DESIGN OF PRIMARY AND SENSITIVITY ANALYSES FOR HANDLING NON-FUTURE DEPENDENCE MISSING DATA IN CLINICAL TRIALS WITH AN EMPHASIS ON THE TYPE-I ERROR RATE USING MULTIPLE IMPUTATION AND PATTERN MIXTURE MODEL APPROACH

By

LIXIAN PENG

A dissertation submitted to the

School of Public Health

and the

Graduate School – New Brunswick

Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Written under the direction of

Professor Weichung Joe Shih

And Approved by

_____

_____

_____

_____

New Brunswick, New Jersey

May, 2015

ABSTRACT OF THE DISSERTATION

DESIGN OF PRIMARY AND SENSITIVITY ANALYSES FOR HANDLING NON-FUTURE DEPENDENCE MISSING DATA IN CLINICAL TRIALS WITH AN EMPHASIS ON THE TYPE-I ERROR RATE USING MULTIPLE IMPUTATION AND PATTERN MIXTURE MODEL APPROACH

by LIXIAN PENG

Dissertation Director:

Professor Weichung Joe Shih, Ph.D

Missing data is a common problem in longitudinal clinical trials. Substantial missing data could introduce potential biases and undermine the scientific credibility of causal conclusions from clinical trials. To handle the missing data issue, it is always required by the regulatory agencies to pre-specify a primary analysis and sensitivity analysis in protocol or statistical analysis plan (SAP). Recent National Research Council (NRC) report questioned the reasonableness of the missing at random (MAR) setting as the primary analysis since MAR is a very special and doubtful assumption for the missing data mechanism, and the report encourages to use not missing at random (NMAR) setting as the primary analysis. One of the NMAR mechanisms is non-future dependence missing data (NFD-NMAR). It is also one of the recommended methods in the NRC report. This dissertation addressed this issue and proposed a process to investigate the mean-shift model with NFD-NMAR mechanism (NFD-Delta method). The goal is to

provide, via the investigation process, a method of finding an appropriate shift parameter to specify the primary NMAR analysis in study protocol or SAP based on the criteria of maintenance of the type-I error rate for late phase trials by simulations. The simulation set-up should be based on either early phase data or information from interim data of the current trial. The shift parameter of the NFD-Delta method constitutes the sensitivity analysis.

Several components were considered for the NFD shift parameter in this dissertation: the metric/unit, magnitude, and the algorithm to place the shift to examine the effect of these components on the type-I error rate (alpha) under the null hypothesis of no treatment effect. For the metric factor, four different metric units were considered: constant STD_1, constant RSD_1, STD_k, RSD_k; for the magnitude factor, different values of shift parameter *f* were considered to investigate which *f* value is the appropriate shift parameter to control the type-I error rate to the nominal level; for the algorithm to implement the delta shift, three different methods were proposed: sequential, non-sequential and single adjustment method. Extensive simulations were conducted to investigate the type-I error rate. Correctness and robustness of the results were examined.

Acknowledgements

I would never have been able to finish my dissertation without the guidance of my committee members and support from my family.

I would like to express my deepest gratitude to my dissertation advisor, Dr. Weichung Joe Shih, for his excellent guidance, instructions, patience. Dr. Shih inspired me and provided me an excellent atmosphere and spirit for doing research. Without his guidance and persistent help this dissertation would not have been possible. I would like to thank Dr. Yong Lin, who provided great help and guidance on the simulation work of my dissertation. I am also grateful to Dr. Shou-En Lu and Dr. Hui Quan for being my committee members, giving inspiration on my dissertation topic, reviewing my dissertation, and providing helpful comments.

I would like to thank Pingshan who took care of my daughter during years of my Ph.D study.

Finally, I would like to thank my mom and my sisters. They were always supporting me and encouraging me with their best wishes.

Dedication

To my mom, my sisters, Pingshan and my daughter Emily.

Table of Contents

List of Tables

Chapter 1

Introduction and Overview

1.1 Missing Data Problem

Missing data is a common problem in longitudinal clinical trials. Major reasons for participants to discontinue the assigned treatment include: adverse events, lack of tolerability, lack of efficacy, lost to follow-up, or simple inconvenience. Substantial missing data could undermine the scientific credibility of causal conclusions from clinical trials. An example to illustrate the possible bias caused by missing data is the following: Suppose that a high value is a bad outcome of an endpoint, e.g., blood pressure or pain score, patients with high values tend to discontinue the therapy and drop out of the study more easily than patients with lower values, especially when the therapy also accompanies tolerability issue. Then, the outcome of the study would be biased when there is a severe imbalance of dropout rates between treatment groups in a randomized clinical trial. The recent Food and Drug Administration-commissioned report The Prevention and Treatment of Missing Data in Clinical Trials by the National Research Council (NRC) panel, and subsequent publications called for increasing awareness of the potential frailty of conclusions in the presence of missing data, and provided recommendations of strategies in design to prevent and in analysis to treat missing data. Nevertheless, the report also acknowledges that missing data continues to be ubiquitous in clinical trials, and more research is needed to deal with the problem, especially methods for sensitivity analysis.

1.2 Sensitivity Analysis

Although many statistical methods exist to deal with missing data in clinical trials, there is not a universal method since each trial has its own set of design and measurement characteristics. The properties of these methods depend heavily on the mechanisms leading to missing data, which are assumptions that cannot be verified or tested completely by the observed data. Therefore, in reports and interpretations of trial results that contain substantial missing data, we need scientifically defensible analyses coupled with a sensitivity analysis to assess robustness of the results. For clinical trials sponsored by the pharmaceutical industry, especially those pivotal trials for the New Drug Application (NDA), a common practice requested by the regulatory authorities is to specify a primary analysis and several sensitivity analyses for the primary endpoint.

There are different types of sensitivity analyses. Some uses different sets of data such as all randomized or treated populations versus per-protocol population that excludes protocol violators, outliers, or influential observations; different analytical approaches about the distributional assumption such as parametric versus non-parametric methods; or different forms of summary statistics such as last time-point versus area under the curve. The most concern of missing data, however, is on the assumption of the missing data mechanism (MDM), since it is untestable in most cases. This dissertation also follows the same thread as in the NRC's report to focus on the sensitivity analysis regarding the untestable assumption of MDM in clinical trial. For example, if the primary analysis is based on the assumption of missing at random (MAR) mechanism, then the sensitivity analysis would be to allow departures from MAR. If the primary analysis is based on a particular missing not at random (MNAR) mechanism, then the sensitivity analysis would

be to assess different degrees of MNAR and their impact on the results. I will continue the discussion of sensitivity analysis in relation to this dissertation in the next section. Then in Section 2.7, I will review the literature and continue more discussions on sensitivity analyses.

1.3 Objectives and Research Plan

Over the last several years, especially after the publications of the European Medicines Agency's Guideline on Missing Data in Confirmatory Clinical Trials and the NRC's report, both in 2010, the old-fashioned, widely used LOCF (last observation carried forward) and BOCF (baseline observation carried forward) methods have pretty much been abandoned as the primary (or even secondary) analysis for most NDA trials, since these methods assume extremely unreasonable models for missing data. In many occasions, for example, in the recent FDA's Guidance for Industry Analgesic Indications, methods assuming MAR mechanism have also been denounced, since the justification of such a missing data mechanism is difficult. Instead, MDM that is 'not missing at random' (NMAR) or equivalently, 'missing not at random' (MNAR) is encouraged to use as the primary analysis. One of the NMAR mechanisms is the so-called non-future dependence missing data proposed by Kenward et al., abbreviated as NFD-NMAR in this dissertation. NFD-NMAR is also one of the recommended methods in the NRC report.

I will investigate the NFD-NMAR mechanism in great detail in this dissertation for continuous endpoints. There are several steps in carrying out an analysis using the NFD-NMAR mechanism. The foremost important step is to know the causal estimand that the analysis intends to address. A causal estimand requires both the target inferential

population and the endpoint of the analysis. The most commonly referenced ITT (intention to Treat) estimand essentially avoids missing data or requires no missing data entirely, hence it is ideal but not very practical or applicable in the presence of missing data (see Lachine, 2000). I will consider two other estimands described in the NRC's report that require imputation of the missing data. One is the AAT (As Assigned Treatment) Estimand. For this estimand, it is assumed that the dropouts would follow the path in the originally assigned treatment group had they stayed in the study. The imputation of missing data does not cross treatment groups. Another one is the ACT (As Control Treatment) Estimand. This estimand assumes that the discontinued subjects in the treatment group would switch to the path of the control group had they continued the study. The NRC report contains discussion of different views of these two causal estimands, but did not explicitly give the specific terms. These two phrases AAT and ACT were given in Shih and Aisner.

Following Daniels and Hogan, the next step in carrying out analysis with the NFD-NMAR mechanism is to assume a mean-shift (delta) model to link the un-identifiable part to the identifiable (observed) data. The mean-shift model is a helpful interpretation of the NFD-NMAR, and has the advantage of including MAR mechanism as a special case. This is called the NFD-Delta method in this dissertation. The shift parameter of the NFD-Delta method is then constitutes the sensitivity analysis.

Furthermore, there are several components to consider for the shift parameter: the metric/unit, magnitude, and the algorithm to place the shift(s). In the literature, there is little discussion of these components, and as far as I know, no study has ever been done to examine the effect of these components on the type-I error rate (alpha) of the analysis.

In most of discussions of sensitivity analyses, other authors emphasized the 'tipping point', where the shift parameter changes the study result substantially, e.g., from statistical significance to non-significance. However, I feel strongly that the selection of an appropriate shift parameter in an analysis can possibly be done only when we first understand their impact on the type-I error rate. Without this assessment, the sense of statistical significance/non-significance has no ground. It is conceivable that these three components: unit, magnitude, and algorithm, interplay in the findings of the 'tipping point' (in the sense from preserving the type-I error rate to losing power) when analyzing a data set.

The above discussion is from the data analysis perspective. A more pertinent issue regarding a study with potential missing data due to dropouts at the study design stage is how to pre-specify in the protocol or the statistical analysis plan (SAP) a primary analysis and a sensitivity analysis strategy, as always required by the regulatory agency. If the primary analysis is based on the assumption that the missing data are MAR, the sensitivity analysis will then (naturally) be the results based on graduate deviation from MAR. There is no much of issues, the delta adjustment for the primary analysis (MAR) is set at zero (i.e., no adjustment). However, recently the FDA and others have questioned the reasonableness of setting MAR as the primary analysis, since MAR is a very special and doubtful assumption for the missing data mechanism. The central question I try to address in this dissertation is that, when the primary analysis cannot be set at zero adjustment under MAR assumption, what would then be an "appropriate" specification of the delta adjustment in the study protocol or SAP? The

"appropriateness" that I advocate to use in this dissertation is the maintenance of the type-I error rate.

This dissertation will be organized as follows.

In Chapter 2, I shall briefly review the missing data framework in the literature, with a focus on the main theory of the NFD-NMAR mechanism and the mean-shift model as expressed in the pattern mixture model (PMM) approach, and the relationship between the PMM and selection model (SM) notions of NFD-NMAR. The SM formulation is more convenient for performing simulations, and the PMM formulation is more useful for performing data analysis. Hence we need both of them.

In Chapter 3, I will focus on the implementation of NFD-NMAR together with the mean-shift model, and the NFD-Delta method in this dissertation. As I alluded to previously, there are steps and components for this implementation. I will use a real data example of clinical trial in a pain study as the proto-type for illustration of the general framework. This data set consists of two treatment groups with about 200 subjects in each. The study duration is 5 weeks and the endpoint is the daily average pain score (DAPS) reported at baseline and weekly after treatment. It was a phase II study to be used for planning a future phase III trial, especially for planning the primary and sensitivity analyses of DAPS in the presence of missing data that is very much expected in the phase III trial with a longer duration. In this phase II data, the dropout rates and reasons for withdrawal will be presented. Multiple imputation (MI) by the SAS^{TM} software will be involved. But there are several versions of the implementation, and I will point out the differences between those in the literature and in what I propose and give my rationale.

In Chapter 4, I will conduct extensive simulation studies. My focus is to investigate the type-I error rate. The simulation plan is structured in Table 1.1.

(a) I first verify the well-known fact that the alpha level should be at the nominal level for the MMRM (mixed effect model for repeated measures by Proc Mixed in SAS) as well as for the multiple imputation (MI) method (by Proc MI plus Proc MIANALYZE in SAS) when the MDM is MAR. This is because they are likelihood-based methods.

(b) When the MDM is NFD-NMAR or mixture of MAR and NFD-NMAR, I hypothesize that the simulated alpha level should still be close to the nominal level for MMRM and MI methods when the dropout rates are the same between treatment groups, but when the treatment groups' dropout rates are different between treatment groups there will be inflation of the alpha level. This hypothesis is based on the theorem in Wei and Shih. The degree of the inflation will be examined.

(c) The most important part of this dissertation is the attempt to find a delta-shift for the NFD-Delta method so that the type-I error rate can be controlled closer to the nominal level when the MDM is NFD-NMAR or mixture of MAR & NFD-NMAR, where the methods of MMRM and MI inflate the type-I error rate. Since the NFD-Delta method involves different metric/units, magnitudes and algorithms (see Chapter 3), the simulation will investigate combinations of these settings.

(d) While a delta, which is expressed in terms of metric/unit and magnitude and depends on the algorithm, may be found for the NFD-Delta method to control the alpha level at the nominal level when the MDM is (correctly) specified, i.e., when

MDM is truly NFD-MNAR or mixture of MAR & NFD-NMAR, I also investigate the robustness of the NFD-Delta when the MDM is actually MAR. Another scenario of the robustness is when the MDM is future-dependent. It would be interesting to examine whether the choice of the delta-shift is robust, or how robust the delta-shift is, under these situations.

(e) The above plan is for the AAT (As assigned Treatment) Estimand. In addition, I will also conduct analysis for the ACT (As Control Treatment) Estimand. Since this estimand assumes that the discontinued subjects in the treatment group would follow the path of the control group, this represents another type of "penalty" other than the "delta shift" model. There are also different varieties of implementation of ACT. I will give my version in contrast with others.

As alluded to earlier, finding an appropriate shift parameter is important for a study protocol, where a primary data analysis plan must be pre-specified, followed by sensitivity analyses planned around this primary analysis. Since each trial with its own design (including endpoint and study duration) and data, it is not possible to establish a universal shift parameter that is appropriate for all types of studies when the primary analysis has to be based on a NMAR scenario. Rather, this dissertation will, via the investigation process, provide a method of finding an appropriate shift parameter for any late phase trial by simulations (Chapter 4) based on either early phase data or information from interim data of the current trial.

Finally, I shall summarize the findings in Chapter 5, and provide additional comments and some potential future research topics.

Table 1.1 Simulation Plan for Investigation of Alpha Levels Under Different Missing Data Mechanism, Missing Data Pattern, and Methods of Data Analysis

Estimand: AAT (See Chapter 2)
Delta Metric: (See Chapter 3)

Missing Data Mechanism (MDM)		Method of Data Analysis						
	Dropout rate between treatment groups	PROC Mixed (MMRM)	PROC MI + MMRM + MIANALYZE (Multiple Imputation)	PMM – NFD Mean-Shift (NFD – Delta)				
				Magnitude		Algorithm		
				Large	Small	Single adjustment	Non-sequential	Sequential
MAR	Same	Should be at the nominal level	Should be at the nominal level					
	Different	Should be at the nominal level	Should be at the nominal level					
NFD-NMAR	Same	Should be close to the nominal level	Should be close to the nominal level					
	Different	Possible inflation	Possible inflation					
Mixture of MAR & NFD-NMAR	Same	Should be close to the nominal level	Should be close to the nominal level					
	Different	Possible inflation	Possible inflation					
Future Dependence	Different	Possible inflation	Possible inflation					

Note: Table will be expanded for AAT estimand for different shift parameters, and revised for ACT Estimand, and Blanks are to be filled-in with observed alpha levels from simulations in Chapter 4.

