu_{it}^M}{\partial\beta}+\frac{\partial\log\boldsymbol{L}2i}{\partial\mu_{it}^C}\frac{\partial\mu_{it}^C}{\partial\eta_{it}^C}\frac{\partial\eta_{it}^C}{\partial\mu_{ij}^M}\frac{\partial\mu_{ij}^M}{\partial\beta}$$

where

$$\begin{aligned}\frac{\partial\log\boldsymbol{L}2}{\partial\mu_{it}^C}\frac{\partial\mu_{it}^C}{\partial\eta_{it}^C} &= \frac{(y_{it}-\mu_{it}^C)}{\mu_{it}^C(1-\mu_{it}^C)}\mu_{it}^C(1-\mu_{it}^C) = y_{it}-\mu_{it}^C\\ \frac{\partial\eta_{it}^C}{\partial\mu_{it}^M} &= \frac{\partial}{\partial\mu_{it}^M}\left[y_{ij}\log(\Psi_t)+\log\frac{\mu_{it}^M-\nu_{itj}}{1-\mu_{it}^M-\mu_{ij}^M+\nu_{itj}}\right]\end{aligned}$$

$$
\begin{aligned}
&= \frac{(1-\frac{\partial \nu_{itj}}{\partial \mu_{it}^M})(1-\mu_{ij}^M)}{(\mu_{it}^M-\nu_{itj})(1-\mu_{it}^M-\mu_{ij}^M+\nu_{itj})} \\
\frac{\partial \mu_{it}^M}{\partial \beta} &= \mu_{it}^M(1-\mu_{it}^M)\mathrm{X}_{it} \\
\frac{\partial \eta_{it}^C}{\partial \mu_{ij}^M} &= \frac{\mu_{it}^M-\nu_{itj}-\frac{\partial \nu_{itj}}{\partial \mu_{ij}^M}(1-\mu_{ij}^M)}{(\mu_{it}^M-\nu_{itj})(1-\mu_{it}^M-\mu_{ij}^M+\nu_{itj})} \\
\frac{\partial \mu_{ij}^M}{\partial \beta} &= \mu_{ij}^M(1-\mu_{ij}^M)\mathrm{X}_{ij}
\end{aligned}
$$

Now for some messy parts:

$$
\begin{aligned}
\frac{\partial \nu_{tj}}{\partial \mu_{it}^M} &= \frac{\partial}{\partial \mu_{it}^M}\frac{1-(\mu_{it}^M+\mu_{ij}^M)(1-\Psi_t)-\left[(1-(1-\Psi_t)(\mu_{it}^M+\mu_{ij}^M))^2-4\Psi_t\mu_{it}^M\mu_{ij}^M(\Psi_t-1)\right]^{1/2}}{2(\Psi_t-1)} \\
&= \frac{1}{2}+\frac{1}{2}a_{tj}^{-1/2}\left[-1+\mu_{it}^M+\mu_{ij}^M+\Psi_t(\mu_{ij}^M-\mu_{it}^M)\right] \\
&\quad \text{where } a_{tj}=\left[(1-(1-\Psi_t)(\mu_{it}^M+\mu_{ij}^M))^2+4\Psi_t\mu_{it}^M\mu_{ij}^M(1-\Psi_t)\right] \\
\frac{\partial \nu_{tj}}{\partial \mu_{ij}^M} &= \frac{\partial}{\partial \mu_{ij}^M}\frac{1-(\mu_{it}^M+\mu_{ij}^M)(1-\Psi_t)-\left[(1-(1-\Psi_t)(\mu_{it}^M+\mu_{ij}^M))^2-4\Psi_t\mu_{it}^M\mu_{ij}^M(\Psi_t-1)\right]^{1/2}}{2(\Psi_t-1)} \\
&= \frac{1}{2}+\frac{1}{2}a_{tj}^{-1/2}\left[-1+\mu_{it}^M+\mu_{ij}^M+\Psi_t(\mu_{it}^M-\mu_{ij}^M)\right]
\end{aligned}
$$

The full score equation for beta:

$$
\begin{aligned}
\frac{\partial \log(\boldsymbol{L}(\beta,\alpha))}{\partial \beta} &= \sum_{i=1}^{N}(y_{i1}-\mu_{i1}^M)\mathrm{X}_{i1}+\sum_{i=1}^{N}\sum_{t=2}^{n_i}(y_{it}-\mu_{it}^C)\frac{\partial \eta_{it}^C}{\partial \mu_{it}^M}\frac{\partial \mu_{it}^M}{\partial \beta} \\
&\quad +\sum_{i=1}^{N}\sum_{t=2}^{n_i}(y_{it}-\mu_{it}^C)\frac{\partial \eta_{it}^C}{\partial \mu_{ij}^M}\frac{\partial \mu_{ij}^M}{\partial \beta}
\end{aligned}
$$

Now for the Hessian (used for Newton-Raphson estimation and standard errors)...

$$
\begin{aligned}
\frac{\partial^2 \log(\boldsymbol{L}(\beta,\alpha))}{\partial \beta^2} &= -\sum_{i=1}^{N}\frac{\partial \mu_{i1}^M}{\partial \beta}\mathrm{X}_{i1}-\sum_{i=1}^{N}\sum_{t=2}^{n_i}\left(\frac{\partial \mu_{it}^C}{\partial \eta_{it}^C}\frac{\partial \eta_{it}^C}{\partial \mu_{it}^M}\frac{\partial \mu_{it}^M}{\partial \beta}+\frac{\partial \mu_{it}^C}{\partial \eta_{it}^C}\frac{\partial \eta_{it}^C}{\partial \mu_{ij}^M}\frac{\partial \mu_{ij}^M}{\partial \beta}\right)\left[\frac{\partial \eta_{it}^C}{\partial \mu_{it}^M}\frac{\partial \mu_{it}^M}{\partial \beta}\right] \\
&\quad +\sum_{i=1}^{N}\sum_{t=2}^{n_i}\left(\frac{\partial^2 \eta_{it}^C}{\partial (\mu_{it}^M)^2}\frac{\partial \mu_{it}^M}{\partial \beta}+\frac{\partial^2 \eta_{it}^C}{\partial \mu_{it}^M\partial \mu_{ij}^M}\frac{\partial \mu_{ij}^M}{\partial \beta}\right)\left[(y_{it}-\mu_{it}^C)\frac{\partial \mu_{it}^M}{\partial \beta}\right]
\end{aligned}
$$

$$
\begin{aligned}
&+\sum_{i=1}^{N}\sum_{t=2}^{n_i}\left(\frac{\partial\mu_{it}^M}{\partial\beta}(1-2\mu_{it}^M)*\mathrm{X}_{it}\right)\left[(y_{it}-\mu_{it}^C)\frac{\partial\eta_{it}^C}{\partial\mu_{it}^M}\right]\\
&-\sum_{i=1}^{N}\sum_{t=2}^{n_i}\left(\frac{\partial\mu_{it}^C}{\partial\eta_{it}^C}\frac{\partial\eta_{it}^C}{\partial\mu_{it}^M}\frac{\partial\mu_{it}^M}{\partial\beta}+\frac{\partial\mu_{it}^C}{\partial\eta_{it}^C}\frac{\partial\eta_{it}^C}{\partial\mu_{ij}^M}\frac{\partial\mu_{ij}^M}{\partial\beta}\right)\left[\frac{\partial\eta_{it}^C}{\partial\mu_{ij}^M}\frac{\partial\mu_{ij}^M}{\partial\beta}\right]\\
&+\sum_{i=1}^{N}\sum_{t=2}^{n_i}\left(\frac{\partial^2\eta_{it}^C}{\partial\mu_{ij}^M\partial\mu_{it}^M}\frac{\partial\mu_{it}^M}{\partial\beta}+\frac{\partial^2\eta_{it}^C}{\partial(\mu_{ij}^M)^2}\frac{\partial\mu_{ij}^M}{\partial\beta}\right)\left[(y_{it}-\mu_{it}^C)\frac{\partial\mu_{ij}^M}{\partial\beta}\right]\\
&+\sum_{i=1}^{N}\sum_{t=2}^{n_i}\left(\frac{\partial\mu_{ij}^M}{\partial\beta}(1-2\mu_{ij}^M)*\mathrm{X}_{ij}\right)\left[(y_{it}-\mu_{it}^C)\frac{\partial\eta_{it}^C}{\partial\mu_{ij}^M}\right]
\end{aligned}
$$

where

$$
\begin{aligned}
\frac{\partial^2\eta_{it}^C}{\partial(\mu_{it}^M)^2} &= \frac{-\frac{\partial^2\nu_{itj}}{\partial(\mu_{it}^M)^2}(1-\mu_{ij}^M)-\frac{\partial\eta_{it}^C}{\partial\mu_{it}^M}(1-\frac{\partial\nu_{itj}}{\partial\mu_{it}^M})(1-2\mu_{it}^M-\mu_{ij}^M+2\nu_{itj})}{(\mu_{it}^M-\nu_{itj})(1-\mu_{it}^M-\mu_{ij}^M+\nu_{itj})}\\
\frac{\partial^2\eta_{it}^C}{\partial\mu_{it}^M\partial\mu_{ij}^M} &= \frac{\frac{\partial^2\nu_{itj}}{\partial\mu_{it}^M\partial\mu_{ij}^M}(1-\mu_{ij}^M)-1+\frac{\partial\nu_{itj}}{\partial\mu_{it}^M}+\frac{\partial\eta_{it}^C}{\partial\mu_{it}^M}\left[\frac{\partial\nu_{itj}}{\partial\mu_{ij}^M}(1-2\mu_{it}^M-\mu_{ij}^M+2\nu_{itj})+\mu_{it}^M-\nu_{itj}\right]}{(\mu_{it}^M-\nu_{itj})(1-\mu_{it}^M-\mu_{ij}^M+\nu_{itj})}\\
\frac{\partial^2\eta_{it}^C}{\partial(\mu_{ij}^M)^2} &= \frac{-\frac{\partial^2\nu_{itj}}{\partial(\mu_{ij}^M)^2}(1-\mu_{ij}^M)+\frac{\partial\eta_{it}^C}{\partial\mu_{ij}^M}\left[\frac{\partial\nu_{itj}}{\partial\mu_{ij}^M}(1-2\mu_{it}^M-\mu_{ij}^M+2\nu_{itj})+\mu_{it}^M-\nu_{itj}\right]}{(\mu_{it}^M-\nu_{itj})(1-\mu_{it}^M-\mu_{ij}^M+\nu_{itj})}
\end{aligned}
$$

and

$$
\begin{aligned}
\frac{\partial^2\nu_{itj}}{\partial(\mu_{it}^M)^2} &= \frac{1}{2}a_{tj}^{-1/2}(1-\Psi_t)\left[1-a_{tj}^{-1}\left(-1+\mu_{it}^M+\mu_{ij}^M+\Psi_t(\mu_{ij}^M-\mu_{it}^M)\right)^2\right]\\
\frac{\partial^2\nu_{itj}}{\partial(\mu_{ij}^M)^2} &= \frac{1}{2}a_{tj}^{-1/2}(1-\Psi_t)\left[1-a_{tj}^{-1}\left(-1+\mu_{it}^M+\mu_{ij}^M+\Psi_t(\mu_{it}^M-\mu_{ij}^M)\right)^2\right]\\
\frac{\partial^2\nu_{itj}}{\partial\mu_{it}^M\partial\mu_{ij}^M} &= \frac{1}{2}a_{tj}^{-1/2}\left[1+\Psi_t-a_{tj}^{-1}(1-\Psi_t)\left[(-1+\mu_{it}^M+\mu_{ij}^M)^2-\Psi_t^2(\mu_{it}^M-\mu_{ij}^M)^2\right]\right]
\end{aligned}
$$

That was a submatrix of the Hessian involving only β. Now find the score for α and the corresponding submatrix of the Hessian. From before,

$$
\log(\boldsymbol{L}(\beta,\alpha)) = \underbrace{\sum_{i=1}^{N} y_{i1}\mathrm{logit}(\mu_{i1}^M)+\log(1-\mu_{i1}^M)}_{\text{doesn't depend on }\Psi}+\underbrace{\sum_{i=1}^{N}\sum_{t=2}^{n_i} y_{it}\mathrm{logit}(\mu_{it}^C)+\log(1-\mu_{it}^C)}_{\log\boldsymbol{L}_2}
$$

$$\begin{aligned}
\frac{\partial \log \boldsymbol{L}2}{\partial \alpha} &= \frac{\partial \log \boldsymbol{L}2}{\partial \eta_{it}^C}\frac{\partial \eta_{it}^C}{\partial \Psi_{it}}\frac{\partial \Psi_{it}}{\partial \alpha} \\
\frac{\partial \log \boldsymbol{L}2}{\partial \eta_{it}^C} &= (y_{it} - \mu_{it}^C) \\
\frac{\partial \Psi_{it}}{\partial \alpha} &= \Psi_t * \mathrm{Z}_{it}
\end{aligned}$$

