

BAYESIAN NONPARAMETRIC AND SEMI-PARAMETRIC METHODS
FOR INCOMPLETE LONGITUDINAL DATA

By

CHENGUANG WANG

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

2010

© 2010 Chenguang Wang

I dedicate this to my family

ACKNOWLEDGMENTS

First and foremost, I would like to express the deepest appreciation to my advisor, Professor Michael J. Daniels. Without his extraordinary guidance and persistent help, I will never be able to be where I am. I admire his wisdom, his knowledge and his commitment to the highest standard. It has been truly an honor to work with him.

I wish to specially thank Professor Daniel O. Scharfstein of Johns Hopkins for his encouragement and crucial contribution to the research. I will always bear in mind the advice he gave: just relax and enjoy the learning process.

I would like to thank my committee members, Professor Malay Ghosh, Dr. Brett Presnell, and Dr. Almut Winterstein, who have provided abundant support and valuable insights over the entire process throughout the classes, exams and dissertation.

Many thanks go in particular to Professor Rongling Wu of Pennsylvania State University. From Professor Wu, I started learning what I wanted for my career. I also gratefully thank Professor Myron Chang, Professor Linda Young, Dr. Meenakshi Devidas and Dr. Gregory Campbell of FDA. I am fortunate to have their support at those critical moments of my career.

Finally, I would like to thank my wife, my son, my soon-to-be-born baby, my parents and my parents-in-law. It is only because of you that I have been able to keep working toward this dream I have.

TABLE OF CONTENTS

	<u>page</u>
ACKNOWLEDGMENTS	4
LIST OF TABLES	8
LIST OF FIGURES	9
ABSTRACT	10
CHAPTER	
1 INTRODUCTION	12
1.1 Missing Data Concepts and Definitions	12
1.2 Likelihood-Based Methods	16
1.3 Non-Likelihood Methods	19
1.4 Intermittent Missingness	20
1.5 Identifying Restrictions in Pattern Mixture Models	22
1.6 Dissertation Goals	24
2 A BAYESIAN SHRINKAGE MODEL FOR LONGITUDINAL BINARY DATA WITH DROP-OUT	26
2.1 Introduction	26
2.1.1 Breast Cancer Prevention Trial	26
2.1.2 Informative Drop-Out in Longitudinal Studies	27
2.1.3 Outline	30
2.2 Data Structure and Notation	30
2.3 Assumptions	31
2.4 Identifiability	32
2.5 Modeling	35
2.6 Prior Specification and Posterior Computation	35
2.6.1 Shrinkage Priors	36
2.6.2 Prior of Sensitivity Parameters	37
2.6.3 Posterior Computation	40
2.7 Assessment of Model Performance via Simulation	40
2.8 Application: Breast Cancer Prevention Trial (BCPT)	42
2.8.1 Model Fit and Shrinkage Results	42
2.8.2 Inference	43
2.9 Summary and Discussion	43
2.10 Acknowledgments	45
2.11 Tables and Figures	45

3	A BETA-BINOMIAL BAYESIAN SHRINKAGE MODEL FOR INTERMITTENT MISSINGNESS LONGITUDINAL BINARY DATA	54
3.1	Introduction	54
3.1.1	Intermittent Missing Data	54
3.1.2	Computational Issues	55
3.1.3	Outline	55
3.2	Notation, Assumptions and Identifiability	56
3.3	Modeling, Prior Specification and Posterior Computation	58
3.3.1	Modeling	58
3.3.2	Shrinkage Prior	58
3.3.3	Prior of Sensitivity Parameters	60
3.3.4	Posterior Computation	61
3.4	Assessment of Model Performance via Simulation	61
3.5	Application: Breast Cancer Prevention Trial (BCPT)	62
3.5.1	Model Fit	62
3.5.2	Inference	63
3.5.3	Sensitivity of Inference to the Priors	64
3.6	Summary and Discussion	65
3.7	Tables and Figures	65
3.8	Appendix	73
4	A NOTE ON MAR, IDENTIFYING RESTRICTIONS, AND SENSITIVITY ANALYSIS IN PATTERN MIXTURE MODELS	77
4.1	Introduction	77
4.2	Existence of MAR under Multivariate Normality within Pattern	79
4.3	Sequential Model Specification and Sensitivity Analysis under MAR	83
4.4	Non-Future Dependence and Sensitivity Analysis under Multivariate Normality within Pattern	85
4.5	MAR and Sensitivity Analysis with Multivariate Normality on the Observed-Data Response	87
4.6	Example: Growth Hormone Study	89
4.7	ACMV Restrictions and Multivariate Normality with Baseline Covariates	91
4.7.1	Bivariate Case	91
4.7.2	Multivariate Case	93
4.8	Summary	95
4.9	Tables	96
4.10	Appendix	98
5	DISCUSSION: FUTURE APPLICATION OF THE BAYESIAN NONPARAMETRIC AND SEMI-PARAMETRIC METHODS	103
5.1	Summary	103
5.2	Extensions to Genetics Mapping	103
5.3	Extensions to Causal Inference	105

5.3.1	Causal Inference Introduction	105
5.3.2	Data and Notation	107
5.3.3	Missing Data Mechanism	108
5.3.4	Causal Inference Assumption	109
5.3.5	Stochastic Survival Monotonicity Assumption	111
5.3.6	Summary of Causal Inference	112
5.4	Figures	112
REFERENCES		115
BIOGRAPHICAL SKETCH		123

LIST OF TABLES

<u>Table</u>	<u>page</u>
2-1 Relative Risks to be Elicited	45
2-2 Percentiles of Relative Risks Elicited	45
2-3 Simulation Scenario	46
2-4 Simulation Results: MSE ($\times 10^3$). P and T represent placebo and tamoxifen arms, respectively.	47
2-5 Patients Cumulative Drop Out Rate	47
3-1 Missingness by Scheduled Measurement Time	65
3-2 Simulation Results: MSE ($\times 10^3$). P and T represent placebo and tamoxifen arms, respectively.	66
3-3 Sensitivity to the Elicited Prior	67
3-4 Sensitivity to the Elicited Prior	68
4-1 Growth Hormone Study: Sample mean (standard deviation) stratified by dropout pattern.	96
4-2 Growth Hormone Study: Posterior mean (standard deviation)	97

LIST OF FIGURES

<u>Figure</u>	<u>page</u>
2-1 Extrapolation of the elicited relative risks.	48
2-2 Prior Density of τp	49
2-3 Model Fit	50
2-4 Model Shrinkage	51
2-5 Posterior distribution of $P[Y_7 = 1 Z = z]$. Black and gray lines represent tamoxifen and placebo arms, respectively. Solid and dashed lines are for MNAR and MAR, respectively.	52
2-6 Posterior mean and 95% credible interval of difference of $P[Y_j = 1 Z = z]$ between placebo and tamoxifen arms. The gray and white boxes are for MAR and MNAR, respectively.	53
3-1 Model Fit	69
3-2 Shrinkage	70
3-3 Posterior distribution of $P[Y_7 = 1 Z = z]$. Black and gray lines represent tamoxifen and placebo arms, respectively. Solid and dashed lines are for MNAR and MAR, respectively.	71
3-4 Posterior mean and 95% credible interval of difference of $P[Y_j = 1 Z = z]$ between placebo and tamoxifen arms. The gray and white boxes are for MAR and MNAR, respectively.	72
5-1 Contour and Perspective Plots of a Bivariate Density	113
5-2 Illustration of $p_{c,t}$	114

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

BAYESIAN NONPARAMETRIC AND SEMI-PARAMETRIC METHODS
FOR INCOMPLETE LONGITUDINAL DATA

By

Chenguang Wang

August 2010

Chair: Michael J. Daniels

Major: Statistics

We consider inference in randomized longitudinal studies with missing data that is generated by skipped clinic visits and loss to follow-up. In this setting, it is well known that full data estimands are not identified unless unverified assumptions are imposed. Sensitivity analysis that assesses the sensitivity of model-based inferences to such assumptions is often necessary.

In Chapters 2 and 3, we posit an exponential tilt model that links non-identifiable distributions and identifiable distributions. This exponential tilt model is indexed by non-identified parameters, which are assumed to have an informative prior distribution, elicited from subject-matter experts. Under this model, full data estimands are shown to be expressed as functionals of the distribution of the observed data. We propose two different saturated models for the observed data distribution, as well as shrinkage priors to avoid the curse of dimensionality. The two procedures provide researchers different strategies for reducing the dimension of parameter space. We assume a non-future dependence model for the drop-out mechanism and partial ignorability for the intermittent missingness. In a simulation study, we compare our approach to a fully parametric and a fully saturated model for the distribution of the observed data. Our methodology is motivated by, and applied to, data from the Breast Cancer Prevention Trial.

In Chapter 4, we discuss pattern mixture models. Pattern mixture modeling is a popular approach for handling incomplete longitudinal data. Such models are not identifiable by construction. Identifying restrictions are one approach to mixture model identification (Daniels and Hogan, 2008; Kenward et al., 2003; Little, 1995; Little and Wang, 1996; Thijs et al., 2002) and are a natural starting point for missing not at random sensitivity analysis (Daniels and Hogan, 2008; Thijs et al., 2002). However, when the pattern specific models are multivariate normal (MVN), identifying restrictions corresponding to missing at random may not exist. Furthermore, identification strategies can be problematic in models with covariates (e.g. baseline covariates with time-invariant coefficients). In this paper, we explore conditions necessary for identifying restrictions that result in missing at random (MAR) to exist under a multivariate normality assumption and strategies for identifying sensitivity parameters for sensitivity analysis or for a fully Bayesian analysis with informative priors. A longitudinal clinical trial is used for illustration of sensitivity analysis. Problems caused by baseline covariates with time-invariant coefficients are investigated and an alternative identifying restriction based on residuals is proposed as a solution.

CHAPTER 1 INTRODUCTION

The problem of incomplete data is frequently confronted by statisticians, especially in longitudinal studies. The most common type of incomplete data is *missing data*, in which each data value is either perfectly known or completely unknown. In other situations, data are partially missing and partially observed. Examples include rounded data and censored data, etc.. This type of incomplete data is referred to as *coarse data*. *Missing data* can be viewed as a special case of *coarse data* (Heitjan and Rubin, 1991). In both cases, the incompleteness occurs because we observe only a subset of the complete data, which includes the true, unobservable data. In this dissertation, *missing data* including the drop-out missingness, in which case subjects missing a measurement will not return to study at the next follow-up, and the intermittent missingness, in which case the missing visit is followed by an observed measurement.

1.1 Missing Data Concepts and Definitions

If the missingness does not happen “at random” and the missingness process is ignored, biased inferences will often occur. Until the 1970s, most of the methods for handling missing values in the statistics literature ignored the missingness mechanism by deleting the incomplete units. Complete-case analysis, also known as case deletion, confines the analysis to cases that have all variables observed. Available-case analysis, also known as partial deletion, uses all values observed for univariate analysis. Both approaches are only valid under the strong assumption that the missingness is completely at random (MCAR), i.e. the missingness is independent of the response. In situations when MCAR doesn’t hold, it is possible to adjust the selection bias caused by case deletion by reweighting the remaining cases (Little and Rubin, 1987, chapter 4); however, this method is inefficient. Another common approach is single imputation, that is, filling in a single value for each missing value. The advantage of single imputation is that it does not delete any units and after the imputation, standard methods for complete

data can be applied on the filled-in data. However, single imputation does not reflect the uncertainty of the missing value. Multiple imputation was proposed to address this flaw by imputing several values for each missing response (Rubin, 1987).

For notation, let $\mathbf{y} = \{y_1, \dots, y_J\}$ denote the full data response vector of outcome, possibly partially observed. Let $\mathbf{r} = \{r_1, r_2, \dots, r_J\}$ denote the missing data indicator, with $r_j = 0$ if y_j is missing and 1 if y_j is observed. Let \mathbf{x} denote the covariates. Let \mathbf{y}_{obs} and \mathbf{y}_{mis} denote the observed and missing response data, respectively. Let ω be the parameters indexing the full data model $p(\mathbf{y}, \mathbf{r})$, $\theta(\omega)$ be the parameters indexing the full data response model $p(\mathbf{y})$, and $\phi(\omega)$ be the parameters indexing the missing data mechanism model $p(\mathbf{r}|\mathbf{y})$.

The common assumptions about the missing data mechanism are as follows.

Little and Rubin's taxonomy: Rubin (1976) and Little and Rubin (1987) developed a hierarchy for missing data mechanisms by classifying the relationship between missingness and the response data.

Definition 1.1. Missing responses are *missing completely at random* (MCAR) if

$$p(\mathbf{r}|\mathbf{y}_{\text{obs}}, \mathbf{y}_{\text{mis}}, \mathbf{x}; \phi(\omega)) = p(\mathbf{r}|\mathbf{x}; \phi(\omega))$$

for all \mathbf{x} and ω .

Definition 1.2. Missing responses are *missing at random* (MAR) if

$$p(\mathbf{r}|\mathbf{y}_{\text{obs}}, \mathbf{y}_{\text{mis}}, \mathbf{x}; \phi(\omega)) = p(\mathbf{r}|\mathbf{y}_{\text{obs}}, \mathbf{x}; \phi(\omega))$$

for all \mathbf{x} and ω .

Note that MAR holds if and only if $p(\mathbf{y}_{\text{mis}}|\mathbf{y}_{\text{obs}}, \mathbf{r}) = p(\mathbf{y}_{\text{mis}}|\mathbf{y}_{\text{obs}})$. The proof is as follows:

Proof:

1. Suppose MAR holds. Then we have

$$p(\mathbf{r}|\mathbf{y}_{\text{mis}}, \mathbf{y}_{\text{obs}}) = p(\mathbf{r}|\mathbf{y}_{\text{obs}})$$

and we can derive that

$$\begin{aligned} \rho(\mathbf{y}_{\text{mis}}|\mathbf{y}_{\text{obs}}, \mathbf{r}) &= \frac{\rho(\mathbf{y}_{\text{mis}}, \mathbf{r}|\mathbf{y}_{\text{obs}})}{\rho(\mathbf{r}|\mathbf{y}_{\text{obs}})} = \frac{\rho(\mathbf{r}|\mathbf{y}_{\text{mis}}, \mathbf{y}_{\text{obs}})\rho(\mathbf{y}_{\text{mis}}|\mathbf{y}_{\text{obs}})}{\rho(\mathbf{r}|\mathbf{y}_{\text{obs}})} \\ &= \frac{\rho(\mathbf{r}|\mathbf{y}_{\text{obs}})\rho(\mathbf{y}_{\text{mis}}|\mathbf{y}_{\text{obs}})}{\rho(\mathbf{r}|\mathbf{y}_{\text{obs}})} = \rho(\mathbf{y}_{\text{mis}}|\mathbf{y}_{\text{obs}}, \mathbf{r}). \end{aligned}$$

2. To show the reverse direction, note that

$$\begin{aligned} \rho(\mathbf{r}|\mathbf{y}_{\text{mis}}, \mathbf{y}_{\text{obs}}) &= \frac{\rho(\mathbf{r}, \mathbf{y}_{\text{mis}}|\mathbf{y}_{\text{obs}})}{\rho(\mathbf{y}_{\text{mis}}|\mathbf{y}_{\text{obs}})} = \frac{\rho(\mathbf{y}_{\text{mis}}|\mathbf{r}, \mathbf{y}_{\text{obs}})\rho(\mathbf{r}|\mathbf{y}_{\text{obs}})}{\rho(\mathbf{y}_{\text{mis}}|\mathbf{y}_{\text{obs}})} \\ &= \frac{\rho(\mathbf{y}_{\text{mis}}|\mathbf{y}_{\text{obs}})\rho(\mathbf{r}|\mathbf{y}_{\text{obs}})}{\rho(\mathbf{y}_{\text{mis}}|\mathbf{y}_{\text{obs}})} = \rho(\mathbf{r}|\mathbf{y}_{\text{obs}}). \end{aligned}$$

This completes the proof. □

Definition 1.3. Missing responses are *missing not at random* (MNAR) if

$$\rho(\mathbf{r}|\mathbf{y}_{\text{obs}}, \mathbf{y}_{\text{mis}}, \mathbf{x}; \phi(\boldsymbol{\omega})) \neq \rho(\mathbf{r}|\mathbf{y}_{\text{obs}}, \mathbf{y}'_{\text{mis}}, \mathbf{x}; \phi(\boldsymbol{\omega}))$$

for some $\mathbf{y}_{\text{mis}} \neq \mathbf{y}'_{\text{mis}}$.

Ignorability: Under certain condition, the missingness process can be left unspecified for the inference on the response model parameter $\theta(\boldsymbol{\omega})$ (Laird, 1988).

This condition is called *ignorability* (Rubin, 1976).

Definition 1.4. The missing data mechanism is *ignorable* if

1. The missing data mechanism is MAR.
2. The parameters of the full data response model, $\theta(\boldsymbol{\omega})$ and the parameters of the missingness model are distinguishable, i.e. the full data parameter $\boldsymbol{\omega}$ can be decomposed as $(\theta(\boldsymbol{\omega}), \phi(\boldsymbol{\omega}))$.
3. The parameters $\theta(\boldsymbol{\omega})$ and $\phi(\boldsymbol{\omega})$ are *a priori* independent, i.e. $p(\theta(\boldsymbol{\omega}), \phi(\boldsymbol{\omega})) = p(\theta(\boldsymbol{\omega}))p(\phi(\boldsymbol{\omega}))$.

Full data models that do not satisfy Definition 1.4 have *non-ignorable* missingness.

Under ignorability, posterior inference on parameters $\theta(\omega)$ can be based on the observed data response likelihood

$$L(\theta(\omega)|\mathbf{y}_{\text{obs}}) \propto \prod_{i=1}^n \int p_i(\mathbf{y}_{\text{obs}}, \mathbf{y}_{\text{mis}}|\theta(\omega)) d\mathbf{y}_{\text{mis}}.$$

We show this below,

$$\begin{aligned} L(\omega|\mathbf{y}_{\text{obs}}, \mathbf{r}) &= L(\theta(\omega), \phi(\omega)|\mathbf{y}_{\text{obs}}, \mathbf{r}) \\ &= \int p(\mathbf{y}_{\text{obs}}, \mathbf{y}_{\text{mis}}, \mathbf{r}|\theta(\omega), \phi(\omega)) d\mathbf{y}_{\text{mis}} \\ &= \int p(\mathbf{r}|\mathbf{y}_{\text{obs}}, \phi(\omega)) p(\mathbf{y}_{\text{obs}}, \mathbf{y}_{\text{mis}}|\theta(\omega)) d\mathbf{y}_{\text{mis}} \\ &= p(\mathbf{r}|\mathbf{y}_{\text{obs}}, \phi(\omega)) \int p(\mathbf{y}_{\text{obs}}, \mathbf{y}_{\text{mis}}|\theta(\omega)) d\mathbf{y}_{\text{mis}} \\ &= p(\mathbf{r}|\mathbf{y}_{\text{obs}}, \phi(\omega)) p(\mathbf{y}_{\text{obs}}|\theta(\omega)) \\ &= L(\phi(\omega)|\mathbf{r}, \mathbf{y}_{\text{obs}}) L(\theta(\omega)|\mathbf{y}_{\text{obs}}), \end{aligned}$$

and furthermore,

$$p(\omega|\mathbf{y}_{\text{obs}}, \mathbf{r}) \propto p(\omega) L(\omega|\mathbf{y}_{\text{obs}}, \mathbf{r}) = p(\phi(\omega)) L(\phi(\omega)|\mathbf{r}, \mathbf{y}_{\text{obs}}) p(\theta(\omega)) L(\theta(\omega)|\mathbf{y}_{\text{obs}}).$$

Therefore,

$$p(\theta(\omega)|\mathbf{y}_{\text{obs}}, \mathbf{r}) \propto p(\theta(\omega)) L(\theta(\omega)|\mathbf{y}_{\text{obs}})$$

and the posterior inference of $\theta(\omega)$ can be based on observed response likelihood

$$L(\theta(\omega)|\mathbf{y}_{\text{obs}}).$$

Non-future Dependence: For cases with monotone missingness, i.e. $r_j = 0$ implies $r_{j'} = 0$ for $j' > j$, [Kenward et al. \(2003\)](#) defined the term *non-future dependence*.

Definition 1.5. If the missingness is monotone, the MDM is *non-future dependent* if

$$p(\mathbf{r}|\mathbf{y}_{\text{obs}}, \mathbf{y}_{\text{mis}}, \mathbf{x}; \phi(\omega)) = p(\mathbf{r}|\mathbf{y}_{\text{obs}}, y_C, \mathbf{x}; \phi(\omega))$$

with $C = \min_j \{r_j = 0\}$.

Non-future dependence assumes that missingness only depends on observed data and the current missing value. It can be viewed as a special case of MNAR and an extension of MAR

1.2 Likelihood-Based Methods

Likelihood based methods handle the missing values by integrating them out of the likelihood function, instead of deletion or explicitly filling in values. The general strategy is to model the joint distribution of a response and the missingness process ([Hogan and Laird, 1997b](#)). Likelihood-based models for missing data are distinguished by the way the joint distribution of the outcome and missing data processes are factorized. They can be classified as selection models, pattern-mixture models, and shared-parameter models.

Selection model: Selection models factor the full-data distribution as

$$p(\mathbf{y}, \mathbf{r}|\boldsymbol{\omega}) = p(\mathbf{r}|\mathbf{y}, \phi(\boldsymbol{\omega}))p(\mathbf{y}|\theta(\boldsymbol{\omega})).$$

The term "selection" was first introduced in the economics literature for modeling sample selection bias; that is different responses have different probabilities of being selected into a sample. [Heckman \(1979a,b\)](#) used a bivariate response \mathbf{Y} with missing Y_2 as an example and showed that in general it's critical to answer the question "why are the data missing" by modeling the missingness of Y_{2i} as a function of observed Y_{1i} (for subject i). [Diggle and Kenward \(1994\)](#) extended the Heckman model to longitudinal studies and modeled the drop-out process by logistic regression such as

$$\text{logit}(r_j = 0 | r_{j-1} = 1, \mathbf{y}) = \mathbf{y}'\boldsymbol{\beta}.$$

The Diggle and Kenward model has been adopted and extended by many researchers by (mostly) proposing different full data response models ([Albert, 2000](#); [Baker, 1995](#); [Fitzmaurice et al., 1995](#); [Heagerty, 2002](#); [Kurland and Heagerty, 2004](#)).

Pattern mixture model: Pattern mixture models factor the full-data distribution as

$$p(\mathbf{y}, \mathbf{r}|\boldsymbol{\omega}) = p(\mathbf{y}|\mathbf{r}, \theta(\boldsymbol{\omega}))p(\mathbf{r}|\phi(\boldsymbol{\omega})).$$

Rubin (1977) introduced the idea of modeling respondents and nonrespondents in surveys separately and using subjective priors to relate respondents' and nonrespondents' model parameters. Little (1993, 1994) explored pattern mixture models in discrete time settings. Specifically, different identifying restrictions (see Section 1.5) were proposed to identify the full-data model. When the number of dropout patterns is large and pattern-specific parameters will be weakly identified by identifying restrictions, Roy (2003) and Roy and Daniels (2008) proposed to use latent-class model for dropout classes. When the dropout time is continuous and the mixture of patterns is infinite, Hogan et al. (2004) proposed to model the response given dropout by a varying coefficient model where regression coefficients were unspecified, non-parametric functions of dropout time. For time-event data with informative censoring, Wu and Bailey (1988, 1989) and Hogan and Laird (1997a) developed random effects mixture models. Fitzmaurice and Laird (2000a) generalized Wu and Bailey and Hogan and Laird approach for discrete, ordinal and count data by using generalized linear mixture models and GEE approach for statistical inference. Daniels and Hogan (2000) proposed a parameterization of the pattern mixture model for continuous data. Sensitivity analysis can be done on the additive (location) and multiplicative (scale) terms. Forster and Smith (1998) considered a pattern mixture model for a single categorical response with categorical covariates. Bayesian approaches were employed for non-ignorable missingness.

Shared parameter model: Shared parameter models factorize the full-data model as

$$p(\mathbf{y}, \mathbf{r}|\boldsymbol{\omega}) = \int p(\mathbf{y}, \mathbf{r}, |b, \boldsymbol{\omega})p(b|\boldsymbol{\omega})db,$$

where b are subject-specific random effects. It is usually assumed that \mathbf{y} and \mathbf{r} are independent conditionally on b .

Wu and Carroll (1988) presented a shared parameter random effects model for continuous responses and informative censoring, in which individual effects are taken into account as intercepts and slopes for modeling the censoring process. DeGruttola and Tu (1994) extended Wu and Carroll's model to allow general covariates. Follmann and Wu (1995) developed generalized linear model for response and proposed an approximation algorithm for the joint full-data model for inference. Faucett and Thomas (1996) and Wulfsohn and Tsiatis (1997) proposed to jointly model the continuous covariate over time and relate the covariates to the response simultaneously. Henderson et al. (2000) generalized the joint modeling approach by using two correlated Gaussian random processes for covariates and response. Ten Have et al. (1998, 2000) proposed a shared parameter mixed effects logistic regression model for longitudinal ordinal data. Recently, Yuan and Little (2009) proposed a mixed-effect hybrid model allows the missingness and response to be conditionally dependent given random effects.

Sensitivity analysis for missing not at random: Sensitivity analysis is critical in longitudinal analysis of incomplete data. The full-data model can be factored into an extrapolation model and an observed data model,

$$p(\mathbf{y}, \mathbf{r}|\boldsymbol{\omega}) = p(\mathbf{y}_{\text{mis}}|\mathbf{y}_{\text{obs}}, \mathbf{r}, \boldsymbol{\omega}_E)p(\mathbf{y}_{\text{obs}}, \mathbf{r}|\boldsymbol{\omega}_I),$$

where $\boldsymbol{\omega}_E$ are parameters indexing the extrapolation model and $\boldsymbol{\omega}_I$ are parameters indexing the observed data model and are identifiable from observed data (Daniels and Hogan, 2008). Full-data model inference requires unverifiable assumptions about the extrapolation model $p(\mathbf{y}_{\text{mis}}|\mathbf{y}_{\text{obs}}, \mathbf{r}, \boldsymbol{\omega}_E)$. A sensitivity analysis explores the sensitivity of inferences of interest about the full data response model to unverifiable assumptions about the extrapolation model. This is typically done by varying sensitivity parameters, which we define next (Daniels and Hogan, 2008).

Definition 1.6. Let $p(\mathbf{y}, \mathbf{r}|\omega)$ be a full data model with extrapolation factorization

$$p(\mathbf{y}, \mathbf{r}|\omega) = p(\mathbf{y}_{\text{mis}}|\mathbf{y}_{\text{obs}}, \mathbf{r}, \omega_E)p(\mathbf{y}_{\text{obs}}, \mathbf{r}|\omega_I).$$

Suppose there exists a reparameterization $\xi(\omega) = (\xi_S, \xi_M)$ such that

1. ξ_S is a non-constant function of ω_E ,
2. The observed likelihood $L(\xi_S, \xi_M|\mathbf{y}_{\text{obs}}, \mathbf{r})$ is a constant as a function of ξ_S ,
3. Given ξ_S fixed, $L(\xi_S, \xi_M|\mathbf{y}_{\text{obs}}, \mathbf{r})$ is a non-constant function of ξ_M

then ξ_S is a *sensitivity parameter*.

Unfortunately, fully parametric selection models and shared parameter models do *not* allow sensitivity analysis as sensitivity parameters cannot be found (Daniels and Hogan, 2008, Chapter 8). Examining sensitivity to distributional assumptions, e.g., random effects, will provide *different* fits to the observed data, $(\mathbf{y}_{\text{obs}}, \mathbf{r})$. In such cases, a sensitivity analysis cannot be done since varying the distributional assumptions does not provide equivalent fits to the observed data (Daniels and Hogan, 2008). It then becomes an exercise in model selection.

Fully Bayesian analysis allows researchers to have a single conclusion by admitting prior beliefs about the sensitivity parameters. For continuous responses, Lee and Berger (2001) built a semiparametric Bayesian selection model which has strong distributional assumption for the response but weak assumption on missing data mechanism. Scharfstein et al. (2003) on the other hand, placed strong parametric assumptions on missing data mechanism but minimal assumptions on the response outcome.

1.3 Non-Likelihood Methods

In non-likelihood approaches, the joint distribution of the outcomes is typically modeled semiparametrically and estimating equations are used for inference. Liang and Zeger (1986) proposed generalized estimating equations (GEE) whose solution

is consistent if the marginal mean of response is correctly specified. However, inference based on GEE is only valid under MCAR. [Robins et al. \(1995\)](#) proposed inverse-probability of censoring weighted generalized estimating equations (IPCW-GEE) approach, which reweights each individual's contribution to the usual GEE by the estimated probability of drop-out. IPCW-GEE will lead to consistent estimation when the missingness is MAR. However, both GEE and IPCW-GEE can result in biased estimation under MNAR.

[Rotnitzky et al. \(1998a, 2001\)](#), [Scharfstein et al. \(2003\)](#) and [Schulman et al. \(1999\)](#) adopted semiparametric selection modeling approaches, in which the model for drop-out is indexed by interpretable sensitivity parameters that express departures from MAR. For such approaches, the inference results depend on the choice of unidentified, yet interpretable, sensitivity analysis parameters.

1.4 Intermittent Missingness

Intermittent missingness occurs when a missing value is followed by an observed value. The existence of intermittent missing values increases exponentially the number of missing patterns that need to be properly modeled. Thus, handling informative intermittent missing data is methodologically and computationally challenging and, as a result, the statistics literature is limited.

One approach to handle intermittent missingness is to consider a “monotonized” dataset, whereby all observed values on an individual after their first missingness are deleted [Land et al. \(2002\)](#). However, this increases the “dropout” rate, loses efficiency, and may introduce bias.

Other methods in the literature often adopt a likelihood approach and rely on strong parametric assumptions. For example, [Troxel et al. \(1998\)](#), [Albert \(2000\)](#) and [Ibrahim et al. \(2001\)](#) suggested a selection model approach. [Albert et al. \(2002\)](#) used a shared latent autoregressive process model. [Lin et al. \(2004\)](#) employed latent class pattern mixture model.

Semiparametric methods have been proposed by [Troxel et al. \(1998\)](#) and [Vansteelandt et al. \(2007\)](#). [Troxel et al. \(1998\)](#) proposed a marginal model and introduced a pseudo-likelihood estimation procedure. [Vansteelandt et al. \(2007\)](#) extended the ideas of [Rotnitzky et al. \(1998b\)](#), [Scharfstein et al. \(1999\)](#) and [Rotnitzky et al. \(2001\)](#) to non-monotone missing data that assume (exponentially tilted) extensions of sequential explainability and specified parametric models for certain conditional means.

Most related to the approach we will use in Chapter 3 are the (partial ignorability) assumptions formalized in [Harel and Schafer \(2009\)](#) that partition the missing data and allow one (or more) of the partitions to be ignored given the other partition(s) and the observed data. Specifically, [Harel and Schafer \(2009\)](#) defined a missing data mechanism to be partially missing at random if

$$p(\mathbf{r}|\mathbf{y}_{\text{obs}}, \mathbf{y}_{\text{mis}}, g(\mathbf{r}), \mathbf{x}; \phi(\boldsymbol{\omega})) = p(\mathbf{r}|\mathbf{y}_{\text{obs}}, g(\mathbf{r}), \mathbf{x}; \phi(\boldsymbol{\omega}))$$

where $g(\cdot)$ denotes a summary function of \mathbf{r} and can be chosen based on what aspects of \mathbf{r} are related to missing values. These assumptions are similar to the sequential explainability assumption reviewed in [Vansteelandt et al. \(2007\)](#).

In this dissertation, we explicitly partition the missing data indicator vector \mathbf{r} into $\{\mathbf{r}_s, s\}$, where $s = \max_t \{r_t = 1\}$ denotes the last time point a response was observed, i.e. the “survival” time, and $\mathbf{r}_s = \{r_t : t < s\}$ denotes the missing data indicators recorded prior to the drop-out time. With this partition, we define partial missing at random as follows:

Definition 1.7. Missing responses are *partially missing at random* if

$$p(\mathbf{r}_s|\mathbf{y}_{\text{obs}}, \mathbf{y}_{\text{mis}}, s, \mathbf{x}; \phi(\boldsymbol{\omega})) = p(\mathbf{r}_s|\mathbf{y}_{\text{obs}}, s, \mathbf{x}; \phi(\boldsymbol{\omega})).$$

This can be viewed as Harel and Schafer’s definition with $g(\mathbf{r})$ chosen to be the survival time.

In the following Chapters, we first propose a model that admits identification of the treatment-specific distributions of the trajectory of longitudinal binary outcomes when there is drop-out, but no intermittent observations. We then obtain identification with intermittent missing observations by assuming, that within drop-out and treatment strata, the intermittent missing responses are missing at random. This is the partial ignorability assumption in Definition 1.7.

1.5 Identifying Restrictions in Pattern Mixture Models

Pattern-mixture models by construction are not identified: the observed data does not provide enough information to identify the distributions for incomplete patterns (Little, 1993, 1994). Additional assumptions about the missing data process are necessary in order to yield identifying restrictions that equate the inestimable parameters to functions of estimable parameters and identify the full-data model.

For example, consider the situation when $\mathbf{y} = (y_1, y_2)$ is a bivariate normal response with missing data only in y_2 . Let s be the survival time, i.e. $s = 1$ if y_2 is missing and $s = 2$ if y_2 is observed. We model $p(s)$ and $p(\mathbf{y}|s)$ as $s \sim \text{Bern}(\phi)$ and $\mathbf{y}|s = i \sim N(\boldsymbol{\mu}^{(s)}, \boldsymbol{\Sigma}^{(s)})$ for $i = 1, 2$, with

$$\boldsymbol{\mu}^{(s)} = \begin{bmatrix} \mu_1^{(s)} \\ \mu_2^{(s)} \end{bmatrix} \quad \text{and} \quad \boldsymbol{\Sigma}^{(s)} = \begin{bmatrix} \sigma_{11}^{(s)} & \sigma_{12}^{(s)} \\ \sigma_{12}^{(s)} & \sigma_{22}^{(s)} \end{bmatrix}.$$

For $s = 1$, only y_1 is observed. Therefore, parameters $\mu_2^{(1)}$, $\sigma_{21}^{(1)}$ (equals $\sigma_{12}^{(1)}$) and $\sigma_{22}^{(1)}$ are not identified. By assuming

$$y_2|y_1, s = 1 \simeq y_2|y_1, s = 2,$$

which is the available case missing value (ACMV) restriction defined later in this section, we have

$$\begin{aligned}\mu_2^{(1)} + \frac{\sigma_{21}^{(1)}}{\sigma_{11}^{(1)}}(y_1 - \mu_1^{(1)}) &= \mu_2^{(2)} + \frac{\sigma_{21}^{(2)}}{\sigma_{11}^{(2)}}(y_1 - \mu_1^{(2)}) \\ \sigma_{22}^{(1)} - \frac{(\sigma_{21}^{(1)})^2}{\sigma_{11}^{(1)}} &= \sigma_{22}^{(2)} - \frac{(\sigma_{21}^{(2)})^2}{\sigma_{11}^{(2)}} \\ \frac{\sigma_{21}^{(1)}}{\sigma_{11}^{(1)}} &= \frac{\sigma_{21}^{(2)}}{\sigma_{11}^{(2)}},\end{aligned}$$

by which all the unidentified parameters are identified.

Understanding (identifying) restrictions that lead to MAR is an important first step for sensitivity analysis under missing not at random (MNAR) (Daniels and Hogan, 2008; Scharfstein et al., 2003; Zhang and Heitjan, 2006). In particular, MAR provides a good starting point for sensitivity analysis and sensitivity analysis are essential for the analysis of incomplete data (Daniels and Hogan, 2008; Scharfstein et al., 1999).

Little (1993) developed several common identifying restrictions. For example, complete case missing value (CCMV) restrictions which equate all missing patterns to the complete cases, i.e.

$$p_k(y_j | \bar{y}_{j-1}) = p_j(y_j | \bar{y}_{j-1});$$

equating parameters to a set of patterns, that is set parameters in pattern k , namely $\theta^{(k)}$, equal to the set of patterns S

$$\theta^{(k)} = \sum_{j \in S} \pi_j \theta^{(j)};$$

or equating pattern distributions to a mixture of a set of patterns S , i.e.

$$p_k(\cdot) = \sum_{j \in S} \pi_j p_j(\cdot).$$

Some special case of the pattern-set mixture models restrictions include nearest-neighbor constraints:

$$p_k(y_j | \bar{y}_{j-1}) = p_j(y_j | \bar{y}_{j-1}),$$

and available case missing value (ACMV) constraints:

$$p_k(y_j|\bar{y}_{j-1}) = p_{\geq j}(y_j|\bar{y}_{j-1}).$$

Molenberghs et al. (1998) proved that for discrete time points and monotone missingness, the ACMV constraint is equivalent to missing at random (MAR).

