

2014

Informative censoring with an imprecise anchor event: estimation of change over time and implications for longitudinal data analysis

<https://hdl.handle.net/2144/14316>

Downloaded from DSpace Repository, DSpace Institution's institutional repository

BOSTON UNIVERSITY
GRADUATE SCHOOL OF ARTS AND SCIENCES

Dissertation

**INFORMATIVE CENSORING WITH AN IMPRECISE ANCHOR EVENT:
ESTIMATION OF CHANGE OVER TIME AND IMPLICATIONS FOR
LONGITUDINAL DATA ANALYSIS**

by

JAMIE ELIZABETH COLLINS

B.A., M.A., Boston University, 2005

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

2014

© Copyright by
JAMIE ELIZABETH COLLINS
2014

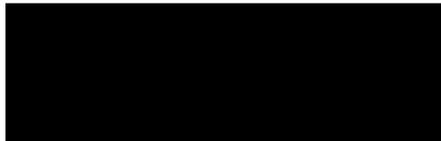
Approved by

First Reader



Elena Losina, PhD
Adjunct Associate Professor of Biostatistics

Second Reader



Lisa Sullivan, PhD
Professor of Biostatistics

Third Reader



Michael P. LaValley, PhD
Professor of Biostatistics

ACKNOWLEDGEMENTS

This dissertation would not be complete without giving thanks to the many people who contributed to its completion.

First, I would like to express my sincere gratitude and appreciation to my dissertation committee members. Elena Losina, my advisor, has been a tireless advocate of mine since the moment I stepped into OrACORe. I have learned so much in the past four years and will be forever grateful for your guidance, patience, and encouragement. It is sometimes a struggle to juggle everything and Elena has taught me to how to focus on what is truly important, work or otherwise, even in the face of the (nonexistent) “research emergencies.” I wouldn’t have started on the path I’m on now without Tim Heeren, who, unbeknownst to him, inspired me to pursue biostatistics after I took MA684 with him as an undergrad. Howard Cabral has been a mentor since the Summer Institute in Biostatistics (SIBS-I). He’s been there for everything over the past almost 10 years, whether it is comments on my dissertation, career or class advice, qualifying exam panic, SAS tips, or just to chat. I’m grateful to Mike LaValley for his thoughtful and thorough comments on my work. Taking BS820 with him as I finished my BA/MA started my sure to be lifelong love affair with SAS. Finally, I would like to thank Lisa Sullivan, of course for her comments and advice on my dissertation work, and for all of the guidance over the last 10 years. It is difficult to put into words how wonderful a mentor Lisa is (though everyone already knows this) and I would just like to express my deepest gratitude.

I've been fortunate to have many mentors along the way, and would like to express my thanks to them. Whenever I need advice I know that I can always count on Lori. I'm not sure where I would be without your help and advice. Whether it is panicked gchats, courtyard café lunches, or discussions over drinks, you've always been there for me (and for all of my SIBSters). I was lucky to spend 5 years learning and growing at DM-STAT, with Kim Dukes as my mentor. Kim trusted me with responsibilities well beyond my years and I grew not only as a biostatistician and a programmer, but as a person. Finally, I would like to thank Jeff Katz for all of his support over the last few years. Jeff handles everything with such grace and optimism and has taught me so much about being both a researcher and a leader. And all of the diet cokes don't hurt either.

I would have never made it this far without help and support from my classmates. To this day I credit Alison's notes as the reason I was able to get through the theory qualifier. Robin helped me focus when I was starting my dissertation, and I am so grateful to her guidance and advice during the point in the process when things are most confusing. Finally, Bill was a wonderful cubemate and it was nice to be able to follow in his footsteps. It's also wonderful knowing another Yankees fan.

My friends have stood by me through all of this, even though juggling everything wasn't always pretty. Emily has been through everything with me, from my first calculus exam to my first job to deciding to go back to school. Becka's support during my first years juggling school and full-time work was

amazing and I wouldn't have gotten through it without her. I'm thankful to the DM-STAT alumni committee for always being there for a DM-STAT drink. I'm especially grateful to have worked with my fellow SAS attackers EQ and Fay; I feel like in some ways we all grew up and became statisticians together. Finally my ORACORE friends Laurel, Doris, Yan, and Liz have been so supportive, especially through this last leg.

I wouldn't be where I am today without the love and support of my family, Mom, Dad & Denise, Tim, Zak, and Pat. The older I get the more I understand and appreciate all of the sacrifices that our parents made for us. They have supported me, cheered me on, and have pushed me to work hard. My brothers inspire me to be creative, passionate, to take risks, and to challenge myself. I love you all very much and I am very lucky to have you.

Finally I would like to thank Brendan. For your encouragement and support, for knowing when I need a break, or a beer, or a surprise trip to Iceland. You always take care of me. Thank you.

**INFORMATIVE CENSORING WITH AN IMPRECISE ANCHOR EVENT:
ESTIMATION OF CHANGE OVER TIME AND IMPLICATIONS FOR
LONGITUDINAL DATA ANALYSIS**

(ORDER NO.)

JAMIE ELIZABETH COLLINS

Boston University Graduate School of Arts and Sciences, 2014

Major Professor: Elena Losina, Associate Professor of Biostatistics

ABSTRACT

A number of methods have been developed to analyze longitudinal data with dropout. However, there is no uniformly accepted approach. Model performance, in terms of the bias and accuracy of the estimator, depends on the underlying missing data mechanism and it is unclear how existing methods will perform when little is known about the missing data mechanism.

Here we evaluate methods for estimating change over time in longitudinal studies with informative dropout in three settings: using a linear mixed effect (LME) estimator in the presence of multiple types of dropout; proposing an update to the pattern mixture modeling (PMM) approach in the presence of imprecision in identifying informative dropouts; and utilizing this new approach in the presence of prognostic factor by dropout interaction.

We demonstrate that amount of dropout, the proportion of dropout that is informative, and the variability in outcome all affect the performance of an LME

estimator in data with a mixture of informative and non-informative dropout.

When the amount of dropout is moderate to large (>20% overall) the potential for relative bias greater than 10% increases, especially with large variability in outcome measure, even under scenarios where only a portion of the dropouts are informative.

Under conditions where LME models do not perform well, it is necessary to take the missing data mechanism into account. We develop a method that extends the PMM approach to account for uncertainty in identifying informative dropouts. In scenarios with this uncertainty, the proposed method outperformed the traditional method in terms of bias and coverage.

In the presence of interaction between dropout and a prognostic factor, the LME model performed poorly, in terms of bias and coverage, in estimating prognostic factor-specific slopes and the interaction between the prognostic factor and time. The update to the PMM approach, proposed here, outperformed both the LME and traditional PMM.

Our work suggests that investigators must be cautious with any analysis of data with informative dropout. We found that particular attention must be paid to the model assumptions when the missing data mechanism is not well understood.

TABLE OF CONTENTS

CHAPTER 1: Introduction.....	1
CHAPTER 2: Estimation of Rate of Change in Longitudinal Studies with Varying Degrees of Missingness and Informative Dropout.....	5
2.1 Introduction.....	5
2.1.1 Estimating Disease Progression in Osteoarthritis.....	10
2.1.2 Motivating Example: Estimating Disease Progression in the Osteoarthritis Initiative.....	11
2.2 Methods.....	15
2.2.1 Overview.....	15
2.2.2 The Linear Mixed Effect Model.....	15
2.2.3 Simulation Study Details.....	16
2.2.3.1 Complete Data Generating Mechanism.....	16
2.2.3.2 Missing Data Mechanism.....	17
2.2.3.3 Parameters Evaluated.....	20
2.2.3.4 Criteria for Evaluation.....	20
2.3 Results.....	21
2.3.1 Missing Data Mechanism 1a.....	21
2.3.1.1 Relative Bias.....	21
2.3.1.2 Standard Error.....	22
2.3.1.3 Coverage.....	23
2.3.2 Other Missing Data Mechanisms.....	24

2.3.2.1 Relative Bias	24
2.3.2.2 Standard Error	24
2.3.2.3 Coverage	25
2.4 Discussion	26
CHAPTER 3: Estimation of Rate of Change in Longitudinal Studies with	
Informative Dropout and an Imprecise Anchor Event	
3.1. Introduction	52
3.1.1 Defining Anchor Event.....	53
3.1.2 Objectives.....	56
3.1.3 Motivating Example: Estimating Change in Joint Space Width in the	
Osteoarthritis Initiative	56
3.2. Methods	60
3.2.1 Overview	60
3.2.2 Notation	60
3.2.3 Pattern Mixture Models.....	61
3.2.3.1 Strategies for Defining Patterns	62
3.2.4 Simulation Study Details.....	65
3.2.4.1 Complete Data Generating Mechanism.....	65
3.2.4.2 Missing Data Mechanism.....	66
3.2.4.3 Anchor Event	68
3.2.4.4 Models considered.....	70
3.2.4.5 Parameters Evaluated	73

3.2.4.6 Criteria for Evaluation	73
3.3 Results	74
3.3.1 Perfect Anchor	74
3.3.1.1 Relative Bias	74
3.3.1.2 Standard Error	75
3.3.1.3 Coverage	75
3.3.2 Imperfect Anchor: Impact of Precision	76
3.3.2.1 Anchor 1: Anchor Associated with Underlying Rate of Change	76
3.3.2.2 Anchor 2: Anchor Randomly Assigned to MNAR Dropouts	79
3.4. Discussion.....	80
CHAPTER 4: Impact of Interaction Between Informative Dropout and a	
Prognostic Factor on Estimating Change Over Time in Longitudinal Studies... 106	
4.1 Introduction	106
4.2 Methods	109
4.2.1 Overview	109
4.2.2 Notation	109
4.2.3 Pattern Mixture Models.....	109
4.2.3.1 Strategies for Defining Patterns	110
4.2.3.2 Incorporating Prognostic Factor.....	112
4.2.4 Simulation Study Details.....	112
4.2.4.1 Complete Data Generating Mechanism.....	112
4.2.4.2 Missing Data Mechanism.....	115

4.2.4.3 Anchor Event	117
4.2.3.4 Models considered.....	118
4.2.4.5 Parameters Evaluated	122
4.2.4.6 Criteria for Evaluation	122
4.3. Results	123
4.3.1 Missing Data Mechanism 1 (Association between outcome and dropout time)	123
4.3.1.1 Relative Bias.....	123
4.3.1.2 Standard Error	125
4.3.1.3 Coverage	126
4.3.2 Impact of Missing Data Mechanism.....	126
4.4 Discussion	127
CHAPTER 5: Summary	151
REFERENCES	158
CURRICULUM VITAE	164

LIST OF TABLES

Table 2.1 Missing data in the Osteoarthritis Initiative (OAI).....	12
Table 2.2 Percent relative bias by sample size, standard deviation of change, total dropout, and nature of dropout; missing data mechanism 1.	33
Table 2.3 Standard Error by sample size, standard deviation of change, total dropout, and nature of dropout; missing data mechanism 1.....	34
Table 2.4 Coverage of 95% CI by sample size, standard deviation of change, total dropout, and nature of dropout; missing data mechanism 1.	35
Table 2.5 Percent relative bias by mechanism, standard deviation of change, total dropout, and nature of dropout; sample size=1000.....	36
Table 3.1 Percent relative bias by precision of anchor, S_b , and model.....	97
Table 3.1 Percent relative bias by precision of anchor, S_b , and model (continued).	98
Table 3.2 Standard error by precision of anchor, S_b , and model.	99
Table 3.2 Standard error by precision of anchor, S_b , and model (continued). ..	100
Table 3.3 Coverage of 95% CI by precision of anchor, S_b , and model.	101
Table 3.3 Coverage of 95% CI by precision of anchor, S_b , and model (continued).	102
Table 3.4 Percent relative bias by anchor, precision of anchor, S_b , total dropout and model.....	103
Table 3.5 Standard error by anchor, precision of anchor, S_b , total dropout and model.....	104

Table 3.6 Coverage of the 95% CI by anchor, precision of anchor, S_b , total dropout and model.....	105
Table 4.1 Percent relative bias by model, missing data mechanism, variability, total dropout and prevalence of PF.....	148
Table 4.2 Standard error by model, missing data mechanism, variability, total dropout and prevalence of PF.	149
Table 4.3 Coverage of 95% CI by model, missing data mechanism, variability, total dropout and prevalence of PF.....	150
Table 5.1 Estimate of yearly change in joint space width in the OAI.	154
Table 5.2 Group-specific estimates of yearly change in joint space width in the OAI.	154
Table 5.3 Estimate of yearly change in joint space width with standard error by obesity status in the OAI.....	156

LIST OF FIGURES

Figure 2.1 Mean joint space width with standard error by dropout and time in the OAI.	13
Figure 2.2 Mean WOMAC pain with standard error by dropout and time in the OAI.	14
Figure 2.3 Relative bias by total dropout and %MCAR (mechanism 1a).	37
Figure 2.4 Relative bias by sample size and %MCAR (mechanism 1a).	38
Figure 2.5 Relative bias by standard deviation and % MCAR (mechanism 1a)..	39
Figure 2.6 Standard error by standard deviation and % MCAR (mechanism 1a).	40
Figure 2.7 Standard error by sample size and % MCAR (mechanism 1a).	41
Figure 2.8 Coverage of 95% CI by standard deviation and sample size (mechanism 1a).	42
Figure 2.9 Estimate of yearly change with 95% CI by standard deviation and sample size (mechanism 1a).	43
Figure 2.10 Relative bias by total dropout and mechanism.	44
Figure 2.11 Standard error by total dropout and mechanism.	45
Figure 2.12 Coverage of 95% CI by total dropout/year and mechanism.	46
Figure 2.13 Coverage of 95% CI by sample size and mechanism.	47
Figure 2.14 Coverage of 95% CI by nature of dropout and mechanism.	48
Figure 2.15 Coverage of 95% CI by standard deviation and mechanism.	49

Figure 2.16 Estimate of yearly change with 95% CI by standard deviation and mechanism.....	50
Figure 2.17 Relative bias for scenarios plausible in the OAI.	51
Figure 3.1 Mean joint space width with standard error by total knee replacement status and visit.....	58
Figure 3.2 Mean joint space width with standard error by total knee replacement and completion status and visit.....	59
Figure 3.3 Relative bias by model and total dropout, perfect anchor ($S_b=1*\beta_1$; $S_e=0.5*\beta_1$).....	86
Figure 3.4 Relative bias by model and total dropout, perfect anchor ($S_b=2*\beta_1$; $S_e=1*\beta_1$).....	87
Figure 3.5 Relative bias by model and S_b , perfect anchor.....	88
Figure 3.6 Relative bias by model and S_e , perfect anchor.....	89
Figure 3.7 Standard error by model and total dropout, perfect anchor.....	90
Figure 3.8 Standard error by model and S_b , perfect anchor.....	91
Figure 3.9 Standard error by model and S_e , perfect anchor.....	92
Figure 3.10 Coverage of 95% CI by model and total dropout, perfect anchor....	93
Figure 3.11 Coverage of 95% CI by model and S_b , perfect anchor.....	94
Figure 3.12 Relative bias by precision of anchor, model, and total dropout ($S_b=1*\beta_1$; $S_e=0.5*\beta_1$).....	95
Figure 3.13 Relative bias by precision of anchor, model, and total dropout ($S_b=2*\beta_1$; $S_e=1*\beta_1$).....	96

Figure 4.1 Relative bias by model and total dropout.....	132
Figure 4.2 Relative bias by model and variability.....	133
Figure 4.3 Relative bias by model and prevalence of PF.	134
Figure 4.4 Relative bias by model and correlation between b_0 and b_1 (CC).	135
Figure 4.5 Relative bias of PF-specific slopes by model and total dropout.....	136
Figure 4.6 Relative bias of PF-specific slopes by model and prevalence of PF.	137
Figure 4.7 Standard error by model and total dropout.	138
Figure 4.8 Standard error by model and variability.....	139
Figure 4.9 Standard error by model and prevalence of PF.....	140
Figure 4.10 Standard error of PF-specific slopes by model and total dropout. .	141
Figure 4.11 Standard error of PF-specific slopes by model and prevalence of PF.	142
Figure 4.12 Coverage of 95% CI by model and total dropout.....	143
Figure 4.13 Coverage of 95% CI for PF-specific slopes by model and total dropout.	144
Figure 4.14 Coverage of 95% CI for PF-specific slopes by model and prevalence of PF.....	145
Figure 4.15 Relative bias by model and missing data mechanism.	146
Figure 4.16 Relative bias of PF-specific slopes by model and missing data mechanism.	147

LIST OF ABBREVIATIONS

CC	Correlation coefficient
CI	Confidence interval
JSW	Joint space width
KL	Kellgren-Lawrence grading scale
LME	Linear mixed effect
LOCF	Last observation carried forward
MAR	Missing at random
MCAR	Missing completely at random
MI	Multiple imputation
MNAR	Missing not at random
OA	Osteoarthritis
OAI	Osteoarthritis Initiative
PF	Prognostic factor
PMM	Pattern mixture model
PMM1	Pattern mixture model 1
PMM2	Pattern mixture model 2
PMM3	Pattern mixture model 3
SE	Standard error
TKR	Total knee replacement
WOMAC	Western Ontario and McMaster Universities Arthritis Index

CHAPTER 1: Introduction

Longitudinal studies often involve the measurement of change of an outcome variable over time. Over the course of a longitudinal study it is common for subjects to miss visits or to dropout before the scheduled end of follow-up. Missing data may lead to biased and imprecise estimates of change. In this dissertation we will investigate the effect of missing data on estimates of change over time in longitudinal data and propose a novel approach to estimate change when there is uncertainty in the nature and mechanism of the missing data.

The effect of missing data on the estimation of the rate of change depends on the underlying missing data mechanism. Following the terminology of Little and Rubin, data are said to be missing completely at random (MCAR) if the missing data mechanism is independent of the outcome [1]. Data are said to be missing at random (MAR) if the missing data mechanism is independent of the unobserved outcome, but may depend on the observed outcome or covariates. Finally, data are said to be missing not at random (MNAR) when the missing data mechanism depends on the unobserved outcome. When dropout is MNAR the missing data process is said to be informative. In this situation, standard statistical models for longitudinal data that fail to take into account the association between dropout patterns and the underlying rate of change can lead to biased and inaccurate estimators [2, 3].

The linear mixed effect model is often used to analyze longitudinal data. The model can easily handle missing data since there is no requirement that each subject is measured at the same time points or that each subject has the same number of measurements [4]. If the missing data are MCAR or MAR then likelihood based methods that correctly specify the entire joint distribution of the outcomes will yield valid estimates [5]. If the missing data are MNAR the linear mixed effect model will yield biased estimates of the mean response trend [6]. While this has been confirmed in simulation studies, in a number of scenarios that looked at low overall dropout (under 20% overall) and small residual variance, models that did not take into account the missing data mechanism tended to perform adequately in terms of relative bias and mean squared error. Schluchter et al. found that under scenarios with low dropout or small residual variance the bias of the estimator from a mixed effect model that assumed ignorable dropout was small (<5%) [7]. Li, et al found similar results with a log-normal informative missing data mechanism for scenarios with low dropout, but found moderate bias under small residual variance [8].

While under certain conditions LME models may be sufficient for data with MNAR dropout, ignoring informative dropout has the potential to lead to substantial bias and imprecision [3, 6, 9]. As a result, many methods have been proposed to account for the informative dropout process when a non-ignorable missing data mechanism is suspected. These methods can be grouped into three general classes: selection models, pattern mixture models, and frailty/shared

parameter models. In this dissertation, we focus on pattern mixture models. These models stratify the population based on the pattern of dropout and separately model each group. The complete data is a mixture of the conditional distributions over the dropout patterns [10, 11]. This requires that missingness can be categorized into distinct patterns and that the number of dropout times is small [12]. A model for the dropout mechanism is not explicitly specified, meaning that the relationship between missingness and unobserved outcome is not explicitly modeled. However, pattern mixture models are underidentified, meaning that all parameters cannot be estimated within each pattern unless additional restrictions are imposed. Little and others have proposed a variety of identifying restrictions in order to estimate all parameters [11].

All methods that incorporate the informative dropout process require knowledge and assumptions about the missing data process [6, 12]. Previous studies comparing various non-ignorable methods have found no uniformly superior method: the underlying missing data mechanism affects model performance [13]. Much of the previous work in this area has focused on diseases with a clear understanding of the association between disease progression and subject dropout; it is unclear how these methods will perform when little is known about the missing data mechanism.

In this dissertation, we will evaluate methods for analyzing longitudinal data in the presence of heterogeneity and uncertainty in the missing data mechanism. In Chapter 2 we will address the accuracy of the estimation of

change over time in longitudinal studies with informative dropout. This project will examine the implications of the nature of dropout and the variability of the outcome measure on the performance of estimators from a linear mixed effect model. Our goal is to determine whether there are scenarios where the “simple” approach of a linear mixed effects model is adequate and we can avoid the potential pitfalls and assumptions associated with more complex models. Under conditions where linear mixed models do not perform well, it is necessary to incorporate more sophisticated methods to account for missing data to improve the accuracy and precision of the estimate. In Chapter 3, we will propose an update to the pattern mixture model paradigm to account for uncertainty in the missing data mechanism. We will use a simulation study to compare the bias, accuracy, and coverage of the estimate of change over time across linear mixed effects models and different strategies for defining patterns in pattern mixture models, including the proposed update. In Chapter 4 we extend the pattern mixture model proposed in Chapter 3 to investigate the effect of interaction between dropout and a prognostic factor in estimating prognostic factor-specific slopes and the interaction between the prognostic factor and time in a longitudinal study. We conclude in Chapter 5 by summarizing the main conclusions and proposing areas for future research. Throughout the dissertation we will apply these methods to a longitudinal cohort study of the progression of knee osteoarthritis.

CHAPTER 2: Estimation of Rate of Change in Longitudinal Studies with Varying Degrees of Missingness and Informative Dropout

2.1 Introduction

Longitudinal studies involve a series of measurements over time on the same individual or observational unit, allowing for the direct study of change over time. Over the course of a longitudinal study it is common for subjects to miss visits or to drop out before the scheduled end of follow-up. Missing data may lead to biased and imprecise estimates.

The effect of missing data on the estimation of the rate of change depends on the underlying missing data mechanism. Following the terminology of Little and Rubin, data is said to be *missing completely at random (MCAR)* if the missing data mechanism is independent of the outcome [1, 14]. For example, if a subject moves out of the area and does not complete the study this would be judged unrelated to the outcome and the data would be considered MCAR. Data is said to be *missing at random (MAR)* if the missing data mechanism is independent of the unobserved outcome, but may depend on the observed outcome or covariates. For example, if older subjects are more likely to miss visits, but given a subject's age dropout is completely at random, this data would be considered MAR. Finally, data is said to be *missing not at random (MNAR)* when the missing data mechanism depends on the unobserved outcome. For example, in a study of the progression of knee osteoarthritis patients may be less

likely to complete follow-up visits as their function worsens. If this worsening is not observed – for example, a patient completes yearly visit 1 and then experiences a dramatic worsening of function and drops out of the study – then the data would be considered MNAR. When dropout is MNAR the missing data process is said to be *informative*. In this situation, standard statistical models for longitudinal data that fail to take into account the association between dropout patterns and the underlying rate of change can lead to biased and inaccurate estimators [2, 3, 14].

Many methods are used to analyze incomplete longitudinal data when the missing data are thought to be MAR or MCAR, including last observation carried forward (LOCF), complete case, imputation, and linear mixed models. LOCF has been widely criticized in the statistical literature. This method requires very strong assumptions including and in addition to MCAR dropout, and in general it is recommended that this method be avoided [5, 15-18]. Complete case is valid if the missing data are MCAR. In this case, dropout is unrelated to outcome and completers are a random sample of the original study sample. However, if the data is MAR complete case analysis will yield biased and inefficient estimators [5]. While imputation can be a useful strategy, especially when covariates are missing, in scenarios when only responses are missing it has been shown to not reduce bias or improve precision as compared to the linear mixed model approach [6, 19]. Estimates from linear mixed models are valid under MCAR and

MAR, provided that the joint distribution of responses is correctly specified [2, 15].

While the estimator from a linear mixed model (SAS PROC MIXED) performs well when data is MAR, under MNAR mechanisms it can demonstrate substantial bias [3, 6, 9]. As a result, many methods have been proposed to account for the informative dropout process when a non-ignorable missing data mechanism is suspected. These methods can be grouped into three general classes: selection models, pattern mixture models, and frailty/shared parameter models.

Selection models model the hypothetical complete data and then append a model for the missing data process [6, 20]. Let y be the outcome, r be the missing data process, and x be the set of covariates, the selection model can be factored as:

$$f(y, r|x) = f(y|x)f(r|y, x)$$

The model for the missing data process can be outcome-dependent, meaning that the process depends on observed and unobserved outcome Y_i , or random-effects dependent, meaning that the process depends on underlying random-effects b_i [20-22].

Pattern mixture models stratify the population based on the pattern of dropout and separately model each group. [10, 11, 23]

$$f(y, r|x) = f(y|r, x)f(r|x)$$

The complete data is a mixture of the conditional distributions, $f(y|r,x)$, over the dropout patterns. Under the PMM framework, the dropout mechanism, $f(r|x)$ does not depend on unobserved outcome; in other words, given the dropout pattern the missing data is ignorable. This requires that missingness can be categorized into distinct patterns and that the number of dropout times is small [10, 12].

Frailty models, also called shared parameter models or random-coefficient dependent dropout models, assume that there is some latent frailty or random effect that is shared between the dropout process and the outcome.

$$f(y, r|x) = \int f(y|\eta, x) f(r|\eta, x) dF(\eta|x)$$

Conditional on the frailties, η , the outcome and dropout process are independent [2, 7, 8].

Non-ignorable methods require that assumptions be made about the missing data process. These assumptions cannot be formally tested and must reflect both clinical and statistical considerations [15, 24]. Previous simulation studies comparing various non-ignorable methods have found no uniformly superior method: model performance depends on the underlying missing data mechanism, and no method is completely robust to misspecification [13].

Simulation studies evaluating these methods have confirmed that ignoring informative dropout can result in biased and imprecise estimators. However, in a few scenarios that looked at low overall dropout (under 20%) and small residual variance, models that ignored the missing data mechanism tended to perform adequately. Schluchter et al. found that under scenarios with low censoring

(approximately 15% cumulative dropout) or small residual variance the relative bias of the estimator from a mixed effect model that assumed ignorable dropout was slight. With a log-normal informative missing data mechanism the relative bias was 1.9% in the low censoring scenario and 5.7% in the small residual variance scenario. With a non-ignorable threshold missing data mechanism the relative biases were 5.8% and 3.9% respectively [7]. In the low censoring scenarios the estimator was more efficient (smaller MSE) than the proposed log-normal informative censoring model (which jointly models outcome and time to informative dropout). Li, et al found similar results with a log-normal informative missing data mechanism for scenarios with low dropout, but found moderate relative bias under small residual variance [8]. While these studies did examine scenarios with low overall dropout and small residual variance, these parameters were not varied together (e.g., small variance and low dropout).

The performance of models for longitudinal data that ignore informative dropout for estimating time trends in the outcome when overall dropout is relatively small or when only a portion of dropout is informative has not been examined. We designed a simulation study to evaluate the effect of different parameters on the estimates of the parameter for time trends in outcome from linear mixed models, ignoring the dropout mechanism. We examined the effect of the percent of dropout, proportion of dropout that is informative, sample size, standard deviation of change, and missing data mechanism on relative bias, standard error, and coverage of the estimate for change over time.

2.1.1 Estimating Disease Progression in Osteoarthritis

Recent estimates suggest that symptomatic knee osteoarthritis (OA) occurs in 6% of adults 30 years of age or older, and in 13% of adults age 60 and over [25, 26]. Knee OA is a slowly progressing disease; recent work by Eckstein et al. found very modest rates of cartilage loss on MRI-based parameters, with the most sensitive parameters showing loss rates of about 2% over a 12 month period [27]. In addition, the rates of loss are highly variable. The maximum standardized response mean in the Eckstein et al. study was 0.3, indicating that the standard deviation of change was approximately three times as large as the mean absolute change. Similar rates of changes were shown in both radiographic and MRI parameters by Duryea et al [28].

In studies of OA progression it is conceivable that as a subject's OA gets worse he or she may be more likely to dropout. A study to assess rates of cartilage loss in patients with knee OA examined patients clinically and obtained magnetic resonance (MR) images at baseline and at two follow-ups approximately 1 and 2 years post-baseline. Fourteen of 40 patients dropped out by the end of the study. Reasons for dropout included death, knee replacement, and difficulty getting to study appointments and patients with more severe OA were more likely to dropout [29]. Zhang et al note that selection bias introduced by loss to follow-up can be a major problem, especially when follow-up time is long [30]. The authors remark that "it is reasonable to assume that subjects who

have radiographic OA progression have a higher proportion of lost to follow-up than those who did not have radiographic OA progression.”

2.1.2 Motivating Example: Estimating Disease Progression in the Osteoarthritis Initiative

The Osteoarthritis Initiative (OAI) is a multi-center, longitudinal, prospective observational study of knee OA [31]. The goal of the OAI is to establish and maintain a natural history database for OA that will include clinical evaluation data, radiological images and image assessments, and a biospecimen repository. Subjects were enrolled into one of three sub-cohorts: the progression sub-cohort included patients with symptomatic, radiographic knee OA at baseline; the incidence sub-cohort included patients without radiographic knee OA at baseline, but with an elevated risk of developing OA (frequent knee symptoms, knee injury, etc.); the healthy controls included subjects without radiographic knee OA and without any risk factors. Subjects were assessed at baseline and then at yearly follow-up visits through four years, with a planned extension to follow subjects through eight years. If a subject could not come in to be evaluated in person, he or she was given the option to complete the clinical assessments over the telephone; image assessment and biospecimen collection were missed while patient reported outcomes, such as pain, function, and medication use, were collected.

We examined the amount of dropout and missing data in the OAI. Our analysis sample included subjects with radiographic, symptomatic knee OA at baseline (n=1,330). Seventy-one percent of patients completed the final study assessment at four years. Six percent of patients had no assessments after baseline. In addition to dropouts there were intermittent missing data: a subject missed a particular visit but came in for a subsequent visit. Seven percent of subjects missed the 12 month visit, but came in for a subsequent visit.

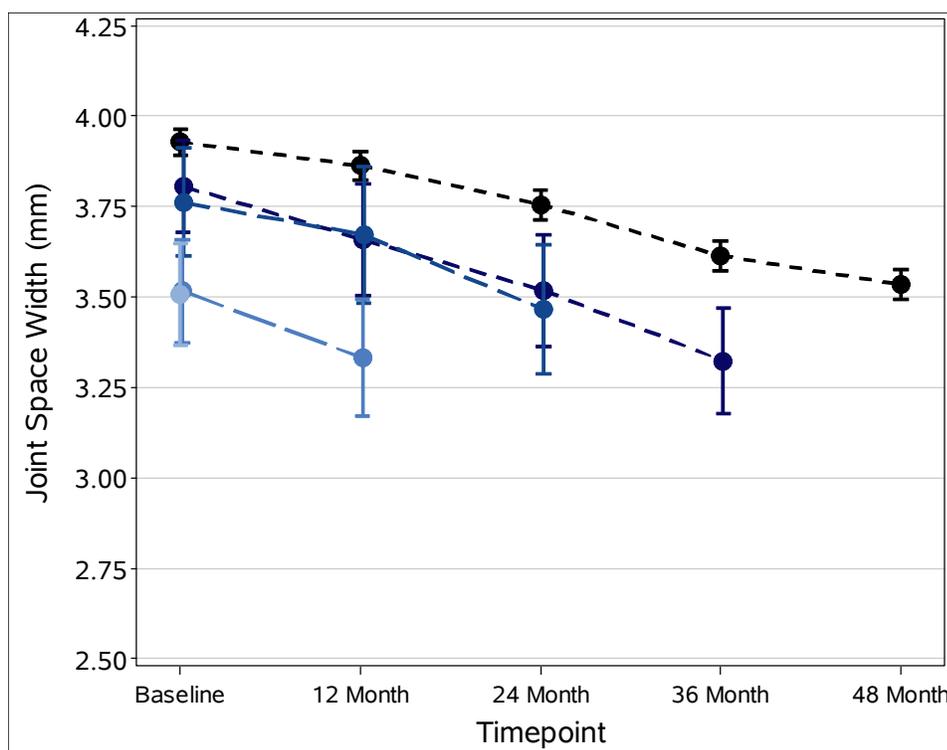
Table 2.1 Missing data in the Osteoarthritis Initiative (OAI)

Time point	Completed Visit	Dropout	Missed Visit
12 Month	1154 (86.8%)	84 (6.3%)	92 (6.9%)
24 Month	1065 (80.1%)	173 (13.0%)	92 (6.9%)
36 Month	992 (74.6%)	280 (21.1%)	58 (4.4%)
48 Month	940 (70.7%)	390 (29.3%)	NA

We grouped subjects by dropout time and examined various outcomes to determine if there was an association between dropout time and outcome. Radiography is used to quantify disease progression by measuring the narrowing of the joint space width (JSW) between the adjacent bones of the knee. Figure 1 shows mean JSW over time stratified by dropout group. On average, all groups experience loss of JSW over time. This loss appears to be slightly more severe for patients that dropout. For example, subjects not dropping out (black line) lose about 0.15mm of JSW on average through the 24 month time point, while subjects dropping out after the 24 month visit lose on average about 0.35mm of

JSW over this time period. There is a clear association between dropout and baseline JSW, with subjects not dropping out (black line) having larger baseline JSW as compared to dropouts.

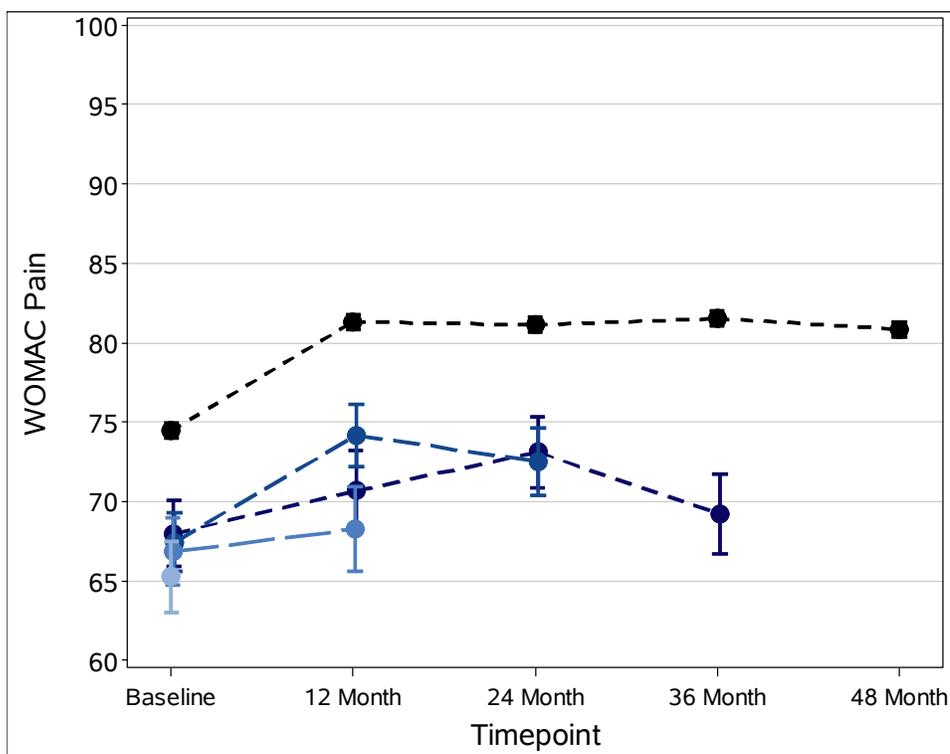
Figure 2.1 Mean joint space width with standard error by dropout and time in the OAI.



We evaluated the WOMAC pain scores of patients by time point and dropout group. The WOMAC pain scale is a 5 item survey that measures knee pain on a variety of daily tasks [32]. It is scored from 0 to 100 with lower scores indicating more severe pain. Figure 2.2 shows mean WOMAC pain over time stratified by dropout group. There is a clear relationship between dropout group and WOMAC pain score. The mean WOMAC pain at baseline was 65 in the group that dropped out after baseline and 74 in the group that completed the

study. The mean WOMAC score for patients dropping out after 24 and 36 months decreased in the visit prior to dropout. If patients are more likely to drop out as their pain worsens, we could be underestimating the rate of disease progression.

Figure 2.2 Mean WOMAC pain with standard error by dropout and time in the OAI.



2.2 Methods

2.2.1 Overview

The goal of this study was to evaluate the performance of general linear mixed models that ignore informative dropout when estimating time trends in outcome in longitudinal studies. Using a simulation study we compared the bias, standard error, and coverage of the model under a wide range of scenarios including variations in the dropout mechanism, the amount of overall dropout, the proportion of dropout that is informative, and a range of sample sizes and standard deviations of change.

2.2.2 The Linear Mixed Effect Model

The linear mixed effect model

$$Y_i = X_i\beta + Z_ib_i + e_i \quad (2.1)$$

was used to estimate the rate of change over time, where β is a vector of fixed effects, b_i is a vector of random effects, and X_i and Z_i are the corresponding matrices of covariates. The sampling error, e_i , is independent of b_i and $E(e_i) = 0$.

The covariance of Y_i has the structure

$$Y_i = Z_iGZ_i' + \sigma^2I_{n_i}$$

Where G is the covariance of the random effects

$$G = Cov(b_i) = \begin{pmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{pmatrix}$$

And $\sigma^2I_{n_i}$ is the sampling error for subject i .

The linear mixed effect model can easily handle missing data since there is no requirement that each subject is measured at the same time points or that subjects have the same number of measurements [4]. If the missing data are MCAR or MAR then likelihood based methods that correctly specify the entire joint distribution of the outcomes will yield valid estimates [5]. However if the missing data are MNAR the linear mixed effect model will yield biased estimates of the mean response trend [6].

2.2.3 Simulation Study Details

2.2.3.1 Complete Data Generating Mechanism

For each subject a continuous response variable was evaluated at baseline and at four fixed follow-up time points. A decline in response indicates clinical worsening. For each subject a vector of correlated responses, $Y_i = (Y_{i0}, Y_{i1}, Y_{i2}, Y_{i3}, Y_{i4})$ was generated. Data were generated under the linear mixed effect model (2.1) where:

$$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}, b_i = \begin{bmatrix} b_{0i} \\ b_{1i} \end{bmatrix}, X_i = \begin{bmatrix} t_1 \\ t_2 \\ \dots \\ t_n \end{bmatrix}, Z_i = \begin{bmatrix} t_1 \\ t_2 \\ \dots \\ t_n \end{bmatrix}$$

and

$$Y_{ij} = \beta_0 + \beta_1 t_{ij} + b_{0i} + b_{1i} t_{ij} + e_{ij}$$

Each subject's baseline value was drawn from a normal distribution with mean 100 and standard deviation 15: $\beta_0 = 100$; $b_{0i} \sim N(0,15)$.

Each subject's slope was drawn from a normal distribution with mean -7.5 and fixed standard deviation: $\beta_1 = -7.5$; $b_{1i} \sim N(0, SD)$. β_1 was used to estimate the rate of change.

The residual variance was 2: $e_{ij} \sim N(0, 2)$.

2.2.3.2 Missing Data Mechanism

For missing data mechanism 1 the response and the change in response from the previous time point were grouped into quintiles at each time point. An ordinal group variable was created for both outcome and change, ranging from 1 to 5 with 1 indicating the best disposition (highest response/least amount of deterioration) and 5 indicating the worst disposition (lowest response/most amount of deterioration). Based on our data generating mechanism, it was possible for subjects to improve over time. These subjects are by definition grouped into the "best" quintile for change. We did observe subjects with improving pain and function scores in the OAI.

$$qo_{it} = 6 - (\text{quintile of response at time } t \text{ for subject } i)$$

$$qc_{it} = (\text{quintile of change from time } t-1 \text{ to time } t \text{ for subject } i)$$

The probability of dropping out between time point $t-1$ and t is modeled as:

$$\pi_{it} = \alpha_{0t} + \alpha_{1t}qo_{it-1} + \alpha_{2t}qc_{it-1}I_{(t>1)} + \alpha_{3t}qo_{it} + \alpha_{4t}qc_{it} \quad (2.2)$$

α_0 is the underlying probability of dropping out at any given time point independent of outcome. α_1 and α_2 are coefficients that correspond to a MAR process: the dropout at time t is associated with outcome at time $t-1$ (α_1) or

change from time t-2 to t-1 (α_2), while α_3 and α_4 correspond to a MNAR process: the dropout at time t is associated with the (possibly unobserved) outcome at time t (α_3) or with the change from time t-1 to t (α_4). For example, $\alpha = (0, 0, 0, 0, 0.05)$ is an MNAR process: each quintile decrease in change from time t-1 to t is associated with an increase in probability of dropout of 5%. Those in the highest quintile group (most change) at time t would have a probability of dropout of 25% compared to a probability of 5% in the lowest group. The alpha value $\alpha_{4t} = 0.05$ was chosen to create an overall dropout rate of 15%.

Within this framework we created 3 MNAR mechanisms:

Mechanism 1a: Dropout at time t depends on (possibly unobserved) change from time t-1 to t. Equal drop-out across all time points.

$$\alpha = (\alpha_0, 0, 0, 0, \alpha_4)$$

$$\alpha_{01} = \alpha_{02} = \alpha_{03} = \alpha_{04} = \alpha_{05}; \alpha_{41} = \alpha_{42} = \alpha_{43} = \alpha_{44} = \alpha_{45}$$

Mechanism 1b: Dropout at time t depends on (possibly unobserved) change from time t-1 to t. Drop-out is twice as likely between baseline and time point 1 than at subsequent time points

$$\alpha = (\alpha_0, 0, 0, 0, \alpha_4)$$

$$\alpha_{01} = 2\alpha_{02} = 2\alpha_{03} = 2\alpha_{04} = 2\alpha_{05}; \alpha_{41} = 2\alpha_{42} = 2\alpha_{43} = 2\alpha_{44} = 2\alpha_{45}$$

Mechanism 1c: Dropout at time t depends on (possibly unobserved) outcome at time t. Equal drop-out across all time points.

$$\alpha = (\alpha_0, 0, 0, \alpha_3, 0)$$

$$\alpha_{01} = \alpha_{02} = \alpha_{03} = \alpha_{04} = \alpha_{05}; \alpha_{31} = \alpha_{32} = \alpha_{33} = \alpha_{34} = \alpha_{35}$$

Mechanism 2: As a sensitivity analysis we included a second missing data mechanism. We modeled the log odds of dropping out between time t-1 and t as:

$$\begin{aligned} \text{logit}(\pi_{it}) = & \alpha_{0t} + \alpha_{1t} \text{outcome}_{it-1} + \alpha_{2t} \text{change}_{it-1} I_{(t>1)} + \alpha_{3t} \text{outcome}_{it} + \\ & \alpha_{4t} \text{change}_{it} \end{aligned} \quad (2.3)$$

The alphas are as described above: α_0 corresponds to an MCAR process; α_1 and α_2 correspond to a MAR process; and α_3 and α_4 correspond to a MNAR process. Dropout at time t depends on (possibly unobserved) change from time t-1 to t. Equal drop-out across all time points. The alphas were chosen so that the odds of dropout increase by 1.5 for each 1 standard deviation increase in change_{it} .

$$\alpha = (\alpha_0, 0, 0, 0, \alpha_4)$$

$$\alpha_{01} = \alpha_{02} = \alpha_{03} = \alpha_{04} = \alpha_{05}; \alpha_{41} = \alpha_{42} = \alpha_{43} = \alpha_{44} = \alpha_{45}$$

At each time point, a Bernoulli random variable, r_{it} , was drawn with p equal to π_{it} to determine if subject i dropped out at time t. Once a subject was determined to dropout that subject was not observed at any subsequent time points, i.e., missing data was only due to dropout.