The difficult element is $\frac{\partial \eta_{it}^C}{\partial \Psi_{it}}$

$$\begin{aligned}
\frac{\partial \eta_{it}^C}{\partial \Psi_{it}} &= \frac{\partial}{\partial \Psi_t}\left[y_{ij}\log\Psi_t + \log\frac{\mu_{it}^M - \nu_{tj}}{1 - \mu_{it}^M - \mu_{ij}^M + \nu_{tj}} \right] \\
&= \frac{y_{ij}}{\Psi_t} - \frac{\frac{\partial \nu_{tj}}{\partial \Psi_t}(1 - \mu_{ij}^M)}{(\mu_{it}^M - \nu_{tj})(1 - \mu_{it}^M - \mu_{ij}^M + \nu_{tj})}
\end{aligned}$$

where

$$\begin{aligned}
a_{tj} &= \left[(1 - (1 - \Psi_t)(\mu_{it}^M + \mu_{ij}^M))^2 + 4\Psi_t \mu_{it}^M \mu_{ij}^M (1 - \Psi_t)\right] \\
\frac{\partial \nu_{tj}}{\partial \Psi_t} &= \frac{\mu_{it}^M + \mu_{ij}^M - \frac{1}{2}a^{-1/2}\frac{\partial a_{tj}}{\partial \Psi_t} - 2\nu_{tj}}{2(\Psi_t - 1)} \\
\frac{\partial a_{tj}}{\partial \Psi_t} &= 2(\mu_{it}^M + \mu_{ij}^M) - 2(\mu_{it}^M + \mu_{ij}^M)^2(1 - \Psi_t) + 4\mu_{it}^M \mu_{ij}^M (1 - 2\Psi_t)
\end{aligned}$$

So the full score for alpha is

$$\begin{aligned}
\frac{\partial \log(\boldsymbol{L}(\beta, \alpha))}{\partial \alpha} &= \sum_{i=1}^{N}\sum_{t=2}^{n_i}(y_{it} - \mu_{it}^C)\Psi_t \left[\frac{y_{ij}}{\Psi_t} - \frac{\frac{\partial \nu_{tj}}{\partial \Psi_t}(1 - \mu_{ij}^M)}{(\mu_{it}^M - \nu_{tj})(1 - \mu_{it}^M - \mu_{ij}^M + \nu_{tj})} \right] * \mathrm{Z}_{it} \\
&= \sum_{i=1}^{N}\sum_{t=2}^{n_i}(y_{it} - \mu_{it}^C)\frac{\partial \eta_{it}^C}{\partial \Psi_{it}}\frac{\partial \Psi_{it}}{\partial \alpha}
\end{aligned}$$

On to the Hessian for alpha

$$\begin{aligned}
\frac{\partial^2 \log(\boldsymbol{L}(\beta, \alpha))}{\partial \alpha^2} &= \left(-\frac{\partial \mu_{it}^C}{\partial \eta_{it}^C}\frac{\partial \eta_{it}^C}{\partial \Psi_t}\frac{\partial \Psi_t}{\partial \alpha} \right)\left[\frac{\partial \eta_{it}^C}{\partial \Psi_{it}}\frac{\partial \Psi_{it}}{\partial \alpha} \right] + \frac{\partial \Psi_t}{\partial \alpha}\left[(y_{it} - \mu_{it}^C)\frac{\partial \eta_{it}^C}{\partial \Psi_{it}} * \mathrm{Z}_{it} \right] \\
&\quad + \left(\frac{\partial^2 \eta_{it}^C}{\partial \Psi_{it}^2}\frac{\partial \Psi_t}{\partial \alpha} \right)\left[(y_{it} - \mu_{it}^C)\frac{\partial \Psi_{it}}{\partial \alpha} \right]
\end{aligned}$$

where

$$\frac{\partial^2 \eta_{it}^C}{\partial \Psi_{it}^2} = -\frac{y_{ij}}{\Psi_t^2} + \frac{-\frac{\partial^2 \nu_{tj}}{\partial \Psi_t^2}(1-\mu_{ij}^M) + \left(\frac{\partial \eta_{it}^C}{\partial \Psi_{it}} - \frac{y_{ij}}{\Psi_t}\right)\left[\frac{\partial \nu_{tj}}{\partial \Psi_t}(1-2\mu_{it}^M - \mu_{ij}^M + 2\nu_{tj})\right]}{(\mu_{it}^M - \nu_{tj})(1-\mu_{it}^M - \mu_{ij}^M + \nu_{tj})}$$

and

$$\frac{\partial^2 \nu_{tj}}{\partial \Psi_t^2} = \frac{\frac{1}{4}a^{-3/2}\left(\frac{\partial a}{\partial \Psi_t}\right)^2 - \frac{1}{2}a^{-1/2}\frac{\partial^2 a_{tj}}{\partial \Psi_t^2} - 4\frac{\partial \nu_{itj}}{\partial \Psi_t}}{2(\Psi_t - 1)}$$

$$\frac{\partial^2 a_{tj}}{\partial \Psi_t^2} = 2(\mu_{it}^M + \mu_{ij}^M)^2 - 8\mu_{it}^M \mu_{ij}^M$$

Finally, $\frac{\partial^2 \log(\boldsymbol{L}(\beta,\alpha))}{\partial \beta, \partial \alpha}$ (which ought to be zero)

$$\begin{aligned}
\frac{\partial \log(\boldsymbol{L}(\beta,\alpha))}{\partial \beta \partial \alpha} &= \frac{\partial}{\partial \alpha}\left[\sum_{i=1}^{N}(y_{i1} - \mu_{i1}^M)\mathrm{X}_{i1} + \sum_{i=1}^{N}\sum_{t=2}^{n_i}(y_{it} - \mu_{it}^C)\frac{\partial \eta_{it}^C}{\partial \mu_{it}^M}\frac{\partial \mu_{it}^M}{\partial \beta}\right. \\
&\qquad \left. + \sum_{i=1}^{N}\sum_{t=2}^{n_i}(y_{it} - \mu_{it}^C)\frac{\partial \eta_{it}^C}{\partial \mu_{ij}^M}\frac{\partial \mu_{ij}^M}{\partial \beta}\right] \\
&= \sum_{i=1}^{N}\sum_{t=2}^{n_i}\left(-\frac{\partial \mu_{it}^C}{\partial \eta_{it}^C}\frac{\partial \eta_{it}^C}{\partial \Psi_t}\frac{\partial \Psi_t}{\partial \alpha}\right)\left[\frac{\partial \eta_{it}^C}{\partial \mu_{it}^M}\frac{\partial \mu_{it}^M}{\partial \beta}\right] \\
&\quad + \sum_{i=1}^{N}\sum_{t=2}^{n_i}\left(\frac{\partial^2 \eta_{it}^C}{\partial \mu_{it}^M \partial \Psi_t}\frac{\partial \Psi_t}{\partial \alpha}\right)\left[(y_{it} - \mu_{it}^C)\frac{\partial \mu_{it}^M}{\partial \beta}\right] \\
&\quad + \sum_{i=1}^{N}\sum_{t=2}^{n_i}\left(-\frac{\partial \mu_{it}^C}{\partial \eta_{it}^C}\frac{\partial \eta_{it}^C}{\partial \Psi_t}\frac{\partial \Psi_t}{\partial \alpha}\right)\left[\frac{\partial \eta_{it}^C}{\partial \mu_{ij}^M}\frac{\partial \mu_{ij}^M}{\partial \beta}\right] \\
&\quad + \sum_{i=1}^{N}\sum_{t=2}^{n_i}\left(\frac{\partial^2 \eta_{it}^C}{\partial \mu_{ij}^M \partial \Psi_t}\frac{\partial \Psi_t}{\partial \alpha}\right)\left[(y_{it} - \mu_{it}^C)\frac{\partial \mu_{ij}^M}{\partial \beta}\right]
\end{aligned}$$

where

$$\frac{\partial^2 \eta_{it}^C}{\partial \mu_{it}^M \partial \Psi_t} = \frac{-\frac{\partial^2 \nu_{itj}}{\partial \mu_{it}^M \partial \Psi_t}(1-\mu_{ij}^M) + \frac{\partial \eta_{it}^C}{\partial \mu_{it}^M}\frac{\partial \nu_{itj}}{\partial \Psi_t}(1-2\mu_{it}^M - \mu_{ij}^M + 2\nu_{itj})}{(\mu_{it}^M - \nu_{tj})(1-\mu_{it}^M - \mu_{ij}^M + \nu_{tj})}$$

$$\frac{\partial^2 \eta_{it}^C}{\partial \mu_{ij}^M \partial \Psi_t} = \frac{-\frac{\partial \nu_{itj}}{\partial \Psi_t} - \frac{\partial^2 \nu_{itj}}{\partial \mu_{ij}^M \partial \Psi_t}(1-\mu_{ij}^M) + \frac{\partial \eta_{it}^C}{\partial \mu_{ij}^M}\frac{\partial \nu_{itj}}{\partial \Psi_t}(1-2\mu_{it}^M - \mu_{ij}^M + 2\nu_{itj})}{(\mu_{it}^M - \nu_{tj})(1-\mu_{it}^M - \mu_{ij}^M + \nu_{tj})}$$

and

$$\frac{\partial^2 \nu_{itj}}{\partial \mu_{it}^M \partial \Psi_t} = -\frac{1}{4}a_{itj}^{-3/2}\frac{\partial a_{itj}}{\partial \Psi_t}\left[-1 + \mu_{it}^M + \mu_{ij}^M + \Psi_t(\mu_{ij}^M - \mu_{it}^M)\right] + \frac{1}{2}a_{itj}^{-1/2}(\mu_{ij}^M - \mu_{it}^M)$$

$$\frac{\partial^2 \nu_{itj}}{\partial \mu_{ij}^M \partial \Psi_t} = -\frac{1}{4} a_{itj}^{-3/2} \frac{\partial a_{itj}}{\partial \Psi_t} \left[-1 + \mu_{it}^M + \mu_{ij}^M + \Psi_t(\mu_{it}^M - \mu_{ij}^M)\right] + \frac{1}{2} a_{itj}^{-1/2} (\mu_{it}^M - \mu_{ij}^M)$$

A.2 Contribution of L2

Let $c = d - 1$ where d is the index of the first timepoint when response Y is missing due to dropout. Other notation is introduced in Section 3.2.1 on page 32. For dropout model $\text{logit}(p_{it}) = \mathrm{D}_{it}\phi + \mathrm{G}_{it}\lambda_1 \cdot \mathrm{Y}_{it} + \mathrm{G}_{it}\lambda_0 \cdot (1 - \mathrm{Y}_{it})$ (λ_0 is not needed, but is included in the accompanying software for maximum flexibility in modeling.),

$$\log \mathrm{L2}_i = \sum_{t=2}^{d-1} \log(p_{it}).$$

L2 contributes to the score for ϕ only, not for α or β.

$$\begin{aligned}
\frac{\partial \log \mathrm{L2}}{\partial \phi} &= \frac{\partial}{\partial \phi} \sum_{i=1}^{N} \sum_{t=2}^{d-1} \log(p_{it}) = \frac{\partial}{\partial \phi} \sum_{i=1}^{N} \sum_{t=2}^{d-1} \log \frac{exp(\mathrm{D}_{it}\phi + \mathrm{G}_{it}\lambda_1 \mathrm{Y}_{it} + \mathrm{G}_{it}\lambda_0(1 - \mathrm{Y}_{it}))}{1 + exp(\mathrm{D}_{it}\phi + \mathrm{G}_{it}\lambda_1 \mathrm{Y}_{it} + \mathrm{G}_{it}\lambda_0(1 - \mathrm{Y}_{it}))} \\
&= \frac{\partial}{\partial \phi} \sum_{i=1}^{N} \sum_{t=2}^{d-1} \mathrm{D}_{it}\phi + \mathrm{G}_{it}\lambda_1 \mathrm{Y}_{it} + \mathrm{G}_{it}\lambda_0(1 - \mathrm{Y}_{it}) \\
&\quad - \log(1 + exp(\mathrm{D}_{it}\phi + \mathrm{G}_{it}\lambda_1 \mathrm{Y}_{it} + \mathrm{G}_{it}\lambda_0(1 - \mathrm{Y}_{it}))) \\
&= \sum_{i=1}^{N} \sum_{t=2}^{d-1} \mathrm{D}_{it}(1 - p_{it}) \\
\frac{\partial^2 \log \mathrm{L2}}{\partial \phi^2} &= \sum_{i=1}^{N} \sum_{t=2}^{d-1} -\mathrm{D}_{it} \frac{\partial p_{it}}{\partial \phi} \\
&= -p_{it}(1 - p_{it}) \mathrm{D}_{it} \mathrm{D}_{it}^T
\end{aligned}$$

A.3 Contribution of L3

$\mathrm{L3} = \int (1 - p_{id}) f(\mathrm{Y}_i^{\mathrm{m}} | \mathrm{Y}_i^{\mathrm{o}}) d(\mathrm{Y}_i^{\mathrm{m}} | \mathrm{Y}_i^{\mathrm{o}})$. Let