Thijs et al. (2002) developed strategies to apply identifying restrictions. That is first fit $p_k(\bar{y}_k)$, then choose an identifying restriction to identify the missing patterns. Multiple imputation can be applied by drawing unobserved components from the identified missing patterns. Kenward et al. (2003) discussed identifying restrictions corresponding to missing non-future dependence.

1.6 Dissertation Goals

There will be two major components to this dissertation. First, we will develop a Bayesian semiparametric model for longitudinal binary responses with non-ignorable missingness, including drop-out and intermittent missingness. Second, we will carefully explore identifying restrictions for pattern mixture models.

Bayesian shrinkage model: We propose two different parameterizations of saturated models for the observed data distribution, as well as corresponding shrinkage priors to avoid the curse of dimensionality. The two procedures provide researchers different strategies for reducing the dimension of parameter space. We assume a non-future dependence model for the drop-out mechanism and partial ignorability for the intermittent missingness. In a simulation study, we compare our approach to a fully parametric and a fully saturated model for the distribution of the observed data. Our methodology is motivated by, and applied to, data from the Breast Cancer Prevention Trial.

Identifying restrictions and sensitivity analysis in pattern mixture models: The normality of response data (if appropriate) for pattern mixture models is desirable as it easily allows incorporation of baseline covariates and introduction of sensitivity

parameters (for MNAR analysis) that have convenient interpretations as deviations of means and variances from MAR ([Daniels and Hogan, 2008](#)). However, multivariate normality within patterns can be overly restrictive when applying identifying restrictions. We explore such issues in [Chapter 4](#).

Furthermore, identification strategies can be problematic in models with covariates (e.g. baseline covariates with time-invariant coefficients). In this Chapter, we also explore conditions necessary for identifying restrictions that result in missing at random (MAR) to exist under a multivariate normality assumption and strategies for sensitivity analysis. Problems caused by baseline covariates with time-invariant coefficients are investigated and an alternative identifying restriction based on residuals is proposed as a solution.

CHAPTER 2
A BAYESIAN SHRINKAGE MODEL FOR LONGITUDINAL BINARY DATA WITH
DROP-OUT

2.1 Introduction

2.1.1 Breast Cancer Prevention Trial

The Breast Cancer Prevention Trial (BCPT) was a large multi-center, double-blinded, placebo-controlled, chemoprevention trial of the National Surgical Adjuvant Breast and Bowel Project (NSABP) designed to test the efficacy of 20mg/day tamoxifen in preventing breast cancer and coronary heart disease in healthy women at risk for breast cancer (Fisher et al., 1998). The study was open to accrual from June 1, 1992 through September 30, 1997 and 13,338 women aged 35 or older were enrolled in the study during this interval. The primary objective was to determine whether long-term tamoxifen therapy is effective in preventing the occurrence of invasive breast cancer. Secondary objectives included quality of life (QOL) assessments to evaluate benefit as well as risk resulting from the use of tamoxifen.

Monitoring QOL was of particular importance for this trial since the participants were healthy women and there had been concerns voiced by researchers about the association between clinical depression and tamoxifen use. Accordingly, data on depression symptoms was scheduled to be collected at baseline prior to randomization, at 3 months, at 6 months and every 6 months thereafter for up to 5 years. The primary instrument used to monitor depressive symptoms over time was the Center for Epidemiologic Studies Depression Scale (CES-D)(Radloff, 1977). This self-test questionnaire is composed of 20 items, each of which is scored on a scale of 0-3. A score of 16 or higher is considered as a likely case of clinical depression.

The trial was unblinded on March 31, 1998, after an interim analysis showed a dramatic reduction in the incidence of breast cancer in the treatment arm. Due to the potential loss of the control arm, we focus on QOL data collected on the 10,982 participants who were enrolled during the first two years of accrual and had their CES-D

score recorded at baseline. All women in this cohort had the potential for three years of follow-up (before the unblinding).

In the BCPT, the clinical centers were not required to collect QOL data on women after they stopped their assigned therapy. This design feature aggravated the problem of missing QOL data in the trial. As reported in [Land et al. \(2002\)](#), more than 30% of the CES-D scores were missing at the 36-month follow-up, with a slightly higher percentage in the tamoxifen group. They also showed that women with higher baseline CES-D scores had higher rates of missing data at each follow-up visit and the mean observed CES-D scores preceding a missing measurement were higher than those preceding an observed measurement; there was no evidence that these relationships differed by treatment group.

While these results suggest that the missing data process is associated with observed QOL outcomes, one cannot rule out the possibility that the process is further related to unobserved outcomes and that this relationship is modified by treatment. In particular, investigators were concerned (a priori) that, between assessments, tamoxifen might be causing depression in some individuals, who then do not return for their next assessment. If this occurs, the data are said to be missing not at random (MNAR); otherwise the data are said to be missing at random (MAR).

2.1.2 Informative Drop-Out in Longitudinal Studies

In this paper, we will concern ourselves with inference in longitudinal studies, where individuals who miss visits do not return for subsequent visits (i.e., drop-out). In such a setting, MNAR is often referred to as informative drop-out. While there were some intermittent responses in the BCPT, we will, as in [Land et al. \(2002\)](#), consider a “monotonized” dataset, whereby all CES-D scores observed on an individual after their first missing score have been deleted (this increases the “dropout” rate).

There are two main inferential paradigms for analyzing longitudinal studies with informative drop-out: likelihood (parametric) and non-likelihood (semi-parametric).

Articles by [Little \(1995\)](#), [Hogan and Laird \(1997b\)](#) and [Kenward and Molenberghs \(1999\)](#) as well as recent books by [Molenberghs and Kenward \(2007\)](#) and [Daniels and Hogan \(2008\)](#) provide a comprehensive review of likelihood-based approaches, including selection models, pattern-mixture models, and shared-parameter models. These models differ in the way the joint distribution of the outcome and missing data processes are factorized. In selection models, one specifies a model for the marginal distribution of the outcome process and a model for the conditional distribution of the drop-out process given the outcome process (see, for example, [Albert, 2000](#); [Baker, 1995](#); [Diggle and Kenward, 1994](#); [Fitzmaurice et al., 1995](#); [Heckman, 1979a](#); [Liu et al., 1999](#); [Molenberghs et al., 1997](#)); in pattern-mixture models, one specifies a model for the conditional distribution of the outcome process given the drop-out time and the marginal distribution of the drop-out time (see, for example, [Birmingham and Fitzmaurice, 2002](#); [Daniels and Hogan, 2000](#); [Fitzmaurice and Laird, 2000b](#); [Hogan and Laird, 1997a](#); [Little, 1993, 1994, 1995](#); [Pauler et al., 2003](#); [Roy, 2003](#); [Roy and Daniels, 2008](#); [Thijs et al., 2002](#)); and in shared-parameter models, the outcome and drop-out processes are assumed to be conditionally independent given shared random effects (see, for example, [DeGruttola and Tu, 1994](#); [Land et al., 2002](#); [Pulkstenis et al., 1998](#); [Ten Have et al., 1998, 2000](#); [Wu and Carroll, 1988](#); [Yuan and Little, 2009](#)). Traditionally, these models have relied on very strong distributional assumptions in order to obtain model identifiability.

Without these strong distributional assumptions, specific parameters from these models would not be identified from the distribution of the observed data. To address this issue within a likelihood-based framework, several authors ([Baker et al., 1992](#); [Daniels and Hogan, 2008](#); [Kurland and Heagerty, 2004](#); [Little, 1994](#); [Little and Rubin, 1999](#); [Nordheim, 1984](#)) have promoted the use of global sensitivity analysis, whereby non- or weakly- identified, interpretable parameters are fixed and then varied to evaluate

the robustness of the inferences. Scientific experts can be employed to constrain the range of these parameters.

Non-likelihood approaches to informative drop-out in longitudinal studies have been primarily developed from a selection modeling perspective. Here, the marginal distribution of the outcome process is modeled non- or semi-parametrically and the conditional distribution of the drop-out process given the outcome process is modeled semi- or fully- parametrically. In the case where the drop-out process is assumed to depend only on observable outcomes (i.e., MAR), [Robins et al. \(1994, 1995\)](#), [van der Laan and Robins \(2003\)](#) and [Tsiatis \(2006\)](#) developed inverse-weighted and augmented inverse-weighted estimating equations for inference. For informative drop-out, [Rotnitzky et al. \(1998a\)](#), [Scharfstein et al. \(1999\)](#) and [Rotnitzky et al. \(2001\)](#) introduced a class of selection models, in which the model for drop-out is indexed by interpretable sensitivity parameters that express departures from MAR. Inference using inverse-weighted estimating equations was proposed.

The problem with the aforementioned sensitivity analysis approaches is that the ultimate inferences can be cumbersome to display. [Vansteelandt et al. \(2006a\)](#) developed a method for reporting ignorance and uncertainty intervals (regions) that contain the true parameter(s) of interest with a prescribed level of precision, when the true data generating model is assumed to fall within a plausible class of models (as an example, see [Scharfstein et al., 2004](#)). An alternative and very natural strategy is specify an informative prior distribution on the non- or weakly- identified parameters and conduct a fully Bayesian analysis, whereby the ultimate inferences are reported in terms of posterior distributions. In the cross-sectional setting with a continuous outcome, [Scharfstein et al. \(2003\)](#) adopted this approach from a semi-parametric selection modeling perspective. [Kaciroti et al. \(2009\)](#) proposed a parametric pattern-mixture model for cross-sectional, clustered binary outcomes . [Lee et al. \(2008\)](#) introduced a fully-parametric pattern-mixture approach in the longitudinal setting with binary

outcomes. In this paper, we consider the same setting as [Lee et al. \(2008\)](#), but offer a more flexible strategy. In the context of BCPT, the longitudinal outcome will be the indicator that the CES-D score is 16 or higher.

2.1.3 Outline

The paper is organized as follows. In [Section 2.2](#), we describe the data structure. In [Section 2.3](#) and [2.4](#), we formalize identification assumptions and prove that the full-data distribution is identified under these assumptions. We introduce a saturated model for the distribution of the observed data in [Section 2.5](#). In [Section 2.6](#), we illustrate how to apply shrinkage priors to high-order interaction parameters in the saturated model to reduce the dimensionality of the parameter space and how to elicit (conditional) informative priors for non-identified sensitivity parameters from experts. In [Section 2.7](#), we assess, by simulation, the behavior of three classes of models for the distribution of observed data; parametric, saturated, and shrinkage. Our analysis of the BCPT trial is presented in [Section 2.8](#). [Section 2.9](#) is devoted to a summary and discussion.

2.2 Data Structure and Notation

Let Z denote the treatment assignment indicator, where $Z = 1$ denotes tamoxifen and $Z = 0$ denotes placebo. Let Y_j denote the binary outcome (i.e., depression) scheduled to be measured at the j th visit ($j = 0(\text{baseline}), \dots, J$) and let $\bar{\mathbf{Y}}_j = (Y_0, \dots, Y_j)$ denote the history of the outcome process through visit j . Let R_j denote the indicator that an individual has her depression status recorded at visit j . We assume that $R_0 = 1$ (i.e., Y_0 is always observed) and $R_j = 0$ implies that $R_{j+1} = 0$ (i.e., monotone missing data). Let $C = \max\{t : R_t = 1\}$ be the last visit at which an individual's depression status is recorded. The full and observed data for an individual are $F = (Z, C, \bar{\mathbf{Y}}_J)$ and $O = (Z, C, \bar{\mathbf{Y}}_C)$, respectively. We assume that we observe n i.i.d., copies of O . We will use the subscript i to denote data for the i th individual.

Our goal is to draw inference about $\mu_{zj}^* = P[Y_j = 1 | Z = z]$ for $j = 1, \dots, J$ and $z = 0, 1$.

2.3 Assumptions

To identify $\mu_{z,j}^*$ from the distribution of the observed data, we make the following two untestable assumptions:

Assumption 1 (Non-Future Dependence): R_j is independent of (Y_{j+1}, \dots, Y_J) given $R_{j-1} = 1$ and $\bar{\mathbf{Y}}_j$, for $j = 1, \dots, J - 1$.

This assumption asserts that for individuals at risk for drop-out at visit j and who share the same history of outcomes up to and including visit j , the distribution of future outcomes is the same for those who are last seen at visit j and those who remain on study past visit j . This assumption has been referred to as non-future dependence (Kenward et al., 2003).

Assumption 2 (Pattern-Mixture Representation): For $j = 1, \dots, J$ and $y_j = 0, 1$,

$$P[Y_j = y_j | R_j = 0, R_{j-1} = 1, \bar{\mathbf{Y}}_{j-1}, Z = z] = \frac{P[Y_j = y_j | R_j = 1, \bar{\mathbf{Y}}_{j-1}, Z = z] \exp\{q_{z,j}(\bar{\mathbf{Y}}_{j-1}, y_j)\}}{E[\exp\{q_{z,j}(\bar{\mathbf{Y}}_{j-1}, Y_j)\} | R_j = 1, \bar{\mathbf{Y}}_{j-1}, Z = z]}$$

where $q_{z,j}(\bar{\mathbf{Y}}_{j-1}, Y_j)$ is a specified function of its arguments.

Assumption 2 links the non-identified conditional distribution of Y_j given $R_j = 0$, $R_{j-1} = 1$, $\bar{\mathbf{Y}}_{j-1}$, and $Z = z$ to the identified conditional distribution of Y_j given $R_j = 1$, $\bar{\mathbf{Y}}_{j-1}$, and $Z = z$ using exponential tilting via the specified function $q_{z,j}(\bar{\mathbf{Y}}_{j-1}, Y_j)$.

Assumption (2) has a selection model representation that is obtained using Bayes' rule.

Assumption 2 (Selection Model Representation): For $j = 1, \dots, J$,

$$\text{logit} \{P[R_j = 0 | R_{j-1} = 1, \bar{\mathbf{Y}}_j, Z = z]\} = h_{z,j}(\bar{\mathbf{Y}}_{j-1}) + q_{z,j}(\bar{\mathbf{Y}}_{j-1}, Y_j)$$

where

$$h_{z,j}(\bar{\mathbf{Y}}_{j-1}) = \text{logit} P[R_j = 0 | R_{j-1} = 1, \bar{\mathbf{Y}}_{j-1}, Z = z] - \log\{E[\exp\{q_{z,j}(\bar{\mathbf{Y}}_{j-1}, Y_j)\} | R_{j-1} = 1, \bar{\mathbf{Y}}_{j-1}, Z = z]\}$$

With this characterization, we see that the function $q_{z,j}(\bar{\mathbf{Y}}_{j-1}, Y_j)$ quantifies the influence (on a log odds ratio scale) of the potentially unobservable outcome Y_j on the conditional odds of dropping at time j .

2.4 Identifiability

The above two assumptions non-parametrically, just-identify $\mu_{z,j}^*$ for all $j = 1, \dots, J$ and $z = 0, 1$. To see this, consider the following representation of this conditional distribution, derived using the laws of total and conditional probability:

$$\begin{aligned} \mu_{z,j}^* &= \sum_{\bar{\mathbf{y}}_{j-1}} P[Y_j = 1 | R_j = 1, \bar{\mathbf{Y}}_{j-1} = \bar{\mathbf{y}}_{j-1}, Z = z] \times \\ &\quad \left\{ \prod_{l=1}^j P[R_l = 1 | R_{l-1} = 1, \bar{\mathbf{Y}}_{l-1} = \bar{\mathbf{y}}_{l-1}, Z = z] \prod_{l=0}^{j-1} P[Y_l = y_l | R_l = 1, \bar{\mathbf{Y}}_{l-1} = \bar{\mathbf{y}}_{l-1}, Z = z] \right\} + \\ &\quad \sum_{k=1}^j \sum_{\bar{\mathbf{y}}_{k-1}} P[Y_j = 1 | R_k = 0, R_{k-1} = 1, \bar{\mathbf{Y}}_{k-1} = \bar{\mathbf{y}}_{k-1}, Z = z] P[R_k = 0 | R_{k-1} = 1, \bar{\mathbf{Y}}_{k-1} = \bar{\mathbf{y}}_{k-1}, Z = z] \\ &\quad \times \left\{ \prod_{l=1}^{k-1} P[R_l = 1 | R_{l-1} = 1, \bar{\mathbf{Y}}_{l-1} = \bar{\mathbf{y}}_{l-1}, Z = z] \prod_{l=0}^{k-1} P[Y_l = y_l | R_l = 1, \bar{\mathbf{Y}}_{l-1} = \bar{\mathbf{y}}_{l-1}, Z = z] \right\} \end{aligned}$$

All quantities on the right hand side of this equation are identified, without appealing to any assumptions, except $P[Y_j = 1 | R_k = 0, R_{k-1} = 1, \bar{\mathbf{Y}}_{k-1} = \bar{\mathbf{y}}_{k-1}, Z = z]$ for $k = 1, \dots, j-1$. Under Assumptions 1 and 2, these probabilities can be shown to be identified, implying that $\mu_{z,j}^*$ is identified for all j and z .

Theorem 1: $P[Y_j = 1 | R_{k-1} = 1, \bar{\mathbf{Y}}_{k-1} = \bar{\mathbf{y}}_{k-1}, Z = z]$ and $P[Y_j = 1 | R_k = 0, R_{k-1} = 1, \bar{\mathbf{Y}}_{k-1} = \bar{\mathbf{y}}_{k-1}, Z = z]$ are identified for $k = 1, \dots, j$.

Proof: The proof follows by backward induction. Consider $k = j$. By Assumption 2,

$$\begin{aligned} P[Y_j = 1 | R_j = 0, R_{j-1} = 1, \bar{\mathbf{Y}}_{j-1} = \bar{\mathbf{y}}_{j-1}, Z = z] &= \\ &= \frac{P[Y_j = 1 | R_j = 1, \bar{\mathbf{Y}}_{j-1} = \bar{\mathbf{y}}_{j-1}, Z = z] \exp\{q_{z,j}(\bar{\mathbf{Y}}_{j-1}, 1)\}}{E[\exp\{q_{z,j}(\bar{\mathbf{Y}}_{j-1}, Y_j)\} | R_j = 1, \bar{\mathbf{Y}}_{j-1} = \bar{\mathbf{y}}_{j-1}, Z = z]} \end{aligned}$$

Since the right hand side is identified, we know that $P[Y_j = 1|R_j = 0, R_{j-1} = 1, \bar{\mathbf{Y}}_{j-1} = \bar{\mathbf{y}}_{j-1}, Z = z]$ is identified. Further, we can write

$$\begin{aligned} & P[Y_j = 1|R_{j-1} = 1, \bar{\mathbf{Y}}_{j-1} = \bar{\mathbf{y}}_{j-1}, Z = z] \\ &= \sum_{r=0}^1 P[Y_j = 1|R_j = r, R_{j-1} = 1, \bar{\mathbf{Y}}_{j-1} = \bar{\mathbf{y}}_{j-1}, Z = z]P[R_j = r|R_{j-1} = 1, \bar{\mathbf{Y}}_{j-1} = \bar{\mathbf{y}}_{j-1}, Z = z] \end{aligned}$$

Since all quantities on the right hand side are identified, $P[Y_j = 1|R_{j-1} = 1, \bar{\mathbf{Y}}_{j-1} = \bar{\mathbf{y}}_{j-1}, Z = z]$ is identified.

Suppose that $P[Y_j = 1|R_k = 0, R_{k-1} = 1, \bar{\mathbf{Y}}_{k-1} = \bar{\mathbf{y}}_{k-1}, Z = z]$ and $P[Y_j = 1|R_{k-1} = 1, \bar{\mathbf{Y}}_{k-1} = \bar{\mathbf{y}}_{k-1}, Z = z]$ are identified for some k where $1 < k < j$. Then, we need to show that these probabilities are identified for $k' = k - 1$. To see this, note that

$$\begin{aligned} & P[Y_j = 1|R_{k'} = 0, R_{k'-1} = 1, \bar{\mathbf{Y}}_{k'-1} = \bar{\mathbf{y}}_{k'-1}, Z = z] \\ &= P[Y_j = 1|R_{k-1} = 0, R_{k-2} = 1, \bar{\mathbf{Y}}_{k-2} = \bar{\mathbf{y}}_{k-2}, Z = z] \\ &= \sum_{y_{k-1}=0}^1 P[Y_j = 1|R_{k-1} = 0, R_{k-2} = 1, \bar{\mathbf{Y}}_{k-1} = \bar{\mathbf{y}}_{k-1}, Z = z] \times \\ &\quad P[Y_{k-1} = y_{k-1}|R_{k-1} = 0, R_{k-2} = 1, \bar{\mathbf{Y}}_{k-2} = \bar{\mathbf{y}}_{k-2}, Z = z] \\ &= \sum_{y_{k-1}=0}^1 P[Y_j = 1|R_{k-1} = 1, \bar{\mathbf{Y}}_{k-1} = \bar{\mathbf{y}}_{k-1}, Z = z] \times \\ &\quad \frac{P[Y_{k-1} = y_{k-1}|R_{k-1} = 1, \bar{\mathbf{Y}}_{k-2} = \bar{\mathbf{y}}_{k-2}, Z = z] \exp\{q_{Z,k-1}(\bar{\mathbf{Y}}_{k-2}, y_{k-1})\}}{E[\exp\{q_{Z,k-1}(\bar{\mathbf{Y}}_{k-2}, Y_{k-1})\}|R_{k-1} = 1, \bar{\mathbf{Y}}_{k-2} = \bar{\mathbf{y}}_{k-2}, Z = z]} \end{aligned}$$

The third equality follows by Assumptions 1 and 2. Since all the quantities on the right hand side of the last equality are identified, $P[Y_j = 1|R_{k'} = 0, R_{k'-1} = 1, \bar{\mathbf{Y}}_{k'-1} = \bar{\mathbf{y}}_{k'-1}, Z = z]$ is identified. Further,

$$\begin{aligned} & P[Y_j = 1|R_{k'-1} = 1, \bar{\mathbf{Y}}_{k'-1} = \bar{\mathbf{y}}_{k'-1}, Z = z] \\ &= P[Y_j = 1|R_{k-2} = 1, \bar{\mathbf{Y}}_{k-2} = \bar{\mathbf{y}}_{k-2}, Z = z] \\ &= \sum_{y_{k-1}=0}^1 P[Y_j = 1|R_{k-1} = 1, \bar{\mathbf{Y}}_{k-1} = \bar{\mathbf{y}}_{k-1}, Z = z] \times \end{aligned}$$

$$\begin{aligned}
& P[Y_{k-1} = y_{k-1} | R_{k-1} = 1, \bar{\mathbf{Y}}_{k-2} = \bar{\mathbf{y}}_{k-2}, Z = z] \times \\
& P[R_{k-1} = 1 | R_{k-2} = 1, \bar{\mathbf{Y}}_{k-2} = \bar{\mathbf{y}}_{k-2}, Z = z] + \\
& \sum_{y_{k-1}=0}^1 P[Y_j = 1 | R_{k-1} = 0, R_{k-2} = 1, \bar{\mathbf{Y}}_{k-1} = \bar{\mathbf{y}}_{k-1}, Z = z] \times \\
& P[Y_{k-1} = y_{k-1} | R_{k-1} = 0, \bar{\mathbf{Y}}_{k-2} = \bar{\mathbf{y}}_{k-2}, Z = z] \times \\
& P[R_{k-1} = 0 | R_{k-2} = 1, \bar{\mathbf{Y}}_{k-2} = \bar{\mathbf{y}}_{k-2}, Z = z] \\
= & \sum_{y_{k-1}=0}^1 P[Y_j = 1 | R_{k-1} = 1, \bar{\mathbf{Y}}_{k-1} = \bar{\mathbf{y}}_{k-1}, Z = z] \times \\
& P[Y_{k-1} = y_{k-1} | R_{k-1} = 1, \bar{\mathbf{Y}}_{k-2} = \bar{\mathbf{y}}_{k-2}, Z = z] \times \\
& P[R_{k-1} = 1 | R_{k-2} = 1, \bar{\mathbf{Y}}_{k-2} = \bar{\mathbf{y}}_{k-2}, Z = z] + \\
& \sum_{y_{k-1}=0}^1 P[Y_j = 1 | R_{k-1} = 1, \bar{\mathbf{Y}}_{k-1} = \bar{\mathbf{y}}_{k-1}, Z = z] \times \\
& \frac{P[Y_{k-1} = y_{k-1} | R_{k-1} = 1, \bar{\mathbf{Y}}_{k-2} = \bar{\mathbf{y}}_{k-2}, Z = z] \exp\{q_{z,k-1}(\bar{\mathbf{Y}}_{k-2}, y_{k-1})\}}{E[\exp\{q_{z,k-1}(\bar{\mathbf{Y}}_{k-2}, Y_{k-1})\} | R_{k-1} = 1, \bar{\mathbf{Y}}_{k-2} = \bar{\mathbf{y}}_{k-2}, Z = z]} \times \\
& P[R_{k-1} = 0 | R_{k-2} = 1, \bar{\mathbf{Y}}_{k-2} = \bar{\mathbf{y}}_{k-2}, Z = z]
\end{aligned}$$

The third equality follows by Assumptions 1 and 2. Since all the quantities on the right hand side of the last equality are identified, $P[Y_j = 1 | R_{k'-1} = 1, \bar{\mathbf{Y}}_{k'-1} = \bar{\mathbf{y}}_{k'-1}, Z = z]$ is identified. \square

The identifiability result shows that, given the functions $q_{z,j}(\bar{\mathbf{Y}}_{j-1}, Y_j)$, $\mu_{z,j}^*$ can be expressed as functional of the distribution of the observed data. In particular, the functional depends on the conditional distributions of Y_j given $R_j = 1$, $\bar{\mathbf{Y}}_{j-1}$, and Z for $j = 0, \dots, J$ and the conditional distributions of R_j given $R_{j-1} = 1$, $\bar{\mathbf{Y}}_{j-1}$ and Z for $j = 1, \dots, J$. Furthermore, the functions $q_{z,j}(\bar{\mathbf{Y}}_{j-1}, Y_j)$ are not identifiable from the distribution of the observed data and their specification places no restrictions on the distribution of the observed data.

2.5 Modeling

We specify saturated models for the observed data via the sequential conditional distributions of $[Y_j | R_j = 1, \bar{\mathbf{Y}}_{j-1}, Z]$ for $j = 0, \dots, J$ and the conditional hazards $[R_j | R_{j-1} = 1, \bar{\mathbf{Y}}_{j-1}, Z]$ for $j = 1, \dots, J$. We parameterize these models as follows:

$$\begin{aligned} \text{logit } P[Y_0 = 1 | R_0 = 1, Z = z] &= \alpha_{z,0,0} \\ \text{logit } P[Y_j = 1 | R_j = 1, \bar{\mathbf{Y}}_{j-1} = \bar{\mathbf{y}}_{j-1}, Z = z] &= \alpha_{z,j,0} + \alpha_{z,j,1}Y_{j-1} + \sum_{k=0}^{j-2} \alpha_{z,j,k}^{(1)}Y_k \\ &+ \sum_{k,l \in A_j^{(2)}} \alpha_{z,j,\cdot}^{(2)}Y_kY_l + \sum_{k,l,m \in A_j^{(3)}} \alpha_{z,j,\cdot}^{(3)}Y_kY_lY_m + \dots + \alpha_{z,j}^{(j-1)}Y_0Y_1 \cdots Y_{j-1} \\ \text{logit } P[R_j = 0 | R_{j-1} = 1, \bar{\mathbf{Y}}_{j-1} = \bar{\mathbf{y}}_{j-1}, Z = z] &= \gamma_{z,j,0} + \gamma_{z,j,1}Y_{j-1} + \sum_{k=0}^{j-2} \gamma_{z,j,k}^{(1)}Y_k \\ &+ \sum_{k,l \in A_j^{(2)}} \gamma_{z,j,\cdot}^{(2)}Y_kY_l + \sum_{k,l,m \in A_j^{(3)}} \gamma_{z,j,\cdot}^{(3)}Y_kY_lY_m + \dots + \gamma_{z,j}^{(j-1)}Y_0Y_1 \cdots Y_{j-1} \end{aligned}$$

for $j = 1, \dots, J$, where $A_j^{(t)}$ is the set of all t -tuples of the integers $0, \dots, j-1$. Let α denote the parameters indexing the conditional distributions $[Y_j | R_j = 1, \bar{\mathbf{Y}}_{j-1}, Z]$, γ denote the parameters indexing the conditional distributions $[R_j | R_{j-1} = 1, \bar{\mathbf{Y}}_{j-1}, Z]$ and $\theta = \{\alpha, \gamma\}$.

Furthermore, we propose to parameterize the functions $q_{z,j}(\bar{\mathbf{Y}}_{j-1}, Y_j)$ with parameters $\tau_{z,j,\bar{\mathbf{y}}_{j-1}} = q_{z,j}((\bar{\mathbf{y}}_{j-1}, 1)) - q_{z,j}((\bar{\mathbf{y}}_{j-1}, 0))$. Here, $\exp(\tau_{z,j,\bar{\mathbf{y}}_{j-1}})$ represents, in the context of the BCPT trial, the conditional odds ratio of dropping out between visits $j-1$ and j for individuals who are depressed vs. not depressed at visit j , but share the mental history $\bar{\mathbf{y}}_{j-1}$ through visit $j-1$. We let τ denote the collection of $\tau_{z,j,\bar{\mathbf{y}}_{j-1}}$'s.

2.6 Prior Specification and Posterior Computation

For specified sensitivity analysis parameters τ , the saturated model proposed in Section 2.5 provides a perfect fit to the distribution of the observed data. In this model, however, the number of parameters increases exponentially in J . In contrast, the number of data points increases linearly in J . As a consequence, there will be many

combinations of $\bar{\mathbf{y}}_{j-1}$ (i.e., "cells") which will be sparsely represented in the dataset. For example, in the BCPT trial, about 50% of the possible realizations of $\bar{\mathbf{Y}}_7$ have less than two observations and about 15% have no observations. For a frequentist perspective, this implies that components of θ will be imprecisely estimated; in turn, this can adversely affect estimation of $\mu_{z,j}^*$. This has been called the curse of dimensionality (Robins and Ritov, 1997).

2.6.1 Shrinkage Priors

To address this problem, we introduce data driven shrinkage priors for higher order interactions to reduce the number of parameters in an automated manner. In particular, we assume

$$\alpha_{z,j,k}^{(t)} \sim N(0, \sigma_\alpha^{(t)}) \quad \text{and} \quad \gamma_{z,j,k}^{(t)} \sim N(0, \sigma_\gamma^{(t)}) \quad k \in A_j^{(t)}, 3 \leq t < j \leq J, z = 0, 1 \quad (2-1)$$

where t is the order of interactions and the hyper-parameters (shrinkage variances) follow distributions

$$\sigma_\alpha^{(t)} \sim \text{Unif}(0, 10) \quad \text{and} \quad \sigma_\gamma^{(t)} \sim \text{Unif}(0, 10).$$

When $\sigma_\alpha^{(t)}$ and $\sigma_\gamma^{(t)}$ equal zero for all interactions, the saturated model is reduced to a first order Markov model,

$$\text{logit } P[Y_0 = 1 | R_0 = 1, Z = z] = \alpha_{z,0,0}$$

$$\text{logit } P[Y_j = 1 | R_j = 1, \bar{\mathbf{Y}}_{j-1} = \bar{\mathbf{y}}_{j-1}, Z = z] = \alpha_{z,j,0} + \alpha_{z,j,1}Y_{j-1}$$

$$\text{logit } P[R_j = 0 | R_{j-1} = 1, \bar{\mathbf{Y}}_{j-1} = \bar{\mathbf{y}}_{j-1}, Z = z] = \gamma_{z,j,0} + \gamma_{z,j,1}Y_{j-1}.$$

The shrinkage priors allow the "neighboring" cells in the observed data model to borrow information from each other and provide more precise estimates.

When the first order Markov model is not true, as n goes to infinity, the posterior means of observed data probabilities will converge to their true values as long as

the shrinkage priors are $O(1)$ (which is the case here) and all the true values of the observed data probabilities, $P[Y_j|R_j = 1, \bar{\mathbf{Y}}_{j-1}, Z]$ for $j = 0, \dots, J$ and are in the open interval, $(0, 1)$. This follows, since under this latter condition, all combinations of depression histories have a positive probability of being observed and the prior will become swamped by the observed data. However, when the true value of any of the observed data probabilities is zero or one, there exists at least one combination of depression history that will never be observed and thus the influence of the prior will not dissipate as n increases.

We specify non-informative priors $N(0, 1000)$ for the non-interaction parameters in θ , namely $\alpha_{z,j,0}$ for $j = 0, \dots, J$ and $z = 0, 1$, $\alpha_{z,j,1}$, $\gamma_{z,j,0}$ and $\gamma_{z,j,1}$ for $j = 1, \dots, J$ and $z = 0, 1$.

2.6.2 Prior of Sensitivity Parameters

The sensitivity parameters in Assumption 2, defined formally in Section 2.5, are (conditional) odds ratios. In our experience, subject matter experts often have difficulty thinking in terms of odds ratios; rather, they are more comfortable expressing beliefs about relative risks (Scharfstein et al., 2006; Shepherd et al., 2007). With this in mind, we asked Dr. Patricia Ganz, a medical oncologist and expert on quality of life outcomes in breast cancer, to express her beliefs about the risk of dropping out and its relationship to treatment assignment and depression. We then translated her beliefs into prior distributional assumptions about the odds ratio sensitivity parameters τ .

Specifically, we asked Dr. Ganz to answer the following question for each treatment group:

Q: Consider a group of women assigned to placebo (tamoxifen), who are on study through visit $j - 1$ and who share the same history of depression. Suppose that the probability that a randomly selected woman in this group drops out before visit j is p (denoted by the columns in Table 1). For each p , what is the minimum, maximum and your best guess (median) representing how much more (e.g. twice) or less (e.g., half)

likely you consider the risk of dropping out before visit j for a woman who would be depressed at year j RELATIVE to a woman who would not be depressed at visit j ?

Implicit in this question is the assumption that, for each treatment group, the relative risk only depends on past history and the visit number only through the risk of dropping out between visits $j - 1$ and j .

For notational convenience, let $r_z(p)$ denote the relative risk of drop-out for treatment group z and drop-out probability p . Further, let $r_{z,min}(p)$, $r_{z,med}(p)$ and $r_{z,max}(p)$ denote the elicited minimum, median, and maximum relative risks (see Table 2-1). Let $p_{z,j}(\bar{\mathbf{y}}_{j-1}) = P[R_j = 0 | R_{j-1} = 1, \bar{\mathbf{Y}}_{j-1} = \bar{\mathbf{y}}_{j-1}, Z = z]$ and let $p_{z,j}^{(y)}(\bar{\mathbf{y}}_{j-1}) = P[R_j = 0 | R_{j-1} = 1, \bar{\mathbf{Y}}_{j-1} = \bar{\mathbf{y}}_{j-1}, Y_j = y, Z = z]$ for $y = 0, 1$.

By definition,

$$r_z(p_{z,j}(\bar{\mathbf{y}}_{j-1})) = p_{z,j}^{(1)}(\bar{\mathbf{y}}_{j-1}) / p_{z,j}^{(0)}(\bar{\mathbf{y}}_{j-1})$$

$$p_{z,j}(\bar{\mathbf{y}}_{j-1}) = \sum_{y=0}^1 p_{z,j}^{(y)}(\bar{\mathbf{y}}_{j-1}) \pi_{z,j}^{(y)}(\bar{\mathbf{y}}_{j-1})$$

where $\pi_{z,j}^{(y)}(\bar{\mathbf{y}}_{j-1}) = P[Y_j = y | R_{j-1} = 1, \bar{\mathbf{Y}}_{j-1} = \bar{\mathbf{y}}_{j-1}, Z = z]$ for $y = 0, 1$. This implies that

$$p_{z,j}^{(0)}(\bar{\mathbf{y}}_{j-1}) = \frac{p_{z,j}(\bar{\mathbf{y}}_{j-1})}{\pi_{z,j}^{(1)}(\bar{\mathbf{y}}_{j-1})(r_z(p_{z,j}(\bar{\mathbf{y}}_{j-1})) - 1) + 1}.$$

Since $\pi_{z,j}^{(1)}(\bar{\mathbf{y}}_{j-1}) \in [0, 1]$, given $p_{z,j}(\bar{\mathbf{y}}_{j-1})$ and $r_z(p_{z,j}(\bar{\mathbf{y}}_{j-1}))$, $p_{z,j}^{(0)}(\bar{\mathbf{y}}_{j-1})$ is bounded as follows:

for $r_z(p_{z,j}(\bar{\mathbf{y}}_{j-1})) \geq 1$,

$$p_{z,j}(\bar{\mathbf{y}}_{j-1}) / r_z(p_{z,j}(\bar{\mathbf{y}}_{j-1})) \leq p_{z,j}^{(0)}(\bar{\mathbf{y}}_{j-1}) \leq \min\{p_{z,j}(\bar{\mathbf{y}}_{j-1}), 1\}$$

and, for $r_z(p_{z,j}(\bar{\mathbf{y}}_{j-1})) \leq 1$,

$$p_{z,j}(\bar{\mathbf{y}}_{j-1}) \leq p_{z,j}^{(0)}(\bar{\mathbf{y}}_{j-1}) \leq \min\{p_{z,j}(\bar{\mathbf{y}}_{j-1}) / r_z(p_{z,j}(\bar{\mathbf{y}}_{j-1})), 1\}.$$

We will use these bounds to construct our prior.