2.2.3.3 Parameters Evaluated

We examined sample sizes of 50, 100, 200, 500, and 1000. The standard deviation of change ($b_{1i} \sim N(0,SD)$) was examined at one half, one, one and a half, two, and three times the rate of change (i.e., 3.75, 7.5, 11.25, 15, 22.5). The total dropout by end of study ranged from approximately 15% to 70%.

The nature of the drop-out was varied so that the percent of drop-out that was random went from 0% to 100%. This was accomplished by varying the alphas in (2.2) and (2.3). The alphas were weighted according to the nature of the dropout by parameter r . For example, for mechanism 1: $\pi_{it} = r * \alpha_0 + (1 - r) * \alpha_4 q c_{it}$. For $r=1$ (100% random model): $\pi_{it} = \alpha_0$. For $r=0.4$ (40% random model): $\pi_{it} = 0.4 * \alpha_0 + (0.6) * \alpha_4 q c_{it}$. r varied from 0% to 100% by 20%. We implemented this in order to create a heterogeneous dropout mechanism: in a clinical study it is likely that there are different reasons for patient dropout, some related to outcome and others unrelated.

2.2.3.4 Criteria for Evaluation

We evaluated absolute bias, relative bias, standard error (SE), coverage probability, and length of the 95% confidence interval (CI) for the estimate of change over time. The absolute bias is calculated by subtracting the estimated progression estimate from the fixed true value and the relative bias is calculated by dividing the absolute bias by the fixed true value of the parameter estimate. The coverage probability is the proportion of times the 95% confidence interval

includes the true progression estimate. Five thousand replicates were run for each scenario. All simulations were conducted in R (<http://cran.r-project.org/>).

2.3 Results

2.3.1 Missing Data Mechanism 1a

2.3.1.1 Relative Bias

Table 2.2 presents relative bias by total dropout, nature of dropout, sample size, and standard deviation of change for low (5% year/overall dropout 18%) moderate (10% year/overall dropout 34%), and severe total dropout (20% year/overall dropout 60%) and small (n=50) moderate (n=200) and large (n=1000) sample sizes for missing data mechanism 1a ($\alpha = (\alpha_0, 0, 0, 0, \alpha_4)$; $\alpha_{01} = \alpha_{02} = \alpha_{03} = \alpha_{04} = \alpha_{05}$). When dropout is completely at random ($r=1$) the relative bias is approximately zero for all combinations of sample size, standard deviation, and total dropout. When bias is present it is negative, indicating that the model is under-estimating mean change. As expected, bias increases (became more negative) with increasing total dropout and as the dropouts go from 100% MCAR ($r=1$) to 0% MCAR ($r=0$) (Figure 2.3). For example, for overall dropout of 10% per year, sample size of 1000, and standard deviation of change at 7.5, relative bias was approximately 0.1% when dropout was 100% completely at random, -2.4% when dropout was 60% completely at random, and -5.9% when dropout was 0% completely at random.

Mean relative bias was approximately the same across the different sample sizes (Figure 2.4). The variability in observed bias increased as the sample size decreased. For example, Figure 2.4 shows that for the 0% MCAR model with moderate dropout (10% year/overall dropout 34%) and standard deviation of change equal to 1.5 times slope, the interquartile range for observed relative bias across all simulations increases from 6.7 for the scenario with a sample size of 1,000 to 30.3 for a sample size of 50. There was a strong association between bias and standard deviation of change. For overall dropout of 10% per year, sample size of 1000, and a 0% random model ($r=0$) the relative bias ranged from -3.8% for a standard deviation of 3.75 (1/2 times change) to -15.2% for a standard deviation of 22.5 (3 times change) (Figure 2.5).

2.3.1.2 Standard Error

There was no difference in $SE(\hat{\beta})$ across different levels of dropout nature: as the mechanism moved from 0% MCAR to 100% MCAR $SE(\hat{\beta})$ did not change considerably (Table 2.3). There was a slight increase in $SE(\hat{\beta})$ as the overall amount of dropout increased: for a 0% MCAR model with a sample size of 200 and standard deviation of change of 11.25 the standard error increased from 1.63 to 1.67 to 1.77 as dropout per year increased from 5% to 10% to 20%.

Sample size and standard deviation of change were both associated with $SE(\hat{\beta})$. As expected, standard error is highest for the smallest sample sizes and largest standard deviations of change. For the 0% MCAR model with 10%

dropout per year the standard error ranged from 3.3 for a sample size of 50 and standard deviation of change 22.5 to 0.15 for a sample size of 1000 and standard deviation of change 3.75. (Figures 2.6-2.7)

2.3.1.3 Coverage

Figure 2.8 shows the coverage of the 95% confidence interval for the slope by nature of dropout and overall amount of dropout for the base cases where standard deviation of change is 11.25 and sample size is 1000. For the 100% MCAR model coverage is maintained at 95% for all dropout amounts. Coverage is maintained at close to 95% for the overall dropout rate of 2.5% across all amounts of informative dropout. As the overall dropout rate increases the effect of the MNAR dropout is seen: when dropout is 5% per year coverage drops from 94.6% for the 80% MCAR model to 87.2% for the 0% MCAR model.

There appeared to be a paradoxical relationship between sample size, standard deviation and coverage. As sample size decreased and standard deviation increased the coverage was closer to the nominal value of 95% (Table 2.4). In the scenario where total dropout is 10% per year, the dropout is 0% at random, and the standard deviation of change is 7.5, coverage is 80% when the sample size is 500 and drops to 64% when the sample size increases to 1000. If the sample size remains at 500 and standard deviation of change decreases to 3.75 the coverage drops to 65%. We examined the length of the 95% confidence interval and found that the length increases dramatically as the sample size

decreases and standard deviation of change increases (Figure 2.9). For a sample size of 50 and standard deviation of 22.5 then length of the 95% confidence interval is 13. The confidence interval is likely to contain the true estimate of the slope simply because the interval is extremely wide.

2.3.2 Other Missing Data Mechanisms

2.3.2.1 Relative Bias

The relationship between overall dropout, nature of dropout, sample size standard deviation of change, and relative bias was similar to mechanism 1 for the other missing data mechanisms: bias increased with increasing standard deviation, with increasing overall amount of dropout, and as the nature of dropout went from 100% MCAR to 0% MCAR. There was not a strong association with sample size. (Table 2.5).

Mechanism 1b (dropout was twice as likely between baseline and time point 1 than at subsequent time points) had the highest bias across all scenarios. The bias was slightly lower for Mechanism 1c (dropout depended on outcome at time t instead of change from time $t-1$ to t). Finally, the bias for mechanism 2 was close to that for Mechanism 1b (Figure 2.10).

2.3.2.2 Standard Error

The relationship between overall dropout, nature of dropout, sample size standard deviation of change, and standard error was similar to mechanism 1 for

the other missing data mechanisms: standard error increases with increasing standard deviation and decreasing sample size. There was a slight association with amount of dropout: as overall amount of dropout increases standard error increases slightly; however there was not a strong association with dropout nature. The standard error is similar across all dropout mechanisms, though slightly higher for mechanism 1b (Figure 2.11).

2.3.2.3 Coverage

The relationship between overall dropout, nature of dropout, sample size and coverage was similar to mechanism 1 for the other missing data mechanisms: Coverage decreases with increasing total dropout and as the percent of dropout that is completely at random decreases. The paradoxical relationship with sample size remains across all mechanisms: as sample size increases from 50 to 1000 coverage decreases. Coverage was highest for mechanism 1c (Figures 2.12 - 2.14).

Mechanisms 1b and 2 demonstrate the same relationship between standard deviation of change and coverage as mechanisms 1a: as the standard deviation of change increases the coverage decreases. For mechanism 1b the coverage is approximately 20% when standard deviation of change is 3.75 and increases to 30% when standard deviation of change increases to 22.5; for mechanism 1c the coverage is approximately 33% and increases to 74% when standard deviation of change increases to 22.5. This relationship is not seen with

mechanism 1c: as the standard deviation of change increases the coverage decreases. (Figure 2.15) Examining the estimate of yearly change and length of the 95% confidence interval explains this trend: Mechanism 1c has the lowest bias across all three mechanisms. Because Mechanisms 1a, 1b, and 2 have more bias, the scenarios with smaller standard deviations have poor coverage because the length of the 95% confidence interval is short and isn't able to reach the true rate of change of -7.5. Mechanism 1c is less biased, and though the length of the interval is short the point estimate of change is close to the true value of -7.5, so the 95% confidence interval includes the true value (Figure 2.16).

2.4 Discussion

We evaluated the impact of sample size, standard deviation, missing data mechanism, amount of overall dropout, and nature of dropout on estimating the rate of change in a longitudinal study using a linear mixed effect model. We found that the standard deviation of change, amount of overall dropout and nature of the dropout were important factors in the performance of the estimator. The missing data mechanism was important in terms of whether the missingness depended on change (worsening) over time, or on the actual outcome value, and whether the dropouts were concentrated at the beginning of the study.

For scenarios with small amounts of dropout (approximately <20% overall/5% per year over 4 years) the bias of the estimator from the linear mixed

model was relatively minor when the standard deviation of change was moderate (≤ 1.5 times the rate of change). Even as the standard deviation of change increased to 3 times the rate of change, relative bias only surpassed 10% for the model 1b, with early dropouts. For scenarios with moderate to large amounts of dropout (>20% overall/5% per year over 4 years) even slight deviations from the 100% random dropout lead to substantial bias, especially if the standard deviation of change is substantial. For example, for missing data mechanism 1a and dropout of 10% per year relative bias increases from 0% for the 100% random model to 5% for the 40% random model (standard deviation of change of 1.5 times the rate of change, sample size=1000).

The results of our study are consistent with other results in the literature. Schlucter et al found that the bias from a maximum likelihood estimator in the presence of informative dropout depends on the amount of censoring [7]. Relative bias ranged from -1.9% in a low censoring scenario to -40.9% in a high censoring scenario. Bias also increased as the within-subject variance increased. The MSE increased dramatically in both cases as well, particularly when the amount of censoring increased. Touloumi et al observed modest bias of -10% for a non-ignorable missing data mechanism and about -5% for a compound ignorable/non-ignorable mechanism when the between subject variance was relatively small and overall dropout was approximately 50%, which is consistent with our results [33].

We found that the standard deviation of change is an important factor in model performance. This makes sense intuitively: as the between-subject variance increases subjects become more different from one another. The subjects that are doing the worst – and are the most likely to drop out – are doing much worse than the population average. If the between-subject variance is small then subjects are more similar to each other, and the subjects doing the worst generally are not doing that much worse than the population average. While the mean relative bias was stable across sample size, we found that the variability in observed bias increased as the sample size decreased. This means that the chance of a single study producing biased results increases with decreasing sample size.

The timing and nature of the dropouts were also important factors. While the timing of dropouts is something that is easy to observe and quantify – investigators should be especially concerned if dropouts are clustered at the beginning of a longitudinal study – the nature of the dropout mechanism is more difficult to measure. Whether the probability of dropout depended on the change from the last time point (mechanism 1a) or the actual outcome value (mechanism 1c) made a clear difference in model performance. Both are plausible scenarios: a patient might drop out of a study if he/she experiences a sudden substantial worsening in symptoms, or a patient might drop out if he/she reaches a threshold in symptoms. Model performance worsened as the amount of dropout that was informative increased. It is reasonable to believe that some dropouts will be

related to outcome while others will not. If an investigator collects reasons for dropout, and can assuredly rule out an informative mechanism in some cases, then confidence in the estimator from the linear mixed model should increase, especially when the overall amount of dropout is not large to begin with.

As with any study that utilizes simulations, our study had several limitations. Missing data in our study resulted only from dropout; there was no intermittent missing data. It is unlikely that a study would have no intermittent missing data, and adding additional missingness may increase bias and decrease precision even more than what we observed. In addition, we only examined four missing data mechanisms. Reasons for dropout in a longitudinal study are complex and it is unlikely that the missing data mechanism would be as simple as described above. Finally, baseline covariates were not incorporated. It is likely that some dropout in a longitudinal study is at least partially dependent on other covariates, such as age. Including these in the dropout mechanism and in the generalized linear model could have decreased bias and improved precision. Data were simulated as multivariate normal with a normal error term. In reality, the distribution of the WOMAC score from our motivating example is bounded between 0 and 100, and is moderately skewed. We chose our data generating mechanism for ease of interpretation, and a different underlying distribution or error distribution could lead to different conclusions. This is a topic for future work.

In designing and carrying out a longitudinal study, investigators should of course aim to limit subject dropout. Study design features could include incentives for completing each follow-up visit and for completing the study, clear protocols for research staff to follow in contacting patients (e.g., make four phone calls at least 3 days apart with a final phone call from the principal investigator), and reducing the respondent burden by limiting the amount of data collected/length of data report forms [34]. Another important consideration in the design phase of the study is developing a way to accurately record reason for dropout. In our study we found that the nature of the dropout is extremely important in model performance. Better understanding of reasons for dropout, and hence the dropout mechanism, will allow the investigator to tailor the analytic approach accordingly. Detail in this area is crucial; many clinical studies use a blanket term “withdrew consent” for dropouts, which tells the investigator nothing about the association between dropout and outcome.

Some dropout is unavoidable; for an investigator to determine the potential impact of informative dropout, it is important to quantify the factors that we found to have the biggest impact on model performance: amount of dropout, nature of dropout, timing of dropout and variability in the outcome. The total amount of dropout and nature of dropout (informative vs. non-informative) can be considered together to determine the total amount of informative dropout. For example, a study with 40% overall dropout of which only 15% is informative may be less problematic than a study with 30% overall dropout of which 75% is

informative. The timing of the dropouts is important as well: early dropouts, especially dropouts occurring before the first follow-up visit, had a larger impact on model performance than later dropouts. The relative bias under dropout mechanism 1b, with dropout twice as likely to occur at baseline as compared to later time points, was approximately 50% higher than the relative bias under dropout mechanism 1a, with dropout equal across time points. Finally, preliminary models and descriptive statistics can be used to understand variability in the outcome measure. Investigators should consider these factors together, and refer to the results of the study (Tables 2.3, 2.4, and 2.5) to determine an analytic approach.

In the OAI, approximately 6% of patients dropped out each year. Dropout was relatively equal across time points, with slightly more occurring early in the study. The sample size was large, with 1,330 subjects in the analysis cohort, and variability in change in JSW over time was high, approximately twice the rate of change. Based on Table 2.3 we estimate that using a linear mixed effects model to estimate annual change could be underestimating change by up to 10%. Because the variability is high, there is potential for even greater relative bias, as shown by the widths of the boxplots in Figures 2.3 and 2.5. Figure 2.17 presents relative bias for scenarios plausible in the OAI: Mechanisms 1a and 1c, dropout per year 5% and 10%, standard deviation equal to twice the slope, and sample size equal to 1000. While most scenarios demonstrate mean relative bias less than 10%, the interquartile range extends to 15%. We may wish to investigate

MNAR methods as a sensitivity analysis to understand how our estimates of change over time change if we model the informative dropout mechanism. With a small number of possible dropout times and uncertainty about the exact nature of the missing data mechanism we would use a pattern mixture modeling approach to investigate incorporating a possible MNAR dropout mechanism.

In this paper we evaluated the performance of a general linear mixed model for evaluating the rate of change in a longitudinal study under a variety of scenarios. We considered the impact of sample size, standard deviation, missing data mechanism, amount of overall dropout, and nature of dropout. We found that a number of factors affect model performance in estimating change over time, both observable (e.g., sample size, overall amount of dropout), and unobservable (e.g., missing data mechanism). Even when the overall amount of dropout is relatively small the model performance can be inadequate depending on other factors. Investigators should proceed with caution in the presence of any missing data and should work to understand the factors influencing model performance

Table 2.2 Percent relative bias by sample size, standard deviation of change, total dropout, and nature of dropout; missing data mechanism 1.

		Total dropout											
		5% year				10% year				20% year			
		Standard deviation of change				Standard deviation of change				Standard deviation of change			
r (% MCAR)	n	$\frac{1}{2}\beta_1$	$1\beta_1$	$2\beta_1$	$3\beta_1$	$\frac{1}{2}\beta_1$	$1\beta_1$	$2\beta_1$	$3\beta_1$	$\frac{1}{2}\beta_1$	$1\beta_1$	$2\beta_1$	$3\beta_1$
0	50	-1.9	-2.7	-5.0	-6.8	-3.9	-6.2	-10.0	-15.4	-8.2	-12.8	-23.6	-33.1
	200	-1.8	-2.8	-4.8	-6.9	-3.8	-5.7	-10.5	-15.0	-8.3	-13.0	-23.2	-34.1
	1000	-1.8	-2.8	-4.9	-7.3	-3.8	-5.9	-10.5	-15.2	-8.3	-13.1	-23.2	-34.2
0.2	50	-1.3	-1.9	-4.0	-6.0	-3.2	-5.1	-8.5	-12.7	-6.6	-10.5	-19.1	-26.9
	200	-1.4	-2.2	-4.0	-5.7	-3.1	-4.8	-8.1	-12.2	-6.7	-10.5	-18.7	-26.9
	1000	-1.4	-2.3	-4.0	-5.8	-3.0	-4.7	-8.4	-12.1	-6.6	-10.5	-18.6	-27.3
0.4	50	-1.3	-2.0	-2.8	-3.7	-2.1	-3.5	-5.9	-8.9	-5.0	-7.8	-14.0	-20.7
	200	-1.1	-1.5	-2.9	-4.3	-2.3	-3.5	-6.2	-9.2	-5.0	-7.8	-14.1	-20.4
	1000	-1.1	-1.7	-2.9	-4.2	-2.3	-3.6	-6.2	-9.0	-4.9	-7.8	-13.9	-20.5
0.6	50	-0.6	-0.8	-1.8	-3.4	-1.6	-2.1	-3.8	-5.2	-3.4	-4.9	-9.6	-13.5
	200	-0.7	-1.1	-2.0	-3.0	-1.5	-2.6	-4.1	-6.2	-3.4	-5.2	-9.8	-14.2
	1000	-0.7	-1.2	-2.0	-2.9	-1.5	-2.4	-4.3	-6.0	-3.3	-5.2	-9.5	-13.5
0.8	50	-0.4	-0.6	-1.2	-1.8	-0.8	-1.2	-2.4	-3.5	-1.6	-2.5	-4.2	-7.6
	200	-0.4	-0.6	-0.8	-1.2	-0.8	-1.2	-2.2	-2.9	-1.7	-2.7	-5.0	-6.5
	1000	-0.4	-0.6	-0.9	-1.2	-0.8	-1.1	-2.0	-3.0	-1.7	-2.6	-4.7	-6.7
1	50	-0.1	0.1	0.1	0.6	0.1	0.0	0.4	-0.1	-0.2	-0.1	-0.1	-0.2
	200	0.0	-0.0	0.2	-0.5	-0.0	-0.0	-0.3	-0.1	-0.0	0.0	-0.2	0.0
	1000	0.0	-0.0	-0.1	-0.1	0.0	-0.0	0.1	0.1	-0.0	0.1	0.0	0.1

Scenarios with relative bias greater than or equal to 10% are shaded

Table 2.3 Standard Error by sample size, standard deviation of change, total dropout, and nature of dropout; missing data mechanism 1.

		Total dropout											
		5% year				10% year				20% year			
		Standard deviation of change				Standard deviation of change				Standard deviation of change			
r (% MCAR)	n	$\frac{1}{2}*\beta_1$	$1*\beta_1$	$2*\beta_1$	$3*\beta_1$	$\frac{1}{2}*\beta_1$	$1*\beta_1$	$2*\beta_1$	$3*\beta_1$	$\frac{1}{2}*\beta_1$	$1*\beta_1$	$2*\beta_1$	$3*\beta_1$
0	50	0.55	1.09	2.17	3.26	0.58	1.12	2.23	3.33	0.62	1.19	2.35	3.51
	200	0.28	0.55	1.09	1.63	0.29	0.56	1.12	1.67	0.31	0.60	1.18	1.77
	1000	0.12	0.25	0.49	0.73	0.13	0.25	0.50	0.75	0.14	0.27	0.53	0.79
0.2	50	0.55	1.09	2.17	3.24	0.58	1.13	2.23	3.33	0.62	1.19	2.35	3.51
	200	0.28	0.55	1.09	1.63	0.29	0.56	1.12	1.68	0.31	0.60	1.19	1.77
	1000	0.12	0.24	0.49	0.73	0.13	0.25	0.50	0.75	0.14	0.27	0.53	0.79
0.4	50	0.55	1.09	2.16	3.26	0.58	1.12	2.23	3.33	0.62	1.19	2.36	3.53
	200	0.28	0.55	1.09	1.63	0.29	0.56	1.12	1.68	0.31	0.60	1.19	1.77
	1000	0.12	0.25	0.49	0.73	0.13	0.25	0.50	0.75	0.14	0.27	0.53	0.79
0.6	50	0.56	1.09	2.17	3.25	0.58	1.12	2.23	3.34	0.62	1.19	2.36	3.53
	200	0.28	0.55	1.09	1.63	0.29	0.56	1.12	1.68	0.31	0.60	1.19	1.78
	1000	0.12	0.25	0.49	0.73	0.13	0.25	0.50	0.75	0.14	0.27	0.53	0.80
0.8	50	0.56	1.09	2.17	3.26	0.58	1.12	2.23	3.34	0.63	1.20	2.37	3.54
	200	0.28	0.55	1.09	1.63	0.29	0.56	1.12	1.68	0.31	0.60	1.19	1.78
	1000	0.12	0.25	0.49	0.73	0.13	0.25	0.50	0.75	0.14	0.27	0.53	0.80
1	50	0.55	1.09	2.17	3.25	0.58	1.12	2.23	3.34	0.62	1.20	2.37	3.55
	200	0.28	0.55	1.09	1.63	0.29	0.56	1.12	1.68	0.31	0.60	1.19	1.78
	1000	0.12	0.25	0.49	0.73	0.13	0.25	0.50	0.75	0.14	0.27	0.53	0.80

Table 2.4 Coverage of 95% CI by sample size, standard deviation of change, total dropout, and nature of dropout; missing data mechanism 1.

		Total dropout											
		5% year				10% year				20% year			
		Standard deviation of change				Standard deviation of change				Standard deviation of change			
r (% MCAR)	n	$\frac{1}{2}*\beta_1$	$1*\beta_1$	$2*\beta_1$	$3*\beta_1$	$\frac{1}{2}*\beta_1$	$1*\beta_1$	$2*\beta_1$	$3*\beta_1$	$\frac{1}{2}*\beta_1$	$1*\beta_1$	$2*\beta_1$	$3*\beta_1$
0	50	93.9	94.6	94.4	94.8	91.5	92.7	93.4	92.9	82.4	86.6	87.2	88.0
	200	92.4	93.4	93.6	93.8	83.4	88.3	89.1	89.4	48.0	62.7	69.2	70.0
	1000	81.0	86.2	88.1	88.1	40.1	58.1	65.0	66.7	0.4	4.2	9.2	10.3
0.2	50	93.7	94.7	93.9	94.5	92.6	93.0	93.5	93.5	87.6	89.3	90.0	90.5
	200	93.1	94.1	93.7	93.8	87.5	89.8	91.8	91.4	63.5	74.1	77.9	79.4
	1000	86.1	89.3	90.7	90.9	58.6	71.4	75.8	77.0	5.6	16.4	25.1	26.8
0.4	50	94.5	94.7	94.2	94.3	93.8	93.5	94.5	93.8	89.8	91.2	91.9	92.6
	200	94.0	94.4	94.3	94.6	90.2	92.3	92.4	92.7	77.6	83.7	85.2	85.6
	1000	90.1	92.3	92.9	92.5	73.8	81.6	84.7	85.2	24.7	42.1	49.7	51.0
0.6	50	94.7	94.8	94.1	94.4	94.0	94.6	94.1	94.5	92.3	93.4	93.5	93.8
	200	94.6	94.7	94.8	94.7	92.8	93.4	93.6	94.4	87.2	89.6	90.4	90.8
	1000	93.0	93.2	94.0	94.1	85.9	89.0	90.0	90.9	57.3	69.9	73.1	74.8
0.8	50	94.7	94.5	94.3	94.5	94.4	94.3	94.4	94.4	93.4	94.0	94.1	93.9
	200	94.9	95.1	94.9	95.0	94.6	94.9	94.8	94.5	93.0	93.7	93.9	94.1
	1000	94.3	94.8	95.0	95.0	92.7	93.5	93.7	93.6	85.7	88.5	89.5	90.3
1	50	94.6	94.0	94.3	94.6	94.6	94.7	94.3	94.8	94.3	94.7	94.0	94.7
	200	94.9	94.6	95.2	94.5	95.0	95.0	95.1	95.1	95.0	95.0	94.8	95.1
	1000	95.0	94.9	95.1	95.0	94.8	94.5	95.3	94.7	94.9	95.1	95.2	94.7

Table 2.5 Percent relative bias by mechanism, standard deviation of change, total dropout, and nature of dropout; sample size=1000.

		Total dropout								
		5% year			10% year			20% year		
		Standard deviation of change			Standard deviation of change			Standard deviation of change		
Mech.	r (% MCAR)	$\frac{1}{2}\beta_1$	$1.5\beta_1$	$3\beta_1$	$\frac{1}{2}\beta_1$	$1.5\beta_1$	$3\beta_1$	$\frac{1}{2}\beta_1$	$1.5\beta_1$	$3\beta_1$
1b	0	-2.3	-6.1	-12.0	-4.9	-13.0	-25.7	-11.9	-32.3	-63.0
	0.2	-1.9	-4.8	-9.4	-3.9	-10.6	-20.4	-9.6	-25.7	-50.9
	0.4	-1.4	-3.6	-6.9	-3.0	-7.9	-15.3	-7.1	-19.2	-37.9
	0.6	-0.9	-2.4	-4.9	-2.0	-5.2	-10.4	-4.8	-12.8	-25.0
	0.8	-0.4	-1.1	-2.3	-1.0	-2.6	-5.3	-2.4	-6.5	-12.5
	1	0.0	0.0	0.1	-0.0	0.0	0.2	0.0	0.1	0.1
1c	0	-0.6	-2.3	-6.0	-1.0	-4.8	-12.3	-2.4	-10.9	-28.4
	0.2	-0.4	-1.8	-5.0	-0.9	-3.9	-9.8	-2.0	-8.6	-22.8
	0.4	-0.3	-1.3	-3.7	-0.7	-2.8	-7.7	-1.5	-6.4	-16.8
	0.6	-0.2	-0.9	-2.1	-0.4	-1.9	-4.8	-0.9	-4.3	-10.7
	0.8	-0.1	-0.4	-1.3	-0.3	-0.9	-2.3	-0.5	-2.1	-5.5
	1	0.1	0.0	-0.1	-0.0	0.0	-0.1	-0.0	0.1	0.1
2	0	-2.2	-3.7	-6.7	-4.1	-7.2	-13.1	-7.7	-13.7	-25.1
	0.2	-1.7	-2.9	-5.6	-3.3	-5.8	-10.3	-6.1	-10.8	-20.3
	0.4	-1.3	-2.2	-4.2	-2.5	-4.3	-8.1	-4.6	-8.1	-15.1
	0.6	-0.9	-1.5	-2.6	-1.7	-2.7	-5.1	-3.1	-5.4	-9.8
	0.8	-0.4	-0.7	-1.3	-0.8	-1.5	-2.6	-1.5	-2.6	-4.8
	1	0.0	0.1	0.2	0.0	0.0	0.1	0.0	-0.1	0.1

Figure 2.3 Relative bias by total dropout and %MCAR (mechanism 1a).

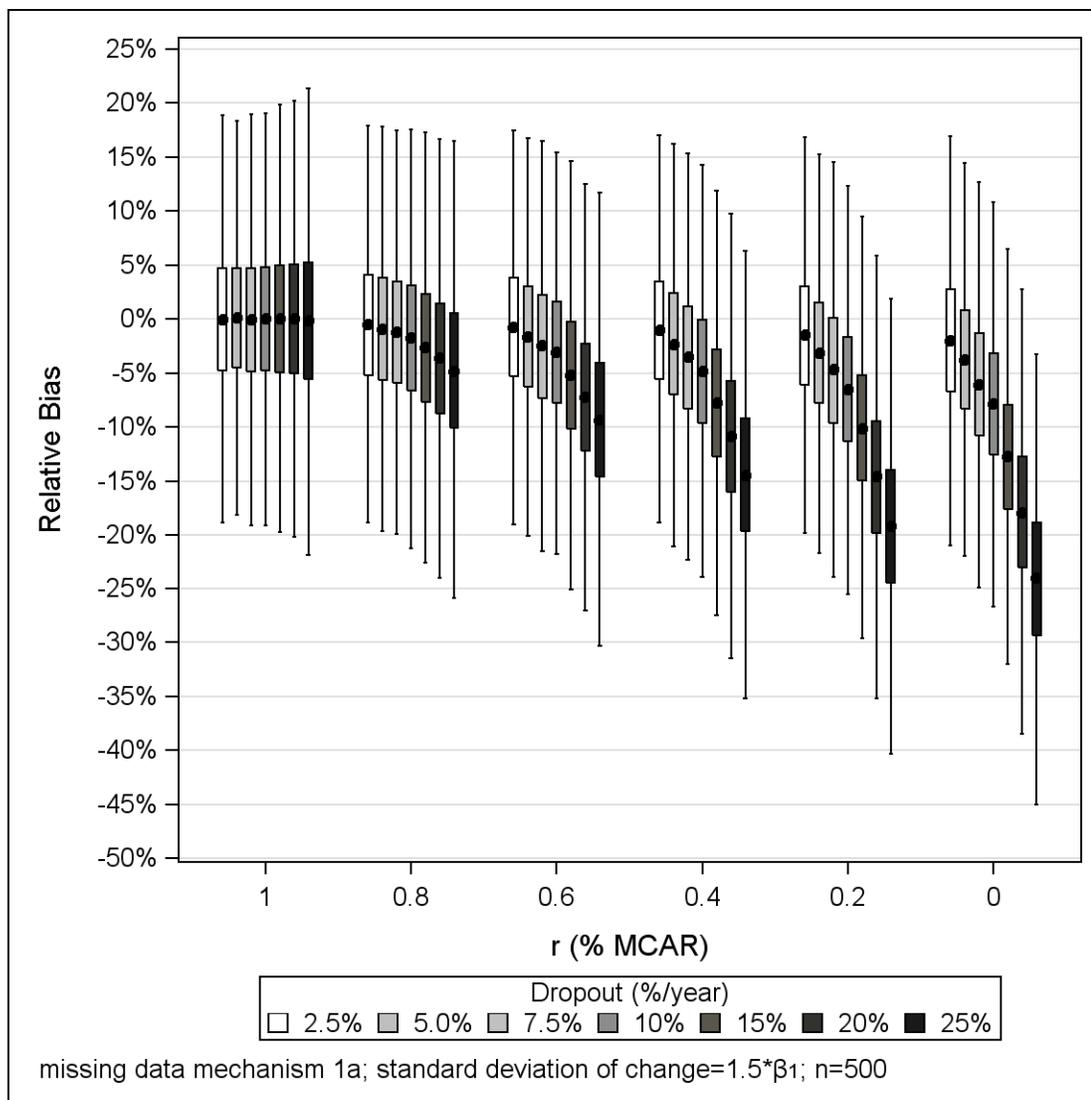


Figure 2.4 Relative bias by sample size and %MCAR (mechanism 1a).

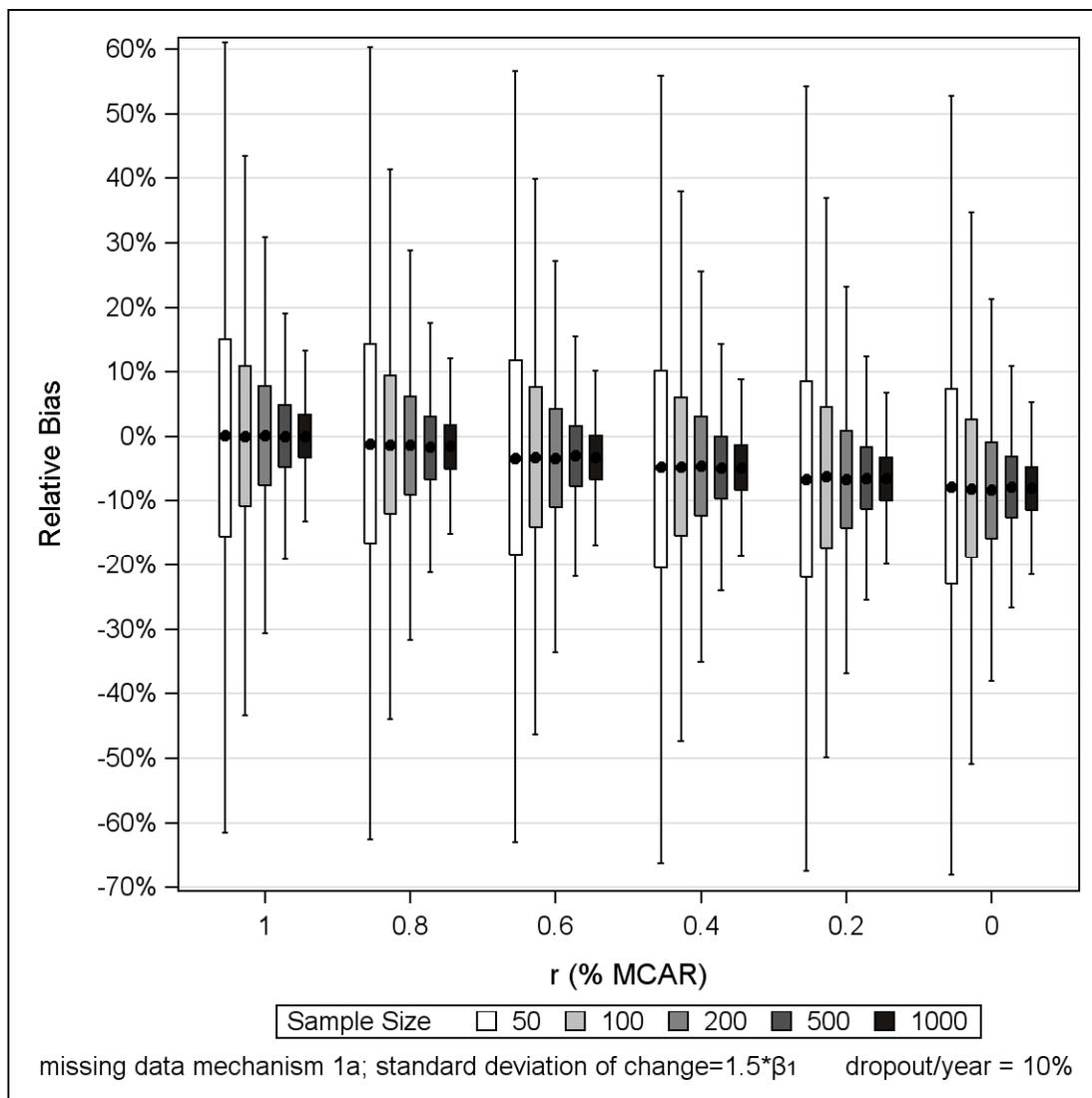


Figure 2.5 Relative bias by standard deviation and % MCAR (mechanism 1a).

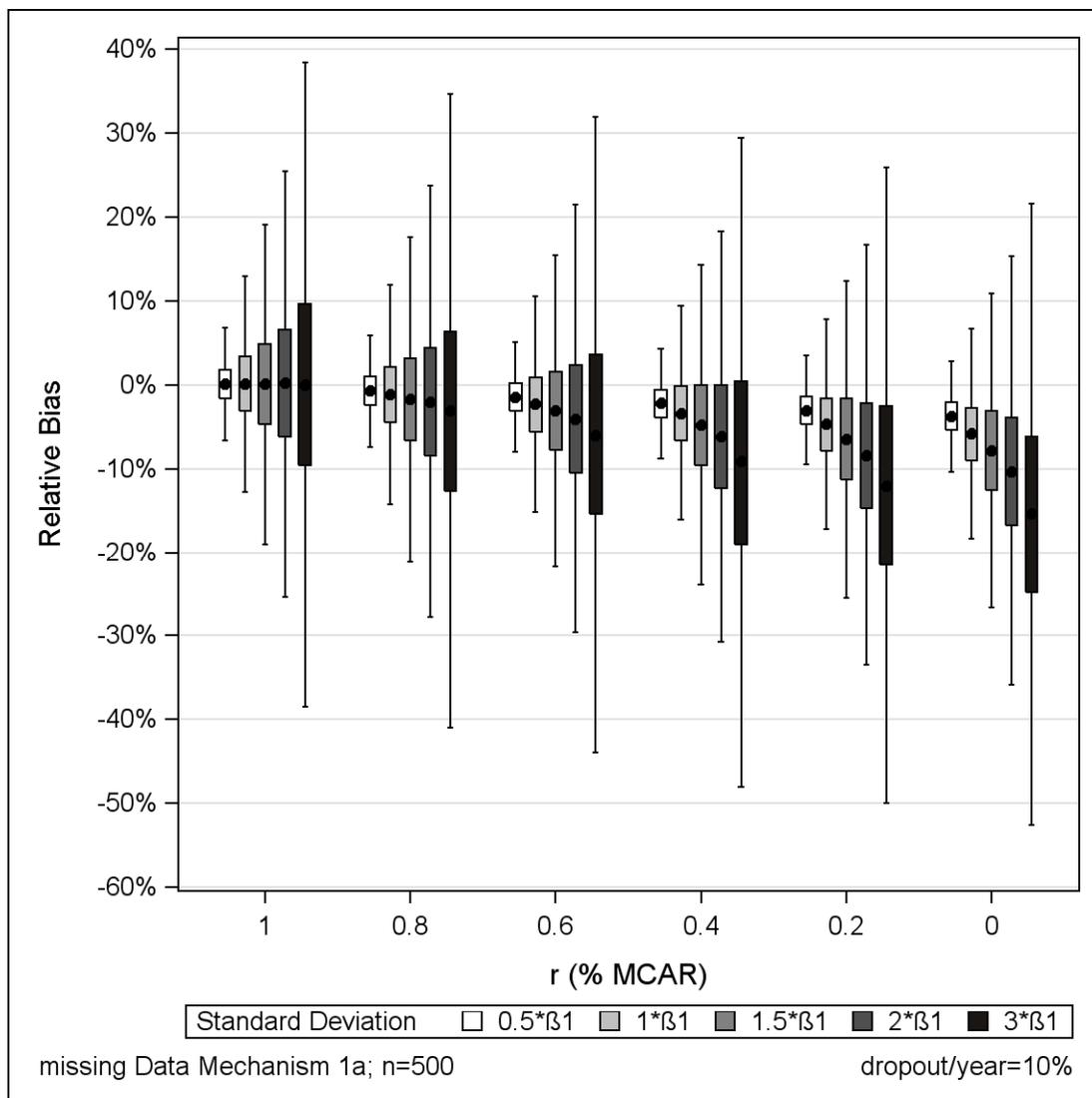


Figure 2.6 Standard error by standard deviation and % MCAR (mechanism 1a).

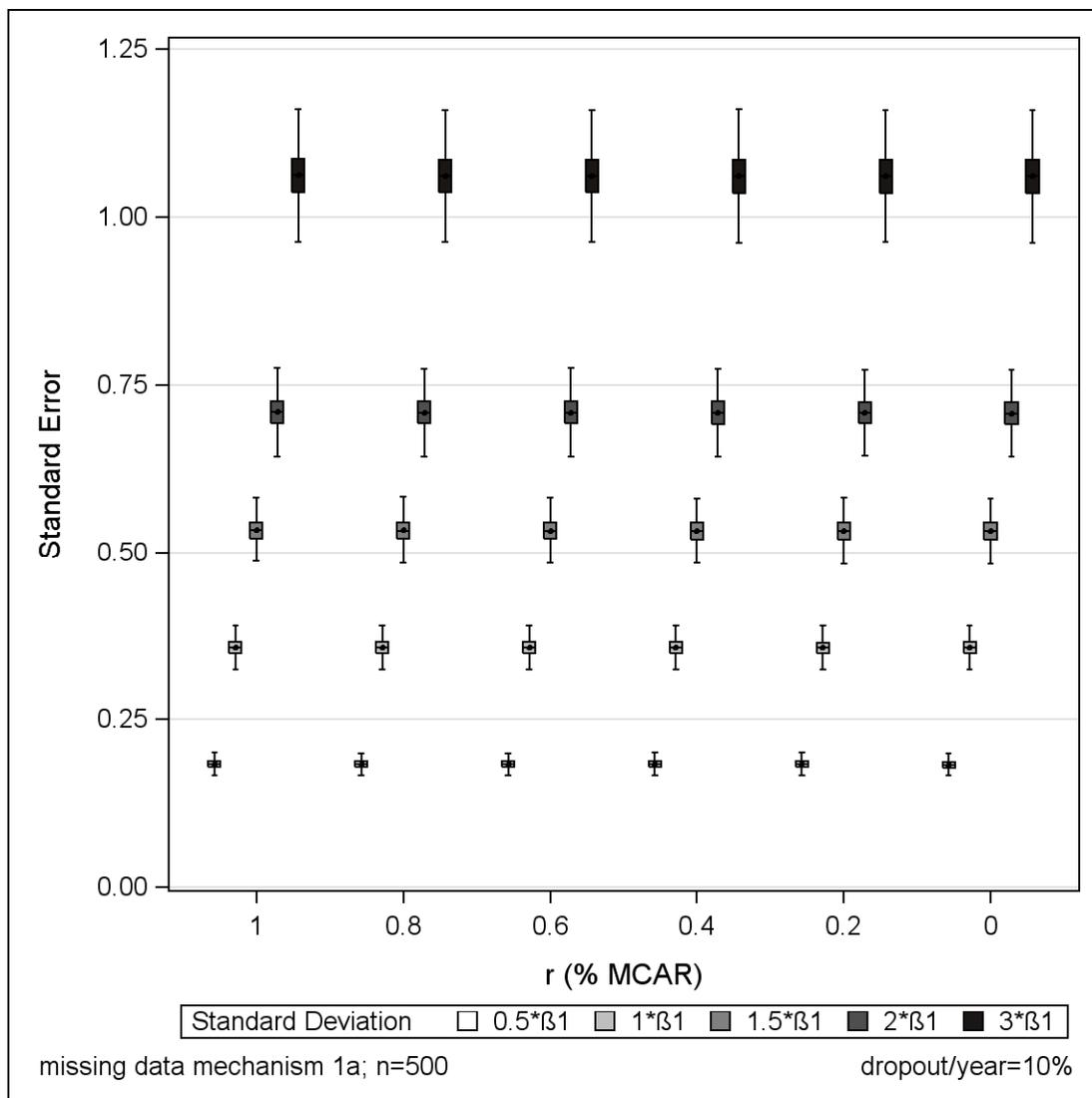


Figure 2.7 Standard error by sample size and % MCAR (mechanism 1a).