$q_{id1} = (1 - p_{id}(\mathrm{Y}_{id} = 1)) = 1/(1 + exp(\mathrm{D}_{id}\phi + \mathrm{G}_{id}\lambda_1))$

$q_{id0} = (1 - p_{id}(\mathrm{Y}_{id} = 0)) = 1/(1 + exp(\mathrm{D}_{id}\phi + \mathrm{G}_{id}\lambda_0))$

$$\frac{\partial \log \mathrm{L3}}{\partial \beta} = \sum_{i=1}^{N} \frac{1}{\mathrm{L3}} \frac{\partial \mathrm{L3}}{\partial \beta}$$

$$\text{where } \frac{\partial \mathrm{L3}}{\partial \beta} = \frac{\partial}{\partial \beta} \left[q_{id1} \mu_{id}^{C} + q_{id0} (1 - \mu_{id}^{C}) \right]$$

$$= (q_{id1} - q_{id0}) \frac{\partial \mu_{id}^{C}}{\partial \beta}$$

$$= (q_{id1} - q_{id0}) \frac{\partial \mu_{id}^{C}}{\partial \eta_{id}^{C}} \frac{\partial \eta_{id}^{C}}{\partial \beta}$$

$$= (q_{id1} - q_{id0}) \mu_{id}^{C} (1 - \mu_{id}^{C}) \frac{\partial \eta_{id}^{C}}{\partial \beta}$$

$$\text{and } \frac{\partial \eta_{id}^{C}}{\partial \beta} = \left(\frac{\partial \eta_{id}^{C}}{\partial \mu_{id}^{M}} \frac{\partial \mu_{id}^{M}}{\partial \beta} + \frac{\partial \eta_{id}^{C}}{\partial \mu_{ic}^{M}} \frac{\partial \mu_{ic}^{M}}{\partial \beta} \right)$$

$$\frac{\partial \log \mathrm{L3}}{\partial \alpha} = \sum_{i=1}^{N} \frac{1}{\mathrm{L3}} \frac{\partial \mathrm{L3}}{\partial \alpha}$$

$$\text{where } \frac{\partial \mathrm{L3}}{\partial \alpha} = (q_{id1} - q_{id0}) \frac{\partial \mu_{id}^{C}}{\partial \alpha}$$

$$\text{and } \frac{\partial \mu_{id}^{C}}{\partial \alpha} = \frac{\partial \mu_{id}^{C}}{\partial \eta_{id}^{C}} \frac{\partial \eta_{id}^{C}}{\partial \Psi_{d}} \frac{\partial \Psi_{d}}{\partial \alpha}$$

$$= \mu_{id}^{C} (1 - \mu_{id}^{C}) \frac{\partial \eta_{id}^{C}}{\partial \Psi_{d}} \frac{\partial \Psi_{d}}{\partial \alpha}$$

$$\frac{\partial \log \mathrm{L3}}{\partial \phi} = \sum_{i=1}^{N} \frac{1}{\mathrm{L3}} \frac{\partial \mathrm{L3}}{\partial \phi}$$

$$\text{where } \frac{\partial \mathrm{L3}}{\partial \phi} = -\mathrm{D}_{id} \left[\mu_{id}^{C} q_{id1} (1 - q_{id1}) + (1 - \mu_{id}^{C}) q_{id0} (1 - q_{id0}) \right]$$

$$\frac{\partial^2 \log \mathrm{L3}}{\partial \beta^2} = \sum_{i=1}^{N} \frac{1}{\mathrm{L3}} \frac{\partial^2 \mathrm{L3}}{\partial \beta^2} - \frac{1}{\mathrm{L3}^2} \left(\frac{\partial \mathrm{L3}}{\partial \beta} \right) \left(\frac{\partial \mathrm{L3}}{\partial \beta} \right)^{T}$$

$$\text{where } \frac{\partial^2 \text{L3}}{\partial \beta^2} = (q_{id1} - q_{id0})(1 - 2\mu_{id}^C)\frac{\partial \mu_{id}^C}{\partial \eta_{id}^C}\frac{\partial \eta_{id}^C}{\partial \beta}\left(\frac{\partial \eta_{id}^C}{\partial \beta}\right)^T + (q_{id1} - q_{id0})\mu_{id}^C(1 - \mu_{id}^C)\frac{\partial^2 \eta_{id}^C}{\partial \beta^2}$$

$$\text{and } \frac{\partial^2 \eta_{id}^C}{\partial \beta^2} = \frac{\partial^2 \eta_{id}^C}{\partial (\mu_{id}^M)^2}\frac{\partial \mu_{id}^M}{\partial \beta}\left(\frac{\partial \mu_{id}^M}{\partial \beta}\right)^T + \frac{\partial \eta_{id}^C}{\partial \mu_{id}^M}\frac{\partial^2 \mu_{id}^M}{\partial \beta^2} + \frac{\partial^2 \eta_{id}^C}{\partial (\mu_{ic}^M)^2}\frac{\partial \mu_{ic}^M}{\partial \beta}\left(\frac{\partial \mu_{ic}^M}{\partial \beta}\right)^T + \frac{\partial \eta_{id}^C}{\partial \mu_{ic}^M}\frac{\partial^2 \mu_{ic}^M}{\partial \beta^2}$$

$$= \frac{\partial^2 \eta_{id}^C}{\partial (\mu_{id}^M)^2}\frac{\partial \mu_{id}^M}{\partial \beta}\left(\frac{\partial \mu_{id}^M}{\partial \beta}\right)^T + \frac{\partial \eta_{id}^C}{\partial \mu_{id}^M}(1 - 2\mu_d^M)\frac{\partial \mu_{id}^M}{\partial \beta}\text{X}_{id}^T + \frac{\partial^2 \eta_{id}^C}{\partial (\mu_{ic}^M)^2}\frac{\partial \mu_{ic}^M}{\partial \beta}\left(\frac{\partial \mu_{ic}^M}{\partial \beta}\right)^T + \frac{\partial \eta_{id}^C}{\partial \mu_{ic}^M}(1 - 2\mu_{ic}^M)\frac{\partial \mu_{ic}^M}{\partial \beta}\text{X}_{ic}^T$$

$$\frac{\partial^2 \log \text{L3}}{\partial \beta \partial \alpha} = \sum_{i=1}^{N} -\frac{1}{\text{L3}^2}\frac{\partial \text{L3}}{\partial \beta}\left(\frac{\partial \text{L3}}{\partial \alpha}\right)^T + \frac{1}{\text{L3}}\frac{\partial^2 \text{L3}}{\partial \beta \partial \alpha}$$

$$\text{where } \frac{\partial^2 \text{L3}}{\partial \beta \partial \alpha} = (q_{id1} - q_{id0})(1 - 2\mu_{id}^C)\frac{\partial \eta_{id}^C}{\partial \beta}\left(\frac{\partial \mu_{id}^C}{\partial \alpha}\right)^T + (q_{id1} - q_{id0})\mu_{id}^C(1 - \mu_{id}^C)\left[\frac{\partial^2 \eta_{id}^C}{\partial \mu_d^M \partial \Psi_d}\frac{\partial \mu_{id}^M}{\partial \beta} + \frac{\partial^2 \eta_{id}^C}{\partial \mu_{ic}^M \partial \Psi_d}\frac{\partial \mu_{ic}^M}{\partial \beta}\right]\left(\frac{\partial \Psi_d}{\partial \alpha}\right)^T$$

$$\frac{\partial^2 \log \text{L3}}{\partial \alpha^2} = \sum_{i=1}^{N} -\frac{1}{\text{L3}^2}\frac{\partial \text{L3}}{\partial \alpha}\left(\frac{\partial \text{L3}}{\partial \alpha}\right)^T + \frac{1}{\text{L3}}\frac{\partial^2 \text{L3}}{\partial \alpha^2}$$

$$\text{where } \frac{\partial^2 \text{L3}}{\partial \alpha^2} = (q_{id1} - q_{id0})(1 - 2\mu_{id}^C)\frac{\partial \mu_{id}^C}{\partial \alpha}\frac{\partial \eta_{id}^C}{\partial \Psi_d}\left(\frac{\partial \Psi_d}{\partial \alpha}\right)^T + (q_{id1} - q_{id0})\mu_{id}^C(1 - \mu_{id}^C)\left[\frac{\partial^2 \eta_{id}^C}{\partial \Psi_d^2}\frac{\partial \Psi_d}{\partial \alpha}\left(\frac{\partial \Psi_d}{\partial \alpha}\right)^T + \frac{\partial \eta_{id}^C}{\partial \Psi_d}\frac{\partial \Psi_d}{\partial \alpha}\text{Z}_{id}^T\right]$$

$$\frac{\partial^2 \log \text{L3}}{\partial \beta \partial \phi} = \sum_{i=1}^{N} -\frac{1}{\text{L3}^2}\frac{\partial \text{L3}}{\partial \phi}\frac{\partial \text{L3}}{\partial \beta} + \frac{1}{\text{L3}}\frac{\partial^2 \text{L3}}{\partial \beta \partial \phi}$$

$$\text{where } \frac{\partial^2 \text{L3}}{\partial \beta \partial \phi} = \mu_{id}^C(1 - \mu_{id}^C)\frac{\partial \eta_{id}^C}{\partial \beta}\left(\frac{\partial (q_{id1} - q_{id0})}{\partial \phi}\right)^T$$

$$\frac{\partial^2 \log \text{L3}}{\partial \alpha \partial \phi} = \sum_{i=1}^{N} -\frac{1}{\text{L3}^2}\frac{\partial \text{L3}}{\partial \alpha}\left(\frac{\partial \text{L3}}{\partial \phi}\right)^T + \frac{1}{\text{L3}}\frac{\partial^2 \text{L3}}{\partial \alpha \partial \phi}$$

$$\text{where } \frac{\partial^2 \text{L3}}{\partial\alpha\partial\phi} = \frac{\partial\mu_{id}^{C}}{\partial\alpha}\left(\frac{\partial(q_{id1}-q_{id0})}{\partial\phi}\right)^{T}$$

$$\text{where } \frac{\partial(q_{id1}-q_{id0})}{\partial\phi} = \text{D}_{id}\left[q_{id0}(1-q_{id0}) - q_{id1}(1-q_{id1})\right]$$

$$\frac{\partial^2 \log \text{L3}}{\partial\phi^2} = \sum_{i=1}^{N} -\frac{1}{\text{L3}^2}\frac{\partial \text{L3}}{\partial\phi}\left(\frac{\partial \text{L3}}{\partial\phi}\right)^{T} + \frac{1}{\text{L3}}\frac{\partial^2 \text{L3}}{\partial\phi^2}$$

$$\text{where } \frac{\partial^2 \text{L3}}{\partial\phi^2} = \text{D}_{id}\text{D}_{id}^{T}\left[\mu_{id}^{C} q_{id1}(1-q_{id1})(1-2q_{id1}) + (1-\mu_{id}^{C})q_{id0}(1-q_{id0})(1-2q_{id0})\right]$$

A.4 Fisher Scoring

The score equation can be written as

$$\sum_{i=1}^{N}(y_{i1}-\mu_{i1}^{M})\text{X}_{i1} + \sum_{i=1}^{N}\left(\frac{\partial\boldsymbol{\mu_i^C}}{\partial\boldsymbol{\theta}}\right)^{\text{T}}\boldsymbol{V_i}^{-1}(\boldsymbol{y_i}-\boldsymbol{\mu_i^C})$$

$$\text{where } \boldsymbol{V_i}^{-1} = \text{diag}\left(\frac{1}{\boldsymbol{\mu_i^C}(1-\boldsymbol{\mu_i^C})}\right)$$

$$\text{since for } \theta=(\alpha,\beta,\phi),\ \frac{\partial\log(\boldsymbol{L})}{\partial\boldsymbol{\theta}} = \frac{\partial\log(\boldsymbol{L})}{\partial\boldsymbol{\mu_i^C}}\frac{\partial\boldsymbol{\mu_i^C}}{\partial\boldsymbol{\theta}} = \frac{\partial\boldsymbol{\mu_i^C}}{\partial\boldsymbol{\theta}}\boldsymbol{V_i}^{-1}(\boldsymbol{y_i}-\boldsymbol{\mu_i^C})$$

$$\text{And Cov}(U) = \frac{\partial\boldsymbol{\mu_i^C}}{\partial\boldsymbol{\theta}}\boldsymbol{V_i}^{-1}\boldsymbol{V_i}\boldsymbol{V_i}^{-1}\frac{\partial\boldsymbol{\mu_i^C}^{T}}{\partial\boldsymbol{\theta}} = \frac{\partial\boldsymbol{\mu_i^C}}{\partial\boldsymbol{\theta}}\frac{\partial\boldsymbol{\mu_i^C}^{T}}{\partial\boldsymbol{\theta}}\frac{1}{\boldsymbol{\mu_i^C}(1-\boldsymbol{\mu_i^C})} - \sum_{i=1}^{N}\mu_{i1}^{M}(1-\mu_{i1}^{M})\text{X}_{i1}\text{X}_{i1}^{T}$$