We construct the conditional prior of $\tau_{z,j,\bar{\mathbf{y}}_{j-1}}$ given $p_{z,j}(\bar{\mathbf{y}}_{j-1})$ using Steps 1-4 given below. The general strategy is to use the elicited information on the relative risk at different drop-out probabilities and the bounds derived above to construct the prior of interest.

Step 1. For $m \in \{min, med, max\}$, interpolate the elicited $r_{z,m}(p)$ at different drop-out probabilities (see Figure 2-1) to find $r_{z,m}(p_{z,j}(\bar{\mathbf{y}}_{j-1}))$ for any $p_{z,j}(\bar{\mathbf{y}}_{j-1})$.

Step 2. Construct the prior of $r_z(p_{z,j}(\bar{\mathbf{y}}_{j-1}))$ given $p_{z,j}(\bar{\mathbf{y}}_{j-1})$ as a 50-50 mixture of

$$\text{Uniform}(r_{z,min}(p_{z,j}(\bar{\mathbf{y}}_{j-1})), r_{z,med}(p_{z,j}(\bar{\mathbf{y}}_{j-1})))$$

and

$$\text{Uniform}(r_{z,med}(p_{z,j}(\bar{\mathbf{y}}_{j-1})), r_{z,max}(p_{z,j}(\bar{\mathbf{y}}_{j-1})))$$

random variables. This preserves the elicited percentiles of the relative risk.

Step 3. Construct a conditional prior of $p_{z,j}^{(0)}(\bar{\mathbf{y}}_{j-1})$ given $p_{z,j}(\bar{\mathbf{y}}_{j-1})$ and $r_z(p_{z,j}(\bar{\mathbf{y}}_{j-1}))$ as a uniform distribution with lower bound

$$\frac{p_{z,j}(\bar{\mathbf{y}}_{j-1})}{\max\{r_z(p_{z,j}(\bar{\mathbf{y}}_{j-1})), 1\}}$$

and upper bound

$$\min\left\{\frac{p_{z,j}(\bar{\mathbf{y}}_{j-1})}{\min\{r_z(p_{z,j}(\bar{\mathbf{y}}_{j-1})), 1\}}, \frac{1}{\max\{r_z(p_{z,j}(\bar{\mathbf{y}}_{j-1})), 1\}}\right\}.$$

The bounds were derived above.

Step 4. Steps (2) and (3) induce a prior for $\tau_{z,j,\bar{\mathbf{y}}_{j-1}}|\theta$ by noting

$$\tau_{z,j,\bar{\mathbf{y}}_{j-1}} = \log\left(\frac{r_z(p_{z,j}(\bar{\mathbf{y}}_{j-1}))(1 - p_{z,j}^{(0)}(\bar{\mathbf{y}}_{j-1}))}{1 - r_z(p_{z,j}(\bar{\mathbf{y}}_{j-1}))p_{z,j}^{(0)}(\bar{\mathbf{y}}_{j-1})}\right),$$

i.e., $\tau_{z,j,\bar{\mathbf{y}}_{j-1}}$ is a deterministic function of $r_z(p_{z,j}(\bar{\mathbf{y}}_{j-1}))$ and $p_{z,j}^{(0)}(\bar{\mathbf{y}}_{j-1})$.

The relative risks elicited from Dr. Ganz are given in Table 2-2. We extrapolated the relative risks outside the ranges given in Table 2-2 as shown in Figure 2-1.

Figure 2-2 shows the density of τ given $p_{z,j}(\bar{\mathbf{y}}_{j-1})$ equal 10% and 25% for the tamoxifen and placebo arms. For two patients with the same response history up to time point $j - 1$, the log odds ratio of dropping out at time point j , for the patient that is

depressed at time point j versus the patient that is not, increases as the overall drop out rate at time point j increases. In general, for a given $p_{z,j}(\bar{\mathbf{y}}_{j-1})$, the log odds ratio is higher for patients in the tamoxifen versus placebo arms.

2.6.3 Posterior Computation

With the shrinkage priors on θ , the elicited conditional priors τ given θ , and the observed data, the following steps are used to simulate draws from the posterior of $\mu_{z,j}^*$:

1. Using the proposed observed data model with the shrinkage priors on θ , we simulate draws from the posterior distributions of $P[Y_j = 1|R_j = 1, \bar{\mathbf{Y}}_{j-1} = \bar{\mathbf{y}}_{j-1}, Z = z]$ and $P[R_j = 0|R_j = 1, \bar{\mathbf{Y}}_{j-1} = \bar{\mathbf{y}}_{j-1}, Z = z]$ for all j, z and $\bar{\mathbf{y}}_{j-1}$ in WinBUGS.
2. For each draw of $P[R_j = 0|R_j = 1, \bar{\mathbf{Y}}_{j-1} = \bar{\mathbf{y}}_{j-1}, Z = z]$, we draw $\tau_{z,j,\bar{\mathbf{y}}_{j-1}}$ based on the conditional priors described in Section 6.2.
3. We compute $\mu_{z,j}^*$ by plugging the draws of $P[Y_j = 1|R_j = 1, \bar{\mathbf{Y}}_{j-1} = \bar{\mathbf{y}}_{j-1}, Z = z]$, $P[R_j = 0|R_j = 1, \bar{\mathbf{Y}}_{j-1} = \bar{\mathbf{y}}_{j-1}, Z = z]$ and $\tau_{z,j,\bar{\mathbf{y}}_{j-1}}$ into the identification algorithm discussed in Section 2.4.

To sample from the posterior distributions of $P[Y_j = 1|R_j = 1, \bar{\mathbf{Y}}_{j-1} = \bar{\mathbf{y}}_{j-1}, Z = z]$ and $P[R_j = 0|R_j = 1, \bar{\mathbf{Y}}_{j-1} = \bar{\mathbf{y}}_{j-1}, Z = z]$ in WinBUGS we stratify the individual binary data (by previous response history) and analyze as Binomial data; this serves to drastically improve the computational efficiency. Sampling $\tau_{z,j,\bar{\mathbf{y}}_{j-1}}$ and computing $\mu_{z,j}^*$ is implemented separately from the first step using R.

2.7 Assessment of Model Performance via Simulation

Via simulation, we compared the performance of the shrinkage model with a correct parametric model (given below), an incorrect parametric model (first order Markov model) and the saturated model with diffuse priors (given below).

The shrinkage model uses the shrinkage priors proposed in Section 2.6.1 (shrink the saturated model toward a first order Markov model). Note that the shrinkage priors shrink the saturated model to an incorrect parametric model.

For the saturated model with diffuse priors, we re-parameterize the model as

$$P[Y_j = 1 | R_j = 1, \bar{\mathbf{Y}}_{j-1} = \bar{\mathbf{y}}_{j-1}, Z = z] = \mu_{z, \bar{\mathbf{y}}_{j-1}}$$

$$P[R_j = 0 | R_{j-1} = 1, \bar{\mathbf{Y}}_{j-1} = \bar{\mathbf{y}}_{j-1}, Z = z] = \rho_{z, \bar{\mathbf{y}}_{j-1}}$$

for $j = 1, \dots, 7$, and specify independent $\text{Unif}(0, 1)$ on μ 's and ρ 's.

We simulated observed data from a "true" parametric model of the following form:

$$\text{logit}P[Y_0 = 1 | R_0 = 1, Z = z] = \alpha_{z,0,0}$$

$$\text{logit}P[Y_1 = 1 | R_1 = 1, Y_0 = y_0, Z = z] = \alpha_{z,1,0} + \alpha_{z,1,1}Y_0$$

$$\text{logit}P[R_1 = 0 | R_0 = 1, Y_0 = y_0, Z = z] = \gamma_{z,1,0} + \gamma_{z,1,1}Y_0$$

$$\text{logit}P[Y_j = 1 | R_j = 1, \bar{\mathbf{Y}}_{j-1} = \bar{\mathbf{y}}_{j-1}, Z = z] = \alpha_{z,j,0} + \alpha_{z,j,1}Y_{j-1} + \alpha_{z,j,2}Y_{j-2}$$

$$\text{logit}P[R_j = 0 | R_{j-1} = 1, \bar{\mathbf{Y}}_{j-1} = \bar{\mathbf{y}}_{j-1}, Z = z] = \gamma_{z,j,0} + \gamma_{z,j,1}Y_{j-1} + \gamma_{z,j,2}Y_{j-2},$$

for $j = 2$ to 7 .

To determine the parameters of the data generating model, we fit this model to the "monotonized" BCPT data in WinBUGS with non-informative priors. We used the posterior mean of the parameters α_z and γ_z as the true parameters. We compute the "true" values of $\mu_{z,j}^*$ by (1) drawing 10,000 values from the elicited prior of τ_z given γ_z given in Table 2-2, (2) computing $\mu_{z,j}^*$ using the identification algorithm in Section 2.4 for each draw, and (3) average the resulting $\mu_{z,j}^*$'s. The model parameters and the "true" depression rates $\mu_{z,j}^*$, are given in Table 2-3.

We considered (relatively) small (3000), moderate (5000), and large (10000) sample sizes for each treatment arm; for each sample size, we simulated 50 datasets. We assessed model performance using the mean squared error (MSE) criterion.

In Table 2-4, we report the MSEs of $P[Y_j = 1 | R_j = 1, \bar{\mathbf{Y}}_{j-1}, Z = z]$ and $P[R_j = 1 | R_{j-1} = 1, \bar{\mathbf{Y}}_{j-1}, Z = z]$ averaged over all j and all $\bar{\mathbf{Y}}_{j-1}$ (see columns 3 and 4, respectively). We also report the MSEs for $\mu_{z,j}^*$ (see columns 6-12). For

reference, the MSEs associated with the true data generating model are bolded. This table demonstrate that the shrinkage model generally outperforms both the incorrectly specified parametric model and the saturated model at all sample sizes. This improved performance is especially noticeable when comparing the MSEs for the rates of depression at times 3-7.

In addition, the MSEs for the shrinkage model compare favorably with those of the true parametric model for all sample sizes considered, despite the fact that the shrinkage priors were specified to shrink toward an incorrect model.

2.8 Application: Breast Cancer Prevention Trial (BCPT)

Table 2-5 displays the treatment-specific monotonized drop-out rates in the BCPT. By the 7th study visit, more than 40% of patients had missed one or more assessments, with a slightly higher percentage in the tamoxifen arm.

We fit the shrinkage model to the observed data using WinBUGS, with four chains of 8000 iterations and 1000 burn-in. Convergence was checked by examining trace plots of the multiple chains.

2.8.1 Model Fit and Shrinkage Results

To assess the model fit, we compared the empirical rates and posterior means (with 95% credible intervals) of $P[Y_j = 1, R_j = 1|Z = z]$ and $P[R_j = 0|Z = z]$. As shown in Figure 2-3, the shrinkage model fits the observed data well. Figure 2-4 illustrates the effect of shrinkage on the model fits by comparing the difference between the empirical rate and posterior mean of $P[Y_j = 1|R_j = 1, \bar{Y}_{j-1}, Z = z]$ for all j, z and \bar{Y}_{j-1} . We can see that for early time points, the difference is close to zero since there is little shrinkage applied to the model parameters. For later time points, more higher order interaction coefficients are shrunk toward zero and the magnitude of difference increases and drifts away from zero line. In general, the empirical estimates are less reliable for the later time points (re: the simulation results in Section 7). In some cases, there are no observations within "cells." By shrinking the high order

interactions (i.e., borrowing information across neighboring cells), we are able to estimate $P[Y_j = 1 | R_j = 1, \bar{Y}_{j-1}, Z = z]$ for all j, z and \bar{Y}_{j-1} with reasonable precision.

2.8.2 Inference

Figure 2-5 shows the posterior of $P[Y_7 = 1 | Z = z]$, the treatment-specific probability of depression at the end of the 36-month follow up (solid lines). For comparison, the posterior under MAR (corresponding to point mass priors for τ at zero) is also presented (dashed lines). The observed depression rates (i.e., complete case analysis) were 0.115 on both the placebo and tamoxifen arms. Under the MNAR analysis (using the elicited priors), the posterior mean of the depression rates at month 36 were 0.126 (95% *CI* : 0.115, 0.138) and 0.130 (95% *CI* : 0.119, 0.143) for the placebo and tamoxifen arms; the difference was 0.004 (95% *CI* : -0.012, 0.021). Under MAR, the rates were 0.125 (95% *CI* : 0.114, 0.136) and 0.126 (95% *CI* : 0.115, 0.138) for the placebo and tamoxifen arms; the difference was 0.001 (95% *CI* : -0.015, 0.018). The posterior probability of depression was higher under the MNAR analysis than the MAR analysis since researchers believed depressed patients were more likely to drop out (see Table 2-2), a belief that was captured by the elicited priors. Figure 2-6 shows that under the two treatments there were no significant differences in the depression rates at every time point (95% credible intervals all cover zero) under both MNAR and MAR. Similar (non-significant) treatment differences were seen when examining treatment comparisons conditional on depression status at baseline.

2.9 Summary and Discussion

In this paper, we have presented a Bayesian shrinkage approach for longitudinal binary data with informative drop-out. Our model provides a framework that incorporates expert opinion about non-identifiable parameters and avoids the curse of dimensionality by using shrinkage priors. In our analysis of the BCPT data, we concluded that there was little (if any) evidence that women on tamoxifen were more depressed than those on placebo.

An important feature of our approach is that the specification of models for the identifiable distribution of the observed data and the non-identifiable parameters can be implemented by separate independent data analysts. This feature can be used to increase the objectivity of necessarily subjective inferences in the FDA review of randomized trials with informative drop-out.

Penalized likelihood ([Fan and Li, 2001](#); [Green and Silverman, 1994](#); [Wahba, 1990](#)) is another approach for high-dimensional statistical modeling. There are similarities between the penalized likelihood approach and our shrinkage model. In fact, the shrinkage priors on the saturated model parameters proposed in our approach can be viewed as a specific form for the penalty.

The ideas in this paper can be extended to continuous outcomes. For example, one could use the mixtures of Dirichlet processes model ([Escobar and West, 1995](#)) for the distribution of observed responses. They can also be extended to multiple cause dropout; in this trial, missed assessments were due to a variety of reasons including patient-specific causes such as experiencing a protocol defined event, stopping therapy, or withdrawing consent and institution-specific causes such as understaffing an staff turnover. Therefore, some missingness is less likely to be informative; extensions will need to account for that. In addition, institutional differences might be addressed by allowing institution-specific parameters with priors that shrink them toward a common set of parameters.

For smaller sample sizes, WinBUGS has difficulty sampling from the posterior distribution of the parameters in the shrinkage model. In addition, the “monotonizing” approach ignores the intermittent missing data and may lead to biased results. These issues will be examined in the next Chapter.

2.10 Acknowledgments

This research was supported by NIH grants R01-CA85295, U10-CA37377, and U10-CA69974. The authors are grateful to oncologist Patricia Ganz at UCLA for providing her expertise for the MNAR analysis.

2.11 Tables and Figures

Table 2-1. Relative Risks to be Elicited

Question	Relative Risk	Drop out Rate p		
		p_1	p_2	...
100% confident the number is above	$r_{z,min}(p)$			
Best Guess	$r_{z,med}(p)$			
100% confident the number is below	$r_{z,max}(p)$			

Table 2-2. Percentiles of Relative Risks Elicited

Treatment	Percentile	Drop out Rate	
		10%	25%
Tamoxifen	Minimum	1.10	1.30
	Median	1.20	1.50
	Maximum	1.30	1.60
Placebo	Minimum	1.01	1.20
	Median	1.05	1.30
	Maximum	1.10	1.40

Table 2-3. Simulation Scenario

Parameter	Time Point							
	0	1	2	3	4	5	6	7
<i>Tamoxifen</i>								
α_0	-2.578	-2.500	-2.613	-2.752	-2.626	-2.789	-2.811	-2.895
α_1		2.460	1.978	1.940	2.023	2.072	1.885	2.007
α_2			1.500	1.599	1.389	1.612	1.639	1.830
γ_0		-2.352	-2.871	-2.625	-2.513	-2.281	-2.217	-2.536
γ_1		0.611	0.397	0.460	0.247	0.320	0.127	0.228
γ_2			0.121	0.422	0.261	0.035	0.293	0.204
<i>Depression Rate</i>	0.066	0.097	0.119	0.124	0.139	0.126	0.126	0.123
<i>Placebo</i>								
α_0	-2.653	-2.632	-2.59	-2.663	-2.598	-2.884	-2.853	-3.035
α_1		2.708	2.304	1.874	2.104	2.068	2.123	2.243
α_2			1.241	1.608	1.471	1.693	1.540	1.989
γ_0		-2.308	-2.970	-2.729	-2.474	-2.410	-2.460	-2.673
γ_1		0.466	0.468	0.469	0.272	0.376	0.088	0.001
γ_2			-0.293	0.323	0.278	0.288	0.241	0.428
<i>Depression Rate</i>	0.071	0.107	0.118	0.120	0.132	0.130	0.126	0.125

Table 2-4. Simulation Results: MSE ($\times 10^3$). P and T represent placebo and tamoxifen arms, respectively.

Model	Treat	Observed		$\mu_{j,z}^*$						
		Y	R	1	2	3	4	5	6	7
<i>Sample Size 3000</i>										
True	P	0.946	0.378	0.034	0.048	0.049	0.052	0.050	0.05	0.065
	T	1.075	0.431	0.026	0.035	0.046	0.056	0.052	0.078	0.061
Parametric	P	30.176	0.451	0.034	0.048	0.048	0.062	0.061	0.067	0.086
	T	28.882	0.385	0.026	0.037	0.056	0.078	0.073	0.096	0.077
Shrinkage	P	6.970	1.999	0.033	0.045	0.051	0.059	0.047	0.052	0.066
	T	6.988	2.401	0.024	0.026	0.056	0.053	0.063	0.119	0.073
Saturated	P	35.678	67.171	0.036	0.050	0.054	0.058	0.101	0.231	0.561
	T	34.654	62.606	0.026	0.033	0.045	0.059	0.097	0.329	0.722
<i>Sample Size 5000</i>										
True	P	0.659	0.238	0.023	0.027	0.027	0.028	0.036	0.043	0.058
	T	0.604	0.261	0.020	0.020	0.032	0.027	0.034	0.035	0.044
Parametric	P	30.029	0.381	0.023	0.027	0.028	0.033	0.044	0.052	0.073
	T	28.571	0.277	0.020	0.023	0.042	0.042	0.049	0.056	0.057
Shrinkage	P	4.628	1.188	0.025	0.028	0.032	0.031	0.035	0.048	0.057
	T	4.448	1.414	0.017	0.023	0.028	0.026	0.038	0.033	0.044
Saturated	P	30.274	54.647	0.023	0.028	0.028	0.033	0.061	0.138	0.290
	T	29.599	51.219	0.020	0.020	0.032	0.028	0.051	0.140	0.392
<i>Sample Size 10000</i>										
True	P	0.314	0.121	0.009	0.009	0.015	0.017	0.013	0.014	0.018
	T	0.278	0.126	0.011	0.013	0.013	0.013	0.015	0.019	0.020
Parametric	P	29.849	0.315	0.009	0.010	0.016	0.023	0.028	0.033	0.043
	T	28.418	0.223	0.011	0.014	0.02	0.026	0.028	0.039	0.043
Shrinkage	P	2.392	0.707	0.008	0.010	0.015	0.017	0.014	0.013	0.014
	T	2.474	0.712	0.011	0.015	0.011	0.015	0.016	0.019	0.023
Saturated	P	22.989	37.716	0.009	0.009	0.016	0.018	0.018	0.038	0.094
	T	22.245	34.791	0.011	0.013	0.014	0.014	0.021	0.048	0.128

Table 2-5. Patients Cumulative Drop Out Rate

Month		3	6	12	18	24	30	36
Tamoxifen	Available	5364	4874	4597	4249	3910	3529	3163
	Drop out	490	767	1115	1454	1835	2201	2447
	Drop Rate(%)	9.13	14.30	20.79	27.11	34.21	41.03	45.62
Placebo	Available	5375	4871	4624	4310	3951	3593	3297
	Drop out	504	751	1065	1424	1782	2078	2304
	Drop Rate(%)	9.38	13.97	19.81	26.49	33.15	38.66	42.87

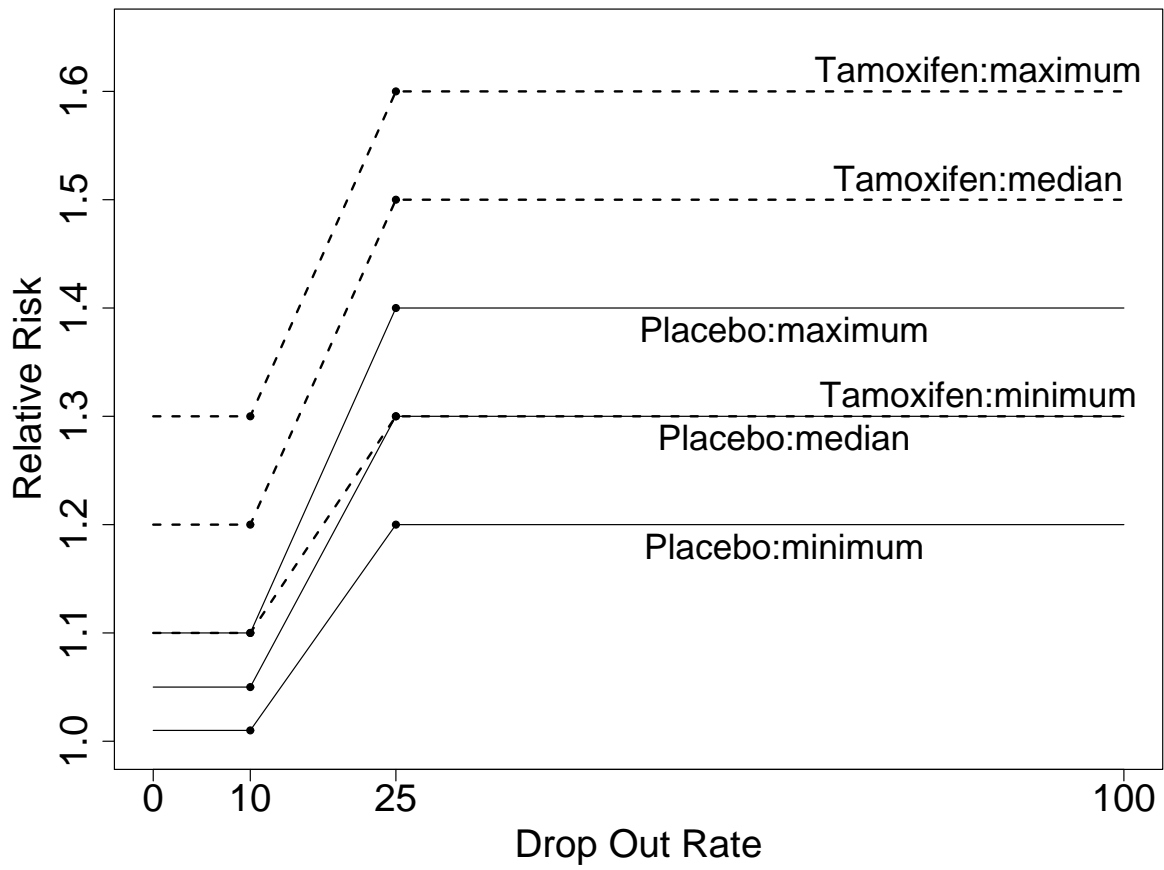


Figure 2-1. Extrapolation of the elicited relative risks.

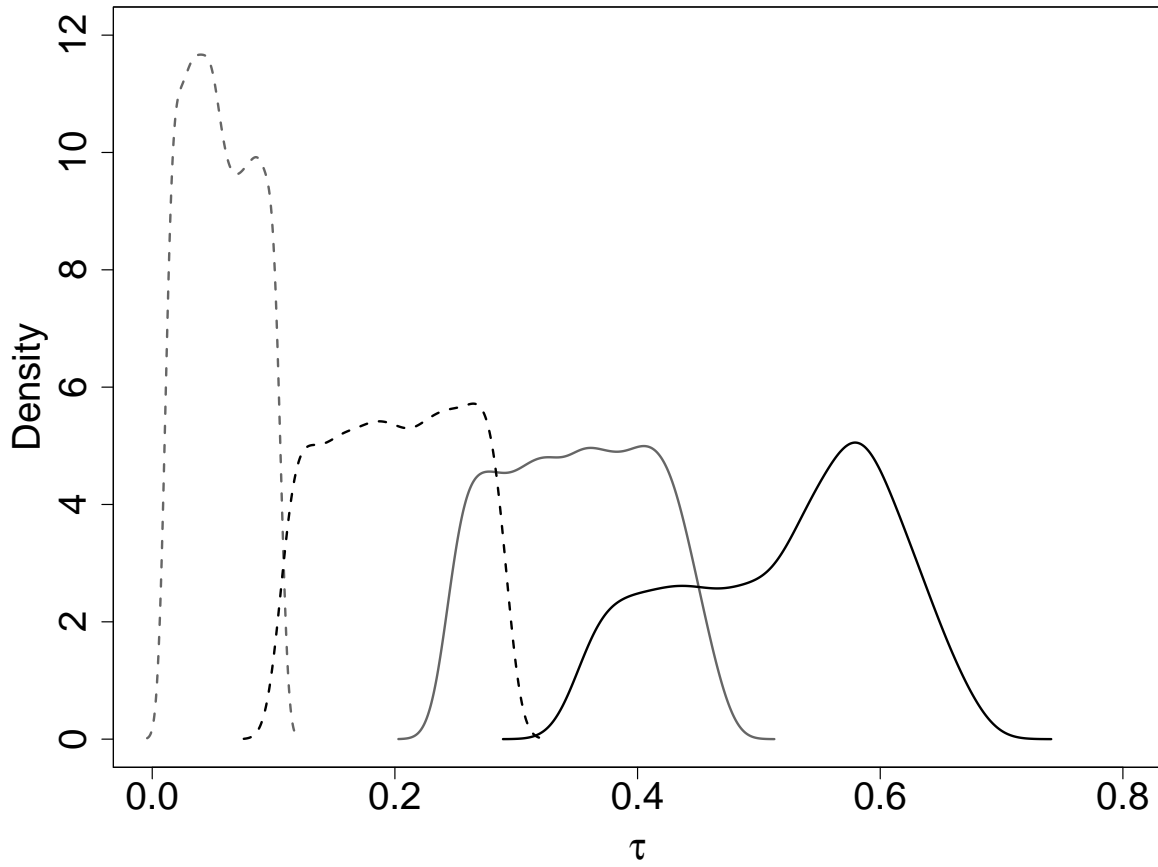


Figure 2-2. Prior conditional density $\tau_{z,j,\bar{\mathbf{y}}_{j-1}}$ given $p_{z,j}(\bar{\mathbf{y}}_{j-1})$. Black and gray lines represent tamoxifen and placebo arms, respectively. Solid and dashed lines are for $p_{z,j}(\bar{\mathbf{y}}_{j-1}) = 0.25$ and $p_{z,j}(\bar{\mathbf{y}}_{j-1}) = 0.10$, respectively.

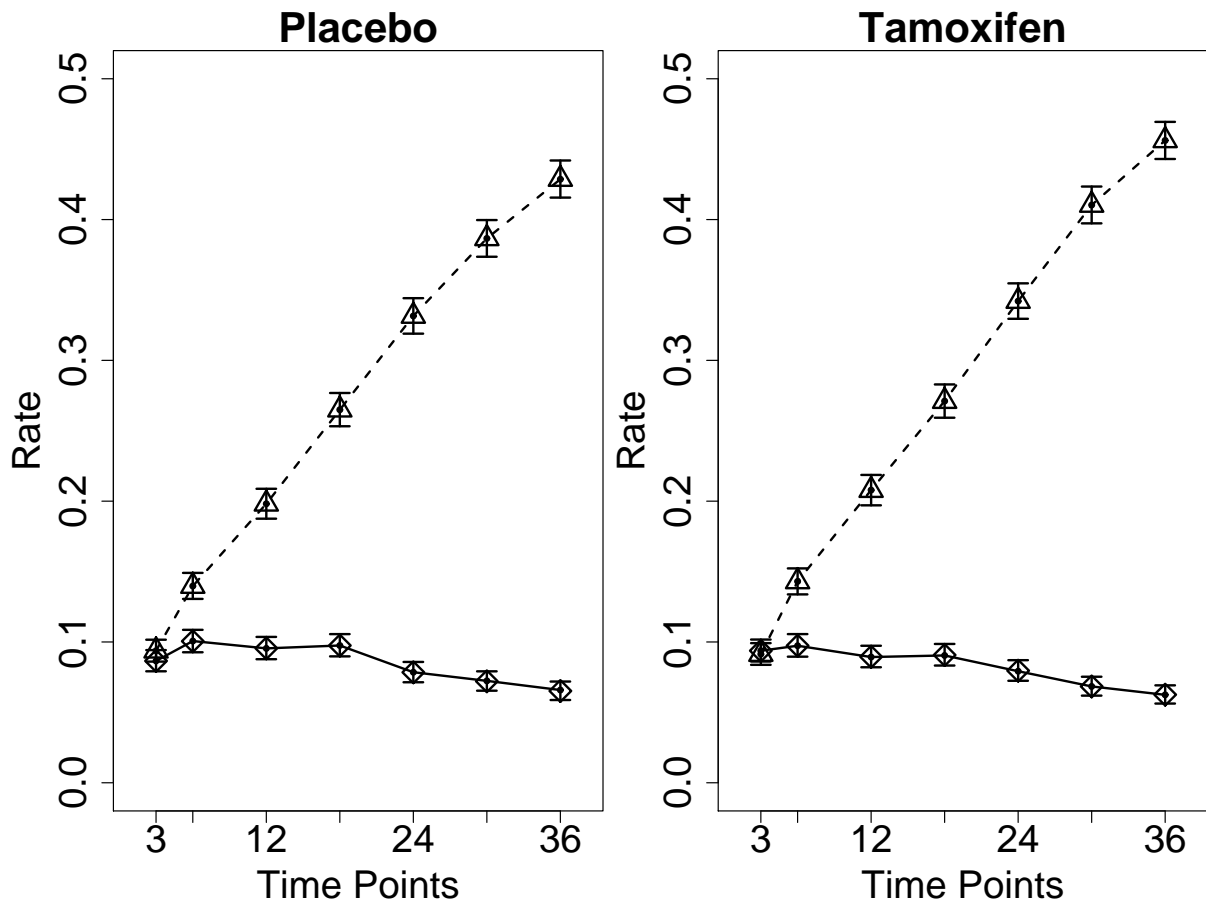


Figure 2-3. Solid and dashed lines represent the empirical rate of $P[Y_j = 1, R_j = 1 | Z = z]$ and $P[R_j = 0 | Z = z]$, respectively. The posterior means of $P[Y_j = 1, R_j = 1 | Z = z]$ (diamond) and $P[R_j = 0 | Z = z]$ (triangle) and their 95% credible intervals are displayed at each time point.

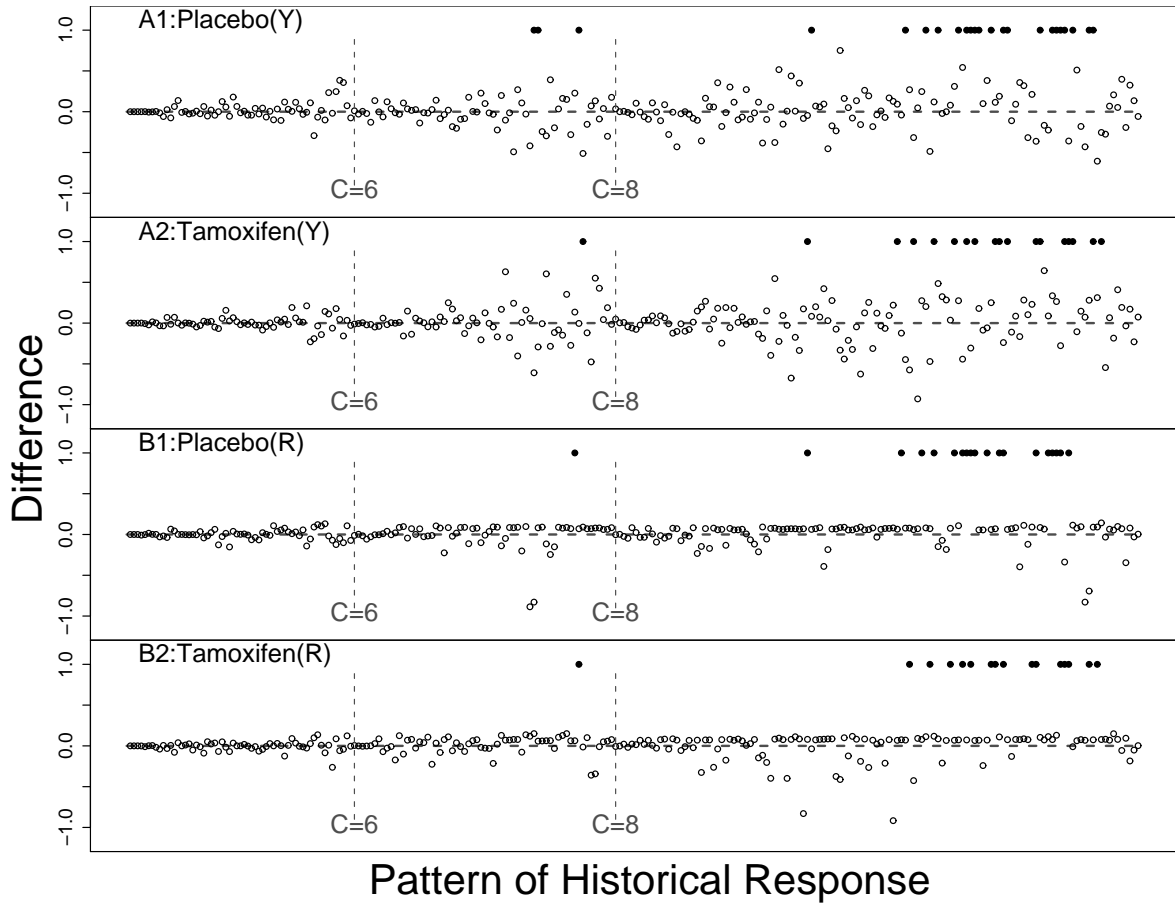


Figure 2-4. Differences between posterior mean and empirical rate of $P[Y_j = 1|R_j = 1, \bar{Y}_{j-1}, Z = z]$ (A1 and A2) and $P[R_j = 0|R_{j-1} = 1, \bar{Y}_{j-1}, Z = z]$ (B1 and B2). The x-axis is ordered by follow up time C ($\max\{t : R_t = 1\}$). The bullets are the posterior mean of $P[Y_j = 1|R_j = 1, \bar{Y}_{j-1}, Z = z]$ and $P[R_j = 0|R_{j-1} = 1, \bar{Y}_{j-1}, Z = z]$ when there are no patients with historical response \bar{Y}_{j-1} .