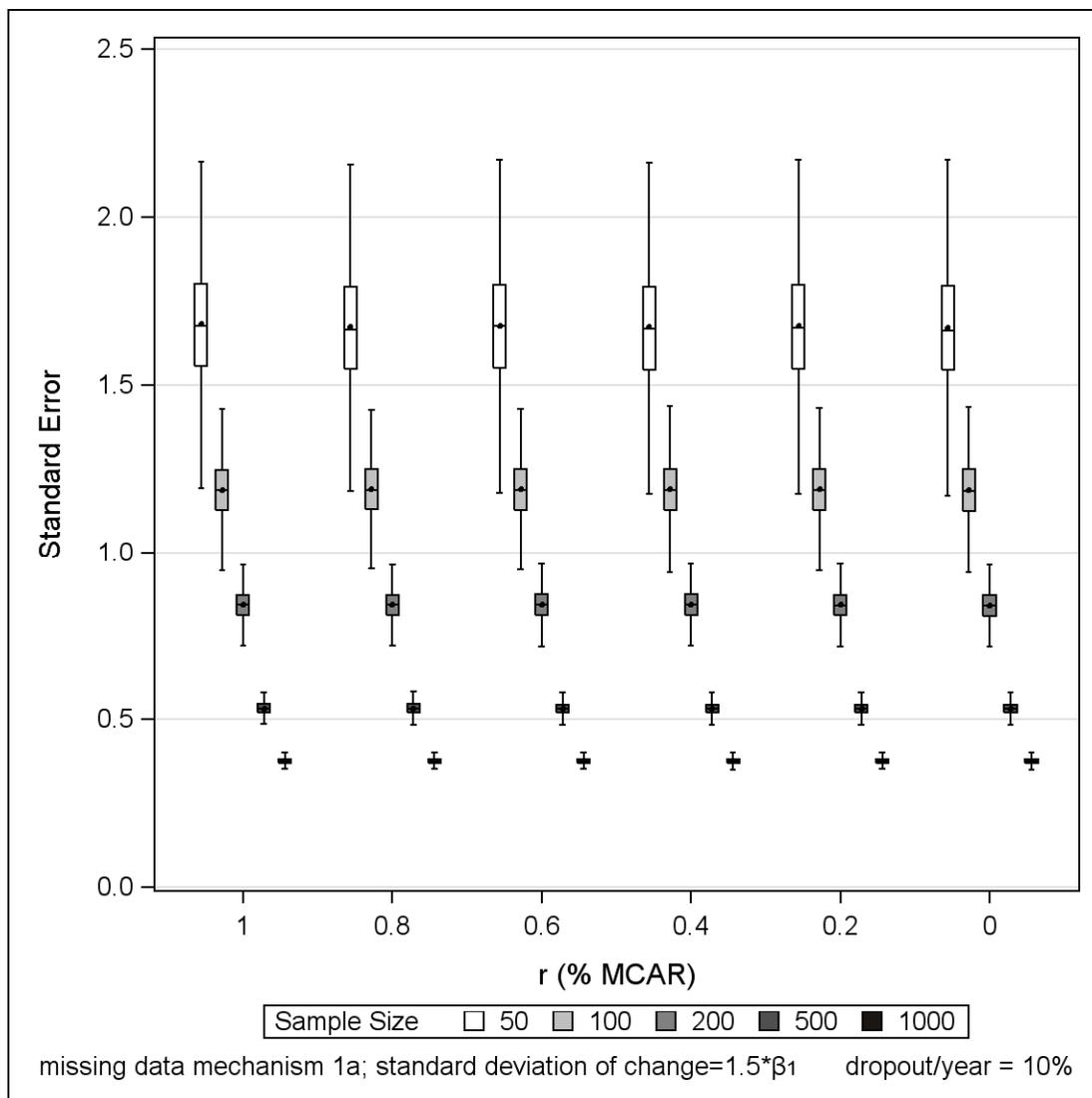
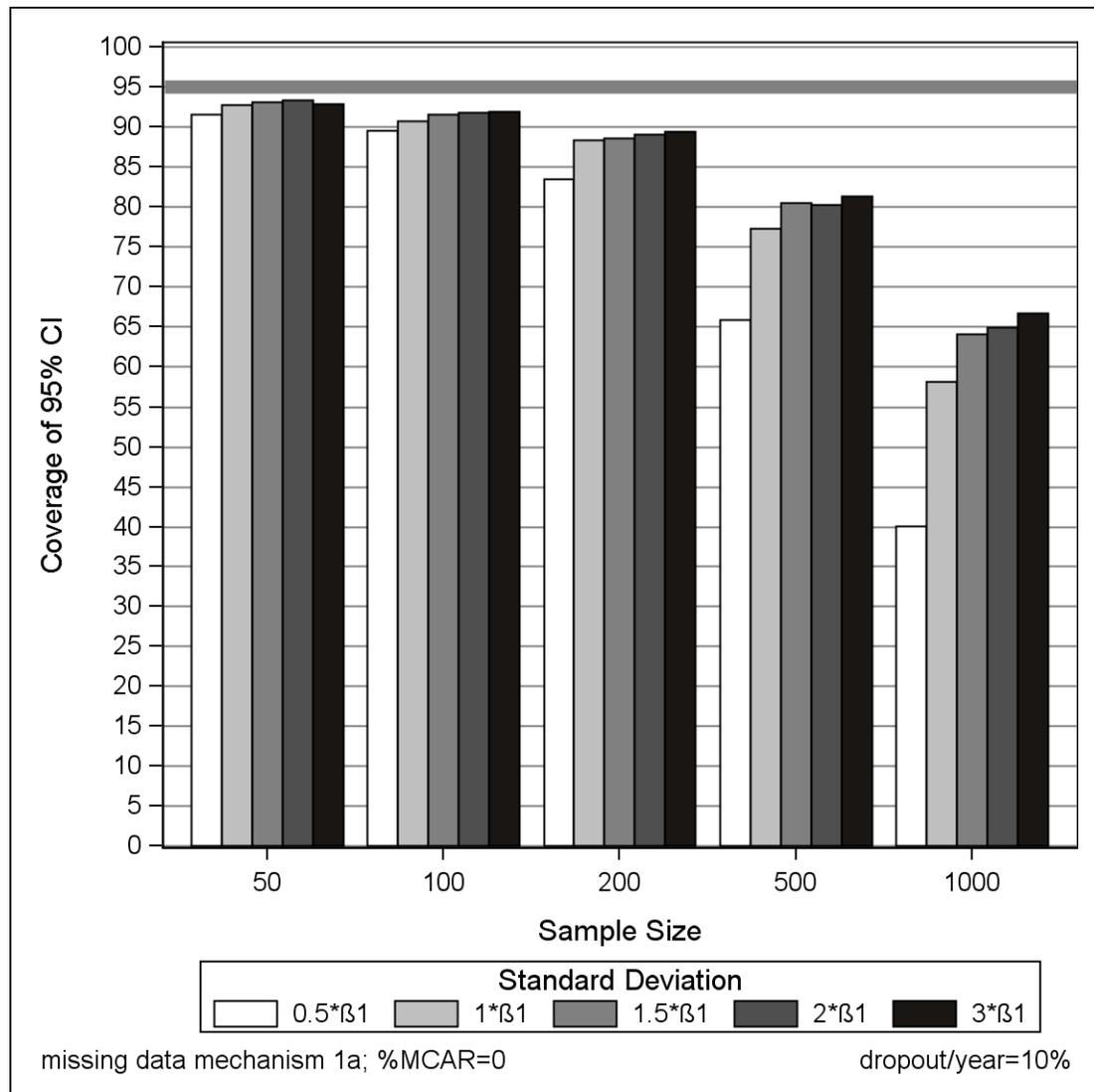


Figure 2.8 Coverage of 95% CI by standard deviation and sample size (mechanism 1a).



Shaded area represents the binomial margin of error based on the number of simulations

Figure 2.9 Estimate of yearly change with 95% CI by standard deviation and sample size (mechanism 1a).

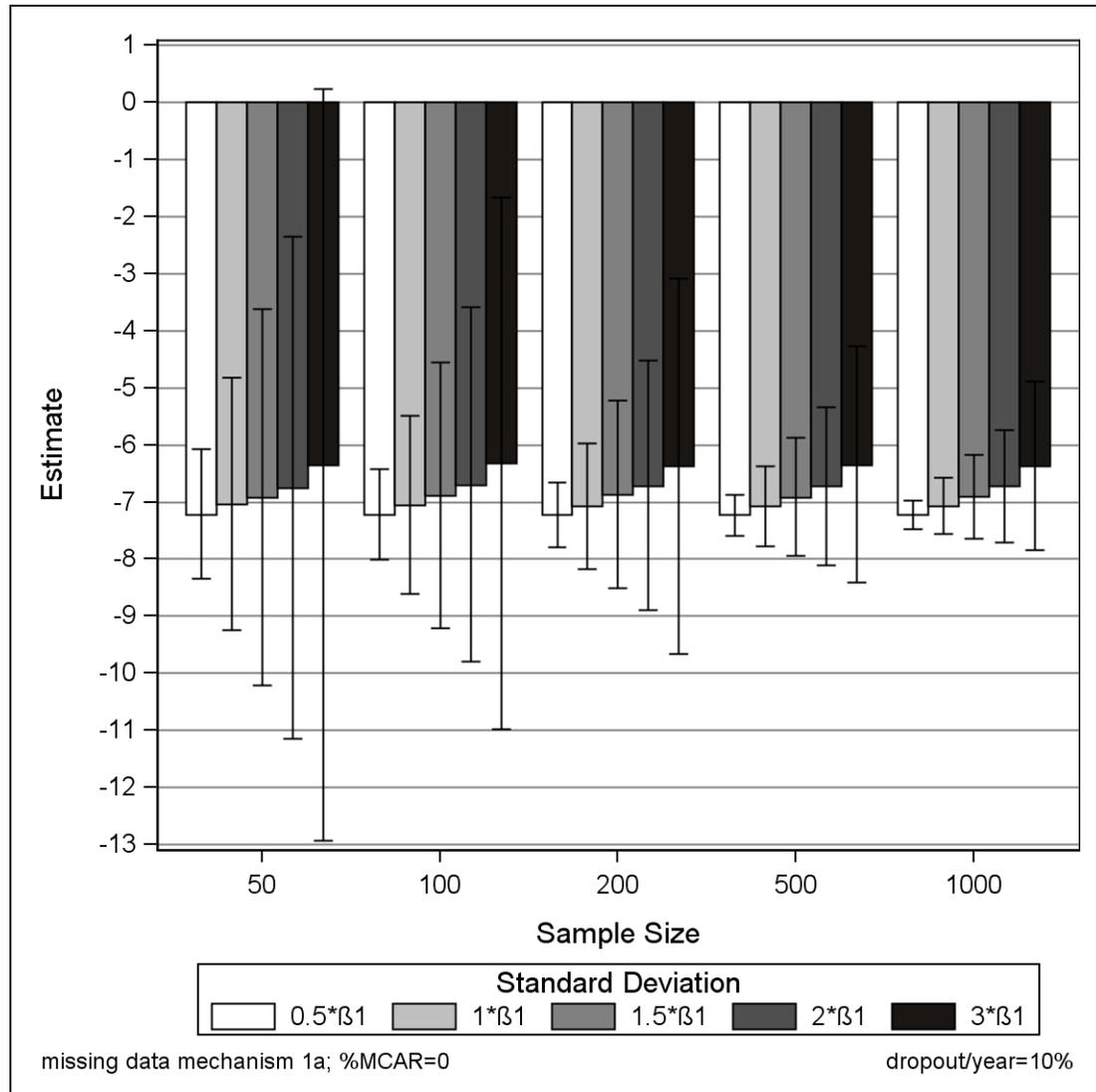


Figure 2.10 Relative bias by total dropout and mechanism.

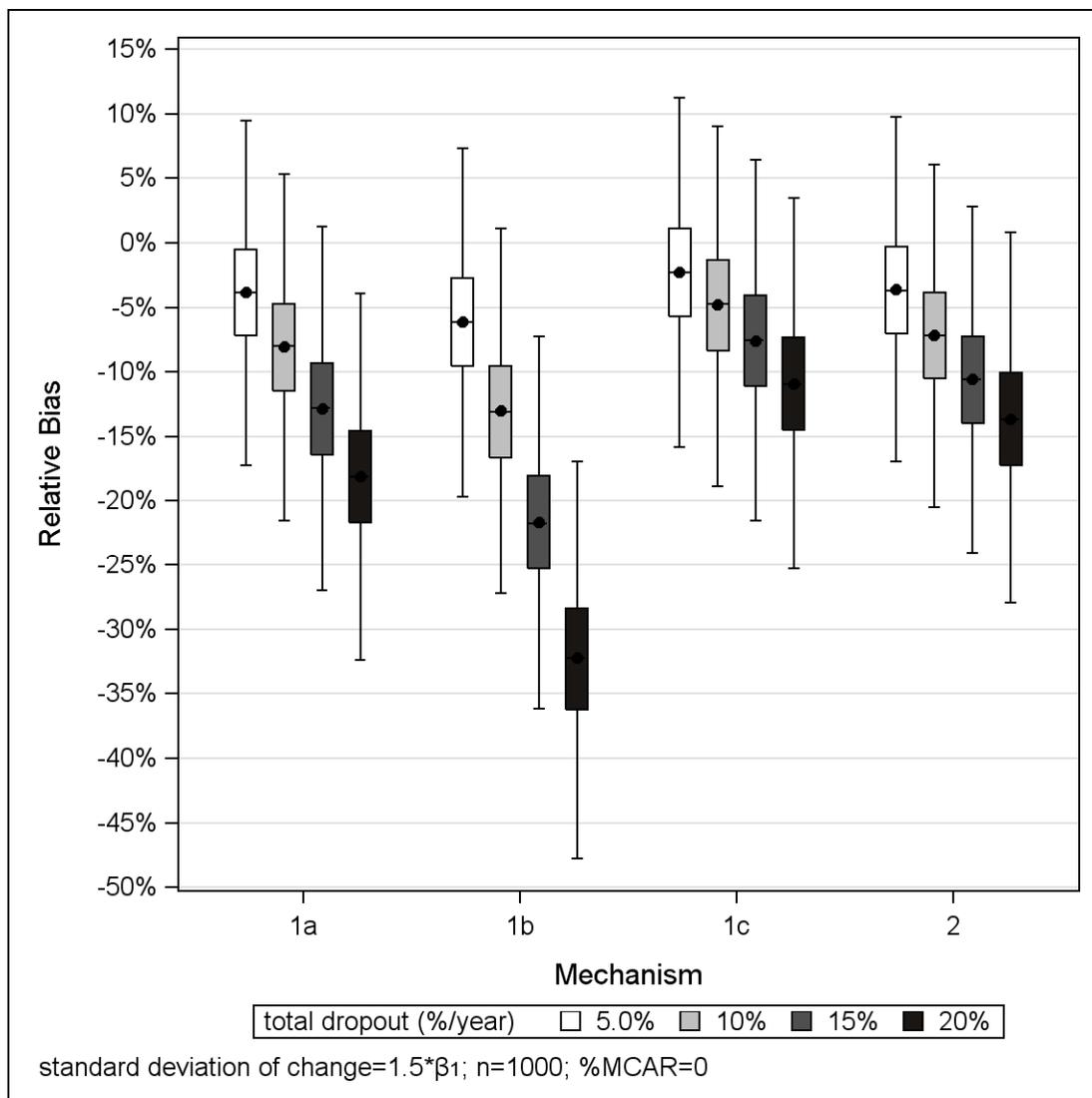


Figure 2.11 Standard error by total dropout and mechanism.

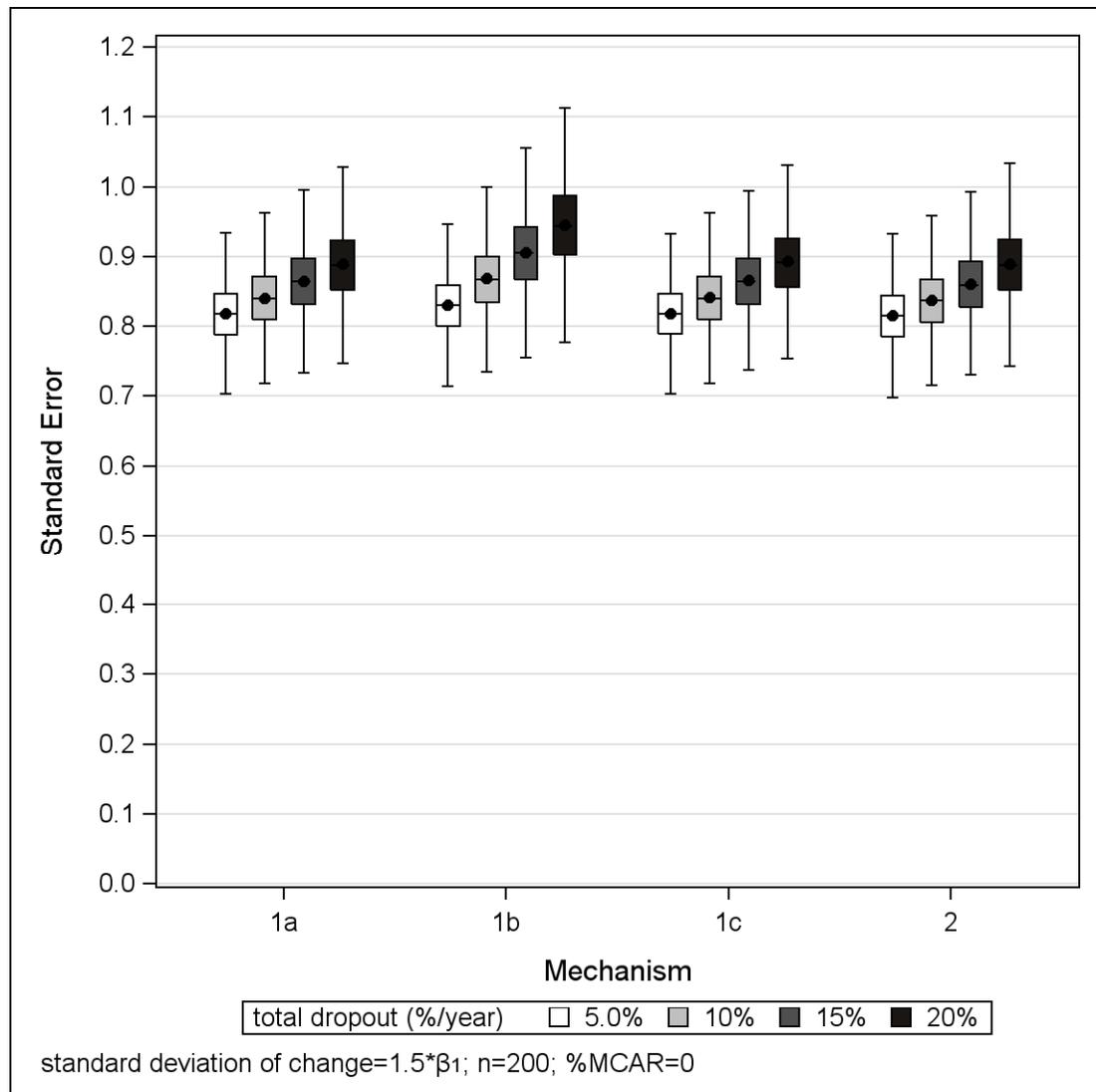
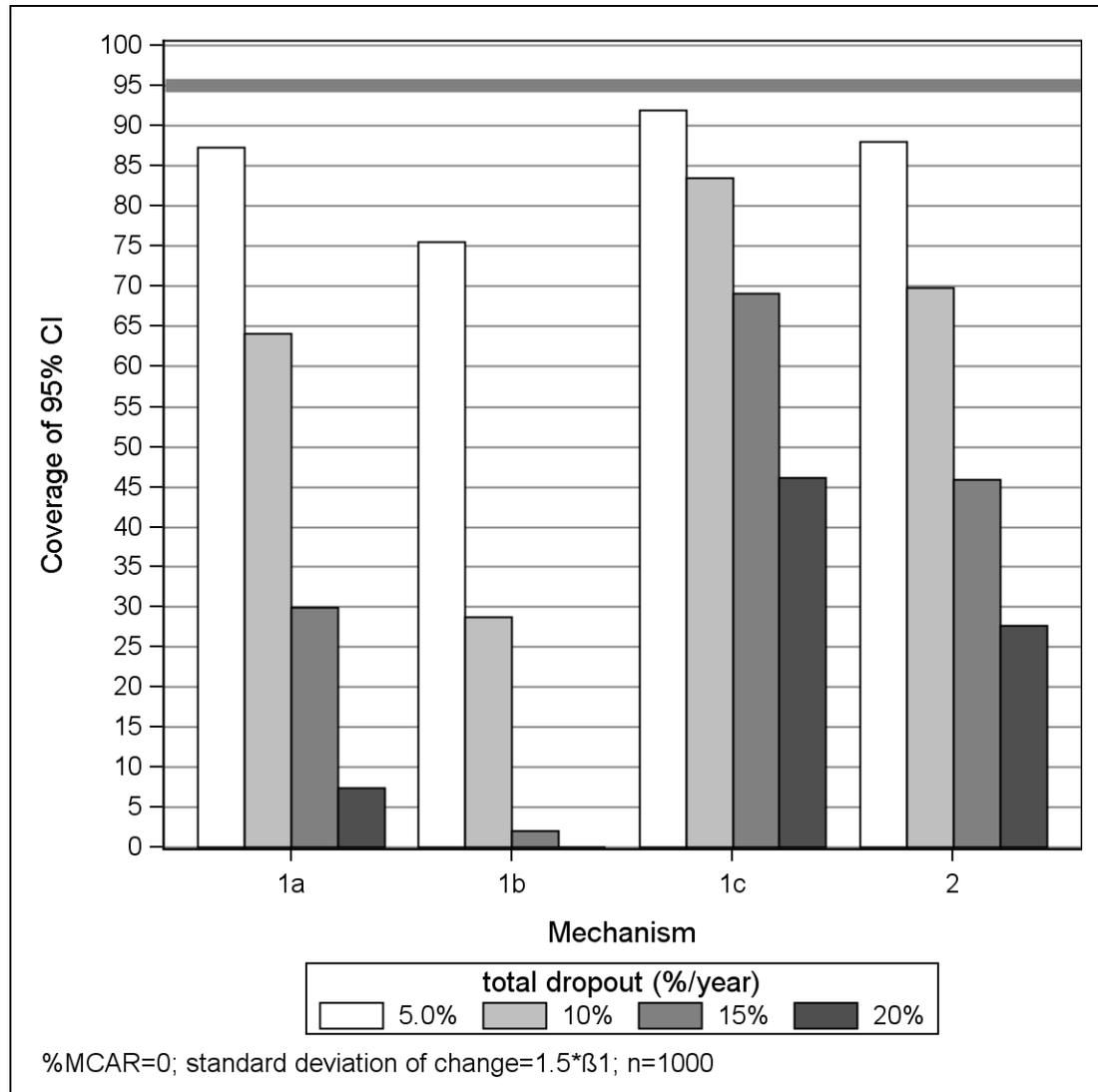
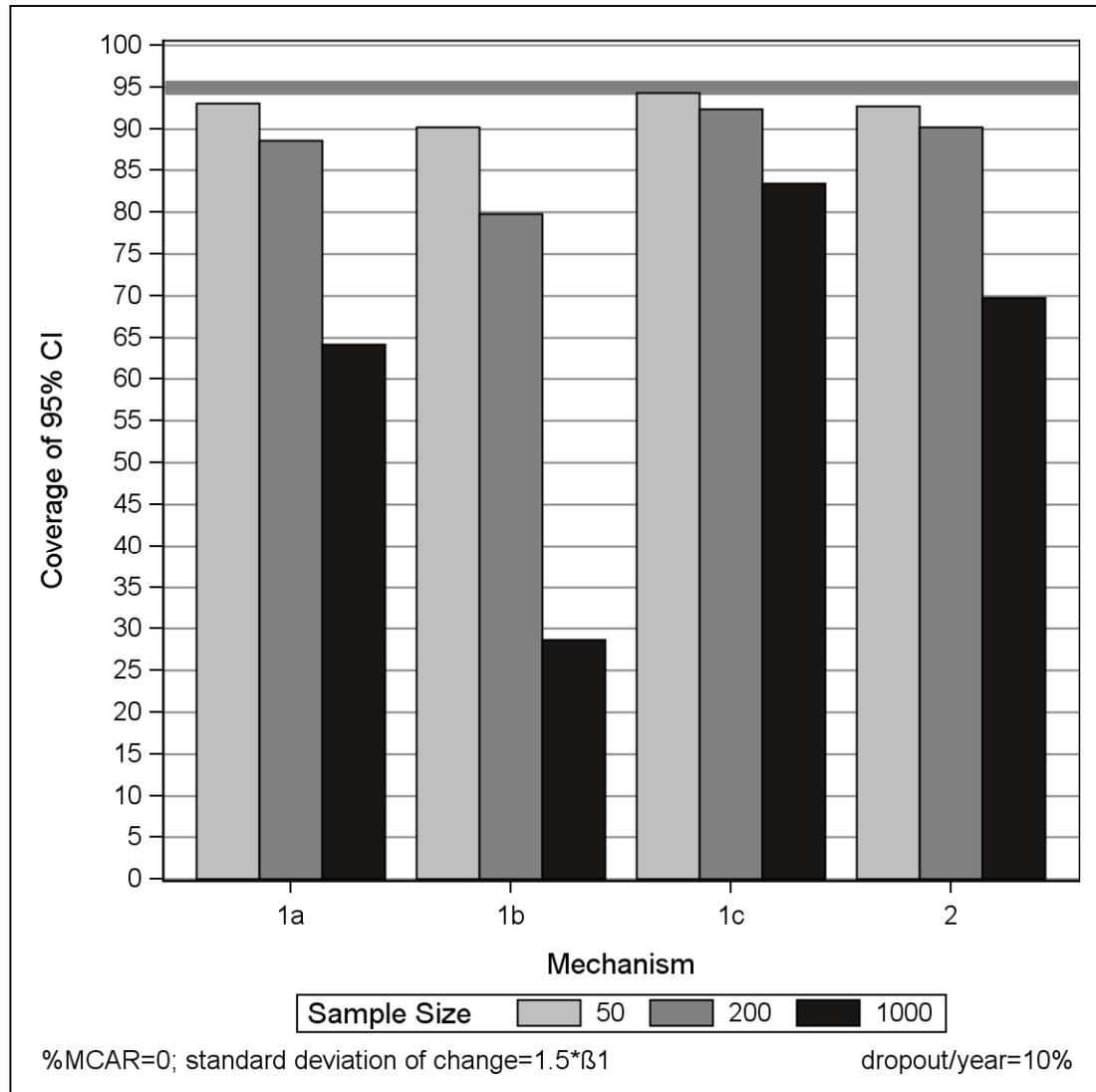


Figure 2.12 Coverage of 95% CI by total dropout/year and mechanism.



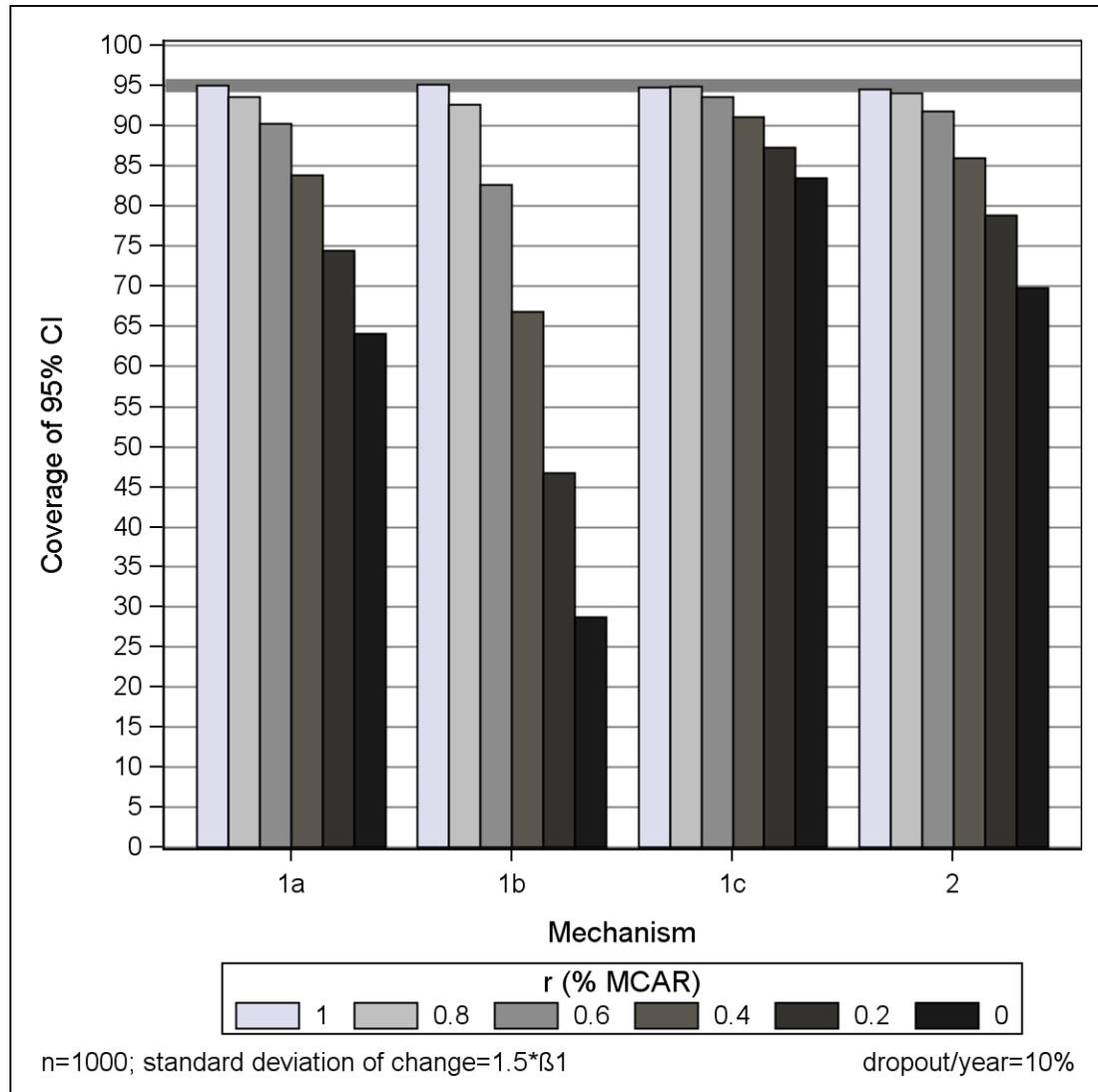
Shaded area represents the binomial margin of error based on the number of simulations

Figure 2.13 Coverage of 95% CI by sample size and mechanism.



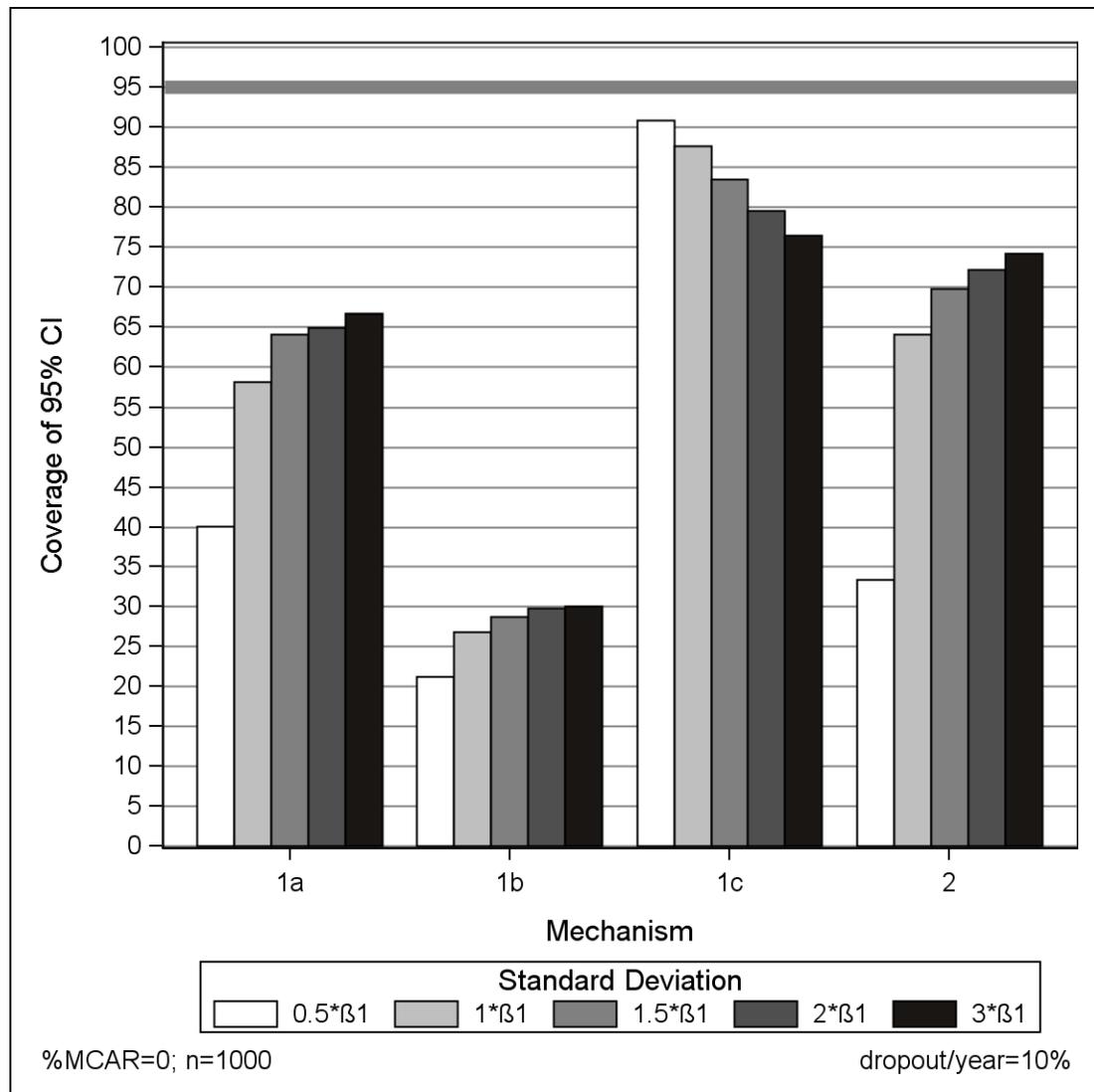
Shaded area represents the binomial margin of error based on the number of simulations

Figure 2.14 Coverage of 95% CI by nature of dropout and mechanism.



Shaded area represents the binomial margin of error based on the number of simulations

Figure 2.15 Coverage of 95% CI by standard deviation and mechanism.



Shaded area represents the binomial margin of error based on the number of simulations

Figure 2.16 Estimate of yearly change with 95% CI by standard deviation and mechanism.

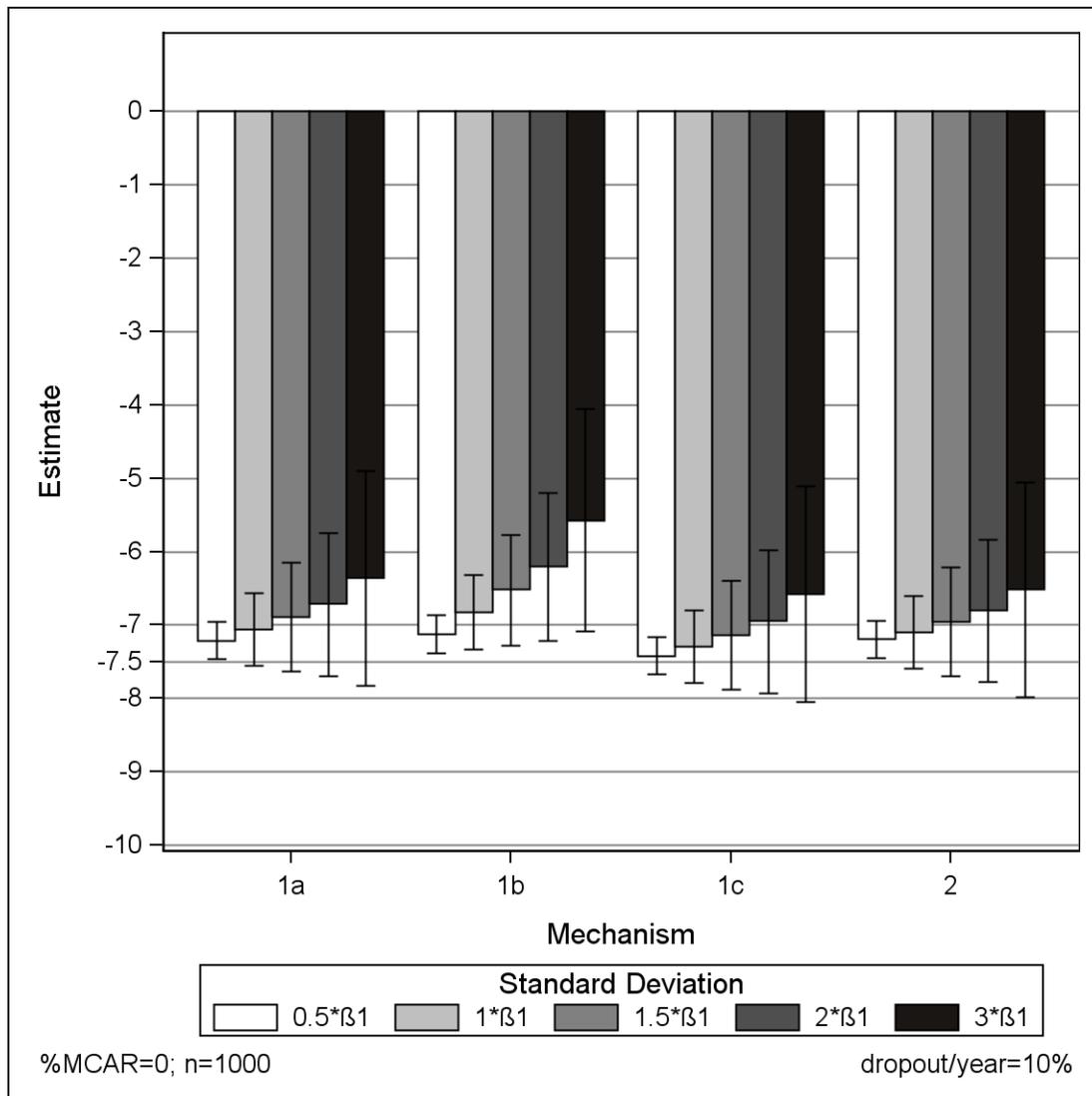
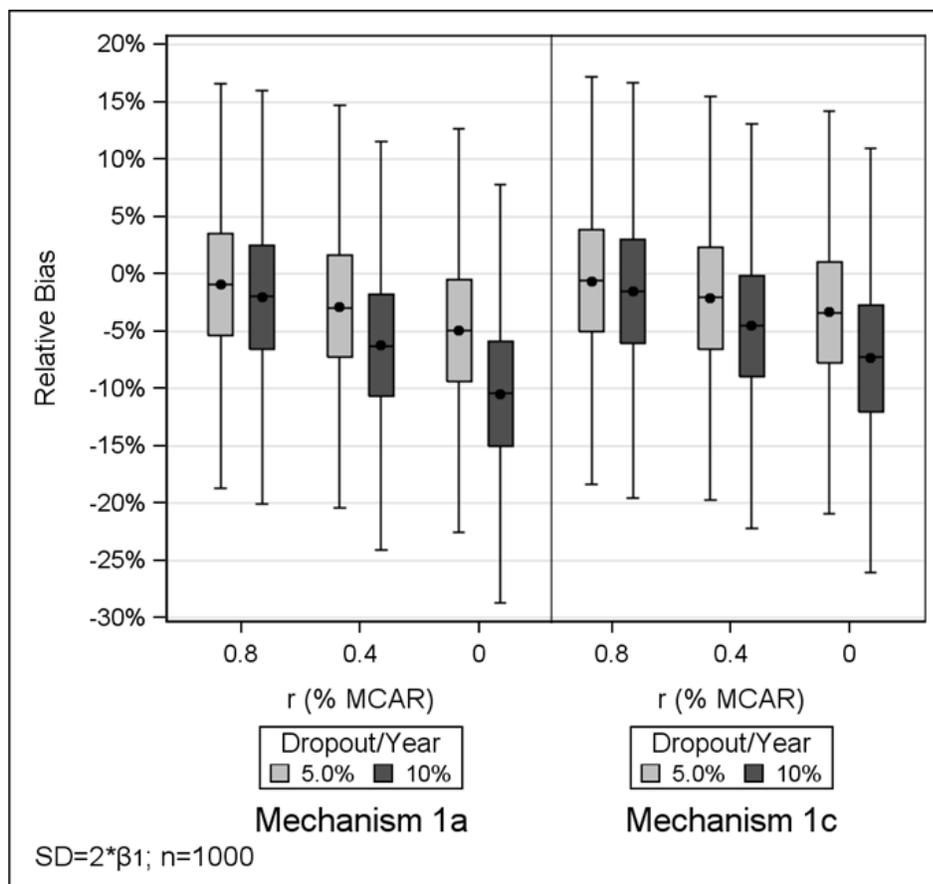


Figure 2.17 Relative bias for scenarios plausible in the OAI.



CHAPTER 3: Estimation of Rate of Change in Longitudinal Studies with Informative Dropout and an Imprecise Anchor Event

3.1. Introduction

Longitudinal studies involve a series of measurements over time on the same individual or observational unit, allowing for the direct study of change over time. Over the course of a longitudinal study it is common for subjects to miss visits or to drop out before the scheduled end of follow-up.

The effect of missing data on the estimation of the rate of change depends on the underlying missing data mechanism. When data are missing not at random (MNAR), standard statistical models for longitudinal data that fail to take into account the association between dropout patterns and the underlying rate of change can lead to biased and inaccurate estimators. [2, 3] In previous work we found that linear mixed models that do not incorporate the dropout mechanism performed adequately when the overall amount of dropout was small (<20% overall) and the variability in the outcome was low to moderate (≤ 1.5 times the rate of change). Under these scenarios, relative bias was modest (<5%) and coverage of the 95% confidence interval was greater than 80%. For scenarios with moderate to large amounts of missing data or with highly variable outcomes, even small amounts of MNAR dropout can lead to substantial bias in the parameter estimates from a linear mixed effects model, as demonstrated in Chapter 2. Under conditions where linear mixed models do not perform well, it is

necessary to incorporate more sophisticated methods to account for missing data to improve the accuracy and precision of the estimate.

3.1.1 Defining Anchor Event

There has been a good deal of work to develop models for longitudinal data when MNAR dropout is suspected. These include pattern mixture, selection, and frailty models, among others. [1, 3, 14, 35, 36] These methods attempt to improve estimates of change over time by incorporating the informative dropout process. This can be done by explicitly modeling the informative dropout process and incorporating this into the model for longitudinal change (e.g., selection models) or by stratifying based on pattern of informative dropout (e.g., pattern mixture). While no universally superior method has been found for all scenarios where MNAR is suspected, these methods generally have been shown improve the estimate of change over time in longitudinal data, in terms of bias and efficiency, when model assumptions are met. [13]

In a longitudinal study it is not uncommon for multiple types of dropout to be present – both related to and unrelated to outcome. For example, in a study of renal disease, investigators classified patients dropping out due to death, kidney transplant, and initiation of dialysis as potentially related to unobserved outcome (informative), while dropouts due to other reasons (e.g., moving away, no longer interested) were classified as not related to study outcome (non-informative). [7]. In analyzing these data, the informative dropout process was incorporated by first

modeling time to informative dropout, and then jointly modeling the longitudinal data with this time-to-event variable. Non-informative dropouts were not included in the time-to-event analysis as this was solely to model the *informative* dropout process.