And $\boldsymbol{\theta} = (\beta,\alpha,\phi)$ is updated as follows:

$$\boldsymbol{\theta}^{(\text{new})} = \boldsymbol{\theta}^{(\text{old})} + \text{Cov}^{-1}(U)U$$

A.5 Newton-Raphson Estimation

Here we use the computed Hessian matrix and

$$\boldsymbol{\theta}^{(\text{new})} = \boldsymbol{\theta}^{(\text{old})} + \left(\frac{-\partial^2 \log(\boldsymbol{L}(\boldsymbol{\theta}))}{\partial \boldsymbol{\theta}^2} \bigg|_{\boldsymbol{\theta}^{(\text{old})}} \right)^{-1} \left(\frac{\partial \log(\boldsymbol{L}(\boldsymbol{\theta}))}{\partial \boldsymbol{\theta}} \bigg|_{\boldsymbol{\theta}^{(\text{old})}} \right)$$

Appendix B

DERIVATION OF SCORE EQUATIONS FOR SECOND-ORDER MARGINALIZED TRANSITION MODEL WITH IGNORABLE AND NONIGNORABLE MONOTONE DROPOUT

Setting: binary outcome, clustered data

Notation and Model:

i=1,...,N (clusters)

t=1,...,n_i (cluster size)

$j = t-1; \quad k = t-2$

B.1 Likelihood under MAR Missingness

$$\begin{aligned} \boldsymbol{L}_i(\beta,\alpha) &= P(\mathrm{Y}_{i1}=y_{i1},\mathrm{Y}_{i2}=y_{i2},\cdots,\mathrm{Y}_{in_i}=y_{in_i}) \\ &= P(\mathrm{Y}_{i1}=y_{i1})P(\mathrm{Y}_{i2}=y_{i2}|\mathrm{Y}_{i1}=y_{i1})P(\mathrm{Y}_{i3}=y_{i3}|\mathrm{Y}_{i1}=y_{i1},\mathrm{Y}_{i2}=y_{i2}) \\ &\quad \cdots P(\mathrm{Y}_{in_i}=y_{in_i}|\mathrm{Y}_{in_i-1}=y_{in_i-1},\mathrm{Y}_{i(n_i-2)}=y_{i(n_i-2)}) \\ &\quad \text{(since lag 2)} \end{aligned}$$

where:

marginal mean

$P(\mathrm{Y}_{i1}=y_{i1}) = (\mu_{i1}^M)^{y_{i1}}(1-\mu_{i1}^M)^{(1-y_{i1})}$

$\mu_{it}^M = P(\mathrm{Y}_{it}=1)$

$\mathrm{logit}(\mu_{it}^M) = \eta_{it}^M = \mathrm{X}_{it}\beta$

conditional mean

$P(\mathrm{Y}_{it}=y_{it}|\mathrm{Y}_{ij}=y_{ij},\mathrm{Y}_{ik}=y_{ik}) = (\mu_{it}^C)^{y_{it}}(1-\mu_{it}^C)^{(1-y_{it})}$

$$\mu_{it}^C = \begin{cases} t=2 & : \quad P(\mathrm{Y}_{it}=1|\mathrm{Y}_{ij}=y_{ij}) \\ t>2 & : \quad P(\mathrm{Y}_{it}=1|\mathrm{Y}_{ij}=y_{ij},\mathrm{Y}_{ik}=y_{ik}) \end{cases}$$

$$\mathrm{logit}(\mu_{it}^C)=\eta_{it}^C = \begin{cases} t=2 & : \quad y_{ij}\log(\Psi_{itj})+\log\frac{\mu_{it}^M-\pi_{(11)}^{t+1}}{1-\mu_{it}^M-\mu_{ij}^M+\pi_{(11)}^{t+1}} \\ t>2 & : \quad \Delta_{it}+\gamma_{it,1}y_{ij}+\gamma_{it,2}y_{ik} \end{cases}$$

$\pi_{(11)}^{t+1}=P(\mathrm{Y}_{it}=1,\mathrm{Y}_{ij}=1)$ (see definition of $\pi_{(jk)}^t$ below)

$\gamma_{i2,1}=\mathrm{Z}_{i2,1}\alpha_0$ (may hold $\alpha_0=\alpha_1$)

$\gamma_{it,1}=\mathrm{Z}_{it,1}\alpha_1$

$\gamma_{it,2}=\mathrm{Z}_{it,2}\alpha_2$

association

$\Psi_{i21}=\frac{P(\mathrm{Y}_{i2}=1,\mathrm{Y}_{i1}=1)P(\mathrm{Y}_{i2}=0,\mathrm{Y}_{i1}=0)}{P(\mathrm{Y}_{i2}=1,\mathrm{Y}_{i1}=0)P(\mathrm{Y}_{i2}=0,\mathrm{Y}_{i1}=1)}=\frac{\pi_{(11)}^3\pi_{(00)}^3}{\pi_{(10)}^3\pi_{(01)}^3}$

$\log(\Psi_{i21})=\mathrm{Z}_{i2,1}\alpha_0$

convolution

$\mu_{it}^M=\sum_{j=0}^1\sum_{k=0}^1\mu_{it(jk)}^C\pi_{(jk)}^t$

$\pi_{(jk)}^t=P(\mathrm{Y}_{it-1}=j,\mathrm{Y}_{it-2}=k) \qquad j,k=0,1$

$\mu_{it(jk)}^C=P(\mathrm{Y}_{it}=y_{it}|\mathrm{Y}_{it-1}=j,\mathrm{Y}_{it-2}=k)$

$\eta_{it(jk)}^C=\mathrm{logit}(\mu_{it(jk)}^C)$

Δ_{it} is not estimated separately, but is found via the convolution equation:

$$\begin{aligned} \mu_{it}^M &= \sum_{j=0}^1\sum_{k=0}^1\mu_{it(jk)}^C\pi_{(jk)}^t \\ f(\Delta_{it})=0 &= \sum_{j=0}^1\sum_{k=0}^1\mathrm{antilogit}(\Delta_{it}+\gamma_{it,1}\cdot j+\gamma_{it,2}\cdot k)\pi_{(jk)}^t-\mu_{it}^M \end{aligned}$$

Taking derivatives to find Δ_{it} by Newton-Raphson:

$$f'(\Delta_{it})=\sum_{j=0}^1\sum_{k=0}^1\mu_{it(jk)}^C(1-\mu_{it(jk)}^C)\pi_{(jk)}^t$$

where $\mu_{it(jk)}^C$ is computed using current parameter values. $\pi_{(jk)}^t$ is discussed below.

Returning to the likelihood:

$$\log(\boldsymbol{L}(\beta,\alpha)) = \sum_{i=1}^N\underbrace{y_{i1}\mathrm{logit}(\mu_{i1}^M)+\log(1-\mu_{i1}^M)}_{\log\boldsymbol{L}_{1i}}+\sum_{i=1}^N\underbrace{y_{i2}\mathrm{logit}(\mu_{i2}^C)+\log(1-\mu_{i2}^C)}_{\log\boldsymbol{L}_{2i}}$$

$$+\sum_{i=1}^{N}\sum_{t=3}^{n_i}\underbrace{y_{it}\text{logit}(\mu_{it}^C)+\log(1-\mu_{it}^C)}_{\log \boldsymbol{L}3i}$$

$\log\mathbf{L}1i$ and $\log\mathbf{L}2i$ contributions dealt with in same manner as for MTM(1) (with lag-1 version of μ^C for $\log\mathbf{L}2i$):

$$\begin{aligned}\frac{\partial \log \boldsymbol{L}1i}{\partial\beta} &= \frac{\partial \log \boldsymbol{L}1i}{\partial\eta_{i1}^M}\frac{\partial\eta_{i1}^M}{\partial\beta}\\ &= (y_{i1}-\mu_{i1}^M)*(\mathrm{X}_{i1})\\ \frac{\partial \log \boldsymbol{L}1i}{\partial\alpha} &= \mathbf{0}\end{aligned}$$

$$\begin{aligned}\frac{\partial \log \boldsymbol{L}2}{\partial\beta} &= \sum_{i=1}^{N}\frac{\partial \log \boldsymbol{L}2i}{\partial\mu_{i2}^C}\frac{\partial\mu_{i2}^C}{\partial\eta_{i2}^C}\frac{\partial\eta_{i2}^C}{\partial\mu_{i2}^M}\frac{\partial\mu_{i2}^M}{\partial\beta}+\frac{\partial \log \boldsymbol{L}2i}{\partial\mu_{i2}^C}\frac{\partial\mu_{i2}^C}{\partial\eta_{i2}^C}\frac{\partial\eta_{i2}^C}{\partial\mu_{i1}^M}\frac{\partial\mu_{i1}^M}{\partial\beta}\\ &= \sum_{i=1}^{N}(y_{i2}-\mu_{i2}^C)\frac{\partial\eta_{i2}^C}{\partial\mu_{i2}^M}\frac{\partial\mu_{i2}^M}{\partial\beta}+\sum_{i=1}^{N}(y_{i2}-\mu_{i2}^C)\frac{\partial\eta_{i2}^C}{\partial\mu_{i1}^M}\frac{\partial\mu_{i1}^M}{\partial\beta}\\ &\quad\text{[see Appendix A]}\\ \frac{\partial \log \boldsymbol{L}2}{\partial\alpha_0} &= \sum_{i=1}^{N}(y_{i2}-\mu_{i2}^C)\frac{\partial\eta_{i2}^C}{\partial\Psi_{i21}}\frac{\partial\Psi_{i21}}{\partial\alpha_0}\\ \frac{\partial \log \boldsymbol{L}2}{\partial\alpha_1} &= \mathbf{0}\\ \frac{\partial \log \boldsymbol{L}2}{\partial\alpha_2} &= \mathbf{0}\end{aligned}$$

$$\begin{aligned}\frac{\partial \log \boldsymbol{L}3i}{\partial\beta} &= \frac{\partial \log \boldsymbol{L}3i}{\partial\mu_{it}^C}\frac{\partial\mu_{it}^C}{\partial\beta}\\ &= \frac{\partial\mu_{it}^C}{\partial\beta}\frac{(y_{it}-\mu_{it}^C)}{\mu_{it}^C(1-\mu_{it}^C)}\end{aligned}$$

where

$$\begin{aligned}
\frac{\partial \mu_{it}^C}{\partial \beta} &= \frac{\partial \mu_{it}^C}{\partial \eta_{it}^C}\frac{\partial \eta_{it}^C}{\partial \beta} \\
&= \mu_{it}^C(1-\mu_{it}^C)\frac{\partial \eta_{it}^C}{\partial \beta} \\
\frac{\partial \eta_{it}^C}{\partial \beta} &= \frac{\partial}{\partial \beta}\left[\,\Delta_{it} + \gamma_{it,1}y_{ij} + \gamma_{it,2}y_{ik}\right] \\
&= \frac{\partial \Delta_{it}}{\partial \beta}
\end{aligned}$$

Find $\frac{\partial \Delta_{it}}{\partial \beta}$ by implicit differentiation of the convolution equation

$$\begin{aligned}
\frac{\partial \mu_{it}^M}{\partial \beta} &= \sum_{j,k}\frac{\partial \mu_{it(jk)}^C}{\partial \Delta_{it}}\frac{\partial \Delta_{it}}{\partial \beta}\pi_{(jk)}^t + \mu_{it(jk)}^C\frac{\partial \pi_{(jk)}^t}{\partial \beta} \\
\Rightarrow \mu_{it}^M(1-\mu_{it}^M)\mathrm{X}_{it} &= \sum_{j,k}\mu_{it(jk)}^C(1-\mu_{it(jk)}^C)\frac{\partial \Delta_{it}}{\partial \beta}\pi_{(jk)}^t + \mu_{it(jk)}^C\frac{\partial \pi_{(jk)}^t}{\partial \beta} \\
\text{and } \frac{\partial \Delta_{it}}{\partial \beta} &= \frac{\mu_{it}^M(1-\mu_{it}^M)\boldsymbol{X}_{it} - \sum_{j,k}\mu_{it(jk)}^C\frac{\partial \pi_{(jk)}^t}{\partial \beta}}{\sum_{j,k}\mu_{it(jk)}^C(1-\mu_{it(jk)}^C)\pi_{(jk)}^t} \\
&= \left[\mu_{it}^M(1-\mu_{it}^M)\mathrm{X}_{it} - \sum_{j,k}\mu_{it(jk)}^C\frac{\partial \pi_{(jk)}^t}{\partial \beta}\right]/A_t
\end{aligned}$$