Chapter 2

Missing Data Framework: Literature Review and Comments

2.1 Notation

The missing data literature is abundant since the problem of missing data has a long history. I will limit the review to only those related to my research topics summarized in Table 1.1. Throughout this dissertation, the following conventions are used. Let X represent treatment indicators and baseline covariates that are fully observed and conditioned on in the primary statistical analysis, or the design variables that would be adjusted for or conditioned on in the final statistical analysis. Let Y denote the primary outcome which is a continuous variable and is repeatedly measured during the trial. Let V represent auxiliary variables. Auxiliary variables are different from design variables X, they may represent subject-level characteristics that can help in drawing inference from incomplete response data. Some variables that are not included in the primary analytic model but could be correlated to some of the variables in the model can be included in V, such as the information on compliance or side effects of treatment, etc.. These variables can be used to reduce the uncertainty caused by the missing data and improve the precision of the estimation because of the correlations. Let M denote the indicator of whether Y is missing or observed. In clinical trials with repeated measures, we usually include a subscript for the repeated measures. That is, if the intended outcome measures are $Y = (Y_0, Y_1, Y_2, ..., Y_T)$, with T+1 scheduled visits, the corresponding missingness indicators are $M = (M_0, M_1, M_2, ..., M_T)$, where $M_j = 1$ if Y_j is missing, and $M_j = 0$ if it is observed for j = 0 to T (j=0 being baseline). In many situations, it is helpful to allow

more than one missing-value code to indicate different types of missing data, such as M = 1 for lack of efficacy, M = 2 for inability to tolerate a drug due to adverse events, M = 3 for a missed clinic visit, and so on. This notation allows for different modeling assumptions for the different reasons of missing data. M is a random variable and is called missing data pattern in the literature.

With multivariate Y, it is helpful to partition Y as (Y_{obs}, Y_{mis}). Particularly, if T post-treatment observations were intended to be taken on each subject after baseline visit (j=0), and the missing data pattern is monotone, then the observed data comprise Y_{obs} = ($Y_0, Y_1,\ldots,Y_j$) for some $j \leq T$, and M = (0,...,0, 1,...,1) when the first $j+1$ elements of M are 0's and the remaining elements are 1's. j varies across subjects. Monotone patterns happen when patients drop out of the study due to death, withdrawal of consent, or moved away. The NRC's report discussed design methods such as allowing concomitant or rescue medications to retain patients in the study (to provide measurements), to minimize dropouts due to "lack of efficacy", and to continue collecting the data in the event of adverse effect even when the study treatment is altered or discontinued. When data are collected for every patient, then ITT (intention to treat) estimand is an option for the data analysis. Otherwise, when missing data are present, ITT is not supported by the design and data.

When missing data are present, the joint distribution of Y and M, written in the bracket notation as [Y, M] for convenience, is the key in studying the missing data problem, as noticed by Shih and Quan (1997). The joint distribution may be factored in two different ways: [Y, M] = [M| Y; φ] [Y; θ], or [Y, M] = [Y| M; θ_M] [M]. The former is called selection model (SM) approach and the latter is called pattern mixture model (PMM)

approach. When conditioning on X, the parameter θ is pertinent to the treatment effect, is our primary interest. The parameter φ is about the MDM. The parameter θ_M is pattern specific treatment effect, hence needs to be integrated over the patterns to relate them to θ.

2.2 Missing Data Mechanisms

While the missing data pattern M is a random variable, the distribution of M in relation to (Y, X, V) is the missing data mechanisms (MDM), which describes the process that causes the missing data. Rubin (1976) first characterized restrictions on the distribution of M given (Y_{obs}, Y_{mis}, X ,V), using the SM approach, written in the bracket notation as [M | X, V_{obs}, V_{mis}, Y_{obs}, Y_{mis}] for convenience and defined three general different MDM as follows. Little (1995) later pointed it out that each of the three MDM in Rubin's SM approach has an equivalent representation in terms of the predictive distribution of missing responses, namely Y_{mis} given (M, Y_{obs}, X, V) based on the PMM approach. Inferences from incomplete data rely on the assumption of MDM which cannot be verified from the data.

The MDM can be classified as missing completely at random (MCAR), missing at random (MAR), and not missing at random (NMAR).

Missing Completely At Random (MCAR)

With the SM approach, missing data are missing completely at random (MCAR) if missingness does not depend on values of the covariates, auxiliary and outcome variables (X, V, Y). That is,

$$[M \mid X, V_{obs}, V_{mis}, Y_{obs}, Y_{mis}] = [M] \quad (2.1)$$

MCAR assumption is very strong and generally unlikely to hold, especially in the missing data settings that are not under the control of investigators in clinical trials. One example of the situation in which MCAR might be suitable is administrative censoring. In this situation, the outcomes are censored because a study is terminated at a pre-planned date, the outcome has not occurred yet for late accruals, the reason for missingness was unrelated to the response model and its covariates.

MCAR has an unique feature that we can test whether the missing outcomes are MCAR if they are at least missing at random (Little, 1988).

When MCAR is applicable, the missing data is ignorable. The inference based on observed data Y_{obs} alone is appropriate. The method of generalized estimating equations (GEE) for unbalanced clustered data including repeated measures (Liang and Zeger, 1986), being a sampling distribution based inference, requires the assumption of MCAR.

Missing At Random (MAR)

Missing at random (MAR) is more realistic than MCAR for many studies. MAR requires that the missingness does not depend on the missing responses Y_{mis} and V_{mis}, conditionally on observed responses (Y_{obs}, V_{obs}) and covariates X. That is, with the SM approach:

$$[M \mid X, V_{obs}, V_{mis}, Y_{obs}, Y_{mis}] = [M \mid X, V_{obs}, Y_{obs}] \quad (2.2)$$

With the PMM approach, the equivalent notion of MAR is:

$$[Y_{mis}, V_{mis} \mid X, V_{obs}, Y_{obs}, M] = [Y_{mis}, V_{mis} \mid X, V_{obs}, Y_{obs}] \quad (2.3)$$

which implies that the predictive distribution of the missing variables given the observed variables does not depend on the pattern of missing values. This PMM notion of MAR is relevant from an analysis perspective since it relates to the predictive distribution of the missing values, which is the basis for principled methods of imputation. The equivalence of (2.2) and (2.3) will be given in Section 2.5.

Many standard analysis methods for incomplete data work under the MAR assumption. Therefore, as emphasized and stated in the NRC's report, it is crucial that "both the MAR assumption and the assumptions underlying the data model (e.g., multivariate normality) be thoroughly justified before the results from these models can be considered valid for treatment comparisons". Generally, with incomplete data, the assumptions about both the MDM and the data model are not verifiable from the observed data (although there are tests that can show that a dataset is not MCAR, there is no test that proves that any dataset is MCAR or MAR.). the inference and decisions about treatment effect often crucially depend upon the missing data mechanism. Even under MAR, different assumptions about the full data model will lead to different predictive distributions. This is the fundamental idea of the need to perform sensitivity analyses when missing data are present, as suggested in the ICH guideline (European Medicines Evaluation Agency, 2010).

When MAR is applicable, Rubin (1976) showed that the likelihood-based inference based on Y_{obs} alone is appropriate, provided that the parameter of interest, θ, and that of the missing data process, φ, are functionally distinct. In this case, MAR together with distinct

parameters, is called "MDM is ignorable". The method of weighted *GEE*, the mixed-effect models (Laird and Ware, 1982) in SAS® PROC MIXED, and the multiple imputation method (Rubin, 1998; Schafer 1997) implemented in SAS® PROC MI, require the MAR assumption (Yuan, 2011).

For a monotone pattern, it is easy to add the following notation. At any given time *j*, let $Y_j^- = (Y_0, Y_1, ..., Y_{j-1})$ denote the past history of measurements up to but not including time *j*, and let $Y_j^+ = (Y_j, ..., Y_T)$ denote the current and future measurements scheduled, including and after time *j*. At time *j*, the predictive distribution of future values given the observed history is denoted by $[Y_j^+|Y_j^-, X, M_j = 0]$. For easy expression, we omit the auxiliary variable V here. With the PMM approach, the MAR condition (2.3) holds if predictions of future measurements for those who drop out at time *j* are equivalent in distribution to predictions for those who have observed data at and after time *j*. That is, MAR is equivalent to

$$[Y_j^+|Y_j^-, X, M_j = 1] = [Y_j^+|Y_j^-, X, M_j = 0] \quad (2.4)$$

Hence, under MAR assumption, the missing values at time *j* and beyond can be predicted sequentially from the histories of subjects who are still in the study at time *j*.

Eq. (2.4) can be generalized by allowing the past measurements to include auxiliary covariates *V*. Let $Z_j^- = (Y_0, Y_1, ..., Y_{j-1}, V_0, V_1, ..., V_{j-1})$ represent the observed history of both outcomes and auxiliaries. Then MAR can be rewritten by replacing Y_j^- with Z_j^- in (2.4). The validity of the MAR assumption can be improved by measuring and including

auxiliary variables V that are predictive of whether the outcome variables are missing and predictive of the values of the missing variables.

Not Missing At Random (NMAR)

With the selection model setting, we factor [Y, M|X] = [M|Y, X][Y|X] = [M|Y_{obs}, Y_{mis}, X][Y|X]. Under NMAR, the missing data mechanism [M|Y, X] = [M|Y_{obs}, Y_{mis}, X] depends not only on Y_{obs} (as for MAR) but also on the missing data Y_{mis}. In other words, if missingness or dropout depends on the values of missing variables Y_{mis} after conditioning on the observed variables $Y_{obs,}$ MAR fails to hold, missing data are said to be not missing at random (NMAR) or missing not at random (MNAR). Withdrawal after a sudden unrecorded drop in efficacy response could be NMAR because it would depend on unobserved data and could not be estimated from the observed data alone.

For a monotone missing data pattern, missingness will be NMAR if there exists, for any *j*, at least one value of Y_j^- for which

$$[Y_j^+|Z_j^-, X, M_j = 1] \neq [Y_j^+|Z_j^-, X, M_j = 0] \quad (2.5)$$

or equivalently, there exists, for any *j*, at least one value of Y_j^+ , such that

$$P(M_j = 1| M_{j-1} = 0, Z_j^-, Y_j^+, X) \neq P(M_j = 1| M_{j-1} = 0, Z_j^-, X) \quad (2.6)$$

For (2.5), we can see, under NMAR, the prediction of future observations for those who drop out cannot be properly predicted using data observed prior to dropping out or that the distribution $[Y_j^+|Y_j^-]$ varies between those who do and do not drop out at time *j*. Since

these differences cannot be predicted from the observed data, they are fully assumption driven. This is the key issue of missing data analysis in many clinical trials.

When the MDM is NMAR, the statistical inference based on the joint distribution [M, V, Y|X] cannot ignore the MDM expressed by [M|Y, X]. In the literature, this is said "the missing data is non-ignorable". As discussed before, although the assumptions used about missing data should be reasonable and thoroughly justified, ultimately most assumptions about missing data are not testable. Since the inference depends on the MDM assumption that cannot be verified from the observed data, a practical strategy is to do 'sensitivity analyses', in which several statistical models are considered simultaneously or a statistical model is further scrutinized using specialized tools. That is, specify a set of plausible models of the missing data mechanism in the protocol and investigate the sensitivity of the results of analysis to the models and method of handling missing values. This is more so with the NMAR situation. Unlike the well-developed literature on drawing inferences from incomplete data, the literature on sensitivity analysis to various modeling assumptions is relatively new, it is an active area of research and is evolving. More discussion on sensitivity analysis will be reviewed in Section 2.7.

2.3 Selection model and Pattern Mixture Model Approaches

There are two broad classes of models for the joint distribution of Y and M: selection models (SM), which factor the joint distribution as

$$[Y_{obs}, Y_{mis}, M \mid X] = [M \mid Y_{obs}, Y_{mis}, X] \times [Y_{obs}, Y_{mis} \mid X] \quad (2.7)$$

and pattern mixture models (PMM), which factor the full-data distribution as

$$[Y_{obs}, Y_{mis}, M \mid X] = [Y_{obs}, Y_{mis} \mid M, X] \times [M \mid X] \quad (2.8)$$

Pattern mixture models (2.8) can be further factored to make the missing data extrapolation explicit within missing data pattern *M*:

$$[Y_{obs}, Y_{mis}, M] = [Y_{mis} \mid Y_{obs}, M, X] \times [Y_{obs} \mid M, X] \times [M \mid X] \quad (2.9)$$

2.3.1 Selection Models

Selection models can be separated into two types, parametric and semi-parametric. Parametric selection models were first introduced by Rubin (1974) and Heckman (1976), based on parametric assumptions for the joint distribution of the full data (mostly, a normal distribution for responses data and a probit or a logit regression for the missing data indicators). For repeated measures, parametric selection models were proposed by Diggle and Kenward (1994), and semi-parametric models were described by Robins et al. (1995) and Rotnitzky et al. (1998). I follow the parametric models in this dissertation.

For an easier illustration of a standard formulation, we assume the full response data comprise (Y_1, Y_2). Assume Y_2 is missing on some subjects, the aim is to obtain the mean of Y_2 in each treatment group. A parametric selection model might assume that the full (hypothetical) complete response data follows a bivariate normal distribution in each treatment group (denoted by X):

$$(Y_1, Y_2) \mid X = x \sim N(\mu(x), \Sigma(x)) \quad (2.10)$$

and the 'selection mechanism' part of the model (MDM) follows a logistic regression

$$\text{logit}\{P(M = 1 \mid Y_1, Y_2, X)\} = \alpha_0 + \alpha_1 Y_1 + \alpha_2 Y_2 \quad (2.11)$$

The regression coefficients (α_0, α_1, α_2) in (2.11) could be different for different treatment groups.

In this example, if $\alpha_1 = \alpha_2 = 0$ then the MDM is MCAR. If $\alpha_2 = 0$, then the MDM is MAR. NMAR is when $\alpha_2 \neq 0$.

Model (2.10) could be fit to observed data under MAR, even though it seems no empirical information about some of the model parameters. The model could be fit because of the parametric and structural assumptions being imposed on the (hypothetical) complete data distribution [Y_1, Y_2 |X] (i.e., bivariate normal). It is beneficial but need to use extreme caution because the assumptions underlying this parametric model cannot be tested from the observed data. The major advantage is convenience. When parametric selection models fit under the MNAR assumption, the parameters identification and the sensitivity to assumptions can cause big issues: examples can be seen in the reference of Kenward (1998), Little and Rubin (2002, Chapter 15), and the discussion of Diggle and Kenward (1994).

Semi-parametric selection models do not assume a parametric model for the hypothetical completer data response distribution [Y_1, Y_2 |X], and hence less sensitive to this part of assumptions.

Selection models are very sensitive to parametric assumptions about the hypothetical complete-data distribution. But the SM approach is useful for conducting simulation studies for incomplete data by probit or logit models such as (2.11) after generating the complete data set from a known distribution (such as multivariate normal distribution).

2.3.2 Pattern Mixture Models

Pattern mixture models for repeated measures data were proposed by Little (1993, 1994) and followed by a number of extensions and generalizations. The connection between pattern mixture and selection models is described in Little and Wang (1996), Birmingham et al., (2003), and in Molenberghs et al. (1998).

The PMM approach is especially useful for data analysis since it can be viewed from an imputation perspective, in which missing values Y_{mis} can be imputed from their predictive distribution given the observed data including M; that is,

$$p(y_{mis} \mid y_{obs}, x, M) \tag{2.12}$$

Under MAR assumption, this is equivalent to $p(y_{mis} \mid y_{obs}, x)$, which is a conditional distribution obtained from the distribution of Y given X. Nevertheless, if the data is not MAR, the predictive distribution (2.12) is a direct by-product of the pattern mixture formulation because it conditions on the missing data pattern M. This more direct relationship between the pattern mixture formulation and the predictive distribution for imputations generates the benefits in transparency and computational simplicity in some situations, as described in Kenward and Carpenter (2008, Section 4.6).

Under NMAR, the SM factorization requires full specification of the model for the missing data mechanism (MDM). Little (1994) and Little and Wang (1996) illustrated that for some PM models, the specification of the model for MDM in NMAR situations can be avoided by using assumptions about the mechanism to yield restrictions on the model parameters. See more discussion in Section 2.4.

Pattern mixture formulations are suitable to sensitivity analysis because they separate the observed data distribution from the predictive distribution of the missing data given observed data. Sensitivity analyses can be formulated in terms of differences in mean or other measures between those with observed data and those with missing responses. The non-future dependent MDM is of this kind of setting, and it is the focus of this dissertation. The next section will review the definitions of non-future dependent MDM in both SM and PMM approaches. Chapter 3 will elaborate it in more detail and relate it to the multiple imputation.