We define *Anchor Event* as an event or dropout reason that informs the researcher as to whether or not the dropout was related to study outcome. In the example above, death, kidney transplant, and initiation of dialysis were the anchor events. Dropouts not experiencing these anchor events were not considered informative, and were not included in the time-to-event analysis for the informative dropout process. Death is often used as an anchor event; for example, studies of advanced colorectal cancer and end-stage heart failure both looked at quality of life over time and both studies modeled deaths separately from other dropouts, since death is clearly associated with quality of life. [37, 38] In these examples – renal disease, advanced colorectal cancer, and end-stage heart failure –there is a clear relationship between anchor event and outcome. Investigators could be confident that patients initiating dialysis, for example, had progressed to end-stage renal disease. In fact, there is a specific threshold of kidney function (glomerular filtration rate of 7 to 8 ml/min) where either dialysis or kidney transplant is medically required. [7] Many methods of modeling informative dropout have been developed in diseases where a clear anchor event exists.

A complication arises when the anchor event is imprecise, that is, the anchor does not perfectly separate the informative and non-informative dropouts. In knee osteoarthritis (OA), total knee replacement (TKR) is a treatment that has historically been offered to patients with end stage disease. If this held true, we could be confident that patients dropping out of a longitudinal study of knee OA to undergo TKR had progressed to end stage OA, if not already there by the last study visit. The TKR would inform us about subjects' disease progression since the last study visit. This is especially important in knee OA since it is a slowly progressing disease and study visits in longitudinal studies of OA are often spaced a year or more apart. [27, 31] However, there are concerns about both the sensitivity and specificity of TKR as an anchor event: TKR is an elective surgery, and not everyone with end stage disease is willing or able to undergo TKR; also, indications for TKR appear to be changing and some subjects with mild to moderate disease that is not end stage undergo the procedure. [39] Therefore indications and acceptance of TKR for end stage knee OA may relax the conditions critical for definition of an anchor event. For this reason, TKR is likely to be an imperfect anchor. The question of whether the use of an imperfect anchor leads to improved model performance had not been studied. If we incorporate TKR into the models of disease progression in knee osteoarthritis, even though we know that it's not a perfect marker of progression, will this improve the estimates? Could it actually make them worse?

3.1.2 Objectives

This project has two objectives. First, we will propose an update for the pattern mixture modeling (PMM) paradigm to better handle uncertainty in the anchor event. Then we will evaluate the effect of an imprecise anchor event in evaluating disease progression in longitudinal studies with informative dropouts with PMMs.

3.1.3 Motivating Example: Estimating Change in Joint Space Width in the Osteoarthritis Initiative

Our clinical motivation is to investigate the utility of incorporating TKR as a method of adjusting for informative dropout. TKR is an elective surgery and multiple variables affect a subject's decision to undergo the operation. One study of knee and hip replacement showed that willingness to consider surgery was a key predictor of knee replacement in two years of follow-up. [40] Another longitudinal study of knee OA found that only 10% of subjects (22 of 227) with end stage radiographic knee OA underwent TKR during 2 year follow-up. [41] TKR captures both clinical necessity and subject preference and may be an imperfect measure of disease progression. [39]

The Osteoarthritis Initiative (OAI) is a multi-center, longitudinal, prospective observational study of knee OA. The goal of the OAI is to establish and maintain a natural history database for knee OA that will include clinical evaluation data, radiological images and image assessments, and a biospecimen repository.

Subjects were assessed at baseline and then at yearly follow-up visits through 6 years, with a planned extension to follow patients through 8 years. We investigated the relationship between TKR and OA progression in OAI patients with radiographic knee OA. We assessed structural OA progression using joint space width (JSW). Radiography is used to quantify disease progression by measuring the narrowing of the joint space width (JSW) between the adjacent bones of the knee. Twenty-nine percent of the cohort dropped out of the study. Approximately one-third of these dropouts had TKR, and two-thirds did not. Figure 3.1 displays JSW over time by TKR status. Each line is the average JSW (with standard error bars) for a particular TKR group. Participants with early TKRs tended to start with less JSW at baseline, and on average participants with TKRs tended to experience a greater loss of JSW over time compared to patients not undergoing TKR. When we further split the groups by anchor event, we see that the trajectory of non-TKR dropouts is much more similar to the completers than the TKRs (Figure 3.2).

Figure 3.1 Mean joint space width with standard error by total knee replacement status and visit

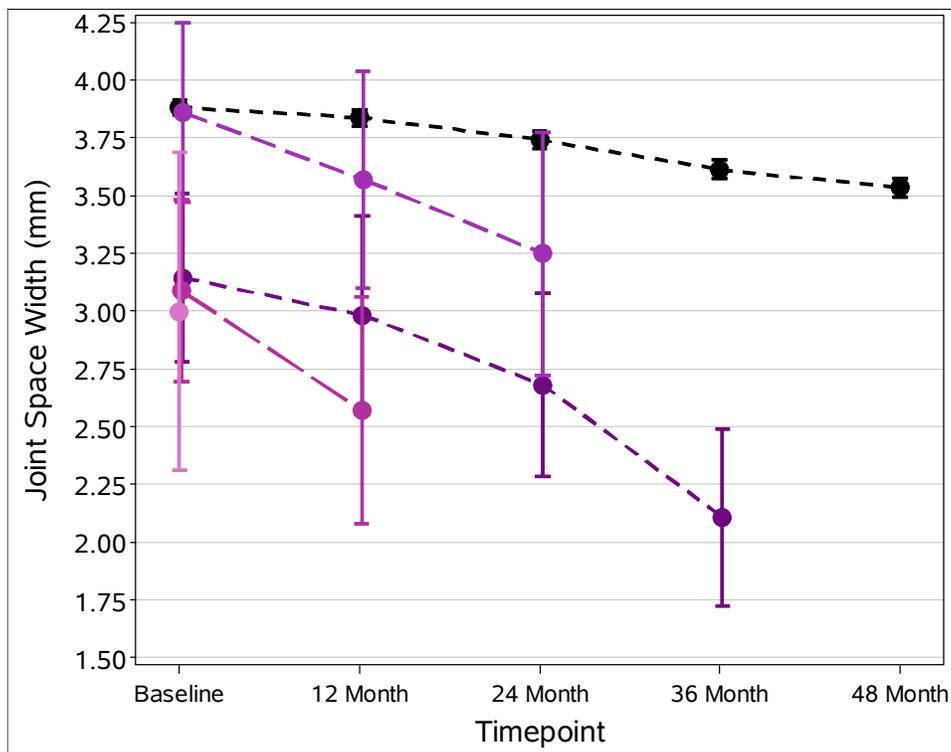
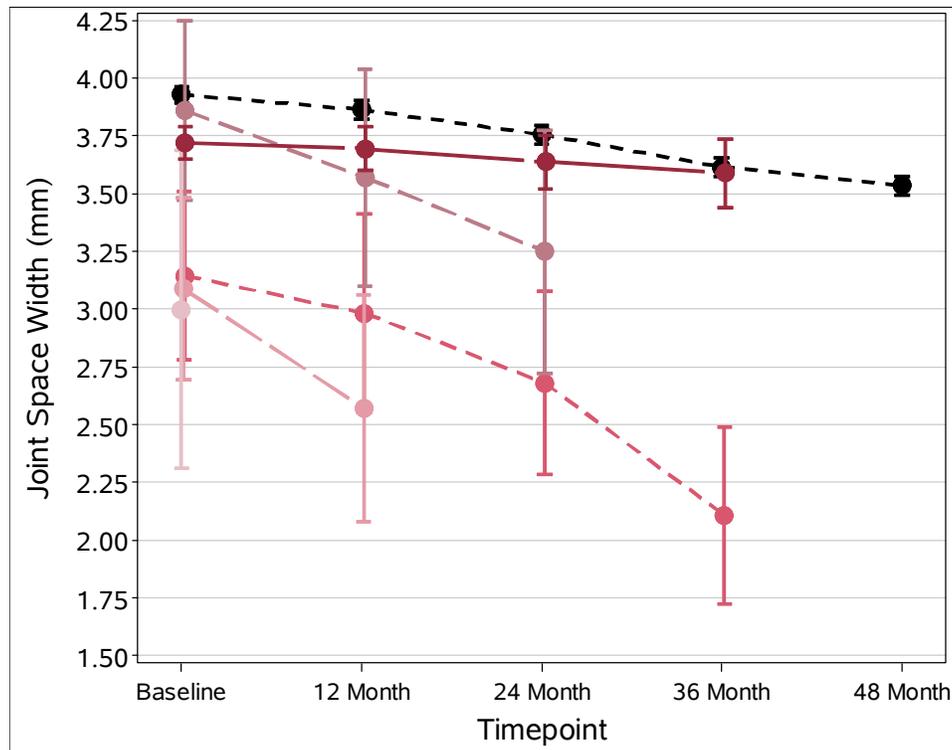


Figure 3.2 Mean joint space width with standard error by total knee replacement and completion status and visit



This suggests that there is a heterogeneous dropout mechanism: some patients are dropping out of the study to undergo TKR, and those dropouts appear to be associated with disease progression, while others are dropping out for other reasons and the relationship with disease progression is unclear. One option is to use a pattern mixture modeling approach and count only the TKRs as informative dropouts (i.e., use TKR as the anchor event). However, we have established that TKR is an imperfect marker of disease progression. Subjects that dropout and do not undergo TKR could also experience disease progression. In this case, we'd like to know if using TKR as an anchor event to

define our dropout groups improves or worsens our estimates of disease progression.

3.2. Methods

3.2.1 Overview

The goal of this study is to investigate the utility of incorporating an imprecise anchor event in PMMs to investigate disease progression over time in longitudinal studies. We propose a modification to the pattern mixture modeling approach to better account for imprecision in the anchor event. Using a simulation study we will compare the relative bias, standard error, and coverage of three different ways of incorporating the anchor event in the pattern mixture modeling framework. We will evaluate the performance of the models under a wide range of scenarios including variations in the percent of dropout, the precision of the anchor event, and variability of the outcome measure.

3.2.2 Notation

Let Y_i be a set of longitudinal measurements on subject i ($i=1,2,\dots,N$) and let Y_{ij} be the measurement taken at the j th time point ($j=1,2,\dots,T$). Let R_i be the vector of observed data indicators, where $R_{ij} = 1$ if Y_{ij} is observed and $R_{ij} = 0$ otherwise. We can partition Y_i into two components: let Y_i^o be the matrix of observed data, Y_i^m be the matrix of missing data. X_i is the set of observed covariates. Let D_i be the vector of dropout indicators, where $D_{ij} = 1$ if subject i drops out of the study

between time j and time $j+1$. Finally, let A_i be the vector of anchor indicators, where $A_{ij} = 1$ if subject i drops out of the study between time j and time $j+1$ and has the anchor event. Note that D_{iT} and A_{iT} are always equal to zero, since a subject cannot dropout of the study after study completion.

3.2.3 Pattern Mixture Models

Pattern mixture models stratify the study population based on the pattern of dropout and separately model each group. The overall estimate is the weighted average of the group-specific estimates, with the weights equal to the proportion of subjects in each group. [11] The joint distribution of the outcome and missingness is factored as:

$$f(Y_i^o, Y_i^m, R_i | X_i) = f(Y_i^o, Y_i^m | R_i, X_i) f(R_i | X_i) \quad (3.1)$$

The distribution of the responses is conditional on the missing data pattern. This implies a unique distribution for each dropout group. If we assume a normal distribution for the outcome Y_i :

$$Y_i | R_i, X_i \sim N(\mu(R_i), \Sigma(R_i)) \quad (3.2)$$

The parameter of interest $\hat{\beta}$, is obtained by averaging over the P missing data patterns:

$$\hat{\beta} = \sum^P \hat{\pi}^{\{p\}} \hat{\beta}^{\{p\}}$$

where $\pi^{\{p\}}$ is the proportion of subjects and $\beta^{\{p\}}$ is the conditional parameter estimate in the p^{th} pattern.

In a study with T assessments there are 2^T possible patterns, so subjects are often combined into groups based on time of last assessment. This strategy results in T groups: a group for each possible dropout time and a group of completers. [24]

3.2.3.1 Strategies for Defining Patterns

3.2.3.1.1 Pattern Mixture Model 1 (PMM1)

Subjects are grouped together based on time of dropout, with no accommodation for dropout reason. Completers are grouped together in a final pattern. Again assuming a normal distribution for the outcome Y_i , we have a different distribution for each dropout group:

$$Y_i | D_i, X_i \sim N(\mu(D_i), \Sigma(D_i)) \quad (3.3)$$

We will refer to this strategy as Pattern Mixture Model 1 (PMM1). This strategy is appropriate when all dropouts are assumed to be not at random, for example a study that follows all subjects until disease progression or death. This approach is often used in end stages of terminal diseases where it is unlikely that a subject would simply move away or lose interest in participating in the study. [42]

3.2.3.1.2 Pattern Mixture Model 2 (PMM2)

The second approach uses an anchor event (e.g., death, relapse, hospitalization) to determine which dropouts are informative and assumes that all other dropouts are MAR or MCAR. The patterns are defined based on time of

dropout for those subjects with an anchor event; completers and MAR/MCAR dropouts are grouped together in a final pattern. The outcome Y_i is normally distributed conditional on an anchor event time and covariates.

$$Y_i|A_i, X_i \sim N(\mu(A_i), \Sigma(A_i)) \quad (3.4)$$

We will refer to this approach as Pattern Mixture Model 2 (PMM2). This strategy is appropriate when there is a mixture of informative and non-informative dropouts and the anchor event is unambiguous.

The key assumption of PMMs is that, conditional on dropout pattern, missing data are MAR. The estimates for each pattern will be valid, assuming that the joint distribution is correctly specified. If the MAR assumption is satisfied, then within each pattern, the future values for a subject that drops out can be predicted from that subject's past history and the specified joint distribution of outcome. [5, 6] Two subjects with the same pattern of observed data will have the same predicted values:

$$E[Y_i^m | Y_i^o] = X_i^m \beta + \Sigma^{m,o} (\Sigma^{o,o})^{-1} (Y_i^o - X_i^o \beta)$$

If informative dropouts are not clearly defined, and some patterns have a mixture of informative and non-informative dropouts, then for both PMM1 and PMM2, the assumption that the missing data are MAR given the dropout pattern is not met. For example, under PMM1, all dropouts are grouped together. If the dropout mechanism is heterogeneous, that is, a mixture of non-informative and informative dropouts, then the MAR assumption within each pattern will not hold. Similarly, for PMM2, if the anchor event does not delineate informative and non-

informative dropouts, then these different types of dropout will be grouped together in the same pattern and the MAR assumption again will not hold. For example, in the OA example subjects that drop out of a study due to disease progression may not undergo TKR. These subjects would be an informative dropout, but would not have the anchor event. Under PMM2 these subjects would be put in the group $A_i = T$. While we continue to model this group based on the assumption in equation 3.1 above, the actual distribution is a mixture of two groups: informative dropouts without anchor and everyone else without anchor

$$(Y_i | A_i = T, X_i) \sim N \begin{pmatrix} \mu(A_i = T0) & \Sigma(A_i = T0) \\ \mu(A_i = T1) & \Sigma(A_i = T1) \end{pmatrix}$$

Due to the model's inability to partition the informative and non-informative dropouts in this group the estimate $\hat{\beta}^{\{T\}}$ will be biased.

3.2.3.1.3 Pattern Mixture Model 3 (PMM3)

We propose a third strategy for grouping dropouts. If the anchor event is imprecise it is possible that some MNAR dropouts will not have the anchor event. For example, a subject in an OA study could experience disease worsening resulting in increased functional impairment and drop out of the study due to difficulty traveling to study appointments, but he or she may not be willing to undergo TKR. In this strategy, dropouts with an anchor event are grouped based on the time of dropout. Another group contains all other non-anchor dropouts, and a final group contains all completers. This ensures that there is at least one

pattern with an unbiased estimate of $\hat{\beta}^{(p)}$, the completers. The bias of the estimates for the other patterns will depend on the precision of the anchor event, as it does for the PMM2 approach.

Let dropout group for subject i be denoted by D_i as in (3.3) and anchor event be denoted by A_i as in (3.4). Equation 3.1 above can be re-written

$$f(Y_i^o, Y_i^m, R_i | X_i) = f(Y_i^o, Y_i^m, A_i, D_i | X_i) = f(Y_i^o, Y_i^m | A_i, D_i, X_i) f(A_i, D_i | X_i)$$

Conditional on both dropout group and anchor event, missing data are assumed to be MAR.

3.2.4 Simulation Study Details

3.2.4.1 Complete Data Generating Mechanism

For each subject a continuous response variable was evaluated at baseline and at four fixed follow-up time points. A decline in response indicates clinical worsening. For each subject we generated a vector of correlated responses, $Y_i = (Y_{i0}, Y_{i1}, Y_{i2}, Y_{i3}, Y_{i4})$ under the linear mixed effect model:

$$Y_i = X_i \beta + Z_i b_i + e_i$$

where

$$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}, b_i = \begin{bmatrix} b_{0i} \\ b_{1i} \end{bmatrix}, X_i = \begin{bmatrix} t_1 \\ t_2 \\ \dots \\ t_n \end{bmatrix}, Z_i = \begin{bmatrix} t_1 \\ t_2 \\ \dots \\ t_n \end{bmatrix}$$

and

$$Y_{ij} = \beta_0 + \beta_1 t_{ij} + b_{0i} + b_{1i} t_{ij} + e_{ij}$$

The fixed effects (population average) for intercept and slope were

$$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} 100 \\ -7.5 \end{bmatrix}$$

Variability was added with random effects for intercept and slope

$$b_i = \begin{bmatrix} b_{0i} \\ b_{1i} \end{bmatrix} \sim N\left(0, \begin{bmatrix} \sigma_0^2 & 0.25 * \sigma_0 \sigma_1 \\ 0.25 * \sigma_0 \sigma_1 & \sigma_1^2 \end{bmatrix}\right)$$

The standard deviation of the random effect for slope, σ_1 was varied at 0.25, 0.5, 1, 1.5, and 2 times the slope. The standard deviation of the random effect for intercept, σ_0^2 , was 15.

Finally, residual variance was generated with an AR(1) correlation structure with correlation $\rho=0.75$ and σ_e varied at 0.25, 0.5, 1, 1.5, and 2 times the slope:

$$e_{ij} = \begin{bmatrix} e_{i0} \\ e_{i1} \\ e_{i2} \\ e_{i3} \\ e_{i4} \end{bmatrix} \sim N \begin{bmatrix} \sigma_e^2 & \rho\sigma_e^2 & \rho^2\sigma_e^2 & \rho^3\sigma_e^2 & \rho^4\sigma_e^2 \\ \rho\sigma_e^2 & \sigma_e^2 & \rho\sigma_e^2 & \rho^2\sigma_e^2 & \rho^3\sigma_e^2 \\ \rho^2\sigma_e^2 & \rho\sigma_e^2 & \sigma_e^2 & \rho\sigma_e^2 & \rho^2\sigma_e^2 \\ \rho^3\sigma_e^2 & \rho^2\sigma_e^2 & \rho\sigma_e^2 & \sigma_e^2 & \rho\sigma_e^2 \\ \rho^4\sigma_e^2 & \rho^3\sigma_e^2 & \rho^2\sigma_e^2 & \rho\sigma_e^2 & \sigma_e^2 \end{bmatrix}$$

3.2.4.2 Missing Data Mechanism

We created a random-effects dependent heterogeneous missing data mechanism, with dropout the result of MNAR, MAR, and MCAR mechanisms. Approximately 60% of the dropouts were not at random, i.e., related to the underlying random effect for slope. First, for each subject we computed the z-score for the random effect for slope, zb_{i1} . This indicates how many standard deviations the subject is above or below the population slope. Then, for each subject we computed the log odds of dropping out as a function of the z-score of the random effect for slope, zb_{i1} :

$$\text{logit}(\pi_{it}) = \alpha_{0t} + \alpha_{1t}zb_{i1}$$

α_{0t} and α_{1t} were defined based on the total amount of dropout, which ranged from 20% to 80% of the cohort, and so that each increase in 1 standard deviation of slope (increase in 1 unit in zb_{i1}) was associated with an increased of odds of dropout of 2.5. 50% of the MNAR dropouts occurred after baseline, 15% after the first time point, 15% after the second time point, and 20% after the third time point.

Approximately 25% of the dropouts were at random, i.e., related to observed outcome, in this case, baseline value. To create a MAR mechanism, we first computed the z-score for the random effect for intercept, zb_{i1} . This z-score indicates how many standard deviations the subject is above or below the population baseline value. Then, for each subject log odds of dropping out was modeled as a function of the z-score of the random effect for intercept, zb_{i0} :

$$\text{logit}(\pi_{it}) = \alpha_{2t} + \alpha_{3t}zb_{i0}$$

α_{2t} and α_{3t} were defined based on the total amount of dropout, which ranged from 20% to 80% of the cohort, and so that each increase in 1 standard deviation of intercept (increase in 1 unit in zb_{i0}) was associated with an increased of odds of dropout of 1.5. 25% of the MAR dropouts occurred after baseline, 25% after the first time point, 25% after the second time point, and 25% after the third time point.

Finally, approximately 15% of the dropouts were MCAR, i.e., not related to outcome. The probability of dropping out was modeled as a Bernoulli random

variable with $p=0.15$. 25% of the MCAR dropouts occurred after baseline, 25% after the first time point, 25% after the second time point, and 25% after the third time point.

3.2.4.3 Anchor Event

A “perfect” anchor event perfectly separates MNAR and non-MNAR dropouts. For example, in the renal disease example presented above, death, receipt of a kidney transplant or initiation of dialysis were considered anchor events indicating informative dropout.

$P(\text{anchor} \mid \text{MNAR Dropout}) = 1$; Subjects with MNAR dropout must have anchor event. A subject would not dropout of the study to initiate dialysis unless he/she experienced significant disease worsening.

$P(\text{anchor} \mid \text{non-MNAR Dropout}) = 0$; Subjects without MNAR dropout cannot have anchor event. A subject that drops out of the study because he/she is too busy would not undergo kidney transplant.

We began with a “perfect” anchor and then deviated from this by allowing some non-MNAR dropouts to have the anchor and allowing some MNAR dropouts to not have an anchor, i.e.,

$P(\text{anchor} \mid \text{MNAR Dropout}) < 1$; Subjects with MNAR dropout may not have anchor event. A subject may drop out of a study of knee OA progression due to increased pain and functional impairment, but may not be willing to undergo TKR.

$P(\text{anchor} \mid \text{non-MNAR Dropout}) > 0$; Subjects without MNAR dropout cannot have anchor event. A patient may drop out of a study of knee OA progression without significant worsening of the disease, but may opt to undergo TKR.

We generated the probabilities guiding the anchor events in two different ways in order to examine the properties of an event that make it a useful anchor :

1. Anchor1: For MNAR Dropouts we modeled the log odds of having an anchor event as a function of the z-score of the random effect for slope, zb_{i1} :

$$\text{logit}(\pi_{it}) = \gamma_{0t} + \gamma_{1t}zb_{i1}$$

γ_{0t} and γ_{1t} were defined based on the total percent of MNAR dropouts with anchor, which ranged from 25% to 100% of the cohort, and so that each increase in 1 standard deviation of slope (increase in 1 unit in zb_{i1}) was associated with an increased of odds of anchor of 2.0. For MAR dropouts the probability of having an anchor was modeled as a Bernoulli random variable with $\pi_{it} = \lambda$, λ ranging from 0% to 100%. The probability of having an anchor event for MCAR dropouts was set to zero.

2. Anchor2: For MNAR dropouts the probability of having an anchor event was modeled as a Bernoulli random variable with $\pi_{it} = \gamma$. γ ranged from 25% to 100%. For MAR dropouts the probability of having an anchor was modeled as a Bernoulli random variable with $\pi_{it} = \lambda$. λ ranged from 0% to

100%. The probability of having an anchor event for MCAR dropouts was set to zero.

3.2.4.4 Models considered

We compared the performance of four models for longitudinal data.

3.2.4.4.1 Linear Mixed Effects Model

We considered a linear mixed model (LME) with no accommodation for informative dropout:

$$Y_i = X_i\beta + Z_ib_i + e_i$$

3.2.4.4.2 Pattern Mixture Models

This section will describe the additional conditions and restrictions necessary for PMM specification and will describe the three PMMs evaluated.

3.2.4.4.2.1 Identifying Restrictions

Additional restrictions are necessary in order to estimate all parameters in the PMMs. We assumed a linear trajectory in each pattern:

$$Y_i^{\{p\}} = X_i^{\{p\}}\beta^{\{p\}} + Z_i^{\{p\}}b_i^{\{p\}} + e_i^{\{p\}}$$

Slope can only be estimated in patterns with at least two observations per subject, thus we combined subjects dropping out after baseline with subjects dropping out after the first time point.

3.2.4.4.2.2 Models

We considered three different PMMs as described in section 2.3.1:

1. PMM1 – ignore the anchor event and group all dropouts together at each time point. Based on the data generating mechanism with five time points (baseline and four yearly follow-ups), this produces four total groups.

- ▶ $D_{ij} = [1\ 0\ 0\ 0\ 0]^T$ or $D_{ij} = [0\ 1\ 0\ 0\ 0]^T$ (dropout between baseline and year 1, or between year 1 and year 2)
- ▶ $D_{ij} = [0\ 0\ 1\ 0\ 0]^T$ (dropout between year 2 and year 3)
- ▶ $D_{ij} = [0\ 0\ 0\ 1\ 0]^T$ (dropout between year 3 and year 4)
- ▶ $D_{ij} = [0\ 0\ 0\ 0\ 0]^T$ (completers)

2. PMM2 – First separate dropouts based on the anchor event. All completers and dropouts without an anchor will be one group. All remaining dropouts are grouped by time point. Based on the data generating mechanism with five time points (baseline and four yearly follow-ups), this produces four total groups.

- ▶ $A_{ij} = [1\ 0\ 0\ 0\ 0]^T$ or $A_{ij} = [0\ 1\ 0\ 0\ 0]^T$ (dropout between baseline and year 1, or between year 1 and year 2 and anchor)
- ▶ $A_{ij} = [0\ 0\ 1\ 0\ 0]^T$ (dropout between year 2 and year 3 and anchor)

- ▶ $A_{ij} = [0\ 0\ 0\ 1\ 0]^T$ (dropout between year 3 and year 4 and anchor)
- ▶ $A_{ij} = [0\ 0\ 0\ 0\ 0]^T$ (no anchor)

3. PMM3 – First separate dropouts based on the anchor event. All completers are grouped together, all dropouts without an anchor event are grouped together, all dropouts with an anchor event are grouped together by time point. Based on the data generating mechanism with five time points (baseline and four yearly follow-ups), this produces five total groups.

- ▶ $A_{ij} = [1\ 0\ 0\ 0\ 0]^T$ or $A_{ij} = [0\ 1\ 0\ 0\ 0]^T$ (dropout between baseline and year 1, or between year 1 and year 2 and anchor)
- ▶ $A_{ij} = [0\ 0\ 1\ 0\ 0]^T$ (dropout between year 2 and year 3 and anchor)
- ▶ $A_{ij} = [0\ 0\ 0\ 1\ 0]^T$ (dropout between year 3 and year 4 and anchor)
- ▶ $A_{ij} = [0\ 0\ 0\ 0\ 0]^T$ and D_{ij} not equal to $[0\ 0\ 0\ 0\ 0]^T$ (dropout, no anchor)
- ▶ $A_{ij} = [0\ 0\ 0\ 0\ 0]^T$ and $D_{ij} = [0\ 0\ 0\ 0\ 0]^T$ (completers)

3.2.4.5 Parameters Evaluated

The standard deviation of change, denoted as S_b ($b_{1i} \sim N(0, \sigma_1)$), and residual standard deviation, denoted as S_e ($e_{ij} \sim N(0, \sigma_e)$), were examined at one quarter, one half, one, one and a half, and two times the rate of change (i.e., 1.875, 3.75, 7.5, 11.25, 15). The total dropout by end of study was varied at 20%, 40%, 60%, and 80%. The precision of the anchor event was investigated by varying the sensitivity and specificity of the anchor.

3.2.4.6 Criteria for Evaluation

We evaluated relative bias, standard error, coverage probability, and length of the 95% confidence interval. The absolute bias is calculated by subtracting the estimated progression estimate from the fixed true value and the relative bias is calculated by dividing the absolute bias by the fixed true value of the parameter estimate. The coverage probability is the proportion of times the 95% confidence interval includes the true progression estimate. We report the number of times that we were unable to compute a progression estimate due to not having enough subjects in each group. Five thousand replicates were run for each scenario. All simulations were conducted in SAS version 9.3 (SAS Institute, Cary, NC).

3.3 Results

3.3.1 Perfect Anchor

3.3.1.1 Relative Bias

We first present results for models with a perfect anchor event, that is, an event that classifies MNAR dropouts with 100% sensitivity and specificity. Holding S_b fixed at 1 times the rate of change and S_e fixed at one half times the rate of change, we examine relative bias by model and total dropout by the end of the study (Figure 3.3). Relative bias increases as dropout increases, and is largest for the LME model, ranging from approximately -6% for 20% total dropout to -11% for 80% total dropout, indicating that the model is moderately underestimating mean change. Relative bias from PMM1 is the next largest and offers an improvement of the LME model with values at approximately half the bias for the LME model. Finally, relative bias from models PMM2 and PMM3 is quite low with all values under 0.5%.

Relative bias increases for the LME and PMM1 models as we increase the variability to S_b of two times change and S_e to 1 times change, reaching approximately -23% for the LME model and -12% for the PMM1 model when dropout is 80%. Relative bias for the PMM2 and PMM3 models remains very small, with the relative bias under -1.5% for both models when dropout is 80% (Figure 3.4).

As we increase S_b , holding total dropout fixed and S_e equal to $\frac{1}{2}$ times slope, relative bias increases for the LME and PMM1 models, while the PMM2

and PMM3 models continue to perform well (Figure 3.5). The variability around the estimates of relative bias increases substantially as S_b increases.

Likewise, if we hold S_b constant at 1 times slope and total dropout constant at 60%, there is an increase in relative bias for the LME and PMM1 models as we increase S_e . PMM2 and PMM3 continue to perform well even under large amounts of variability (Figure 3.6).

3.3.1.2 Standard Error

Standard error is lowest for the LME model and similar for the 3 PMMs. Holding S_b constant at 1 times slope and S_e constant at $\frac{1}{2}$ times slope, the standard error of the estimate increases as we increase total dropout from 20% to 80% (Figure 3.7). As expected, standard error increases as S_b and S_e increase. It remains lowest for the LME model and similar for models PMM1, PMM2, and PMM3 (Figures 3.8-3.9).

3.3.1.3 Coverage

Holding S_b and S_e constant at 1 and $\frac{1}{2}$ times slope respectively, the PMM2 and PMM3 models are able to maintain coverage at 95% across all amounts of dropout. The coverage for the PMM2 model is approximately 92% for total dropout of 20% and dropped to 80% when total dropout increases to 80%. The coverage is lowest for the LME, ranging from 60% for total dropout of 20% to 20% for total dropout of 80%. These levels of coverage are maintained for PMM1, PMM2, and PMM3, even under increasing variability (Figures 3.10-3.11).

3.3.2 Imperfect Anchor: Impact of Precision

We changed the precision of the anchor event by decreasing the percent of informative dropouts with anchor event and increasing the amount of random dropouts with anchor event and evaluated the performance of the PMM2 and PMM3 models. The LME and PMM1 models do not incorporate the anchor event and therefore the performance will not change with changing precision of the anchor. Results for the LME and PMM1 models are presented for the selected combination of S_b , S_e and total dropout for comparison.

3.3.2.1 Anchor 1: Anchor Associated with Underlying Rate of Change

3.3.2.1.1 Relative Bias

For both the PMM2 and PMM3 models, relative bias increases with decreasing precision of the anchor event. Holding S_e constant at $0.5\beta_1$ and total dropout constant at 60%, relative bias increases both as the percent of MNAR dropouts with anchor decreased and as the percent of MAR dropouts with anchor increased. The first column of Table 3.1 displays the perfect anchor scenario – 100% of MNAR dropouts with anchor and 0% of MAR dropouts with anchor. As we increase the percent of MAR dropouts with anchor the relative bias increases, but across all values of S_b the relative bias never reaches that of the LME or PMM1 models. As we decrease the percent of MNAR dropouts with the anchor event the bias for both the PMM2 and PMM3 models increases. Changing the percent of MNAR dropouts with anchor (sensitivity) has a greater effect than

changing the amount of MAR dropouts with anchor (specificity); For 100% MNAR/100% MAR the bias is -3.1% for both the PMM2 and PMM3 models. For 25% MNAR/0% MAR the bias for the PMM2 models is -6.0% and for the PMM3 model is -2.5%.

Figure 3.12 contains all combinations of sensitivity and specificity for the scenario where S_b is $1 \cdot \beta_1$ and S_e is $0.5 \cdot \beta_1$. The first column displays the relative bias of PMM2 and the second column displays the relative bias of PMM3. As the quality of the anchor deteriorates, the relative bias for the PMM2 model eventually surpasses that for the PMM1 model and approaches that for the LME model. The relative bias for the PMM3 model approaches that for the PMM1 model but never surpasses it. In figure 3.13 the variability is increased to S_b equal to $2 \cdot \beta_1$ and S_e equal to $1 \cdot \beta_1$. The same patterns are evident, with the relative bias for PMM2 surpassing that for PMM1 as the quality of the anchor deteriorates.

3.3.2.1.2 Standard Error

For both the PMM2 and PMM3 models, standard error decreases with decreasing precision of the anchor event. Allowing MAR dropouts to have the anchor event does not greatly affect standard error, while allowing MNAR dropouts to not have the anchor resulted in lower standard errors. For S_b of $1 \cdot \beta_1$, S_e of $0.5 \cdot \beta_1$, total dropout of 60%, and 0% of MAR dropouts with anchor, the standard error for the PMM2 model is 0.22 when 100% of MNAR dropouts have

the anchor event and 0.18 when 25% of MNAR dropouts have the anchor event. A similar trend is seen for the PMM3 model (Table 3.2).

3.3.2.1.3 Coverage

Coverage of the 95% confidence interval decreases with decreasing precision of the anchor event for both the PMM2 and PMM3 models. For the PMM2 model the decrease in coverage is larger as the percent of MNAR dropouts with anchor decreased than as the percent of MAR dropouts with anchor increased. For S_b of $1 \cdot \beta_1$, S_e of $0.5 \cdot \beta_1$, total dropout of 60%, coverage is 94.6% for a perfect anchor event, 86.9% when 100% of both MNAR and MAR dropouts have the anchor, and 67.5% when 25% of MNAR and 0% of MAR dropouts have the anchor (Table 3.3). As the percent of MNAR dropouts with anchor approaches 50%, the PMM1 model begins to outperform the PMM2 model. The PMM3 model is able to maintain coverage close to 90% as long as 0% of MAR dropouts have the anchor event. Coverage begins to drop as more MAR dropouts are allowed to have the anchor. When 50% or less of the MNAR dropouts have the anchor and 100% of MAR dropouts have the anchor event the coverage from the PMM3 model is comparable to that of the PMM1 model.

3.3.2.2 Anchor 2: Anchor Randomly Assigned to MNAR Dropouts

3.3.2.2.1 Relative Bias

The PMM2 and PMM3 models display more relative bias when the anchor event is randomly assigned to subjects with MNAR dropout (Anchor 2) than when the probability of having anchor is related to underlying change (Anchor 1). In scenarios where only 25% of MNAR dropouts have the anchor event, the relative bias from the PMM2 models with Anchor 2 approaches the relative bias from the LME model and surpasses the bias from the PMM1 model. Table 3.4 displays relative bias for scenarios with S_e equal to $0.5\beta_1$ and with 25% of MAR dropouts having an anchor event. With Anchor 1, the relative bias from the PMM2 model approaches that from the PMM1 model when about 50% of the MNAR dropouts have an anchor event; with Anchor 2 this happens when 75% of the dropouts have an anchor event. When the anchor event is not associated with underlying change even small decreases in precision of the anchor can lead to substantial bias under PMM2. The PMM3 model maintains relative bias less than or comparable to the bias from the PMM1 model under both Anchor 1 and Anchor 2, under varying levels of anchor precision.

3.3.2.2.2 Standard Error

The standard error for both the PMM2 and PMM3 models is similar for scenarios with Anchor 1 and Anchor 2 (Table 3.5).

3.3.2.2.3 Coverage

Coverage for both the PMM2 and PMM3 models decreases for scenarios with Anchor 2 vs. Anchor 1. When the anchor event is related to underlying change (Anchor 1), the PMM2 model performs better than the PMM1 model, in terms of coverage, when at least 50% of MNAR dropouts have the anchor event. If the anchor event is not related to underlying change, then at least 75% of the MNAR dropouts must have the anchor event for PMM2 to outperform PMM1. Under Anchor 2, when only 25% of MNAR dropouts have the anchor event the performance of PMM2 is close to that of the LME. The PMM3 model outperforms the LME model under all scenarios. PMM3 outperforms PMM2 under all scenarios with Anchor 1. Under Anchor 2, the coverage of PMM3 is close to, and sometimes less than, the coverage of the PMM1 model when less than 50% of the MNAR dropouts have the anchor event (Table 3.6).

3.4. Discussion

We evaluated the impact of an imprecise anchor event in evaluating the rate of change in longitudinal studies with informative dropouts using PMMs. We showed that PMMs that incorporate an anchor event perform well when the anchor event is precise, but the performance deteriorates as the anchor event becomes less precise, especially when the sensitivity of the anchor is poor. Our proposed update to the grouping of dropout patterns had the best performance across all four models evaluated.

We evaluated four methods for estimating the rate of change in longitudinal studies. The first method was a linear mixed model that did not incorporate the dropout mechanism (LME). We have shown previously that this method performs adequately when the overall amount of dropout and the variability in the outcome are small. The second method was a PMM with separate patterns for each dropout group (PMM1). This model did not incorporate the anchor event and assumed that every dropout was informative. The third method was a PMM with separate patterns for each dropout group, with only dropouts with an anchor event counted in each dropout group (PMM2). All completers and dropouts without an anchor event were grouped together. The fourth model was a PMM using the same patterns as the third model, except completers and dropouts without an anchor event were grouped separately (PMM3). When the amount of variability was large or the total amount of dropout was large, the LME and PMM1 models demonstrated large relative bias and low coverage. The PMM2 and PMM3 models performed well under these scenarios, as long as the anchor event was precise. As the anchor event became less precise the performance of the PMM2 model deteriorated quickly, falling behind that of the PMM1 model. This deterioration was even more rapid when the anchor event was randomly assigned to MNAR dropouts, instead of being associated with underlying change.

The PMM3 model is a new approach with respect to the grouping of the dropouts. When the anchor event is imprecise it may be possible that some

subjects drop out of the study but do not have the anchor (e.g., a subject with OA drops out of a study because of functional impairment and the associated difficulty getting to study visits, but does not undergo TKR). The PMM3 model demonstrated the lowest bias and highest coverage across all scenarios examined. The model was able to handle imprecision in the anchor event better than the PMM2 model, but nevertheless for some scenarios still demonstrated substantial bias and low coverage.

Advantages to using the PMM3 approach include that it is easy to implement, and the performance was equal to or better than that from the LME and PMM1 models, which disregarded the anchor event completely. A disadvantage is that it requires $T+1$ groups, where T is the total number of measurements. In studies with small sample size or many time points it may not be possible to estimate parameters for $T+1$ groups. Hogan et al. propose examining the covariate effects of each group graphically to determine if any can be combined. [12] To use the PMM3 approach it is essential that there are enough drop outs without an anchor event to form a separate group, and this may not always be practical.

Investigators should work to understand the association between anchor event and informative dropout. Sensitivity of the anchor should be assessed: we can review reason for dropout among subjects that dropout and do not have anchor and approximate how many of these are potential informative dropouts. This may involve review of medical records to determine if any procedures took

place after study discontinuation. Expert opinion from a clinician can also provide valuable information about this relationship. For example, in the OA study we would review reasons for dropout among subjects that did not have TKR and look for things such as difficulty getting to study appointments. We would also review medical records to see if any of these subjects went on to have TKR in year after dropout. We may also ask surgeons to estimate how many of their patients qualify for a TKR based on pain/functional impairment but are unwilling or unable to undergo the procedure. We know, for example, that there are racial disparities in utilization of TKR, with African-Americans less willing to consider the surgery when it is clinically recommended. [39] Specificity of the anchor can also be assessed. Examining the association between anchor and outcome will give some indication of the specificity of the anchor, though it is important to keep in mind that with an MNAR missing data mechanism we may not observe this association. The amount of dropout and variability in outcome are also important factors and these can be easily quantified; more dropout and more variability in outcome led to worse performance of all four models and would imply that extra attention should be paid to the quality of the anchor event.

As with any study that utilizes simulations, our study had several limitations. Missing data in our study resulted only from dropout; there was no intermittent missing data. In addition, we examined only one missing data mechanism where 60% of the dropouts were not at random, and many subjects dropped out after the baseline visit. Decreasing the amount of not at random

dropouts and having dropout later in the study would likely improve the performance of all four methods, while increasing the amount of not at random dropouts would likely cause the performance to worsen. We attempted to model the heterogeneity in dropout reason by including three types of dropout: MNAR, MAR, and MCAR. Reasons for dropout in a longitudinal study are complex and it is unlikely that the underlying missing data mechanism would as simple as described above. Data were simulated as multivariate normal with a normal error term. In reality, the distribution of the WOMAC score from the motivating example is bounded between 0 and 100, and is moderately skewed. We chose the data generating mechanism for ease of interpretation, and a different underlying distribution or error distribution could lead to different conclusions. This is a topic for future work. Finally, we only examined scenarios with 1000 subjects and 5 time points. The PMM described required at least one dropout group per time point. Decisions about how to group sparse dropout patterns were not examined in this analysis.

In this paper we compared the performance of four different models for evaluating the rate of change in longitudinal studies when there is a heterogeneous informative dropout mechanism and imprecision in the anchor event. The proposed approach, which separates completers from dropouts that do not have an anchor event, had the superior performance across the 4 models. The model had good performance even under scenarios with high variability in the outcome and large amounts of dropout and was able to maintain this

performance even as the anchor event became less precise; however, the performance began to deteriorate as the anchor event became very imprecise.

Figure 3.3 Relative bias by model and total dropout, perfect anchor ($S_b=1*\beta_1$; $S_e=0.5*\beta_1$).