$\frac{\partial \pi_{(jk)}^3}{\partial \beta}$ is not hard to find:

$$\begin{aligned}
\pi_{(11)}^3 &= \text{solution to } \Psi_{i21} = \frac{\pi_{(11)}^3(1-\mu_{i2}^M-\mu_{i1}^M+\pi_{(11)}^3)}{(\mu_{i1}^M-\pi_{(11)}^3)(\mu_{i2}^M-\pi_{(11)}^3)} \\
\frac{\partial \pi_{(11)}^3}{\partial \beta} &= \frac{\partial \pi_{(11)}^3}{\partial \mu_{i2}^M}\frac{\partial \mu_{i2}^M}{\partial \beta} + \frac{\partial \pi_{(11)}^3}{\partial \mu_{i1}^M}\frac{\partial \mu_{i1}^M}{\partial \beta} \\
pi_{(00)}^3 &= 1-\mu_{i2}^M-\mu_{i1}^M+\pi_{(11)}^3 \\
\frac{\partial \pi_{(00)}^3}{\partial \beta} &= -\frac{\partial \mu_{i2}^M}{\partial \beta} - \frac{\partial \mu_{i1}^M}{\partial \beta} + \frac{\partial \pi_{(11)}^3}{\partial \mu_{i2}^M}\frac{\partial \mu_{i2}^M}{\partial \beta} + \frac{\partial \pi_{(11)}^3}{\partial \mu_{i1}^M}\frac{\partial \mu_{i1}^M}{\partial \beta} \\
\pi_{(01)}^3 &= \mu_{i1}^M - \pi_{(11)}^3 \\
\frac{\partial \pi_{(01)}^3}{\partial \beta} &= \frac{\partial \mu_{i1}^M}{\partial \beta} - \frac{\partial \pi_{(11)}^3}{\partial \mu_{i2}^M}\frac{\partial \mu_{i2}^M}{\partial \beta} - \frac{\partial \pi_{(11)}^3}{\partial \mu_{i1}^M}\frac{\partial \mu_{i1}^M}{\partial \beta}
\end{aligned}$$

$$\begin{aligned}
\pi^3_{(10)} &= \mu^M_{i2} - \pi^3_{(11)} \\
\frac{\partial \pi^3_{(10)}}{\partial \beta} &= \frac{\partial \mu^M_{i2}}{\partial \beta} - \frac{\partial \pi^3_{(11)}}{\partial \mu^M_{i2}}\frac{\partial \mu^M_{i2}}{\partial \beta} - \frac{\partial \pi^3_{(11)}}{\partial \mu^M_{i1}}\frac{\partial \mu^M_{i1}}{\partial \beta}
\end{aligned}$$

where $\frac{\partial \mu^M_{i2}}{\partial \beta} = \mu^M_{i2}(1-\mu^M_{i2})\mathrm{X}_{i2}$ and $\frac{\partial \pi^3_{(11)}}{\partial \mu^M_{i2}}$, $\frac{\partial \pi^3_{(11)}}{\partial \mu^M_{i1}}$ found by implicit differentiation of

$$\Psi_{i21} = \frac{\pi^3_{(11)}(1-\mu^M_{i2}-\mu^M_{i1}+\pi^3_{(11)})}{(\mu^M_{i2}-\pi^3_{(11)})(\mu^M_{i1}-\pi^3_{(11)})}$$

But how are $\pi^t_{(jk)}$ found for higher values of t? Rather than using $\Psi_{it\tilde{j}}$, each $\pi^{t+1}_{(jk)}$ is found through it's relationship with the previous value, $\pi^t_{(jk)}$.

$$\begin{aligned}
\pi^{t+1}_{(11)} &= P(\mathrm{Y}_{it}=1, \mathrm{Y}_{ij}=1) \\
&= P(\mathrm{Y}_{it}=1, \mathrm{Y}_{ij}=1, \mathrm{Y}_{ik}=1) + P(\mathrm{Y}_{it}=1, \mathrm{Y}_{ij}=1, \mathrm{Y}_{ik}=0) \\
&= P(\mathrm{Y}_{it}=1|\mathrm{Y}_{ij}=1, \mathrm{Y}_{ik}=1)P(\mathrm{Y}_{ij}=1, \mathrm{Y}_{ik}=1) \\
&\quad + P(\mathrm{Y}_{it}=1|\mathrm{Y}_{ij}=1, \mathrm{Y}_{ik}=0)P(\mathrm{Y}_{ij}=1, \mathrm{Y}_{ik}=0) \\
&= \mu^C_{it(11)}\pi^t_{(11)} + \mu^C_{it(10)}\pi^t_{(10)}
\end{aligned}$$

similarly,

$$\begin{aligned}
\pi^{t+1}_{(00)} &= (1-\mu^C_{it(01)})\pi^t_{(01)} + (1-\mu^C_{it(00)})\pi^t_{(00)} \\
\pi^{t+1}_{(01)} &= (1-\mu^C_{it(11)})\pi^t_{(11)} + (1-\mu^C_{it(10)})\pi^t_{(10)} \\
\pi^{t+1}_{(10)} &= \mu^C_{it(01)}\pi^t_{(01)} + \mu^C_{it(00)}\pi^t_{(00)}
\end{aligned}$$

$\frac{\partial \pi^{t+1}_{(jk)}}{\partial \beta}$ is also found through its relation to previous values.

$$\begin{aligned}
\frac{\partial \pi^{t+1}_{(11)}}{\partial \beta} &= \frac{\partial \mu^C_{it(11)}}{\partial \beta}\pi^t_{(11)} + \mu^C_{it(11)}\frac{\partial \pi^t_{(11)}}{\partial \beta} + \frac{\partial \mu^C_{it(10)}}{\partial \beta}\pi^t_{(10)} + \mu^C_{it(10)}\frac{\partial \pi^t_{(10)}}{\partial \beta} \\
&= \mu^C_{it(10)}(1-\mu^C_{it(10)})\frac{\partial \Delta_{it}}{\partial \beta}\pi^t_{(10)} + \mu^C_{it(11)}(1-\mu^C_{it(11)})\frac{\partial \Delta_{it}}{\partial \beta}\pi^t_{(11)} + \mu^C_{it(10)}\frac{\partial \pi^t_{(10)}}{\partial \beta} + \mu^C_{it(11)}\frac{\partial \pi^t_{(11)}}{\partial \beta} \\
\frac{\partial \pi^{t+1}_{(00)}}{\partial \beta} &= -\mu^C_{it(01)}(1-\mu^C_{it(01)})\frac{\partial \Delta_{it}}{\partial \beta}\pi^t_{(01)} - \mu^C_{it(00)}(1-\mu^C_{it(00)})\frac{\partial \Delta_{it}}{\partial \beta}\pi^t_{(00)}
\end{aligned}$$

$$+(1-\mu^C_{it(01)})\frac{\partial \pi^t_{(01)}}{\partial \beta}+(1-\mu^C_{it(00)})\frac{\partial \pi^t_{(00)}}{\partial \beta}$$

$$\frac{\partial \pi^{t+1}_{(01)}}{\partial \beta} = -\mu^C_{it(10)}(1-\mu^C_{it(10)})\frac{\partial \Delta_{it}}{\partial \beta}\pi^t_{(10)}-\mu^C_{it(11)}(1-\mu^C_{it(11)})\frac{\partial \Delta_{it}}{\partial \beta}\pi^t_{(11)}$$
$$+(1-\mu^C_{it(10)})\frac{\partial \pi^t_{(10)}}{\partial \beta}+(1-\mu^C_{it(11)})\frac{\partial \pi^t_{(11)}}{\partial \beta}$$

$$\frac{\partial \pi^{t+1}_{(10)}}{\partial \beta} = \mu^C_{it(00)}(1-\mu^C_{it(00)})\frac{\partial \Delta_{it}}{\partial \beta}\pi^t_{(00)}+\mu^C_{it(01)}(1-\mu^C_{it(01)})\frac{\partial \Delta_{it}}{\partial \beta}\pi^t_{(01)}$$
$$+\mu^C_{it(00)}\frac{\partial \pi^t_{(00)}}{\partial \beta}+\mu^C_{it(01)}\frac{\partial \pi^t_{(01)}}{\partial \beta}$$

Now on to $\frac{\partial \log \boldsymbol{L3i}}{\partial \alpha}$.

$$\frac{\partial \log \boldsymbol{L3i}}{\partial \alpha_1} = \frac{\partial \log \boldsymbol{L3i}}{\partial \mu^C_{it}}\frac{\partial \mu^C_{it}}{\partial \alpha_1}$$
$$= \frac{\partial \eta^C_{it}}{\partial \alpha_1}(y_{it}-\mu^C_{it})$$
$$\frac{\partial \eta^C_{it}}{\partial \alpha_1} = \frac{\partial}{\partial \alpha_1}[\,\Delta_{it}+\gamma_{it,1}y_{ij}+\gamma_{it,2}y_{ik}]$$
$$= \frac{\partial \Delta_{it}}{\partial \alpha_1}+\mathrm{Z}_{it,1}y_{ij}$$

Again using implicit differentiation of the convolution equation,

$$\frac{\partial \mu^M_{it}}{\partial \alpha_1}=0 = \sum_{j,k}\frac{\partial \mu^C_{it(jk)}}{\partial \eta^C_{it(jk)}}\left(\frac{\partial \eta^C_{it(jk)}}{\partial \Delta_{it}}\frac{\partial \Delta_{it}}{\partial \alpha_1}+\frac{\partial \eta^C_{it(jk)}}{\partial \gamma_{it,1}}\frac{\partial \gamma_{it,1}}{\partial \alpha_1}\right)\pi^t_{(jk)}+\mu^C_{it}\frac{\partial \pi^t_{(jk)}}{\partial \alpha_1}$$
$$= \sum_{j,k}\mu^C_{it(jk)}(1-\mu^C_{it(jk)})\left[\frac{\partial \Delta_{it}}{\partial \alpha_1}+\mathrm{Z}_{it,1}y_{ij}\right]\pi^t_{(jk)}+\mu^C_{it(jk)}\frac{\partial \pi^t_{(jk)}}{\partial \alpha_1}$$
$$\Rightarrow \frac{\partial \Delta_{it}}{\partial \alpha_1} = \frac{-\sum_{j,k}\left[\mu^C_{it(jk)}\frac{\partial \pi^t_{(jk)}}{\partial \alpha_1}+\mu^C_{it(jk)}(1-\mu^C_{it(jk)})\mathrm{Z}_{it,1}y_{ij}\pi^t_{(jk)}\right]}{A_t}$$

where, since $\gamma_{12}=\log(\Psi_{i21})$,

$$\frac{\partial \pi^3_{(00)}}{\partial \alpha_1}=\frac{\partial \pi^3_{(11)}}{\partial \alpha_1} = \frac{\partial \pi^3_{(11)}}{\partial \Psi_{i21}}\frac{\partial \Psi_{i21}}{\partial \alpha_1}$$
$$\frac{\partial \pi^3_{(01)}}{\partial \alpha_1}=\frac{\partial \pi^3_{(10)}}{\partial \alpha_1} = -\frac{\partial \pi^3_{(11)}}{\partial \Psi_{i21}}\frac{\partial \Psi_{i21}}{\partial \alpha_1}$$

$\frac{\partial \pi^3_{(11)}}{\partial \Psi_{i21}}$ found by implicit differentiation, $\frac{\partial \Psi_{i21}}{\partial \alpha_1} = \Psi_{i21} Z_{i2,1}$

For higher values of t,

$$\begin{aligned}
\frac{\partial \pi^{t+1}_{(11)}}{\partial \alpha_1} &= \frac{\partial \mu^C_{it(11)}}{\partial \alpha_1}\pi^t_{(11)} + \mu^C_{it(11)}\frac{\partial \pi^t_{(11)}}{\partial \alpha_1} + \frac{\partial \mu^C_{it(10)}}{\partial \alpha_1}\pi^t_{(10)} + \mu^C_{it(10)}\frac{\partial \pi^t_{(10)}}{\partial \alpha_1} \\
&= \mu^C_{it(10)}(1-\mu^C_{it(10)})(\frac{\partial \Delta_{it}}{\partial \alpha_1} + 1 * Z_{it,1})\pi^t_{(10)} + \mu^C_{it(11)}(1-\mu^C_{it(11)})(\frac{\partial \Delta_{it}}{\partial \alpha_1} + 1 * Z_{it,1})\pi^t_{(11)} \\
&\quad + \mu^C_{it(11)}\frac{\partial \pi^t_{(11)}}{\partial \alpha_1} + \mu^C_{it(10)}\frac{\partial \pi^t_{(10)}}{\partial \alpha_1} \\
\frac{\partial \pi^{t+1}_{(00)}}{\partial \alpha_1} &= -\mu^C_{it(00)}(1-\mu^C_{it(00)})(\frac{\partial \Delta_{it}}{\partial \alpha_1} + 0 * Z_{it,1})\pi^t_{(00)} - \mu^C_{it(01)}(1-\mu^C_{it(01)})(\frac{\partial \Delta_{it}}{\partial \alpha_1} + 0 * Z_{it,1})\pi^t_{(01)} \\
&\quad + (1-\mu^C_{it(00)})\frac{\partial \pi^t_{(00)}}{\partial \alpha_1} + (1-\mu^C_{it(01)})\frac{\partial \pi^t_{(01)}}{\partial \alpha_1} \\
\frac{\partial \pi^{t+1}_{(01)}}{\partial \alpha_1} &= -\mu^C_{it(10)}(1-\mu^C_{it(10)})(\frac{\partial \Delta_{it}}{\partial \alpha_1} + 1 * Z_{it,1})\pi^t_{(10)} - \mu^C_{it(11)}(1-\mu^C_{it(11)})(\frac{\partial \Delta_{it}}{\partial \alpha_1} + 1 * Z_{it,1})\pi^t_{(11)} \\
&\quad + (1-\mu^C_{it(10)})\frac{\partial \pi^t_{(10)}}{\partial \alpha_1} + (1-\mu^C_{it(11)})\frac{\partial \pi^t_{(11)}}{\partial \alpha_1} \\
\frac{\partial \pi^{t+1}_{(10)}}{\partial \alpha_1} &= \mu^C_{it(00)}(1-\mu^C_{it(00)})(\frac{\partial \Delta_{it}}{\partial \alpha_1} + 0 * Z_{it,1})\pi^t_{(00)} + \mu^C_{it(01)}(1-\mu^C_{it(01)})(\frac{\partial \Delta_{it}}{\partial \alpha_1} + 0 * Z_{it,1})\pi^t_{(01)} \\
&\quad + \mu^C_{it(01)}\frac{\partial \pi^t_{(01)}}{\partial \alpha_1} + \mu^C_{it(00)}\frac{\partial \pi^t_{(00)}}{\partial \alpha_1}
\end{aligned}$$

α_2 is similar, except that $\frac{\partial \pi^3_{(jk)}}{\partial \alpha_2}$ is 0 for all 4 combinations of j,k. And the coefficients (0,1) are different for $Z_{it,2}$ compared to $Z_{it,1}$ above.