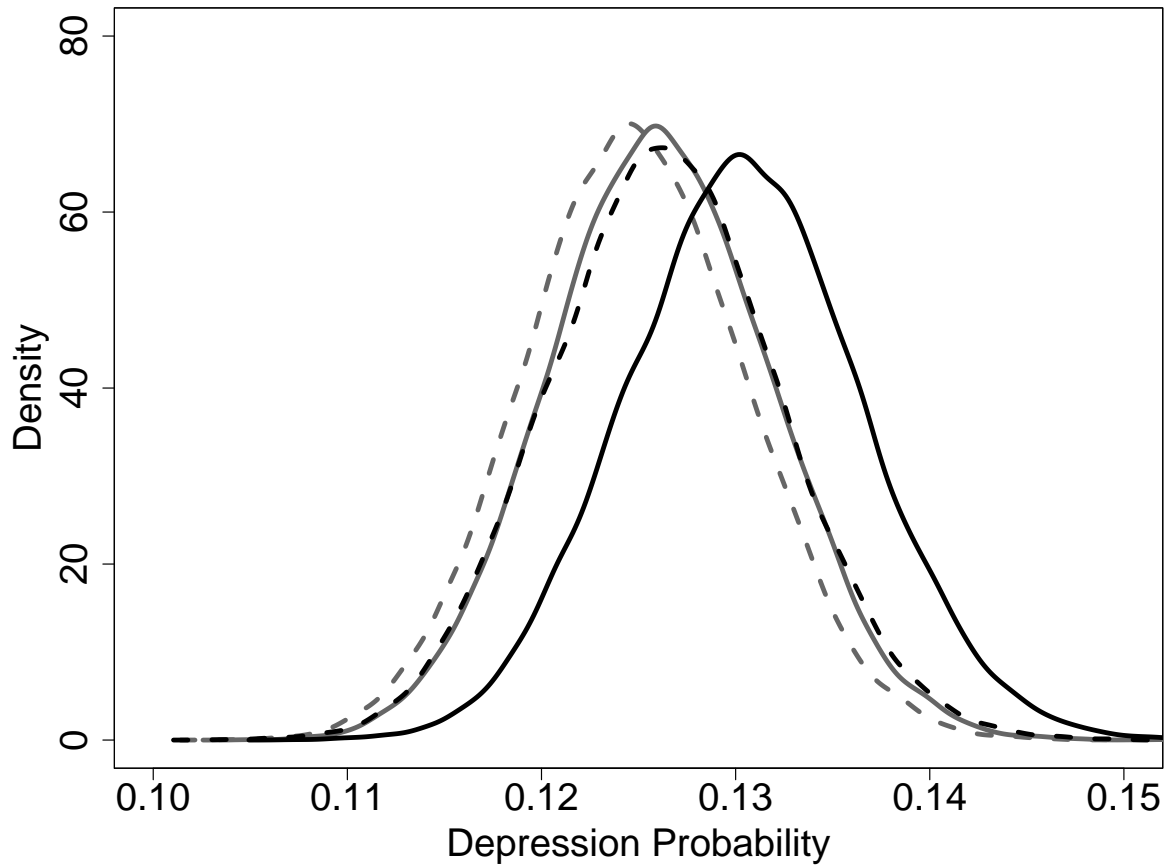


Figure 2-5. Posterior distribution of $P[Y_7 = 1|Z = z]$. Black and gray lines represent tamoxifen and placebo arms, respectively. Solid and dashed lines are for MNAR and MAR, respectively.

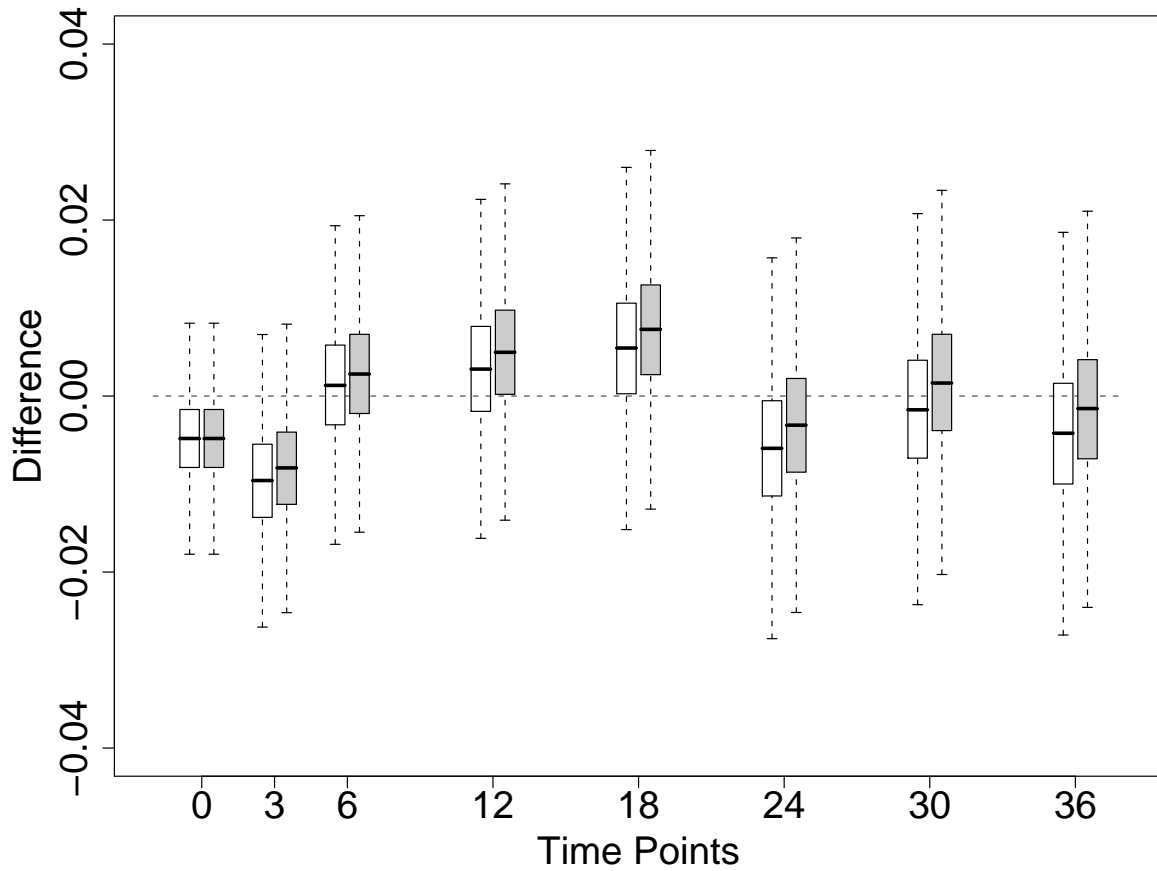


Figure 2-6. Posterior mean and 95% credible interval of difference of $P[Y_j = 1|Z = z]$ between placebo and tamoxifen arms. The gray and white boxes are for MAR and MNAR, respectively.

CHAPTER 3 A BETA-BINOMIAL BAYESIAN SHRINKAGE MODEL FOR INTERMITTENT MISSINGNESS LONGITUDINAL BINARY DATA

3.1 Introduction

We proposed in Chapter 2 a Bayesian shrinkage approach for longitudinal binary data with informative drop-out. The saturated observed data models were constructed sequentially via conditional distributions for response and for drop out time and parameterized on the logistic scale using all interaction terms. However, two issues were not addressed: the “ignored” intermittent missing data and the intrinsic computational challenge with the “interaction” parameterization. This Chapter proposes solutions to these two issues.

3.1.1 Intermittent Missing Data

In the BCPT, approximately 15% of the responses were intermittently missing, i.e. there are missing values prior to drop-out. One approach to handle intermittent missingness is to consider a “monotonized” dataset, whereby all CES-D scores observed on an individual after their first missing score are deleted, as in [Land et al. \(2002\)](#); we did this in Chapter 2. However, this increases the “drop-out” rate, throws away information and thus loses efficiency, and may introduce bias.

Handling informative intermittent missing data is methodologically and computationally challenging and, as a result, the statistics literature is relatively limited. Most methods adopt a likelihood approach and rely on strong parametric assumptions (see, for example, [Albert, 2000](#); [Albert et al., 2002](#); [Ibrahim et al., 2001](#); [Lin et al., 2004](#); [Troxel et al., 1998](#)). Semiparametric methods have been proposed by [Troxel et al. \(1998\)](#) and [Vansteelandt et al. \(2007\)](#). [Troxel et al. \(1998\)](#) proposed a marginal model and introduced a pseudo-likelihood estimation procedure. [Vansteelandt et al. \(2007\)](#) extended the ideas of [Rotnitzky et al. \(1998b\)](#), [Scharfstein et al. \(1999\)](#) and [Rotnitzky et al. \(2001\)](#) to non-monotone missing data.

Most related to our approach are the (partial ignorability) assumptions proposed in [Harel and Schafer \(2009\)](#) that partition the missing data and allow one (or more) of the partitions to be ignored given the other partition(s) and the observed data. In this Chapter, we apply a partial ignorability assumption such that the intermittent missing data mechanism can be ignored given drop-out and treatment strata.

3.1.2 Computational Issues

WinBUGS is a popular software package that allows convenient application of MCMC techniques. However, there are major drawbacks. For the shrinkage model proposed in Chapter 2, Section 2.5, WinBUGS has difficulty sampling from the posterior distribution of the parameters when sample size is relatively small (less than 3000 per arm). Tailored sampling algorithms can be written to overcome this difficulty, however, WinBUGS lacks the flexibility to incorporate modifications and/or extensions to its existing algorithms.

In this Chapter, we will provide an alternative parameterizations of the saturated model for the observed data as well as alternative shrinkage prior specifications to improve computational efficiency. This alternative approach to posterior sampling can easily be programmed in R.

3.1.3 Outline

This Chapter is organized as follows. In Section 3.2, we describe the data structure, formalize identification assumptions and prove that the treatment-specific distribution of the full trajectory of longitudinal outcomes is identified under these assumptions. In Section 3.3, we introduce a saturated model for the distribution of the data that would be observed when there is drop-out, but no intermittent observations. We then introduce shrinkage priors to parameters in the saturated model to reduce the dimensionality of the parameter space. In Section 3.4, we assess, by simulation, the behavior of three classes of models: parametric, saturated, and shrinkage. Our analysis of the BCPT trial is presented in Section 3.5. Section 3.6 is devoted to a summary and discussion.

3.2 Notation, Assumptions and Identifiability

To address the intermittent missingness, we redefine the notation in Chapter 2, Section 2.2, as well as introduce some additional notation in this Section. The following notation is defined for a random individual. When necessary, we use the subscript i to denote data for the i th individual.

Let Z denote the treatment assignment indicator, where $Z = 1$ denotes tamoxifen and $Z = 0$ denotes placebo. Let \mathbf{Y} be the complete response data vector with elements Y_j denoting the binary outcome (i.e., depression) scheduled to be measured at the j th visit ($j = 0(\text{baseline}), \dots, J$) and let $\bar{\mathbf{Y}}_j = (Y_0, \dots, Y_j)$ denote the history of the outcome process through visit j . Let \mathbf{R} be the vector of missing data indicators with the same dimension as \mathbf{Y} , such that $R_j = 1$ indicates Y_j is observed and $R_j = 0$ indicates Y_j is missing. Let $S = \max\{t : R_t = 1\}$ be the last visit at which an individual's depression status is recorded. If $S < J$, then we say that the individual has dropped out and S is referred to as the drop-out time. Let $\mathbf{R}_S = \{R_j : j < S\}$ be the collection of intermittent missing data indicators recorded prior to S .

We will find it useful to distinguish three sets of data for an individual: the complete data $\mathcal{C} = (Z, S, \mathbf{R}_S, \mathbf{Y})$, the full data $\mathcal{F} = (Z, S, \mathbf{R}_S, \bar{\mathbf{Y}}_S)$, and the observed data $\mathcal{O} = (Z, S, \mathbf{R}_S, \mathbf{Y}_{\text{obs}})$, where \mathbf{Y}_{obs} is the subset of \mathbf{Y} for which $R_j = 1$. It is useful to also define $\mathbf{Y}_{\text{mis}} = (\mathbf{Y}_{\text{mis}}^I, \mathbf{Y}_{\text{mis}}^C, \mathbf{Y}_{\text{mis}}^F)$, where $\mathbf{Y}_{\text{mis}}^I = \{Y_j : R_j = 0, j < S\}$ denotes the “intermittent” missing responses, $\mathbf{Y}_{\text{mis}}^C = \{Y_j : j = S + 1, j \leq J\}$ denotes the missing response at the time right after drop-out, and $\mathbf{Y}_{\text{mis}}^F = \{Y_j : S + 1 < j \leq J\}$ denotes the “future” missing responses. Note that $\bar{\mathbf{Y}}_S = (\mathbf{Y}_{\text{mis}}^I, \mathbf{Y}_{\text{obs}})$.

We assume that individuals are drawn as a simple random sample from a super-population so that we have an i.i.d. data structure for \mathcal{C} , \mathcal{F} and \mathcal{O} . We let the parameters θ_z index a model for the joint conditional distribution of S and $\bar{\mathbf{Y}}_S$ given $Z = z$ and the parameters $\phi_{s,z}$ index a model for the conditional distribution of \mathbf{R}_S given

$S = s, \bar{\mathbf{Y}}_S$ and $Z = z$. We assume that the parameters θ_z and $\phi_z = (\phi_{1,z}, \dots, \phi_{J,z})$ are distinct.

Our goal is to draw inference about $\mu_{z,j}^* = P[Y_j = 1|Z = z]$ for $j = 1, \dots, J$ and $z = 0, 1$. To identify $\mu_{z,j}^*$ from the distribution of the observed data, we make the following three (untestable) assumptions:

Assumption 1: Given Z and S , the intermittent missing data are missing at random, i.e.,

$$\mathbf{R}_S \perp\!\!\!\perp \mathbf{Y}_{\text{mis}}^I \mid Z, S, \mathbf{Y}_{\text{obs}}.$$

Under this assumption the parameters of the joint conditional distribution of S and $\bar{\mathbf{Y}}_S$ given $Z = z$ are estimable from the distribution of the observed data.

This assumption plus the assumption that θ_z is *a priori* independent of ϕ_z implies that the intermittent missingness mechanism is ancillary or ignorable. Specifically, this means that when considering inferences about θ_z from a likelihood perspective, as we are in this paper, the conditional distribution of \mathbf{R}_S given Z, S and \mathbf{Y}_{obs} does not contribute to the likelihood and can be ignored (Harel and Schafer, 2009).

Assumptions 2 and 3 are the same as Assumptions 1 and 2 in Chapter 2, Section 2.3, respectively. We restate below the two assumption using the “survival” time S notation (instead of missing indicators R in Chapter 2).

Assumption 2 (Non-Future Dependence): For $j = 1, \dots, J$,

$$P[S = j - 1 \mid S \geq j - 1, \mathbf{Y}] = P[S = j - 1 \mid S \geq j - 1, \bar{\mathbf{Y}}_j]$$

Assumption 3 (Pattern-Mixture Representation): For $j = 1, \dots, J$ and $y_j = 0, 1$,

$$P[Y_j = y_j \mid S = j - 1, \bar{\mathbf{Y}}_{j-1}, Z = z] = \frac{P[Y_j = y_j \mid S \geq j, \bar{\mathbf{Y}}_{j-1}, Z = z] \exp\{q_{z,j}(\bar{\mathbf{Y}}_{j-1}, y_j)\}}{E[\exp\{q_{z,j}(\bar{\mathbf{Y}}_{j-1}, Y_j)\} \mid S \geq j, \bar{\mathbf{Y}}_{j-1}, Z = z]}$$

where $q_{z,j}(\bar{\mathbf{Y}}_{j-1}, Y_j)$ is a specified function of its arguments.

Theorem 1: $P[Y_j = 1 \mid S \geq k - 1, \bar{\mathbf{Y}}_{k-1} = \bar{\mathbf{y}}_{k-1}, Z = z]$ and $P[Y_j = 1 \mid S = k - 1, \bar{\mathbf{Y}}_{k-1} = \bar{\mathbf{y}}_{k-1}, Z = z]$ are identified for $k = 1, \dots, j$ under Assumptions 1-3.

Proof: Under Assumption 1, we know that the parameters of the conditional joint distribution of S and \bar{Y}_S given $Z = z$ are estimable from the distribution of the observed data. The rest of the proof is the same as in Chapter 2.

The identifiability result shows that, given the functions $q_{z,j}(\bar{\mathbf{Y}}_{j-1}, Y_j)$, $\mu_{z,j}^*$ can be expressed as functional of the distribution of the observed data.

3.3 Modeling, Prior Specification and Posterior Computation

3.3.1 Modeling

We reparameterize the saturated observed data model in Chapter 2 as follows:

$$\begin{aligned}
 P[Y_0 = 1] &= \alpha_{z,0} \\
 P[Y_1 = 1 | S \geq 1, Y_0 = y, Z = z] &= \alpha_{z,1,y} \\
 P[Y_j = 1 | S \geq j, Y_{j-1} = y, \bar{\mathbf{Y}}_{j-2} = \bar{\mathbf{y}}_{j-2}, Z = z] &= \alpha_{z,j,\bar{\mathbf{y}}_{j-2},y} \\
 P[S = 0 | Y_0 = y, Z = z] &= \gamma_{z,0,y} \\
 P[S = j - 1 | S \geq j - 1, Y_{j-1} = y, \bar{\mathbf{Y}}_{j-2} = \bar{\mathbf{y}}_{j-2}, Z = z] &= \gamma_{z,j-1,\bar{\mathbf{y}}_{j-2},y}
 \end{aligned} \tag{3-1}$$

for $j = 2, \dots, J$ and $y = 0, 1$.

Let α_z denote the parameters indexing the first set of models for response and γ_z denote the parameters indexing the second set of models for drop-out. Recall that we defined θ_z to denote the parameters of the conditional distribution of S and \bar{Y}_S given $Z = z$; thus, $\theta_z = (\alpha_z, \gamma_z)$.

This saturated model avoids the complex interaction term model parameterization. As a result, the (conditional) posterior distributions of θ_z will have simple forms and efficient posterior sampling is possible even when the sample size is moderate or small.

We use the same parameterization of the functions $q_{z,j}(\bar{\mathbf{Y}}_{j-1}, Y_j)$ as in Chapter 2, Section 2.5.

3.3.2 Shrinkage Prior

In Chapter 2, the strategy to avoid the curse of dimensionality was to apply shrinkage priors for higher order interactions to reduce the number of parameters

(i.e., shrink them to zero). For the directly parameterized model 3–1, we use a different shrinkage strategy. In particular, we propose to use Beta priors for shrinkage as follows:

$$\begin{aligned}\alpha_{z,j,\bar{\mathbf{y}}_{j-2,y}} &\sim \text{Beta}\left(m_{z,j,y}^{(\alpha)}/\eta_{z,j,y}^{(\alpha)}, (1 - m_{z,j,y}^{(\alpha)})/\eta_{z,j,y}^{(\alpha)}\right) \\ \gamma_{z,j-1,\bar{\mathbf{y}}_{j-2,y}} &\sim \text{Beta}\left(m_{z,j-1,y}^{(\gamma)}/\eta_{z,j-1,y}^{(\gamma)}, (1 - m_{z,j-1,y}^{(\gamma)})/\eta_{z,j-1,y}^{(\gamma)}\right).\end{aligned}\tag{3–2}$$

for $j = 2, \dots, J$ and $y = 0, 1$. For $\alpha_{z,0}$, $\alpha_{z,1,y}$ and $\gamma_{z,0,y}$ for $y = 0, 1$, we assign $\text{Unif}(0, 1)$ priors. Let $\mathbf{m}_z^{(\alpha)}$ ($\mathbf{m}_z^{(\gamma)}$) and $\boldsymbol{\eta}_z^{(\alpha)}$ ($\boldsymbol{\eta}_z^{(\gamma)}$) denote the parameters $m_{z,j,y}^{(\alpha)}$ ($m_{z,j-1,y}^{(\gamma)}$) and $\eta_{z,j,y}^{(\alpha)}$ ($\eta_{z,j-1,y}^{(\gamma)}$) respectively.

Note that for a random variable X that follows a $\text{Beta}(m/\eta, (1 - m)/\eta)$ distribution, we have

$$E[X] = m \quad \text{and} \quad \text{Var}[X] = m(1 - m) \times \frac{\eta}{\eta + 1}.$$

For fixed m , $\text{Var}[X] \rightarrow 0$ as $\eta \rightarrow 0$, indicating shrinkage of the distribution of X toward the mean. Thus, $\eta_{z,j,y}^{(\alpha)}$ and $\eta_{z,j-1,y}^{(\gamma)}$ serve as shrinkage parameters for $\alpha_{z,j,\bar{\mathbf{y}}_{j-2,y}}$ and $\gamma_{z,j-1,\bar{\mathbf{y}}_{j-2,y}}$ respectively. As the shrinkage parameters go to zero, the distribution of the probabilities $\alpha_{z,j,\bar{\mathbf{y}}_{j-2,y}}$ and $\gamma_{z,j-1,\bar{\mathbf{y}}_{j-2,y}}$ are shrunk toward the mean of the probabilities that do not depend on $\bar{\mathbf{y}}_{j-2}$, namely $m_{z,j,y}^{(\alpha)}$ and $m_{z,j-1,y}^{(\gamma)}$, respectively. In essence, the model is being shrunk toward a first-order Markov model. The shrinkage priors allow “neighboring cells” to borrow information from each other and provide more precise inferences.

Theorem 2: When there is no intermittent missingness, the proposed model yields consistent posterior means of the observed data probabilities, as long as all the true values of the observed data probabilities are in the open interval $(0, 1)$.

Proof: See Appendix.

We specify independent $\text{Unif}(0, 1)$ priors for $m_{z,j,y}^{(\alpha)}$ and $m_{z,j-1,y}^{(\gamma)}$. For the shrinkage parameters $\eta_{z,j,y}^{(\alpha)}$ and $\eta_{z,j-1,y}^{(\gamma)}$, we specify independent, uniform shrinkage priors (Daniels,

1999) as follows

$$\eta_{z,j,y}^{(\alpha)} \sim \frac{g\left(\mathbf{E}_{z,j,y}^{(\alpha)}\right)}{\left(g\left(\mathbf{E}_{z,j,y}^{(\alpha)}\right)\eta_{z,j,y}^{(\alpha)} + 1\right)^2} \quad \text{and} \quad \eta_{z,j-1,y}^{(\gamma)} \sim \frac{g\left(\mathbf{E}_{z,j-1,y}^{(\gamma)}\right)}{\left(g\left(\mathbf{E}_{z,j-1,y}^{(\gamma)}\right)\eta_{z,j-1,y}^{(\gamma)} + 1\right)^2}, \quad (3-3)$$

where

- $g(\cdot)$ is a summary function (e.g., minimum, median or maximum, as suggested in Christiansen and Morris (1997)).
- $\mathbf{E}_{z,j,y}^{(\alpha)} = \{e_{z,j,\bar{\mathbf{y}}_{j-2},y}^{(\alpha)} : \text{the expected number of subjects with } S \geq j, Y_{j-1} = y, \bar{\mathbf{Y}}_{j-2} = \bar{\mathbf{y}}_{j-2}, Z = z\}$.
- $\mathbf{E}_{z,j-1,y}^{(\gamma)} = \{e_{z,j-1,\bar{\mathbf{y}}_{j-2},y}^{(\gamma)} : \text{the expected number of subjects with } S \geq j-1, Y_{j-1} = y, \bar{\mathbf{Y}}_{j-2} = \bar{\mathbf{y}}_{j-2}, Z = z\}$.

The expected number of subjects with $S \geq j, Y_{j-1} = y, \bar{\mathbf{Y}}_{j-2} = \bar{\mathbf{y}}_{j-2}, Z = z$ and with $S \geq j-1, Y_{j-1} = y, \bar{\mathbf{Y}}_{j-2} = \bar{\mathbf{y}}_{j-2}, Z = z$ can be computed as:

$$\begin{aligned} e_{z,j,\bar{\mathbf{y}}_{j-2},y}^{(\alpha)} &= n_z \sum_{s=j}^J \sum_{y_j, y_{j+1}, \dots, y_s} P[S = s, Y_s = y_s, \dots, Y_j = y_j, Y_{j-1} = y, \bar{\mathbf{Y}}_{j-2} = \bar{\mathbf{y}}_{j-2} | Z = z] \\ e_{z,j-1,\bar{\mathbf{y}}_{j-2},y}^{(\gamma)} &= n_z \sum_{s=j-1}^J \sum_{y_j, y_{j+1}, \dots, y_s} P[S = s, Y_s = y_s, \dots, Y_j = y_j, Y_{j-1} = y, \bar{\mathbf{Y}}_{j-2} = \bar{\mathbf{y}}_{j-2} | Z = z] \end{aligned} \quad (3-4)$$

where the probabilities on the right hand side of the above equations are estimable under Assumption 1.

The expected sample sizes above are used in the prior instead of the observed binomial sample sizes which are not completely determined due to the intermittent missingness. Thus, our formulation of these priors induces a small additional amount of data dependence beyond its standard dependence on the binomial sample sizes. This additional dependence affects the median of the prior but not its diffuseness.

3.3.3 Prior of Sensitivity Parameters

We use the same approach as in Chapter 2, Section 2.6.2 for constructing priors of τ_z given θ_z .

3.3.4 Posterior Computation

Compared to Chapter 2, posterior computations for the observed data model are much easier and more efficient under the reparameterized model 3–1 and the Beta shrinkage priors. The posterior sampling algorithms can be implemented in R with no sample size restrictions.

The following steps are used to simulate draws from the posterior of $\mu_{z,j}^*$:

1. Sample $P(\theta_z, \mathbf{Y}_{\text{mis}}^l | \mathbf{Y}_{\text{obs}}, S, \mathbf{R}_S, Z = z)$ using Gibbs sampling with data augmentation (see details in Appendix). Continue sampling until convergence.
2. For each draw of $\gamma_{z,j-1, \bar{Y}_{j-2, Y_{j-1}}}$, draw $\tau_{z,j, \bar{Y}_{j-1}}$ based on the conditional priors described in Section 2.6.2.
3. Compute $\mu_{z,j}^*$ by plugging the draws of $\alpha_{z, \bar{Y}_{j-2, Y}}$, $\gamma_{z, \bar{Y}_{j-2, Y}}$, and $\tau_{z,j, \bar{Y}_{j-1}}$ into the identification algorithm discussed in Section 2.4.

3.4 Assessment of Model Performance via Simulation

For assessment of model performance, we use the same “true” parametric model as in Chapter 2, Section 2.7 to simulate observed data (no intermittent missingness). We again compared the performance of our shrinkage model with (1) a correct parametric model, (2) an incorrect parametric model (first order Markov model) and (3) a saturated model (with diffuse priors). Our shrinkage model uses the shrinkage priors proposed in Section 3.3.2.

We considered small (500), moderate (2000), large (5000) and very large (1,000,000) sample sizes for each treatment arm; for each sample size, we simulated 500 datasets. We assessed model performance using mean squared error (MSE).

In Table 3-2 (sample size 1,000,000 not shown), we report the MSE’s of $P[Y_j = 1 | S \geq j, \bar{Y}_{j-1}, Z = z]$ and $P[S = j - 1 | S \geq j - 1, \bar{Y}_{j-1}, Z = z]$ averaged over all j and all \bar{Y}_{j-1} (see columns 3 and 4, respectively). We also report the MSE’s for $\mu_{z,j}^*$ (see columns 6-12). For reference, the MSE’s associated with the true data generating model are bolded. At all sample sizes, the shrinkage model has lower MSE’s for the

rates of depression at times 3-7 than the incorrectly specified parametric model and the saturated model. Our simulation results show that as sample size goes to infinity (e.g. very large, 1,000,000), both the shrinkage model and the saturated model converge to the true values of $\mu_{z,j}^*$, whereas the incorrectly specified parametric model yields biased estimates.

In addition, the MSE's for the parameters $\mu_{z,j}^*$ in the shrinkage model compare favorably with those of the true parametric model for all sample sizes considered, despite the fact that the shrinkage priors were specified to shrink toward an incorrect model.

3.5 Application: Breast Cancer Prevention Trial (BCPT)

Table 3-1 displays the treatment-specific drop-out and intermittent missing rates in the BCPT. By the 7th study visit (36 months), more than 30% of patients had dropped out in each treatment arm, with a slightly higher percentage in the tamoxifen arm.

3.5.1 Model Fit

We fit the shrinkage model to the observed data using R, with multiple chains of 5000 iterations and 1000 burn-in. Convergence was checked by examining trace plots of the multiple chains. We defined $g(\cdot)$ in the priors for the hyperparameters (Equation 3-3) to be the maximum function. To compute the expected number of subjects $e_{z,j,\bar{\mathbf{y}}_{j-2},y}^{(\alpha)}$ and $e_{z,j-1,\bar{\mathbf{y}}_{j-2},y}^{(\gamma)}$ in Equation (3-4), we assigned a point mass prior at 0.5 to all $\mathbf{m}_z^{(\alpha)}$, $\mathbf{m}_z^{(\gamma)}$, $\boldsymbol{\eta}_z^{(\alpha)}$ and $\boldsymbol{\eta}_z^{(\gamma)}$ (which corresponds to Unif(0, 1) priors on $\alpha_{z,\bar{\mathbf{y}}_{j-2},y}$ and $\gamma_{z,\bar{\mathbf{y}}_{j-2},y}$) and sampled $\alpha_{z,\bar{\mathbf{y}}_{j-2},y}$ and $\gamma_{z,\bar{\mathbf{y}}_{j-2},y}$ using Step 1 in the algorithm described in Section 3.3.4. To avoid data sparsity, we calculated $P[S = s, \bar{\mathbf{Y}}_s = \bar{\mathbf{y}}_s]$ using the posterior mean of $\alpha_{z,\bar{\mathbf{y}}_{j-2},y}$ and $\gamma_{z,\bar{\mathbf{y}}_{j-2},y}$ rather than the empirical probabilities.

To assess model fit, we compared the empirical rates and posterior means (with 95% credible intervals) of $P[Y_j = 1, S \geq j | Z = z]$ and $P[S < j | Z = z]$. As shown in Figure 3-1, the shrinkage model fits the observed data well.

Figure 3-2 illustrates the effect of shrinkage on the model fit by comparing the difference between the empirical rates and posterior means of $P[Y_j = 1|S \geq j, \bar{\mathbf{Y}}_{j-1} = \bar{\mathbf{y}}_{j-1}, Z = z]$ for the tamoxifen arm ($Z = 1$) and $j = 6, 7$. We use the later time points to illustrate this since the observed data were more sparse and the shrinkage effect was more apparent. The empirical depression rates often reside on the boundary (0 or 1). In some cases, there are no observations within "cells", thus the empirical rates were undefined. From the simulation results in Section 2.7, we know that the empirical estimates are less reliable for later time points. Via the shrinkage priors, the probabilities $P[Y_j = 1|S \geq j, Y_{j-1} = y_{j-1}, \bar{\mathbf{Y}}_{j-2} = \bar{\mathbf{y}}_{j-2}, Z = z]$ with the same $\bar{\mathbf{y}}_{j-2}$ are shrunk together and away from the boundaries. By borrowing information across neighboring cells, we are able to estimate $P[Y_j = 1|S \geq j, \bar{\mathbf{Y}}_{j-1}, Z = z]$ for all j, z and $\bar{\mathbf{Y}}_{j-1}$ with better precision. The differences between the empirical rates and the posterior means illustrate the magnitude of the shrinkage effect. In the BCPT, the depression rate was (relatively) low and there were few subjects at the later times that were observed with a history of mostly depression at the earlier visits; as a result, the differences were larger when $\bar{\mathbf{Y}}_{j-1}$ had a lot of 1's (depression).

3.5.2 Inference

Figure 3-3 shows the posterior of $P[Y_7 = 1|Z = z]$, the treatment-specific probability of depression at the end of the 36-month follow up (solid lines). For comparison, the posterior under MAR (corresponding to point mass priors for τ at zero) is also presented (dashed lines). The observed depression rates (i.e., complete case analysis) were 0.124 and 0.112 for the placebo and tamoxifen arms, respectively. Under the MNAR analysis (using the elicited priors), the posterior mean of the depression rates at month 36 were 0.133 (95% CI : 0.122, 0.144) and 0.125 (95% CI : 0.114, 0.136) for the placebo and tamoxifen arms; the difference was -0.007 (95% CI : $-0.023, 0.008$). Under MAR, the rates were 0.132 (95% CI : 0.121, 0.143) and 0.122 (95% CI : 0.111, 0.133) for the placebo and tamoxifen arms; the difference was -0.01 (95% CI : $-0.025, 0.005$).

The posterior probability of depression was higher under the MNAR analysis than the MAR analysis since researchers believed depressed patients were more likely to drop out (see Table 2-2), a belief that was captured by the elicited priors. Figure 3-4 shows that under the two treatments there were no significant differences in the depression rates at any measurement time (95% credible intervals all cover zero) under both MNAR and MAR. Similar (non-significant) treatment differences were seen when examining treatment comparisons conditional on depression status at baseline.

3.5.3 Sensitivity of Inference to the Priors

To assess the sensitivity of inference on the 36 month depression rates to the elicited (informative) priors $\{r_{min}, r_{med}, r_{max}\}$, we considered several alternative scenarios based on Table 1. In the first scenario, we made the priors more or less informative by scaling the range, but leaving the median unchanged. That is, we considered increasing (or decreasing) the range by a scale factor v to $\{r_{med} - v(r_{med} - r_{min}), r_{med}, r_{med} + v(r_{max} - r_{med})\}$. In the second scenario, we shifted the prior by a factor u , $\{u + r_{min}, u + r_{med}, u + r_{max}\}$.

The posterior mean and between-treatment difference of the depression rate at month 36 with 95% CI are given in Tables 3-3 and 3-4. None of the scenarios considered resulted in the 95% CI for the difference in rates of depression at 36 months that excluded zero except for the (extreme) scenario where the elicited tamoxifen intervals were shifted by 0.5 and the elicited placebo intervals were shifted by -0.5 .

We also assessed the impact of switching the priors for the placebo and tamoxifen arms; in this case, the posterior means were 0.135 (95% CI : 0.124, 0.146) and 0.123 (95% CI : 0.112, 0.134) for the placebo and tamoxifen arms respectively, while the difference was -0.012 (95% CI : -0.027 , 0.004).

3.6 Summary and Discussion

In this Chapter, we extended the Bayesian shrinkage approach proposed in Chapter 2 for intermittent missingness. In addition, we reparameterized the saturated observed data model and dramatically improved the computational efficiency.

WinBUGS can still be applied for the reparameterized model when there is no intermittent missing data. However, with the intermittent missingness, the augmentation step in the posterior computation requires extensive programming in WinBUGS. Nevertheless, the approach in Chapter 2 may still be preferred in certain cases, e.g. for directly shrinking the interaction terms.

As an extension, we might consider alternatives to the partial ignorability assumption (Assumption 1) which has been widely used, but questioned by some (Robins, 1997).

3.7 Tables and Figures

Table 3-1. Missingness by Scheduled Measurement Time

	Time Point j (Month)						
	1(3)	2(6)	3(12)	4(18)	5(24)	6(30)	7(36)
<i>Tamoxifen (Total N = 5364, Overall Missing 34.94%)</i>							
Intermittent Missing	330	224	190	200	203	195	
Drop-out at j	160	122	259	280	332	352	369
Cumulative Drop-out	160	282	541	821	1153	1505	1874
<i>Placebo (Total N = 5375, Overall Missing 31.83%)</i>							
Intermittent Missing	347	215	153	181	199	197	
Drop-out at j	157	106	247	287	309	272	333
Cumulative Drop-out	157	263	510	797	1106	1378	1711

Table 3-2. Simulation Results: MSE ($\times 10^3$). P and T represent placebo and tamoxifen arms, respectively.