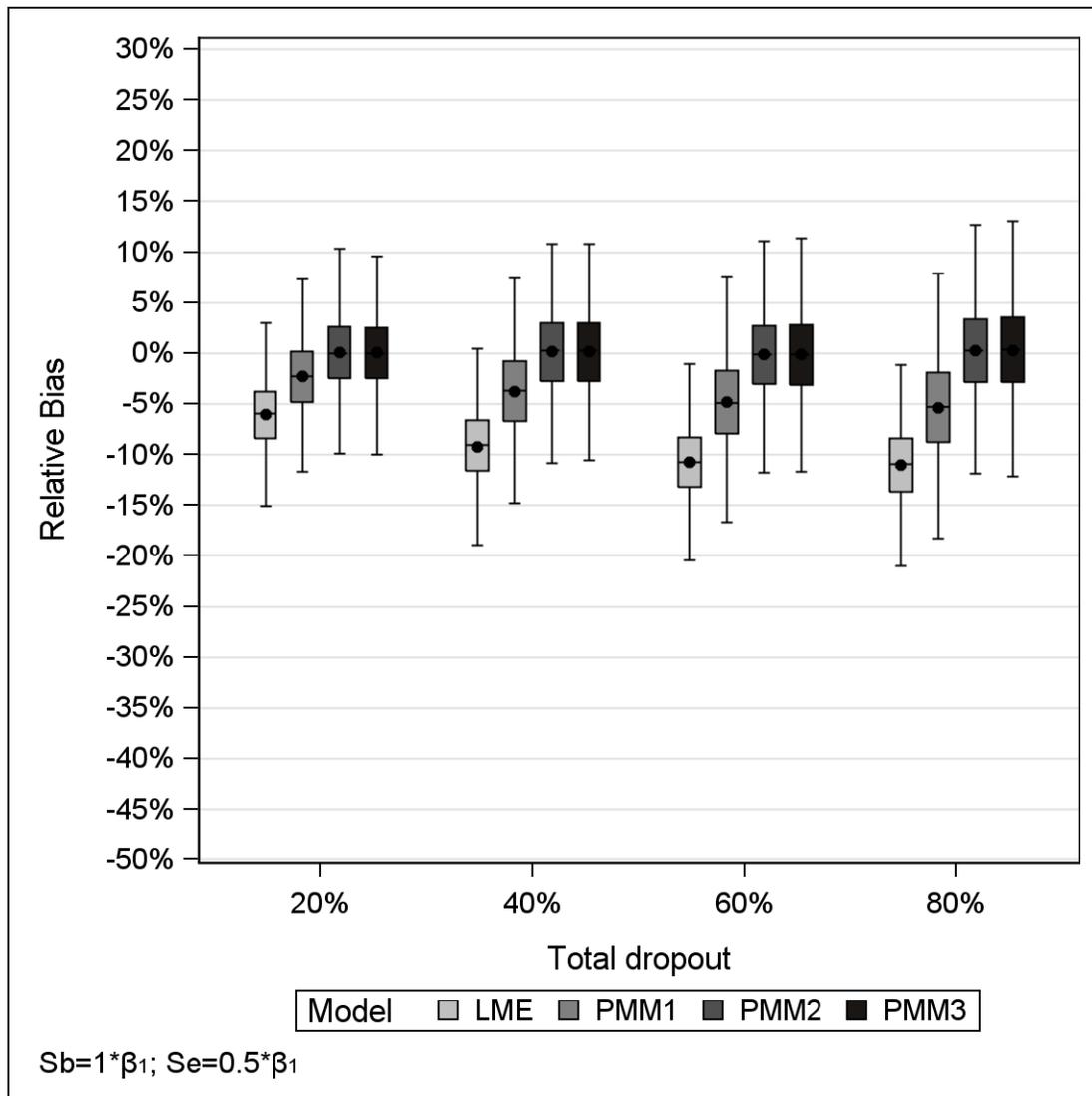


Figure 3.4 Relative bias by model and total dropout, perfect anchor ($S_b=2*\beta_1$; $S_e=1*\beta_1$).

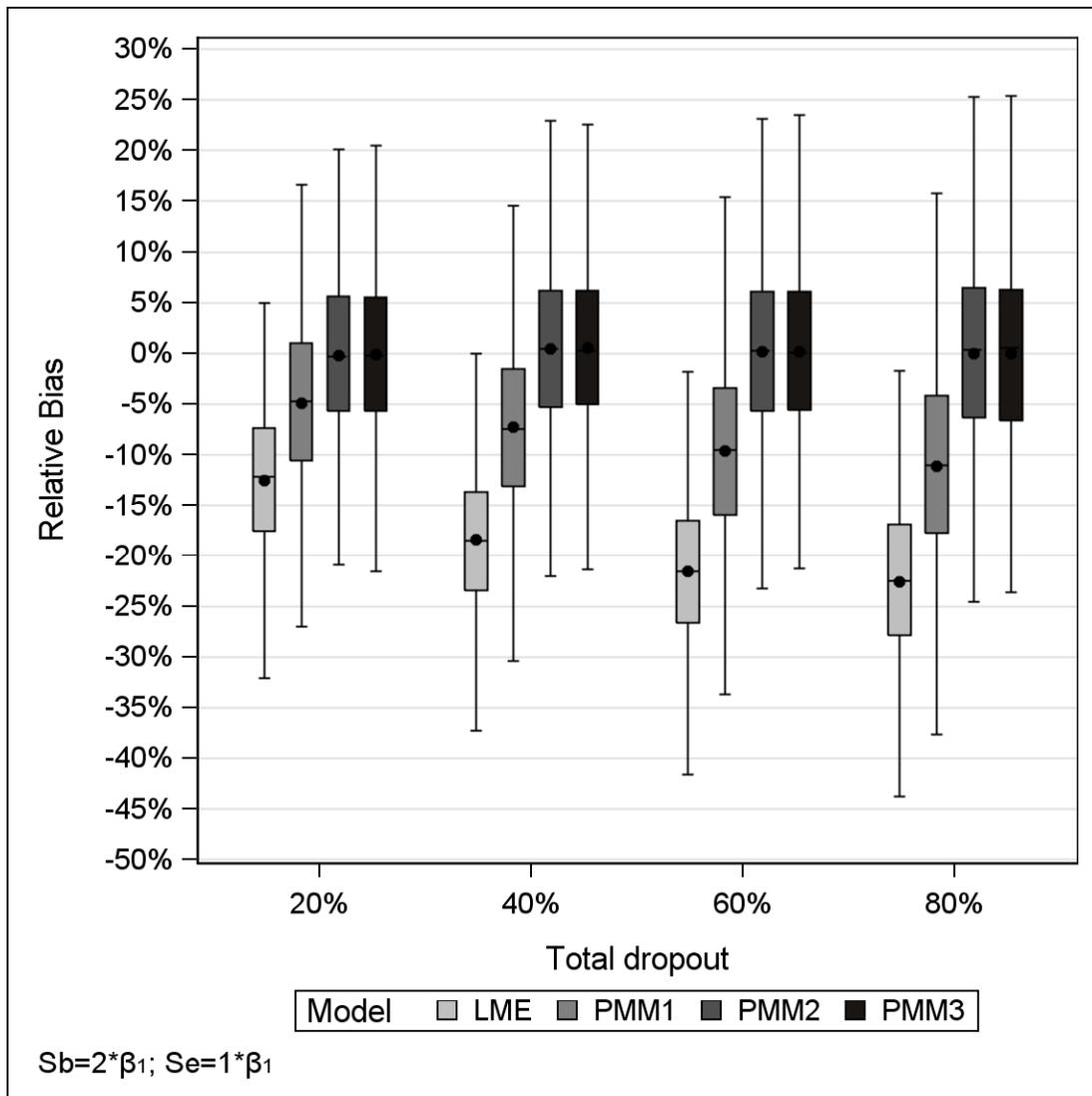


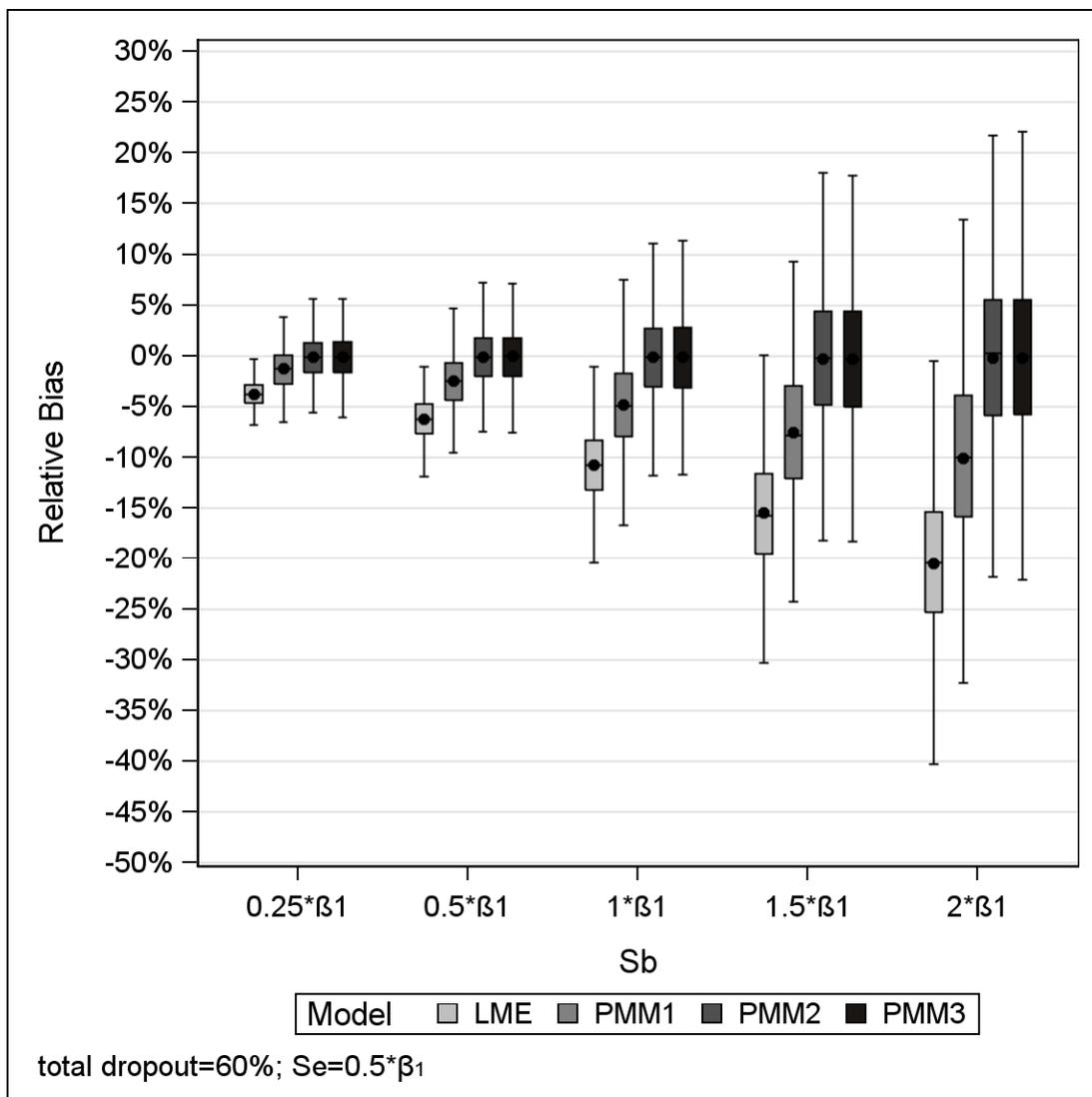
Figure 3.5 Relative bias by model and S_b , perfect anchor.

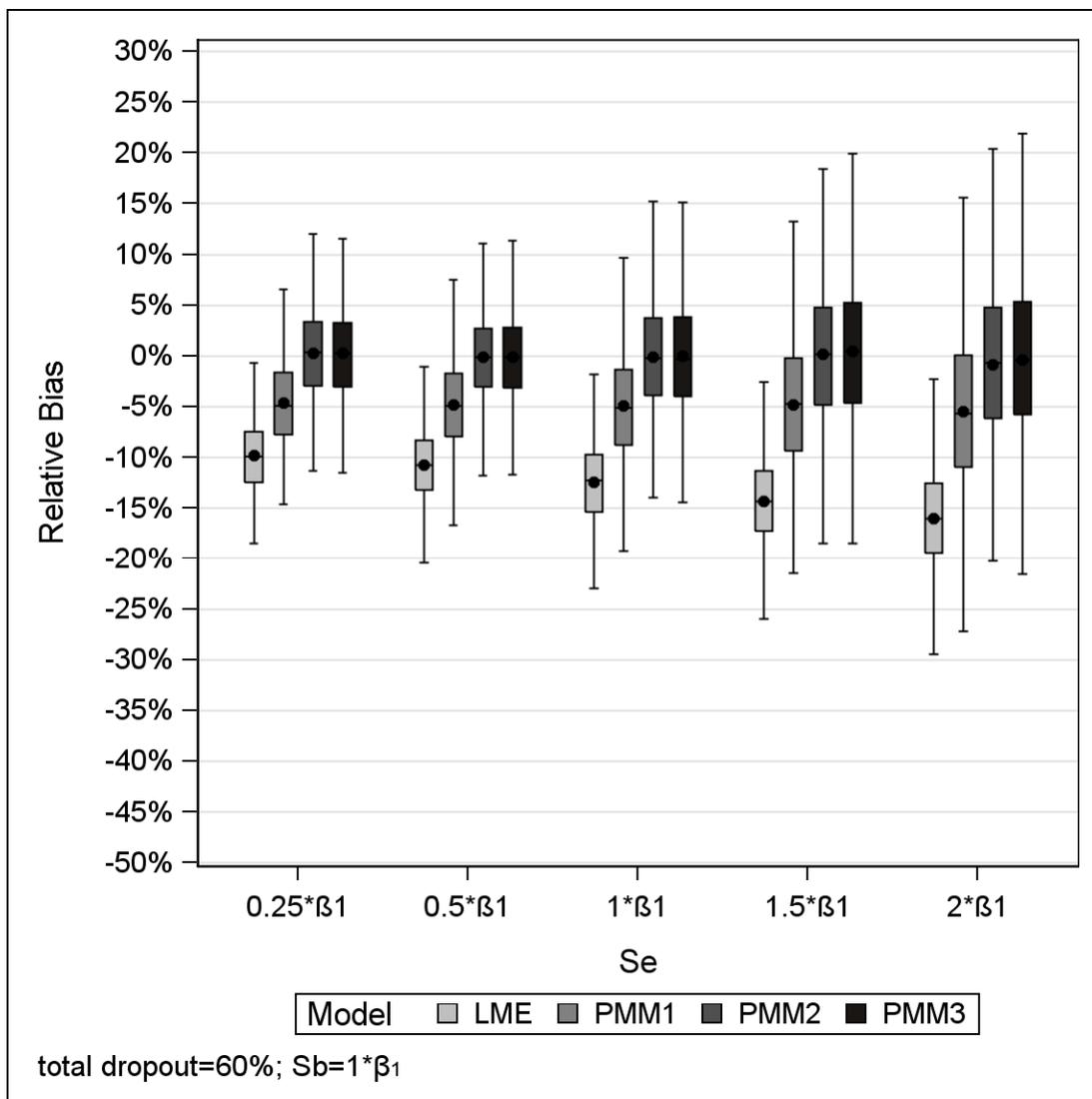
Figure 3.6 Relative bias by model and S_e , perfect anchor.

Figure 3.7 Standard error by model and total dropout, perfect anchor.

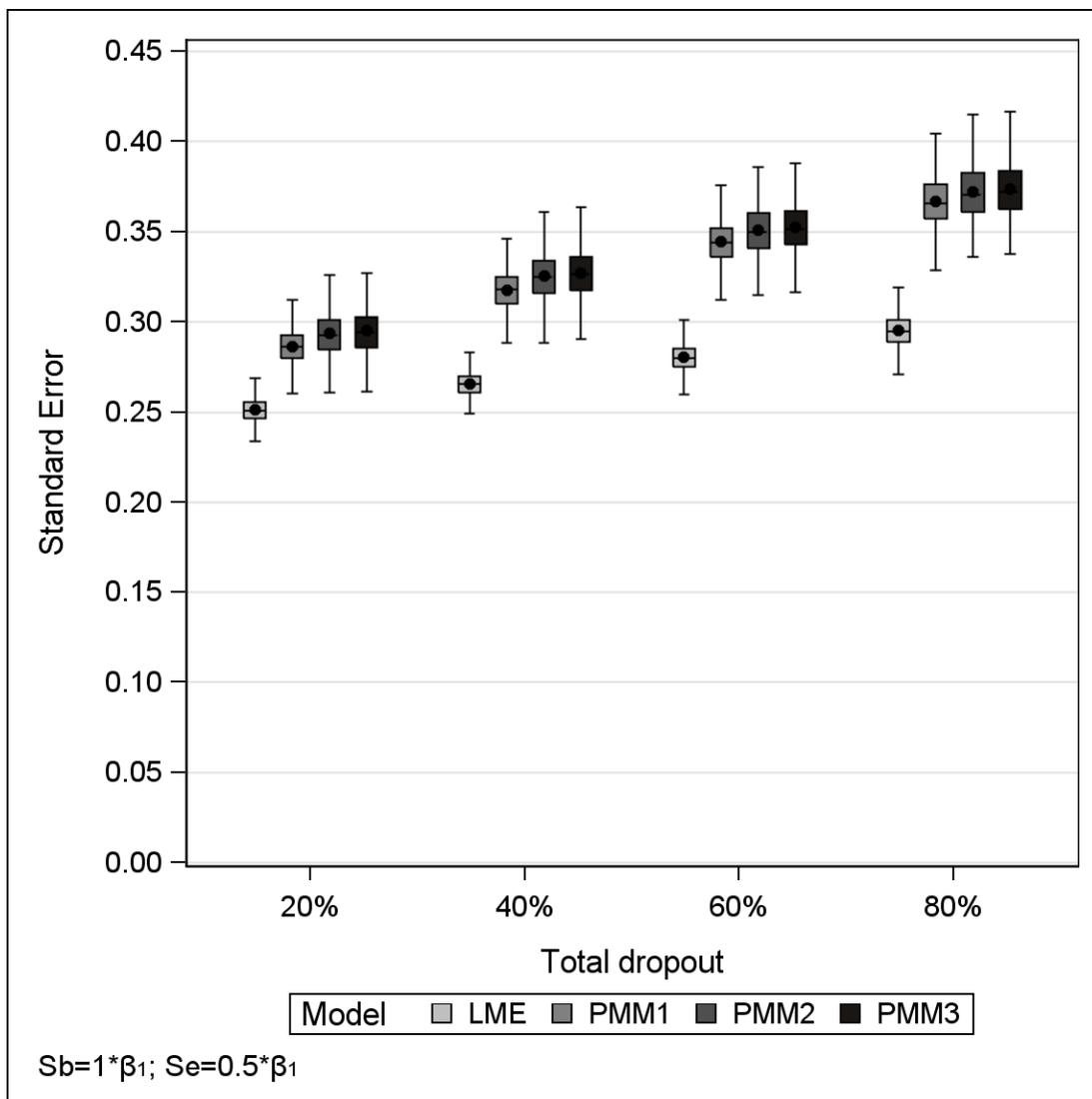


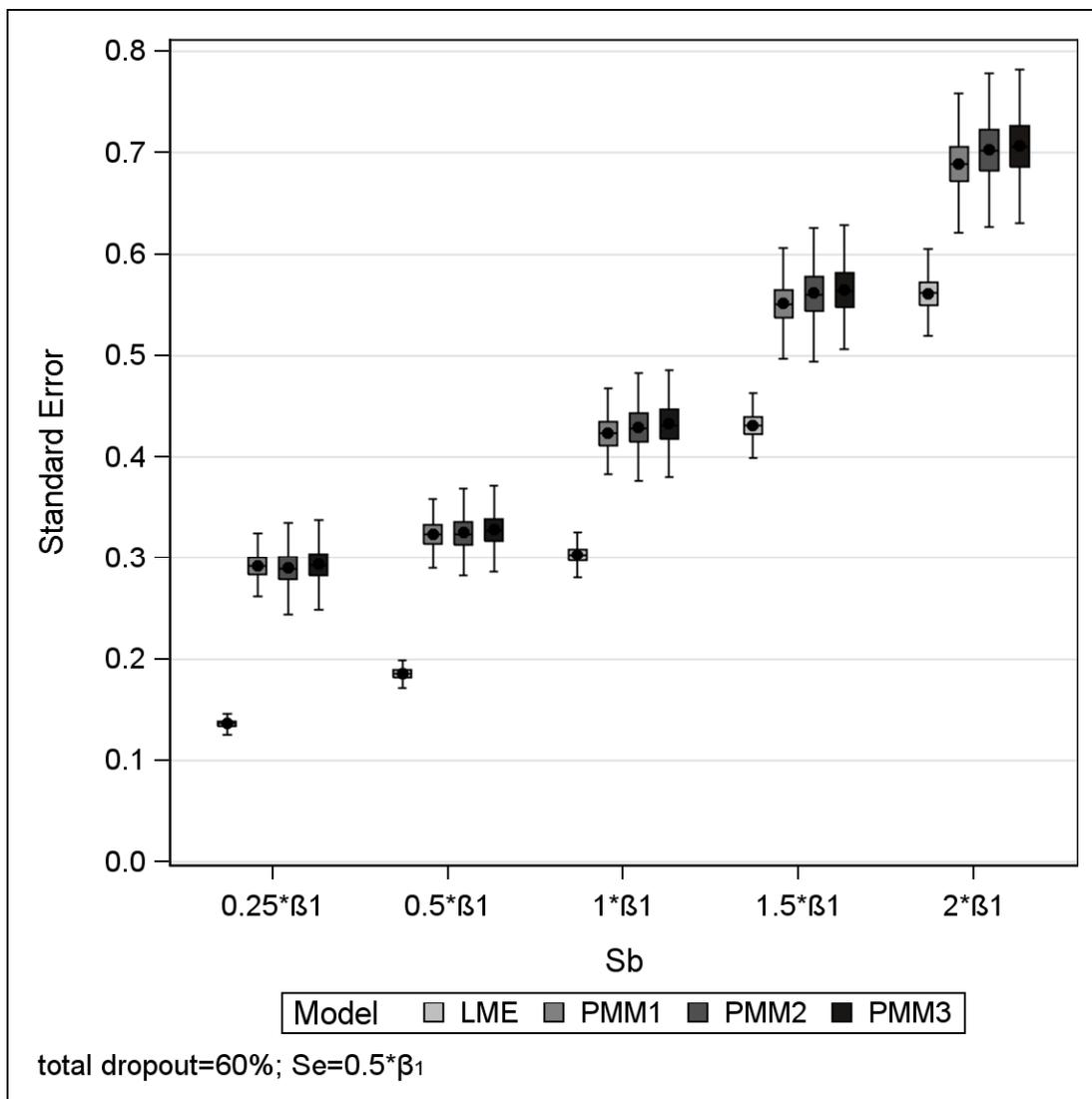
Figure 3.8 Standard error by model and S_b , perfect anchor.

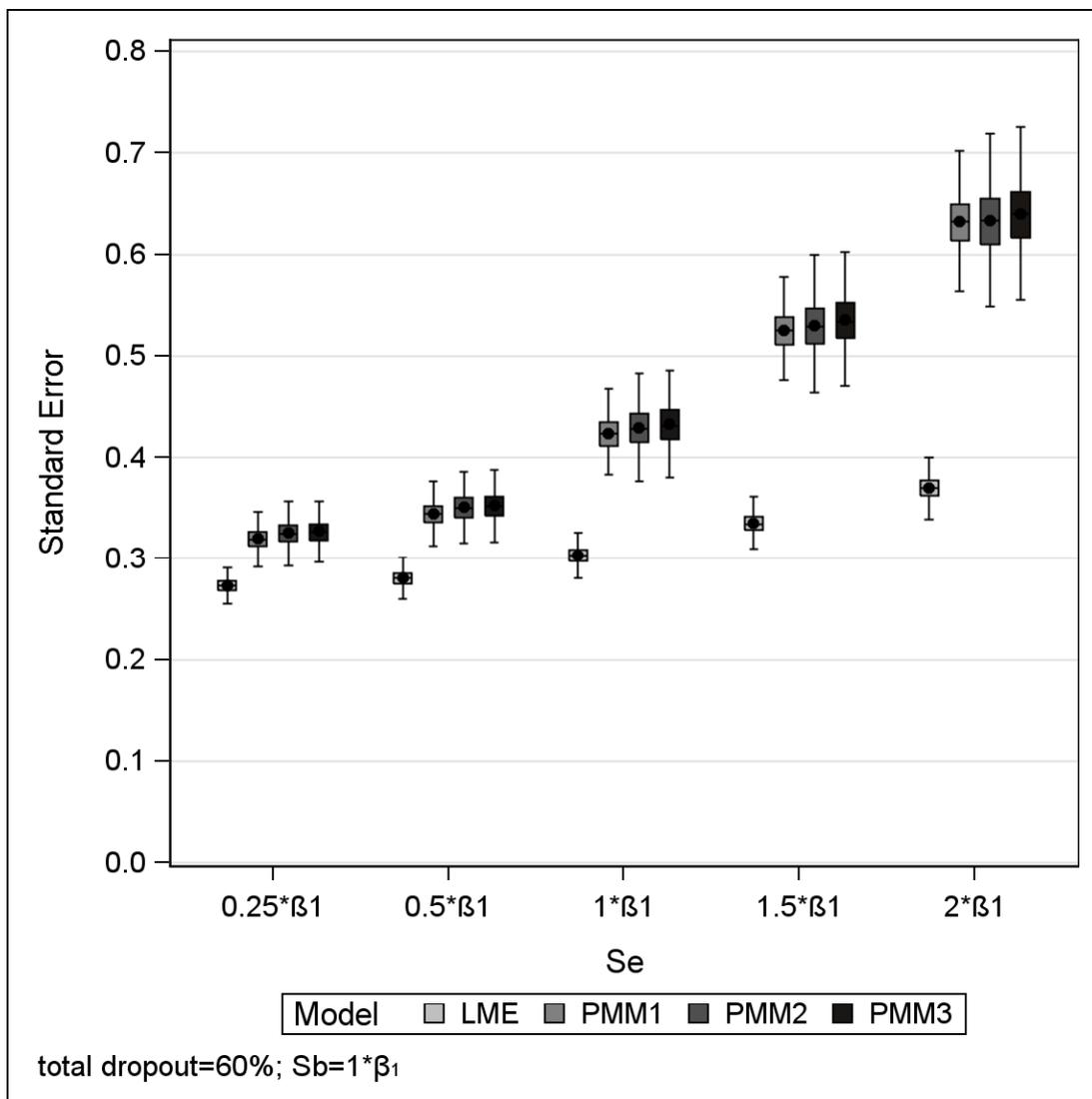
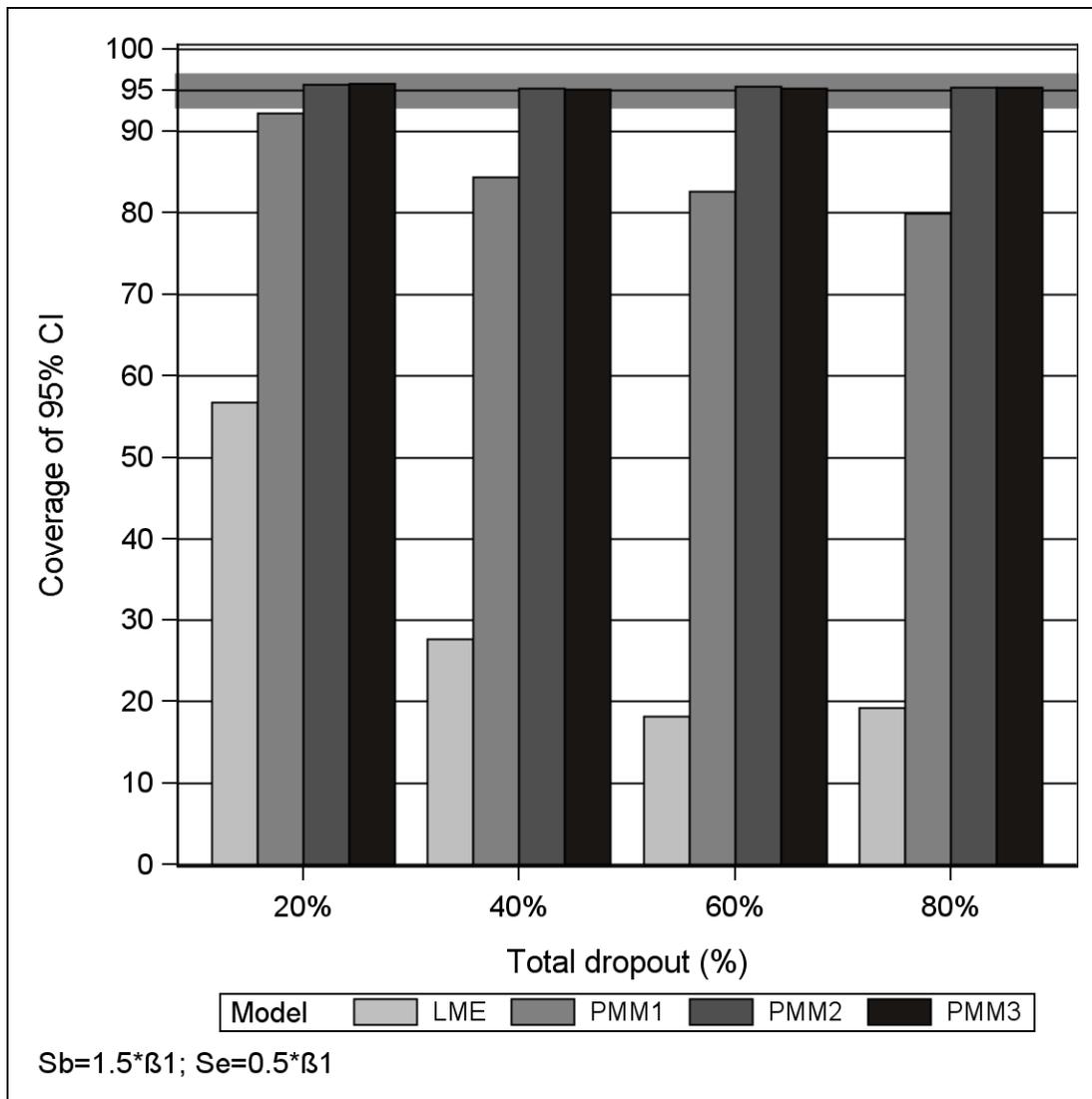
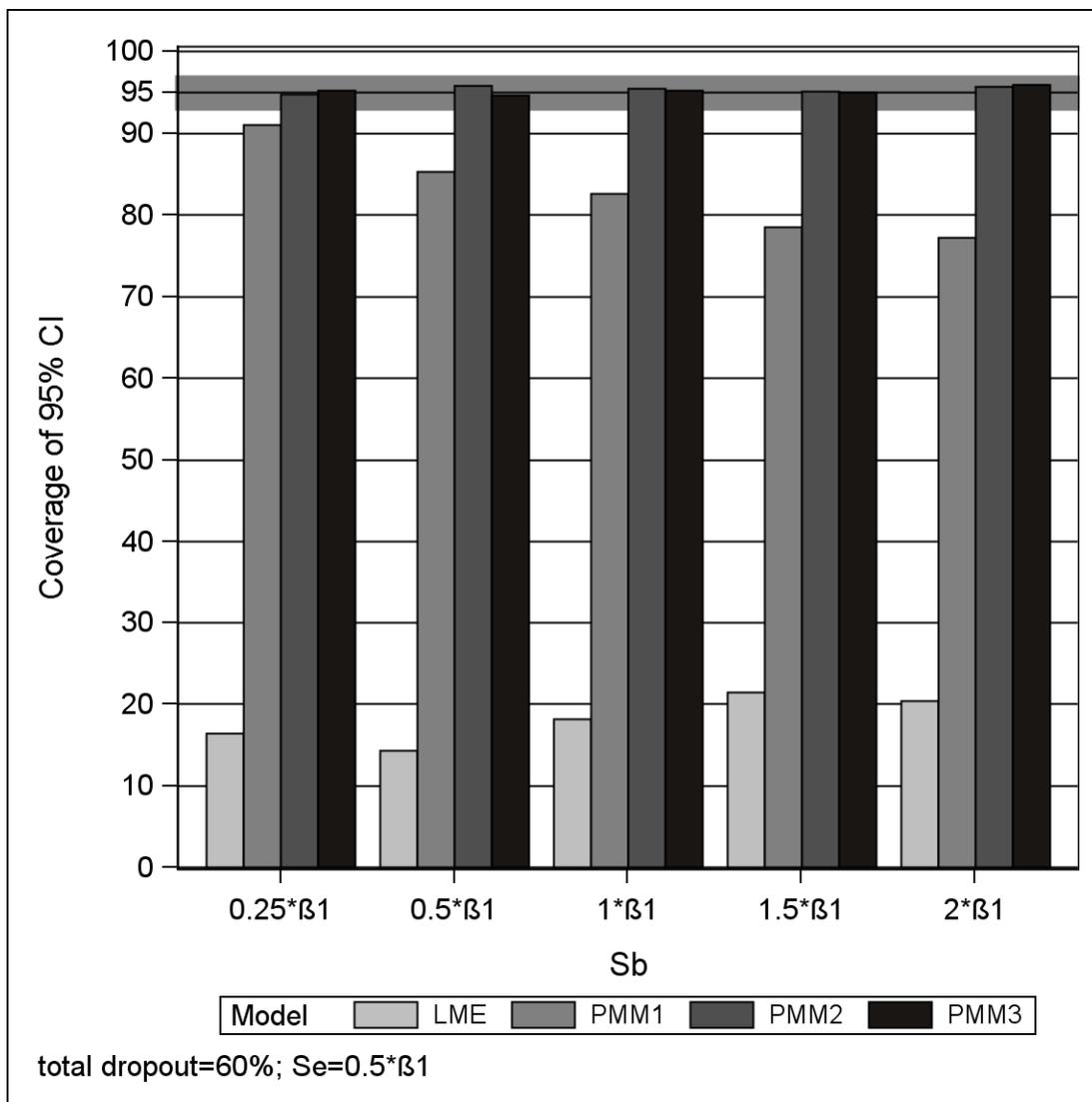
Figure 3.9 Standard error by model and S_e , perfect anchor.

Figure 3.10 Coverage of 95% CI by model and total dropout, perfect anchor.

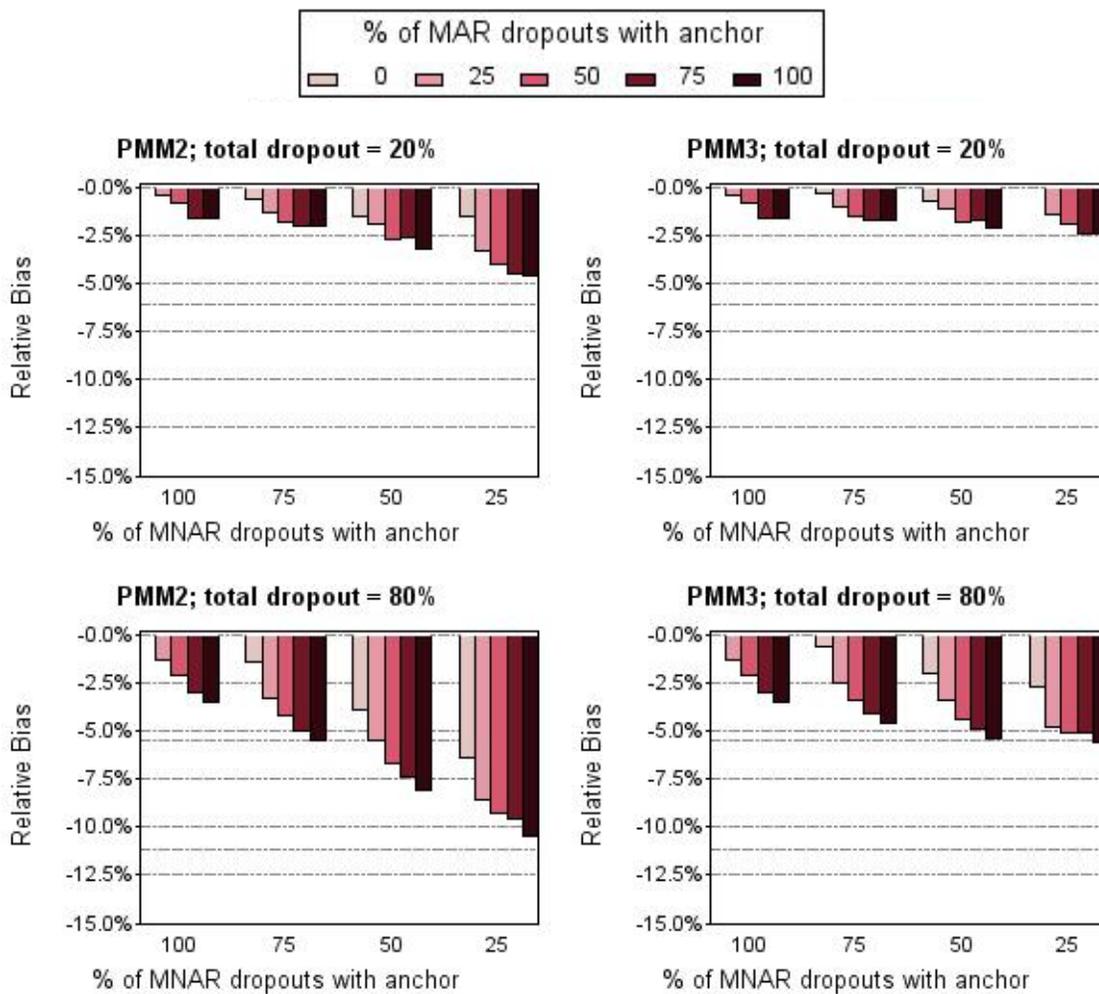


Shaded area represents the binomial margin of error based on the number of simulations

Figure 3.11 Coverage of 95% CI by model and S_b , perfect anchor.

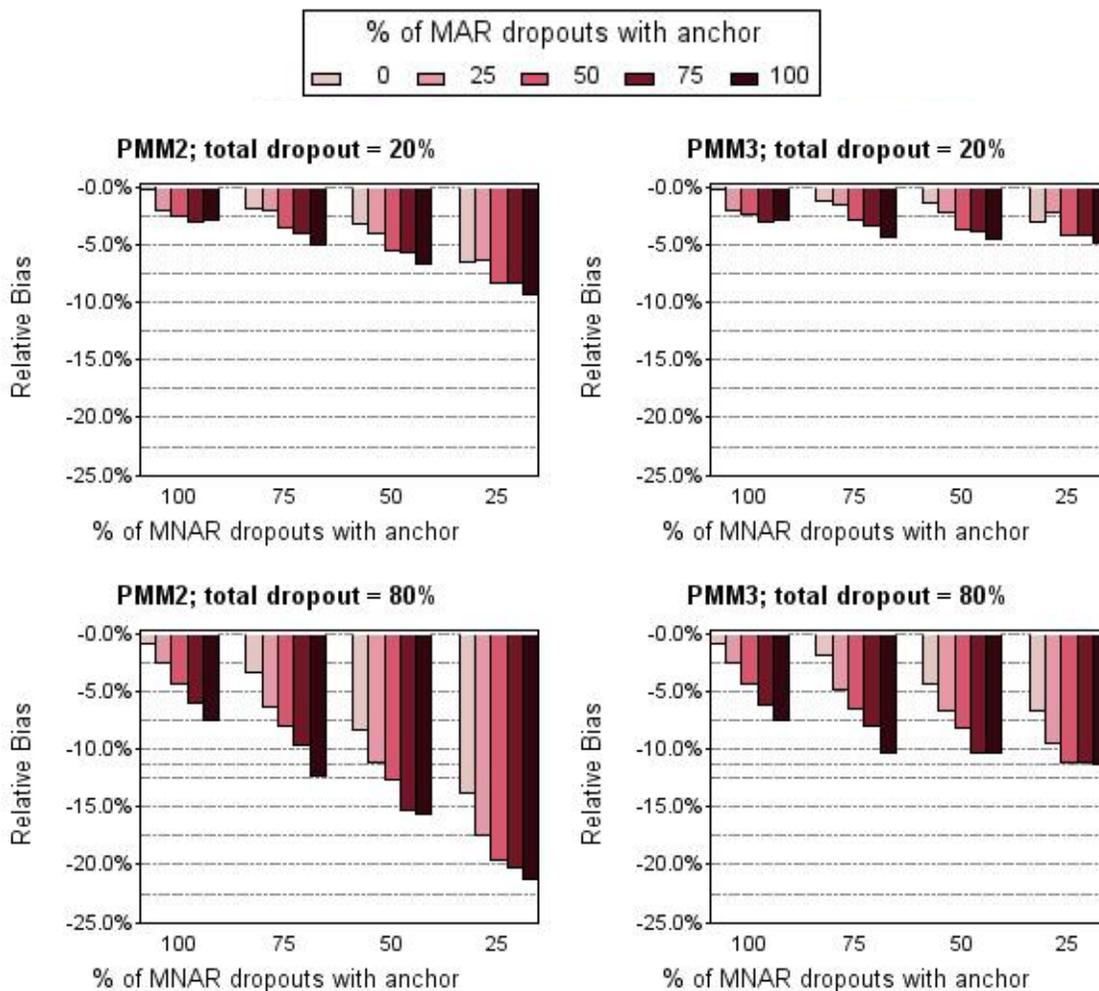
Shaded area represents the binomial margin of error based on the number of simulations

Figure 3.12 Relative bias by precision of anchor, model, and total dropout ($S_b=1*\beta_1$; $S_e=0.5*\beta_1$).



Dashed line indicates bias for PMM1; Dotted line indicates bias for LME
 $S_b=1*\beta_1$ and $S_e=0.5*\beta_1$

Figure 3.13 Relative bias by precision of anchor, model, and total dropout ($S_b=2*\beta_1$; $S_e=1*\beta_1$).



Dashed line indicates bias for PMM1; Dotted line indicates bias for LME
 $S_b=2*\beta_1$ and $S_e=1*\beta_1$

Table 3.1 Percent relative bias by precision of anchor, S_b , and model.

S_b	Model	100% MNAR					75% MNAR					
		0% MAR	25% MAR	50% MAR	75% MAR	100% MAR	0% MAR	25% MAR	50% MAR	75% MAR	100% MAR	
0.25* β_1	LME											-3.8
	PMM1											-1.2
	PMM2	-0.3	-0.1	-0.1	-0.5	-0.6	-0.4	-0.5	-0.9	-1.0	-1.6	
	PMM3	-0.3	-0.1	-0.1	-0.4	-0.7	-0.0	-0.3	-0.6	-0.6	-1.3	
0.5* β_1	LME											-6.2
	PMM1											-2.5
	PMM2	-0.1	-0.6	-1.2	-1.5	-1.6	-0.8	-1.4	-2.0	-2.1	-2.5	
	PMM3	-0.1	-0.6	-1.2	-1.5	-1.6	-0.4	-0.9	-1.6	-1.6	-2.0	
1* β_1	LME											-10.7
	PMM1											-4.9
	PMM2	0.1	-0.9	-2.1	-2.7	-3.1	-1.7	-2.6	-3.1	-4.5	-5.3	
	PMM3	0.1	-0.9	-2.1	-2.7	-3.1	-1.1	-1.9	-2.4	-3.8	-4.6	
1.5* β_1	LME											-15.5
	PMM1											-7.5
	PMM2	-0.4	-1.3	-3.2	-4.5	-4.8	-1.7	-3.9	-5.3	-6.4	-7.4	
	PMM3	-0.4	-1.3	-3.3	-4.6	-4.8	-0.9	-3.1	-4.4	-5.5	-6.3	
2* β_1	LME											-20.2
	PMM1											-9.7
	PMM2	-0.2	-2.3	-3.4	-5.1	-7.3	-2.2	-5.5	-6.5	-8.6	-9.3	
	PMM3	-0.2	-2.3	-3.4	-5.0	-7.3	-1.2	-4.3	-5.3	-7.4	-8.0	

$S_e=0.5*\beta_1$ and total dropout=60%

Table 3.1 Percent relative bias by precision of anchor, S_b , and model (continued).