And the contribution of log L3i to the expected information matrix is $\frac{\partial \mu^C_{it}}{\partial \theta}\frac{\partial \mu^{C^T}_{it}}{\partial \theta}\frac{1}{\mu^C_{it}(1-\mu^C_{it})}$ where $\theta = (\beta, \alpha)$.

B.2 Contribution of Selection Model

The contribution of L2 and L3 to the score equations is almost identical to that described for MTM(1) in Appendix A. The exception is terms dealing with α_2. The Hessian for MTM(2) is found by numerical differentiation for models with nonignorable dropout, so second derivatives are not given analytically.

$$\frac{\partial \log \mathrm{L3}}{\partial \beta} = \sum_{i=1}^{N} \frac{1}{\mathrm{L3}i} \frac{\partial \mathrm{L3}}{\partial \beta}$$

$$= \sum_{i=1}^{N} \frac{1}{\mathrm{L3}i} (q_{id1} - q_{id0}) \mu_{id}^{C} (1 - \mu_{id}^{C}) \frac{\partial \Delta_{id}}{\partial \beta}$$

$$\frac{\partial \log \mathrm{L3}}{\partial \alpha_1} = \sum_{i=1}^{N} \frac{1}{\mathrm{L3}i} (q_{id1} - q_{id0}) \mu_{id}^{C} (1 - \mu_{id}^{C}) \frac{\partial \Delta_{id}}{\partial \alpha_1}$$

$$\frac{\partial \log \mathrm{L3}}{\partial \alpha_2} = \sum_{i=1}^{N} \frac{1}{\mathrm{L3}i} (q_{id1} - q_{id0}) \mu_{id}^{C} (1 - \mu_{id}^{C}) \frac{\partial \Delta_{id}}{\partial \alpha_2}$$

Note that $\frac{\partial \log \mathrm{L3}}{\partial \phi}$ is unchanged.

Appendix C

ADDITIONAL PLOTS AND TABLES: CHAPTER 4

Table C.1: Efficiency comparing IPCW-GEE models to MTM(2): main effects MAR selection model, 20% missing data. Independence (ind), exchangeable (exch), and autoregressive (AR) working correlation. 1000 simulations with sample size of 500.

covariate	TRUE (ind)	TRUE (exch)	TRUE (AR)	OVER (ind)	OVER (exch)	OVER (AR)
	LEVEL (ϕ_2, ϕ_3) = (-0.75, -0.75)					
INTERCEPT	94	94	95	96	96	97
GROUP	89	89	89	96	95	96
TIME	95	95	95	96	95	95
GROUP:TIME	92	92	91	95	95	94
	CURRENT (ϕ_2, ϕ_3) = (-1.5, 0.0)					
INTERCEPT	93	93	93	96	96	96
GROUP	83	83	83	92	91	92
TIME	94	94	93	95	95	94
GROUP:TIME	90	90	89	94	94	93
	INCREMENT (ϕ_2, ϕ_3) = (-0.75, 0.75)					
INTERCEPT	97	97	97	99	98	99
GROUP	96	95	96	98	98	99
TIME	98	97	97	98	98	98
GROUP:TIME	97	97	96	99	98	98

correct selection (TRUE):

$$\text{logit}[\text{E}(\text{R}_{it}|\text{R}_{it-1}=1, \text{Y}_{it}, \text{H}_{it}^{(\text{Y})})] = \hat{\phi}_0 + \hat{\phi}_2 \cdot y_{it-1} + \hat{\phi}_3 \cdot y_{it-2}$$

overspecified selection (OVER):

$$\text{logit}[\text{E}(\text{R}_{it}|\text{R}_{it-1}=1, \text{Y}_{it}, \text{H}_{it}^{(\text{Y})})] = \hat{\phi}_0 + \hat{\phi}_2 \cdot y_{it-1} + \hat{\phi}_3 \cdot y_{it-2} + \hat{\phi}_5 \cdot \mathbf{group}_i \cdot y_{it-1} + \hat{\phi}_6 \cdot \mathbf{group}_i \cdot y_{it-2}$$

Table C.2: Efficiency comparing IPCW-GEE models to MTM(2): main effects MAR selection model, 50% missing data. Independence (ind), exchangeable (exch), and autoregressive (AR) working correlation. 1000 simulations with sample size of 500.

covariate	TRUE (ind)	TRUE (exch)	TRUE (AR)	OVER (ind)	OVER (exch)	OVER (AR)
	LEVEL (ϕ_2, ϕ_3) = (-0.75, -0.75)					
INTERCEPT	62	64	59	65	66	60
GROUP	54	54	51	62	63	58
TIME	55	55	52	55	54	51
GROUP:TIME	50	50	47	54	54	50
	CURRENT (ϕ_2, ϕ_3) = (-1.5, 0.0)					
INTERCEPT	55	56	53	58	58	55
GROUP	44	44	43	52	52	50
TIME	47	46	44	47	46	44
GROUP:TIME	42	41	40	45	45	43
	INCREMENT (ϕ_2, ϕ_3) = (-0.75, 0.75)					
INTERCEPT	83	85	80	88	89	84
GROUP	79	79	76	87	87	83
TIME	73	72	70	75	74	72
GROUP:TIME	71	70	68	75	74	72

correct selection (TRUE):

$$\text{logit}[\text{E}(\text{R}_{it}|\text{R}_{it-1}=1, \text{Y}_{it}, \text{H}_{it}^{(\text{Y})})] = \hat{\phi}_0 + \hat{\phi}_2 \cdot y_{it-1} + \hat{\phi}_3 \cdot y_{it-2}$$

overspecified selection (OVER):

$$\text{logit}[\text{E}(\text{R}_{it}|\text{R}_{it-1}=1, \text{Y}_{it}, \text{H}_{it}^{(\text{Y})})] = \hat{\phi}_0 + \hat{\phi}_2 \cdot y_{it-1} + \hat{\phi}_3 \cdot y_{it-2} + \hat{\phi}_5 \cdot \mathbf{group}_i \cdot y_{it-1} + \hat{\phi}_6 \cdot \mathbf{group}_i \cdot y_{it-2}$$

Table C.3: Efficiency comparing IPCW-GEE models to MTM(2): MAR selection model with group-by-response interaction. 20% and 50% missing data, with IPCW-GEE weighted by correctly specified selection model. Independence (ind), exchangeable (exch), and autoregressive (AR) working correlation. 1000 simulations with sample size of 500.

	20% MISSING			50% MISSING		
covariate	ind	exch	AR	ind	exch	AR
LEVEL $(\phi_2, \phi_3, \phi_5, \phi_6)$ = (-0.75, -0.75, 0.75, 0.75)						
INTERCEPT	96	96	96	61	64	57
GROUP	98	97	99	77	79	72
TIME	95	95	93	52	51	47
GROUP:TIME	99	99	97	65	64	59
CURRENT $(\phi_2, \phi_3, \phi_5, \phi_6)$ = (-1.5, 0.0, 1.5, 0.0)						
INTERCEPT	93	92	92	54	55	50
GROUP	95	94	96	65	66	60
TIME	91	90	89	45	45	42
GROUP:TIME	95	95	95	56	55	52
INCREMENT $(\phi_2, \phi_3, \phi_5, \phi_6)$ = (-0.75, 0.75, 0.75, -0.75)						
INTERCEPT	99	98	99	87	89	81
GROUP	98	98	99	87	89	82
TIME	99	98	97	76	74	70
GROUP:TIME	98	98	99	79	78	74

selection: $\text{logit}[\text{E}(\text{R}_{it}|\text{R}_{it-1}=1, \text{Y}_{it}, \text{H}_{it}^{(\text{Y})})] =$

$\phi_0 + \phi_2 \cdot y_{it-1} + \phi_3 \cdot y_{it-2} + \phi_5 \cdot \texttt{group}_i \cdot y_{it-1} + \phi_6 \cdot \texttt{group}_i \cdot y_{it-2}$

Table C.4: Percent bias for independence estimating equation (IEE), inverse probability of censoring weighted generalized estimating equations (IPCW-GEE) and marginalized transition models (MTM) fitted to data generated under a second-order MTM. Percent bias is based on 1000 simulated data sets with 500 subjects. 20% and 50% missing data generated by MAR selection model with main effects of first and second-order lagged responses.

		20% MISSING			50% MISSING		
pattern	(ϕ_2, ϕ_3)	IEE	IPW1	MTM(1)	IEE	IPW1	MTM(1)
				INTERCEPT ($\beta_0 = 0.40$)			
level	(-0.75, -0.75)	-38	-7	-9	-85	-22	-16
current	(-1.5, 0.0)	-54	0	-8	-125	-5	-19
increment	(-0.75, 0.75)	-13	5	1	-45	17	-5
				GROUP ($\beta_1 = -1.25$)			
level	(-0.75, -0.75)	-1	0	0	1	3	4
current	(-1.5, 0.0)	-1	1	1	-3	0	1
increment	(-0.75, 0.75)	0	1	1	0	2	1
				TIME ($\beta_2 = 0.65$)			
level	(-0.75, -0.75)	-6	-1	-1	-13	-4	-2
current	(-1.5, 0.0)	-8	1	-1	-18	0	-3
increment	(-0.75, 0.75)	-1	2	1	-6	5	0
				GROUP:TIME ($\beta_3 = -0.31$)			
level	(-0.75, -0.75)	-2	-1	0	-1	2	3
current	(-1.5, 0.0)	-1	1	2	-5	0	1
increment	(-0.75, 0.75)	0	1	1	0	3	1

true selection model:$\text{logit}[E(R_{it}|R_{it-1}=1, Y_{it}, H_{it}^{(Y)})] = \phi_0 + \phi_2 \cdot y_{it-1} + \phi_3 \cdot y_{it-2}$

IPW1 selection model:$\text{logit}[E(R_{it}|R_{it-1}=1, Y_{it}, H_{it}^{(Y)})] = \hat{\phi}_0 + \hat{\phi}_2 \cdot y_{it-1}$

Table C.5: Percent bias for independence estimating equation (IEE), inverse probability of censoring weighted generalized estimating equations (IPCW-GEE) and marginalized transition models (MTM) fitted to data generated under second-order MTM. Percent bias based on 1000 simulated data sets with 500 subjects. 20 and 50% missing data generated by MAR selection model with a group-by-response interaction for both first and second-order lagged responses.