2.4 Non-future Dependence Missing Data Mechanism

2.4.1 Non-future Dependence Missing Data Mechanism Definition with Selection Model Approach

Obviously there are many NMAR models as seen from the selection model setting: The missing data mechanism $[M|Y, X] = [M|Y_{obs}, Y_{mis}|X]$ depends both Y_{obs} (as for MAR) and on Y_{mis}. As there are infinitely many different ways M can depend on Y_{mis}, we shall further subclass the NMAR models. First, for longitudinal data with Y_j, j = 0, 1, 2, …, T, T+1 discrete time points, where j=0 is the baseline visit, the monotone missing data pattern can be characterized by the follow-up time, i.e., the last visit that a subject has a measurement observed, as L=0, 1, ..., T. Patients with L = j-1 has Y_j missing for all j, j+1, …,T. Patients with the pattern L = T are the completers.

A useful subclass of NMAR is called non-future dependence (NFD) missing introduced by Kenward et al. (2003). In simple terms, the non-future dependence assumption states

that the probability of dropout at time *j* can only depend on observed data up to time *j* and the possibly missing value of Y_j, and not future values beyond time *j*.

In notation, for any time-point j, the selection model factorization for L= j is then written as

$$[Y_0, Y_1, \ldots, Y_j, \ldots, Y_T, L=j] = [L= j \mid Y_0, Y_1, \ldots, Y_j, \ldots, Y_T]\,[Y_0, Y_1, \ldots, Y_j, \ldots, Y_T]$$

At the current time pint j+1, we may re-write the above expression in terms of the past history (observed): $Y_{j+1}^{-} = (Y_0, Y_1, \ldots, Y_j)$, and the current and future responses: $Y_{j+1}^{+} = (Y_{j+1}, \ldots, Y_T)$ as:

$$[Y_0, Y_1, \ldots, Y_j, \ldots, Y_T, L=j] = [L= j \mid Y_{j+1}^{-}, Y_{j+1}^{+}]\,[Y_{j+1}^{-}, Y_{j+1}^{+}] \quad (2.13)$$

If the MDM is such that

$$[L= j \mid Y_{j+1}^{-}, Y_{j+1}^{+}] = [L= j \mid Y_{j+1}^{-}, Y_{j+1}] \quad (2.14)$$

it is then called non-future dependence (NFD) missing, since the current, unobserved outcome Y_{j+1} (in the pattern L = j) depends on the past observed outcomes $Y_{j+1}^{-} = (Y_0, Y_1, \ldots, Y_j)$ as well as the current, unobserved response Y_{j+1}, but not on the future responses $Y_{j+1}^{+} = (Y_{j+2}, \ldots, Y_T)$ (Kenward et al., 2003).

2.4.2 Non-future Dependence Missing under Pattern Mixture Model

With the pattern mixture framework, NFD missing also has a corresponding form, expressed by Kenward et al. (2003) in terms of "restrictions" or "constraints" as follows:

Assume the baseline value Y_0 is always observed, and that the first missing data occurs at the first follow-up visit. (If the first missing data occurs at a visit that is beyond the first follow-up visit, then we can absorb the always observed visits into Y_0 as a vector for easy notation.) For visits $k \geq 2$ and all pattern $j < k-1$

$$[Y_k \mid Y_0, Y_1 \ldots, Y_{k-1}, L= j] = [Y_k \mid Y_0, Y_1, \ldots, Y_{k-1}, L \geq k-1] \quad (2.15)$$

Kenward (2003) showed that (2.15) is equivalent to

$$[Y_k \mid Y_0, Y_1 \ldots, Y_{k-1}, L= j] = [Y_k \mid Y_0, Y_1, \ldots, Y_{k-1}] \quad (2.16)$$

Notice that in the above condition (2.15), the distribution $[Y_k \mid Y_0, Y_1, \ldots, Y_{k-1}, L = k-1]$ was left unspecified for $k = 1, \ldots, T$.

For easy comprehension, I illustrate the conditions in (2.15) in Table 2.1 for $T = 4$.

Table 2.1 Illustration of Non-future Dependence MDM with T = 4

	k=0	k=1	k=2	k=3	k=4
L=0	x	??	?=$[Y_2 \mid Y_0, Y_1, L\geq 1]$	?=$[Y_3 \mid Y_0, Y_1, Y_2, L\geq 2]$	?=$[Y_4 \mid Y_0, Y_1, Y_3, L\geq 3]$
L=1	x	$[Y_1 \mid Y_0, L=1]$	??	?=$[Y_3 \mid Y_0, Y_1, Y_2, L\geq 2]$	?=$[Y_4 \mid Y_0, Y_1, Y_3, L\geq 3]$
L=2	x	$[Y_1 \mid Y_0, L=2]$	$[Y_2 \mid Y_0, Y_1, L=2]$	??	?=$[Y_4 \mid Y_0, Y_1, Y_3, L\geq 3]$
L=3	x	$[Y_1 \mid Y_0, L=3]$	$[Y_2 \mid Y_0, Y_1, L=3]$	$[Y_3 \mid Y_0, Y_1, Y_2, L=3]$	??
L=4	x	$[Y_1 \mid Y_0, L=4]$	$[Y_2 \mid Y_0, Y_1, L=4]$	$[Y_3 \mid Y_0, Y_1, Y_2, L=4]$	$[Y_4 \mid Y_0, Y_1, Y_3, L=4]$

?? is unidentified distribution (which implies that we need more assumption, and will relate to sensitivity analysis)

Proof of Equivalence of (2.15) and (2.16)

The proof is shown in the following. Note that there is a mistake in Kenward (2003) and is corrected in the following proof.

First, observe that, for $k \geq 2$

$[Y_k \mid Y_0, Y_1, \ldots, Y_{k-1}]$

$= \sum_{i=0}^{k-2} [Y_k \mid Y_0, Y_1, \ldots, Y_{k-1} \mid L = i][L = i] + [Y_k \mid Y_0, Y_1, \ldots, Y_{k-1} \mid L \geq k-1]\,[L \geq k-1]$ (2.17)

(a) To Show that (2.15) implies (2.16):

From Eq. (2.17),

$[Y_k \mid Y_0, Y_1, \ldots, Y_{k-1}]$

$= \sum_{i=0}^{k-2} [Y_k \mid Y_0, Y_1, \ldots, Y_{k-1} \mid L = i][L = i] + [Y_k \mid Y_0, Y_1, \ldots, Y_{k-1} \mid L \geq k-1]\,[L \geq k-1]$

$= \sum_{i=1}^{k-2} [Y_k \mid Y_0, Y_1, \ldots, Y_{k-1} \mid L \geq k-1][L = i] + [Y_k \mid Y_0, Y_1, \ldots, Y_{k-1} \mid L \geq k-1]\,[L \geq k-1]$

(from Eq. (2.15))

(Note: The above step and the following steps are corrections to that in Kenward (2003))

$= [Y_k \mid Y_0, Y_1, \ldots, Y_{k-1} \mid L \geq k-1] \left(\sum_{i=0}^{k-2} [L = i] + [L \geq k-1]\right)$

$= [Y_k \mid Y_0, Y_1, \ldots, Y_{k-1} \mid L \geq k-1]$

$= [Y_k \mid Y_0, Y_1, \ldots, Y_{k-1} \mid L = j]$ for $j < k-1$ (from Eq. (2.15)).

(b) To Show that (2.16) implies (2.15)

From (2.17), we have, for $k \geq 2$:

$[Y_k \mid Y_0, Y_1, \ldots, Y_{k-1}, L \geq k-1]\,[L \geq k-1]$

$= [Y_k \mid Y_0, Y_1, \ldots, Y_{k-1}] - \sum_{i=0}^{k-2} [Y_k \mid Y_0, Y_1, \ldots, Y_{k-1} \mid L = i][L = i]$

$= [Y_k \mid Y_0, Y_1, \ldots, Y_{k-1}] - \sum_{i=0}^{k-2} [Y_k \mid Y_0, Y_1, \ldots, Y_{k-1}][L = i]$ (from Eq. (2.16))

$= [Y_k \mid Y_0, Y_1, \ldots, Y_{k-1}] (1 - \sum_{i=0}^{k-2} [L = i])$

$= [Y_k \mid Y_0, Y_1, \ldots, Y_{k-1}, L=j] [L \geq k-1]$ for $j < k-1$ (from Eq. (2.16))

Hence, $[Y_k \mid Y_0, Y_1 \ldots, Y_{k-1}, L= j] = [Y_k \mid Y_0, Y_1, \ldots, Y_{k-1}, L \geq k-1]$ for $j < k-1$

2.5 Equivalence of Definitions using SM and PMM Approaches

In this Section, for simplicity we can omit the conditioning on X and just focus on the joint distribution [Y, M]. All the arguments will go through exactly the same when the conditioning on X is placed. For the SM approach, the factorization is [Y, M] = [M|Y] [Y], and for the PMM approach, the factorization is [Y, M] = [Y|M] [M].

If M is entirely independent of $Y = (Y_{obs}, Y_{mis})$, then the MDM is MCAR. It follows that

[Y, M] = [M|Y] [Y] = [M] [Y] = [Y|M] [M].

Therefore, MCAR definition is equivalent between SM and PMM approaches. The next two sections reviews how the MAR and NFD NMAR representations are equivalent between the SM and PMM approaches. The equivalence of representations is important since the SM approach is useful for simulation set ups, while the PMM approach is the basis used in the multiple imputation for data analysis.

2.5.1 Equivalence of MAR Definitions Between SM and PMM Approaches

(The proof in this section is shown in Shih and Aisner, 2014)

For selection models, the MAR representation is $[M|Y_{obs}, Y_{mis}] = [M|Y_{obs}]$. Thus,

$$[Y, M] = [M|Y][Y] = [M|Y_{obs}, Y_{mis}][Y] = [M|Y_{obs}][Y] \quad (2.18)$$

Under this representation, (2.18) can be re-written as

$$[Y_{obs}, Y_{mis}, M] = [M|Y_{obs}][Y_{obs}, Y_{mis}]$$

$$= \frac{[Y_{obs}|M][M]}{[Y_{obs}]}[Y_{obs}, Y_{mis}]$$

$$= [Y_{obs}|M][M]\frac{[Y_{obs}, Y_{mis}]}{[Y_{obs}]}$$

$$= [Y_{obs}, M][Y_{mis}|Y_{obs}] \quad (2.19)$$

However, $[Y_{obs}, Y_{mis}, M] = [Y_{obs}, M][Y_{mis}|Y_{obs}, M]$. Equating this to (2.19), we obtain

$$[Y_{mis}|Y_{obs}, M] = [Y_{mis}|Y_{obs}]. \quad (2.20)$$

(2.20) is the representation or condition of MAR with the PMM approach.

Reversely, when (2.20) is true, multiplying both sides of (2.20) by $[Y_{obs}, M]$ we obtain

$[Y_{obs}, M][Y_{mis}|Y_{obs}, M] = [Y_{obs}, M][Y_{mis}|Y_{obs}]$.

This leads to $[Y_{obs}, Y_{mis}, M] = [Y_{obs}, M][Y_{mis}|Y_{obs}]$

$$= [Y_{obs}|M][M]\frac{[Y_{obs}, Y_{mis}]}{[Y_{obs}]}$$

$$= \frac{[Y_{obs}|M][M]}{[Y_{obs}]}[Y_{obs}, Y_{mis}]$$

$$= [M|Y_{obs}][Y_{obs}, Y_{mis}] \quad (2.21)$$

Hence, we have [Y, M] = [M|Y][Y] = [M|Y_{obs}][Y], implying that [M|Y] = [M|Y_{obs}], which is the representation or condition of MAR with the SM approach.

2.5.2 Equivalence of NFD-NMAR Definitions Between SM and PMM Approaches

The condition or representation of NFD NMAR in the SM approach was given in Eq. (2.14):

$$[L= j \mid Y_{j+1}^{-}, Y_{j+1}^{+}] = [L= j \mid Y_{j+1}^{-}, Y_{j+1}]$$

That is, the probability of the current measurement Y_{j+1} being missing (for the pattern L=j) depends on the past observed outcomes Y_{j+1}^{-} = (Y_0, Y_1, …,Y_j) and the presently missing outcome Y_{j+1}, but not on the future outcomes (Y_{j+2}, …,Y_T).

For PMM approach, NFD missing representation is expressed in Eq. (2.15) or equivalently Eq. (2.16): For visit k ≥ 2 and pattern j < k-1,

$$[Y_k \mid Y_0, Y_1 \ldots, Y_{k-1}, L= j] = [Y_k \mid Y_0, Y_1, \ldots, Y_{k-1}]$$

Proof of the equivalence was given in Kenward et al. (2003), however, with some typos. The following is a re-write with corrections applied.

(a) Show that Eq. (2.14) implies Eq. (2.16)

Proof: Use the notation Y = (Y_0, Y_1, …, Y_T), Y_{j+1}^{-} = (Y_0, Y_1, …, Y_j), and Y_{j+1}^{+} = (Y_{j+1}, Y_{j+2}, …, Y_T) for any i ≥ 0. We first need to establish an identity shown in Eq. (2.25) as follows. With the SM factorization:

$$[Y, L= i] = [Y]\,[L= i \mid Y] = [Y]\,[L= i \mid Y_{i+1}^{-}, Y_{i+1}^{+}]$$

$$= [Y]\,[L= j \mid \bar{Y}_{i+1}, Y_{i+1}] \quad \text{(from Eq. (2.14))} \qquad (2.22)$$

Also, with the PMM factorization:

$$[Y, L= i] = [Y \mid L= i]\,[L= i]$$

$$= [Y_{i+2}, \ldots, Y_T \mid Y_0, Y_1, \ldots, Y_{i+1}, L= i]\,[Y_0, Y_1, \ldots, Y_{i+1} \mid L= i]\,[L= i]$$

$$= [Y_{i+2}, \ldots, Y_T \mid Y_0, Y_1, \ldots, Y_{i+1}, L= i]\,[L= i, Y_0, Y_1, \ldots, Y_{i+1}]$$

$$= [Y_{i+2}, \ldots, Y_T \mid Y_0, Y_1, \ldots, Y_{i+1}, L= i]\,[L= i \mid Y_0, Y_1, \ldots, Y_{i+1}]\,[Y_0, Y_1, \ldots, Y_{i+1}]$$

$$= [Y_{i+2}, \ldots, Y_T \mid \bar{Y}_{i+1}, Y_{i+1}, L= i]\,[L= i \mid \bar{Y}_{i+1}, Y_{i+1}]\,[\bar{Y}_{i+1}, Y_{i+1}] \qquad (2.23)$$

Equating (2.22) and (2.23), we have:

$$[Y] = [Y_{i+2}, \ldots, Y_T \mid \bar{Y}_{i+1}, Y_{i+1}, L= i]\,[\bar{Y}_{i+1}, Y_{i+1}] \qquad (2.24)$$

Or $\quad [\bar{Y}_{i+1}, Y_{i+1}, Y_{i+2}, \ldots, Y_T] = [Y_{i+2}, \ldots, Y_T \mid \bar{Y}_{i+1}, Y_{i+1}, L= i]\,[\bar{Y}_{i+1}, Y_{i+1}]$

This leads to

$$\frac{[\bar{Y}_{i+1}, Y_{i+1}, Y_{j+2}, \ldots, Y_T]}{[\bar{Y}_{i+1}, Y_{i+1}]} = [Y_{i+2}, \ldots, Y_T \mid \bar{Y}_{i+1}, Y_{i+1}, L= i]$$

That is, $[Y_{i+2}, \ldots, Y_T \mid \bar{Y}_{i+1}, Y_{i+1}] = [Y_{i+2}, \ldots, Y_T \mid \bar{Y}_{i+1}, Y_{i+1}, L= i]$ (2.25)

The proof of Eq. (2.16) is done by induction on the visit time-point k. Starting k=1 is trivial because there is no restriction with k=1, only the unidentified conditional distribution $[Y_1|Y_0, L=0]$ (j has to be < k-1, for $k \geq 2$ in Eq. (2.16)).

For a given j, j<k-1 implies k ≥ j+2. Assume (2.16) is true for k ≥ j+2, …, i, we then need to show that (2.16) is true for k=i+1.