S_b	Model	50% MNAR					25% MNAR					
		0% MAR	25% MAR	50% MAR	75% MAR	100% MAR	0% MAR	25% MAR	50% MAR	75% MAR	100% MAR	
0.25* β_1	LME											-3.8
	PMM1											-1.2
	PMM2	-1.3	-1.5	-1.5	-2.0	-2.2	-2.3	-2.5	-2.8	-2.9	-2.6	
	PMM3	-0.5	-0.7	-0.7	-1.2	-1.3	-0.8	-0.9	-1.2	-1.2	-0.7	
0.5* β_1	LME											-6.2
	PMM1											-2.5
	PMM2	-2.1	-2.5	-3.0	-3.4	-3.6	-3.5	-4.0	-4.4	-5.0	-5.4	
	PMM3	-1.0	-1.4	-1.7	-2.1	-2.1	-1.2	-1.6	-1.8	-2.4	-2.6	
1* β_1	LME											-10.7
	PMM1											-4.9
	PMM2	-3.2	-4.5	-5.4	-6.5	-6.9	-6.0	-7.5	-8.5	-9.0	-9.2	
	PMM3	-1.5	-2.7	-3.5	-4.3	-4.5	-2.5	-3.8	-4.7	-4.8	-4.5	
1.5* β_1	LME											-15.5
	PMM1											-7.5
	PMM2	-4.6	-7.5	-8.6	-9.3	-10.4	-9.1	-11.3	-12.3	-13.5	-13.6	
	PMM3	-2.2	-5.0	-6.0	-6.4	-7.2	-4.4	-6.2	-6.8	-7.5	-7.1	
2* β_1	LME											-20.2
	PMM1											-9.7
	PMM2	-6.4	-8.9	-10.8	-12.6	-13.4	-11.8	-14.4	-16.1	-16.8	-18.1	
	PMM3	-3.4	-5.7	-7.4	-8.8	-9.1	-5.5	-7.8	-9.1	-9.1	-9.7	

$S_e=0.5*\beta_1$ and total dropout=60%

Table 3.2 Standard error by precision of anchor, S_b , and model.

S_b	Model	100% MNAR					75% MNAR				
		0% MAR	25% MAR	50% MAR	75% MAR	100% MAR	0% MAR	25% MAR	50% MAR	75% MAR	100% MAR
0.25* β_1	LME					0.09					
	PMM1					0.16					
	PMM2	0.16	0.16	0.16	0.16	0.16	0.15	0.15	0.15	0.15	0.15
	PMM3	0.16	0.16	0.16	0.16	0.16	0.15	0.15	0.15	0.15	0.15
0.5* β_1	LME					0.15					
	PMM1					0.21					
	PMM2	0.22	0.21	0.21	0.21	0.21	0.20	0.20	0.20	0.20	0.20
	PMM3	0.22	0.22	0.21	0.21	0.21	0.21	0.21	0.20	0.20	0.20
1* β_1	LME					0.28					
	PMM1					0.34					
	PMM2	0.35	0.35	0.35	0.35	0.35	0.34	0.33	0.33	0.33	0.33
	PMM3	0.35	0.35	0.35	0.35	0.35	0.34	0.34	0.34	0.33	0.33
1.5* β_1	LME					0.41					
	PMM1					0.49					
	PMM2	0.50	0.50	0.49	0.49	0.49	0.48	0.48	0.48	0.47	0.47
	PMM3	0.50	0.50	0.49	0.49	0.49	0.49	0.48	0.48	0.48	0.48
2* β_1	LME					0.55					
	PMM1					0.64					
	PMM2	0.66	0.65	0.65	0.64	0.64	0.63	0.63	0.63	0.62	0.62
	PMM3	0.66	0.65	0.65	0.64	0.64	0.64	0.63	0.63	0.63	0.63

$S_e=0.5*\beta_1$ and total dropout=60%

Table 3.2 Standard error by precision of anchor, S_b , and model (continued).

S_b	Model	50% MNAR					25% MNAR				
		0% MAR	25% MAR	50% MAR	75% MAR	100% MAR	0% MAR	25% MAR	50% MAR	75% MAR	100% MAR
0.25* β_1	LME					0.09					
	PMM1					0.16					
	PMM2	0.14	0.14	0.14	0.14	0.14	0.12	0.12	0.12	0.12	0.12
	PMM3	0.14	0.14	0.14	0.14	0.14	0.13	0.13	0.13	0.13	0.13
0.5* β_1	LME					0.15					
	PMM1					0.21					
	PMM2	0.19	0.19	0.19	0.19	0.19	0.18	0.17	0.17	0.17	0.17
	PMM3	0.20	0.19	0.19	0.19	0.19	0.18	0.18	0.18	0.18	0.18
1* β_1	LME					0.28					
	PMM1					0.34					
	PMM2	0.32	0.32	0.32	0.32	0.32	0.31	0.30	0.30	0.30	0.30
	PMM3	0.33	0.33	0.32	0.32	0.32	0.31	0.31	0.31	0.31	0.31
1.5* β_1	LME					0.41					
	PMM1					0.49					
	PMM2	0.47	0.46	0.46	0.46	0.46	0.45	0.44	0.44	0.44	0.44
	PMM3	0.47	0.47	0.47	0.47	0.46	0.46	0.45	0.45	0.45	0.45
2* β_1	LME					0.55					
	PMM1					0.64					
	PMM2	0.61	0.61	0.60	0.60	0.60	0.59	0.58	0.58	0.58	0.58
	PMM3	0.62	0.61	0.61	0.61	0.61	0.60	0.59	0.59	0.59	0.59

$S_e=0.5*\beta_1$ and total dropout=60%

Table 3.3 Coverage of 95% CI by precision of anchor, S_b , and model.

S_b	Model	100% MNAR					75% MNAR				
		0% MAR	25% MAR	50% MAR	75% MAR	100% MAR	0% MAR	25% MAR	50% MAR	75% MAR	100% MAR
0.25* β_1	LME					14.9					
	PMM1					91.9					
	PMM2	93.7	93.8	97.6	92.0	94.0	94.7	94.7	90.7	93.3	85.0
	PMM3	94.3	94.7	97.6	94.0	94.0	95.3	96.0	89.3	96.0	92.0
0.5* β_1	LME					13.7					
	PMM1					85.5					
	PMM2	95.5	97.7	91.8	93.6	89.6	94.8	93.9	88.2	88.6	84.5
	PMM3	93.9	96.8	91.8	93.6	90.1	95.8	95.2	91.6	92.3	89.6
1* β_1	LME					18.2					
	PMM1					80.9					
	PMM2	94.6	93.9	92.6	92.5	86.9	92.0	90.9	90.1	82.3	76.0
	PMM3	94.4	94.1	92.4	93.6	87.5	94.4	93.7	92.4	86.0	79.5
1.5* β_1	LME					19.6					
	PMM1					79.2					
	PMM2	95.3	93.5	91.3	89.4	86.4	95.6	89.9	85.5	82.3	77.6
	PMM3	95.4	93.5	91.0	88.7	86.1	96.1	93.0	88.6	86.0	82.6
2* β_1	LME					21.0					
	PMM1					78.9					
	PMM2	95.2	93.5	91.7	90.4	87.6	95.1	90.4	89.3	80.5	80.0
	PMM3	95.2	93.6	91.2	90.2	88.0	94.5	92.7	91.3	85.4	85.3

$S_e=0.5*\beta_1$ and total dropout=60%

Table 3.3 Coverage of 95% CI by precision of anchor, S_b , and model (continued).

S_b	Model	50% MNAR					25% MNAR				
		0% MAR	25% MAR	50% MAR	75% MAR	100% MAR	0% MAR	25% MAR	50% MAR	75% MAR	100% MAR
0.25* β_1	LME					14.9					
	PMM1					91.9					
	PMM2	88.0	89.3	85.3	77.3	68.0	74.0	66.7	61.3	56.0	58.7
	PMM3	94.0	92.0	93.3	90.7	87.0	92.7	96.0	94.7	92.0	97.3
0.5* β_1	LME					13.7					
	PMM1					85.5					
	PMM2	85.4	81.7	75.8	70.1	70.1	67.1	57.1	52.6	42.8	35.1
	PMM3	94.3	90.3	90.0	88.9	86.7	92.5	88.4	85.5	83.1	82.0
1* β_1	LME					18.2					
	PMM1					80.9					
	PMM2	89.2	83.6	76.4	64.2	62.4	67.5	56.4	42.9	38.6	38.6
	PMM3	95.3	91.2	87.2	78.4	82.1	89.2	82.7	77.9	80.7	79.1
1.5* β_1	LME					19.6					
	PMM1					79.2					
	PMM2	89.2	76.5	70.7	65.5	59.3	65.0	53.0	43.2	37.7	33.8
	PMM3	93.8	88.6	83.8	80.6	78.1	89.0	79.4	80.0	76.6	79.5
2* β_1	LME					21.0					
	PMM1					78.9					
	PMM2	87.1	79.5	72.9	63.0	62.9	68.0	55.3	46.0	41.0	35.1
	PMM3	93.8	89.3	84.3	77.4	78.7	87.8	83.3	79.8	77.2	75.8

$S_e=0.5*\beta_1$ and total dropout=60%

Table 3.4 Percent relative bias by anchor, precision of anchor, S_b , total dropout and model.

S_b	% drop out	Model	Anchor 1				Anchor 2			
			100% MNAR	75% MNAR	50% MNAR	25% MNAR	100% MNAR	75% MNAR	50% MNAR	25% MNAR
$1*\beta_1$	20%	LME				-6.0				
	20%	PMM1				-2.3				
	20%	PMM2	-0.4	-1.2	-1.9	-3.3	-0.2	-2.4	-4.1	-6.4
	20%	PMM3	-0.4	-1.0	-1.0	-1.4	-0.1	-1.5	-2.3	-3.5
	40%	LME				-9.2				
	40%	PMM1				-3.8				
	40%	PMM2	-1.1	-1.5	-3.6	-6.4	-1.1	-3.5	-6.0	-7.7
	40%	PMM3	-1.1	-1.0	-2.2	-3.3	-1.2	-2.1	-3.1	-3.3
	80%	LME				-11.1				
	80%	PMM1				-5.5				
	80%	PMM2	-1.3	-3.2	-5.4	-8.6	-1.5	-5.0	-8.0	-9.8
	80%	PMM3	-1.3	-2.4	-3.4	-4.8	-1.5	-3.2	-4.4	-4.4
$2*\beta_1$	20%	LME				-11.4				
	20%	PMM1				-4.6				
	20%	PMM2	-1.0	-2.1	-4.1	-6.5	-1.6	-4.3	-7.5	-9.0
	20%	PMM3	-1.0	-1.7	-2.6	-3.1	-1.6	-2.8	-3.9	-2.8
	40%	LME				-17.3				
	40%	PMM1				-7.7				
	40%	PMM2	-1.1	-4.4	-6.8	-11.4	-1.3	-6.5	-10.7	-15.0
	40%	PMM3	-1.1	-3.5	-4.3	-6.0	-1.4	-4.2	-5.5	-7.0
	80%	LME				-21.2				
	80%	PMM1				-11.1				
	80%	PMM2	-3.0	-6.8	-10.5	-15.3	-3.1	-8.2	-15.8	-19.7
	80%	PMM3	-3.1	-5.4	-6.9	-8.3	-3.1	-5.2	-9.0	-10.0

$S_e=0.5*\beta_1$ and %MAR with Anchor=25%

Table 3.5 Standard error by anchor, precision of anchor, S_b , total dropout and model.

S_b	% drop out	Model	Anchor 1				Anchor 2			
			100% MNAR	75% MNAR	50% MNAR	25% MNAR	100% MNAR	75% MNAR	50% MNAR	25% MNAR
$1*\beta_1$	20%	LME					0.25			
	20%	PMM1					0.29			
	20%	PMM2	0.29	0.28	0.28	0.27	0.29	0.28	0.27	0.26
	20%	PMM3	0.29	0.29	0.28	0.27	0.29	0.29	0.28	0.27
	40%	LME					0.27			
	40%	PMM1					0.32			
	40%	PMM2	0.32	0.31	0.30	0.29	0.32	0.31	0.30	0.28
	40%	PMM3	0.32	0.32	0.31	0.29	0.32	0.31	0.30	0.29
	80%	LME					0.30			
	80%	PMM1					0.37			
	80%	PMM2	0.37	0.35	0.34	0.32	0.37	0.36	0.33	0.32
	80%	PMM3	0.37	0.36	0.34	0.33	0.37	0.36	0.34	0.32
$2*\beta_1$	20%	LME					0.50			
	20%	PMM1					0.55			
	20%	PMM2	0.56	0.55	0.54	0.52	0.56	0.54	0.53	0.50
	20%	PMM3	0.56	0.55	0.54	0.53	0.56	0.55	0.54	0.52
	40%	LME					0.52			
	40%	PMM1					0.63			
	40%	PMM2	0.61	0.59	0.57	0.55	0.61	0.59	0.57	0.54
	40%	PMM3	0.61	0.60	0.58	0.56	0.61	0.59	0.57	0.56
	80%	LME					0.57			
	80%	PMM1					0.68			
	80%	PMM2	0.69	0.66	0.64	0.61	0.69	0.67	0.64	0.60
	80%	PMM3	0.69	0.67	0.64	0.62	0.69	0.67	0.65	0.61

$S_e=0.5*\beta_1$ and %MAR with Anchor=25%

Table 3.6 Coverage of the 95% CI by anchor, precision of anchor, S_b , total dropout and model.

S_b	% drop out	Model	Anchor 1				Anchor 2			
			100% MNAR	75% MNAR	50% MNAR	25% MNAR	100% MNAR	75% MNAR	50% MNAR	25% MNAR
$1*\beta_1$	20%	LME				56.8				
	20%	PMM1				90.9				
	20%	PMM2	95.0	94.3	91.6	69.7	97.6	93.6	76.3	51.7
	20%	PMM3	95.0	94.5	93.4	93.0	97.6	95.2	85.9	74.1
	40%	LME				27.0				
	40%	PMM1				83.9				
	40%	PMM2	94.3	92.8	86.3	61.3	92.0	82.4	62.7	43.4
	40%	PMM3	94.8	92.8	93.0	87.0	92.0	86.4	84.0	86.0
	80%	LME				19.8				
	80%	PMM1				80.2				
	80%	PMM2	94.0	88.0	76.3	47.5	95.2	82.4	54.0	40.7
	80%	PMM3	94.2	90.0	88.5	83.3	95.2	90.4	86.7	81.3
$2*\beta_1$	20%	LME				59.1				
	20%	PMM1				90.0				
	20%	PMM2	94.5	94.3	88.5	71.6	92.0	92.0	80.6	50.0
	20%	PMM3	94.1	94.9	91.5	92.0	92.0	95.2	91.4	60.0
	40%	LME				29.9				
	40%	PMM1				83.2				
	40%	PMM2	96.0	89.3	88.0	66.9	94.4	85.6	74.7	42.4
	40%	PMM3	96.4	91.6	92.6	86.9	93.6	94.4	90.0	82.6
	80%	LME				20.3				
	80%	PMM1				76.8				
	80%	PMM2	94.1	87.6	78.3	51.4	92.8	82.4	53.3	32.0
	80%	PMM3	93.7	91.4	86.9	81.3	93.6	89.6	86.7	75.3

$S_e=0.5*\beta_1$ and %MAR with Anchor=25%

CHAPTER 4: Impact of Interaction Between Informative Dropout and a Prognostic Factor on Estimating Change Over Time in Longitudinal Studies

4.1 Introduction

Longitudinal studies involve a series of measurements over time on the same individual or observational unit, allowing for the direct study of change over time. Over the course of a longitudinal study it is common for subjects to miss visits or to drop out before the scheduled end of follow-up. The effect of missing data on the estimate of the rate of change depends on the underlying missing data mechanism. [6, 13] *Informative dropout* is used to describe the scenario where dropout depends on the unobserved outcomes. In this case, the analysis must consider both the longitudinal nature of the data and the missing data mechanism.

Many methods have been developed to analyze longitudinal data when informative dropout is suspected, but few have examined scenarios with a heterogeneous missing data mechanism, that is, when the relationship between outcome and dropout is not the same for all subjects. Ten Have et al. proposed a selection model to analyze ordinal outcome data when the data are subject to multiple sources of informative dropout (death and unknown loss-to-follow up), assuming that reasons for dropout are fully understood. [43] Selection models explicitly specify the informative dropout process and incorporate it into the model for longitudinal change. [22]The authors specify separate models for

dropout for each of the sources of dropout. Crawford et al. proposed a test for detecting multiple sources of informative dropout. The test is based on a random-effects model relating dropout time and change in outcome, and accommodates the situation where reasons for dropout are not fully known. However, they do not provide a method to estimate change over time in the presence of multiple sources of informative dropout. [44] A number of models have been proposed to jointly model longitudinal and survival data in the presence of competing causes of dropout. [45-49] Such an approach is useful when the time to event is of primary interest, particularly the effect of longitudinal measurements on the binary outcome, and when the causes of dropout are well defined.

It is unclear how to estimate change over time with multiple sources of dropout and unclear or unknown dropout reasons. In Chapter 3, we investigated the use of an anchor event to improve the estimate of change over time. We define *Anchor Event* as an event or dropout reason that informs the researcher as to whether or not the dropout was related to study outcome. This event will aid the researcher in categorizing dropouts in order to separately model different dropout mechanisms. In the motivating example we examined the effect of using total knee replacement (TKR) as an anchor event to improve the estimate of progression of knee osteoarthritis (OA). Heterogeneity in the dropout mechanism may arise from an underlying relationship between a prognostic factor and dropout, for example, younger subjects may drop out from a clinical study as their outcome improves and they feel the intervention is no longer necessary,

while older subjects may drop out as outcome worsens and they are unable to get back and forth to study visits. Thus, there is a heterogeneous missing data mechanism that is different for different levels of a prognostic factor. The interaction could be qualitative, patients in one group are likely to dropout as outcome improves while patients in another group are likely to dropout as outcome deteriorates, or quantitative, both groups drop out as outcome deteriorates, but the association between deterioration and dropout is not the same in both groups. In Chapter 3 we proposed an update to the pattern mixture modeling approach to handle uncertainty in the anchor event. We will extend the use of this method to account for interaction between the prognostic factor of the outcome and dropout mechanism. We chose pattern mixture models vs. selection or shared parameter models for two reasons: 1) The analysis of interest is in estimating longitudinal change over time. Informative dropouts are a nuisance parameter, not a primary outcome. 2) There is uncertainty in the dropout mechanism. Selection and shared parameter models require explicit modeling of the dropout process, while pattern mixture models do not. Instead, in pattern mixture models the piece of the model specifying the missing data mechanism does not depend on the unobserved outcome. [24]

The objective of this project is to examine the impact of interaction between a prognostic factor of the outcome and dropout in estimating the prognostic factor-specific slopes and the interaction between the prognostic factor and time in a longitudinal study. We will compare four methods for

analyzing change: the linear mixed effects model, and 3 pattern mixture models, each with a different strategy for defining patterns.

4.2 Methods

4.2.1 Overview

Using a simulation study we will compare the relative bias, standard error, and coverage of a linear mixed effects model and three different ways of incorporating the anchor event in the pattern mixture modeling framework. We will evaluate the performance of the models, in terms of relative bias, standard error, and coverage, under a wide range of scenarios including variations in the percent of dropout, the prevalence of the prognostic factor, and variability of the outcome measure.

4.2.2 Notation

Notation was established in Chapter 3, Section 3.2.2 .An additional parameter describes the prognostic factor: let $PF_i = 0$ if subject i does not have the prognostic factor and $PF_i = 1$ if subject i does have the prognostic factor.

4.2.3 Pattern Mixture Models

Pattern mixture models stratify the study population based on the pattern of dropout and separately model each group. The models were described in detail in Chapter 3, Section 3.2.3. Briefly, in pattern mixture modeling the overall

estimate is the weighted average of the group-specific estimates, with the weights equal to the proportion of subjects in each group. [11] The joint distribution of the outcome and missingness is factored as:

$$f(Y_i^o, Y_i^m, R_i | X_i) = f(Y_i^o, Y_i^m | R_i, X_i) f(R_i | X_i)$$

The distribution of the responses is conditional on the missing data pattern. This implies a unique distribution for each dropout group. If we assume a normal distribution for the outcome Y_i ,

$$Y_i | R_i, X_i \sim N(\mu(R_i), \Sigma(R_i)) \quad (4.1)$$

The parameter of interest $\hat{\beta}$, is obtained by averaging over the P missing data patterns:

$$\hat{\beta} = \sum^P \hat{\pi}^{\{p\}} \hat{\beta}^{\{p\}} \quad (4.2)$$

where $\pi^{\{p\}}$ is the proportion of subjects and $\beta^{\{p\}}$ is the conditional parameter estimate in the pth pattern.

4.2.3.1 Strategies for Defining Patterns

Strategies for defining patterns were described in detail in Chapter 3, Section 3.2.3.1. They are reviewed briefly below.

4.2.3.1.1 Pattern Mixture Model 1 (PMM1)

Subjects are grouped together based on time of dropout, with no accommodation for dropout reason. Completers are grouped together in a final pattern. Again assuming a normal distribution for the outcome Y_i , we have a different distribution for each dropout group:

$$Y_i | D_i, X_i \sim N(\mu(D_i), \Sigma(D_i))$$

4.2.3.1.2 Pattern Mixture Model 2 (PMM2)

The second approach uses an anchor event (e.g., death, relapse, hospitalization) to determine which dropouts are informative and assumes that all other dropouts are missing completely at random (MCAR), that is, independent of outcome, or missing at random (MAR), that is independent of unobserved outcome, but may depend on observed outcome or covariates. The patterns are defined based on time of dropout for those subjects with an anchor event; completers and MAR/MCAR dropouts are grouped together in a final pattern. The outcome Y_i is normally distributed conditional on an anchor event time and covariates.

$$Y_i | A_i, X_i \sim N(\mu(A_i), \Sigma(A_i))$$

4.2.3.1.3 Pattern Mixture Model 3 (PMM3)

We propose a third strategy for grouping dropouts. In this strategy, dropouts with an anchor event are grouped based on the time of dropout. Another group contains all other non-anchor dropouts, and a final group contains all completers. This ensures that there is at least one pattern with an unbiased estimate of $\hat{\beta}^{(p)}$, the completers. The bias of the estimates for the other patterns will depend on the precision of the anchor event, as it does for the PMM2 approach.

4.2.3.2 Incorporating Prognostic Factor

We allowed the dropout distribution to be a function of the prognostic factor, allowing for each pattern to have different intercepts and slopes depending on prognostic factor. [11, 50] Following from equation (4.1) above

$$Y_i | R_i, X_i \sim N_2 \left(\mu_{PF_i}(R_i), \Sigma_{PF_i}(R_i) \right)$$

$$p(R_i = r | PF_i) = \pi_{PF_i}$$

The estimates of parameters of interest are obtained by averaging over the P missing data patterns, as in equation (4.2), within each level of the prognostic factor:

$$\hat{\beta}_{PF=0} = \sum^P \hat{\pi}_{PF=0}^{\{P\}} \hat{\beta}_{PF=0}^{\{P\}}$$

$$\hat{\beta}_{PF=1} = \sum^P \hat{\pi}_{PF=1}^{\{P\}} \hat{\beta}_{PF=1}^{\{P\}}$$

$$\hat{\beta}_{PF1-PF0} = \sum^P (\hat{\pi}_{PF=1}^{\{P\}} \hat{\beta}_{PF=1}^{\{P\}} - \hat{\pi}_{PF=0}^{\{P\}} \hat{\beta}_{PF=0}^{\{P\}})$$

4.2.4 Simulation Study Details

4.2.4.1 Complete Data Generating Mechanism

For each subject a dichotomous prognostic factor was generating using the Bernoulli distribution, a continuous response variable was generated at baseline and at four fixed follow-up time points. A decline in response indicates clinical worsening. Separately for each level of the prognostic factor, we

generated a vector of correlated responses, $Y_i = (Y_{i0}, Y_{i1}, Y_{i2}, Y_{i3}, Y_{i4})$ for each subject under the linear mixed effect model:

$$Y_i = X_i\beta + Z_ib_i + e_i$$

where

$$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}, b_i = \begin{bmatrix} b_{0i} \\ b_{1i} \end{bmatrix}, X_i = \begin{bmatrix} 1 & t_1 \\ 1 & t_2 \\ \dots & \dots \\ 1 & t_n \end{bmatrix}, Z_i = \begin{bmatrix} 1 & t_1 \\ 1 & t_2 \\ \dots & \dots \\ 1 & t_n \end{bmatrix}$$

and

$$Y_{ij} = \beta_0 + \beta_1 t_{ij} + b_{0i} + b_{1i} t_{ij} + e_{ij}$$

The fixed effect for intercept was 100. This was chosen to avoid a floor effect.

This was modeled after the Osteoarthritis Initiative (OAI), a longitudinal observational study of patients with or at high risk for developing knee OA, where patients that were unlikely to demonstrate measureable loss of joint space during the study, defined as having advanced radiographic OA in both knees, were excluded. [31] The fixed effects (population average) for intercept and slope were

$$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} 100 \\ -5 \end{bmatrix} \text{ for } PF = 0$$

$$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} 100 \\ -10 \end{bmatrix} \text{ for } PF = 1$$

The interaction effect was the difference in change over time between levels of the prognostic factor:

$$\delta = \mu_{PF0} - \mu_{PF1} = (-5) - (-10) = 5$$

In other words, subjects with the prognostic factor decline at twice the rate of subjects without the prognostic factor. In a study of cartilage loss in patients with

knee osteoarthritis (OA), Wirth et al. found that subjects in a high risk sample, defined as subjects with moderate radiographic OA and obesity, lost cartilage at twice the rate as a non-high risk sample over 1 year of follow-up. [51]

Variability was added with random effects for intercept and slope

$$b_i = \begin{bmatrix} b_{0i} \\ b_{1i} \end{bmatrix} \sim N\left(0, \begin{bmatrix} \sigma_0^2 & cc * \sigma_0 \sigma_1 \\ cc * \sigma_0 \sigma_1 & \sigma_1^2 \end{bmatrix}\right)$$

The standard deviation of the random effect for slope, σ_1 was varied at 0.5, 1, and 2 times the average slope and the standard deviation of the random effect for intercept, σ_0^2 , was 15. We chose these parameters to examine the range of change over time that has been reported in the OA literature. For example, in the Wirth et al. study, investigators found rates of cartilage loss ranging from one third to one half of a standard deviation (standardized response mean 0.3 – 0.5). The correlation between b_{0i} and b_{1i} was varied at 0, 0.25, and 0.5 to correspond to small, moderate, and large correlation. In some outcomes of interest in OA, such as joint space width, it is likely that there exists a strong correlation between baseline and slope – patients with low baseline values that have already lost most of the joint space will not have much change over time, simply because they do not have much room to change. In other outcomes, such as pain and function, the relationship between baseline and slope is less clear. We chose these 3 values to examine a range of plausible associations.

Finally, residual variance was generated with an AR(1) correlation structure with correlation $\rho=0.5$ to create a moderate AR structure, and σ_e varied equal to one half of the standard deviation of the random effect for slope:

$$e_{ij} = \begin{bmatrix} e_{i0} \\ e_{i1} \\ e_{i2} \\ e_{i3} \\ e_{i4} \end{bmatrix} \sim N \begin{bmatrix} \sigma_e^2 & \rho\sigma_e^2 & \rho^2\sigma_e^2 & \rho^3\sigma_e^2 & \rho^4\sigma_e^2 \\ \rho\sigma_e^2 & \sigma_e^2 & \rho\sigma_e^2 & \rho^2\sigma_e^2 & \rho^3\sigma_e^2 \\ \rho^2\sigma_e^2 & \rho\sigma_e^2 & \sigma_e^2 & \rho\sigma_e^2 & \rho^2\sigma_e^2 \\ \rho^3\sigma_e^2 & \rho^2\sigma_e^2 & \rho\sigma_e^2 & \sigma_e^2 & \rho\sigma_e^2 \\ \rho^4\sigma_e^2 & \rho^3\sigma_e^2 & \rho^2\sigma_e^2 & \rho\sigma_e^2 & \sigma_e^2 \end{bmatrix}$$

4.2.4.2 Missing Data Mechanism

For each level of the prognostic factor we created a random-effects dependent heterogeneous missing data mechanism, with dropout the result of MNAR and MAR mechanisms. Approximately 70% of the dropouts were not at random, i.e., related to the underlying random effect for slope. We chose this value based on our evaluation of dropouts in the OAI. Approximately 40% of the dropouts underwent TKR. We assumed that in addition to this, some subjects drop out due to an MNAR mechanism and do not have the anchor event, and some subjects dropout due to a different missing data mechanism. We generated the missing data mechanism as follows. First, we created an MNAR mechanism where dropout depends on the underlying random effect for slope. We did this separately for each level of the prognostic factor in order to create a prognostic factor level specific missing data mechanism. First, for each subject we computed the z-score for the random effect for slope, zb_{i1} . This indicates how many standard deviations the subject is above or below the population slope.

Then, for each subject we computed the log odds of dropping out at any point in the study as a function of the z-score of the random effect for slope, zb_{i1} :

$$\text{logit}(\pi_i) = \alpha_{0t} + \alpha_{1t}zb_{i1}$$

α_{0t} and α_{1t} were defined based on the total amount of dropout, which ranged from 25% to 75% of the cohort, and so that each increase in 1 standard deviation of slope (increase in 1 unit in zb_{i1}) was associated with dropout. For subjects without the prognostic factor, each increase in 1 standard deviation of slope was associated with an odds of dropout of 0.5, i.e., as a subject's slope worsened he/she is less likely to dropout, or as a subject's slope improves, he/she is more likely to dropout. For subjects with the prognostic factor, each increase in 1 standard deviation of slope was associated with an increased of odds of dropout of 2, i.e., as a subject's slope worsened he/she was more likely to dropout.

Approximately 30% of the dropouts were at random, i.e., related to observed outcome, in this case, baseline value. To create a MAR mechanism, we first computed the z-score for the random effect for intercept, zb_{i0} . This z-score indicates how many standard deviations the subject is above or below the population baseline value. Then, for each subject log odds of dropping out at any point in the study was modeled as a function of the z-score of the random effect for intercept, zb_{i0} :

$$\text{logit}(\pi_i) = \alpha_{2t} + \alpha_{3t}zb_{i0}$$

α_{2t} and α_{3t} were defined based on the total amount of dropout, which ranged from 25% to 75% of the cohort, and so that each increase in 1 standard deviation

of intercept (increase in 1 unit in zb_{i0}) was associated with an increased of odds of dropout of 1.5.

The preceding steps described how we determined which subjects dropped out. Following this, we determined at which time point subjects dropped out. We did this in two ways:

1. Dropout time was associated with the dropout mechanism and with the random effect for change: subjects with MNAR dropout and with greater worsening were more likely to drop out early, while subjects with MAR dropout were more likely to drop out late.
2. Subjects were randomly assigned a dropout time using the multinomial distribution, with equal dropout at all four time points.

4.2.4.3 Anchor Event

We modeled the anchor event after the non-ignorable threshold model described by Schluchter, et al. [7] We assume that there is a fixed but unknown threshold, and when a subject's outcome first drops below this threshold the subject has the anchor event. Based on the data generating mechanism described above, the mean outcome at time point 4, if no subjects dropped out, would be 70. We chose the threshold to be 1 standard deviation below this (e.g., if $S_b=7.5$ then the threshold was 62.5). Both subjects with and without the prognostic factor are eligible to have the anchor event. Under the definition of a "perfect" anchor given in Chapter 3, Section 3.2.4.3, this anchor is imperfect;

some subjects dropping out due to the MNAR mechanism may never drop below the threshold and will not have the anchor, and some subjects dropping out due to the MAR mechanism may drop below the threshold and have the anchor. However, the anchor event supplies additional information about the dropouts, beyond what we already know by knowing the level of the prognostic factor. The anchor will give us some additional insight into a subject's disease progression, regardless of level of prognostic factor.

The model makes sense in the context of the OA example – for example, men may generally be more likely to dropout as they get better, and women may generally be more likely to dropout as they get worse, but regardless of sex, subjects that reach some threshold of disease worsening will choose to undergo TKR. This threshold may be reached after a long slow decline, or after a short fast decline.

4.2.3.4 Models considered

We compared the performance of four models for longitudinal data in terms of relative bias, standard error, and coverage of the 95% confidence interval. These four models were described in Chapter 3, Section 3.2.4.4 and were modified to include the prognostic factor.

4.2.3.4.1 Linear Mixed Effects Model

We considered a linear mixed model (LME) to estimate prognostic-factor specific slopes and the interaction between prognostic factor and time. The LME has no accommodation for informative dropout:

$$Y_i = X_i\beta + Z_ib_i + e_i$$

Where

$$X_i = \begin{bmatrix} 1 & t_1 & PF & PF * t_1 \\ 1 & t_2 & PF & PF * t_2 \\ 1 & t_3 & PF & PF * t_3 \\ \dots & \dots & \dots & \dots \\ 1 & t_n & PF & PF * t_n \end{bmatrix}, \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix}, Z_i = \begin{bmatrix} 1 & t_1 \\ 1 & t_2 \\ 1 & t_3 \\ \dots & \dots \\ 1 & t_n \end{bmatrix}, b_i = \begin{bmatrix} b_{0i} \\ b_{1i} \end{bmatrix}$$

and

$$Y_{ij} = \beta_0 + \beta_1 t_{ij} + \beta_2 PF_i + \beta_3 t_{ij} * PF_i + b_{0i} + b_{1i} t_{ij} + e_{ij}$$

4.2.3.4.2 Pattern Mixture Models

This section will describe the additional conditions and restrictions necessary for PMM specification and will describe the three PMMs evaluated.

4.2.3.4.2.1 Identifying Restrictions

Additional restrictions are necessary in order to estimate all parameters in the PMMs. We assumed a linear trajectory in each pattern:

$$Y_i^{\{p\}} = X_i^{\{p\}}\beta^{\{p\}} + Z_i^{\{p\}}b_i^{\{p\}} + e_i^{\{p\}}$$

Slope can only be estimated in patterns with at least two observations per subject, thus we combined subjects dropping out after baseline with subjects dropping out after the first time point.

4.2.3.4.2.2 Models

In addition to the LME model, we considered three different PMMs as described in section 3.2.3:

1. PMM1 – ignore the anchor event and group all dropouts together at each time point. Based on the data generating mechanism with five time points (baseline and four yearly follow-ups), this produces four total groups.
 - ▶ $D_{ij} = [1\ 0\ 0\ 0\ 0]^T$ or $D_{ij} = [0\ 1\ 0\ 0\ 0]^T$ (dropout between baseline and year 1, or between year 1 and year 2)
 - ▶ $D_{ij} = [0\ 0\ 1\ 0\ 0]^T$ (dropout between year 2 and year 3)
 - ▶ $D_{ij} = [0\ 0\ 0\ 1\ 0]^T$ (dropout between year 3 and year 4)
 - ▶ $D_{ij} = [0\ 0\ 0\ 0\ 0]^T$ (completers)

2. PMM2 – First separate dropouts based on the anchor event. All completers and dropouts without an anchor will be one group. All remaining dropouts are grouped by time point. Based on the data generating mechanism with five time points (baseline and four yearly follow-ups), this produces four total groups.

- ▶ $A_{ij} = [1\ 0\ 0\ 0\ 0]^T$ or $A_{ij} = [0\ 1\ 0\ 0\ 0]^T$ (dropout between baseline and year 1, or between year 1 and year 2 and anchor)
- ▶ $A_{ij} = [0\ 0\ 1\ 0\ 0]^T$ (dropout between year 2 and year 3 and anchor)
- ▶ $A_{ij} = [0\ 0\ 0\ 1\ 0]^T$ (dropout between year 3 and year four and anchor)
- ▶ $A_{ij} = [0\ 0\ 0\ 0\ 0]^T$ (no anchor)

3. PMM3 – First separate dropouts based on the anchor event. All completers are grouped together, all dropouts without an anchor event are grouped together, all dropouts with an anchor event are grouped together by time point. Based on the data generating mechanism with five time points (baseline and four yearly follow-ups), this results in 5 total groups.

- ▶ $A_{ij} = [1\ 0\ 0\ 0\ 0]^T$ or $A_{ij} = [0\ 1\ 0\ 0\ 0]^T$ (dropout between baseline and year 1, or between year 1 and year 2 and anchor)
- ▶ $A_{ij} = [0\ 0\ 1\ 0\ 0]^T$ (dropout between year 2 and year 3 and anchor)

- ▶ $A_{ij} = [0\ 0\ 0\ 1\ 0]^T$ (dropout between year 3 and year four and anchor)
- ▶ $A_{ij} = [0\ 0\ 0\ 0\ 0]^T$ and D_{ij} not equal to $[0\ 0\ 0\ 0\ 0]^T$ (dropout, no anchor)
- ▶ $A_{ij} = [0\ 0\ 0\ 0\ 0]^T$ and $D_{ij} = [0\ 0\ 0\ 0\ 0]^T$ (completers)

4.2.4.5 Parameters Evaluated

The standard deviation of change, denoted as S_b ($b_{1i} \sim N(0, \sigma_1)$), was examined at one half, one and two times the rate of change (i.e., 3.75, 7.5, 15) and the residual standard deviation, denoted as S_e ($e_{ij} \sim N(0, \sigma_e)$), was equal to one half of S_b . The total dropout by end of study was varied at 25%, 50%, and 75%. The prevalence of the prognostic factor was varied at 25%, 50%, and 75%. The correlation between b_{0i} and b_{1i} (denoted as CC) was varied at 0, 0.25, and 0.5. Finally, two missing data mechanism were implemented as described above.

4.2.4.6 Criteria for Evaluation

We evaluated relative bias, standard error, coverage probability, and length of the 95% confidence interval for the estimate of change over time within each level of the prognostic factor and for the estimate of time by prognostic factor interaction. The absolute bias is calculated by subtracting the estimated progression estimate from the fixed true value and the relative bias is calculated by dividing the absolute bias by the fixed true value of the parameter estimate.

The coverage probability is the proportion of times the 95% confidence interval includes the true progression estimate. We report the number of times that we were unable to compute a progression estimate due to not having enough subjects in each group. Five thousand replicates were run for each scenario. All simulations were conducted in SAS version 9.3 (SAS Institute, Cary, NC).

4.3. Results

4.3.1 Missing Data Mechanism 1 (Association between outcome and dropout time)

4.3.1.1 Relative Bias

4.3.1.1.1 Prognostic Factor by Time Interaction

We first present results for missing data mechanism 1, the scenario with dropout time associated with the dropout mechanism and with the random effect for change. Figure 4.1 displays relative bias by model and total dropout, holding S_b , the standard deviation of change ($b_{1i} \sim N(0, Sb)$), fixed at the average slope, the prevalence of the prognostic factor at 50%, and the correlation between b_{0i} and b_{1i} at 0.25. Relative bias increases as the amount of dropout increases, and is largest for the LME model. Relative bias for the PMM1 and PMM2 models ranges between -5 and -12%, while relative bias for the PMM3 model stays close to -5%. The interquartile range (IQR) of observed relative bias is similar across all four models, approximately 12%.

As we increase the variability in outcome by increasing S_b and S_e , relative bias increases for all four models (Figure 4.2). As S_b is increased from $0.5\beta_1$ to $2\beta_1$ (and S_e correspondingly increased from $0.25\beta_1$ to $1\beta_1$), relative bias increases from approximately -12% to -50% for the LME model, and -3% to -5% for PMM3. As demonstrated by the widths of the box plots, the variability around the relative bias estimate increases substantially as well, with the IWR increasing from approximately 6% to 25% for all models.

There did not appear to be an association between mean relative bias and prevalence of the prognostic factor (Figure 4.3). Finally, mean relative biased decreases slightly with increasing correlation between b_0 and b_1 (Figure 4.4).

4.3.1.1.2 Relative Bias of Prognostic Factor-Specific Slopes

Figure 4.5 displays relative bias of the prognostic factor-specific slopes, by model and total dropout, holding S_b fixed at the average slope, the prevalence of the prognostic factor at 50%, and the correlation between b_{0i} and b_{1i} at 0.25. In general, the slope for PF=0 overestimated (positive relative bias) while the slope for PF=1 is underestimated (negative relative bias). The PMM1 and PMM3 models have the least relative bias in estimating the slope for PF=0, while the PMM2 and PMM3 models have the least relative bias in estimating the slope for PF=1.

Within each level of the prognostic factor, there is less variability in the estimate of mean relative bias as the percent of subjects in each group increases

(Figure 4.6). For PF=0, the width of the boxplots increases as more subjects have the prognostic factor (and there are fewer subjects in the PF=0 group), with for PF=1 the width of the boxplots increases as there are fewer subjects with the prognostic factor.

4.3.1.2 Standard Error

4.3.1.2.1 Prognostic Factor by Time Interaction

Standard error is lowest for the LME model and similar for the 3 PMMs, with slightly higher standard error for PMM1. Holding S_b fixed at the average slope, the prevalence of the prognostic factor at 50%, and the correlation between b_{0i} and b_{1i} at 0.25, the standard error of the estimate increases as we increase total dropout from 25% to 75% (Figure 4.7).

As expected, standard error increases as S_b and S_e increase, and as the prevalence of the prognostic factor changes from 50% (Figures 4.8-4.9).

4.3.1.2.2 Standard Error of Prognostic Factor-Specific Slopes

There is slightly higher mean standard error and more variability around the standard error estimate in estimating the slope for PF=1 vs. PF=0 (Figure 4.10). Within each level of the prognostic factor, there is larger standard error as the percent of subjects in that group decreases (Figure 4.11)

4.3.1.3 Coverage

4.3.1.3.1 Prognostic Factor by Time Interaction

PMM3 maintains coverage close to 95% across all amounts of dropout, in the scenario with S_b fixed at the average slope, the prevalence of the prognostic factor at 50%, and the correlation between b_{0i} and b_{1i} at 0.25 (Figure 4.12). The coverage of PMM1 and PMM2 decreases with increasing dropout, though coverage is maintained above 80%. Coverage was the lowest for the LME model, dropping from approximately 50% at 25% total dropout to 15% at 75% total dropout.

4.3.1.3.2 Prognostic Factor-Specific Slopes

Paralleling the mean relative bias for prognostic factor-specific slopes, the PMM1 and PMM3 models have the highest coverage in estimating the slope for PF=0, maintaining coverage close to 95%, while the PMM2 and PMM3 models have the highest coverage in estimating the slope for PF=1. This result holds across all levels of dropout and prevalence of the prognostic factor (Figures 4.13-4.14).