Model	Treat	Observed		$\mu_{j,z}^*$ (Month)						
		Y	R	1(3)	2(6)	3(12)	4(18)	5(24)	6(30)	7(36)
<i>Sample Size 500</i>										
True	P	6.209	2.474	0.199	0.225	0.258	0.313	0.319	0.352	0.390
	T	6.790	2.789	0.205	0.228	0.297	0.344	0.331	0.405	0.428
Parametric	P	33.351	1.511	0.199	0.227	0.26	0.317	0.323	0.349	0.388
	T	32.323	1.602	0.205	0.226	0.292	0.345	0.333	0.403	0.425
Shrinkage	P	29.478	2.310	0.202	0.226	0.252	0.303	0.312	0.337	0.372
	T	28.410	2.365	0.212	0.232	0.294	0.336	0.330	0.390	0.419
Saturated	P	57.107	111.263	0.202	0.228	0.302	0.490	1.083	2.401	4.427
	T	55.582	104.882	0.211	0.245	0.383	0.657	1.352	3.167	5.782
<i>Sample Size 2000</i>										
True	P	1.474	0.586	0.052	0.058	0.063	0.078	0.081	0.086	0.097
	T	1.610	0.634	0.050	0.063	0.062	0.080	0.093	0.095	0.101
Parametric	P	30.507	0.543	0.051	0.055	0.064	0.081	0.090	0.091	0.108
	T	29.168	0.495	0.050	0.064	0.071	0.086	0.101	0.110	0.121
Shrinkage	P	23.545	0.647	0.053	0.056	0.063	0.078	0.082	0.084	0.095
	T	22.598	0.615	0.050	0.063	0.063	0.080	0.093	0.095	0.102
Saturated	P	40.322	77.627	0.053	0.057	0.069	0.100	0.188	0.457	0.946
	T	38.943	72.731	0.050	0.064	0.067	0.110	0.218	0.560	1.223
<i>Sample Size 5000</i>										
True	P	0.594	0.234	0.020	0.024	0.026	0.033	0.031	0.040	0.036
	T	0.623	0.265	0.024	0.024	0.028	0.035	0.033	0.039	0.040
Parametric	P	29.983	0.379	0.020	0.025	0.029	0.037	0.043	0.049	0.055
	T	28.616	0.298	0.024	0.025	0.035	0.048	0.045	0.060	0.059
Shrinkage	P	18.83	0.394	0.020	0.024	0.026	0.033	0.031	0.039	0.036
	T	18.055	0.322	0.024	0.024	0.028	0.036	0.034	0.040	0.041
Saturated	P	30.071	54.454	0.020	0.024	0.027	0.038	0.052	0.13	0.270
	T	29.156	50.590	0.024	0.024	0.029	0.039	0.059	0.148	0.373

Table 3-3. Sensitivity to the Elicited Prior

Treatment	Percentile	Scenario (T:Tamoxifen, P:Placebo)							
		$v^T = 5, v^P = 5$		$v^T = 0.2, v^P = 0.2$		$u^T = 0.5, u^P = 0.5$		$u^T = -0.5, u^P = -0.5$	
		10%	25%	10%	25%	10%	25%	10%	25%
Tamoxifen	Minimum	0.79	0.50	1.18	1.46	1.60	1.80	0.60	0.80
	Median	1.20	1.50	1.20	1.50	1.70	2.00	0.70	1.00
	Maximum	1.70	2.00	1.22	1.52	1.80	2.10	0.80	1.10
	$P[Y_7 = 1](95\% \text{ CI})$	0.125(0.114, 0.136)		0.125(0.114, 0.136)		0.132(0.120, 0.143)		0.117(0.107, 0.128)	
Placebo	Minimum	0.85	0.80	1.04	1.28	1.51	1.70	0.51	0.70
	Median	1.05	1.30	1.05	1.30	1.55	1.80	0.55	0.80
	Maximum	1.30	1.80	1.06	1.32	1.60	1.90	0.60	0.90
	$P[Y_7 = 1](95\% \text{ CI})$	0.133(0.122, 0.144)		0.133(0.122, 0.144)		0.139(0.128, 0.150)		0.125(0.114, 0.135)	
	Difference of $P[Y_7 = 1](95\% \text{ CI})$	-0.008(-0.024, 0.008)		-0.007(-0.023, 0.008)		-0.007(-0.023, 0.009)		-0.008(-0.023, 0.007)	

Table 3-4. Sensitivity to the Elicited Prior

Treatment	Percentile	Scenario (T:Tamoxifen, P:Placebo)							
		$v^T = 5, v^P = 0.2$		$v^T = 0.2, v^P = 5$		$u^T = 0.5, u^P = -0.5$		$u^T = -0.5, u^P = 0.5$	
		10%	25%	10%	25%	10%	25%	10%	25%
Tamoxifen	Minimum	0.79	0.50	1.18	1.46	1.60	1.80	0.60	0.80
	Median	1.20	1.50	1.20	1.50	1.70	2.00	0.70	1.00
	Maximum	1.70	2.00	1.22	1.52	1.80	2.10	0.80	1.10
	$P[Y_7 = 1](95\% \text{ CI})$	0.125(0.114, 0.136)		0.125(0.114, 0.136)		0.132(0.121, 0.143)		0.117(0.107, 0.128)	
Placebo	Minimum	1.04	1.28	0.85	0.80	0.51	0.70	1.51	1.70
	Median	1.05	1.30	1.05	1.30	0.55	0.80	1.55	1.80
	Maximum	1.06	1.32	1.30	1.80	0.60	0.90	1.60	1.90
	$P[Y_7 = 1](95\% \text{ CI})$	0.133(0.122, 0.144)		0.133(0.122, 0.144)		0.125(0.114, 0.135)		0.139(0.128, 0.150)	
	Difference of $P[Y_7 = 1](95\% \text{ CI})$	-0.008(-0.024, 0.008)		-0.008(-0.023, 0.008)		0.007(-0.008, 0.023)		-0.022(-0.037, -0.006)	

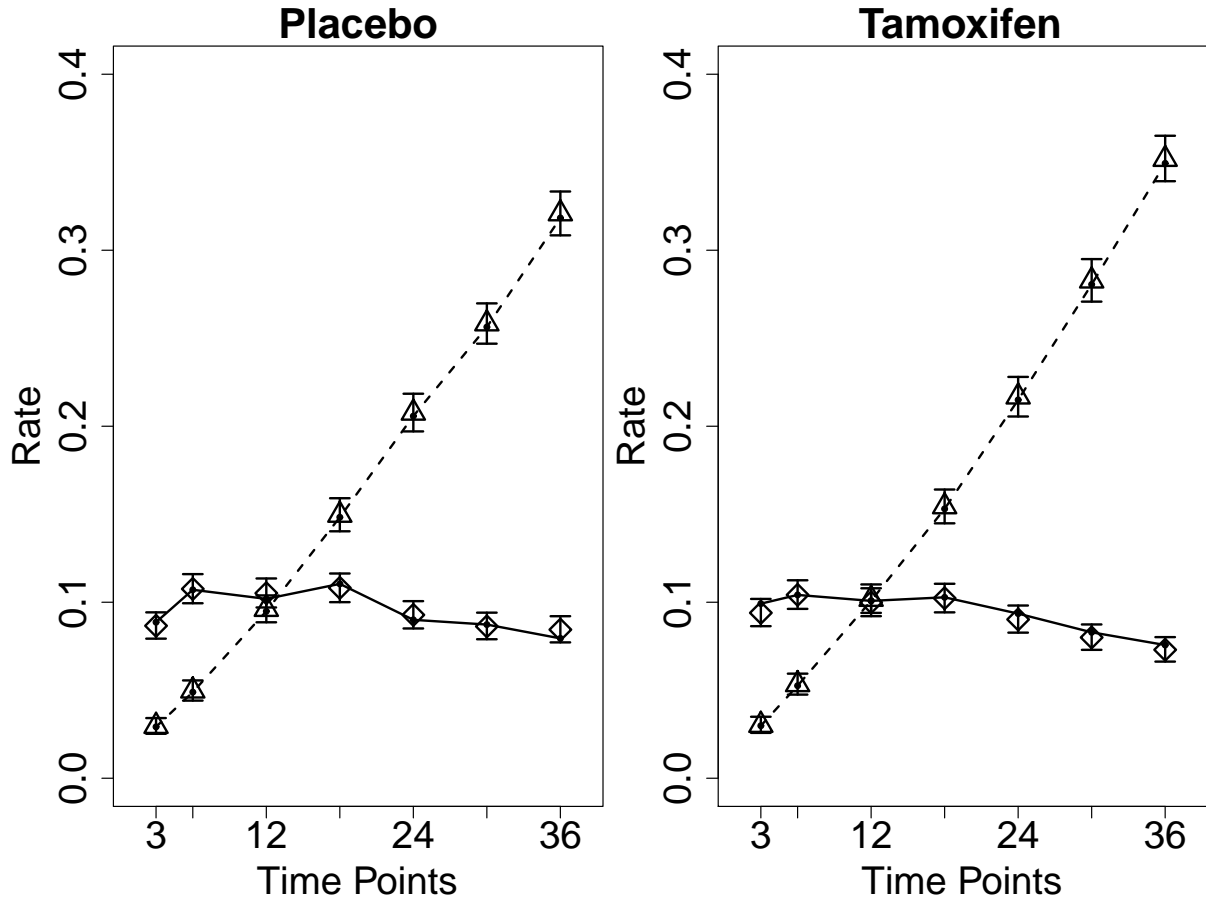


Figure 3-1. Solid and dashed lines represent the empirical rate of $P[Y_j = 1, S \geq j|Z = z]$ and $P[S < j|Z = z]$, respectively. The posterior means of $P[Y_j = 1, S \geq j|Z = z]$ (diamond) and $P[S < j|Z = z]$ (triangle) and their 95% credible intervals are displayed at each time point.

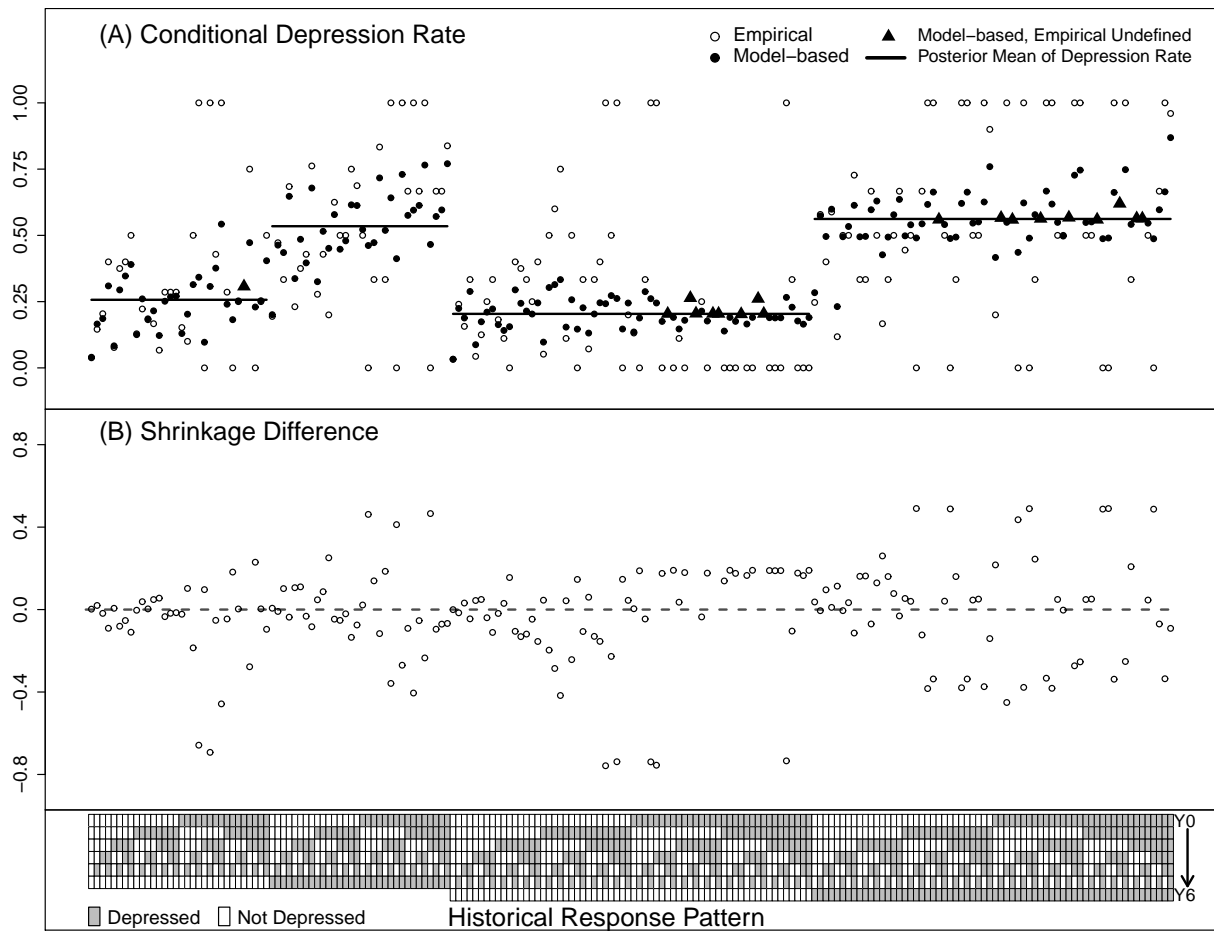


Figure 3-2. (A) The empirical rate and model-based posterior mean of $P[Y_j = 1 | S \geq j, \bar{Y}_{j-1} = \bar{y}_{j-1}, Z = z]$ for $Z = 2$ and $j = 6, 7$. (B) The difference between the empirical and model-based posterior mean of the depression rate. The x-axis is the pattern of historical response data \bar{Y}_{j-1} .

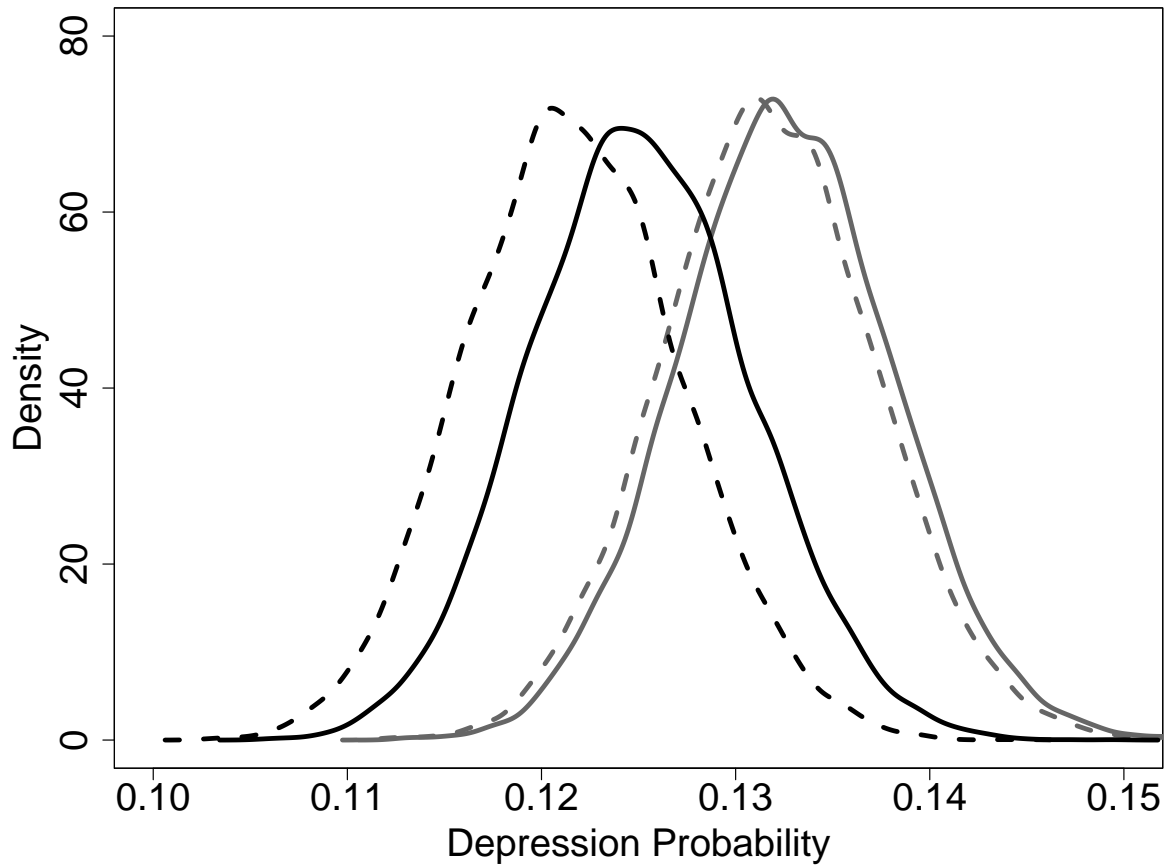


Figure 3-3. Posterior distribution of $P[Y_7 = 1|Z = z]$. Black and gray lines represent tamoxifen and placebo arms, respectively. Solid and dashed lines are for MNAR and MAR, respectively.

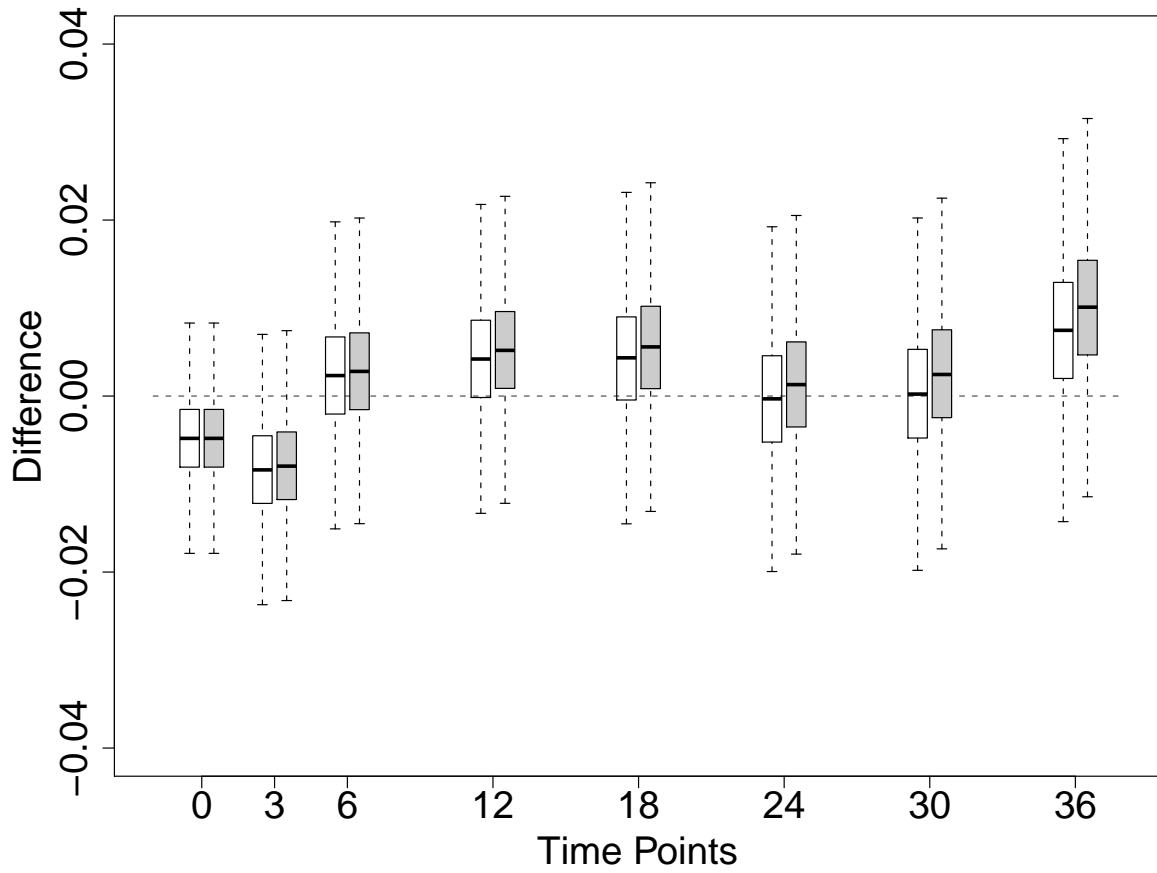


Figure 3-4. Posterior mean and 95% credible interval of difference of $P[Y_j = 1|Z = z]$ between placebo and tamoxifen arms. The gray and white boxes are for MAR and MNAR, respectively.

3.8 Appendix

Gibbs sampler for posterior computation: In the first step of the Gibbs sampler, we draw, for each subject with intermittent missing data, from the full conditional of $\mathbf{Y}_{\text{mis}}^{\text{I}}$ given $\alpha_z, \gamma_z, \mathbf{m}_z^{(\alpha)}, \eta_z^{(\alpha)}, \mathbf{m}_z^{(\gamma)}, \eta_z^{(\gamma)}, \mathbf{Y}_{\text{obs}}, S, \mathbf{R}_S$ and $Z = z$. The full conditional distribution can be expressed as

$$\begin{aligned} P[\mathbf{Y}_{\text{mis}}^{\text{I}} = \mathbf{y}_{\text{mis}}^{\text{I}} | \alpha_z, \gamma_z, \mathbf{m}_z^{(\alpha)}, \eta_z^{(\alpha)}, \mathbf{m}_z^{(\gamma)}, \eta_z^{(\gamma)}, \mathbf{Y}_{\text{obs}} = \mathbf{y}_{\text{obs}}, S = s, \mathbf{R}_S = \mathbf{r}_s, Z = z] \\ = \frac{P[\mathbf{Y}_{\text{mis}}^{\text{I}} = \mathbf{y}_{\text{mis}}^{\text{I}}, \mathbf{Y}_{\text{obs}} = \mathbf{y}_{\text{obs}}, S = s | \alpha_z, \gamma_z, \mathbf{m}_z^{(\alpha)}, \eta_z^{(\alpha)}, \mathbf{m}_z^{(\gamma)}, \eta_z^{(\gamma)}, Z = z]}{\sum_{\text{all } \mathbf{y}_{\text{mis}}^{\text{I}}} P[\mathbf{Y}_{\text{mis}}^{\text{I}} = \mathbf{y}_{\text{mis}}^{\text{I}}, \mathbf{Y}_{\text{obs}} = \mathbf{y}_{\text{obs}}, S = s | \alpha_z, \gamma_z, \mathbf{m}_z^{(\alpha)}, \eta_z^{(\alpha)}, \mathbf{m}_z^{(\gamma)}, \eta_z^{(\gamma)}, Z = z]} \end{aligned}$$

where the right hand side can be expressed as a function of $\mathbf{y}_{\text{mis}}^{\text{I}}, \mathbf{y}_{\text{obs}}, s$ and α_z and γ_z .

In the second step, we draw from the full conditional of $\mathbf{m}_z^{(\alpha)}$ given $\{\mathbf{Y}_{\text{mis}}^{\text{I}}\}, \alpha_z, \gamma_z, \eta_z^{(\alpha)}, \mathbf{m}_z^{(\gamma)}, \eta_z^{(\gamma)}, \{\mathbf{Y}_{\text{obs}}\}, \{S\}, \{\mathbf{R}_S\}$ and $\{Z\} = z$, where the notation $\{\mathcal{D}\}$ denotes data \mathcal{D} for all the individuals on the study. The full conditional can be expressed as

$$\prod_{j=2}^J \prod_{y=0}^1 f(m_{z,j,y}^{(\alpha)} | \{\mathbf{Y}_{\text{mis}}^{\text{I}}\}, \alpha_z, \eta_{z,j,y}^{(\alpha)}, \{\mathbf{Y}_{\text{obs}}\}, \{S\}, \{Z\} = z)$$

where

$$\begin{aligned} f(m_{z,j,y}^{(\alpha)} | \{\mathbf{Y}_{\text{mis}}^{\text{I}}\}, \alpha_z, \eta_{z,j,y}^{(\alpha)}, \{\mathbf{Y}_{\text{obs}}\}, \{S\}, \{Z\} = z) \\ \propto \prod_{\substack{S_i \geq j, Y_{i,j-1} = y \\ i: Z_i = z}} B(\alpha_{z,j}, \bar{\mathbf{Y}}_{i,j-2,y}; m_{z,j,y}^{(\alpha)} / \eta_{z,j,y}^{(\alpha)}, (1 - m_{z,j,y}^{(\alpha)}) / \eta_{z,j,y}^{(\alpha)}) \end{aligned}$$

and $B(\alpha; c, d)$ is a Beta density with parameters c and d .

In the third step, we draw from the full conditional of $\mathbf{m}_z^{(\gamma)}$ given $\{\mathbf{Y}_{\text{mis}}^{\text{I}}\}, \alpha_z, \gamma_z, \eta_z^{(\alpha)}, \mathbf{m}_z^{(\alpha)}, \eta_z^{(\gamma)}, \{\mathbf{Y}_{\text{obs}}\}, \{S\}, \{\mathbf{R}_S\}$ and $\{Z\} = z$. The full conditional can be expressed as

$$\prod_{j=2}^J \prod_{y=0}^1 f(m_{z,j-1,y}^{(\gamma)} | \{\mathbf{Y}_{\text{mis}}^{\text{I}}\}, \gamma_z, \eta_{z,j-1,y}^{(\gamma)}, \{\mathbf{Y}_{\text{obs}}\}, \{S\}, \{Z\} = z)$$

where

$$f(m_{z,j-1,y}^{(\gamma)} | \{\mathbf{Y}_{\text{mis}}^{\text{l}}\}, \gamma_z, \eta_{z,j-1,y}^{(\gamma)}, \{\mathbf{Y}_{\text{obs}}\}, \{S\}, \{Z\} = z) \\ \propto \prod_{\substack{s_i \geq j-1, Y_{i,j-1}=y \\ i:Z_i=z}} B(\gamma_{z,j-1}, \bar{\mathbf{Y}}_{i,j-2,y}; m_{z,j-1,y}^{(\gamma)} / \eta_{z,j-1,y}^{(\gamma)}, (1 - m_{z,j-1,y}^{(\gamma)}) / \eta_{z,j-1,y}^{(\gamma)}).$$

In the fourth step, we draw from the full conditional of $\eta_z^{(\alpha)}$ given $\mathbf{m}_z^{(\alpha)}$, $\{\mathbf{Y}_{\text{mis}}^{\text{l}}\}$, α_z , γ_z , $\mathbf{m}_z^{(\gamma)}$, $\eta_z^{(\gamma)}$, $\{\mathbf{Y}_{\text{obs}}\}$, $\{S\}$, $\{\mathbf{R}_S\}$ and $\{Z\} = z$. The full conditional can be expressed as

$$\prod_{j=2}^J \prod_{y=0}^1 f(\eta_{z,j,y}^{(\alpha)} | \{\mathbf{Y}_{\text{mis}}^{\text{l}}\}, \alpha_z, m_{z,j,y}^{(\alpha)}, \{\mathbf{Y}_{\text{obs}}\}, \{S\}, \{Z\} = z)$$

where

$$f(\eta_{z,j,y}^{(\alpha)} | \{\mathbf{Y}_{\text{mis}}^{\text{l}}\}, \alpha_z, m_{z,j,y}^{(\alpha)}, \{\mathbf{Y}_{\text{obs}}\}, \{S\}, \{Z\} = z) \\ \propto \frac{g(\mathbf{E}_{z,j,y}^{(\alpha)})}{(g(\mathbf{E}_{z,j,y}^{(\alpha)}) \eta_{z,j,y}^{(\alpha)} + 1)^2} \prod_{\substack{s_i \geq j, Y_{i,j-1}=y \\ i:Z_i=z}} B(\alpha_{z,j}, \bar{\mathbf{Y}}_{i,j-2,y}; m_{z,j,y}^{(\alpha)} / \eta_{z,j,y}^{(\alpha)}, (1 - m_{z,j,y}^{(\alpha)}) / \eta_{z,j,y}^{(\alpha)}).$$

In the fifth step, we draw from the full conditional of $\eta_z^{(\gamma)}$ given $\mathbf{m}_z^{(\alpha)}$, $\{\mathbf{Y}_{\text{mis}}^{\text{l}}\}$, α_z , γ_z , $\mathbf{m}_z^{(\gamma)}$, $\eta_z^{(\alpha)}$, $\{\mathbf{Y}_{\text{obs}}\}$, $\{S\}$, $\{\mathbf{R}_S\}$ and $\{Z\} = z$. The full conditional can be expressed as

$$\prod_{j=2}^J \prod_{y=0}^1 f(\eta_{z,j-1,y}^{(\gamma)} | \{\mathbf{Y}_{\text{mis}}^{\text{l}}\}, \gamma_z, m_{z,j-1,y}^{(\gamma)}, \{\mathbf{Y}_{\text{obs}}\}, \{S\}, \{Z\} = z)$$

where

$$f(\eta_{z,-1,j,y}^{(\gamma)} | \{\mathbf{Y}_{\text{mis}}^{\text{l}}\}, \gamma_z, m_{z,j-1,y}^{(\gamma)}, \{\mathbf{Y}_{\text{obs}}\}, \{S\}, \{Z\} = z) \propto \frac{g(\mathbf{E}_{z,j-1,y}^{(\gamma)})}{(g(\mathbf{E}_{z,j-1,y}^{(\gamma)}) \eta_{z,j-1,y}^{(\gamma)} + 1)^2} \\ \times \prod_{\substack{s_i \geq j-1, Y_{i,j-1}=y \\ i:Z_i=z}} B(\gamma_{z,j-1}, \bar{\mathbf{Y}}_{i,j-2,y}; m_{z,j-1,y}^{(\gamma)} / \eta_{z,j-1,y}^{(\gamma)}, (1 - m_{z,j-1,y}^{(\gamma)}) / \eta_{z,j-1,y}^{(\gamma)}).$$

To draw from the full conditionals for steps two to five, we use slice sampling (Neal, 2003).

In the sixth step, we draw from the full conditional of α_z given $\{\mathbf{Y}_{\text{mis}}^I\}$, γ_z , $\mathbf{m}_z^{(\alpha)}$, $\eta_z^{(\alpha)}$, $\mathbf{m}_z^{(\gamma)}$, $\eta_z^{(\gamma)}$, $\{\mathbf{Y}_{\text{obs}}\}$, $\{S\}$, $\{\mathbf{R}_s\}$ and $\{Z\} = z$. The full conditional can be expressed as

$$\prod_{j=2}^J \prod_{y=0}^1 \prod_{\text{all } \bar{\mathbf{y}}_{j-2}} f(\alpha_{z,j,\bar{\mathbf{y}}_{j-2,y}} | \{\mathbf{Y}_{\text{mis}}^I\}, m_{z,j,y}^{(\alpha)}, \eta_{z,j,y}^{(\alpha)}, \{\mathbf{Y}_{\text{obs}}\}, \{S\}, \{Z\} = z)$$

where

$$\begin{aligned} & f(\alpha_{z,j,\bar{\mathbf{y}}_{j-2,y}} | \{\mathbf{Y}_{\text{mis}}^I\}, m_{z,j,y}^{(\alpha)}, \eta_{z,j,y}^{(\alpha)}, \{\mathbf{Y}_{\text{obs}}\}, \{S\}, \{Z\} = z) \\ &= B(\alpha_{z,j,\bar{\mathbf{y}}_{j-2,y}}; m_{z,j,y}^{(\alpha)}/\eta_{z,j,y}^{(\alpha)} + o_{z,j,\bar{\mathbf{y}}_{j-2,y}}^{(\alpha)}, (1 - m_{z,j,y}^{(\alpha)})/\eta_{z,j,y}^{(\alpha)} + n_{z,j,\bar{\mathbf{y}}_{j-2,y}}^{(\alpha)} - o_{z,j,\bar{\mathbf{y}}_{j-2,y}}^{(\alpha)}), \end{aligned}$$

$n_{z,j,\bar{\mathbf{y}}_{j-2,y}}^{(\alpha)}$ is the number of subjects with $S \geq j$, $Y_{j-1} = y$, $\bar{\mathbf{Y}}_{j-2} = \bar{\mathbf{y}}_{j-2}$ and $Z = z$, and

$o_{z,j,\bar{\mathbf{y}}_{j-2,y}}^{(\alpha)}$ is the number of subjects with $S \geq j$, $Y_{j-1} = y$, $\bar{\mathbf{Y}}_{j-2} = \bar{\mathbf{y}}_{j-2}$, $Z = z$ and $Y_j = 1$.

Finally, we draw from the full conditional of γ_z given $\{\mathbf{Y}_{\text{mis}}^I\}$, α_z , $\mathbf{r}_z^{(\alpha)}$, $\eta_z^{(\alpha)}$, $\mathbf{r}_z^{(\gamma)}$, $\eta_z^{(\gamma)}$, $\{\mathbf{Y}_{\text{obs}}\}$, $\{S\}$, $\{\mathbf{R}_s\}$ and $\{Z\} = z$. The full conditional can be expressed as

$$\prod_{j=2}^J \prod_{y=0}^1 \prod_{\text{all } \bar{\mathbf{y}}_{j-2}} f(\gamma_{z,j-1,\bar{\mathbf{y}}_{j-2,y}} | \{\mathbf{Y}_{\text{mis}}^I\}, m_{z,j-1,y}^{(\gamma)}, \eta_{z,j-1,y}^{(\gamma)}, \{\mathbf{Y}_{\text{obs}}\}, \{S\}, \{Z\} = z)$$

where

$$\begin{aligned} & f(\gamma_{z,j-1,\bar{\mathbf{y}}_{j-2,y}} | \{\mathbf{Y}_{\text{mis}}^I\}, m_{z,j-1,y}^{(\gamma)}, \eta_{z,j-1,y}^{(\gamma)}, \{\mathbf{Y}_{\text{obs}}\}, \{S\}, \{Z\} = z) \\ &= B(\gamma_{z,j-1,\bar{\mathbf{y}}_{j-2,y}}; m_{z,j-1,y}^{(\gamma)}/\eta_{z,j-1,y}^{(\gamma)} + o_{z,j-1,\bar{\mathbf{y}}_{j-2,y}}^{(\gamma)}, (1 - m_{z,j-1,y}^{(\gamma)})/\eta_{z,j-1,y}^{(\gamma)} + n_{z,j-1,\bar{\mathbf{y}}_{j-2,y}}^{(\gamma)} - o_{z,j-1,\bar{\mathbf{y}}_{j-2,y}}^{(\gamma)}), \end{aligned}$$

$n_{z,j-1,\bar{\mathbf{y}}_{j-2,y}}^{(\gamma)}$ is the number of subjects with $S \geq j - 1$, $Y_{j-1} = y$, $\bar{\mathbf{Y}}_{j-2} = \bar{\mathbf{y}}_{j-2}$ and $Z = z$,

and $o_{z,j-1,\bar{\mathbf{y}}_{j-2,y}}^{(\gamma)}$ is the number of subjects with $S = j - 1$, $Y_{j-1} = y$, $\bar{\mathbf{Y}}_{j-2} = \bar{\mathbf{y}}_{j-2}$ and $Z = z$.

Proof of Theorem 2: We will show that, with no intermittent missingness and the shrinkage priors (Equation 3–2), the posterior distributions of $P[Y_j = 1 | S \geq j, Y_{j-1}, \bar{\mathbf{Y}}_{j-2}, Z]$ and $P[S = j - 1 | S \geq j - 1, Y_{j-1}, \bar{\mathbf{Y}}_{j-2}, Z]$, modeled as $\alpha_{z,\bar{\mathbf{y}}_{j-2,y}}$ and $\gamma_{z,\bar{\mathbf{y}}_{j-2,y}}$ (re: Equation 3–1), for all Z, j and $\bar{\mathbf{Y}}_{j-1}$ are consistent under the condition that all the true values of the probabilities are in the open interval $(0, 1)$.

We use $n_{z,j,\bar{\mathbf{y}}_{j-1}}^{(\alpha)}$ to denote the number of subjects with $S \geq j$, $Y_{j-1} = y_{j-1}$, $\bar{\mathbf{Y}}_{j-2} = \bar{\mathbf{y}}_{j-2}$, $Z = z$, and use $n_{z,j,\bar{\mathbf{y}}_{j-1}}^{(\gamma)}$ denote the number of subjects with $S \geq j - 1$, $Y_{j-1} = y_{j-1}$, $\bar{\mathbf{Y}}_{j-2} = \bar{\mathbf{y}}_{j-2}$, $Z = z$. The condition that all the true values of $\alpha_{z,\bar{\mathbf{y}}_{j-2},y}$ and $\gamma_{z,\bar{\mathbf{y}}_{j-2},y}$ are in the open interval $(0, 1)$ holds if and only if as the number of subjects goes to infinity, all the $n_{z,j,\bar{\mathbf{y}}_{j-1}}^{(\alpha)}$ and $n_{z,j,\bar{\mathbf{y}}_{j-1}}^{(\gamma)}$ go to infinity. This indicates that to prove Theorem 2, we can just prove that given $\mathbf{Y} = \{Y_1, \dots, Y_k\}$ and $\mathbf{N} = \{n_1, \dots, n_k\}$, where $Y_j \sim \text{Bin}(n_j, p_j)$, $p_j \sim \text{Beta}(\alpha, \beta)$ for $j = 1, \dots, k$ and (α, β) has proper prior density $\pi(\alpha, \beta)$, the posterior distributions for all p_j are consistent as all n_j go to infinity, with regard to the distributions under $Y_j \sim \text{Bin}(n_j, p_j^*)$.