So, for a given j, write (2.16) for k= j+2, …, i,

$$[Y_{j+2} \mid Y_0, Y_1\ldots, Y_{j+1}, L= j]= [Y_{j+2} \mid Y_0, Y_1, \ldots, Y_{j+1}]$$

$$[Y_{j+3} \mid Y_0, Y_1\ldots, Y_{j+2}, L= j]= [Y_{j+3} \mid Y_0, Y_1, \ldots, Y_{j+2}]$$

:

:

$$[Y_i \mid Y_0, Y_1\ldots, Y_{i-1}, L= j]= [Y_i \mid Y_0, Y_1, \ldots, Y_{i-1}]$$

Multiplying the above identities their left-hand sides and right-hand sides,

$$[Y_{j+2} \mid Y_0, Y_1\ldots, Y_{j+1}, L= j]\,[Y_{j+3} \mid Y_0, Y_1\ldots, Y_{j+2}, L= j]\ldots..[Y_i \mid Y_0, Y_1\ldots, Y_{i-1}, L= j]$$

$$= [Y_{j+2} \mid Y_0, Y_1, \ldots, Y_{j+1}]\,[Y_{j+3} \mid Y_0, Y_1, \ldots, Y_{j+2}]\ldots.[Y_i \mid Y_0, Y_1, \ldots, Y_{i-1}]$$

we have:

$$[Y_{j+2}, Y_{j+3}, \ldots,Y_i \mid Y_0, Y_1\ldots, Y_{j+-1}, L= j] = [Y_{j+2}, Y_{j+3}, \ldots,Y_i \mid Y_0, Y_1\ldots, Y_{j+-1}]$$

Or, using the abbreviated notation:

$$[Y_{j+2}, \ldots,Y_i \mid \bar{Y_{j+1}}, Y_{j+1}, L= j] = [Y_{j+2}, \ldots,Y_i \mid \bar{Y_{j+1}}, Y_{j+1}] \qquad (2.26)$$

Next, notice that Eq. (2.25) is true for any i, hence is true for the given j; switching sides of (2.25) and applying it to the given j, we have:

$$[Y_{j+2}, \ldots,Y_{i+1}, \ldots,Y_T \mid \bar{Y_{j+1}}, Y_{j+1}, L= j] = [Y_{j+2}, \ldots,Y_{i+1}, \ldots,Y_T \mid \bar{Y_{j+1}}, Y_{j+1}]$$

Integrate the above equation over $(Y_{i+2}, \ldots, Y_T)$, we have

$$[Y_{j+2}, \ldots, Y_{i+1} | Y_{j+1}^{-}, Y_{j+1}, L= j] = [Y_{j+2}, \ldots, Y_{i+1} | Y_{j+1}^{-}, Y_{j+1}] \quad (2.27)$$

Dividing (2.27) by (2.26) gives:

$$\frac{[Y_{j+2},\ldots,Y_{i+1} | Y_{j+1}^{-},Y_{j+1},L= j]}{[Y_{j+2},\ldots,Y_i \mid Y_{j+1}^{-},Y_{j+1},L= j]} = \frac{[Y_{j+2},\ldots,Y_{i+1} | Y_{j+1}^{-},Y_{j+1}]}{[Y_{j+2},\ldots,Y_i \mid Y_{j+1}^{-}, Y_{j+1}]}$$

This leads to $[Y_{i+1} | Y_{j+1}^{-}, Y_{j+1}, L= j] = [Y_{i+1} | Y_{j+1}^{-}, Y_{j+1}]$

The induction is done.

(b) Show that Eq. (2.16) implies Eq. (2.14)

Starting from factorization

$$[Y, L= j] = [Y_0, Y_1, \ldots, Y_{j+1}, L= j] [Y_{j+2}, \ldots, Y_T | Y_0, Y_1, \ldots, Y_{j+1}, L= j]$$

$$= [Y_0, Y_1, \ldots, Y_{j+1}, L= j] [Y_{j+2} | Y_0, Y_1, \ldots, Y_{j+1}, L= j] [Y_{j+3}, \ldots, Y_T | Y_0, Y_1, \ldots, Y_{j+2}, L= j]$$

$$= [Y_0, Y_1, \ldots, Y_{j+1} \mid L= j] [L= j] \prod_{k=j+2}^{T} [Y_k | Y_0, Y_1, \ldots, Y_{k-1}, L= j] \quad (2.28)$$

Applying Eq. (2.16) to the above, we have

$$[Y, L= j] = [Y_0, Y_1, \ldots, Y_{j+1} \mid L= j] [L= j] \prod_{k=j+2}^{T} [Y_k | Y_0, Y_1, \ldots, Y_{k-1}]$$

$$= [Y_{j+1}^{-}, Y_{j+1} \mid L= j] [L= j] [Y_{j+2}, \ldots, Y_T | Y_0, Y_1, \ldots, Y_{j+1}]$$

$$= \frac{[Y_{j+1}^{-},Y_{j+1} \mid L= j] [L= j]}{[Y_{j+1}^{-},Y_{j+1}]} [Y_{j+1}^{-}, Y_{j+1}] [Y_{j+2}, \ldots, Y_T | Y_{j+1}^{-}, Y_{j+1}]$$

$$= \frac{[\,Y_{j+1}^{-}, Y_{j+1} \mid L= j]\,[L= j]}{[\,Y_{j+1}^{-}, Y_{j+1}]}\,[Y]$$

$$\ldots\ldots = [L= j \mid Y_{j+1}^{-}, Y_{j+1}^{+}]\,[Y] \qquad (2.29)$$

Since [Y, L= j] = [Y] [L= j |Y], we have from (2.29):

$$[L= j \mid Y] = [L= j \mid Y_{j+1}^{-}, Y_{j+1}],$$

which is the SM representation, Eq. (2.14).

2.6 Multiple Imputation Method with MAR Assumption

In the following, I will review the multiple imputation (MI) method under the MAR assumption due to Rubin (1987). This will then be the basis for the main analysis under the NFD NMAR assumption in Chapter 3. We assume that the missing data pattern is monotone as the situation of dropouts in clinical trials.

For a data set with monotone missing data pattern, the multiple imputation method is generally carried out in three steps: (a) Use a multivariate model to impute (fill in) the missing values *m* times to obtain *m* complete data sets; (b) For each of the *m* complete data sets, perform the usual analysis for complete data (such as mixed-effect model for repeated measures, or MMRM in short), obtain the mean and within-imputation and between-imputation variances of the outcome measurements; and (c) Combine the *m* complete-data estimates by incorporating the within and between-imputation variances.

The multivariate normal distribution is usually assumed for step (a) for the continuous response variables (with missing data) and other continuous covariates (such as age) that

are observed. This assumption can be relaxed with categorical covariates without missing data (such as sex, center, etc.)

In step (a), for a variable with missing values, a model is fitted using observations with observed values for that variable. With this resulting model, a new model is fitted and used to impute missing values for the variable. In SAS procedure Proc MI, missing values are imputed sequentially for variables with missing values in the order specified by VAR statement. That is, for a variable Y_j with missing values, the missing values are imputed (drawn) from the predictive distribution:

$$Y_j \sim P(Y_j \mid Y_0, Y_1, Y_2, \ldots, Y_{j-1})$$

One example is regression method for monotone missing data. In the regression method, a regression model is fitted for each variable with missing values, with the previous variables as covariates (we can also include other covariates, as people usually do). Based on the fitted regression model and its coefficients, a new regression model is drawn from the posterior predictive distribution of the parameters and it is used to impute the missing values for the variable (Rubin 1987). Repeat this process sequentially for each variable with missing values.

$$Y_j = \beta_0 + \beta_1 u_1 + \beta_2 u_2 + \ldots + \beta_k u_k$$

where $u_1, u_2, \ldots, u_k$ are the covariates generated from preceding variables $Y_0, Y_1, \ldots, Y_{j-1}$.

The following steps are carried out to impute missing values for Y_j at each imputation and implemented in SAS® proc MI (referred to SAS/STAT manual):

1. Fit a regression model using observed values for the variable Y_j and its covariates $u_1, u_2, \ldots, u_k$. The regression parameter estimates for the fitted regression model is $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k)$, and the associated covariance matrix is $\hat{\sigma}^2_j \mathrm{V_j}$, where $\mathrm{V_j}$ is the usual $u'u$ inverse matrix derived from the intercept and covariates $u_1, u_2, \ldots, u_k$.
2. New parameters $\beta_* = (\beta_{*0}, \beta_{*1}, \ldots, \beta_{*k})$ and $\hat{\sigma}^2_{*j}$ are drawn from the posterior predictive distribution of the parameters (Rubin 1987). They are simulated from $(\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k)$, $\hat{\sigma}^2_j$, and $\mathrm{V_j}$. The variance is derived as

   $$\hat{\sigma}^2_{*j} = \hat{\sigma}^2_j (n_j - k - 1) / g$$

   where g is a $\chi^2_{n_j - k - 1}$ random variate and n_j is the number of observations with non-missing values for Y_j. The regression coefficients are derived as

   $$\beta_* = \hat{\beta} + \sigma_{*\mathrm{j}} \, \mathrm{V'_{hj}} \, \mathrm{Z}$$

   where $\mathrm{V'_{hj}}$ is the upper triangular matrix in the Cholesky decomposition, $\mathrm{V_j} = \mathrm{V'_{hj}} \mathrm{V_{hj}}$, and Z is a vector of $k + 1$ independent random normal variates.
3. The missing values are then replaced by

   $$\beta_{*0} + \beta_{*1} u_1 + \beta_{*2} u_2 + \ldots + \beta_{*k} u_k + z_i \sigma_{*\mathrm{j}}$$

   where $u_1, u_2, \ldots, u_k$ are the values of the covariates and z_i is a simulated normal deviate.

Similar to the regression method, the predictive mean matching method (regpmm) can also be used for imputation. For each missing value, it imputes an observed value that is selected from the specified number of closest observations to the predicted value from the simulated regression model (Rubin 1987). When the normality assumption is violated,

the predictive mean matching method is more appropriate than the regression method (Horton and Lipsitz 2001).

Multiple imputation has some advantages over single imputation methods. It overcomes the major drawbacks of single imputation methods which assumes that each missing values are exactly known and fails to account for any potential variability. For example, mixed effect model based on PROC MIXED is a single imputation method and does not take into account the uncertainty associated with imputed values and may underestimate the standard errors of statistical point estimates. In multiple imputation, data are imputed based on a statistical model, and the parameters of that imputation model are not certain quantities but rather have a distribution associated with them. MI generates multiple imputed dataset and each of these datasets is imputed using predictions from a slightly different imputation model, parameters of which are sampled from a common estimated Bayesian posterior distribution, that is, imputed data are not certain. MI incorporates this uncertainty through multiple 'plausible' imputations for each missing value.

In step (b), once multiple imputed datasets are obtained, each dataset is analyzed using a method chosen for complete data (e.g. analysis of covariance or MMRM with PROC MIXED in SAS).

In step (c), use Rubin's rule to combine the results of these analyses from individual datasets to perform an overall inference that takes into account the uncertainty and variability of imputations.

With *m* imputations, *m* different sets of the point and variance estimates for a parameter *Q* can be computed. Suppose $\hat{Q}_i$ and $\widehat{W}_i$ are the point and variance estimates from the *i*th

imputed data set, i=1, 2, ..., m. Then the combined point estimate for Q from multiple imputations is the average of the m complete-data estimates:

$$\bar{Q} = \frac{1}{m} \sum_{i=1}^{m} \hat{Q}_I \tag{2.30}$$

Suppose $\bar{W}$ is the within-imputation variance, which is the average of the m complete-data estimates:

$$\bar{W} = \frac{1}{m} \sum_{i=1}^{m} \hat{W}_i \tag{2.31}$$

and B is the between-imputation variance

$$B = \frac{1}{m-1} \sum_{i=1}^{m} (\hat{Q}_i - \bar{Q})^2 \tag{2.32}$$

Then the variance estimate associated with Q is the total variance (Rubin 1987)

$$T = \bar{W} + (1 + \frac{1}{m}) B \tag{2.33}$$

where the multiplier $(1 + \frac{1}{m})$ is the finite number (m) correction. As m increases, this correction approaches to 1.

The ratio

$$r = \frac{\left(1 + \frac{1}{m}\right) B}{\bar{W}}$$

is called the relative increase in variance due to non-response (Rubin 1987).

In the clinical trial setting, MI has another advantage. It allows to use an imputation model that is different from the analysis model applied to the imputed data. This allows to account for factors that are potentially correlated with missingness, but are not of interest for the inference regarding the treatment effect on the primary efficacy outcome.

For analyses with MAR assumption, it can be directly implemented in SAS PROC MI and PROC MIANALYZE. For analyses with a particular NMAR assumption, the imputation step in SAS® needs to be augmented by a specific NMAR model.

2.7 Sensitivity Analysis

As mentioned in Chapter 1, there are different types of sensitivity analyses. It is important to first qualify what may be and what may not be a sensitivity analysis in the trial protocol or the data analysis plan of a study. Morris, Kahan and White (2014) provided the following principles in terms of three questions to consider in practice:

1. *Does the proposed sensitivity analysis address the same question as the primary analysis?*
2. *Is it possible for the proposed sensitivity analysis to arrive at a different result to the primary analysis?*
3. *If the proposed sensitivity analysis leads to a different result, is there a genuine degree of uncertainty as to which will be believed?*

To qualify as a sensitivity analysis, the answer to all of the three questions should be 'yes'.

If an analysis is to address an important but different question to the primary analysis, it should be classified as a secondary analysis instead of a sensitivity analysis. An example is the analysis of different causal estimand. ITT analysis and AAT or ACT analyses are to answer different questions regarding the treatment effect for different target population. If one of the three estimand is the primary analysis, then the others are secondary, not sensitivity analyses. Considering them as sensitivity analyses may lead to false anxiety about the robustness of the results.

It is not helpful or sometimes even dangerous for analyses that will always lead to the same conclusion as the primary analysis, since they then give a false measure of the robustness of results. For example, if we use multiple imputation with the same regression model, but only vary the number of imputations from 5 to 10, then only a minor degree of the results will turn out, but the conclusion will always be the same since all analyses make the same assumptions.

Assume the first two answers are affirmative: the proposed analysis addresses the same question as the primary analysis and can lead to different conclusions. Then we consider the third question. If it is clear that one analysis would always be preferred over the other, then the preferred one should be the primary analysis and the other one should not be done. For example, a multiple imputation analysis is obviously always preferred compared to a single imputation analysis with the same models. A single imputation analysis may give a different result since it may give an inappropriate, smaller standard error.

Scharfstein, et al. (2014) classified three main types of sensitivity analysis. (a) Ad-hoc sensitivity analysis involves analyzing the data using a few different methods such as LOCF, complete-case analysis (CCA), MMRM, and MI. The underlying assumptions are very different for some of these methods. LOCF is a very strong, almost unreasonable NMAR model that assumes no variability after withdrawal for a subject. CCA assumes MCAR, while both MMRM and MI assume MAR. (b) Local sensitivity analysis; and (c) Global sensitivity analysis.

Local sensitivity analysis was due to Troxel, Ma and Heitjan (2004) and Ma, Troxel and Heitjan (2005), in which they developed an index of local sensitivity to the non-ignorable dropout process, i.e., MDM. It is to use with the selection model approach, $[Y, M|X] = [M|Y, X; \varphi] [Y|X; \theta] = [M|Y_{obs}, Y_{mis}, X; \varphi] [Y|X; \theta]$, where a perturbation is applied to the parameter φ of the dropout process $[M|Y_{obs}, Y_{mis}, X; \varphi]$ and derives its 'local influence' on the parameter of interest (i.e., treatment effect) θ. An example is seen in the MDM in (2.11):

$$\text{logit}\{P(M = 0 \mid Y_1, Y_2, X)\} = \alpha_0 + \alpha_1 Y_1 + \alpha_2 Y_2$$

A perturbation can be applied to the regression coefficient α_2.

In this framework, it seems that the non-ignorability of the MDM comes from the violation of the distinct parameter condition, as defined in Rubin (1976). That is, somehow, φ is not separated from θ, hence perturbing φ will affect θ. In fact, in the above example, the notion of the (local) sensitivity analysis of Troxel, Ma and Heitjan (2004) is to evaluate an estimate of θ (e.g., MLE) as a function of α_2 when α_2 departs from 0 (MAR). On the other hand, even when φ and θ are sets of distinct parameters, since

$$[Y, M|X; \theta, \varphi] = [M|Y_{obs}, Y_{mis}, X; \varphi]\,[Y_{obs}|X; \theta]\,[Y_{mis}|\, Y_{obs}, X; \theta, \varphi]$$

the full-data log-likelihood function of (θ, φ) can be written as the sum of the above corresponding term as the following:

$$\ell(\theta, \varphi) = \ell_1(\varphi) + \ell_2(\theta) + \ell_3(\theta, \varphi)$$

where the third term is the log-likelihood based on the conditional probability of the unobserved response giving the observed response history of the individual, which depends on the missing data process, hence involves both (θ, φ). It is through this term that a perturbation of φ will influence the estimation of θ.