		20% MISSING					50% MISSING				
pattern	$(\phi_2, \phi_3, \phi_5, \phi_6)$	IEE	IPW1	IPW2	IPW3	MTM(1)	IEE	IPW1	IPW2	IPW3	MTM(1)
						INTERCEPT ($\beta_0 = 0.40$)					
level	(-0.75, -0.75, 0.75, 0.75)	-40	-22	-20	-6	-9	-89	-61	-58	-25	-18
current	(-1.5, 0.0, 1.5, 0.0)	-59	-30	-30	-1	-10	-132	-88	-90	-9	-18
increment	(-0.75, 0.75, 0.75, -0.75)	-13	-3	-5	4	0	-44	-12	-19	16	-3
						GROUP ($\beta_1 = -1.25$)					
level	(-0.75, -0.75, 0.75, 0.75)	-13	-13	-14	-2	-3	-27	-34	-36	-7	-5
current	(-1.5, 0.0, 1.5, 0.0)	-18	-20	-20	0	-3	-41	-53	-51	-2	-5
increment	(-0.75, 0.75, 0.75, -0.75)	-4	-3	-3	2	1	-13	-12	-12	6	0
						TIME ($\beta_2 = 0.65$)					
level	(-0.75, -0.75, 0.75, 0.75)	-6	-4	-3	-1	-1	-13	-10	-9	-5	-3
current	(-1.5, 0.0, 1.5, 0.0)	-9	-4	-4	0	-2	-19	-13	-13	-1	-3
increment	(-0.75, 0.75, 0.75, -0.75)	-1	0	0	1	0	-6	0	-2	4	0
						GROUP:TIME ($\beta_3 = -0.31$)					
level	(-0.75, -0.75, 0.75, 0.75)	-14	-14	-14	-3	-3	-28	-35	-38	-10	-6
current	(-1.5, 0.0, 1.5, 0.0)	-18	-20	-19	0	-4	-41	-50	-48	-2	-6
increment	(-0.75, 0.75, 0.75, -0.75)	-4	-3	-3	2	0	-12	-10	-10	8	-1

true selection model: $\text{logit}[E(R_{it}|R_{it-1}=1, Y_{it}, H_{it}^{(Y)})] = \phi_0 + \phi_2 \cdot y_{it-1} + \phi_3 \cdot y_{it-2} + \phi_5 \cdot \text{group}_i \cdot y_{it-1} + \phi_6 \cdot \text{group}_i \cdot y_{it-2}$

IPW1 selection model: $\text{logit}[E(R_{it}|R_{it-1}=1, Y_{it}, H_{it}^{(Y)})] = \hat{\phi}_0 + \hat{\phi}_2 \cdot y_{it-1}$

IPW2 selection model: $\text{logit}[E(R_{it}|R_{it-1}=1, Y_{it}, H_{it}^{(Y)})] = \hat{\phi}_0 + \hat{\phi}_2 \cdot y_{it-1} + \hat{\phi}_3 \cdot y_{it-2}$

IPW3 selection model: $\text{logit}[E(R_{it}|R_{it-1}=1, Y_{it}, H_{it}^{(Y)})] = \hat{\phi}_0 + \hat{\phi}_2 \cdot y_{it-1} + \hat{\phi}_5 \cdot \text{group}_i \cdot y_{it-1}$

Table C.6: Percent bias for independence estimating equation (IEE), inverse probability of censoring weighted generalized estimating equations (IPCW-GEE) and marginalized transition models (MTM) fitted to data generated under second-order MTM. "Level" situation with 50% of data MAR, for original model, long series, strong association, and different regression model. Percent bias based on 1000 simulated data sets with 500 subjects.

case	MAIN EFFECTS SELECTION			GROUP*RESPONSE SELECTION				
	IEE	IPW1	MTM(1)	IEE	IPW1	IPW2	IPW3	MTM(1)
				INTERCEPT ($\beta_0 = 0.40$)				
Level50	-85	-22	-16	-89	-61	-58	-25	-18
$(\alpha_1, \alpha_2) = (\log(16), \log(8))$	-132	-41	-34	-133	-92	-89	-38	-29
cluster size 10	-50	-14	-13	-46	-33	-31	-9	-8
β = c(1.2, -2.5, 0.65, -0.31)	-39	-10	-8	-41	-21	-17	-10	-9
				GROUP ($\beta_1 = -1.25$)				
Level50	1	3	4	-27	-34	-36	-7	-5
$(\alpha_1, \alpha_2) = (\log(16), \log(8))$	7	5	8	-42	-56	-61	-11	-9
cluster size 10	1	3	3	-14	-18	-19	-2	-2
β = c(1.2, -2.5, 0.65, -0.31)	-8	-1	0	-19	-18	-19	-4	-3
				TIME ($\beta_2 = 0.65$)				
Level50	-13	-4	-2	-13	-10	-9	-5	-3
$(\alpha_1, \alpha_2) = (\log(16), \log(8))$	-20	-8	-5	-20	-14	-14	-7	-5
cluster size 10	-4	-1	-1	-3	-2	-2	0	0
β = c(1.2, -2.5, 0.65, -0.31)	-19	-6	-4	-19	-11	-8	-6	-5
				GROUP:TIME ($\beta_3 = -0.31$)				
Level50	-1	2	3	-28	-35	-38	-10	-6
$(\alpha_1, \alpha_2) = (\log(16), \log(8))$	7	6	8	-43	-57	-64	-14	-9
cluster size 10	-1	1	1	-7	-9	-9	0	1
β = c(1.2, -2.5, 0.65, -0.31)	-18	-4	-1	-40	-40	-41	-13	-9

Table C.7: Percent bias for independence estimating equation (IEE), inverse probability of censoring weighted generalized estimating equations (IPCW-GEE) and marginalized transition models (MTM) fitted to data generated under second-order MTM. Percent bias based on 1000 simulated data sets with 500 subjects. True NI selection model includes main effects for both first and second-order lagged responses.

		20% MISSING						50% MISSING					
pattern	(ϕ_1, ϕ_2)	IEE	IPW2	IPW4	MAR	NI1	NI2	IEE	IPW2	IPW4	MAR	NI1	NI2
						INTERCEPT ($\beta_0 = 0.40$)							
level	(-0.75, -0.75)	-78	-39	-36	-39	-57	-21	-217	-124	-117	-133	-186	-76
current	(-1.5, 0.0)	-107	-79	-75	-81	-120	-41	-342	-261	-250	-285	-407	-150
increment	(-0.75, 0.75)	-25	-32	-31	-35	-52	-16	-90	-104	-101	-119	-176	-58
						GROUP ($\beta_1 = -1.25$)							
level	(-0.75, -0.75)	-1	0	2	1	0	1	-1	3	7	2	0	3
current	(-1.5, 0.0)	-1	0	3	0	-3	2	1	3	10	5	-1	8
increment	(-0.75, 0.75)	2	2	3	2	1	2	2	1	3	1	-1	2
						TIME ($\beta_2 = 0.65$)							
level	(-0.75, -0.75)	-10	-4	-4	-4	-7	-2	-29	-14	-13	-17	-24	-9
current	(-1.5, 0.0)	-13	-9	-8	-10	-15	-4	-46	-30	-28	-38	-55	-20
increment	(-0.75, 0.75)	-2	-3	-3	-3	-6	-1	-11	-11	-10	-15	-23	-7
						GROUP:TIME ($\beta_3 = -0.31$)							
level	(-0.75, -0.75)	-1	0	2	1	0	1	-1	4	9	3	1	4
current	(-1.5, 0.0)	-1	0	3	0	-3	2	4	8	15	8	2	10
increment	(-0.75, 0.75)	3	3	3	2	1	2	3	3	5	2	0	2

true selection model: $\text{logit}[E(R_{it}|R_{it-1}=1, Y_{it}, H_{it}^{(Y)})] = \phi_0 + \phi_1 \cdot y_{it} + \phi_2 \cdot y_{it-1}$

IPW2 selection model: $\hat{\phi}_0 + \hat{\phi}_2 \cdot y_{it-1} + \hat{\phi}_3 \cdot y_{it-2}$

IPW4 selection model: $\hat{\phi}_0 + \hat{\phi}_2 \cdot y_{it-1} + \hat{\phi}_3 \cdot y_{it-2} + \hat{\phi}_5 \cdot \text{group}_i \cdot y_{it-1} + \hat{\phi}_6 \cdot \text{group}_i \cdot y_{it-2}$

NI1 selection model: $\hat{\phi}_0 + \lambda_1 \cdot y_{it} + \hat{\phi}_2 \cdot y_{it-1}$ $\quad \lambda_1 = -\phi_1/2$

NI2 selection model: $\hat{\phi}_0 + \lambda_1 \cdot y_{it} + \hat{\phi}_2 \cdot y_{it-1}$ $\quad \lambda_1 = \phi_1/2$

Table C.8: Percent bias for independence estimating equation (IEE), inverse probability of censoring weighted generalized estimating equations (IPCW-GEE) and marginalized transition models (MTM) fitted to data generated under second-order MTM. Percent bias based on 1000 simulated data sets with 500 subjects. True NI selection model includes a group-by-response interaction for both first and second-order lagged responses.

		20% MISSING						50% MISSING					
pattern	$(\phi_1, \phi_2, \phi_4, \phi_5)$	IEE	IPW2	IPW4	MAR	NI1	NI2	IEE	IPW2	IPW4	MAR	NI1	NI2
						INTERCEPT ($\beta_0 = 0.40$)							
level	(-0.75, -0.75, 0.75, 0.75)	-87	-64	-39	-42	-63	-21	-235	-191	-124	-140	-201	-73
current	(-1.5, 0.0, 1.5, 0.0)	-129	-111	-88	-99	-146	-50	-377	-328	-273	-314	-462	-151
increment	(-0.75, 0.75, 0.75, -0.75)	-27	-30	-31	-36	-52	-18	-92	-93	-97	-119	-175	-59
						GROUP ($\beta_1 = -1.25$)							
level	(-0.75, -0.75, 0.75, 0.75)	-27	-28	-11	-13	-19	-6	-75	-76	-37	-44	-63	-22
current	(-1.5, 0.0, 1.5, 0.0)	-41	-40	-26	-32	-46	-16	-119	-111	-82	-99	-147	-47
increment	(-0.75, 0.75, 0.75, -0.75)	-8	-7	-8	-10	-16	-4	-29	-26	-28	-37	-55	-18
						TIME ($\beta_2 = 0.65$)							
level	(-0.75, -0.75, 0.75, 0.75)	-11	-8	-4	-5	-7	-2	-32	-24	-13	-18	-27	-9
current	(-1.5, 0.0, 1.5, 0.0)	-16	-13	-9	-11	-18	-5	-51	-41	-33	-43	-64	-20
increment	(-0.75, 0.75, 0.75, -0.75)	-2	-3	-3	-4	-6	-2	-11	-10	-10	-15	-23	-7
						GROUP:TIME ($\beta_3 = -0.31$)							
level	(-0.75, -0.75, 0.75, 0.75)	-25	-25	-9	-11	-16	-5	-68	-65	-27	-39	-56	-20
current	(-1.5, 0.0, 1.5, 0.0)	-34	-33	-19	-25	-39	-12	-107	-93	-64	-89	-134	-41
increment	(-0.75, 0.75, 0.75, -0.75)	-5	-4	-5	-8	-12	-3	-24	-17	-19	-33	-49	-16

true selection model: $\text{logit}[E(R_{it}|R_{it-1}=1, Y_{it}, H_{it}^{(Y)})] = \phi_0 + \phi_1 \cdot y_{it} + \phi_2 \cdot y_{it-1} + \phi_4 \cdot \text{group}_i \cdot y_{it} + \phi_5 \cdot \text{group}_i \cdot y_{it-1}$

IPW2 selection model: $\hat{\phi}_0 + \hat{\phi}_2 \cdot y_{it-1} + \hat{\phi}_3 \cdot y_{it-2}$

IPW4 selection model: $\hat{\phi}_0 + \hat{\phi}_2 \cdot y_{it-1} + \hat{\phi}_3 \cdot y_{it-2} + \hat{\phi}_5 \cdot \text{group}_i \cdot y_{it-1} + \hat{\phi}_6 \cdot \text{group}_i \cdot y_{it-2}$

NI1 selection model: $\hat{\phi}_0 + \lambda_1 \cdot y_{it} + \hat{\phi}_2 \cdot y_{it-1} + \lambda_2 \cdot \text{group}_i \cdot y_{it} + \hat{\phi}_5 \cdot \text{group}_i \cdot y_{it-1}$ $\quad \lambda_1 = -\phi_1/2, \lambda_2 = -\phi_4/2$

NI2 selection model: $\hat{\phi}_0 + \lambda_1 \cdot y_{it} + \hat{\phi}_2 \cdot y_{it-1} + \lambda_2 \cdot \text{group}_i \cdot y_{it} + \hat{\phi}_5 \cdot \text{group}_i \cdot y_{it-1}$ $\quad \lambda_1 = \phi_1/2, \lambda_2 = \phi_4/2$

Table C.9: Percent bias for independence estimating equation (IEE), inverse probability of censoring weighted generalized estimating equations (IPCW-GEE) and marginalized transition models (MTM) fitted to data generated under second-order MTM. Percent bias based on 1000 simulated data sets with 500 subjects. "Level" situation with 50% NI missing data, for original model, long series, strong association, and different regression model.