To see this, note that

$$\pi(p_j | \mathbf{Y}, \mathbf{N}) \propto p_j^{Y_j} (1 - p_j)^{n_j - Y_j} \int_{\alpha, \beta} \pi(p_j | \alpha, \beta) \pi(\alpha, \beta) d\alpha d\beta.$$

Note that

$$\pi(p_j) = \int_{\alpha, \beta} \pi(p_j | \alpha, \beta) \pi(\alpha, \beta) d\alpha d\beta = \int_{\alpha, \beta} p_j^{\alpha-1} (1 - p_j)^{\beta-1} \pi(\alpha, \beta) d\alpha d\beta = M < \infty,$$

$\pi(p_j)$ is $O(1)$. As n_j and Y_j go to infinity (this occurs since $p_j \in (0, 1)$), by the Bernstein-von Mises theorem, we have

$$\pi(\sqrt{n_j}(p_j - p_j^*) | Y_j, n_j) \longrightarrow \text{N}(0, p_j^*(1 - p_j^*))$$

in distribution, which further implies that

$$\begin{aligned} E[p_j | \mathbf{Y}, \mathbf{N}] &\rightarrow p_j^* && \text{a.s.} \\ \text{Var}[p_j | \mathbf{Y}, \mathbf{N}] &\rightarrow 0 && \text{a.s.} \end{aligned}$$

CHAPTER 4
A NOTE ON MAR, IDENTIFYING RESTRICTIONS, AND SENSITIVITY ANALYSIS IN
PATTERN MIXTURE MODELS

4.1 Introduction

For analyzing longitudinal studies with informative missingness, popular modeling frameworks include pattern mixture models, selection models and shared parameter models, which differ in the way the joint distribution of the outcome and missing data process are factorized (for a comprehensive review, see [Daniels and Hogan, 2008](#); [Hogan and Laird, 1997b](#); [Kenward and Molenberghs, 1999](#); [Little, 1995](#); [Molenberghs and Kenward, 2007](#)). In this paper, we will concern ourselves with pattern mixture models with monotone missingness (i.e., drop-out). For pattern mixture models with non-monotone (i.e., intermittent) missingness (details go beyond the scope of this paper), one approach is to partition the missing data and allow one (or more) or the partitions to be ignored given the other partition(s) ([Harel and Schafer, 2009](#); [Wang et al., 2010](#)).

It is well known that pattern-mixture models are not identified: the observed data does not provide enough information to identify the distributions for incomplete patterns. The use of identifying restrictions that equate the inestimable parameters to functions of estimable parameters is an approach to resolve the problem ([Daniels and Hogan, 2008](#); [Kenward et al., 2003](#); [Little, 1995](#); [Little and Wang, 1996](#); [Thijs et al., 2002](#)). Common identifying restrictions include complete case missing value (CCMV) constraints and available case missing value (ACMV) constraints. [Molenberghs et al. \(1998\)](#) proved that for discrete time points and monotone missingness, the ACMV constraint is equivalent to missing at random (MAR), as defined by [Rubin \(1976\)](#) and [Little and Rubin \(1987\)](#). A key and attractive feature of identifying restrictions is that they do not affect the fit of the model to the observed data. Understanding (identifying) restrictions that lead to MAR is an important first step for sensitivity analysis under missing not at random (MNAR) ([Daniels and Hogan, 2008](#); [Scharfstein et al., 2003](#); [Zhang and Heitjan, 2006](#)).

In particular, MAR provides a good starting point for sensitivity analysis and sensitivity analysis are essential for the analysis of incomplete data ([Daniels and Hogan, 2008](#); [Scharfstein et al., 1999](#); [Vansteelandt et al., 2006b](#)).

The normality of response data (if appropriate) for pattern mixture models is desirable as it easily allows incorporation of baseline covariates and introduction of sensitivity parameters (for MNAR analysis) that have convenient interpretations as deviations of means and variances from MAR ([Daniels and Hogan, 2008](#)). However, multivariate normality within patterns can be overly restrictive. We explore such issues in this paper.

One criticism of mixture models is that they often induce missing data mechanisms that depend on the future ([Kenward et al., 2003](#)). We explore such non-future dependence in our context here and show how mixture models that have such missing data mechanisms have fewer sensitivity parameters.

In [Section 4.2](#), we show conditions under which MAR exists and does not exist when the full-data response is assumed multivariate normal within each missing pattern. In [Section 4.3](#) and [Section 4.4](#) in the same setting, we explore sensitivity analysis strategies under MNAR and under non-future dependent MNAR respectively. In [Section 4.5](#), we propose a sensitivity analysis approach where only the *observed* data within pattern are assumed multivariate normal. In [Section 4.6](#), we apply the frameworks described in previous sections to a randomized clinical trial for estimating the effectiveness of recombinant growth hormone for increasing muscle strength in the elderly. In [Section 4.7](#), we show that in the presence of baseline covariates with time-invariant coefficients, standard identifying restrictions cause over-identification of the baseline covariate effects and we propose a remedy. We provide conclusions in [Section 8](#).

4.2 Existence of MAR under Multivariate Normality within Pattern

Let Y be a J -dimensional longitudinal response vector with components scheduled to be measured at time points t_j ($j \in \{1, \dots, J\}$); this is the full data response. Without loss of generality, we assume Y_1 is always observed. Let $S = s$ denote the number of observed responses ($s = 1, 2, \dots, J$) corresponding to the follow up time t_s . Let \bar{Y}_j denote the historical response vector (Y_1, Y_2, \dots, Y_j) . Finally, we define $p_s(\cdot) = p(\cdot | S = s)$.

We show that MAR does not necessarily exist when it is assumed that

$$\mathbf{Y} | S = s \sim N(\boldsymbol{\mu}^{(s)}, \boldsymbol{\Sigma}^{(s)}) \quad (4-1)$$

for all s .

To see this, we introduce some further notation. Let

$$\boldsymbol{\mu}^{(s)}(j) = E(\bar{Y}_j | S = s) = \begin{bmatrix} \mu_1^{(s)}(j) \\ \mu_2^{(s)}(j) \end{bmatrix}$$

and

$$\boldsymbol{\Sigma}^{(s)}(j) = \text{Var}(\bar{Y}_j | S = s) = \begin{bmatrix} \boldsymbol{\Sigma}_{11}^{(s)}(j) & \boldsymbol{\Sigma}_{12}^{(s)}(j) \\ \boldsymbol{\Sigma}_{21}^{(s)}(j) & \boldsymbol{\Sigma}_{22}^{(s)}(j) \end{bmatrix}$$

where $\mu_1^{(s)}(j) = E(\bar{Y}_{j-1} | S = s)$, $\mu_2^{(s)}(j) = E(Y_j | S = s)$, $\boldsymbol{\Sigma}_{11}^{(s)}(j) = \text{Var}(\bar{Y}_{j-1} | S = s)$, $\boldsymbol{\Sigma}_{22}^{(s)}(j) = \text{Var}(Y_j | S = s)$, $\boldsymbol{\Sigma}_{12}^{(s)}(j) = \text{Cov}(\bar{Y}_{j-1}, Y_j | S = s)$ and $\boldsymbol{\Sigma}_{21}^{(s)}(j)$ is the transpose of $\boldsymbol{\Sigma}_{12}^{(s)}(j)$.

Lemma 4.1. *For monotone dropout, under the model given in (4-1), define*

$$\begin{aligned} \kappa_1^{(s)}(j) &= \boldsymbol{\Sigma}_{21}^{(s)}(j) \left(\boldsymbol{\Sigma}_{11}^{(s)}(j) \right)^{-1} \\ \kappa_2^{(s)}(j) &= \mu_2^{(s)}(j) - \kappa_1^{(s)}(j) \boldsymbol{\mu}_1^{(s)}(j) \\ \kappa_3^{(s)}(j) &= \boldsymbol{\Sigma}_{22}^{(s)}(j) - \boldsymbol{\Sigma}_{21}^{(s)}(j) \left(\boldsymbol{\Sigma}_{11}^{(s)}(j) \right)^{-1} \boldsymbol{\Sigma}_{12}^{(s)}(j). \end{aligned}$$

The condition that for a given j , the conditional distributions $p_s(y_j|\bar{y}_{j-1})$ are identical for all s is equivalent to $\kappa_1^{(s)}(j)$, $\kappa_2^{(s)}(j)$ and $\kappa_3^{(s)}(j)$ being constant in s .

Proof. The proof is trivial since

$$Y_j|\bar{Y}_{j-1} \sim N\left(\kappa_2^{(s)}(j) - \kappa_1^{(s)}(j)\bar{Y}_{j-1}, \kappa_3^{(s)}(j)\right).$$

□

In other words, if the condition in Lemma 4.1 is satisfied, then there exists a conditional distribution $p_{\geq s}(y_j|\bar{y}_{j-1})$ such that $p_s(y_j|\bar{y}_{j-1}) = p_{\geq s}(y_j|\bar{y}_{j-1})$ for all $s \geq j$. We now state a theorem that gives the restrictions on the model given in (4-1) for MAR to exist.

Theorem 4.2. For pattern mixture models with monotone dropout, under the model given in (4-1), identification via MAR constraints exists if and only if $\mu^{(s)}$ and $\Sigma^{(s)}$ satisfy Lemma 4.1 for $s \geq j$ and $1 < j < J$.

Proof. By Lemma 4.1, we only need to show that MAR constraints exist if and only if for all $1 < j < J$, the conditional distributions $p_s(y_j|\bar{y}_{j-1})$ are identical for $s \geq j$.

Molenberghs et al. (1998) proved that MAR holds if and only if

$$p_k(y_j|\bar{y}_{j-1}) = p_{\geq j}(y_j|\bar{y}_{j-1}) = \sum_{s=j}^J \frac{P(S=s)}{\sum_{s=j}^J P(S=s)} p_s(y_j|\bar{y}_{j-1}) \quad (4-2)$$

for all $j \geq 2$ and $k < j$. These conditionals are normal distributions since we assume $\mathbf{Y}|S$ is multivariate normal.

Suppose that there exists j such that $p_s(y_j|\bar{y}_{j-1})$ is not the same for all $s \geq j$. Then from (4-2), $p_{\geq j}(y_j|\bar{y}_{j-1})$ will be a mixture of normals whereas $p_k(y_j|\bar{y}_{j-1})$ will be a normal distribution. Thus, Molenbergh's condition will not be satisfied, i.e. the MAR constraints do not exist.

On the other hand, if for all $1 < j < J$, the conditional distributions $p_s(y_j|\bar{y}_{j-1})$ are identical for $s \geq j$, then $p_k(y_j|\bar{y}_{j-1})$ and $p_{\geq j}(y_j|\bar{y}_{j-1})$ are both normally distributed and the identification restrictions $p_k(y_j|\bar{y}_{j-1}) = p_{\geq j}(y_j|\bar{y}_{j-1})$ will result in MAR. □

So, a default approach for continuous Y , assuming the full data response is multivariate normal within pattern, does not allow an MAR restriction (unless the restrictions in Theorem 4.2 are imposed).

We now examine the corresponding missing data mechanism (MDM), $S|\mathbf{Y}$. We use “ \simeq ” to denote equality in distribution.

Corollary 4.3. *For pattern mixture models of the form (4–1) with monotone dropout, MAR holds if and only if $S|\mathbf{Y} \simeq S|Y_1$.*

Proof. Since Y_1 is always observed (by assumption), $S|\mathbf{Y} \simeq S|Y_1$ implies that $S|\mathbf{Y}_{\text{mis}}, \mathbf{Y}_{\text{obs}} \simeq S|\mathbf{Y}_{\text{obs}}$, where \mathbf{Y}_{mis} and \mathbf{Y}_{obs} denote the missing and observed data respectively. This shows that MAR holds.

On the other hand, MAR implies that

$$p(S = s|\mathbf{Y}) = p(S = s|\mathbf{Y}_{\text{obs}}) = p(S = s|\bar{Y}_s).$$

By Theorem 4.2, we have that MAR holds only if for all $1 < j < J$, the conditional distributions $p_s(y_j|\bar{y}_{j-1})$ are identical for $s \geq j$. Thus, under MAR

$$p_k(y_j|\bar{y}_{j-1}) = p_{\geq j}(y_j|\bar{y}_{j-1}) = p_s(y_j|\bar{y}_{j-1})$$

for all $j \geq 2$, $k < j$ and $s \geq j$. This implies that for all $j \geq 2$

$$p(y_j|\bar{y}_{j-1}) = \sum_{s=1}^J p_s(y_j|\bar{y}_{j-1})p(S = s) = p_s(y_j|\bar{y}_{j-1})$$

for all s .

Therefore,

$$\begin{aligned} p(S = s|\mathbf{Y}) &= p(S = s|\bar{y}_s) = \frac{p_s(\bar{y}_s)}{p(\bar{y}_s)} p(S = s) \\ &= \frac{p_s(y_s|\bar{y}_{s-1}) \cdots p_s(y_2|y_1)p_s(y_1)}{p(y_s|\bar{y}_{s-1}) \cdots p(y_2|y_1)p(y_1)} p(S = s) \\ &= \frac{p_s(y_1)}{p(y_1)} p(S = s) = p(S = s|y_1). \end{aligned}$$

□

Thus, the implicit MDM is very restrictive and does not depend on the entire history, \bar{Y}_s . We now show connections to missing completely at random (MCAR) and other common identifying restrictions.

Corollary 4.4. *For pattern mixture models of the form (4–1) with monotone dropout, MCAR is equivalent to MAR if $p_s(y_1) = p(y_1)$ for all s .*

Proof. First, MCAR implies MAR. Second, in the proof of Corollary 4.3, we showed that MAR holds if

$$p(S = s|\mathbf{Y}) = \frac{p_s(y_1)}{p(y_1)}p(S = s).$$

Thus under the assumption that $p_s(y_1) = p(y_1)$, MAR implies that $p(S = s|\mathbf{Y}) = p(S = s)$, i.e. MCAR. □

Corollary 4.5. *For pattern mixture models of the form (4–1) with monotone dropout, MAR constraints are identical to complete case missing value (CCMV) and nearest-neighbor constraints (NCMV).*

Proof. By Theorem 4.2, the MAR constraints imply

$$p_j(y_j|\bar{y}_{j-1}) = p_J(y_j|\bar{y}_{j-1}) = p_{\geq j}(y_j|\bar{y}_{j-1}).$$

Therefore for all $k < j$, the MAR constraints

$$p_k(y_j|\bar{y}_{j-1}) = p_{\geq j}(y_j|\bar{y}_{j-1})$$

are identical to CCMV restrictions

$$p_k(y_j|\bar{y}_{j-1}) = p_J(y_j|\bar{y}_{j-1})$$

and to NCMV restrictions

$$p_k(y_j|\bar{y}_{j-1}) = p_j(y_j|\bar{y}_{j-1}).$$

□

The results in this section were all based on specifying the mixture model in (4–1) and demonstrate that MAR only exists under the fairly strict conditions given in Theorem 1.

4.3 Sequential Model Specification and Sensitivity Analysis under MAR

Due to the structure of $\mu^{(s)}$ and $\Sigma^{(s)}$ under MAR constraints as outlined in Section 4.2, we propose to follow the approach in Daniels and Hogan (2008, Chapter 8) and specify distributions of observed Y within pattern as:

$$\begin{aligned} p_s(y_1) &\sim N(\mu_1^{(s)}, \sigma_1^{(s)}) & 1 \leq s \leq J \\ p_s(y_j | \bar{y}_{j-1}) &\sim N(\mu_{j|j^-}^{(\geq j)}, \sigma_{j|j^-}^{(\geq j)}) & 2 \leq j \leq s \leq J \end{aligned} \quad (4-3)$$

where $j^- = \{1, 2, \dots, j-1\}$. Note that by construction, we assume $p_s(y_j | \bar{y}_{j-1})$ are identical for all $j \leq s \leq J$. Consequently, we have $p_s(y_j | \bar{y}_{j-1}) = p(y_j | \bar{y}_{j-1}, S \geq s)$, denoted as $p_{\geq s}(y_j | \bar{y}_{j-1})$.

Corollary 4.6. *For pattern mixture models of the form (4–1) with monotone dropout, identification via MAR constraints exists if and only the observed data can be modeled as (4–3).*

Proof. Theorem 4.2 shows that identification via MAR constraints exists if and only if conditional distributions $p_s(y_j | \bar{y}_{j-1})$ are identical for $s \geq j$ and $j \geq 2$. That is, for observed data, we have

$$p_s(y_j | \bar{y}_{j-1}) \sim N(\mu_{j|j^-}^{(\geq j)}, \sigma_{j|j^-}^{(\geq j)}).$$

□

Corollary 4.6 implies that under the multivariate normality assumption in (4–1) and the MAR assumption, a sequential specification as in (4–3) always exists.

We provide some details for MAR in model (4–1) (which implies the specification in (4–3) as stated in Corollary 4.6) next. Distributions for missing data (which are not

identified) are specified as:

$$p_s(y_j | \bar{y}_{j-1}) \sim N(\mu_{j|j^-}^{(j)}, \sigma_{j|j^-}^{(j)}) \quad 1 \leq s < j \leq J.$$

The conditional mean structure of $\mu_{j|j^-}^{(\geq j)}$ and $\mu_{j|j^-}^{(j)}$ is parameterized as follows:

$$\begin{aligned} \mu_{j|j^-}^{(\geq j)} &= \beta_0^{(\geq j)} + \sum_{l=1}^{j-1} \beta_l^{(\geq j)} y_l \\ \mu_{j|j^-}^{(j)} &= \beta_0^{(j)} + \sum_{l=1}^{j-1} \beta_l^{(j)} y_l. \end{aligned}$$

To identify the full-data model, the MAR constraints require that

$$p_k(y_j | \bar{y}_{j-1}) = p_{\geq j}(y_j | \bar{y}_{j-1})$$

for $k < j$, which implies that

$$\mu_{j|j^-}^{(j)} = \mu_{j|j^-}^{(\geq j)} \quad \text{and} \quad \sigma_{j|j^-}^{(j)} = \sigma_{j|j^-}^{(\geq j)}$$

for $2 \leq j \leq J$. Since the equality of the conditional means need to hold for all \mathbf{Y} , this further implies that the MAR assumption requires that

$$\beta_l^{(j)} = \beta_l^{(\geq j)}$$

for $0 \leq l < j \leq J$.

The motivation of the proposed sequential model is to allow a straightforward extension of the MAR specification to a large class of MNAR models indexed by parameters measuring departures from MAR, as well as the attraction of doing sensitivity analysis on means and/or variances in normal models.

For example, we can let

$$\beta_l^{(j)} = \Delta_l^{(j)} + \beta_l^{(\geq j)} \quad \text{and} \quad \log \sigma_{j|j^-}^{(j)} = \Delta_\sigma^{(j)} + \log \sigma_{j|j^-}^{(\geq j)}$$

for all $j > 1$ and $0 \leq l < j$. Sensitivity analysis can be done on these Δ parameters that capture the information about the missing data mechanism. For example, in a Bayesian framework, we may assign informative priors elicited from experts to these sensitivity parameters Δ . Note in general we may have separate $\Delta_l^{(j)}$ and $\Delta_\sigma^{(j)}$ for each pattern s ($s \leq j$), but in practice it is necessary to limit the dimensionality of these (Daniels and Hogan, 2008). Indeed, we could make $\Delta_l^{(j)}$ and $\Delta_\sigma^{(j)}$ independent of j to further reduce the number of sensitivity parameters.

To see the impact of the Δ parameters on the MDM, we introduce notation $\Delta_{j|j-}^{(j)} = \Delta_0^{(j)} + \sum_{l=1}^{j-1} \Delta_l^{(j)} Y_l$ and then for $k < j$ we have

$$Y_j | \bar{Y}_{j-1}, S = k \sim N \left(\mu_{j|j-}^{(\geq j)} + \Delta_{j|j-}^{(j)}, e^{\Delta_\sigma^{(j)}} \sigma_{j|j-}^{(\geq j)} \right).$$

The conditional probability (hazard) of observing the first s observations given at least s observations is then given by:

$$\begin{aligned} \log \frac{P(S = s | \mathbf{Y})}{P(S \geq s | \mathbf{Y})} &= \log P(S = s) - \frac{\sigma_1^{(s)}}{2} + \frac{(Y_1 - \mu_1^{(s)})^2}{2\sigma_1^{(s)}} + \sum_{l=s+1}^J \left\{ -\frac{e^{\Delta_\sigma^{(l)}}}{2} + \right. \\ &\quad \left. \frac{(Y_l - \mu_{l|l-}^{(\geq l)} - \Delta_{l|l-}^{(l)})^2}{2e^{\Delta_\sigma^{(l)}} \sigma_{l|l-}^{(\geq l)}} \right\} - \log \sum_{k=s}^J \left\{ P(S = k) (\sigma_1^{(k)})^{-\frac{1}{2}} \exp \left\{ \frac{(Y_1 - \mu_1^{(k)})^2}{2\sigma_1^{(k)}} \right\} \right. \\ &\quad \left. \prod_{l=s+1}^k \exp \left\{ \frac{(Y_l - \mu_{l|l-}^{(\geq l)})^2}{2\sigma_{l|l-}^{(\geq l)}} \right\} \times \prod_{l=k+1}^J (e^{\Delta_\sigma^{(l)}})^{-\frac{1}{2}} \exp \left\{ \frac{(Y_l - \mu_{l|l-}^{(\geq l)} - \Delta_{l|l-}^{(l)})^2}{2e^{\Delta_\sigma^{(l)}} \sigma_{l|l-}^{(\geq l)}} \right\} \right\}. \end{aligned}$$

In general the MDM depends on \bar{Y}_J , i.e. MNAR. However, one might want hazard at time t_s to only depend on \bar{Y}_{s+1} , in which case we need to have different distributions and assumptions on $[Y_j | \bar{Y}_{j-1}, S = k]$ for $k < j - 1$ and $j > 2$, as shown in the next section.

4.4 Non-Future Dependence and Sensitivity Analysis under Multivariate Normality within Pattern

Non-future dependence assumes that missingness only depends on observed data and the current missing value, i.e.

$$[S = s | \mathbf{Y}] \simeq [S = s | \bar{Y}_{s+1}],$$

and can be viewed as a special case of MNAR and an extension of MAR (Kenward et al., 2003). Kenward et al. (2003) showed that non-future dependence holds if and only if for each $j \geq 3$ and $k < j - 1$,

$$p_k(y_j|\bar{y}_{j-1}) = p_{\geq j-1}(y_j|\bar{y}_{j-1}).$$

An approach to implement non-future dependence within the framework of Section 4.3 is as follows. We model the observed data as in (4-3). For the conditional distribution of the current missing data (Y_{s+1}), we assume that

$$p_s(y_{s+1}|\bar{y}_s) \sim N \left(\beta_0^{(\geq s+1)} + \Delta_0^{(s+1)} + \sum_{l=1}^s (\beta_l^{(\geq s+1)} + \Delta_l^{(s+1)})y_l, e^{\Delta_\sigma^{(s+1)}} \sigma_{s|s-}^{(\geq s+1)} \right) \quad 2 \leq s < J$$

and for the conditional distribution of the future missing data (Y_{s+2}, \dots, Y_J), we assume that

$$p_s(y_j|\bar{y}_{j-1}) = p_{\geq j-1}(y_j|\bar{y}_{j-1}) \quad 2 \leq s < j - 1 \leq J - 1,$$

where

$$p_{\geq j-1}(y_j|\bar{y}_{j-1}) = \frac{p(S = j - 1)}{p(S \geq j - 1)} p_{j-1}(y_j|\bar{y}_{j-1}) + \frac{p(S \geq j)}{p(S \geq j - 1)} p_{\geq j}(y_j|\bar{y}_{j-1}).$$

Note that by this approach, although the model for future missing data is a mixture of normals, the sensitivity parameters are kept the same as in Section 4.3 ($\Delta_l^{(j)}$ and $\Delta_\sigma^{(j)}$, $j = 2, \dots, J$ and $l = 0, \dots, j - 1$). In addition, this significantly reduces the number of potential sensitivity parameters. For J -dimensional longitudinal data, the total number of sensitivity parameters, $(2J^3 + 3J^2 + J)/6 - J$ is reduced to $(J^2 + 3J - 4)/2$; for $J=3$ (6), from 11 (85) to 7 (25). Further reduction is typically needed. See the data example in Section 4.6 as an illustration. If all the remaining sensitivity parameters are set to zero, then we have

$$p_s(y_{s+1}|\bar{y}_s) = p_{\geq s+1}(y_{s+1}|\bar{y}_s), \quad 2 \leq s < J$$

and

$$p_s(y_j|\bar{y}_{j-1}) = p_{\geq j}(y_j|\bar{y}_{j-1}), \quad 2 \leq s < j - 1 \leq J - 1,$$

which implies

$$p_s(y_j|\bar{y}_{j-1}) = p_{\geq j}(y_j|\bar{y}_{j-1})$$

for all $s < j$, i.e. MAR.

4.5 MAR and Sensitivity Analysis with Multivariate Normality on the Observed-Data Response

If we assume multivariate normality only on observed data response, $\mathbf{Y}_{\text{obs}}|S$ instead of the full data response, $\mathbf{Y}|S$, we can weaken the restrictions on $p_s(y_j|\bar{y}_{j-1})$ for $s \geq j$ and allow the MDM to incorporate all observed data under MAR (cf. Corollary 4.3).

For example, we may specify distributions $\mathbf{Y}_{\text{obs}}|S$ as follows:

$$\begin{aligned} p_s(y_1) &\sim N(\mu_1^{(s)}, \sigma_1^{(s)}) \quad 1 \leq s \leq J \\ p_s(y_j|\bar{y}_{j-1}) &\sim N(\mu_{j|j-}^{(s)}, \sigma_{j|j-}^{(s)}) \quad 2 \leq j \leq s \leq J \end{aligned}$$

where

$$\mu_{j|j-}^{(s)} = \beta_{j,0}^{(s)} + \sum_{l=1}^{j-1} \beta_{j,l}^{(s)} Y_l.$$

To identify the full-data model, recall the MAR constraints imply that

$$p_s(y_j|\bar{y}_{j-1}) = p_{\geq j}(y_j|\bar{y}_{j-1}) = \sum_{k=j}^J \frac{P(S=k)}{P(S \geq j)} p_k(y_j|\bar{y}_{j-1}) \quad (4-4)$$

for $s < j$, which are mixture of normals. For sensitivity analysis in this setting of mixture of normals, we propose to introduce sensitivity parameters Δ_μ (location) and Δ_σ (scale) such that for $s < j$

$$p_s(y_j|\bar{y}_{j-1}) = e^{-\Delta_\sigma^{(j)}} \sum_{k=j}^J \varpi_{j,k} p_k\left(\frac{y_j - \Delta_\mu^{(j)} - (1 - e^{\Delta_\sigma^{(j)}})\mu_{j|j-}^{(k)}}{e^{\Delta_\sigma^{(j)}}} \mid \bar{y}_{j-1}\right) \quad (4-5)$$

where $\varpi_{j,k} = \frac{P(S=k)}{P(S \geq j)}$. The rationale for this parameterization is that each $p_k(\cdot|\bar{y}_{j-1})$ in the summation will have mean $\Delta_\mu^{(j)} + \mu_{j|j-}^{(k)}$ and variance $e^{2\Delta_\sigma^{(j)}} \sigma_{j|j-}^{(k)}$. To reduce the dimension of the sensitivity parameters, we could make $\Delta_\mu^{(j)}$ and $\Delta_\sigma^{(j)}$ common for all j (namely Δ_μ and Δ_σ).

In this set up, we have

$$\mu_{j|j-}^{(s),\text{MNAR}} = \Delta_{\mu}^{(j)} + \sum_{k=j}^J \varpi_{j,k} \mu_{j|j-}^{(k)}$$

and

$$\sigma_{j|j-}^{(s),\text{MNAR}} = e^{2\Delta_{\sigma}^{(j)}} \left\{ \sum_{k=j}^J \varpi_{j,k} \left(\sigma_{j|j-1}^{(k)} + (\mu_{j|j-1}^{(k)})^2 \right) - \left(\sum_{k=j}^J \varpi_{j,k} \mu_{j|j-}^{(k)} \right)^2 \right\} + (1 - e^{2\Delta_{\sigma}^{(j)}}) \mathcal{M}$$

where

$$\mathcal{M} = \sum_{k=j}^J \varpi_{j,k} (\mu_{j|j-}^{(k)})^2 - \left(\sum_{k=j}^J \varpi_{j,k} \mu_{j|j-}^{(k)} \right)^2.$$

Note that \mathcal{M} does not depend on $\sigma_{j|j-1}^{(k)}$ for $k = j, \dots, J$.

Under an MAR assumption (4–4), for $[Y_j | \bar{Y}_{j-1}, S = s]$, we have

$$\mu_{j|j-}^{(s),\text{MAR}} = \sum_{k=j}^J \varpi_{j,k} \mu_{j|j-}^{(k)}$$

and

$$\sigma_{j|j-}^{(s),\text{MAR}} = \sum_{k=j}^J \varpi_{j,k} \left(\sigma_{j|j-1}^{(k)} + (\mu_{j|j-1}^{(k)})^2 \right) - (\mu_{j|j-}^{(s),\text{MAR}})^2.$$

Therefore, under MNAR assumption (4–5), the two sensitivity parameters control the departure of the mean and variance from MAR in the following way,

$$\mu_{j|j-}^{(s),\text{MNAR}} = \Delta_{\mu}^{(j)} + \mu_{j|j-}^{(s),\text{MAR}} \quad \text{and} \quad \sigma_{j|j-}^{(s),\text{MNAR}} = e^{2\Delta_{\sigma}^{(j)}} \sigma_{j|j-}^{(s),\text{MAR}} + (1 - e^{2\Delta_{\sigma}^{(j)}}) \mathcal{M},$$

with $\Delta_{\mu}^{(j)}$ being a location parameter and $\Delta_{\sigma}^{(j)}$ being a scale parameter. The MNAR class allows MAR when $\Delta_{\mu}^{(j)} = \Delta_{\sigma}^{(j)} = 0$ for all $j \geq 2$.

By assuming non-future dependence, we obtain

$$\begin{aligned} p_s(y_j | \bar{y}_{j-1}) &= p_{\geq j-1}(y_j | \bar{y}_{j-1}) = \frac{p(S = j-1)}{p(S \geq j-1)} e^{-\Delta_{\sigma}^{(j)}} \sum_{k=j}^J \varpi_{j,k} p_k \left(\frac{y_j - \Delta_{\mu}^{(j)} - (1 - e^{\Delta_{\sigma}^{(j)}}) \mu_{j|j-}^{(k)}}{e^{\Delta_{\sigma}^{(j)}}} | \bar{y}_{j-1} \right) \\ &\quad + \sum_{k=j}^J \frac{P(S = k)}{p(S \geq j-1)} p_k(y_j | \bar{y}_{j-1}) \quad 2 \leq s < j-1 \leq J-1, \end{aligned}$$

for the future data and (4–5) for the current data ($j = s + 1$). The number of sensitivity parameters in this setup is reduced from $J(J - 1)$ to $(J - 2)(J - 1)$; so, for $J = 3$ (6), from 6 (30) to 2 (20). Further reductions are illustrated in the next section.

4.6 Example: Growth Hormone Study

We analyze a longitudinal clinical trial using the framework from Sections 4.4 and 4.5 that assume multivariate normality for the full-data response within pattern (MVN) or multivariate normality for the observed data response within pattern (OMVN). We assume non-future dependence for the missing data mechanism to minimize the number of sensitivity parameters.

The growth hormone (GH) trial was a randomized clinical trial conducted to estimate the effectiveness of recombinant human growth hormone therapy for increasing muscle strength in the elderly. The trial had four treatment arms: placebo (P), growth hormone only (G), exercise plus placebo (EP), and exercise plus growth hormone (EG). Muscle strength, here mean quadriceps strength (QS), measured as the maximum foot-pounds of torque that can be exerted against resistance provided by a mechanical device, was measured at baseline, 6 months and 12 months. There were 161 participants enrolled on this study, but only (roughly) 75% of them completed the 12 month follow up. Researchers believed that dropout was related to the unobserved strength measures at the dropout times.

For illustration, we confine our attention to the two arms using exercise: exercise plus placebo (EP) and exercise plus growth hormone (EG). Table 4-1 contains the observed data.

Let (Y_1, Y_2, Y_3) denote the full-data response corresponding to baseline, 6 months, and 12 months. Let Z be the treatment indicator ($1 = EG, 0 = EP$). Our goal is to draw inference about the mean difference of QS between the two treatment arms at month

12. That is, the treatment effect

$$\theta = E(Y_3|Z = 1) - E(Y_3|Z = 0).$$

In the full-data model for each treatment under non-future dependence, there are seven sensitivity parameters for the MVN model: $\{\Delta_0^{(2)}, \Delta_1^{(2)}, \Delta_0^{(3)}, \Delta_1^{(3)}, \Delta_2^{(3)}, \Delta_\sigma^{(2)}, \Delta_\sigma^{(3)}\}$, and four sensitivity parameters for OMVN model: $\{\Delta_\mu^{(2)}, \Delta_\mu^{(3)}, \Delta_\sigma^{(2)}, \Delta_\sigma^{(3)}\}$ (see Appendix). For the MNAR analysis, we reduced the number of sensitivity parameters as follows:

- $\Delta_\sigma^{(2)}$ and $\Delta_\sigma^{(3)}$ do not appear in the posterior distribution of $E(Y_3|Z)$ for $Z = 0, 1$, and thus are not necessary for inference on θ .
- We restrict to MNAR departures from MAR in terms of the intercept terms by assuming $\Delta_1^{(2)} = \Delta_1^{(3)} = \Delta_2^{(3)} \equiv 0$.
- We assume the sensitivity parameters are identical between treatments.

This reduces the set of sensitivity parameters to $\{\Delta_0^{(2)}, \Delta_0^{(3)}\}$ for MVN model and $\{\Delta_\mu^{(2)}, \Delta_\mu^{(3)}\}$ for the OMVN model.

There are a variety of ways to specify priors for the sensitivity parameters $\Delta_0^{(2)}$ and $\Delta_0^{(3)}$,

$$\Delta_0^{(2)} = E(Y_2|Y_1, S = 1) - E(Y_2|Y_1, S \geq 2)$$

$$\Delta_0^{(3)} = E(Y_3|Y_2, Y_1, S = 2) - E(Y_3|Y_2, Y_1, S = 3).$$

Both represent the difference of conditional means between the observed and unobserved responses. $\Delta_\mu^{(2)}$ and $\Delta_\mu^{(3)}$ have (roughly) the same interpretation as $\Delta_0^{(2)}$ and $\Delta_0^{(3)}$ respectively.

Based on discussion with investigators, we made the assumption that dropouts do worse than completers; thus, we restrict the Δ 's to be less than zero. To do a fully Bayesian analysis to fairly characterize the uncertainty associated with the missing data mechanism, we assume a uniform prior for the Δ 's as a default choice. Subject matter considerations gave an upper bound of zero for the uniform distributions. We set

the lower bound using the variability of the observed data as follows. We estimate the residual variances of $Y_2|Y_1$ and $Y_3|Y_2, Y_1$ using the observed data; we denote these by $\tau_{2|1}$ and $\tau_{3|2,1}$ respectively. We use the square root of these estimates as the lower bounds. In particular, we specify the priors for $\{\Delta_0^{(2)}, \Delta_0^{(3)}\}$ as well as $\{\Delta_\mu^{(2)}, \Delta_\mu^{(3)}\}$ as $\text{Unif}(\mathcal{D}(\boldsymbol{\tau}))$, where

$$\mathcal{D}(\boldsymbol{\tau}) = \left[-\tau_{2|1}^{-1/2}, 0 \right] \times \left[-\tau_{3|2,1}^{-1/2}, 0 \right].$$

Based on the estimates $\tau_{2|1}^{-1/2} = 18$ and $\tau_{3|2,1}^{-1/2} = 12$, the priors are $[-18, 0] \times [-12, 0]$ for $\{\Delta_0^{(2)}, \Delta_0^{(3)}\}$ and for $\{\Delta_\mu^{(2)}, \Delta_\mu^{(3)}\}$. For the other parameters in the full-data model, we assign $N(0, 10^6)$ for mean parameters (μ, β) and $N(0, 10^4)$ for variance parameters (σ) ; see the Appendix for further details on the models.

We fit the model using WinBUGS, with multiple chains of 25,000 iterations and 4000 burn-in. Convergence was checked by examining trace plots of the multiple chains.

The results of the MVN model, OMVN model, and the observed data analysis are presented in Table 4-2. Under MNAR, the posterior mean (posterior standard deviation) of the difference in quadriceps strength at 12 months between the two treatment arms was 4.0 (8.9) and 4.4 (10) for the MVN and OMVN models. Under MAR the differences were 5.4 (8.8) and 5.8 (9.9) for the MVN and OMVN models, respectively. The smaller differences under MNAR were due to quadriceps strength at 12 months being lower under MNAR due to the assumption that dropouts do worse than completers. We conclude that the treatment difference, θ was not significantly different from zero.

4.7 ACMV Restrictions and Multivariate Normality with Baseline Covariates

In this section, we show that common identifying restrictions over-identify estimable parameters in the presence of baseline covariates with time invariant coefficients and offer a solution.