Other local sensitivity analyses such as Verbeke et al. (2001) and Molenberghs et al (2003) use essentially the same method as Troxel et al. (2004) and Ma et al. (2005), only to make the perturbation at the individual subject's level. For example, in the example of model (2.11), make α_2 as α_{2i} for each patient *i* who drops out. Then the influence on θ is aggregated by using Cook's (1986) local influence assessment method.

The local sensitivity analysis is to assess of inferences in a neighborhood of MAR.

The so-called global sensitivity is not very different, in my opinion, from the local sensitivity. The hypothetical complete data model part, $[Y|X; \theta]$, is still fully specified, and the sensitivity analysis is still to perturb on the MDM part. However, the literature on global sensitivity analysis mostly focused mainly on the PMM approach, although there is always an equivalent representation of the MDM for the SM approach. In the global sensitivity analysis, as in the local sensitivity analysis, we have a set of identified

parameters (that can be estimated from the observed data), and a set of unidentifiable parameters (that cannot be estimated from the observed data). In the example of the MDM (2.11), the parameter α_2 is an unidentifiable parameter as it is the regression coefficient of the possibly missing data Y_2. The only additional part of global sensitivity analysis, from what I can tell, based on discussions in Daniel and Hogan (2008) and Scharfstein et al. (2014) is that the observed-data model part could be semi-parametric and that a Bayesian prior distribution could be put on the sensitivity parameter to reflect the uncertainty or certainty about the MDM part.

This dissertation also follows the same thread as in the NRC's report to focus on the sensitivity analysis regarding the untestable assumption of MDM in clinical trial. For example, if the primary analysis is based on the assumption of missing at random (MAR) mechanism, then the sensitivity analysis would be to allow departures from MAR. If the primary analysis is based on a particular not missing not at random (NMAR) mechanism, then the sensitivity analysis would be to assess different degrees of NMAR and their impact on the results.

Chapter 3

NFD-Delta Method and Sensitivity Analysis Using Multiple Imputation with Pattern Mixture Model Approach

3.1 Preliminary

NFD-Delta Method is to address the non-future dependent missing data. In addition to the NFD-NMAR model for the missing data mechanism, it also assumes the mean-shift model. The first step is to determine the causal estimand (AAT), then the delta shift as the next step. The delta shift involves options of the unit, the magnitude, and the algorithm. I will systematically investigate these options using simulations in Chapter 4. In this Chapter, I will detail the procedures for AAT first and ACT next using a prototype data structure as shown in Table 3.1. For ACT estimand, the control-based imputation will be used without the delta-shift (see Section 3.4).

In general, consider a study with T scheduled post-baseline visits. Let Y_k denote the outcome at visit k, k =0, 1, 2, … T, where k=0 is the baseline visit. For longitudinal data with missing data caused by dropouts, as commonly the case in clinical trials, a monotone pattern is formed naturally. Under the monotone missing data pattern, if the subject's outcome Y_k is missing at visit k, then the outcome Y_j at the future visit j>k is also missing. For this discrete time model, let L be the follow-up time, i.e., the last visit that a subject has a measurement observed. The monotone missing data pattern is then characterized by L=0, 1, …, T, where L=T is when the subject completes the study visits. We assume the baseline values are always observed. A feasible analysis in this situation

is to base it on the PMM approach's factorization. We discuss the method with the case T=5 as displayed in Table 3.1.

Table 3.1. Monotone missing data for T=5

	Y_k					
	k=0	k=1	k=2	k=3	k=4	K=5
L=0	x	?	?	?	?	?
L=1	x	x	?	?	?	?
L=2	x	x	x	?	?	?
L=3	x	x	x	x	?	?
L=4	x	x	x	x	x	?
L=5	x	x	x	x	x	x

x denotes observed, ? denotes missing.

With pattern mixed model under MAR, the conditional distributions of the variables with missing data (Y_{mis}) given observed variables (Y_{obs}) are the same across the missing data pattern (i.e., independent of the pattern). For example, in the case of Table 3.1, for the observations in pattern L=0 only Y_0 (baseline) is observed, we need to impute all other follow-up Y_1 to Y_5. Under MAR with PMM approach:

$[Y_1, Y_2, Y_3, Y_4, Y_5|Y_0, L=0] = [Y_1, Y_2, Y_3, Y_4, Y_5|Y_0]$

$= [Y_5| Y_0, Y_1, Y_2, Y_3, Y_4][Y_1, Y_2, Y_3\ Y_4|Y_0]$

$= [Y_5| Y_0, Y_1, Y_2, Y_3, Y_4, L=5][Y_4|Y_0, Y_1, Y_2, Y_3][Y_1, Y_2, Y_3|Y_0]$

$= [Y_5| Y_0, Y_1, Y_2, Y_3, Y_4, L=5][Y_4|Y_0, Y_1, Y_2, Y_3, L\geq 4][Y_3|Y_0, Y_1, Y_2][\ Y_1, Y_2, |Y_0]$

$= [Y_5| Y_0, Y_1, Y_2, Y_3, Y_4, L=5][Y_4|Y_0, Y_1, Y_2, Y_3, L\geq 4][Y_3|Y_0, Y_1, Y_2,$

$L\geq 3][Y_2|Y_0,Y_1][Y_1|Y_0]$

$$= [Y_5| Y_0, Y_1, Y_2, Y_3, Y_4, L=5][Y_4|Y_0, Y_1, Y_2, Y_3, L\geq 4][Y_3|Y_0, Y_1, Y_2, L\geq 3][Y_2|Y_0,Y_1, L\geq 2][Y_1|Y_0, L\geq 1] \quad (3.1)$$

Hence the joint distribution of $[Y_1, Y_2, Y_3, Y_4, Y_5|Y_0, L=0]$ is estimable since all the factors in the right hand side of (1.6) are identifiable from the observed data. Same factorization and application of the MAR to the variables with missing values are performed for patterns L=1, 2, 3, 4. Missing values are sampled from the predictive/conditional distributions. This is the basis for the multiple imputation method. The analysis based on the multiple imputation procedures (PROC MI and PROC MIANALYZE) are directly applicable and most useful for imputing continuous outcome variables, as reviewed in Chapter 2.

3.2 Data Analysis Under NFD using PMM: NFD-Delta Method

The NFD NMAR defined in Eq. (2.15) or Eq. (2.16) is not a comprehensive set of restrictions. Instead, there is one (and only one) conditional distribution (i.e., regression model) left unidentified, that is: $[Y_{k+1} | Y_0, Y_1\ldots, Y_k, L= k]$ for all follow-up visits $k \geq 1$, since for L=k, Y_{k+1} is the current, unobserved outcome. See the explanation in the next paragraph. Therefore, we need to make an assumption for this unidentified model to proceed. As in the MAR case, we link it with the L>k patterns where Y_{k+1} is observed. One such an assumption is the mean shift model as a simple departure from MAR. This assumption is not verifiable from the observed data, and requires sensitivity analyses based on varying the shift values.

Formally, we write the whole joint distribution with the pattern mixture factorization:

$$[Y_0, Y_1, \ldots, Y_j, \ldots, Y_T] = [Y_0, Y_1, \ldots, Y_j, \ldots, Y_T | L= j]\,[L= j]$$

The pattern-specific joint distribution $[Y_0, Y_1, \ldots, Y_j, \ldots, Y_T | L= j]$ can be written as

$$[Y_0, Y_1, \ldots, Y_j, \ldots, Y_T | L= j] = [Y_0, Y_1, \ldots, Y_j | L= j]\,[Y_{j+1} | Y_0, Y_1, \ldots, Y_j, L= j] \times \prod_{k=j+2}^{T} [Y_k | Y_0, \ldots, Y_{k-1}, L = j] \quad (3.2)$$

The first factor of (3.2) is clearly identifiable from the observed data. The second and beyond are not, due to missing data and require additional assumptions. The second factor could be identified by linking it to the observed $[Y_{j+1} | Y_0, Y_1, \ldots, Y_j | L \geq j+1]$ by a shift parameter model (see Eq. (3.7) later). The third and beyond factors could be identified with the help of the NFD mechanism assumption using (2.15) together with the same link as for the second factor as follows: Via (2.15), for $k \geq j+2$,

$$[Y_k | Y_0, Y_1, \ldots, Y_{k-1}, L= j] = [Y_k | Y_0, Y_1, \ldots, Y_{k-1}, L \geq k-1].$$

Furthermore, the right hand side $[Y_k | Y_0, Y_1, \ldots, Y_{k-1}, L \geq k-1]$

$$= \sum_{s=k-1}^{T} \frac{[Y_0, Y_1, \ldots, Y_{k-1}, L=s]}{[Y_0, Y_1, \ldots, Y_{k-1}, L \geq k-1]} [Y_k | Y_0, Y_1, \ldots, Y_{k-1}, L=s] \quad (3.3)$$

and $\frac{[Y_0, Y_1, \ldots, Y_{k-1}, L=s]}{[Y_0, Y_1, \ldots, Y_{k-1}, L \geq k-1]} = \frac{P(L=s)[Y_0, Y_1, \ldots, Y_{k-1} | L=s]}{\sum_{s=k-1}^{T} P(L=s)[Y_0, Y_1, \ldots, Y_{k-1} | L=s]}$

Notice that $[Y_0, Y_1, \ldots, Y_{k-1} | L=s]$ is identifiable from the observed data for $s \geq k-1$. $[Y_k | Y_0, Y_1, \ldots, Y_{k-1}, L=s]$ is identifiable from the observed data for $s \geq k$. The unidentified $[Y_k | Y_0, Y_1, \ldots, Y_{k-1}, L=k-1]$ is linked to $[Y_k | Y_0, Y_1, \ldots, Y_{k-1}, L \geq k]$ by the same shift model as for the second factor. Thus, all factors in (3.2) are identifiable. The mean shift model as a simple yet useful way to describe the departure from MAR is expressed as

$$E[Y_k \mid Y_0, Y_1,\ldots, Y_{k-1}, L=k-1] = E[Y_k \mid Y_0, Y_1,\ldots, Y_{k-1}, L\geq k] + \Delta_k. \quad (3.4)$$

An illustration of the Specific Analysis under NFD with T=5

We continue the pattern mixture approach to the analysis of data with a monotone pattern, since we shall make a link between the NFD missing to the MAR model. We still use Table 3.1 example with T=5 for illustration. First, for MAR, we have

$$[Y_1|Y_0, L=0] = [Y_1|Y_0, L>0]$$

$$[Y_2|Y_0, Y_1, L=1] = [Y_2|Y_0, Y_1, L>1] \quad (3.5)$$

$$[Y_3|Y_0, Y_1, Y_2, L=2] = [Y_3|Y_0, Y_1, Y_2, L>2]$$

$$[Y_4|Y_0, Y_1, Y_2, Y_3, L=3] = [Y_4|Y_0, Y_1, Y_2, Y_3, L>3]$$

$$[Y_5|Y_0, Y_1, Y_2, Y_3, Y_4, L=4] = [Y_5|Y_0, Y_1, Y_2, Y_3, Y_4, L>4]$$

In the above, for the pattern L=j, the predictive/conditional distribution at the right-hand side can be obtained from the observed data in the pattern L>j. Missing values in the pattern L=j (left hand size) are sampled from the predictive/conditional distribution of the right hand size, as performed by the multiple imputation procedures (PROC MI and PROC MIANALYZE) as discussed in previous sections.

With NFD missing, we shall make additional assumptions to link between the distribution of the dropouts and that of the observed. The assumption about the observables can and should be checked with the data, but the link itself cannot be verified with the data and requires sensitivity analyses. A simple way to describe the departure from MAR is by a shift model in the conditional means. This mean shift-model

assumption cannot be verified from the data itself. The justification for this choice is based mainly in interpretability. It works as follows. First, assume linear regression (i.e. conditional mean) models for the observed (right hand side of 3.5), which can and should be checked with the observed data:

$$E(Y_1|Y_0, L>0) = \alpha_1 + \beta_1 Y_0$$

$$E(Y_2|Y_0, Y_1, L>1) = \alpha_2 + \beta_{20} Y_0 + \beta_{21} Y_1 \qquad (3.6)$$

$$E(Y_3|Y_0, Y_1, Y_2, L>2) = \alpha_3 + \beta_{30} Y_0 + \beta_{31} Y_1 + \beta_{32} Y_2$$

$$E(Y_4|Y_0, Y_1, Y_2, Y_3, L>3) = \alpha_4 + \beta_{40} Y_0 + \beta_{41} Y_1 + \beta_{42} Y_2 + \beta_{43} Y_3$$

$$E(Y_5|Y_0, Y_1, Y_2, Y_3, Y_4, L>4) = \alpha_5 + \beta_{50} Y_0 + \beta_{51} Y_1 + \beta_{52} Y_2 + \beta_{53} Y_3 + \beta_{54} Y_4$$

All the intercepts and slope coefficients are directly obtainable from the observed data. Next, to embed the MAR specification in this large class of NMAR models, the class is indexed by shifts in linking the dropouts with the observed conditional means. However, assume that the shift is in the intercept, not in the slops or the residual variances (Daniels and Hogan, 2008, page 240 [9]; National Research Council, 2010, page 100). That is,

$$E(Y_1|Y_0, L= 0) = (\alpha_1+\Delta_1) + \beta_1 Y_0 = E(Y_1|Y_0, L>0) + \Delta_1 \text{ for all } Y_0$$

$$E(Y_2|Y_0, Y_1, L=j) = (\alpha_2 + \Delta_{2j}) + \beta_{20} Y_0 + \beta_{21} Y_1$$

$$= E(Y_2|Y_0, Y_1, L>1) + \Delta_{2j} \text{ for } j=0, 1 \text{ and all } Y_0, Y_1$$

$$E(Y_3|Y_0, Y_1, Y_2, L=j) = (\alpha_3 + \Delta_{3j}) + \beta_{30} Y_0 + \beta_{31} Y_1 + \beta_{32} Y_2 \qquad (3.7)$$

$$= E(Y_3|Y_0, Y_1, Y_2, L>2) + \Delta_{3j} \quad \text{for } j=0, 1, 2 \text{ and all } Y_0, Y_1, Y_2$$

$$E(Y_4|Y_0, Y_1, Y_2, Y_3, L=j) = (\alpha_4 + \Delta_{4j}) + \beta_{40}Y_0 + \beta_{41}Y_1 + \beta_{42}Y_2 + \beta_{43}Y_3$$

$$= E(Y_4|Y_0, Y_1, Y_2, Y_3, L>3) + \Delta_{4j} \quad \text{for } j=0, 1, 2, 3 \text{ and all } Y_0, Y_1, Y_2, Y_3$$

$$E(Y_5|Y_0, Y_1, Y_2, Y_3, Y_4, L=j) = (\alpha_5 + \Delta_{5j}) + \beta_{50}Y_0 + \beta_{51}Y_1 + \beta_{52}Y_2 + \beta_{53}Y_3 + \beta_{54}Y_4$$

$$= E(Y_5|Y_0, Y_1, Y_2, Y_3, Y_4, L>4) + \Delta_{5j} \quad \text{for } j=0, 1, 2, 3, 4 \text{ and all } Y_0, Y_1, Y_2, Y_3, Y_4$$

Furthermore, an additional simplification assumption is that $\Delta_{2j} = \Delta_2$ for j=0, 1; $\Delta_{3j} = \Delta_3$ for j=0, 1, 2; $\Delta_{4j} = \Delta_4$ for j=0, 1, 2, 3; and $\Delta_{5j} = \Delta_5$ for j=0, 1, 2, 3, 4. That is, shift is equivalent across patterns for each visit with missing data. This is derived from the characteristic of the NFD missing; See Eq. (2.15) and Eq. (2.16) in Chapter 2. Since the regression parameters are all obtainable, the left-hand side of (3.7) are estimable when the shift parameters $\Delta = (\Delta_1, \Delta_2, \Delta_3, \Delta_4, \Delta_5)$ are specified. The missing data in pattern L=j will be filled in sequentially with the conditional mean in (3.7). Computationally, since (3.7) (for NFD missing) is only a shift of (3.6) (for MAR), we can utilize the multiple imputation procedure based on MAR and apply shifts to the imputed values obtained from PROC MI.