4.3.2 Impact of Missing Data Mechanism

The relative bias in estimating the prognostic factor by time interaction was substantially smaller for the LME, PMM1, and PMM3 models under missing data mechanism 2 vs. missing data mechanism 1 (Table 4.1). In fact, under missing data mechanism 2 relative bias was under 2% for both the PMM1 and PMM3 models, for all scenarios examined. Relative bias ranged as high as -11%

for PMM3 and -27% for PMM1 under missing data mechanism 1. While the relative bias for PMM2 was smaller under missing data mechanism 2, it was not substantially smaller. Under missing data mechanism 2, the PMM1 and PMM3 models clearly outperform PMM2 in terms of relative bias. The variability around the relative bias estimates was similar for both mechanisms (Figure 4.15). Figure 4.16 displays the prognostic-factor specific slopes for each missing data mechanism. The LME and PMM2 models overestimate the slope for PF=0, while the PMM1 and PMM3 models have mean bias close to 0%. Under missing data mechanism 1, all models underestimate the slope for PF=1, with the LME and PMM1 models displaying the most bias. Under missing data mechanism 2, all three pattern mixture models have relative bias close to 0%. Standard error was approximately the same between the missing data mechanisms while coverage mirrored relative bias and was smaller for the LME, PMM1, and PMM3 models under missing data mechanism 2 vs. missing data mechanism 1 (Tables 4.2-4.3).

4.4 Discussion

We evaluated the impact of an interaction between the missing data mechanism and a prognostic factor in evaluating the rate of change in longitudinal studies with informative dropouts using PMMs. In estimating the prognostic factor by time interaction, mean relative bias increased and coverage decreased with increasing variability in outcome and total dropout, while the prevalence of the prognostic factor did not have an impact on mean relative bias

or coverage. Our proposed update to the grouping of dropout patterns had the lowest relative bias across all four models evaluated, while the estimated standard error was larger than that estimated from the LME model. In estimating prognostic-factor specific slopes, the SE was higher for the PMM3 model than the LME model, and higher than the PMM2 model under certain conditions.

We evaluated four methods for estimating the rate of change in longitudinal studies. In Chapter 3 we demonstrated that the PMM3 model had the best performance, in terms of relative bias and coverage, in estimating change over time in a scenario with a heterogeneous missing data mechanism and imprecise anchor event. When the amount of variability was large or the total amount of dropout was large, the LME and PMM1 models demonstrated large relative bias and low coverage, while the performance of the PMM2 model depended on the precision of the anchor event. In this analysis, we found that the PMM3 performed best in terms of bias and coverage, while the standard error was large under certain conditions, particularly in estimating the prognostic factor-specific slopes.

The performance of PMM1 and PMM2 depended on the underlying missing data mechanism and the parameter of interest. Under missing data mechanism 1, the scenario with time of dropout associated with underlying change, the PMM1 model had less bias and higher coverage than the PMM2 model in estimating the slope for the subgroup without the prognostic factor. In our scenarios, subjects with the prognostic factor were more likely to dropout as

outcome worsened, and were therefore more likely to reach the anchor event threshold. The PMM2 model does better than the PMM1 model for patients with the prognostic factor because it incorporates the anchor event. In estimating the prognostic factor by time interaction, the PMM2 model had lower relative bias and higher coverage than the PMM1 model under missing data mechanism 1 but had higher relative bias and lower coverage under missing data mechanism 2. Under mechanism 1, patients with more rapid decline in outcome (larger random effect for slope) were more likely to dropout out of the study early. The assumption we made in our identifying restrictions, that the subjects dropping out after baseline had the same slope and could be combined with subjects dropping out after the first time point, does not hold. The anchor event provides some additional information about which dropouts have the worst outcome and should be modeled together and helps us overcome the violation of this assumption. Under mechanism 2, on the other hand, all dropouts were equally likely to dropout at each time point, regardless of mechanism. The assumption that subjects dropping out after baseline can be combined with subjects dropping out after the first time point holds. Here, simply knowing the prognostic factor tells us everything we need to know in order to model the different dropout mechanisms.

This paper examined the situation where multiple sources of informative dropout are present, and these sources are different for different levels of a baseline prognostic factor. Of particular concern is the case of qualitative interaction, where the different directions of informativeness (some subjects

dropping out as outcome improves, others dropping out as outcome worsens) may mask the informativeness of the missing data mechanism and increase the chance that this informative dropout is not detected in standard tests for MCAR or MAR dropout. The test proposed by Crawford et al. can be used to investigate whether there are multiple, heterogeneous, sources of informative dropout. [44] Investigators can also look at the association between anchor event and baseline prognostic factors – if some subgroup is more likely or less likely to undergo anchor event, then it may be an indication that the dropout mechanism is different for that subgroup.

As with any study that utilizes simulations, our study had several limitations. Missing data in our study resulted only from dropout; there were no intermittent missing data. We examined only two missing data mechanisms and did not vary the percent of dropouts that were due to the MNAR vs. MAR mechanism. We only examined scenarios with 1000 subjects and 5 time points. The PMM described required at least one dropout group per time point per level of the PF. With a smaller sample it is possible that groups would be too small to estimate pattern and PF specific slopes. Even with the large sample size that we examined, the standard error for the PMMs was higher than the LME. Decisions about how to group sparse dropout patterns were not examined in this analysis.

In this paper we compared the performance of four different models for evaluating the rate of change in longitudinal studies when there is a heterogeneous informative dropout mechanism and interaction between the

dropout mechanism and prognostic factor. The approach proposed in Chapter 3, which separates completers from dropouts that do not have an anchor event, had the superior performance across the four models in terms of relative bias and coverage. However, estimating pattern-specific slopes for many patterns comes with a drawback in terms of the large standard error. There is more uncertainty when the model estimating many parameters. The model had good performance across varying degrees of dropout and prevalence of the prognostic factor and performed especially well when the dropout time was not associated with outcome.

Figure 4.1 Relative bias by model and total dropout.

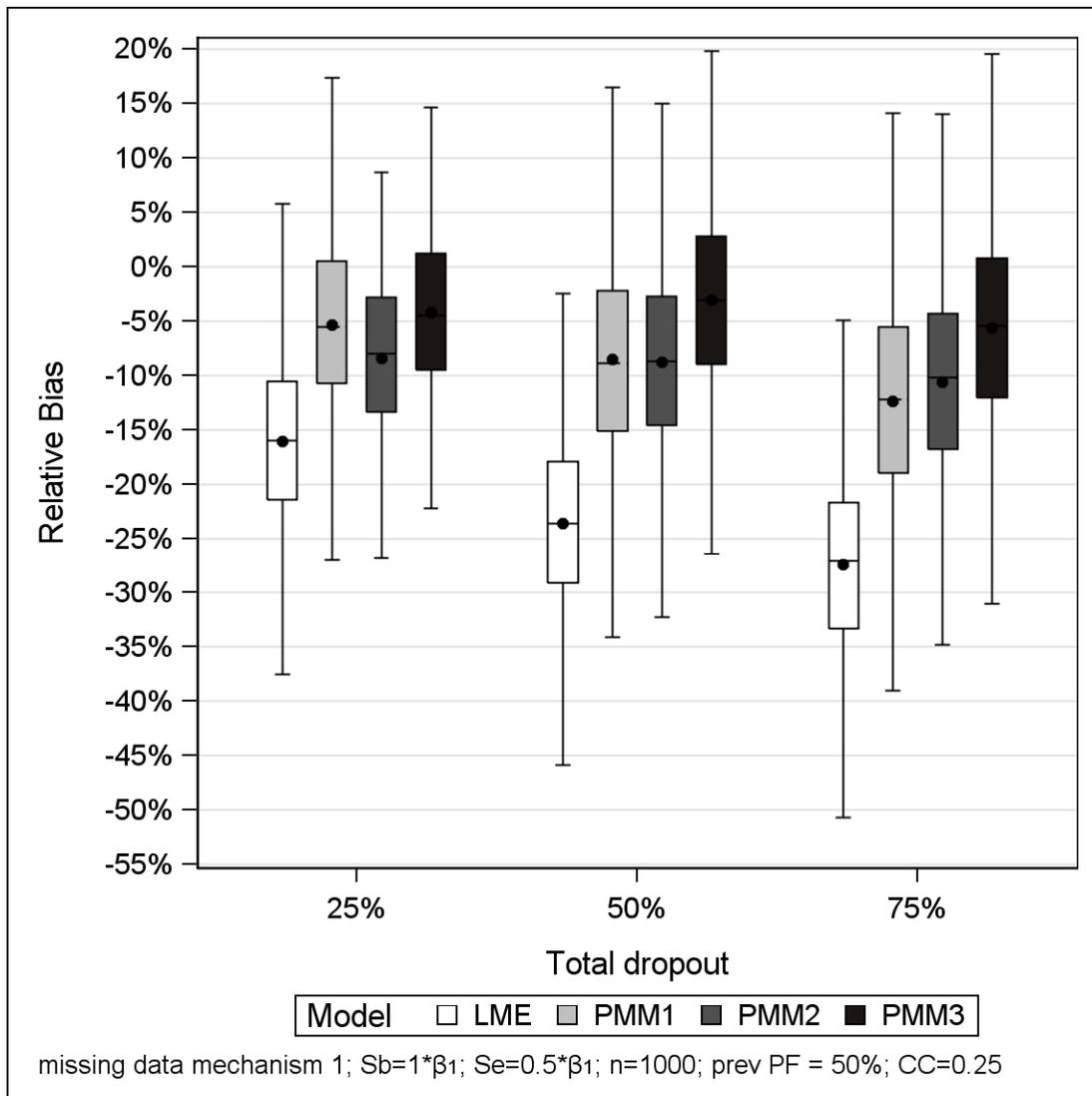


Figure 4.2 Relative bias by model and variability.

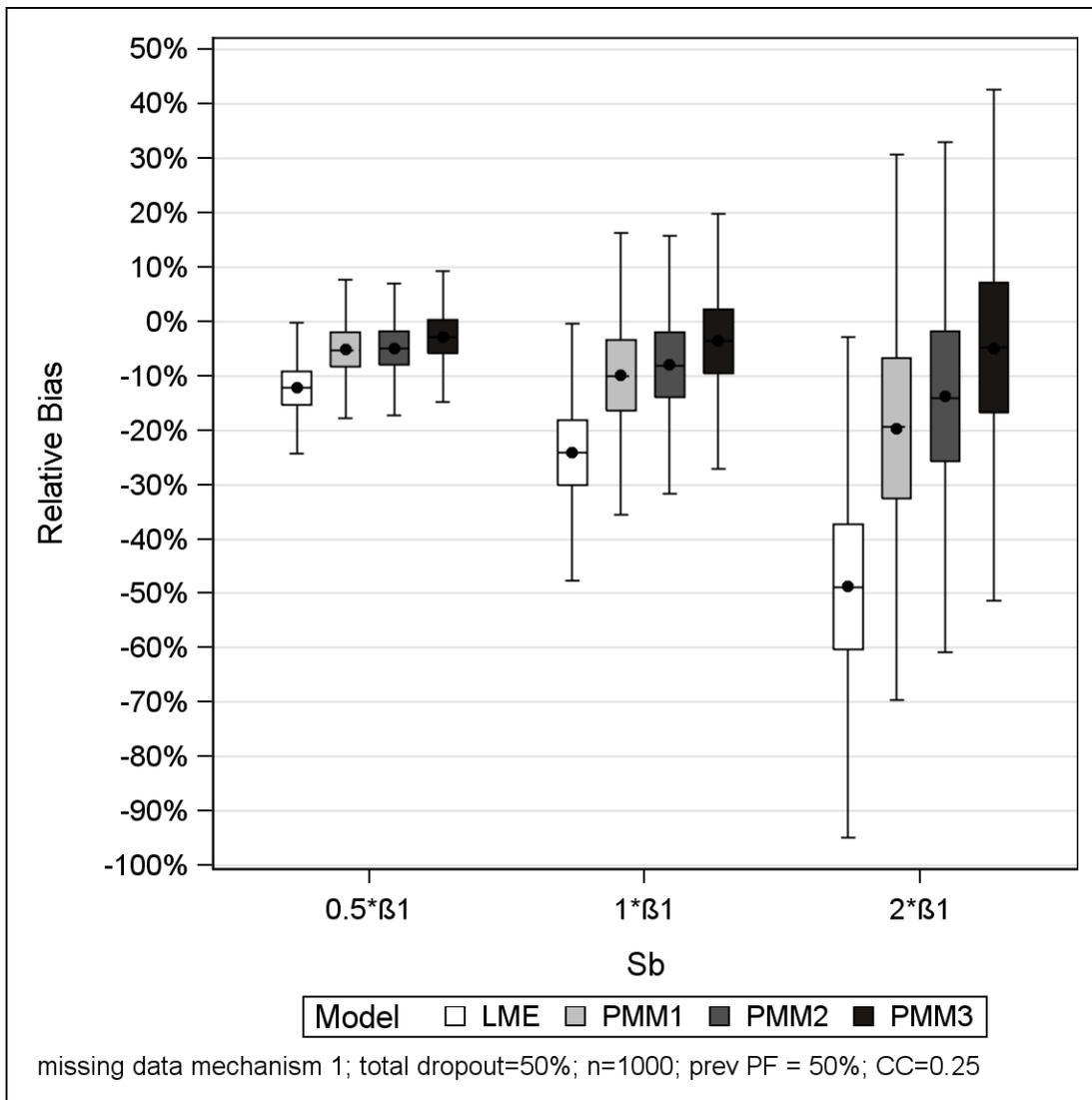


Figure 4.3 Relative bias by model and prevalence of PF.

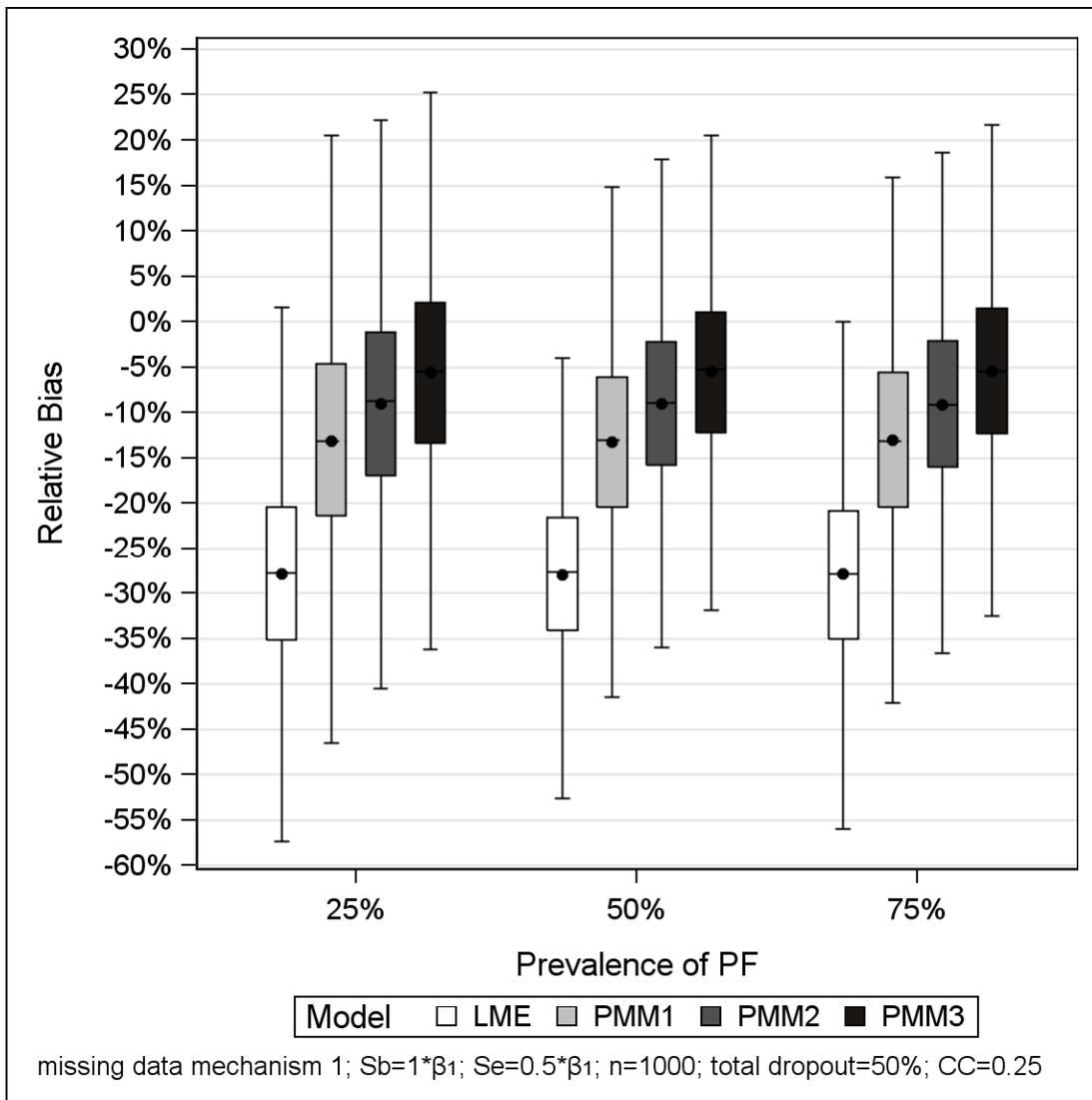


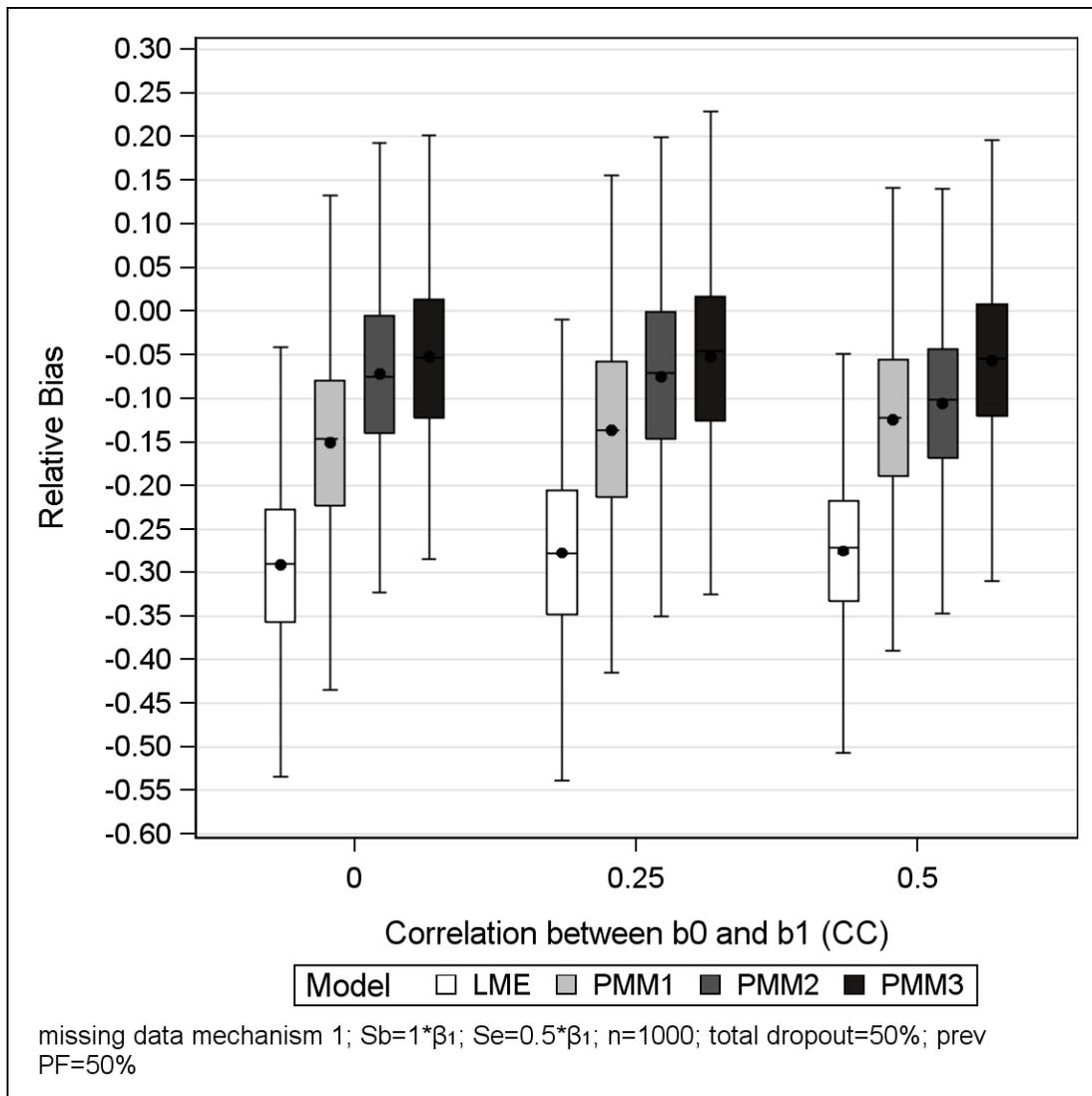
Figure 4.4 Relative bias by model and correlation between b_0 and b_1 (CC).

Figure 4.5 Relative bias of PF-specific slopes by model and total dropout.

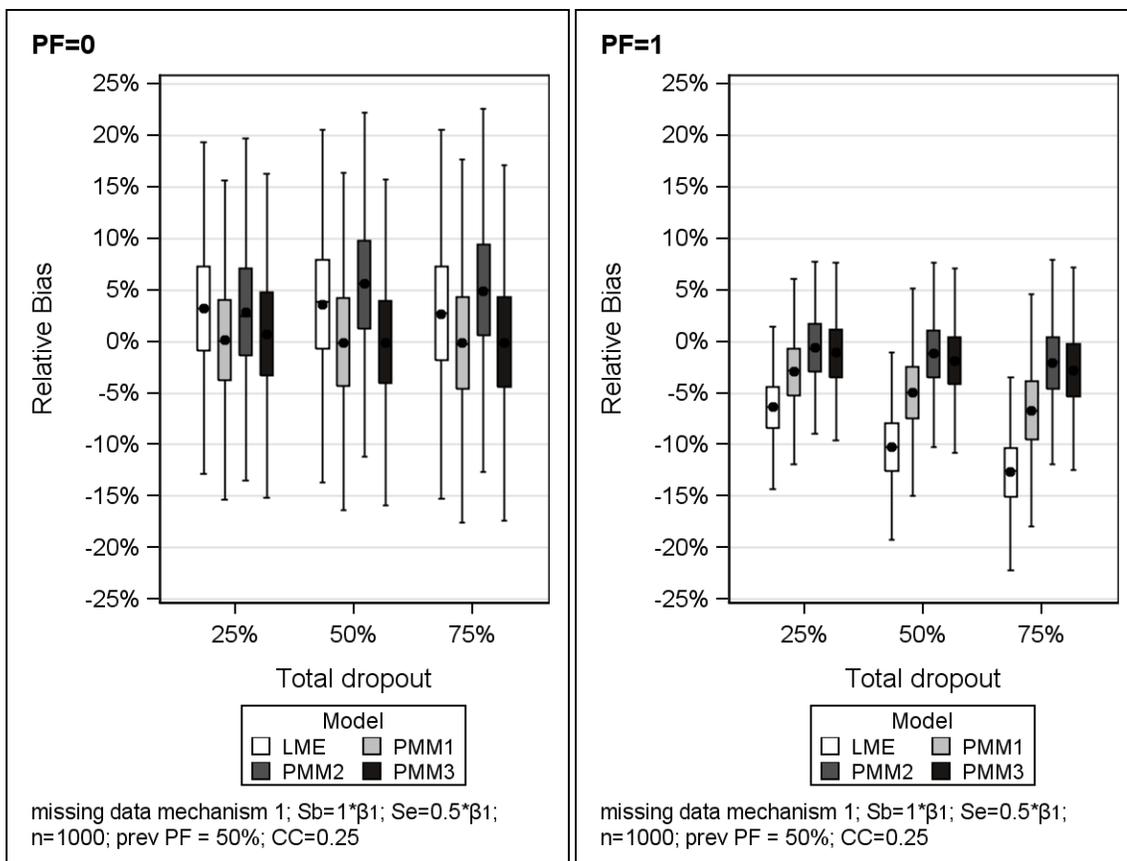


Figure 4.6 Relative bias of PF-specific slopes by model and prevalence of PF.

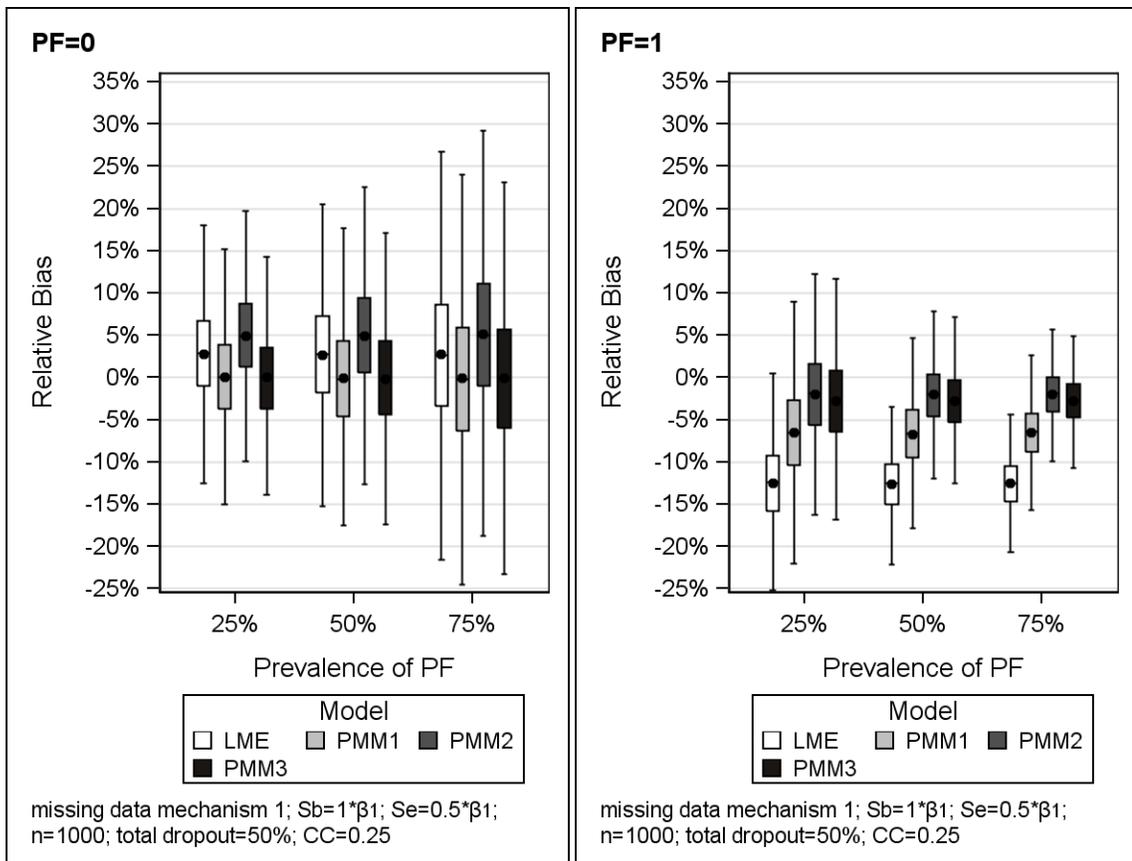


Figure 4.7 Standard error by model and total dropout.

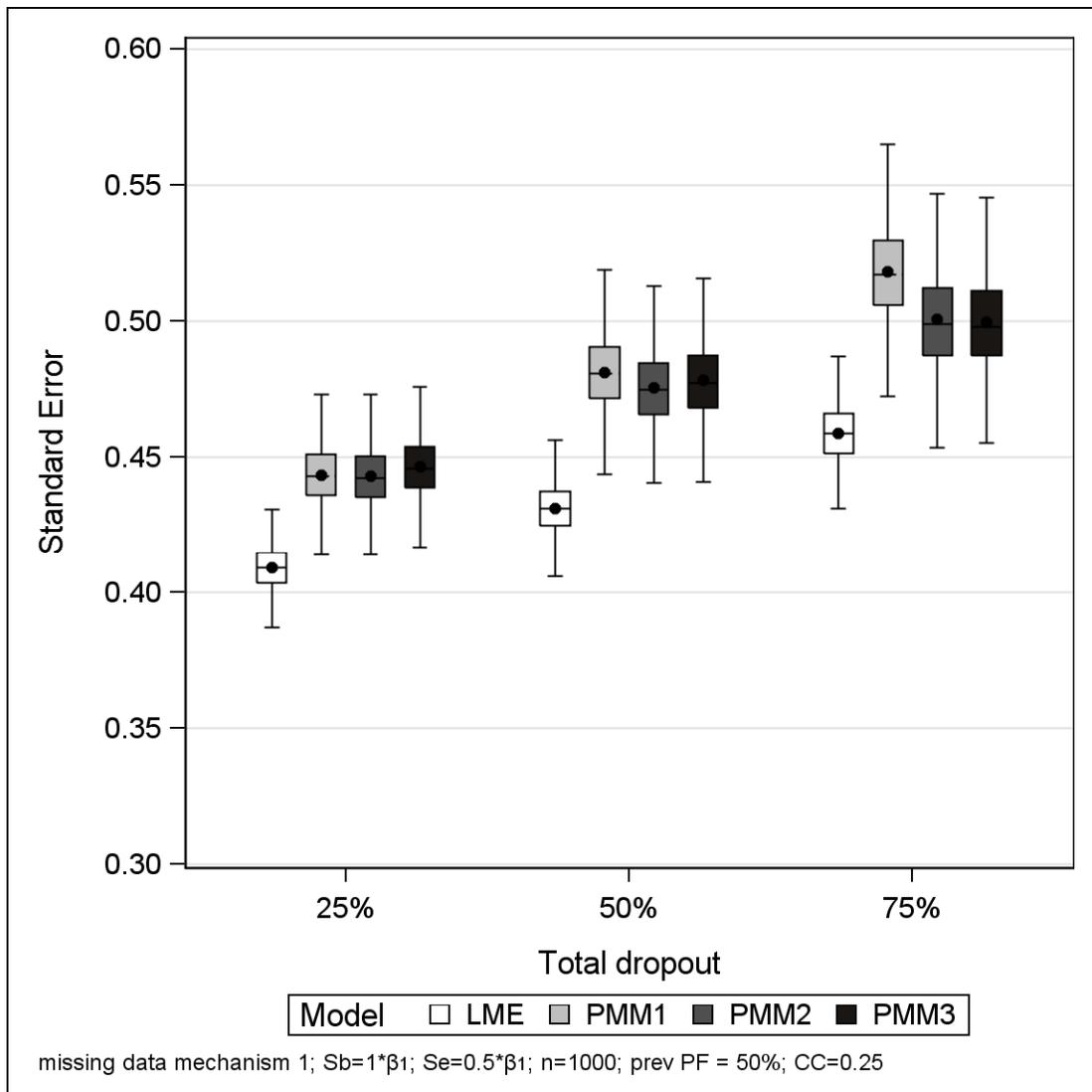


Figure 4.8 Standard error by model and variability.

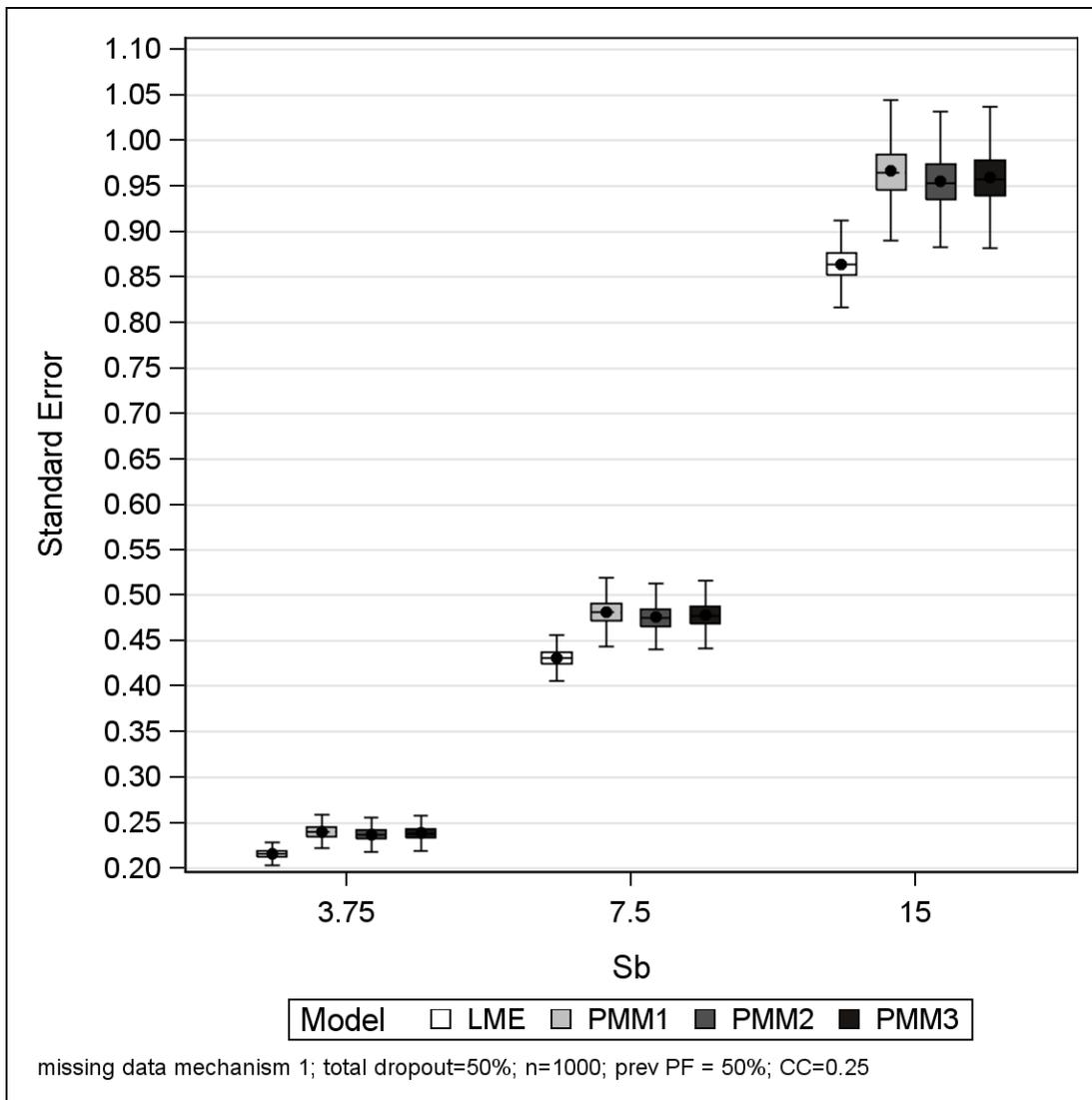


Figure 4.9 Standard error by model and prevalence of PF.

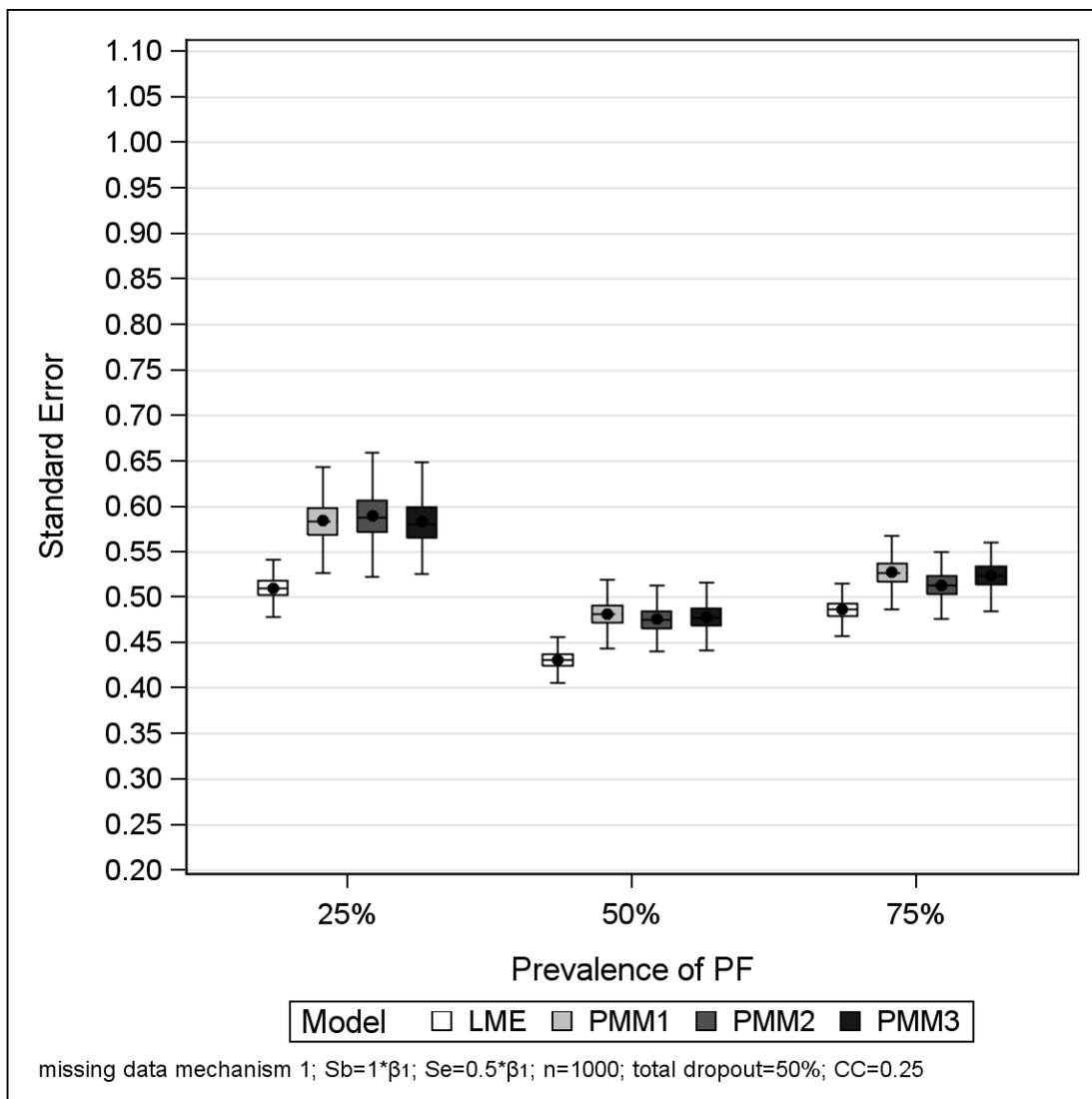


Figure 4.10 Standard error of PF-specific slopes by model and total dropout.

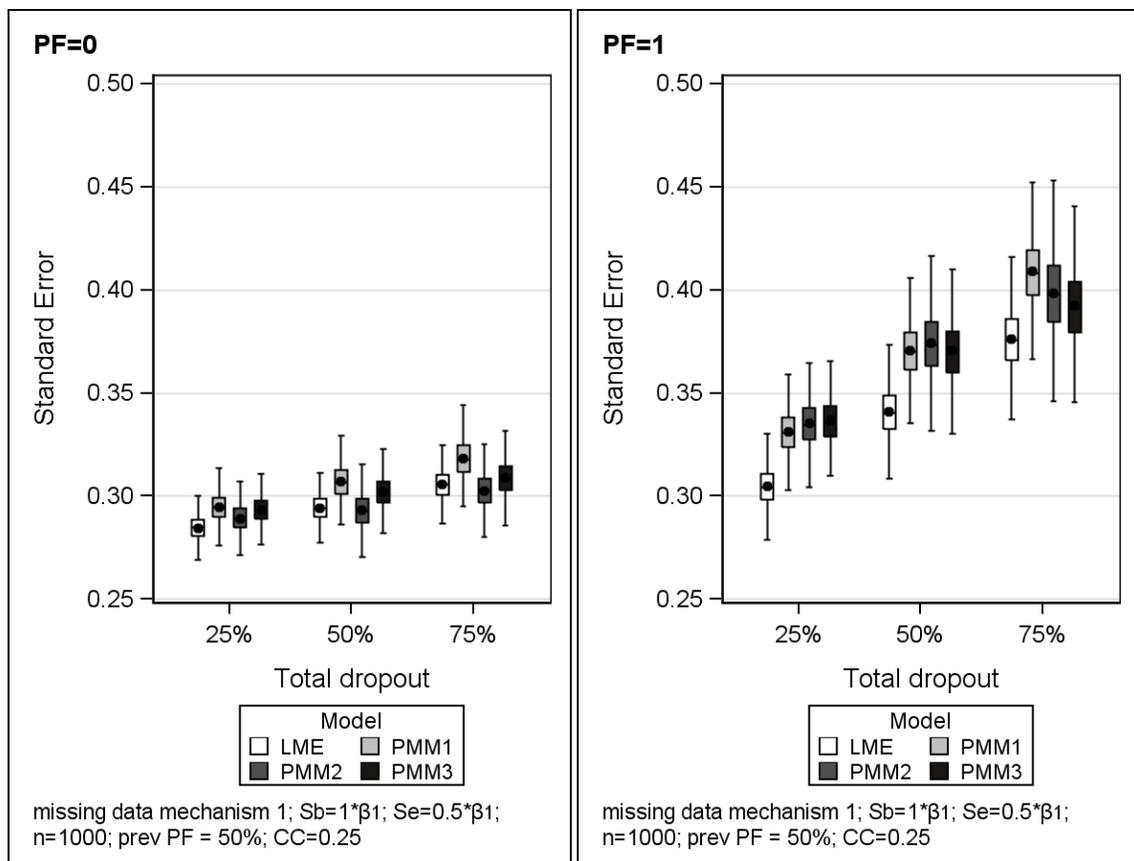


Figure 4.11 Standard error of PF-specific slopes by model and prevalence of PF.