4.7.1 Bivariate Case

Consider the situation when $\mathbf{Y} = (Y_1, Y_2)$ is a bivariate normal response ($J = 2$) with missing data only in Y_2 , i.e. $S = 1$ or 2 . Assume there are baseline covariates X

with time invariant coefficients α . We model $p(S)$ and $p(\mathbf{Y}|S)$ as follows:

$$S|X \sim \text{Bern}(\phi(X))$$

$$\mathbf{Y}|S = s \sim N(\boldsymbol{\mu}^{(s)}(X), \boldsymbol{\Sigma}^{(s)}) \quad s = 1, 2$$

where

$$\boldsymbol{\mu}^{(s)} = \begin{bmatrix} \mu_1^{(s)} + X\alpha^{(s)} \\ \mu_2^{(s)} + X\alpha^{(s)} \end{bmatrix} \quad \text{and} \quad \boldsymbol{\Sigma}^{(s)} = \begin{bmatrix} \sigma_{11}^{(s)} & \sigma_{12}^{(s)} \\ \sigma_{21}^{(s)} & \sigma_{22}^{(s)} \end{bmatrix}.$$

MAR (ACMV) implies the following restriction

$$[Y_2|Y_1, S = 1] \simeq [Y_2|Y_1, S = 2].$$

This implies that conditional means, $E(Y_2|Y_1, X, S = s)$ for $s = 1, 2$, are equal, i.e.

$$\mu_2^{(1)} + X\alpha^{(1)} + \frac{\sigma_{21}^{(1)}}{\sigma_{11}^{(1)}}(Y_1 - \mu_1^{(1)} - X\alpha^{(1)}) = \mu_2^{(2)} + X\alpha^{(2)} + \frac{\sigma_{21}^{(2)}}{\sigma_{11}^{(2)}}(Y_1 - \mu_1^{(2)} - X\alpha^{(2)}). \quad (4-6)$$

For (4-6) to hold for all Y_1 and X , we need that

$$\alpha^{(1)} = \alpha^{(2)}.$$

However, both $\alpha^{(1)}$ and $\alpha^{(2)}$ are already identified by the observed data Y_1 . Thus the ACMV (MAR) restriction affects the model fit to the observed data. This is against the principle of applying identifying restrictions (Little and Wang, 1996; Wang and Daniels, 2009).

To resolve the over-identification issue, we propose to apply MAR constraints on residuals instead of directly on the responses. In the bivariate case, the corresponding restriction is

$$[Y_2 - X\alpha^{(1)}|Y_1 - X\alpha^{(1)}, X, S = 1] \simeq [Y_2 - X\alpha^{(2)}|Y_1 - X\alpha^{(2)}, X, S = 2]. \quad (4-7)$$

Since the conditional distributions

$$[Y_2 - X\alpha^{(s)} | Y_1 - X\alpha^{(s)}, X, S = s] \sim N \left(\mu_2^{(s)} + \frac{\sigma_{21}^{(s)}}{\sigma_{11}^{(s)}} (Y_1 - \mu_1^{(s)}), \sigma_{22}^{(s)} - \frac{(\sigma_{21}^{(s)})^2}{\sigma_{11}^{(s)}} \right)$$

are independent of $\alpha^{(s)}$ for $s = 1, 2$, the restriction (4–7) places no constraints on $\alpha^{(s)}$, thus avoiding over-identification.

The MDM corresponding to the ACMV(MAR) on the residuals is given by

$$\begin{aligned} \log \frac{P(S = 1 | \mathbf{Y}, X)}{P(S = 2 | \mathbf{Y}, X)} &= \log \frac{\phi(X)}{1 - \phi(X)} - \frac{1}{2\sigma^*} \left\{ (1 - B)^2 X(\alpha^{(2)}\alpha^{(2)\top} - \alpha^{(1)}\alpha^{(1)\top})X^\top \right. \\ &\quad \left. - 2(1 - B)(Y_2 - \Delta(Y_1))X(\alpha^{(2)} - \alpha^{(1)}) \right\} - \frac{1}{2} \log \frac{\sigma_{11}^{(2)}}{\sigma_{11}^{(1)}} \\ &\quad - \frac{(Y_1 - X\alpha^{(2)} - \mu_1^{(2)})^2}{2(\sigma_{11}^{(2)})^2} + \frac{(Y_1 - X\alpha^{(1)} - \mu_1^{(1)})^2}{2(\sigma_{11}^{(1)})^2}, \end{aligned}$$

where $\sigma^* = \sigma_{22}^{(2)} - \frac{(\sigma_{21}^{(2)})^2}{\sigma_{11}^{(2)}}$, $B = \frac{\sigma_{21}^{(1)}}{\sigma_{11}^{(1)}}$ and $\Delta(Y_1) = \mu_2^{(2)} + \frac{\sigma_{21}^{(2)}}{\sigma_{11}^{(2)}}(Y_1 - \mu_1^{(2)})$. Hence by assuming MAR on the residuals, we have the MDM being a quadratic form of Y_1 , but independent of Y_2 if and only if $\alpha^{(2)} = \alpha^{(1)}$. In other words, assumption (4–7) implies MAR if and only if $\alpha^{(2)} = \alpha^{(1)}$. So in general, MAR on residuals does not imply that missingness in Y_2 is MAR. However, it is an identifying restriction that does not affect the fit of the model to the observed data. CCMV and NCMV restrictions can be applied similarly to the residuals.

Remark: In general, $\mu_j^{(s)}$ can be replaced by $\mu_{ij}^{(s)}$ if there are subject-specific covariates with time varying coefficients.

4.7.2 Multivariate Case

To incorporate baseline covariates in the multivariate case and apply similar MAR restrictions, we specify the model for the observed data as follows:

$$\begin{aligned} p_s(y_1 | X) &\sim N(\mu_1^{(s)} + X\alpha^{(s)}, \sigma_1^{(s)}) \quad 1 \leq s \leq J \\ p_s(y_j | \bar{y}_{j-1}, X) &\sim N(\mu_{j|j-}^{(s)}, \sigma_{j|j-}^{(\geq j)}) \quad 2 \leq j \leq s \leq J, \end{aligned}$$

where

$$\mu_{j|j^-}^{(s)} = \mu_j^{(\geq j)} + X\alpha^{(s)} + \sum_{l=1}^{j-1} \beta_l^{(\geq j)} (Y_l - \mu_l^{(\geq j)} - X\alpha^{(s)}). \quad (4-8)$$

For the missing data, the conditional distributions are specified as

$$p_s(Y_j | \bar{Y}_{j-1}) \sim N(\mu_{j|j^-}^{(s)}, \sigma_{j|j^-}^{(j)}) \quad 1 \leq s < j \leq J$$

where

$$\mu_{j|j^-}^{(s)} = \mu_j^{(s)} + X\alpha^{(s)} + \sum_{l=1}^{j-1} \beta_{j,l}^{(s)} (Y_l - \mu_l^{(s)} - X\alpha^{(s)}). \quad (4-9)$$

The conditional mean structures in (4-8) and (4-9) induce the following form for the marginal mean response

$$E(Y_j | S = s) = U_j^{(s)} + X\alpha^{(s)},$$

where $U_j^{(s)}$ is a function of intercept (e.g. $\mu_j^{(\geq j)}$) and slope (e.g. $\beta_l^{(\geq j)}$) parameters from $\mu_{j|j^-}^{(s)}$, but not X or α . This marginal mean response reflects the fact that X is the baseline covariates and α is its time-invariant coefficient. This form is also necessary for resolving over-identification of α via the MAR on the residuals restrictions as shown later.

Note that since Y_1 is always observed, $\alpha^{(s)}$ ($1 \leq s \leq J$) are identified by the observed data. However, in the model given by (4-8) and (4-9), there is a two-fold over-identification of $\alpha^{(s)}$ under MAR:

1. For MAR constraints to exist under the model given in (4-1), $\mu_{j|j^-}^{(s)}$ as defined in (4-8) must be equal for $2 \leq j \leq s \leq J$ and for all X . This requires that $\alpha^{(s)} = \alpha^*$ for $2 \leq j \leq s \leq J$.
2. MAR constraints also imply that $\mu_{j|j^-}^{(s)}$ as defined in (4-9) must be equal to $\mu_{j|j^-}^{(\geq j)}$ for $1 \leq s < j$. This places another restriction on $\alpha^{(s)}$.

Similar over-identification exists under CCMV and NCMV.

Similar to the bivariate case, to avoid the over-identification, we again use the MAR on the residuals restriction,

$$p_k(y_j - X\alpha^{(k)} | y_1 - X\alpha^{(k)}, \dots, y_{j-1} - X\alpha^{(k)}, X) = \sum_{s=j}^J \frac{P(S=s)}{P(S \geq j)} p_s(y_j - X\alpha^{(s)} | y_1 - X\alpha^{(s)}, \dots, y_{j-1} - X\alpha^{(s)}, X) \quad k < j. \quad (4-10)$$

With the conditional mean structures specified as (4-8) and (4-9), the MAR on the residuals restriction places no assumptions on $\alpha^{(s)}$.

The corresponding MDM is

$$\begin{aligned} \log \frac{P(S=s | \mathbf{Y}, X)}{P(S \geq s | \mathbf{Y}, X)} &= \log \frac{P(S=s)p(\mathbf{Y} | S=s, X)}{P(\mathbf{Y}, S \geq s | X)} \\ &= \log \frac{P(S=s)p_s(Y_J | \bar{Y}_{J-1}, X)p_s(Y_{J-1} | \bar{Y}_{J-2}, X) \dots p_s(Y_1 | X)}{\sum_{l=s}^J p_l(Y_J | \bar{Y}_{J-1}, X)p_l(Y_{J-1} | \bar{Y}_{J-2}, X) \dots p_l(Y_1 | X)P(S=l)}. \end{aligned}$$

It does not have a simple form in general. However, if $\alpha^{(s)} = \alpha^*$ for all s , then

$$\log \frac{P(S=s | \mathbf{Y}, X)}{P(S \geq s | \mathbf{Y}, X)} = \log \frac{p_s(Y_1 | X)P(S=s)}{\sum_{l=s}^J p_l(Y_1 | X)P(S=l)},$$

i.e. the MDM only depends on Y_1 and X . Otherwise, the missingness is MNAR.

4.8 Summary

Most pattern mixture models allow the missingness to be MNAR, with MAR as a unique point in the parameter space. The magnitude of departure from MAR can be quantified via a set of sensitivity parameters. For MNAR analysis, it is critical to find scientifically meaningful and dimensionally tractable sensitivity parameters. For this purpose, (multivariate) normal distributions are often found attractive since the MNAR departure from MAR can be parsimoniously defined by deviations in the mean and (co-)variance.

However, a simple pattern mixture model based on multivariate normality for the full data response within patterns does not allow MAR without special restrictions that themselves, induce a very restrictive missing data mechanism. We explored this

fully and proposed alternatives based on multivariate normality for the observed data response within patterns. In both these contexts, we proposed strategies for specifying sensitivity parameters.

In addition, we showed that when introducing baseline covariates with time invariant coefficients, standard identifying restrictions result in over-identification of the model. This is against the principle of applying identifying restriction in that they should not affect the model fit to the observed data. We proposed a simple alternative set of restrictions based on residuals that can be used as an 'identification' starting point for an analysis using mixture models.

In the growth hormone study data example, we showed how to reduce the number of sensitivity parameters in practice and a default way to construct informative priors for sensitivity parameters based on limited knowledge about the missingness. In particular, all the values in the range, \mathcal{D} were weighted equally via a uniform distribution. If there is additional external information from expert opinion or historical data, informative priors may be used to incorporate such information (for example, see [Ibrahim and Chen, 2000](#); [Wang et al., 2010](#)). Finally, an important consideration in sensitivity analysis and constructing informative priors is that they should avoid extrapolating missing values outside of a reasonable range (e.g., negative quadriceps strength).

4.9 Tables

Table 4-1. Growth Hormone Study: Sample mean (standard deviation) stratified by dropout pattern.

Treatment	Dropout Pattern	Number of Participants	Month		
			0	6	12
EG	1	12	58(26)		
	2	4	57(15)	68(26)	
	3	22	78(24)	90(32)	88(32)
	All	38	69(25)	87(32)	88(32)
EP	1	7	65(32)		
	2	2	87(52)	86(51)	
	3	31	65(24)	81(25)	73(21)
	All	40	66(26)	82(26)	73(21)

Table 4-2. Growth Hormone Study: Posterior mean (standard deviation)

Treatment	Month	Observed Data	MAR Analysis		MNAR Analysis	
			MVN	OMVN	MVN	OMVN
EP	0	66(9.9)	66(6.0)	66(6.0)	66(6.0)	66(6.0)
	6	82(18)	82(5.9)	81(8.2)	80(6.0)	80(8.3)
	12	72(3.8)	73(4.9)	73(6.1)	72(5.0)	71(6.1)
EG	0	69(7.3)	69(4.9)	69(4.9)	69(4.9)	69(4.9)
	6	87(16)	81(6.8)	82(7.7)	78(7.1)	79(8.0)
	12	88(6.8)	78(7.2)	79(7.8)	76(7.5)	76(8.0)
Difference at 12 mos.		12(7.8)	5.4(8.8)	5.8(9.9)	4.0(8.9)	4.4(10)

4.10 Appendix

Missing data mechanism under missing not at random and multivariate

normality: The MDM in Section 4.3 is derived as follows:

$$\begin{aligned}
\log \frac{P(S = s | \mathbf{Y})}{P(S \geq s | \mathbf{Y})} &= \log \frac{P(S = s) p_s(\mathbf{Y})}{P(\mathbf{Y}, S \geq s)} \\
&= \log \frac{P(S = s) p_s(Y_1) \prod_{l=2}^J p_s(Y_l | \bar{Y}_{l-1})}{\sum_{k=s}^J \left\{ P(S = k) p_k(Y_1) \prod_{l=2}^J p_k(Y_l | \bar{Y}_{l-1}) \right\}} \\
&= \log \frac{\prod_{l=2}^s p_{\geq l}(Y_l | \bar{Y}_{l-1}) P(S = s) p_s(Y_1) \prod_{l=s+1}^J p_s(Y_l | \bar{Y}_{l-1})}{\prod_{l=2}^s p_{\geq l}(Y_l | \bar{Y}_{l-1}) \sum_{k=s}^J \left\{ P(S = k) p_k(Y_1) \prod_{l=s+1}^J p_k(Y_l | \bar{Y}_{l-1}) \right\}} \\
&= \log \frac{P(S = s) p_s(Y_1) \prod_{l=s+1}^J p_s(Y_l | \bar{Y}_{l-1})}{\sum_{k=s}^J \left\{ P(S = k) p_k(Y_1) \prod_{l=s+1}^k p_k(Y_l | \bar{Y}_{l-1}) \prod_{l=k+1}^J p_k(Y_l | \bar{Y}_{l-1}) \right\}} \\
&= \log P(S = s) - \frac{\sigma_1^{(s)}}{2} + \frac{(Y_1 - \mu_1^{(s)})^2}{2\sigma_1^{(s)}} + \sum_{l=s+1}^J \left\{ -\frac{e^{\Delta_{\sigma}^{(l)}}}{2} + \frac{(Y_l - \mu_{l|l-}^{(\geq l)} - \Delta_{l|l-}^{(l)})^2}{2e^{\Delta_{\sigma}^{(l)}} \sigma_{l|l-}^{(\geq l)}} \right\} \\
&\quad - \log \sum_{k=s}^J \left\{ P(S = k) (\sigma_1^{(k)})^{-\frac{1}{2}} \exp \left\{ \frac{(Y_1 - \mu_1^{(k)})^2}{2\sigma_1^{(k)}} \right\} \prod_{l=s+1}^k \exp \left\{ \frac{(Y_l - \mu_{l|l-}^{(\geq l)})^2}{2\sigma_{l|l-}^{(\geq l)}} \right\} \right. \\
&\quad \left. \times \prod_{l=k+1}^J (e^{\Delta_{\sigma}^{(l)}})^{-\frac{1}{2}} \exp \left\{ \frac{(Y_l - \mu_{l|l-}^{(\geq l)} - \Delta_{l|l-}^{(l)})^2}{2e^{\Delta_{\sigma}^{(l)}} \sigma_{l|l-}^{(\geq l)}} \right\} \right\}.
\end{aligned}$$

Mean and variance of $[Y_j | \bar{Y}_{j-1}, S = s]$: The mean and variance of $[Y_j | \bar{Y}_{j-1}, S = s]$ under MNAR assumption in Section 4.5 are derived as follows:

$$\begin{aligned}
\mu_{j|j-}^{(s), \text{MNAR}} &= E(Y_j | \bar{Y}_{j-1}, S = s) = e^{-\Delta_{\sigma}^{(j)}} \sum_{k=j}^J \varpi_{j,k} \int y_j p_k \left(\frac{y_j - \Delta_{\mu}^{(j)} - (1 - e^{\Delta_{\sigma}^{(j)}}) \mu_{j|j-}^{(k)}}{e^{\Delta_{\sigma}^{(j)}}} | \bar{y}_{j-1} \right) dy_j \\
&= e^{-\Delta_{\sigma}^{(j)}} \sum_{k=j}^J \varpi_{j,k} \int \left(e^{\Delta_{\sigma}^{(j)}} y_j^* + \Delta_{\mu}^{(j)} + (1 - e^{\Delta_{\sigma}^{(j)}}) \mu_{j|j-}^{(k)} \right) p_k(y_j^* | \bar{y}_{j-1}) e^{\Delta_{\sigma}^{(j)}} dy_j^* \\
&= \Delta_{\mu}^{(j)} + \sum_{k=j}^J \varpi_{j,k} \mu_{j|j-}^{(s)}
\end{aligned}$$

and

$$\begin{aligned}
\sigma_{j|j-}^{(s),\text{MNAR}} &= \text{Var}(Y_j | \bar{Y}_{j-1}, S = s) \\
&= e^{-\Delta_\sigma^{(j)}} \sum_{k=j}^J \varpi_{j,k} \int y_j^2 p_k \left(\frac{y_j - \Delta_\mu^{(j)} - (1 - e^{\Delta_\sigma^{(j)}}) \mu_{j|j-}^{(k)}}{e^{\Delta_\sigma^{(j)}}} \middle| \bar{y}_{j-1} \right) dy_j - \text{E}^2(Y_j | \bar{Y}_{j-1}, S = k) \\
&= e^{-\Delta_\sigma^{(j)}} \sum_{k=j}^J \varpi_{j,k} \int \left(e^{\Delta_\sigma^{(j)}} y_j^* + \Delta_\mu^{(j)} + (1 - e^{\Delta_\sigma^{(j)}}) \mu_{j|j-}^{(k)} \right)^2 p_k(y_j^* | \bar{y}_{j-1}) e^{\Delta_\sigma^{(j)}} dy_j^* \\
&\quad - \left(\Delta_\mu^{(j)} + \sum_{k=j}^J \varpi_{j,k} \mu_{j|j-}^{(k)} \right)^2 \\
&= e^{2\Delta_\sigma^{(j)}} \sum_{k=j}^J \varpi_{j,k} E((y_j^*)^2 | \bar{y}_{j-1}, S = k) + \sum_{k=j}^J \varpi_{j,k} (\Delta_\mu^{(j)} + (1 - e^{\Delta_\sigma^{(j)}}) \mu_{j|j-}^{(k)})^2 \\
&\quad + 2e^{\Delta_\sigma^{(j)}} \sum_{k=j}^J \varpi_{j,k} (\Delta_\mu^{(j)} + (1 - e^{\Delta_\sigma^{(j)}}) \mu_{j|j-}^{(k)}) E(y_j^* | \bar{y}_{j-1}, S = k) - \left(\Delta_\mu^{(j)} + \sum_{k=j}^J \varpi_{j,k} \mu_{j|j-}^{(k)} \right)^2 \\
&= e^{2\Delta_\sigma^{(j)}} \left\{ \sum_{k=j}^J \varpi_{j,k} (\sigma_{j|j-1}^{(k)} + (\mu_{j|j-1}^{(k)})^2) - \left(\sum_{k=j}^J \varpi_{j,k} \mu_{j|j-}^{(k)} \right)^2 \right\} \\
&\quad + (1 - e^{2\Delta_\sigma^{(j)}}) \left\{ \sum_{k=j}^J \varpi_{j,k} (\mu_{j|j-}^{(k)})^2 - \left(\sum_{k=j}^J \varpi_{j,k} \mu_{j|j-}^{(k)} \right)^2 \right\}
\end{aligned}$$

where $y_j^* = \frac{y_j - \Delta_\mu^{(j)} - (1 - e^{\Delta_\sigma^{(j)}}) \mu_{j|j-}^{(k)}}{e^{\Delta_\sigma^{(j)}}}$.

Full-data model for the growth hormone example (Section 4.6): We specify a pattern mixture model with sensitivity parameters for the two treatment arms. For compactness, we suppress subscript treatment indicator z from all the parameters in the following models.

For missing pattern S , we specify

$$S \sim \text{Mult}(\phi)$$

with the multinomial parameter $\phi = (\phi_1, \phi_2, \phi_3)$, $\phi_s = P(S = s)$ for $s \in \{1, 2, 3\}$, and

$$\sum_{s=1}^3 \phi_s = 1.$$

For the observed response data model $[\mathbf{Y}_{\text{obs}}|S]$, we specify the same MVN and OMVN model for $[Y_1|S]$ as follows:

$$\begin{aligned} Y_1|S = 1 &\sim N(\mu_1^{(1)}, \sigma_1^{(1)}) \\ Y_1|S = 2 &\sim N(\mu_1^{(2)}, \sigma_1^{(2)}) \\ Y_1|S = 3 &\sim N(\mu_1^{(3)}, \sigma_1^{(3)}). \end{aligned}$$

For MVN model, we specify

$$\begin{aligned} \left. \begin{aligned} Y_2|Y_1, S = 2 \\ Y_2|Y_1, S = 3 \end{aligned} \right\} &\sim N\left(\beta_0^{(\geq 2)} + \beta_1^{(\geq 2)} Y_1, \sigma_{2|2-}^{(\geq 2)}\right) \\ Y_3|Y_2, Y_1, S = 3 &\sim N\left(\beta_0^{(\geq 3)} + \beta_1^{(\geq 3)} Y_1 + \beta_2^{(\geq 3)} Y_2, \sigma_{3|3-}^{(\geq 3)}\right). \end{aligned}$$

For OMVN model, we specify

$$\begin{aligned} Y_2|Y_1, S = 2 &\sim N\left(\beta_{2,0}^{(2)} + \beta_{2,1}^{(2)} Y_1, \sigma_{2|2-}^{(2)}\right) \\ Y_2|Y_1, S = 3 &\sim N\left(\beta_{2,0}^{(3)} + \beta_{2,1}^{(3)} Y_1, \sigma_{2|2-}^{(3)}\right) \\ Y_3|Y_2, Y_1, S = 3 &\sim N\left(\beta_{3,0}^{(3)} + \beta_{3,1}^{(3)} Y_1 + \beta_{3,2}^{(3)} Y_2, \sigma_{3|3-}^{(3)}\right). \end{aligned}$$

For missing response data model $[\mathbf{Y}_{\text{mis}}|\mathbf{Y}_{\text{obs}}, S]$, we specify for MVN model

$$\begin{aligned} Y_2|Y_1, S = 1 &\sim N\left(\beta_0^{(\geq 2)} + \Delta_0^{(2)} + (\beta_1^{(\geq 2)} + \Delta_1^{(2)}) Y_1, e^{\Delta_\sigma^{(2)}} \sigma_{2|2-}^{(\geq 2)}\right) \\ Y_3|Y_2, Y_1, S = 2 &\sim N\left(\beta_0^{(\geq 3)} + \Delta_0^{(3)} + (\beta_1^{(\geq 3)} + \Delta_1^{(3)}) Y_1 + (\beta_2^{(\geq 3)} + \Delta_2^{(3)}) Y_2, e^{\Delta_\sigma^{(3)}} \sigma_{3|3-}^{(\geq 3)}\right) \\ Y_3|Y_2, Y_1, S = 1 &\sim \frac{\phi_3}{\phi_2 + \phi_3} N\left(\beta_0^{(\geq 3)} + \beta_1^{(\geq 3)} Y_1 + \beta_2^{(\geq 3)} Y_2, \sigma_{3|3-}^{(\geq 3)}\right) \\ &\quad + \frac{\phi_2}{\phi_2 + \phi_3} N\left(\beta_0^{(\geq 3)} + \Delta_0^{(3)} + (\beta_1^{(\geq 3)} + \Delta_1^{(3)}) Y_1 + (\beta_2^{(\geq 3)} + \Delta_2^{(3)}) Y_2, e^{\Delta_\sigma^{(3)}} \sigma_{3|3-}^{(\geq 3)}\right). \end{aligned}$$

For OMVN model, we specify

$$\begin{aligned}
Y_2|Y_1, S = 1 &\sim \frac{\phi_3}{\phi_2 + \phi_3} N\left(\Delta_u^{(2)} + \beta_{2,0}^{(3)} + \beta_{2,1}^{(3)} Y_1, e^{\Delta_\sigma^{(2)}} \sigma_{2|2-}^{(3)}\right) \\
&\quad + \frac{\phi_2}{\phi_2 + \phi_3} N\left(\Delta_u^{(2)} + \beta_{2,0}^{(2)} + \beta_{2,1}^{(2)} Y_1, e^{\Delta_\sigma^{(2)}} \sigma_{2|2-}^{(2)}\right) \\
Y_3|Y_2, Y_1, S = 2 &\sim N\left(\Delta_u^{(3)} + \beta_{3,0}^{(3)} + \beta_{3,1}^{(3)} Y_1 + \beta_{3,2}^{(3)} Y_2, e^{\Delta_\sigma^{(3)}} \sigma_{3|3-}^{(3)}\right) \\
Y_3|Y_2, Y_1, S = 1 &\sim \frac{\phi_3}{\phi_2 + \phi_3} N\left(\beta_{3,0}^{(3)} + \beta_{3,1}^{(3)} Y_1 + \beta_{3,2}^{(3)} Y_2, \sigma_{3|3-}^{(3)}\right) \\
&\quad + \frac{\phi_2}{\phi_2 + \phi_3} N\left(\Delta_u^{(3)} + \beta_{3,0}^{(3)} + \beta_{3,1}^{(3)} Y_1 + \beta_{3,2}^{(3)} Y_2, e^{\Delta_\sigma^{(3)}} \sigma_{3|3-}^{(3)}\right).
\end{aligned}$$

MAR on residuals constraints: Here we show that in multivariate case (Section 4.7.2), the MAR on the residuals restriction puts no constraints on $\alpha^{(s)}$.

Let $[Z_j|S] \simeq [Y_j - X\alpha^{(s)}]$. The MAR on the residuals constraints are

$$p_k(z_j|\bar{z}_{j-1}, X) = \sum_{s=j}^J \frac{P(S=s)}{P(S \geq j)} p_s(z_j|\bar{z}_{j-1}, X).$$

Note that

$$\begin{aligned}
p_s(y_j, \dots, y_1) &= p_s(y_1) \prod_{l=2}^j p_s(y_l|\bar{y}_{l-1}) \\
&= p_s(y_1) \prod_{l=2}^j \frac{\exp\left\{\frac{1}{2\sigma_{l|l-}^{(s)}} \left(y_l - \mu_l^{(\geq l)} - X\alpha^{(s)} - \sum_{t=1}^{j-1} \beta_t^{(\geq j)} (y_t - \mu_t^{(\geq l)} - X\alpha^{(s)})\right)^2\right\}}{\sqrt{2\pi\sigma_{l|l-}^{(s)}}}.
\end{aligned}$$

Thus,

$$\begin{aligned}
p_s(z_j, \dots, z_1) &= p_s(z_1) \prod_{l=2}^j p_s(z_l|\bar{z}_{l-1}) \\
&= p_s(z_1) \prod_{l=2}^j \frac{\exp\left\{\frac{1}{2\sigma_{l|l-}^{(s)}} \left(z_l - \mu_l^{(\geq l)} - \sum_{t=1}^{j-1} \beta_t^{(\geq j)} (z_t - \mu_t^{(\geq l)})\right)^2\right\}}{\sqrt{2\pi\sigma_{l|l-}^{(s)}}}.
\end{aligned}$$

We can further show that

$$[Z_j|\bar{Z}_{j-1}, S = s, X] \sim N\left(\mu_j^{(\geq j)} + \sum_{l=1}^{j-1} \beta_{j,l}^{(\geq j)} (Z_l - \mu_l^{(\geq j)}), \sigma_{j|j-}^{(\geq j)}\right),$$

which is independent of s . Therefore,

$$\sum_{s=j}^J \frac{P(S = s)}{P(S \geq j)} p_s(z_j | \bar{z}_{j-1}, X) = N \left(\mu_j^{(\geq j)} + \sum_{l=1}^{j-1} \beta_{j,l}^{(\geq j)} (Z_l - \mu_l^{(\geq j)}), \sigma_{j|j^-}^{(\geq j)} \right).$$

Similarly, we may derive that

$$p_k(z_j | \bar{z}_{j-1}, X) = N \left(\mu_j^{(s)} + \sum_{l=1}^{j-1} \beta_{j,l}^{(s)} (Z_l - \mu_l^{(s)}), \sigma_{j|j^-}^{(j)} \right).$$

The constraints (4–10) thus imply

$$\begin{aligned} \beta_{j,l}^{(s)} &= \beta_{j,l}^{(\geq j)} \\ \mu_j^{(s)} &= \mu_j^{(\geq j)} + \sum_{l=1}^{j-1} \beta_{j,l}^{(\geq j)} (\mu_l^{(s)} - \mu_l^{(\geq j)}) \\ \sigma_{j|j^-}^{(j)} &= \sigma_{j|j^-}^{(\geq j)}, \end{aligned}$$

which places no restrictions on $\alpha^{(s)}$.

CHAPTER 5
DISCUSSION: FUTURE APPLICATION OF THE BAYESIAN NONPARAMETRIC AND
SEMI-PARAMETRIC METHODS

5.1 Summary

In this dissertation, we explored the utility of Bayesian shrinkage methods for the analysis of incomplete longitudinal data with informative missingness that includes both drop-out and intermittent missingness. We considered two different saturated model parameterizations and corresponding parameter space reduction strategies. By simulation, we showed that the proposed methods outperform a saturated model without parameter shrinkage and a misspecified parametric model while being very competitive with the correct parametric model. Furthermore, the incomplete data analysis framework developed in this dissertation allows straightforward incorporation of experts' opinions in the form of informative priors, as well as flexible sensitivity analysis. Second, we explored conditions necessary for identifying restrictions that result in missing at random (MAR) to exist under a multivariate normality assumption and strategies for identifying sensitivity parameters for sensitivity analysis or for a fully Bayesian analysis with informative priors, with application to a longitudinal clinical trial. In the following sections, we will discuss further applications of Bayesian nonparametric and semi-parametric methods we are currently considering.

5.2 Extensions to Genetics Mapping

For identifying genes involved in human diseases and their inheritance, multi-generational family-based designs (including parents and offspring) have become popular ([Farnir et al., 2002](#); [Liu et al., 2006](#)). In such analysis, researchers are often interested in simultaneously estimating linkage that describes the tendency of certain alleles to be inherited together, and linkage disequilibrium (LD) that measures the non-random association between different markers. However, if the LD is close to zero, the linkage recombination fraction is hard to estimate (or not estimable at all). We use a two marker scenario to illustrate this dilemma.

Consider two markers A and B , each with two alleles A and a and B and b , respectively. Four possible haplotypes, $[AB]$, $[Ab]$, $[aB]$, and $[ab]$, can be formed by the two markers, with the frequencies denoted as p_{11} , p_{10} , p_{01} , and p_{00} , respectively. We use p , $1 - p$, q and $1 - q$ to denote the allele frequencies of A , a , B and b , respectively. Then, we have the following relationships:

$$\begin{aligned}
 p_{11} &= pq + D \\
 p_{10} &= p(1 - q) - D \\
 p_{01} &= (1 - p)q - D \\
 p_{00} &= (1 - p)(1 - q) + D,
 \end{aligned}
 \tag{5-1}$$

where D is the LD parameter. By simple algebra, we can show that

$$D = p_{11}p_{00} - p_{10}p_{01}.$$

Now, let r denote the linkage recombination fraction, the frequency that a chromosomal crossover will take place between the two markers A and B during meiosis. To estimate r , only offspring with at least a double-heterozygous parent, i.e. parent with genotype Aa/Bb , contribute to the likelihood (proportionally) by

$$p_{11}p_{00}r + p_{10}p_{01}(1 - r).$$

Therefore, r is not estimable when D is zero.

One possible solution is to incorporate more markers in the linkage analysis. By doing this, the number of parents with no less than two heterozygous markers increases. Consequently, more offspring contribute to the likelihood for estimating the linkage recombination fraction. However, the number of haplotype frequencies to be estimated also increases (exponentially) as the number of markers increases. Bayesian shrinkage methods can be applied to address this problem.

5.3 Extensions to Causal Inference

5.3.1 Causal Inference Introduction

For clinical studies of terminal diseases, incomplete data is often caused by death. If a response is truncated by death, the missingness should not be classified as "censored" because "censoring" implies the masking of a value that is potentially observable. It is also not appropriate to handle these cases by traditional non-mortality missing data approaches such as models assuming ignorability or models assuming non-ignorable missing data mechanism which implicitly "impute" missing responses. In randomized studies with no missingness, causal relationships are well established and the treatment causal effects can be estimated directly (Rubin, 1974). However, in non-randomized trials or in presence of missing data, these methods are limited if the research interest demands estimation of causally interpretable effects.

To define causal effects, we first introduce the concept of potential outcomes, which are sometimes used exchangeably with the term counterfactual (but not always, see Rubin, 2000). The use of term "potential outcome" can be traced at least to Neyman (1923). Neyman used "potential yields" U_{ik} to indicate the yield of a plot k if exposed to a variety i . Rubin (1974) defines the causal effect of one treatment, E , over another, C , for a particular unit as the difference between what would have happened if the unit had been exposed to E , namely $Y(E)$, and what would have happened if the unit had been exposed to C , namely $Y(C)$.

Using potential outcomes, Frangakis and Rubin (2002) introduce a framework for comparing treatment effects based on principle stratification, which is a cross-classification of units defined by their potential outcomes with respect to (post)treatment variables, such as treatment noncompliance or drop-out. The treatment comparison adjustment for posttreatment variables is necessary because such variables encode the characteristics of both the treatment and the patient. For example, a patient with diagnosed cancer in a cancer prevention trial may have depression caused by the treatment or by the diagnosis

itself. Furthermore, the comparison may be meaningless without the posttreatment variable adjustment. For example, a response such as depression is no longer defined after a non-response cause such as death happens.

A stratum \mathcal{A} is defined by the joint potential response $S(Z)$ with respect to the posttreatment variable Z (e.g., $Z = 0, 1$). For example, let $S(Z)$ be the potential survival status and let $S(Z) = 1$ and 0 denote alive and dead respectively. Then the stratum $\mathcal{A} = \{S(0) = 1, S(1) = 1\}$ defines the patients who will (potentially) survive on both arms.

A stratum is unaffected by treatment. That is, for subject i , $i \in \mathcal{A}$ or $i \notin \mathcal{A}$ does not depend on the actual treatment i is assigned. Consequently, the treatment effect defined as the difference between

$$\{Y_i(0)|i \in \mathcal{A}\} \quad \text{and} \quad \{Y_i(1)|i \in \mathcal{A}\}$$

is a causal effect.

On the contrary, a standard adjustment for posttreatment variables uses the treatment comparison between

$$\{Y_i(0)|S_i(0) = s\} \quad \text{and} \quad \{Y_i(1)|S_i(1) = s\}.$$

Such an estimand is not a causal effect when $S(z)$ is affected by z , which results in the fact that the group of patients with $S(0) = s$ is not identical to the group of patients with $S(1) = s$.

Consistent with Frangakis and Rubin's framework, [Rubin \(2000\)](#) introduced the concept of survivors average causal effect (SACE), that is the causal effects of treatment on endpoints that are defined only for survivors, i.e. the group of patients who would live regardless of their treatment assignment.

Within the principal strata framework, the identification of SACE or other principal stratum causal effects usually depends on untestable assumptions. To address the

uncertainty of the untestable assumptions, sensitivity analysis is carried out, and/or bounds of the causal effects are derived. For example, [Zhang and Rubin \(2003\)](#) derived large sample bounds for causal effects without assumptions and with assumptions such as monotonicity on death rate on different treatment arms. [Gilbert et al. \(2003\)](#) used a class of logistic selection bias models to identify the causal estimands and carried out sensitivity analysis for the magnitude of selection bias. [Hayden et al. \(2005\)](#) assumed "explainable nonrandom noncompliance" ([Robins, 1998](#)) and outlined a sensitivity analysis for exploring the robustness of the assumption. [Cheng and Small \(2006\)](#) derived sharp bounds for the causal effects and constructed confidence intervals to cover the identification region. [Egleston et al. \(2007\)](#) proposed a similar method to [Zhang and Rubin \(2003\)](#), but instead of identifying the full joint distribution of potential outcomes, they only identify features of the joint distribution that are necessary for identifying the SACE estimand. [Lee et al. \(2010\)](#) replaced the common deterministic monotonicity assumption by a stochastic one that allows incorporation of subject specific effects and generalized the assumptions to more complex trials.