Analytically, we show how to obtain the marginal means of the outcome at each visit in each pattern and the marginal mean at the final visit over the patterns. Denote $\mu_k^{(j)} = E(Y_k| L = j)$. It follows from (3.7) that for each follow-up visit post baseline we can sequentially obtain the patter-specific marginal means:

$$\mu_1^{(o)} = E(Y_1| L= 0) = (\alpha_1+\Delta_1) + \beta_1 \mu_0^{(o)} \qquad (3.8)$$

where $\mu_0^{(0)} = E(Y_0 | L = 0)$ is estimated directly from the patter-specific observed data.

$$\mu_2^{(j)} = E(Y_2 | L = j) = (\alpha_2 + \Delta_2) + \beta_{20}\mu_0^{(j)} + \beta_{21}\mu_1^{(j)} \quad \text{for } j=0, 1 \tag{3.9}$$

where $\mu_0^{(0)}, \mu_0^{(1)}, \mu_1^{(1)}$ are estimated directly from data, $\mu_1^{(0)}$ is obtained from (3.8)

$$\mu_3^{(j)} = E(Y_3 | L = j) = (\alpha_3 + \Delta_3) + \beta_{30}\mu_0^{(j)} + \beta_{31}\mu_1^{(j)} + \beta_{32}\mu_2^{(j)} \quad \text{for } j=0, 1, 2 \tag{3.10}$$

where $\mu_0^{(0)}, \mu_0^{(1)}, \mu_0^{(2)} \mu_1^{(1)}, \mu_1^{(2)}, \mu_2^{(2)}$ are estimated directly from data, $\mu_1^{(0)}$ is obtained from (1.15) and $\mu_2^{(0)}, \mu_2^{(1)}$ are obtained from (3.9)

$$\mu_4^{(j)} = E(Y_4 | L = j) = (\alpha_4 + \Delta_4) + \beta_{40}\mu_0^{(j)} + \beta_{41}\mu_1^{(j)} + \beta_{42}\mu_2^{(j)} + \beta_{43}\mu_3^{(j)} \quad \text{for } j=0, 1, 2, 3 \tag{3.11}$$

where $\mu_0^{(0)}, \mu_0^{(1)}, \mu_0^{(2)}, \mu_0^{(3)}, \mu_1^{(1)}, \mu_1^{(2)}, \mu_1^{(3)}, \mu_2^{(2)}, \mu_2^{(3)}, \mu_3^{(3)}$ are estimated directly from data, while $\mu_1^{(0)}$ is obtained from (3.8), $\mu_2^{(0)}, \mu_2^{(1)}$ are obtained from (3.9), and $\mu_3^{(0)}, \mu_3^{(1)}, \mu_3^{(2)}$ are obtained from (3.10).

$$\mu_5^{(j)} = E(Y_5 | L = j) = (\alpha_5 + \Delta_5) + \beta_{50}\mu_0^{(j)} + \beta_{51}\mu_1^{(j)} + \beta_{52}\mu_2^{(j)} + \beta_{53}\mu_3^{(j)} + \beta_{54}\mu_4^{(j)} \quad \text{for } j=0, 1, 2, 3, 4 \tag{3.12}$$

where $\mu_0^{(0)}, \mu_0^{(1)}, \mu_0^{(2)}, \mu_0^{(3)}, \mu_0^{(4)}, \mu_1^{(1)}, \mu_1^{(2)}, \mu_1^{(3)}, \mu_1^{(4)}, \mu_2^{(2)}, \mu_2^{(3)}, \mu_2^{(4)}, \mu_3^{(3)}, \mu_3^{(4)}, \mu_4^{(4)}$ are estimated directly from data, while $\mu_1^{(0)}$ is obtained from (3.8), $\mu_2^{(0)}, \mu_2^{(1)}$ are obtained from (3.9), $\mu_3^{(0)}, \mu_3^{(1)}, \mu_3^{(2)}$ are obtained from (3.10), and $\mu_4^{(0)}, \mu_4^{(1)}, \mu_4^{(2)}, \mu_4^{(3)}$ are obtained from (3.11).

In many clinical trials, the primary interest is on the marginal mean of the final visit at the last time-point T. For the illustration of T=5,

$$E(Y_5) = \sum_{j=0}^{5} \omega_j E(Y_5 | L = j) = \sum_{j=0}^{5} \omega_j \mu_5^{(j)} \quad (3.13)$$

where ω_j = P(L=j) is also estimated by the proportion of subjects in the pattern L=j.

3.3 Sensitivity Analyses with NFD-Delta Method

We continue the proto-type Table 3.1 for illustrative discussion. With the NFD-Delta method for the NFD-NMAR model, the shift parameters $\Delta = (\Delta_1, \Delta_2, \Delta_3, \Delta_4, \Delta_5)$ can vary over a space of Δ, $D(\Delta)$, to perform sensitivity analysis to examine how the result changes from the MAR case $\Delta = (0, 0, 0, 0, 0)$. In general we need to consider three factors which interplay in making the delta shift: the metric (unit or scale), the magnitude, and the way (algorithm) to apply the delta shift as an adjustment to the imputed values obtained from the MAR model. The delta added to the conditional mean value from MI (under MAR) will be a positive value if higher Y means worse response (such as the blood pressure or the pain score).

3.3.1 Metric of Delta

I will consider the following different metrics (units/scales) for the delta. Let STD_k be the (marginal) standard deviation of Y_k at the k-th visit (pooling treatment groups). STD_k can be obtained from the complete data after MI (assuming MAR). The second choice is RSD_k, which is the residual standard deviation of Y_k conditioning on the previous responses ($Y_0, \ldots, Y_{k-1}$), also obtained after the MI. We can see that STD_k usually increases as the time-point k moves from 1 to T. The rationale for increasing the delta is

that, when a subject drops out, the treatment effect would wear off more and more as time goes by, hence more adjustment is needed for later time-points than the earlier time-points. On the other hand, RSD_k would decrease as the time-point k moves toward T, since more (previous) responses will be conditioned on. The interpretation is that we should "penalize" more the earlier dropouts than the later dropouts. For completers, there is no penalty at all. In addition to these two non-constant (increasing for STD_k and decreasing for RSD_k), another possibility is to set a constant delta unit/scale. We can solicit it from the medical investigator to give a subject-matter expert's unit scale and use it for every time-point. Or, we can simply use STD_1 or RSD_1 throughout the time-points. All the four options: STD_1, RSD_1, STD_k, and RSD_k will be considered in Chapter 4 when I perform simulations for sensitivity analyses.

3.3.2 Magnitude of Delta: Shift Parameter

After the metric is chosen, for example, RSD_k, we may set the maximum range of departure from MAR at visit k equal to within a factor f_k of RSD_k. If prior belief about the value of Δ_k is confined to within $f_k \times RSD_k$, then $D(\Delta) = (f_1 \times RSD_1, \ldots, f_5 \times RSD_5)$. Each missing value at visit k is then replaced with $E(Y_k | Y_0, \ldots, Y_{k-1}, L > k-1) + Unif \times f_k \times RSD_k$, where $E(Y_k | Y_0, \ldots, Y_{k-1}, L > k-1)$ is estimated from the imputation steps under MAR and Unif is a random sample from the uniform (0,1) distribution. Daniel and Hogan (2008, page 241 [9]), for example, used $f_k = 1$ for all k = 1,.., T. Since the metric already varies with time, setting $f_k = f$ independent of time is quite reasonable. This f is called "shift parameter" in this dissertation. In the sensitivity analysis, f will be increased from 0 (i.e., no adjustment of delta) to a value where the result changes drastically. This

will then be the "tipping point" or "break-down point". Sometimes, the sensitivity analysis is also called "stress analysis".

3.3.3 Three Implementation Procedures - Single Adjustment, Sequential and Non-sequential Algorithm

Single Adjustment

There are also different ways to place the delta at each time point. The single adjustment algorithm is the simplest way: Only Unif x f x $RSD_{k\text{-}first}$, or Unif x f x $STD_{k\text{-}first}$ is applied to the time-point where missing value first occurred for the patient. The adjustments for the rest time-points will be taken care of by the correlations through the MI process. Taking an example of $RSD_{k\text{-}first}$ as the metric of delta and assuming the first time-point with missing value is Week 1, single adjustment can be implemented in steps as follows:

i. Impute missing values at time-point 1 assuming MAR, using PROC MI (TRTN is the treatment group variable with 0 representing control group and 1 representing the experimental group):

```
proc mi data=datain out=week1_imput nimpute=5
             seed=1234;
   var TRTN week0 week1;
   class TRTN;
   monotone reg(week1= TRTN week0);
run;
```

As a result of this step, missing data in both control and experimental treatment groups are imputed using a model estimated from subjects with non-missing

values at time-point 1. At this stage, the imputed values are not yet delta-adjusted, which will be done in the next 2 steps.

ii. Use the imputed dataset obtained in step (i), obtain $RSD_{k\text{-}first}$ for the time point 1 using the same regression model specified in MI procedure in step (i). For the example in step (i), obtain RSD_1 (=$RSD_{k\text{-}first}$, if first time point with missing data is Week 1):

```
proc reg data=week1_imput outest=RSD_wk1 covout;
  model week1= TRTN week0;
  by _imputation_;
run;
```

Get the $RSD_{k\text{-}first}$ value from the output dataset RSD_wk1 and save it as variable RSD1.

iii. After step (ii), make the imputed value at time point 1 worse by a value of Unif x f x RSD_1 for those who dropped out at Week 1. Assuming the macro variable &f1 contains the value of f for Week 1 adjustment:

```
data week1_imput;
  merge week1_imput RSD_wk1;
  by _imputation_;
run;

data week1_imput_adjusted;
  set week1_imput;
  f1=&f1.;
  if LASTVIS=0 then week1=week1+f1*ranuni(1234)*rsd1;
run;
```

In this step, for the above example, only the imputed values at Week 1 was adjusted by delta. For subsequent visits (Week 2, Week 3, ...) of the patient who dropped out at Week 1, no delta-adjustment was implemented. The adjustments for these time-points will be taken care by the correlations through the subsequent MI process. Be aware here, the delta-adjustment is applied to both control group and experimental group. In Ratitch et al. (2013), delta-adjustment was only applied to the subjects in the experimental group.

iv. After the delta-adjustment in step (iii), impute all the remaining time-points sequentially by repeating steps (i-iii) for each time-point. Use output dataset from step (iii) as input to the next call to PROC MI. Include only the variable corresponding to the time-point that needs to be imputed in the VAR statement plus predictors (treatment group, baseline, and previous time-points). For example, for time point 2 (Week 2), the following code can be used for MI process (use nimpute=1 for subsequent time points):

```
proc mi data=week1_imput_adjusted out= week2_imput
              nimpute=1 seed=1234;
   var TRTN week0 week1 week2;
   class TRTN;
   monotone reg(week2= TRTN week0 week1);
run;
```

v. At the end of process, the final multiply-imputed dataset where all missing values are filled is ready for analysis by standard SAS procedures with the results combined using PROC MIANALYZE.

If STD is chosen as the metric of delta, in step (iii), use PROC MEANS on the imputed dataset from PROC MI to obtain the STD for that time point and use it when perform the delta-adjustment and then follow the same process specified above.

Non-sequential Delta Adjustment

The non-sequential algorithm is to apply Unif $x \, f \, x$ (STD_1, RSD_1, STD_k, or RSD_k) at each time-point k where missing data occur, only once, after the MI (under MAR) is completed. Taking an example of RSD_k as the metric of delta, it can be implemented in the steps as follows:

i. Impute all the missing values for all the time-points assuming MAR, using PROC MI:

```
proc mi data=datain out=datain_imput nimpute=5
             seed=1234;
   class TRTN;
   var TRTN week0 week1 week2 week3 week4 week5;
   monotone method=REG;
run;
```

As a result of this step, all the missing data in both control and experimental treatment arms are predicted and imputed in turn, given the observed values of all the other variables.

ii. Use the imputed dataset obtained in step (i), obtain RSD_k for each time point using a regression model with the variable corresponding to the time point that

needs to be imputed as the response variable, treatment group and previous time points (including baseline) as the predictors. For example:

```
proc reg data=datain_imput outest=RSD_wk1 covout;
  model week1= TRTN week0;
  by _imputation_;
run;

proc reg data=datain_imput outest=RSD_wk2 covout;
  model week2= TRTN week0 week1;
  by _imputation_;
run;

proc reg data=datain_imput outest=RSD_wk3 covout;
  model week3= TRTN week0 week1 week2;
  by _imputation_;
run;
```

Similarly, derive RSD_k for subsequent visits and save them as variables RSD1, RSD2, ...

iii. After step (ii), merge the imputed dataset from step (i) and the derived RSD_k for each time point from step (ii), then adjust the imputed values for each time point worse by a value of Unif x f x RSD_k. Assuming the macro variables &f1, &f2, &f3, &f4, and &f5 contains the value of f for delta adjustments:

```
data datain_adj;
  set datain;
  by _imputation_;
  f1=&f1.; f2=&f2.; f3=&f3.; f4=&f4.; f5=&f5.;
```

```
  if lastvis=0 then do;
     week1=week1+f1*ranuni(1234)*rsd1;
     week2=week2+f2*ranuni(1234)*rsd2;
     week3=week3+f3*ranuni(1234)*rsd3;
     week4=week4+f4*ranuni(1234)*rsd4;
     week5=week5+f5*ranuni(1234)*rsd5;
  end;

  else if lastvis=1 then do;
     week2=week2+f2*ranuni(1234)*rsd2;
     week3=week3+f3*ranuni(1234)*rsd3;
     week4=week4+f4*ranuni(1234)*rsd4;
     week5=week5+f5*ranuni(1234)*rsd5;
  end;

  else if lastvis=2 then do;
     week3=week3+f3*ranuni(1234)*rsd3;
     week4=week4+f4*ranuni(1234)*rsd4;
     week5=week5+f5*ranuni(1234)*rsd5;
  end;

  else if lastvis=3 then do;
     week4=week4+f4*ranuni(1234)*rsd4;
     week5=week5+f5*ranuni(1234)*rsd5;
  end;

  else if lastvis=4 then do;
     week5=week5+f5*ranuni(1234)*rsd5;
  end;
run;
```

iv. At the end of this process, the final multiply-imputed dataset where all missing values are filled and adjusted is ready for analysis by standard SAS procedures with the results combined using PROC MIANALYZE.

Sequential Delta Adjustment

The sequential algorithm is to apply Unif x *f* x (STD_1, RSD_1, STD_k, or RSD_k) sequentially at each time-point k where missing data occur during the MI process. Taking an example of RSD_k as the metric of delta and assuming the first time-point with missing value is Week 1, sequential delta adjustment can be implemented in steps as follows:

i. Impute missing values at time-point 1 assuming MAR, using PROC MI:

```
proc mi data=datain out=week1_imput nimpute=5
             seed=1234;
   var TRTN week0 week1;
   class TRTN;
   monotone reg(week1= TRTN week0);
run;
```

As a result of this step, missing data in both control and experimental treatment arms are imputed using a model estimated from subjects with non-missing values at time-point 1. At this stage, the imputed values are not yet delta-adjusted, which will be done in the next 2 steps.

ii. Use the imputed dataset obtained in step (i), obtain RSD_k for the time point using the same regression model specified in MI procedure in step (i). For the example in step (i), obtain RSD_1 (RSD_k for Week 1):

```
proc reg data=week1_imput outest=RSD_wk1 covout;
  model week1= TRTN week0;
  by _imputation_;
run;
```

Get the RSD_1 value from the output dataset RSD_wk1 and save it as variable RSD1.

iii. After step (ii), make the imputed value at time point 1 worse by a value of Unif x f x RSD_1. Assuming the macro variable &f1 contains the value of f for Week 1 adjustment:

```
data week1_imput;
  merge week1_imput RSD_wk1;
  by _imputation_;
run;

data week1_imput_adjusted;
  set week1_imput;
  f1=&f1.;
  if LASTVIS=0 then week1=week1+f1*ranuni(1234)*rsd1;
run;
```

In this step, the imputed values at Week 1 were adjusted by delta. Be aware here, the delta-adjustment is applied to both control group and experimental group. In Ratitch et al. (2013), delta-adjustment was only applied to the subjects in the experimental group.

vi. After the delta-adjusting in step (iii), impute all the remaining time-points sequentially by repeating steps (i-iii) for each time-point. Use output dataset from

step (iii) as input to the next call to PROC MI. Include only the variable corresponding to the time-point that needs to be imputed in the VAR statement plus predictors (treatment arm, baseline, and previous time-points). For example, for time point 2 (Week 2), the following code can be used for MI process (use nimpute=1 for subsequent time points):

```
proc mi data=week1_imput_adjusted out= week2_imput
              nimpute=1 seed=1234;
   var TRTN week0 week1 week2;
   class TRTN;
   monotone reg(week2= TRTN week0 week1);
run;
```

Different than the single-adjustment which only adjusts the imputed values of the time-point where missing value first occurred for the patient and the adjustments for the rest time-points are taken care by the correlations through the MI process, sequential delta adjustment algorithm will apply the delta adjustment to all the subsequent time points. For example, for the time point Week 2, delta adjustment of Unif x f x RSD_2 will be applied to the imputed Week 2 values for all patients who dropped out at and before Week 2, which including those dropped out at Week 1 as well. For example, the following SAS code can be used to implement the delta-adjustment after PROC MI for Week 2:

```
data week2_imput_adjusted;
 set week2_imput;
 f2=&f2.;
 if LASTVIS in (0,1) then
    week2=week2+f2*ranuni(1234)*rsd2;
```

```
run;
```

Similarly for the subsequent time points.

iv. At the end of this process, the final multiply-imputed dataset where all missing values are filled is ready for analysis by standard SAS procedures with the results combined using PROC MIANALYZE.