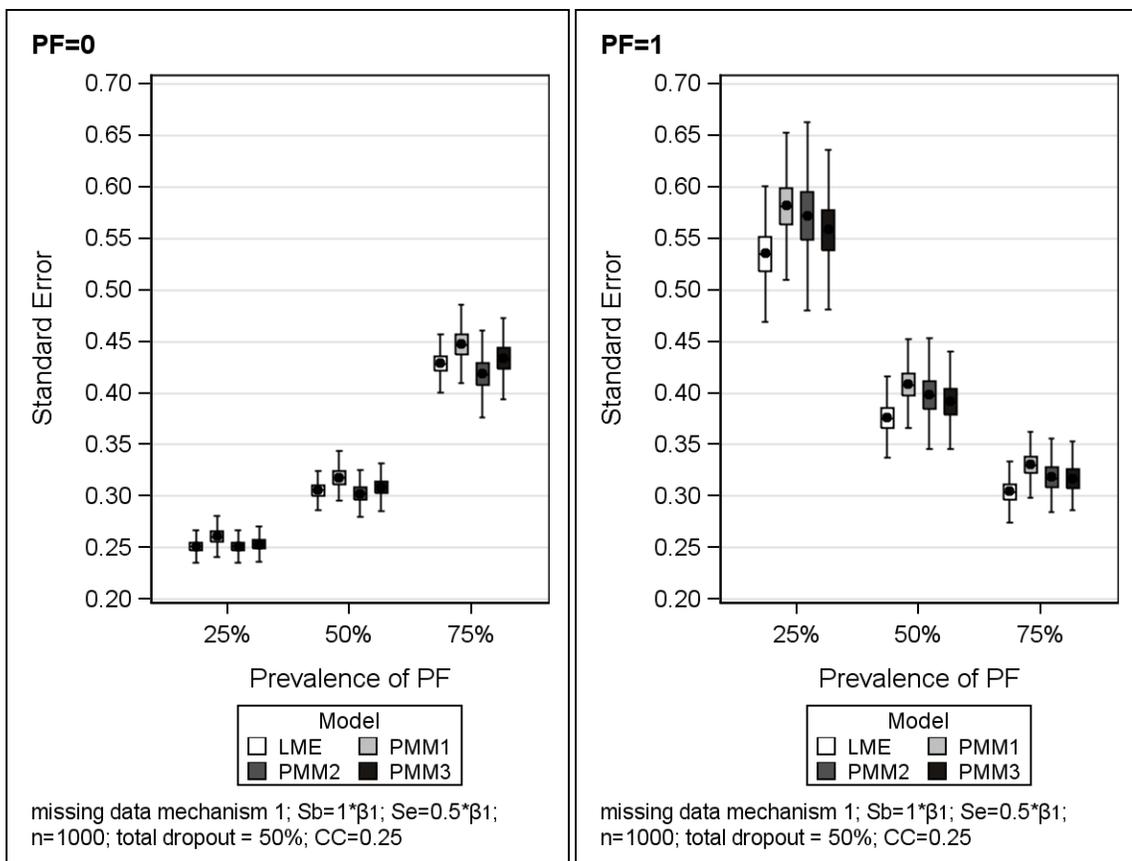
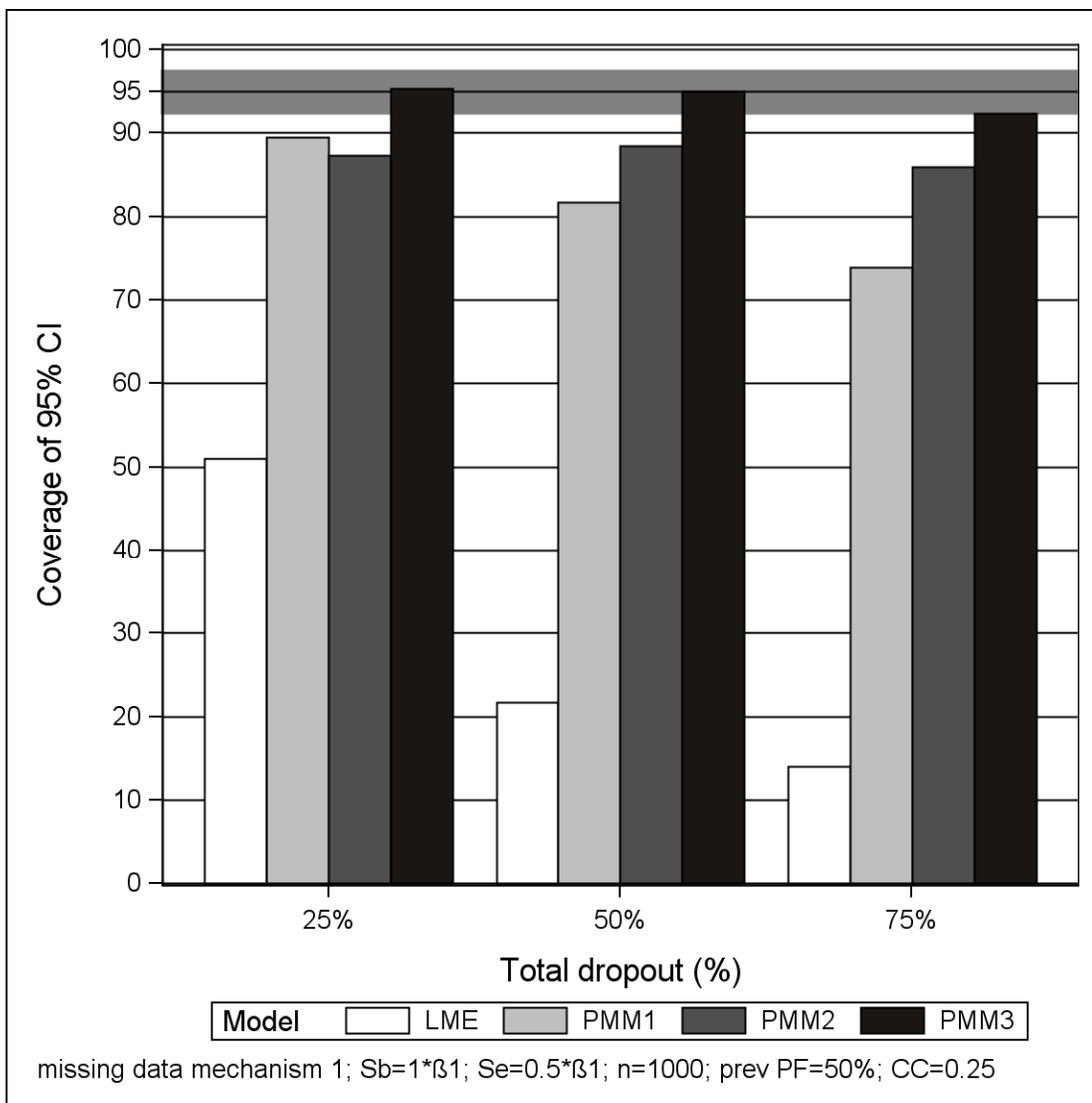
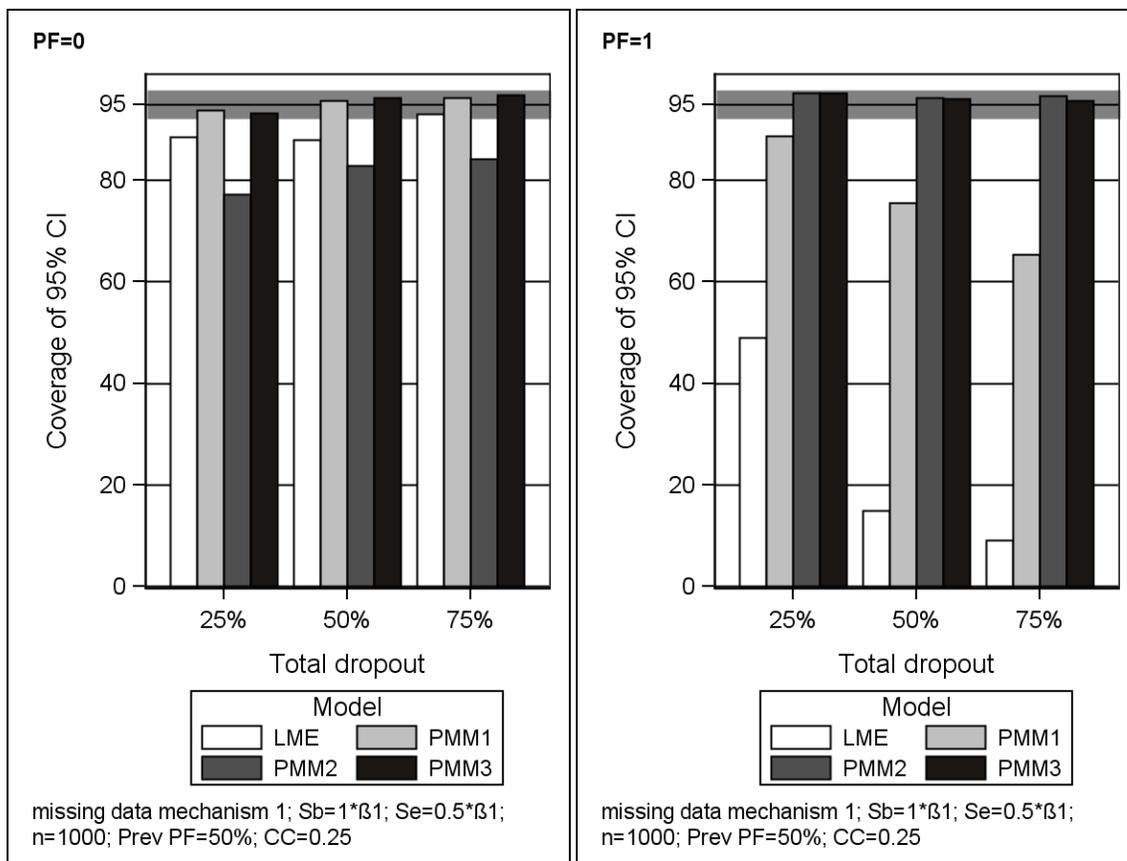


Figure 4.12 Coverage of 95% CI by model and total dropout.



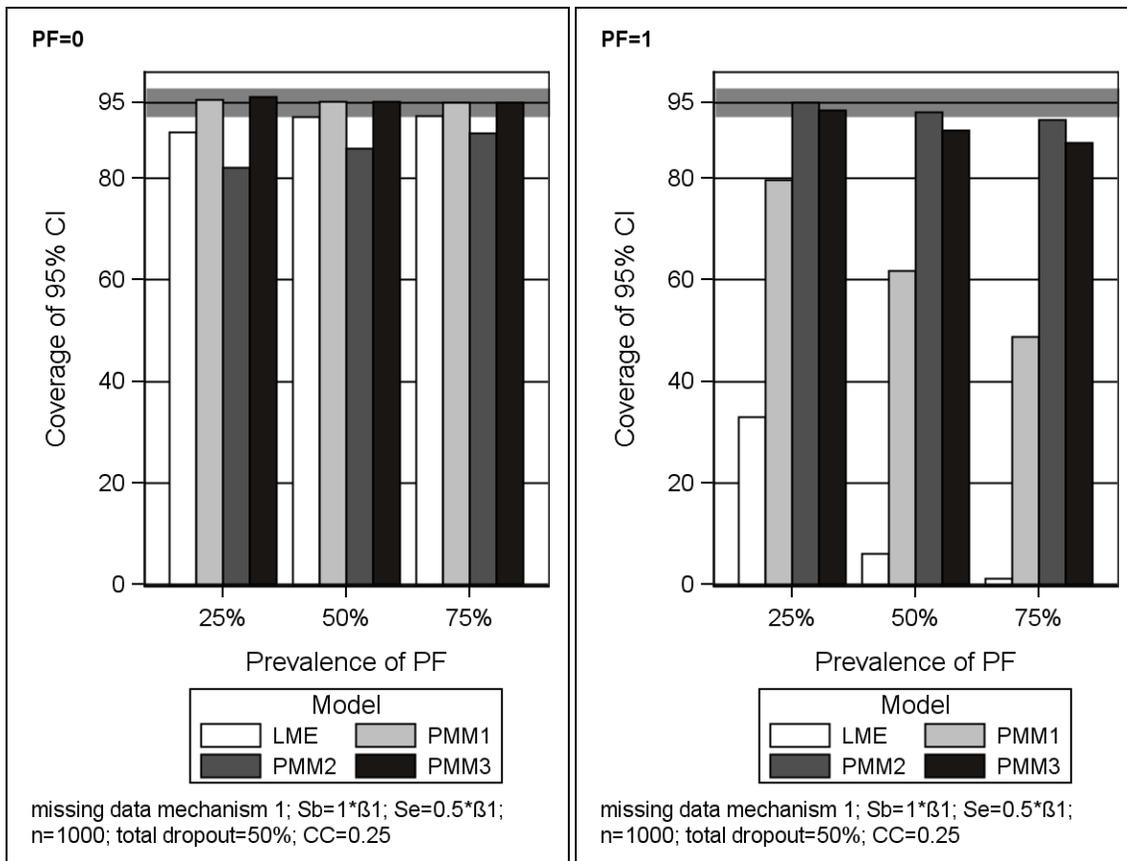
Shaded area represents the binomial margin of error based on the number of simulations

Figure 4.13 Coverage of 95% CI for PF-specific slopes by model and total dropout.



Shaded area represents the binomial margin of error based on the number of simulations

Figure 4.14 Coverage of 95% CI for PF-specific slopes by model and prevalence of PF.



Shaded area represents the binomial margin of error based on the number of simulations

Figure 4.15 Relative bias by model and missing data mechanism.

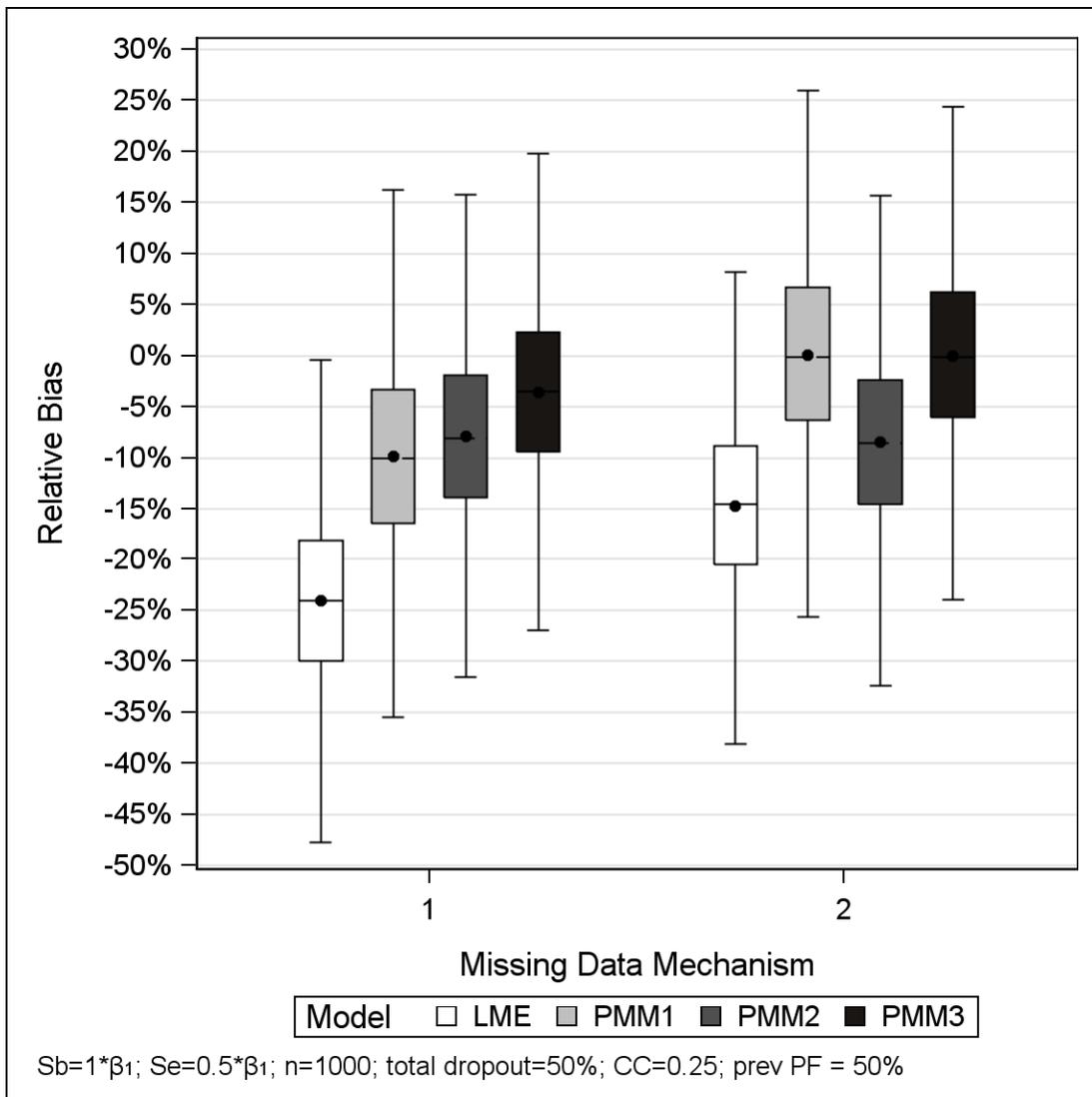


Figure 4.16 Relative bias of PF-specific slopes by model and missing data mechanism.

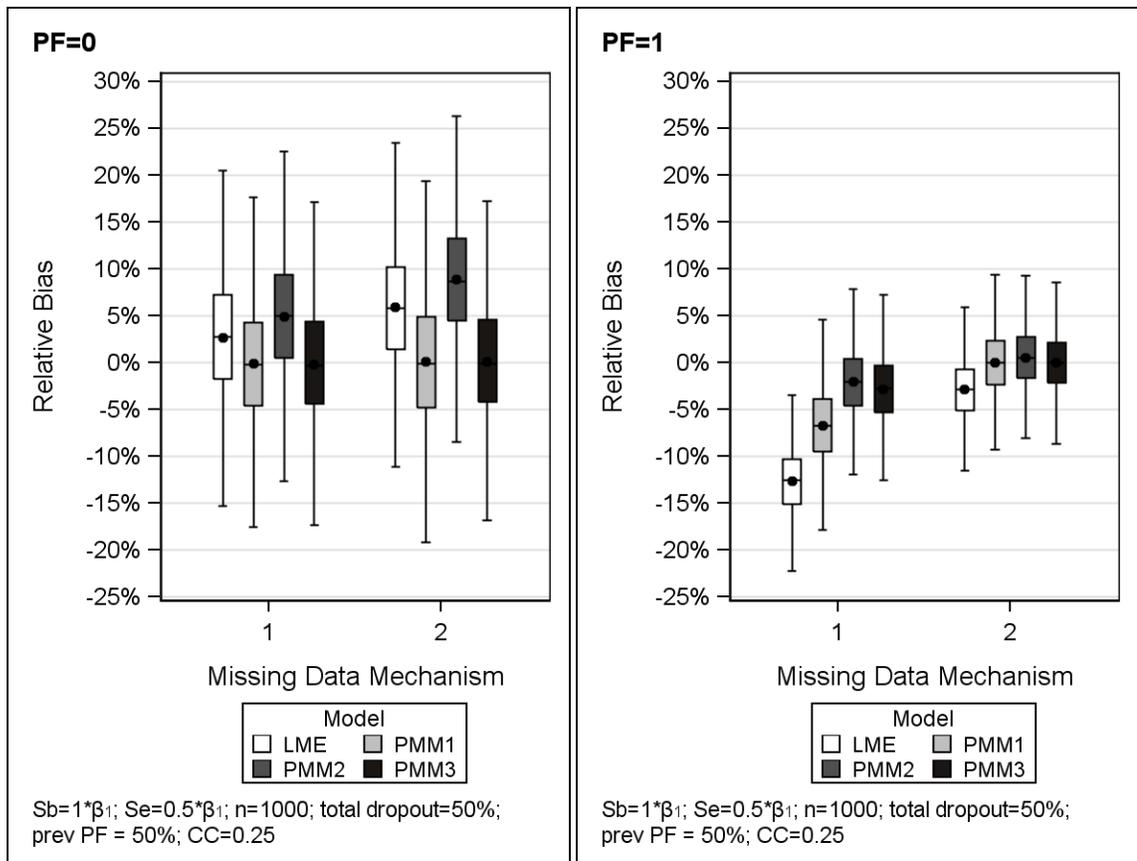


Table 4.1 Percent relative bias by model, missing data mechanism, variability, total dropout and prevalence of PF.

Parameters			50% Total dropout				75% Total Dropout			
Mech.	%PF	S _b	LME	PMM 1	PMM 2	PMM 3	LME	PMM 1	PMM 2	PMM 3
1	0.25	3.75	-12.0	-4.3	-5.4	-2.4	-13.3	-5.7	-6.1	-3.6
1	0.25	7.5	-24.6	-9.2	-9.5	-3.8	-27.4	-12.3	-10.6	-5.7
1	0.25	15	-49.7	-18.8	-16.8	-5.7	-57.1	-26.3	-19.2	-9.3
1	0.5	3.75	-12.2	-4.6	-5.5	-2.6	-13.6	-6.0	-6.4	-3.8
1	0.5	7.5	-23.9	-8.5	-8.9	-3.1	-27.6	-12.2	-10.7	-5.6
1	0.5	15	-50.0	-18.7	-16.6	-4.9	-57.8	-26.0	-19.7	-9.3
1	0.75	3.75	-12.0	-4.2	-8.4	-5.5	-13.4	-5.7	-6.4	-3.7
1	0.75	7.5	-24.6	-8.9	-12.7	-6.8	-27.7	-12.2	-10.8	-5.5
1	0.75	15	-50.4	-18.2	-20.1	-8.2	-58.9	-27.2	-21.6	-10.7
2	0.25	3.75	-6.6	0.0	-3.9	0.1	-5.3	-0.0	-3.6	-0.1
2	0.25	7.5	-13.2	0.0	-7.6	0.2	-11.1	-0.7	-7.9	-0.6
2	0.25	15	-28.0	-1.2	-16.5	-0.7	-20.0	2.2	-13.1	1.8
2	0.5	3.75	-6.5	0.0	-3.9	0.1	-5.3	-0.2	-3.8	-0.2
2	0.5	7.5	-13.2	0.1	-7.7	0.1	-10.8	-0.2	-7.5	-0.2
2	0.5	15	-28.1	-0.6	-17.0	-0.7	-20.5	0.7	-14.2	1.0
2	0.75	3.75	-6.5	0.1	-7.7	-3.6	-5.3	-0.1	-4.1	-0.4
2	0.75	7.5	-13.7	-0.3	-12.6	-4.5	-11.0	-0.7	-8.1	-0.5
2	0.75	15	-26.2	0.3	-17.9	-1.6	-20.7	1.6	-15.0	1.1

correlation between b_{0i} and b_{1i} (CC)=0.25

Table 4.2 Standard error by model, missing data mechanism, variability, total dropout and prevalence of PF.

Parameters			50% Total dropout				75% Total Dropout			
Mech.	%PF	S _b	LME	PMM 1	PMM 2	PMM 3	LME	PMM 1	PMM 2	PMM 3
1	0.25	3.75	0.25	0.29	0.30	0.30	0.27	0.32	0.32	0.32
1	0.25	7.5	0.51	0.59	0.60	0.59	0.55	0.65	0.64	0.63
1	0.25	15	1.02	1.20	1.21	1.19	1.10	1.31	1.30	1.26
1	0.5	3.75	0.22	0.24	0.24	0.24	0.23	0.26	0.25	0.25
1	0.5	7.5	0.43	0.49	0.48	0.48	0.46	0.53	0.51	0.51
1	0.5	15	0.86	0.98	0.97	0.98	0.92	1.06	1.03	1.02
1	0.75	3.75	0.24	0.27	0.25	0.26	0.26	0.28	0.27	0.27
1	0.75	7.5	0.49	0.54	0.51	0.53	0.51	0.57	0.53	0.55
1	0.75	15	0.97	1.08	1.02	1.06	1.03	1.14	1.07	1.10
2	0.25	3.75	0.25	0.27	0.27	0.27	0.26	0.29	0.28	0.28
2	0.25	7.5	0.50	0.55	0.55	0.54	0.52	0.58	0.57	0.56
2	0.25	15	1.00	1.11	1.10	1.09	1.05	1.17	1.15	1.12
2	0.5	3.75	0.21	0.24	0.22	0.23	0.23	0.25	0.23	0.24
2	0.5	7.5	0.43	0.47	0.45	0.46	0.45	0.50	0.47	0.47
2	0.5	15	0.86	0.95	0.90	0.92	0.91	1.01	0.94	0.94
2	0.75	3.75	0.25	0.27	0.25	0.26	0.26	0.29	0.26	0.27
2	0.75	7.5	0.50	0.55	0.50	0.52	0.52	0.58	0.51	0.53
2	0.75	15	1.00	1.10	0.99	1.04	1.05	1.17	1.03	1.07

correlation between b_{0i} and b_{1i} (CC)=0.25

Table 4.3 Coverage of 95% CI by model, missing data mechanism, variability, total dropout and prevalence of PF.

Parameters			50% Total dropout				75% Total Dropout			
Mech.	%PF	S _b	LME	PMM 1	PMM 2	PMM 3	LME	PMM 1	PMM 2	PMM 3
1	0.25	3.75	40.0	86.0	80.0	88.1	31.3	84.0	81.3	90.3
1	0.25	7.5	31.2	88.0	75.6	90.2	28.0	77.6	84.0	92.8
1	0.25	15	32.0	84.8	82.5	99.0	22.4	80.8	89.6	94.4
1	0.5	3.75	24.7	82.0	81.3	91.7	18.7	77.3	77.3	88.7
1	0.5	7.5	29.6	86.4	88.8	92.0	18.4	81.6	84.0	91.2
1	0.5	15	16.0	82.4	90.4	96.8	8.8	76.8	91.2	96.0
1	0.75	3.75	38.7	84.0	87.3	94.0	34.7	82.7	86.7	91.3
1	0.75	7.5	32.0	80.8	88.8	94.4	28.8	81.6	88.0	93.6
1	0.75	15	39.2	91.2	93.5	96.7	24.0	81.6	89.6	94.4
2	0.25	3.75	79.0	94.0	82.0	97.0	84.3	94.7	88.8	95.0
2	0.25	7.5	74.7	95.0	79.5	96.2	84.0	95.0	87.7	95.3
2	0.25	15	81.3	97.3	88.7	97.2	87.7	97.0	90.0	96.7
2	0.5	3.75	74.3	95.0	91.2	95.6	83.0	94.3	91.7	95.0
2	0.5	7.5	71.0	95.0	88.3	96.0	82.7	95.3	90.3	96.7
2	0.5	15	69.7	95.0	89.3	95.7	81.7	95.7	91.3	96.3
2	0.75	3.75	79.7	96.0	93.3	95.7	85.0	94.3	94.0	96.7
2	0.75	7.5	78.7	95.7	93.7	97.0	81.3	95.3	91.0	96.7
2	0.75	15	74.0	96.0	89.0	96.0	86.3	95.3	94.3	95.0

correlation between b_{0i} and b_{1i} (CC)=0.25

CHAPTER 5: Summary

In this dissertation, we evaluated methods for analyzing longitudinal data in the presence of heterogeneity and uncertainty in the missing data mechanism. In the presence of informative censoring there is potential for bias, increased SE, and decreased coverage, even with moderate dropout. Careful consideration should be given to understanding the missing data mechanism, in terms of whether the dropout is MNAR vs. MCAR/MAR and the nature of the association between dropout and (potentially unobserved) outcome. Careful thought must also be given to a potential anchor event.

In Chapter 2 we evaluated the performance of a general linear mixed model (LME) when estimating rate of change in longitudinal studies, when the dropout was result of a combination of MNAR and MCAR mechanisms. We found that mean relative bias was associated with all factors with the exception of sample size. Bias increased with increasing total dropout, increasing informative dropout, and increasing standard deviation of change. While the mean relative bias across all simulations did not change as we varied sample size, the variation around the estimate of relative bias was much larger for smaller sample sizes. Bias was higher under a mechanism with more early dropouts, and was lower when dropout depended on the outcome value rather than change. Standard error increased with increasing standard deviation of change and decreasing sample size, but was not highly associated with dropout mechanism or informative dropout. Coverage probability was associated with both dropout

mechanism and the amount of informative dropout. When the overall amount of dropout was very small (<20% overall), the performance of the linear mixed model was adequate in terms of relative bias, standard error, and coverage, under most scenarios with relative bias less than 7.5% and coverage greater than 80%. However, when the amount of dropout was moderate to large ($\geq 20\%$ overall) the potential for relative bias greater than 10% increases, especially when the standard deviation of change is large, and even under scenarios where only a portion of dropout is informative.

In the motivating example we sought to describe the progression of knee Osteoarthritis (OA) in the Osteoarthritis Initiative (OAI). Based on the results from Chapter 2 we are hesitant to use LME models for this analysis, since dropout is approximately 30% and variability in the outcome, joint space width (JSW), is high at approximately twice the rate of change. When we considered scenarios similar to what we observed in the OAI, we found that while most scenarios demonstrate mean relative bias less than 10%, the interquartile range extends to 15% in some scenarios. In other words, the estimate from a linear mixed effects model to analyze change in JSW over time in the OAI has a non-trivial chance of being biased by greater than 15%. Total knee replacement (TKR) provides some additional information about disease progression. In Chapter 3 we sought to understand if incorporating this additional information would lead to improved estimates of change over time. We proposed an update to the pattern mixture modeling approach to account for uncertainty in the anchor event. This method

separates the data into patterns based both on time of dropout and whether or not a subject experienced the anchor event. We showed that with no uncertainty in the anchor event this method performed as well as the traditional PMM approaches of creating patterns based only on dropout or only on anchor event in terms of bias and coverage. In scenarios with uncertainty in the anchor event, that is, some MNAR dropouts do not have the anchor and some MAR dropouts do have the anchor, the proposed method has lower bias and higher coverage than all other methods evaluated. We also found that the relationship between the anchor event and outcome of interest is important – the performance of the models that incorporate the anchor deteriorated when the anchor was randomly assigned to informative dropouts instead of assigned to the dropouts with the most disease progression.

We implemented each of the four approaches examined in Chapter 3 to analyze change in JSW over time in the OAI. We used TKR as the anchor event. Overall, 34% of dropouts in the OAI underwent TKR. The PMM3 approach yielded the highest estimate of change over time, though all three PMM approaches were quite close (Table 5.1). Using the estimate of yearly change, we estimated total change over 4 years.

Table 5.1 Estimate of yearly change in joint space width in the OAI.

Model	Estimate	Standard Error	p-value	4-year change
LME	-0.1035	0.0058	<0.001	-0.4140
PMM1 (dropout time)	-0.1127	0.0123	<0.001	-0.4508
PMM2 (TKR time)	-0.1170	0.0069	<0.001	-0.4680
PMM3 (dropout + TKR time)	-0.1208	0.0075	<0.001	-0.4832

To further investigate change over time, we calculated the group-specific slopes for each dropout pattern (Table 5.2). The groups undergoing TKR have demonstrate much greater change than those groups not undergoing TKR; the slopes in these groups are approximately three times the slopes in the non-TKR groups. This suggests that TKR may in fact be quite a good anchor event. However, because overall dropout is relatively modest and only approximately one-third of dropouts undergo TKR, this dramatic difference in the change over time in the TKR groups doesn't translate to a substantial difference in the overall estimate of change over time across the entire cohort.

Table 5.2 Group-specific estimates of yearly change in joint space width in the OAI.

Group	Estimate	% of cohort
No dropout/no TKR	-0.09232	72.0%
Dropout/no TKR	-0.1238	19.9%
TKR after 36 months	-0.3288	2.8%
TKR after 24 months	-0.4423	2.9%
TKR between baseline and 24 months	-0.3237	2.4%

In the OAI, in addition to data on TKR we also had information on obesity. Obesity has been shown to be a risk factor for OA progression and could potentially provide additional information about the dropout mechanism. [51] In the OAI, obese subjects were more likely to undergo TKR compared to non-obese subjects, though the dropout rate was similar for the two groups. This suggests that there may be different missing data mechanisms for obese and non-obese subjects. In Chapter 4 we extended the PMM proposed in Chapter 3 to incorporate a baseline prognostic factor. We evaluated the performance of this PMM, in terms of bias, standard error, and coverage, and compared it to the methods evaluated in Chapter 3. The proposed PMM approach had the lowest relative bias and highest coverage among the methods evaluated, both for estimating PF-specific slopes and for estimating the interaction effect. However, the standard error for this method was higher than that for the LME, and under certain conditions was higher than that for the other PMM approaches, reflecting the uncertainty in estimating a large number of parameters.

We once again implemented each of the four approaches – LME, PMM1, PMM2, and PMM3 – and incorporated information on baseline obesity status to analyze change in JSW over time in the OAI. The four approaches were similar in the estimate of yearly change for the non-obese subjects, while for the obese subjects the PMM2 and PMM3 approaches yielded higher estimates of yearly change. This translated to a larger interaction effect estimate in the PMMs, with the LME model estimating an interaction effect of 0.038 millimeters per year

compared to the largest interaction effect estimate of 0.068 using the PMM3 approach.

Table 5.3 Estimate of yearly change in joint space width with standard error by obesity status in the OAI

Model	Estimate: Non-Obese	Estimate: Obese	Estimate: Interaction Effect
LME	-0.086 (0.0078)	-0.124 (0.0084)	0.038 (0.0115)
PMM1 (dropout time)	-0.084 (0.0145)	-0.138 (0.0152)	0.054 (0.0210)
PMM2 (TKR time)	-0.089 (0.0085)	-0.150 (0.0106)	0.061 (0.0136)
PMM3 (dropout + TKR time)	-0.090 (0.0096)	-0.158 (0.0114)	0.068 (0.0149)

The OAI analysis sample included 1,330 subjects with radiographic, symptomatic knee OA at baseline. Consequently, we did not consider the issue of sample size in defining patterns for the pattern mixture models. This is an interesting topic for future research, particularly the question of when it is necessary to separate by time of dropout vs. dichotomous completer/dropout – is it possible to improve standard error without sacrificing bias? Subjects in the OAI who were unable to attend in-clinic assessments were given the option to complete patient-reported outcome questionnaires over the phone. Some patients dropped out of in-clinic assessments, and therefore did not have JSW measured, but did complete questionnaires over the phone. Whether this ancillary information could be incorporated in a meaningful way to better define the patterns is a topic for future research.

We suggest that investigators should work to understand and quantify the association between any potential anchor event and informative dropout. Both sensitivity and specificity of the anchor are important – how many patients dropping out due to disease progression and not having the anchor event, and how many patients that are not experiencing disease progression but are having the anchor event. As demonstrated in Chapter 3, choosing a “bad” anchor – that is, an anchor that does a poor job of differentiating informative and non-informative dropouts – can lead to estimates that have more relative bias and lower coverage than simply utilizing a linear mixed effects model. Other important factors, including variability in outcome, amount of dropout, and timing of dropout are easily quantified and should be evaluated. In general, the performance of the proposed PMM model was superior, in terms of bias and coverage, to the other models evaluated. However, a price was paid in higher SE. The proposed PMM is recommended when there is uncertainty around the anchor event and missing data mechanism.

REFERENCES

1. Little RJA, Rubin DB: **Statistical analysis with missing data**, vol. 4: Wiley New York; 1987.
2. Wu MC, Bailey KR: **Estimation and comparison of changes in the presence of informative right censoring: conditional linear model.** *Biometrics* 1989, **45**(3):939-955.
3. Wang-Clow F, Lange M, Laird NM, Ware JH: **A simulation study of estimators for rates of change in longitudinal studies with attrition.** *Statistics in Medicine* 1995, **14**(3):283-297.
4. Cnaan A, Laird NM, Slasor P: **Tutorial in biostatistics: using the general linear mixed model to analyse unbalanced repeated measures and longitudinal data.** *Statistics in Medicine* 1997, **16**:2349-2380.
5. Fitzmaurice GM, Laird NM, Ware JH: **Applied longitudinal analysis:** Wiley; 2004.
6. Molenberghs G, Kenward MG: **Missing data in clinical studies**, vol. 26: John Wiley & Sons Inc; 2007.
7. Schluchter MD, Greene T, Beck GJ: **Analysis of change in the presence of informative censoring: application to a longitudinal clinical trial of progressive renal disease.** *Statistics in Medicine* 2001, **20**(7):989-1007.
8. Li J, Schluchter MD: **Conditional mixed models adjusting for non-ignorable drop-out with administrative censoring in longitudinal studies.** *Statistics in medicine* 2004, **23**(22):3489-3503.
9. Saha C, Jones MP: **Asymptotic bias in the linear mixed effects model under non-ignorable missing data mechanisms.** *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 2005, **67**(1):167-182.
10. Little RJ: **A class of pattern-mixture models for normal incomplete data.** *Biometrika* 1994, **81**(3):471-483.

11. Little RJ, Wang Y: **Pattern-mixture models for multivariate incomplete data with covariates**. *Biometrics* 1996:98-111.
12. Hogan JW, Roy J, Korkontzelou C: **Handling drop-out in longitudinal studies**. *Statistics in Medicine* 2004, **23**(9):1455-1497.
13. Philipson PM, Ho WK, Henderson R: **Comparative review of methods for handling drop-out in longitudinal studies**. *Statistics in Medicine* 2008, **27**(30):6276-6298.
14. Rubin DB: **Inference and missing data**. *Biometrika* 1976, **63**(3):581-592.
15. Molenberghs G, Thijs H, Jansen I, Beunckens C, Kenward MG, Mallinckrodt C, Carroll RJ: **Analyzing incomplete longitudinal clinical trial data**. *Biostatistics* 2004, **5**(3):445-464.
16. Siddiqui O, Hung HM, O'Neill R: **MMRM vs. LOCF: a comprehensive comparison based on simulation study and 25 NDA datasets**. *Journal of Biopharmaceutical Statistics* 2009, **19**(2):227-246.
17. Lane P: **Handling drop-out in longitudinal clinical trials: a comparison of the LOCF and MMRM approaches**. *Pharmaceutical statistics* 2008, **7**(2):93-106.
18. Mallinckrodt CH, Clark WS, David SR: **Accounting for dropout bias using mixed-effects models**. *Journal of biopharmaceutical statistics* 2001, **11**(1-2):9-21.
19. Peters SAE, Bots ML, Ruijter HM, Palmer MK, Grobbee DE, Crouse JR, O'Leary DH, Evans GW, Raichlen JS, Moons KGM: **Multiple imputation of missing repeated outcome measurements did not add to linear mixed-effects models**. *Journal of Clinical Epidemiology* 2012.
20. Little RJ: **Modeling the drop-out mechanism in repeated-measures studies**. *Journal of the American Statistical Association* 1995, **90**(431):1112-1121.
21. Hogan JW, Laird NM: **Model-based approaches to analysing incomplete longitudinal and failure time data**. *Statistics in Medicine* 1997, **16**(1-3):259-272.

22. Diggle P, Kenward MG: **Informative drop-out in longitudinal data analysis**. *Applied statistics* 1994:49-93.
23. Rubin DB: **Formalizing subjective notions about the effect of nonrespondents in sample surveys**. *Journal of the American Statistical Association* 1977, **72**(359):538-543.
24. Fairclough DL: **Design and analysis of quality of life studies in clinical trials**: Chapman & Hall/CRC.
25. Dunlop DD, Manheim LM, Song J, Chang RW: **Arthritis prevalence and activity limitations in older adults**. *Arthritis & Rheumatism* 2001, **44**(1):212-221.
26. Felson DT, Lawrence RC, Dieppe PA, Hirsch R, Helmick CG, Jordan JM, Kington RS, Lane NE, Nevitt MC, Zhang Y: **Osteoarthritis: new insights. Part 1: the disease and its risk factors**. *Annals of Internal Medicine* 2000, **133**(8):635-646.
27. Eckstein F, Maschek S, Wirth W, Hudelmaier M, Hitzl W, Wyman B, Nevitt M, Le Graverand MPH: **One year change of knee cartilage morphology in the first release of participants from the Osteoarthritis Initiative progression subcohort: association with sex, body mass index, symptoms and radiographic osteoarthritis status**. *Annals of the rheumatic diseases* 2009, **68**(5):674-679.
28. Duryea J, Neumann G, Niu J, Totterman S, Tamez J, Dabrowski C, Le Graverand MP, Luchi M, Beals CR, Hunter DJ: **Comparison of radiographic joint space width with magnetic resonance imaging cartilage morphometry: analysis of longitudinal data from the Osteoarthritis Initiative**. *Arthritis Care Res (Hoboken)* 2010, **62**(7):932-937.
29. Phan CM, Link TM, Blumenkrantz G, Dunn TC, Ries MD, Steinbach LS, Majumdar S: **MR imaging findings in the follow-up of patients with different stages of knee osteoarthritis and the correlation with clinical symptoms**. *European radiology* 2006, **16**(3):608-618.
30. Zhang Y, Niu J, Felson DT, Choi HK, Nevitt M, Neogi T: **Methodologic challenges in studying risk factors for progression of knee osteoarthritis**. *Arthritis care & research* 2010, **62**(11):1527-1532.

31. **Osteoarthritis Initiative** [<http://oai.epi-ucsf.org>]
32. Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt LW: **Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee.** *The Journal of rheumatology* 1988, **15**(12):1833-1840.
33. Touloumi G, Babiker AG, Pocock SJ, Darbyshire JH: **Impact of missing data due to drop-outs on estimators for rates of change in longitudinal studies: a simulation study.** *Statistics in Medicine* 2001, **20**(24):3715-3728.
34. Scharfstein DO, Hogan J, Herman A: **On the prevention and analysis of missing data in randomized clinical trials: the state of the art.** *The Journal of bone and joint surgery American volume* 2012, **94 Suppl 1**:80-84.
35. Hogan JW, Laird NM: **Mixture models for the joint distribution of repeated measures and event times.** *Statistics in Medicine* 1997, **16**(1-3):239-257.
36. Daniels MJ, Hogan JW: **Missing data in longitudinal studies: Strategies for Bayesian modeling and sensitivity analysis:** Chapman and Hall/CRC; 2008.
37. Pauler DK, McCoy S, Moinpour C: **Pattern mixture models for longitudinal quality of life studies in advanced stage disease.** *Statistics in Medicine* 2003, **22**(5):795-809.
38. Li H, Heitjan DF: **A pattern-mixture model for the analysis of censored quality-of-life data.** *Statistics in Medicine* 2006, **25**(9):1533-1546.
39. Ibrahim SA, Franklin PD: **Race and Elective Joint Replacement: Where a Disparity Meets Patient Preference.** *American journal of public health* 2013(0):e1-e2.
40. Hawker GA, Wright JG, Coyte PC, Williams JI, Harvey B, Glazier R, Wilkins A, Badley EM: **Determining the need for hip and knee arthroplasty: the role of clinical severity and patients' preferences.** *Medical care* 2001, **39**(3):206-216.

41. Riddle DL, Kong X, Jiranek WA: **Two-year incidence and predictors of future knee arthroplasty in persons with symptomatic knee osteoarthritis: preliminary analysis of longitudinal data from the osteoarthritis initiative.** *The Knee* 2009, **16**(6):494-500.
42. Kenward MG, Molenberghs G: **Parametric models for incomplete continuous and categorical longitudinal data.** *Statistical Methods in Medical Research* 1999, **8**(1):51-83.
43. Ten Have TR, Miller ME, Reboussin BA, James MK: **Mixed effects logistic regression models for longitudinal ordinal functional response data with multiple-cause drop-out from the longitudinal study of aging.** *Biometrics* 2000, **56**(1):279-287.
44. Crawford SB, Hanfelt JJ: **Testing for the presence of multiple sources of informative dropout in longitudinal data.** *Statistics in Medicine* 2008, **27**(21):4175-4189.
45. Williamson PR, Kolamunnage-Dona R, Philipson P, Marson AG: **Joint modelling of longitudinal and competing risks data.** *Statistics in medicine* 2008, **27**(30):6426-6438.
46. Gueorguieva R, Rosenheck R, Lin H: **Joint modelling of longitudinal outcome and interval-censored competing risk dropout in a schizophrenia clinical trial.** *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 2012, **175**(2):417-433.
47. Li N, Elashoff RM, Li G: **Robust joint modeling of longitudinal measurements and competing risks failure time data.** *Biometrical journal Biometrische Zeitschrift* 2009, **51**(1):19-30.
48. Li N, Elashoff RM, Li G, Saver J: **Joint modeling of longitudinal ordinal data and competing risks survival times and analysis of the NINDS rt-PA stroke trial.** *Statistics in Medicine* 2010, **29**(5):546-557.
49. Elashoff RM, Li G, Li N: **An approach to joint analysis of longitudinal measurements and competing risks failure time data.** *Statistics in Medicine* 2007, **26**(14):2813-2835.

50. Daniels MJ, Hogan JW: **Reparameterizing the pattern mixture model for sensitivity analyses under informative dropout.** *Biometrics* 2000, **56**(4):1241-1248.
51. Wirth W, Hellio Le Graverand MP, Wyman BT, Maschek S, Hudelmaier M, Hitzl W, Nevitt M, Eckstein F: **Regional analysis of femorotibial cartilage loss in a subsample from the Osteoarthritis Initiative progression subcohort.** *Osteoarthritis and cartilage / OARS, Osteoarthritis Research Society* 2009, **17**(3):291-297.

CURRICULUM VITAE