5.3.2 Data and Notation

The following notation is defined for a random individual. When necessary, we use the subscript i to denote data for the i th individual.

We consider a controlled randomized clinical study with treatment arm ($Z = 1$) and control arm ($Z = 0$). A longitudinal binary outcome \mathbf{Y} is scheduled to be measured at visits $j = 1, \dots, J$, i.e. $\mathbf{Y} = (Y_1, \dots, Y_J)$ is a J -dimensional vector. Let $\mathbf{R} = (R_1, \dots, R_J)$ be the missing indicator vector with $R_j = 1$ if Y_j is observed and $R_j = 0$ if Y_j is missing. We assume the missingness is monotone.

We assume there are multiple events that will cause drop out for a patient on this trial, and categorize the events as non-response events (e.g. death) and missing events (e.g. withdraw of consent). We assume that non-response events may happen after the

occurrence of a missing event but not vice versa. We further assume all the events are observed.

Let C denote the “survival” time for a patient. That is, $C = c$ implies that a non-response event happened to the patient between visit c and $c + 1$ and caused the patient to drop out on and after visit $c + 1$. Let $\mathbf{R}_c = \{R_1, \dots, R_C\}$ be the missing data indicator recorded prior to patient drop-out that is caused by a non-response event.

We use $\bar{\mathbf{Y}}_j = (Y_1, \dots, Y_j)$ to denote the historical data up to time point j and \mathbf{Y}_{obs} to denote the observed response data. We use $\mathbf{Y}(z)$, $C(z)$, $\mathbf{R}_c(z)$ and $\mathbf{Y}_{\text{obs}}(z)$ to denote the value of \mathbf{Y} , C , \mathbf{R}_c and \mathbf{Y}_{obs} of a patient, possibly counterfactual, if the patient is assigned to treatment z .

The full data \mathcal{F} of a patient thus consists of

$$\{Z, C(0), \bar{\mathbf{Y}}_{C(0)}(0), C(1), \bar{\mathbf{Y}}_{C(1)}(1)\},$$

and the observed data \mathcal{O} contains

$$\{Z, C(Z), \mathbf{R}_c(Z), \mathbf{Y}_{\text{obs}}(Z)\}.$$

One goal is to measure the causal effect of treatment by estimating the treatment effect for those who would not have dropped out due to non-response reasons under either treatment or control. That is, to estimate the “survivor” average causal effect

$$\begin{aligned} SACE_j &= E(Y_j(1) - Y_j(0) | C(0) \geq j, C(1) \geq j) \\ &= P(Y_j(1) = 1 | C(0) \geq j, C(1) \geq j) - P(Y_j(0) = 1 | C(0) \geq j, C(1) \geq j) \quad (5-2) \end{aligned}$$

for all j . Note that the group of patients of interest $\{C(0) \geq j, C(1) \geq j\}$ form a principal stratum.

5.3.3 Missing Data Mechanism

To make causal inferences, we first need to estimate $\mu_{j,z,c}^* = E[Y_j | C = c, Z = z]$ for all z and $j \leq c$, which are not identifiable without unverifiable assumptions. We make the

same partial missing at random assumption as in Chapter 3, Section 3.2, that

$$\mathbf{R}_c \perp \bar{\mathbf{Y}}_C \mid Z, C, \mathbf{Y}_{\text{obs.}}$$

We have shown in Chapter 3 that $\mu_{j,z,c}^*$ is identified by the observed data under this partial missing at random assumption.

5.3.4 Causal Inference Assumption

The causal effect (5-2) is not identifiable from the observed data

$$\mathcal{O} = \{Z, C(Z), \mathbf{R}_j(Z), \mathbf{Y}_{\text{obs.}}(Z)\}.$$

We propose the following assumptions to identify boundaries for the causal effect:

I *Stable Unit Treatment Value Assumption (SUTVA).*

Let $\mathbf{Z} = (Z_1, \dots, Z_N)$ be the vector of treatment assignment for all the patients. SUTVA means

$$Z_i = Z'_i \Rightarrow (\mathbf{Y}_i(Z_i), \mathbf{C}_i(Z_i)) = (\mathbf{Y}_i(Z'_i), \mathbf{C}_i(Z'_i)),$$

regardless of what \mathbf{Z} is. That is, the potential outcome of patient i is unrelated to the treatment assignment of other patients. This allows us to write $\mathbf{Y}_i(\mathbf{Z})$ and $\mathbf{C}_i(\mathbf{Z})$ as $\mathbf{Y}_i(Z_i)$ and $\mathbf{C}_i(Z_i)$ respectively.

II *Random Assignment*

The treatment assignment \mathbf{Z} is random, i.e.

$$\mathbf{Z} \perp (\mathbf{Y}(0), \mathbf{Y}(1), \mathbf{C}(0), \mathbf{C}(1)),$$

which holds in a controlled randomized clinical trial. This assumption allows us to write $Y_j(z)$ and $C_j(z)$ as $Y_j|Z = z$ and $C_j|Z = z$ respectively.

III *Mean Monotonicity*

$$E[Y_j(z)|C(z) = c, C(1-z) = t] \leq E[Y_j(z)|C(z) = c', C(1-z) = t']$$

for $c' \geq c \geq j, t' \geq t, z = 0, 1$.

This assumption provides an ordering of the mean potential response at visit j under treatment z for all the principal cohorts of individuals who would be on study at visit j under treatment z . The means are assumed to not be worse for cohorts who remain on-study longer under both treatments. That is, the individuals who would be last seen at time c' ($c' \geq j$) under treatment z and time t' under treatment $1 - z$ will not have a worse mean potential response at time j under treatment z than individuals

who would last be seen at a time less than c' (but still greater than or equal to time j) under treatment z or a time less than t' under treatment $1 - z$.

The mean monotonicity assumption is often reasonable in clinical studies. For example, in a cardiovascular stent implantation trial, multiple endpoints including all-cause mortality free survival and 6-minute walk test score are used to evaluate the effectiveness of the device. Since the two endpoints are positively correlated, it is plausible to assume that patients will potentially perform better with their 6-minute walk tests if they have a longer survival time, i.e. remain on the study longer.

We introduce some further notation

1. $p_{c,t} = P(C(0) = c, C(1) = t)$.
2. $\gamma_{z,c} = P(C(z) = c)$.
3. $m_{j,z,c,t} = E[Y_j(z)|C(z) = c, C(1-z) = t] (c \geq j)$.
4. $\mu_{j,z,c} = E[Y_j(z)|C(z) = c] (c \geq j)$.

Note that under Assumption II (randomization), both $\gamma_{z,c} = P(C(z) = c) = P(C = c|Z = z)$ and $\mu_{j,z,c} = E[Y_j(z)|C(z) = c] = E[Y_j|C = c, Z = z]$ are identified by the observed data under the partial missing at random assumption.

The causal effect of interest $SACE_j$ can be expressed as

$$\begin{aligned}
 SACE_j &= E[Y_j(1) - Y_j(0)|C(1) \geq j, C(0) \geq j] \\
 &= E[Y_j(1) - Y_j(0), C(1) \geq j, C(0) \geq j]P(C(1) \geq j, C(0) \geq j)^{-1} \\
 &= \left(\sum_{c=j}^J \sum_{t=j}^J p_{c,t} \right)^{-1} \left\{ \sum_{c=j}^J \sum_{t=j}^J (m_{j,1,c,t} - m_{j,0,c,t}) p_{c,t} \right\}.
 \end{aligned} \tag{5-3}$$

The boundaries of $SACE_j$ in (5-3) can be found subject to the following restrictions:

1. $0 \leq p_{c,t} \leq 1$ for all c and t , and $\sum_{c=1}^J \sum_{t=1}^J p_{c,t} = 1$,
2. $\sum_{c=1}^J p_{c,t} = \gamma_{1,t}$ and $\sum_{t=1}^J p_{c,t} = \gamma_{0,c}$,
3. $\sum_{t=1}^J m_{j,0,c,t} p_{c,t} = \mu_{j,0,c} (c \geq j)$ and $\sum_{c=1}^J m_{j,1,t,c} p_{c,t} = \mu_{j,1,t} ((t \geq j) \text{ for all } j)$,
4. $m_{j,z,c',t'} \geq m_{j,z,c,t}$ for $c' \geq c \geq j$, $t' \geq t$ and all j and z .

where restrictions (1)-(2) are for $p_{s,t}$ to be a distribution with (identified) marginals, restriction (3) satisfies the identified conditional means, and restriction (4) comes from Assumption III.

Finding the boundaries of the *SACE*, i.e. finding the minimum and the maximum of the objective function (5-3), can be approximated (by ignoring the normalizing constant) as a non-convex quadratically constrained quadratic problem (QCQP) (Boyd and Vandenberghe, 1997, 2004). For a QCQP, a standard approach is to optimize a semidefinite relaxation of the QCQP and get lower and upper bounds on local optimal of the objective function (Boyd and Vandenberghe, 1997).

The uncertainty of the estimated bounds can be characterized in a Bayesian framework. The joint posterior distribution of the bounds can be constructed by implementing the optimization for each posterior sample of $\mu_{j,z,c}^*$, identified by the algorithm proposed in Section 5.3.3. The result can be presented as in Figure 5-1. A study decision might be based on the mode of the posterior joint distribution of the bounds.

5.3.5 Stochastic Survival Monotonicity Assumption

Under Assumption II, the marginal distributions $P(C(0))$ and $P(C(1))$ of $P(C(0) = c, C(1) = t)$ (re: 0 and 1 represent the placebo and treatment arm, respectively) are identified. However, the joint distribution remains unidentified without further assumption. We outline several Assumptions that will identify $p_{c,t}$ beyond the identified margins (Figure 5-2). These assumptions, when reasonable, will simplify the optimization of the objective function and yield more precise results.

1. $P(C(0) = m|C(1) = c) = q^{n-m}P(C(0) = n|C(1) = c)$ for $c \geq m \geq n$ and $q > 1$. That is, given a patient will “survive” until time point c on the treatment arm, the probability the patient will “survive” until time point $n - 1$ is q times the probability that the patient will “survive” until n for $n \leq c$ on the placebo arm. The parameter q is a sensitivity parameter.
2. $P(C(1) = t|C(0) = c) = 0$ for $c > t$.

That is, the chance that a patient will “survive” longer on the placebo arm than the active treatment arm is zero. This assumes the lower-triangle (excluding the diagonal) in Figure 5-2 is zero.

These assumptions may be incorporated in the optimization Bayesian framework to improve the precision of the posterior joint distribution of the bounds.

5.3.6 Summary of Causal Inference

We have outlined an approach to estimate the causal effect of treatment where there is dropout due to non-response reasons such as death. We also outlined an approach for posterior inference. We need to further explore point estimation of the intervals/bounds for the causal effect and characterizing their uncertainty in a Bayesian framework.

5.4 Figures

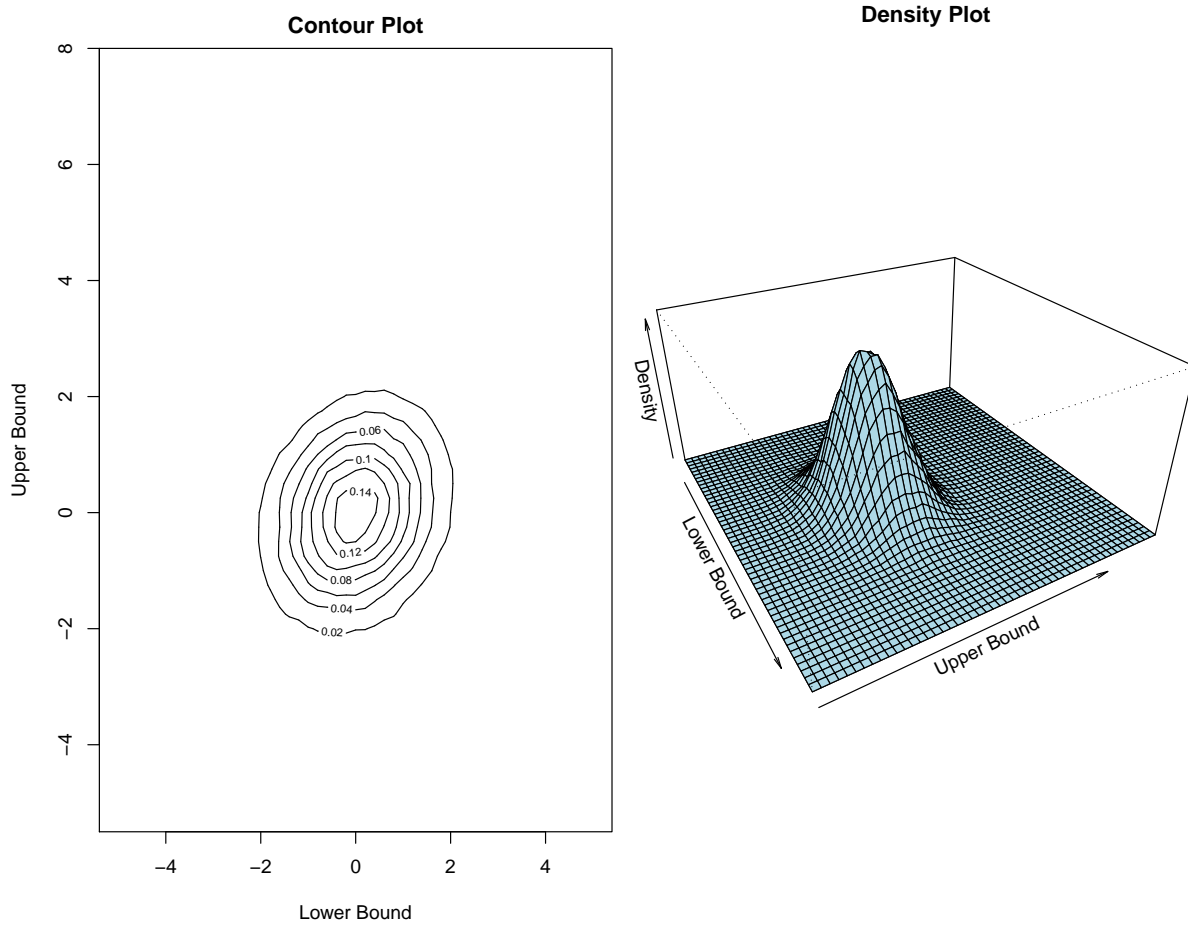


Figure 5-1. Contour and Perspective Plots of a Bivariate Density

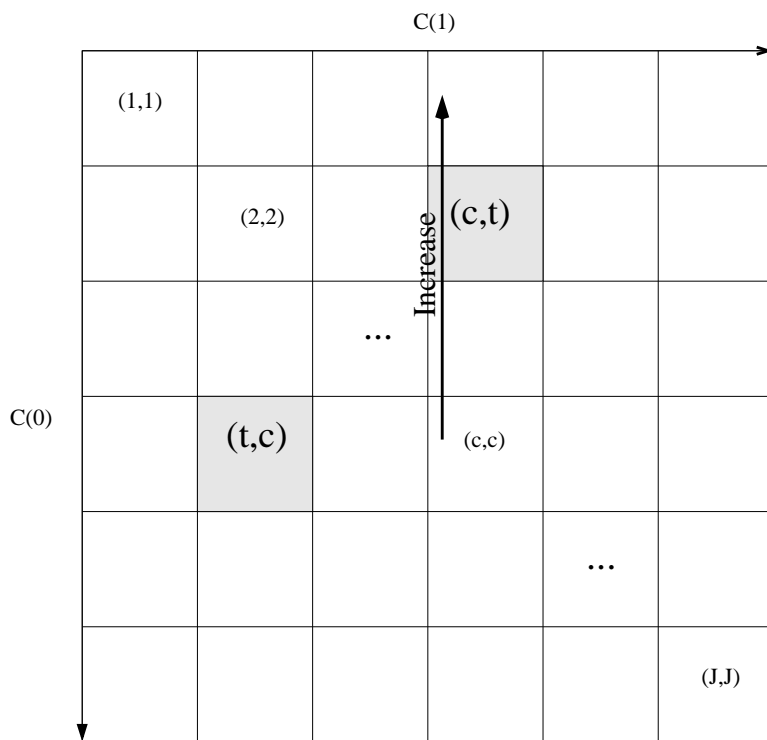


Figure 5-2. Illustration of $p_{c,t}$

REFERENCES

- Albert, P. (2000). A Transitional Model for Longitudinal Binary Data Subject to Nonignorable Missing Data. *Biometrics* **56**, 602–608.
- Albert, P., Follmann, D., Wang, S., and Suh, E. (2002). A latent autoregressive model for longitudinal binary data subject to informative missingness. *Biometrics* **58**, 631–642.
- Baker, S. (1995). Marginal regression for repeated binary data with outcome subject to non-ignorable non-response. *Biometrics* **51**, 1042–1052.
- Baker, S., Rosenberger, W., and DerSimonian, R. (1992). Closed-form estimates for missing counts in two-way contingency tables. *Statistics in Medicine* **11**, 643–657.
- Birmingham, J. and Fitzmaurice, G. (2002). A Pattern-Mixture Model for Longitudinal Binary Responses with Nonignorable Nonresponse. *Biometrics* **58**, 989–996.
- Boyd, S. and Vandenberghe, L. (1997). Semidefinite programming relaxations of non-convex problems in control and combinatorial optimization. *communications, computation, control and signal processing: a tribute to Thomas Kailath*.
- Boyd, S. and Vandenberghe, L. (2004). *Convex optimization*. Cambridge Univ Pr.
- Cheng, J. and Small, D. (2006). Bounds on causal effects in three-arm trials with non-compliance. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **68**, 815–836.
- Christiansen, C. and Morris, C. (1997). Hierarchical Poisson regression modeling. *Journal of the American Statistical Association* pages 618–632.
- Daniels, M. (1999). A prior for the variance in hierarchical models. *Canadian Journal of Statistics* **27**,
- Daniels, M. and Hogan, J. (2000). Reparameterizing the Pattern Mixture Model for Sensitivity Analyses Under Informative Dropout. *Biometrics* **56**, 1241–1248.
- Daniels, M. and Hogan, J. (2008). *Missing Data in Longitudinal Studies: Strategies for Bayesian Modeling and Sensitivity Analysis*. Chapman & Hall/CRC.
- DeGruttola, V. and Tu, X. (1994). Modelling Progression of CD4-lymphocyte Count and its Relationship to Survival Time. *Biometrics* **50**, 1003–1014.
- Diggle, P. and Kenward, M. (1994). Informative Drop-out Longitudinal Data Analysis. *Applied Statistics* **43**, 49–93.
- Egleston, B. L., Scharfstein, D. O., Freeman, E. E., and West, S. K. (2007). Causal inference for non-mortality outcomes in the presence of death. *Biostatistics* **8**, 526 – 545.

- Escobar, M. and West, M. (1995). Bayesian density estimation and inference using mixtures. *Journal of the American Statistical Association* pages 577–588.
- Fan, J. and Li, R. (2001). Variable Selection Via Nonconcave Penalized Likelihood and Its Oracle Properties. *Journal of the American Statistical Association* **96**, 1348–1361.
- Farnir, F., Grisart, B., Coppieters, W., Riquet, J., Berzi, P., Cambisano, N., Karim, L., Mni, M., Moisisio, S., Simon, P., et al. (2002). Simultaneous mining of linkage and linkage disequilibrium to fine map quantitative trait loci in outbred half-sib pedigrees: revisiting the location of a quantitative trait locus with major effect on milk production on bovine chromosome 14. *Genetics* **161**, 275.
- Faucett, C. and Thomas, D. (1996). Simultaneously modelling censored survival data and repeatedly measured covariates: a Gibbs sampling approach. *Statistics in Medicine* **15**,
- Fisher, B., Costantino, J., Wickerham, D., Redmond, C., Kavanah, M., Cronin, W., Vogel, V., Robidoux, A., Dimitrov, N., Atkins, J., Daly, M., Wieand, S., Tan-Chiu, E., Ford, L., Wolmark, N., other National Surgical Adjuvant Breast, and Investigators, B. P. (1998). Tamoxifen for prevention of breast cancer: report of the National Surgical Adjuvant Breast and Bowel Project P-1 study. *Journal of the National Cancer Institute* **90**, 1371–1388.
- Fitzmaurice, G. and Laird, N. (2000a). Generalized linear mixture models for handling nonignorable dropouts in longitudinal studies. *Biostatistics* **1**, 141–156.
- Fitzmaurice, G. and Laird, N. (2000b). Generalized Linear Mixture Models for Handling Nonignorable Dropouts in Longitudinal Studies. *Biostatistics* **1**, 141–156.
- Fitzmaurice, G., Molenberghs, G., and Lipsitz, S. (1995). Regression Models for Longitudinal Binary Responses with Informative Drop-Outs. *Journal of the Royal Statistical Society. Series B. Methodological* **57**, 691–704.
- Follmann, D. and Wu, M. (1995). An approximate generalized linear model with random effects for informative missing data. *Biometrics* pages 151–168.
- Forster, J. and Smith, P. (1998). Model-Based Inference for Categorical Survey Data Subject to Non-Ignorable Non-Response. *Journal of the Royal Statistical Society: Series B: Statistical Methodology* **60**, 57–70.
- Frangakis, C. E. and Rubin, D. B. (2002). Principal stratification in causal inference. *Biometrics* **58**, 21–29.
- Gilbert, P., Bosch, R., and Hudgens, M. (2003). Sensitivity analysis for the assessment of causal vaccine effects on viral load in HIV vaccine trials. *Biometrics* **59**, 531–541.
- Green, P. J. and Silverman, B. (1994). *Nonparametric Regression and Generalized Linear Models*. Chapman & Hall.

- Harel, O. and Schafer, J. (2009). Partial and latent ignorability in missing-data problems. *Biometrika* **96**, 37.
- Hayden, D., Pauler, D., and Schoenfeld, D. (2005). An estimator for treatment comparisons among survivors in randomized trials. *Biometrics* **61**, 305–310.
- Heagerty, P. (2002). Marginalized transition models and likelihood inference for longitudinal categorical data. *Biometrics* pages 342–351.
- Heckman, J. (1979a). Sample Selection Bias as a Specification Error. *Econometrica* **47**, 153–161.
- Heckman, J. (1979b). Sample selection bias as a specification error. *Econometrica: Journal of the econometric society* pages 153–161.
- Heitjan, D. and Rubin, D. (1991). Ignorability and coarse data. *The Annals of Statistics* pages 2244–2253.
- Henderson, R., Diggle, P., and Dobson, A. (2000). Joint modelling of longitudinal measurements and event time data. *Biostatistics* **1**, 465–480.
- Hogan, J. and Laird, N. (1997a). Mixture Models for the Joint Distribution of Repeated Measures and Event Times. *Statistics in Medicine* **16**, 239–257.
- Hogan, J. and Laird, N. (1997b). Model-Based Approaches to Analysing Incomplete Longitudinal and Failure Time Data. *Statistics in Medicine* **16**, 259–272.
- Hogan, J., Lin, X., and Herman, B. (2004). Mixtures of varying coefficient models for longitudinal data with discrete or continuous nonignorable dropout. *Biometrics* **60**, 854–864.
- Ibrahim, J. and Chen, M. (2000). Power prior distributions for regression models. *Statistical Science* pages 46–60.
- Ibrahim, J., Chen, M., and Lipsitz, S. (2001). Missing responses in generalised linear mixed models when the missing data mechanism is nonignorable. *Biometrika* **88**, 551.
- Kaciroti, N., Schork, M., Raghunathan, T., and Julius, S. (2009). A Bayesian Sensitivity Model for Intention-to-treat Analysis on Binary Outcomes with Dropouts. *Statistics in Medicine* **28**, 572–585.
- Kenward, M. and Molenberghs, G. (1999). Parametric Models for Incomplete Continuous and Categorical Longitudinal Data. *Statistical Methods in Medical Research* **8**, 51.
- Kenward, M., Molenberghs, G., and Thijs, H. (2003). Pattern-mixture models with proper time dependence. *Biometrika* **90**, 53–71.
- Kurland, B. and Heagerty, P. (2004). Marginalized Transition Models for Longitudinal Binary Data with Ignorable and Non-Ignorable Drop-Out. *Statistics in Medicine* **23**, 2673–2695.

- Laird, N. (1988). Missing data in longitudinal studies. *Statistics in Medicine* **7**,.
- Land, S., Wieand, S., Day, R., Ten Have, T., Costantino, J., Lang, W., and Ganz, P. (2002). Methodological Issues In the Analysis of Quality of Life Data in Clinical Trials: Illustrations from the National Surgical Adjuvant Breast And Bowel Project (NSABP) Breast Cancer Prevention Trial. *Statistical Methods for Quality of Life Studies* pages 71–85.
- Lee, J. and Berger, J. (2001). Semiparametric Bayesian Analysis of Selection Models. *Journal of the American Statistical Association* **96**, 1397–1409.
- Lee, J., Hogan, J., and Hitsman, B. (2008). Sensitivity analysis and informative priors for longitudinal binary data with outcome-related drop-out. *Technical Report, Brown University* .
- Lee, K., Daniels, M. J., and Sargent, D. J. (2010). Causal effects of treatments for informative missing data due to progression. *To Appear in JASA* .
- Liang, K.-Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73**, 13–22.
- Lin, H., McCulloch, C., and Rosenheck, R. (2004). Latent pattern mixture models for informative intermittent missing data in longitudinal studies. *Biometrics* **60**, 295–305.
- Little, R. (1993). Pattern-Mixture Models for Multivariate Incomplete Data. *Journal of the American Statistical Association* **88**, 125–134.
- Little, R. (1994). A Class of Pattern-Mixture Models for Normal Incomplete Data. *Biometrika* **81**, 471–483.
- Little, R. (1995). Modeling the drop-out mechanism in repeated-measures studies. *Journal of the American Statistical Association* **90**,.
- Little, R. and Rubin, D. (1987). *Statistical Analysis with Missing Data*. Wiley.
- Little, R. and Rubin, D. (1999). Comment on Adjusting for Non-Ignorable Drop-out Using Semiparametric Models by D.O. Scharfstein, A. Rotnitzky and J.M. Robins. *Journal of the American Statistical Association* **94**, 1130–1132.
- Little, R. and Wang, Y. (1996). Pattern-mixture models for multivariate incomplete data with covariates. *Biometrics* **52**, 98–111.
- Liu, T., Todhunter, R., Lu, Q., Schoettinger, L., Li, H., Littell, R., Burton-Wurster, N., Acland, G., Lust, G., and Wu, R. (2006). Modelling extent and distribution of zygotic disequilibrium: Implications for a multigenerational canine pedigree. *Genetics* .

- Liu, X., Waternaux, C., and Petkova, E. (1999). Influence of Human Immunodeficiency Virus Infection on Neurological Impairment: An Analysis of Longitudinal Binary Data with Informative Drop-Out. *Journal of the Royal Statistical Society (Series C): Applied Statistics* **48**, 103–115.
- Molenberghs, G., Kenward, M., and Lesaffre, E. (1997). The Analysis of Longitudinal Ordinal Data with Nonrandom Drop-Out. *Biometrika* **84**, 33–44.
- Molenberghs, G. and Kenward, M. G. (2007). *Missing Data in Clinical Studies*. Wiley.
- Molenberghs, G., Michiels, B., Kenward, M., and Diggle, P. (1998). Monotone Missing Data and Pattern-Mixture Models. *Statistica Neerlandica* **52**, 153–161.
- Neal, R. (2003). Slice sampling. *The Annals of Statistics* **31**, 705–741.
- Neyman, J. (1923). On the application of probability theory to agricultural experiments. *Statistical Science* **5**, 465–472.
- Nordheim, E. (1984). Inference from Nonrandomly Missing Categorical Data: an Example From a Genetic Study of Turner's Syndrome. *Journal of the American Statistical Association* **79**, 772–780.
- Pauler, D., McCoy, S., and Moinpour, C. (2003). Pattern Mixture Models for Longitudinal Quality of Life Studies in Advanced Stage Disease. *Statistics in Medicine* **22**, 795–809.
- Pulkstenis, E., Ten Have, T., and Landis, J. (1998). Model for the Analysis of Binary Longitudinal Pain Data Subject to Informative Dropout Through Remedication. *Journal of the American Statistical Association* **93**, 438–450.
- Radloff, L. (1977). The CES-D Scale: A Self-Report Depression Scale for Research in the General Population. *Applied Psychological Measurement* **1**, 385.
- Robins, J. (1997). Non-response models for the analysis of non-monotone non-ignorable missing data. *Statistics in Medicine* **16**, 21–37.
- Robins, J. (1998). Correction for non-compliance in equivalence trials. *Statistics in Medicine* **17**,.
- Robins, J. and Ritov, Y. (1997). Toward a curse of dimensionality appropriate(coda) asymptotic theory for semi-parametric models. *Statistics in Medicine* **16**, 285–319.
- Robins, J., Rotnitzky, A., and Zhao, L. (1994). Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association* **89**, 846–866.
- Robins, J., Rotnitzky, A., and Zhao, L. (1995). Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *Journal of the American Statistical Association* **90**,.

- Rotnitzky, A., Robins, J., and Scharfstein, D. (1998a). Semiparametric regression for repeated outcomes with nonignorable nonresponse. *Journal of the American Statistical Association* **93**, 1321–1322.
- Rotnitzky, A., Robins, J., and Scharfstein, D. (1998b). Semiparametric regression for repeated outcomes with nonignorable nonresponse. *Journal of the American Statistical Association* **93**, 1321–1322.
- Rotnitzky, A., Scharfstein, D. O., Su, T.-L., and Robins, J. M. (2001). Methods for conducting sensitivity analysis of trials with potentially nonignorable competing causes of censoring. *Biometrics* **57**, 103–113.
- Roy, J. (2003). Modeling Longitudinal Data with Nonignorable Dropouts Using a Latent Dropout Class Model. *Biometrics* **59**, 829–836.
- Roy, J. and Daniels, M. J. (2008). A general class of pattern mixture models for nonignorable dropout with many possible dropout times. *Biometrics* **64**, 538–545.
- Rubin, D. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology* **66**, 688–701.
- Rubin, D. (1976). Inference and missing data. *Biometrika* **63**, 581–592.
- Rubin, D. (1977). Formalizing subjective notions about the effect of nonrespondents in sample surveys. *Journal of the American Statistical Association* pages 538–543.
- Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*. Wiley.
- Rubin, D. B. (2000). Causal inference without counterfactuals: comment. *Journal of the American Statistical Association* pages 435–438.
- Scharfstein, D., Daniels, M., and Robins, J. (2003). Incorporating Prior Beliefs about Selection Bias into the Analysis of Randomized Trials with Missing Outcomes. *Biostatistics* **4**, 495.
- Scharfstein, D., Halloran, M., Chu, H., and Daniels, M. (2006). On estimation of vaccine efficacy using validation samples with selection bias. *Biostatistics* **7**, 615.
- Scharfstein, D., Manski, C., and Anthony, J. (2004). On the Construction of Bounds in Prospective Studies with Missing Ordinal Outcomes: Application to the Good Behavior Game Trial. *Biometrics* **60**, 154–164.
- Scharfstein, D., Rotnitzky, A., and Robins, J. (1999). Adjusting for Nonignorable Drop-Out Using Semiparametric Nonresponse Models. *Journal of the American Statistical Association* **94**, 1096–1146.

- Schulman, K., Berlin, J., Harless, W., Kerner, J., Sistrunk, S., Gersh, B., Dube, R., Taleghani, C., Burke, J., Williams, S., et al. (1999). The effect of race and sex on physicians recommendations for cardiac catheterization. *New England Journal of Medicine* **340**, 618–26.
- Shepherd, B., Gilbert, P., and Mehrotra, D. (2007). Eliciting a Counterfactual Sensitivity Parameter. *American Statistician* **61**, 56.
- Ten Have, T., Kunselman, A., Pulkstenis, E., and Landis, J. (1998). Mixed effects logistic regression models for longitudinal binary response data with informative drop-out. *Biometrics* **54**, 367–383.
- Ten Have, T., Miller, M., Reboussin, B., and James, M. (2000). Mixed Effects Logistic Regression Models for Longitudinal Ordinal Functional Response Data with Multiple-Cause Drop-Out from the Longitudinal Study of Aging. *Biometrics* **56**, 279–287.
- Thijs, H., Molenberghs, G., Michiels, B., Verbeke, G., and Curran, D. (2002). Strategies to fit pattern-mixture models. *Biostatistics* **3**, 245.
- Troxel, A., Harrington, D., and Lipsitz, S. (1998). Analysis of longitudinal data with non-ignorable non-monotone missing values. *Journal of the Royal Statistical Society. Series C (Applied Statistics)* **47**, 425–438.
- Troxel, A., Lipsitz, S., and Harrington, D. (1998). Marginal models for the analysis of longitudinal measurements with nonignorable non-monotone missing data. *Biometrika* **85**, 661.
- Tsiatis, A. A. (2006). *Semiparametric theory and missing data*. Springer, New York.
- van der Laan, M. J. and Robins, J. (2003). *Unified Methods for Censored Longitudinal Data and Causality*. Springer.
- Vansteelandt, S., Goetghebeur, E., Kenward, M., and Molenberghs, G. (2006a). Ignorance and uncertainty regions as inferential tools in a sensitivity analysis. *Statistica Sinica* **16**, 953–979.
- Vansteelandt, S., Goetghebeur, E., Kenward, M., and Molenberghs, G. (2006b). Ignorance and uncertainty regions as inferential tools in a sensitivity analysis. *Statistica Sinica* **16**, 953–979.
- Vansteelandt, S., Rotnitzky, A., and Robins, J. (2007). Estimation of regression models for the mean of repeated outcomes under nonignorable nonmonotone nonresponse. *Biometrika* **94**, 841.
- Wahba, G. (1990). *Spline models for observational data*. Society for Industrial Mathematics.

- Wang, C. and Daniels, M. (2009). Discussion of “Missing Data in longitudinal studies: A review” by Ibrahim and Molenberghs. *TEST* **18**, 51–58.
- Wang, C., Daniels, M., D.O., S., and Land, S. (2010). A Bayesian shrinkage model for incomplete longitudinal binary data with application to the breast cancer prevention trial. *To Appear in JASA* .
- Wu, M. and Bailey, K. (1988). Analysing changes in the presence of informative right censoring caused by death and withdrawal. *Statistics in Medicine* **7**,.
- Wu, M. and Bailey, K. (1989). Estimation and comparison of changes in the presence of informative right censoring: conditional linear model. *Biometrics* pages 939–955.
- Wu, M. and Carroll, R. (1988). Estimation and comparison of changes in the presence of informative right censoring by modeling the censoring process. *Biometrics* **44**, 175–188.
- Wulfsohn, M. and Tsiatis, A. (1997). A joint model for survival and longitudinal data measured with error. *Biometrics* pages 330–339.
- Yuan, Y. and Little, R. J. (2009). Mixed-effect hybrid models for longitudinal data with nonignorable drop-out. *Biometrics (in press)* .
- Zhang, J. and Heitjan, D. (2006). A simple local sensitivity analysis tool for nonignorable coarsening: application to dependent censoring. *Biometrics* **62**, 1260–1268.
- Zhang, J. and Rubin, D. (2003). Estimation of Causal Effects via Principal Stratification When Some Outcomes are Truncated by” Death”. *Journal of Educational and Behavioral Statistics* **28**, 353.

BIOGRAPHICAL SKETCH

Chenguang Wang received his bachelor's and master's degrees in computer science from Dalian University of Technology, China. Chenguang later joined the biometry program of and received his master's degree in statistics from University of Nebraska-Lincoln. At University of Florida, Chenguang's major was statistics while simultaneously working for the Children's Oncology Group Statistics and Data Center (2004-2009) and Center for Devices and Radiological Health, FDA (2009-2010). Chenguang received his Ph.D. from University of Florida in the summer of 2010.

Chenguang's research has focused on constructing a Bayesian framework for incomplete longitudinal data that identifies the parameters of interest and assesses sensitivity of the inference via incorporating expert opinions. Such a framework can be broadly used in clinical trials to provide health care professionals more accurate understanding of the statistical or causal relationship between clinical interventions and human diseases.

Chenguang is a member of American Statistical Association, a member of Eastern North American Region/International Biometric Society, and a member of Children's Oncology Group.