In addition, when the missing values are multiply imputed by the NMAR model, other restrictions may also be considered. For example, for a pain score that ranges from 0 to 10, the imputed value should also be restricted in this range. Suppose that a clinical interpretation of withdrawal due to ineffective therapy means that the missing value should not be better than the baseline value, the imputed value should then also adopt this restriction.

3.4 Methods for ACT Estimand

There is another type of analysis often requested by the regulatory authority and performed by the pharmaceutical company: the control-based analysis. The control-based analysis is to address the ACT (as control treatment) Estimand. Strictly speaking, as discussed in Section 2.7, ACT is a different estimand from the AAT estimand, and addressed a different question from that of the AAT estimand, hence should not really be viewed as a 'sensitivity analysis' per se. Rather, the analysis of the ACT estimand should be regarded as an analysis for a secondary aim. Nevertheless, I will present simulation studies for the ACT estimand analysis.

The ACT Estimand assumes that all discontinued patients would follow the path of the control patients. For a double blind trial, patients may discontinue the therapy or take recue medication (hence invalidate the response) for different reasons, but not for the reason of "treatment group" – since "treatment group" is masked. Hence, this model of MDM is different from the "response" non-future dependence or "response" dependence missingness. It is rather a "covariate" (i.e., treatment group) missing. For the simulation studies, I consider three situations, first: only MAR, second: NFD-NMAR, third: mixture of MAR and NFD-NMAR. In AAT analysis, the AE or LOE are considered to be NMAR, and any other reasons including lost-to-follow-up are considered to be MAR. The basic idea here is that the delta shift penalty used previously in AAT analyses is now replaced by using the control group's patients to impute the missing data for the treatment group's patients. Obviously this is a conservative approach. However, there is no simulation work in the literature, as far as I know, to investigate the degree of conservativeness of this approach. Since, unlike AAT analysis where there is shift parameter to adjust the level of penalty, there is no such a device in this ACT analysis to adjust for the level penalty. Hence it is important to investigate how the differential dropout rate affects the type-I error rate of the ACT analysis – and that is purpose of the simulation for ACT analysis.

In a typical data set with missing data, we include reason of missing data: LOE (lack of efficacy) and AE (adverse event) in the NMAR category and assume NFD missingness, and any other reason ("Other") such as lost-to-follow-up in the MAR category. The idea of ACT analysis can be illustrated by the following diagrams by an example of mixture data of MAR and NMAR, the following steps illustrate the implementation of ACT

analysis. Notice that the imputation for the NMAR category involves time-point by time-point steps:

1. For MAR Category:

   In order to impute MAR missing data for all time-points, first separate the input dataset DATAIN into two datasets: NMAR, containing all subjects with NMAR missing data from both treatment groups; and DATAIN_MAR, containing the rest of the subjects which includes the completed subjects and subjects with any MAR missing data from both treatment groups. Then perform PROC MI on the dataset DATAIN_MAR to impute the MAR missing data using regression based MI for all the time points and save the imputed values obtained from PROC MI for both control and experimental groups.

   <DATAIN>:

SubjID	Treatment	Y_0	Y_1	Y_2	Y_3	Y_4	Y_5	Reason
#001	PBO	X	X	X	X	X	x	
#002	PBO	X	X					Other
#003	PBO	X	X	X	X			LOE
#004	PBO	X						AE
#005	DRG	X	X	X	X	X	x	
#006	DRG	X	X					Other
#007	DRG	X	X					AE
#008	DRG	X						AE
...								

<DATAIN_MAR>: (Block out the NMAR cases in the grey area that are AE and LOE dropouts)

SubjID	Treatment	Y_0	Y_1	Y_2	Y_3	Y_4	Y_5	Reason
#001	PBO	X	X	X	X	X	x	
#002	PBO	X	X	O	O	O	O	Other
#003	PBO	X	X	X	X			LOE
#004	PBO	X						AE
#005	DRG	X	X	X	X	X	x	
#006	DRG	X	X	O	O	O	O	Other
#007	DRG	X	X					AE
#008	DRG	X						AE
...								

The following SAS code can be used:

```
data datain_mar nmar;;
  set datain;
  if label='NMAR' then output nmar;
  else output datain_mar;
run;

proc mi data=datain_mar out=mar_out nimpute=1 seed=1234;
  class TRTN;
   monotone method=REG;
  var TRTN week0 week1 week2 week3 week4 week5;
run;
```

<MAR_OUT>:

SubjID	Treatment	Y_0	Y_1	Y_2	Y_3	Y_4	Y_5	Reason
#001	PBO	X	X	X	X	X	x	
#002	PBO	X	X	Y_{22}	Y_{23}	Y_{24}	Y_{25}	Other
#003	PBO	X	X	X	X			LOE
#004	PBO	X						AE
#005	DRG	X	X	X	X	X	x	
#006	DRG	X	X	Y_{62}	Y_{63}	Y_{64}	Y_{65}	Other
#007	DRG	X	X					AE
#008	DRG	X						AE
…								

```
data mar_missing;
   set mar_out;
   if lastvis ne 5;
run;
```

<MAR_MISSING>:

SubjID	Treatment	Y_0	Y_1	Y_2	Y_3	Y_4	Y_5	Reason
#001	PBO	X	X	X	X	X	x	
#002	PBO	X	X	Y_{22}	Y_{23}	Y_{24}	Y_{25}	Other
#003	PBO	X	X	X	X			LOE
#004	PBO	X						AE
#005	DRG	X	X	X	X	X	x	
#006	DRG	X	X	Y_{62}	Y_{63}	Y_{64}	Y_{65}	Other
#007	DRG	X	X					AE
#008	DRG	X						AE
…								

2. For NMAR for both control and experimental treatment groups:

   Similar to the delta-adjustment methods, we break the imputation process into a sequence of MIs, where each MI is intended to impute missing values at one time-point only, as follows.

   i. In order to impute NMAR missing data at time-point 1 (assume Week 1 in this example), first prepare an input dataset, making sure that it will contain only the intended donor and recipient patterns.

```
data datain_nmar;
  set datain;
 if label='MAR' then output mar;
 else output datain_nmar;
run;
```

<DATAIN>

<DATAIN_NMAR>: (Block out the MAR cases in the grey area that are "Other" dropouts)

SubjID	Treatment	Y_0	Y_1	Y_2	Y_3	Y_4	Y_5	Reason
#001	PBO	X	X	X	X	X	x	
#002	PBO	X	X					Other
#003	PBO	X	X	X	X			LOE
#004	PBO	X						AE
#005	DRG	X	X	X	X	X	x	

#006	DRG	X	X					Other
#007	DRG	X	X					AE
#008	DRG	X						AE
...								

Separate the input dataset DATAIN_NMAR into two datasets: DATAIN_IMP1, containing all control subjects and those subjects from the experimental group that have values at time-point 1 missing (LASTVIS=0); and DATAIN_REST1, containing the rest of the subjects from the experimental group with non-missing value at time-point 1 (e.g., Week 1).

```
data datain_imp1 datain_rest1;
  set datain_nmar;
  if trtabn=1 and lastvis>0 then output datain_rest1;
  else output datain_imp1;
run;
```

<DATAIN_NMAR>

<DATAIN_IMP1>:

SubjID	Treatment	Y_0	Y_1	Y_2	Y_3	Y_4	Y_5	Reason
#001	PBO	X	X	X	X	X	x	
#002	PBO	X	X					Other
#003	PBO	X	X	X	X			LOE
#004	PBO	X						AE
#005	DRG	X	X	X	X	X	x	
#006	DRG	X	X					Other

#007	DRG	X	X					AE
#008	DRG	X						AE
…								

ii. Call PROC MI to impute missing data at time-point 1 (Week 1) using dataset DATAIN_IMP1 as input.

```
proc mi data=datain_imp1 out=datain_reg_imp1 nimpute=1
        seed=1234;
    var week0 week1;
    monotone reg(week1 = week0);
run;
```

<DATAIN_REG_IMP1>:

SubjID	Treatment	Y_0	Y_1	Y_2	Y_3	Y_4	Y_5	Reason
#001	PBO	X	X	X	X	X	x	
#002	PBO	X	X					Other
#003	PBO	X	X	X	X			LOE
#004	PBO	X	Y_{41}					AE
#005	DRG	X	X	X	X	X	x	
#006	DRG	X	X					Other
#007	DRG	X	X					AE
#008	DRG	X	Y_{81}					AE
…								

Note that treatment group is not included as an effect in the model (it is not included in the VAR statement). Since subjects from the experimental arm with non-missing values at time-point t are not included in the input dataset, they will not contribute to the estimation of an imputation model for time-point t. The imputation model will be estimated using control subjects only, while this call to PROC MI will impute missing data at time-point t for all subjects who need imputation at that time-point. This way, subjects from experimental arm will be imputed based on the control subjects' model.

iii. Assemble back a dataset containing all subjects.

```
data datain1;
   set datain_reg_imp1 datain_rest1;
run;
```

iv. Repeat steps (i)-(iii) for all other time-points sequentially using a reconstructed dataset (e.g., DATAIN1) as a starting point in step (i) to get the final imputed NMAR dataset DATAIN5 which includes the subjects with NMAR missing data and the subjects with no any missing data (completers) from both treatment groups.

<DATAIN_IMP2>:

SubjID	Treatment	Y_0	Y_1	Y_2	Y_3	Y_4	Y_5	Reason
#001	PBO	X	X	X	X	X	x	
#002	PBO	X	X					Other

#003	PBO	X	X	X	X			LOE
#004	PBO	X	Y_{41}	○				AE
#005	DRG	X	X	X	X	X	x	
#006	DRG	X	X					Other
#007	DRG	X	X	○				AE
#008	DRG	X	Y_{81}	○				AE
...								

(Note that case #007 is put back)

…

…

<DATAIN_REG_IMP5>:

SubjID	Treatment	Y_0	Y_1	Y_2	Y_3	Y_4	Y_5	Reason
#001	PBO	X	X	X	X	X	x	
#002	PBO	X	X					Other
#003	PBO	X	X	X	X	Y_{34}	Y_{35}	LOE
#004	PBO	X	Y_{41}	Y_{42}	Y_{43}	Y_{44}	Y_{45}	AE
#005	DRG	X	X	X	X	X	x	
#006	DRG	X	X					Other
#007	DRG	X	X	Y_{72}	Y_{73}	Y_{74}	Y_{75}	AE
#008	DRG	X	Y_{81}	Y_{82}	Y_{83}	Y_{84}	Y_{85}	AE
...								

3. Final complete imputed data with MAR and NMAR:

Assemble back a dataset containing all imputed MAR and NMAR data obtained from step (1) and (2):

```
data all;
  set mar_missing datain5;
run;
```

<ALL>:

SubjID	Treatment	Y_0	Y_1	Y_2	Y_3	Y_4	Y_5	Reason
#001	PBO	X	X	X	X	X	x	
#002	PBO	X	X	Y_{22}	Y_{23}	Y_{24}	Y_{25}	Other
#003	PBO	X	X	X	X	Y_{34}	Y_{35}	LOE
#004	PBO	X	Y_{41}	Y_{42}	Y_{43}	Y_{44}	Y_{45}	AE
#005	DRG	X	X	X	X	X	x	
#006	DRG	X	X	Y_{62}	Y_{63}	Y_{64}	Y_{65}	Other
#007	DRG	X	X	Y_{72}	Y_{73}	Y_{74}	Y_{75}	AE
#008	DRG	X	Y_{81}	Y_{82}	Y_{83}	Y_{84}	Y_{85}	AE
...								

4. At the end of this process, the final control-based imputed dataset where all missing values are filled is ready for analysis by standard SAS procedures with the results combined using PROC MIANALYZE.

Chapter 4

Simulations

In this Chapter, I will conduct simulation studies to evaluate the different methods. The criteria of the evaluation will be based on the alpha level (i.e., type-I error rate) in two aspects: correctness and robustness. A good method should not only hold the alpha at the nominal level when the MDM assumption is correct (i.e., correctness), but should also maintain the alpha close to the nominal level even when the missing data mechanism (MDM) assumption is not exactly correct (i.e., robustness).

4.1 Simulation Models, Configurations, and Dropout Patterns

I will consider various scenarios of the MDM and perform 5,000 simulation runs for each scenario. Given the nominal level of alpha= 0.025 (one-sided test), the width of the 95% C.I. of the alpha level based on 5,000 simulation runs will be 1.96*sqrt(0.025*0.975/5000) = 0.00433.

For each simulation run, the complete data is generated from multivariate normal distributions of sample size N=200 for each treatment group A (Control) and group B, with mean vector μ_A and μ_B respectively, and a common variance-covariance matrix, Σ. The numerical configurations mimic the real Pain Study data example, which has baseline y_0 and five weekly follow-up visits y_1 to y_5.

Example Data

The example pain study is a phase II, double-blind, randomized, placebo-controlled clinical trial. The primary efficacy parameter is the pain total score ranging from 0 to 10, with lower values corresponding to more favorable outcomes. Study visits are scheduled at baseline (pre-dose) and post-baseline at weeks 1, 2, 3, 4, and 5. The primary efficacy endpoint is the change from baseline to week 5 in the pain total score. The following table presents the percentages of subjects discontinued and completed by treatment group:

Disposition	Control Group (N=213)	Active Group (N=220)
Subjects Completed	189 (88.7%)	189 (85.9%)
Subjects Discontinued	24 (11.3%)	31 (14.1%)
Adverse Event	8 (3.8%)	21 (9.6%)
Lack of Efficacy	1 (0.5%)	1 (0.5%)
Other	15 (7.0%)	9 (4.1%)

There were more dropouts in the active group compared with control group (14.1% vs. 11.3%). In the control group, 3.8% out of 11.3% of subjects discontinued due to AEs, compared with 9.6% out of 14.1% in the active group.