case	MAIN EFFECTS SELECTION						GROUP*RESPONSE SELECTION					
	IEE	IPW2	IPW4	MAR	NI1	NI2	IEE	IPW2	IPW4	MAR	NI1	NI2
						INTERCEPT ($\beta_0 = 0.40$)						
Level50	-217	-124	-117	-133	-186	-76	-235	-191	-124	-140	-201	-73
$(\alpha_1, \alpha_2) = (\log(16), \log(8))$	-273	-120	-112	-120	-166	-67	-285	-227	-112	-121	-173	-63
cluster size 10	-118	-72	-65	-68	-96	-39	-123	-108	-69	-71	-104	-36
β = c(1.2, -2.5, 0.65, -0.31)	-86	-41	-38	-47	-67	-27	-90	-60	-38	-48	-71	-25
						GROUP ($\beta_1 = -1.25$)						
Level50	-1	3	7	2	0	3	-75	-76	-37	-44	-63	-22
$(\alpha_1, \alpha_2) = (\log(16), \log(8))$	5	2	9	0	-1	1	-90	-101	-32	-37	-54	-19
cluster size 10	-1	1	5	2	1	3	-39	-42	-20	-22	-33	-11
β = c(1.2, -2.5, 0.65, -0.31)	-12	-1	3	-2	-5	0	-42	-39	-16	-22	-33	-11
						TIME ($\beta_2 = 0.65$)						
Level50	-29	-14	-13	-17	-24	-9	-32	-24	-13	-18	-27	-9
$(\alpha_1, \alpha_2) = (\log(16), \log(8))$	-39	-14	-13	-16	-22	-9	-41	-31	-13	-16	-23	-8
cluster size 10	-7	-2	-2	-2	-3	-1	-7	-6	-1	-2	-4	0
β = c(1.2, -2.5, 0.65, -0.31)	-37	-15	-14	-20	-28	-11	-38	-24	-14	-20	-30	-10
						GROUP:TIME ($\beta_3 = -0.31$)						
Level50	-1	4	9	3	1	4	-68	-65	-27	-39	-56	-20
$(\alpha_1, \alpha_2) = (\log(16), \log(8))$	6	4	11	1	0	2	-85	-94	-25	-34	-48	-17
cluster size 10	-2	1	3	2	1	2	-15	-16	-3	-4	-9	-1
β = c(1.2, -2.5, 0.65, -0.31)	-26	-4	4	-5	-12	0	-79	-70	-24	-41	-62	-20

Appendix D

ADDITIONAL TABLES: CHAPTER 5

Table D.1: Models for schizophrenia data (N=420): Parameter estimates and sandwich standard errors for: generalized estimating equations with independence working correlation (IEE); GEE weighted by inverse probability of censoring modeled by first-and second-order response transitions varying by treatment group (IPW2). Parameter estimates and model-based standard errors for: first-order(MTM(1)) and second-order (MTM(2)) marginalized transition models.

Model	IEE	IPW2 (AR)	MTM(1)	MTM(2)
Marginal Mean				
Intercept (β_0)	-0.476 (0.324)	-0.691 (0.259)	-0.634 (0.313)	-0.703 (0.306)
PLAC (β_1)	0.046 (0.514)	-0.096 (0.436)	-0.130 (0.487)	-0.184 (0.483)
RISP (β_2)	0.472 (0.360)	0.522 (0.288)	0.474 (0.349)	0.524 (0.340)
Week (β_3)	-0.055 (0.139)	-0.065 (0.142)	-0.033 (0.158)	-0.072 (0.138)
Week^2 (β_4)	-0.031 (0.018)	-0.028 (0.017)	-0.025 (0.021)	-0.029 (0.019)
PLAC*Week (β_5)	0.425 (0.214)	0.409 (0.221)	0.419 (0.261)	0.431 (0.240)
PLAC*Week^2 (β_6)	0.060 (0.026)	0.060 (0.026)	0.062 (0.034)	0.065 (0.031)
RISP*Week (β_7)	-0.059 (0.155)	-0.075 (0.157)	-0.089 (0.176)	-0.043 (0.153)
RISP*Week^2 (β_8)	-0.006 (0.020)	-0.009 (0.020)	-0.010 (0.024)	-0.004 (0.021)
Serial Dependence				
Y_{i2} Intercept (α_0)			($\alpha_0 = \alpha_1$)	2.401 (0.292)
Y_{it-1} Intercept (α_1)			2.769 (0.156)	2.378 (0.199)
Y_{it-2} Intercept (α_2)				1.322 (0.216)
Autoregression		0.61		
Log-Likelihood			-783.17	-763.94

Table D.2: Sensitivity analysis models for schizophrenia data (N=420): Parameter estimates, with model-based standard errors.

Model	MTM (MAR)	MTM (NI1)	MTM (NI2)	MTM (NI3)	MAR (NI4)
$(\lambda_0, \lambda_1, \lambda_2)$	(0,0,0)	(1.5, 0, 0)	(-1.5, 0, 0)	(0.75, 0, 1)	(1.5, 0, -1)
Marginal Mean					
Intercept (β_0)	-0.703 (0.306)	-0.973 (0.301)	-0.281 (0.311)	-0.862 (0.300)	-0.97 (0.304)
PLAC (β_1)	-0.184 (0.483)	-0.286 (0.461)	0.076 (0.483)	-0.277 (0.467)	-0.268 (0.465)
RISP (β_2)	0.524 (0.340)	0.585 (0.333)	0.395 (0.347)	0.45 (0.333)	0.706 (0.337)
Week (β_3)	-0.072 (0.138)	-0.067 (0.130)	-0.095 (0.149)	-0.067 (0.130)	-0.067 (0.135)
Week^2 (β_4)	-0.029 (0.019)	-0.023 (0.017)	-0.041 (0.020)	-0.025 (0.018)	-0.023 (0.018)
PLAC*Week (β_5)	0.431 (0.240)	0.355 (0.222)	0.513 (0.246)	0.380 (0.226)	0.358 (0.227)
PLAC*Week^2 (β_6)	0.065 (0.031)	0.056 (0.029)	0.072 (0.032)	0.060 (0.029)	0.056 (0.030)
RISP*Week (β_7)	-0.043 (0.153)	-0.060 (0.143)	-0.013 (0.166)	-0.064 (0.143)	-0.051 (0.148)
RISP*Week^2 (β_8)	-0.004 (0.021)	-0.008 (0.019)	0.003 (0.022)	-0.006 (0.019)	-0.009 (0.020)
Serial Dependence					
Y_{i2} Intercept (α_0)	2.401 (0.292)	2.485 (0.288)	2.184 (0.290)	2.47 (0.288)	2.467 (0.291)
Y_{it-1} Intercept (α_1)	2.378 (0.199)	2.644 (0.200)	2.024 (0.195)	2.631 (0.200)	2.535 (0.200)
Y_{it-2} Intercept (α_2)	1.322 (0.216)	1.495 (0.221)	1.083 (0.212)	1.488 (0.220)	1.423 (0.219)
Dropout					
Y_{i2} Intercept (ϕ_0)	2.024 (0.152)				
Intercept (ϕ_1)	1.325 (0.201)	1.288 (0.170)	1.811 (0.178)	1.361 (0.170)	1.257 (0.170)
PLAC (ϕ_2)	-0.293 (0.269)	-0.281 (0.229)	-0.407 (0.240)	-0.306 (0.227)	-0.278 (0.230)
RISP (ϕ_3)	0.425 (0.236)	0.283 (0.203)	0.499 (0.210)	0.167 (0.202)	0.443 (0.204)
Y_{it-1} (ϕ_4)	1.609 (0.288)	0.478 (0.232)	2.076 (0.231)	0.558 (0.231)	0.818 (0.232)
Y_{it-2} (ϕ_5)	-0.706 (0.250)				

Appendix E

SUPPLEMENTAL PROOF, CHAPTER 7

SHOW THAT: $P(\mathrm{R}_{ij} = 1|\mathrm{R}_{ij-1} = 1, S_i) = P(\mathrm{R}_{ij} = 1|\mathrm{R}_{ij-1} = 1, S_i > t_j)$

MAR assumption:	$P(\mathrm{R}_{ij} = 1\|S_i) = P(\mathrm{R}_{ij} = 1\|S_i > t_j)$	(suppressed Y_i)
other assumptions:	$P(\mathrm{R}_{i1} = 1) = 1$	(Y_{i1} never missing)
	$P(\mathrm{R}_{ij} = 1\|\mathrm{R}_{ij-1} = 0) = 0$	(monotone dropout)
	$P(\mathrm{R}_{ij} = 1\|\mathrm{H}_{ij}^{(\mathrm{R})}) = P(\mathrm{R}_{ij} = 1\|\mathrm{R}_{ij-1})$	(monotone dropout)

LEMMA: for $k > 0$, $P(\mathrm{R}_{ij} = 1|S_i > t_{j+k}) = P(\mathrm{R}_{ij} = 1|S_i > t_j)$

$$
\begin{aligned}
P(\mathrm{R}_{ij} = 1|S_i > t_{j+k}) &= \frac{\sum_{m>j+k} P(\mathrm{R}_{ij} = 1|S_i = t_m)P(S_i = t_m)}{\sum_{m>j+k} P(S_i = t_m)} \\
&\underset{\text{MAR}}{=} \frac{\sum_{m>j+k} P(\mathrm{R}_{ij} = 1|S_i > t_j)P(S_i = t_m)}{\sum_{m>j+k} P(S_i = t_m)} \\
&= P(\mathrm{R}_{ij} = 1|S_i > t_j) \quad \blacksquare
\end{aligned}
$$

PROOF BY INDUCTION:

j=2

$$P(\mathrm{R}_{i2}=1|S_i) \overset{\text{MAR}}{=} P(\mathrm{R}_{i2}=1|S_i > t_2)$$

j=3

$$P(\mathrm{R}_{i3}=1|S_i) \overset{\text{MAR}}{=} P(\mathrm{R}_{i3}=1|S_i > t_3)$$

$$\Longrightarrow$$

$$P(\mathrm{R}_{i3}{=}1|\mathrm{R}_{i2}{=}1, S_i) \cdot P(\mathrm{R}_{i2}{=}1|S_i) = P(\mathrm{R}_{i3}{=}1|\mathrm{R}_{i2}{=}1, S_i > t_3) \cdot P(\mathrm{R}_{i2}{=}1|S_i > t_3)$$

$$\Longrightarrow$$

$$P(\mathrm{R}_{i3}{=}1|\mathrm{R}_{i2}{=}1, S_i) \cdot \underbrace{P(\mathrm{R}_{i2}{=}1|S_i > t_2)}_{\text{MAR}} = P(\mathrm{R}_{i3}{=}1|\mathrm{R}_{i2}{=}1, S_i > t_3) \cdot \underbrace{P(\mathrm{R}_{i2}{=}1|S_i > t_2)}_{\text{Lemma}}$$

$$\Longrightarrow P(\mathrm{R}_{i3}=1|\mathrm{R}_{i2}=1, S_i) = P(\mathrm{R}_{i3}=1|\mathrm{R}_{i2}=1, S_i > t_3)$$

Assume that for all $j < j+1$,

$$P(\mathrm{R}_{ij}=1|\mathrm{R}_{ij-1}=1, S_i) = P(\mathrm{R}_{ij}=1|\mathrm{R}_{ij-1}=1, S_i > t_j).$$

Show that:

$$P(\mathrm{R}_{ij+1}=1|\mathrm{R}_{ij}=1, S_i) = P(\mathrm{R}_{ij+1}=1|\mathrm{R}_{ij}=1, S_i > t_{j+1}).$$

$$\begin{aligned}
P(\mathrm{R}_{ij+1}=1|\mathrm{R}_{ij}=1, S_i) &= \frac{P(\mathrm{R}_{ij+1}=1, \mathrm{R}_{ij}=1|S_i)}{P(\mathrm{R}_{ij}=1|S_i)} \\
&= \frac{P(\mathrm{R}_{ij+1}=1|S_i)}{P(\mathrm{R}_{ij}=1|S_i)} \\
&\overset{\text{MAR}}{=} \frac{P(\mathrm{R}_{ij+1}=1|S_i > t_{j+1})}{P(\mathrm{R}_{ij}=1|S_i > t_j)} \\
&= \frac{P(\mathrm{R}_{ij+1}=1|\mathrm{R}_{ij}=1, S_i > t_{j+1}) \cdots P(\mathrm{R}_{i2}=1|S_i > t_{j+1})}{P(\mathrm{R}_{ij}=1|\mathrm{R}_{ij-1}=1, S_i > t_j) \cdots P(\mathrm{R}_{i2}=1|S_i > t_j)} \\
&\overset{\text{Lemma}}{=} \frac{P(\mathrm{R}_{ij+1}=1|\mathrm{R}_{ij}=1, S_i > t_{j+1}) \cdots P(\mathrm{R}_{i2}=1|\mathrm{R}_{i1}=1, S_i > t_2)}{P(\mathrm{R}_{ij}=1|\mathrm{R}_{ij-1}=1, S_i > t_j) \cdots P(\mathrm{R}_{i2}=1|\mathrm{R}_{i1}=1, S_i > t_2)} \\
&= P(\mathrm{R}_{ij+1}=1|\mathrm{R}_{ij}=1, S_i > t_{j+1}) \quad \blacksquare
\end{aligned}$$

VITA

Brenda F. Kurland received her Bachelor of Arts degree in Psychology *magna cum laude* from Harvard/Radcliffe Colleges in 1992; an Ed.M. in Human Development and Psychology, with a specialization in Methodology in Developmental Research in 1994 from the Harvard Graduate School of Education; and a M.S. in Biostatistics from the University of Washington in 2000.