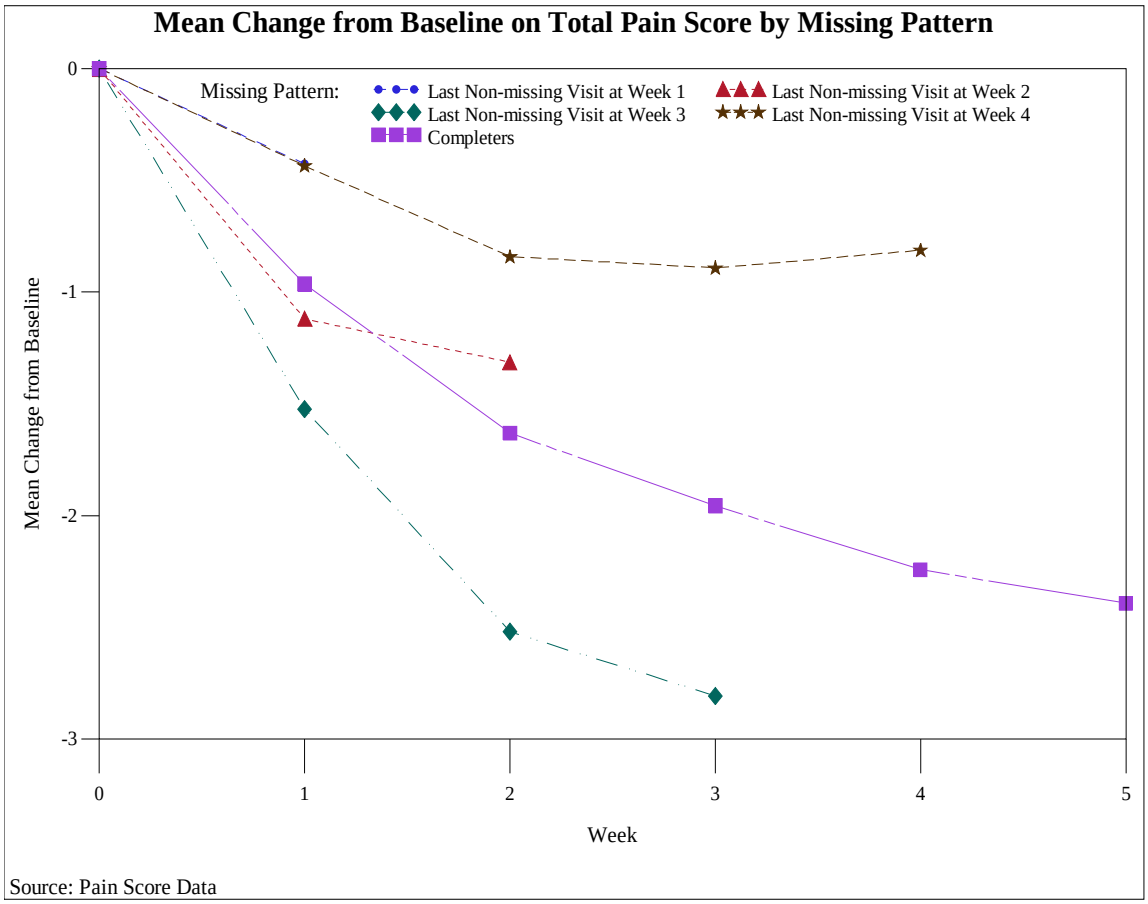
The pattern of the missing data by treatment group is presented as follows:

Treatment Group	Pattern	Week 0	Week 1	Week 2	Week 3	Week 4	Week 5	Freq	Percent
Control Group	1	x	x	x	x	x	x	189	88.73
	2	x	x	x	x	x	o	7	3.29
	3	x	x	x	x	o	o	4	1.88
	4	x	x	x	o	o	o	9	4.23
	5	x	x	o	o	o	o	4	1.88
Active Group	1	x	x	x	x	x	x	189	85.91
	2	x	x	x	x	x	o	3	1.36
	3	x	x	x	x	o	o	4	1.82

Treatment Group	Pattern	Week 0	Week 1	Week 2	Week 3	Week 4	Week 5	Freq	Percent
	4	x	x	x	o	o	o	15	6.82
	5	x	x	o	o	o	o	9	4.09

“x”denotes observed; “o” denotes missing;
"." denotes missing but to be imputed to achieve the monotone missingness.

All subjects have non-missing values at baseline (Week 0) and Week 1. The following figure depicts the missing data pattern:



Simulation Models, Configurations, and Dropout Patterns

The simulation models, configurations and dropout patterns will mimic the Pain Study data example.

Under the null hypothesis, i.e., treatment groups have the same mean pain scores, let

$$\mu_A = \mu_B = (6.88, 6.15, 5.52, 5.25, 5.07, 4.92) \quad (4.1)$$

Or, expressed in term of the *mean changes from baseline* ($\Delta\mu$) for both treatment group:

$$\Delta\mu_A = \Delta\mu_B = (-0.73, -1.36, -1.63, -1.81, -1.96) \quad (4.2)$$

The common variance-covariance (correlation) matrix, Σ, for (Y_0, Y_1, Y_2, Y_3, Y_4, Y_5) is:

2.27	1.85	1.70	1.70	1.69	1.60
(1.00)	(0.66)	(0.52)	(0.48)	(0.46)	(0.42)
1.85	3.50	3.45	3.45	3.28	3.13
(0.66)	(1.00)	(0.85)	(0.78)	(0.71)	(0.65)
1.70	3.45	4.70	4.48	4.34	4.13
(0.52)	(0.85)	(1.00)	(0.89)	(0.83)	(0.76)
1.70	3.45	4.48	5.48	5.17	5.14
(0.48)	(0.78)	(0.89)	(1.00)	(0.91)	(0.88)
1.69	3.28	4.34	5.17	5.87	5.65
(0.46)	(0.71)	(0.83)	(0.91)	(1.00)	(0.92)
1.60	3.13	4.13	5.14	5.64	6.29
(0.42)	(0.65)	(0.76)	(0.88)	(0.92)	(1.00)

The incomplete data is generated with different MDM scenarios using selection models. The baseline y_0 is always observed.

For MAR, the model of missing data is based on the following logistic model:

$$Pr(m_{ij} = 1 | y_{i0}, y_{i1}, y_{i2}, y_{i3}, y_{i4}, y_{i5}) = \exp(a+b*y_{i,j-1})/[1+ \exp(a+b*y_{i,j-1})] \quad (4.3)$$

j =1 to 5. That is, the probability of the (current) y_{ij} being missing (m_{ij}=1) or observed (m_{ij}=0) depends on the value of the immediate visit $y_{i,j-1}$. We then make a monotone

pattern: $m_{ij}=1$ implying $m_{ij'}=1$ for all j'>j. Eq.(4.3) can be written as: $Logit[L = j - 1|L \geq j-1,\ Y_j^-, y_j, y_{j+1}, \dots, y_T] = a + b_{j-1}$.

The values of a and b are chosen to consider four different proportions of dropouts under MAR: A and B groups with the same a and b; with different a but same b; with same a but different b; and with different a and b. Table 4.1 displays the average dropout proportions over 5000 simulation runs. Note that, when a and b are the same, there is no differential dropout rates between treatment groups. The dropout rate is more sensitive to the intercept a than to the slope b. The sample mean differences between A and B groups are also shown in Table 4.1 for each of the scenarios of dropout patterns. These differences are not zero (null hypothesis setting) as a consequence of the differential dropouts in A and B. Positive b implies that higher chance to drop out for higher Y_{j-1} values. Within each treatment group, when high Y_{j-1} tends to dropout and low value tends to stay, the change from baseline (high value) would tend to be more negative. In the simulations I let the multivariate normal play out its full range without restricting the pain score within 0 and 10.

Table 4.1 Cumulative dropout rates at each time point (%) by treatment group, resulting mean change from baseline, and difference ($\Delta\mu_A$ - $\Delta\mu_B$) departure from the null hypothesis – MAR

Treatment Group	a	b	Week1	Week2	Week3	Week4	Week5
A	-6.91	0.58	6.9	11.7	15.4	18.5	21.1
$\Delta\mu_A$			-0.71	-1.38	-1.71	-1.94	-2.15
B	-6.91	0.58	6.9	11.8	15.4	18.5	21.1
$\Delta\mu_B$			-0.71	-1.38	-1.70	-1.94	-2.14
$\Delta\mu_B$ - $\Delta\mu_A$			0.00	0.00	0.01	0.01	0.01
A	-6.21	0.58	12.4	20.6	26.1	30.4	34.0
$\Delta\mu_A$			-0.70	-1.40	-1.77	-2.03	-2.26
B	-5.81	0.58	17.0	27.2	33.8	39.0	43.0

$\Delta\mu_B$			-0.69	-1.41	-1.80	-2.09	-2.33
$\Delta\mu_B - \Delta\mu_A$			0.01	-0.01	-0.03	-0.06	-0.08
A	-6.91	0.58	6.9	11.7	15.4	18.5	21.1
$\Delta\mu_A$			-0.71	-1.38	-1.71	-1.94	-2.15
B	-6.91	0.68	13.1	21.3	26.6	30.7	34.1
$\Delta\mu_B$			-0.70	-1.40	-1.77	-2.04	-2.26
$\Delta\mu_B - \Delta\mu_A$			0.02	-0.02	-0.06	-0.09	-0.11
A	-6.21	0.58	12.4	20.6	26.1	30.4	34.0
$\Delta\mu_A$			-0.70	-1.40	-1.77	-2.03	-2.26
B	-5.81	0.68	28.1	41.9	49.4	54.5	58.4
$\Delta\mu_B$			-0.66	-1.44	-1.89	-2.22	-2.49
$\Delta\mu_B - \Delta\mu_A$			0.04	-0.03	-0.13	-0.19	-0.23

For NFD-NMAR, the model of missing data is based on the following logistic model:

$$\Pr(m_{ij} = 1 | y_{i0}, y_{i1}, y_{i2}, y_{i3}, y_{i4}, y_{i5}) = \exp(a+b*y_{ij})/[1+ \exp(a+b*y_{ij})] \quad (4.4)$$

j=1 to 5. That is, the probability of the (current) y_{ij} being missing (m_{ij}=1) or observed (m_{ij}=0) depends on the value of the current visit y_{ij} itself. With a monotone pattern: m_{ij}=1 implying $m_{ij'}$=1 for all j'>j. Eq.(4.4) can be written as: $Logit[L = j - 1 | L \geq j - 1, Y_j^-, y_j, y_{j+1}, \dots, y_T] = a + b_j$.

The values of a and b are chosen to consider four different proportions of dropouts under NFD-NMAR: A and B groups with the same a and b; with different a but same b; with same a but different b; and with different a and b. Table 4.2 displays the average dropout proportions over 5,000 simulation runs. Note that when a and b are the same there is no differential dropout rate between treatment groups. The dropout rate is more sensitive to the intercept a than to the slope b. The sample mean differences between A and B groups are also shown in Table 4.2 for each of the scenarios of dropout patterns. These differences are not zero (null hypothesis setting) as a consequence of the differential

dropout rates in A and B. (Positive b implies that higher chance to drop out for higher Y_j values.)

Table 4.2 Cumulative dropout rates at each time point (%) by treatment group, resulting mean change from baseline, and difference ($\Delta\mu_A$ - $\Delta\mu_B$) departure from the null hypothesis – NFD-NMAR

Treatment Group	a	b	Week1	Week2	Week3	Week4	Week5
A	-6.91	0.68	10.3	16.8	21.8	25.8	29.1
$\Delta\mu_A$			-0.83	-1.57	-1.95	-2.21	-2.44
B	-6.91	0.68	10.3	16.8	21.8	25.8	29.1
$\Delta\mu_B$			-0.83	-1.56	-1.94	-2.20	-2.43
$\Delta\mu_B$ - $\Delta\mu_A$			0.00	0.01	0.00	0.00	0.00
A	-6.21	0.68	17.0	26.6	33.2	38.1	42.1
$\Delta\mu_A$			-0.88	-1.69	-2.11	-2.39	-2.65
B	-5.81	0.68	21.9	33.4	40.8	46.2	50.4
$\Delta\mu_B$			-0.93	-1.76	-2.21	-2.51	-2.78
$\Delta\mu_B$ - $\Delta\mu_A$			-0.04	-0.08	-0.10	-0.12	-0.13
A	-6.91	0.58	5.4	9.3	12.7	15.6	18.1
$\Delta\mu_A$			-0.78	-1.47	-1.80	-2.03	-2.25
B	-6.91	0.68	10.3	16.9	21.8	25.7	29.0
$\Delta\mu_B$			-0.83	-1.57	-1.95	-2.21	-2.43
$\Delta\mu_B$ - $\Delta\mu_A$			-0.05	-0.11	-0.15	-0.17	-0.19
A	-6.21	0.58	9.9	16.5	21.7	25.9	29.4
$\Delta\mu_A$			-0.81	-1.55	-1.92	-2.18	-2.41
B	-5.81	0.68	22.0	33.2	40.7	46.1	50.3
$\Delta\mu_B$			-0.93	-1.76	-2.22	-2.51	-2.78
$\Delta\mu_B$ - $\Delta\mu_A$			-0.12	-0.22	-0.30	-0.34	-0.37

We also consider mixture of MAR and NFD-NMAR in a data set by using both models (4.3) and (4.4) as shown in Table 4.3. I consider three scenarios: the first one where the MAR and NFD-NMAR proportions are the same for both treatment groups (both groups have more NMAR than MAR); the second scenario where NMAR models have different slopes but both groups has more NMAR than MAR; and the third scenario where NMAR models have different intercepts and slopes (so that group A has more NMAR and less

MAR cases while group B has the reverse). The sample mean differences between A and B groups are also shown in Table 4.3 for each of the scenarios of dropout patterns. These between-group differences in changes from baseline are not zero (null hypothesis setting) for the second and third scenarios as a consequence of the differential dropout rates in A and B.

Table 4.3 Cumulative dropout rates at each time point (%) by treatment group, resulting mean change from baseline, and difference ($\Delta\mu_A$ - $\Delta\mu_B$) deviation from the null hypothesis – Mixture of MAR and NFD-NMAR

Treatment Group		a	b	Week1	Week2	Week3	Week4	Week5
A	MAR	-7.25	0.50	2.8	5.0	6.7	8.3	9.8
	NFD-NMAR	-6.11	0.57	8.2	13.9	18.4	22.1	25.3
$\Delta\mu_A$				-0.79	-1.51	-1.89	-2.18	-2.45
B	MAR	-7.25	0.50	2.8	5.0	6.7	8.3	9.8
	NFD-NMAR	-6.11	0.57	8.3	13.9	18.4	22.2	25.3
$\Delta\mu_B$				-0.79	-1.51	-1.89	-2.17	-2.44
$\Delta\mu_B$ - $\Delta\mu_A$				0.00	0.01	0.01	0.01	0.01
A	MAR	-7.25	0.50	2.8	5.0	6.7	8.3	9.8
	NFD-NMAR	-6.91	0.58	4.4	7.6	10.4	12.8	15.0
$\Delta\mu_A$				-0.76	-1.45	-1.80	-2.05	-2.29
B	MAR	-7.25	0.50	2.8	5.0	6.7	8.3	9.8
	NFD-NMAR	-6.91	0.68	8.4	13.8	17.9	21.4	24.2
$\Delta\mu_B$				-0.80	-1.52	-1.91	-2.20	-2.46
$\Delta\mu_B$ - $\Delta\mu_A$				-0.03	-0.07	-0.11	-0.14	-0.17
A	MAR	-7.25	0.50	2.8	5.0	6.7	8.3	9.8
	NFD-NMAR	-7.91	0.58	1.7	3.1	4.3	5.4	6.4
$\Delta\mu_A$				-0.74	-1.40	-1.72	-1.95	-2.15
B	MAR	-7.25	0.50	2.8	5.0	6.7	8.3	9.8
	NFD-NMAR	-6.91	0.68	8.4	13.8	17.9	21.4	24.2
$\Delta\mu_B$				-0.80	-1.52	-1.91	-2.20	-2.46
$\Delta\mu_B$ - $\Delta\mu_A$				-0.05	-0.12	-0.19	-0.25	-0.31

As introduced in Chapter 1, the foremost step in analyzing an incomplete data from clinical trials in to consider the causal estimand. I consider the AAT Estimand in Section 4.2, then the ACT Estimand in Section 4.3. The AAT estimand here is the mean difference of the changes from baseline at the end of the final visit (Week 5) between treatment groups A and B, assuming patients who discontinued the therapy would follow the path as the originally assigned treatment. The ACT estimand, on the other hand, assumes that all patients who discontinued the therapy (NMAR) would follow the path of the control group.

4.2 Consideration of Methods for the AAT Estimand: MMRM, MI and NFD-Delta Methods

For the AAT Estimand, each incomplete data set is analyzed under the null structure (4.2) by the mixed effect model for repeated measures (MMRM), the multiple imputation (MI) method, and a NFD-NMAR mean-shift method (NFD-Delta) to test the null hypothesis at the last time-point:

$$H_0: \quad \Delta\mu_A \text{ (Week 5)} - \Delta\mu_B \text{ (Week 5)} = 0 \tag{4.5}$$

$$\text{versus } H_1: \quad \Delta\mu_A \text{ (Week 5)} - \Delta\mu_B \text{ (Week 5)} > 0 \tag{4.6}$$

The objective of the simulations is to examine the respective alpha level for each method under different scenarios. I will set the one-sided test at the nominal level of 0.025.

The MMRM and MI methods need no extra explanation. For the NFD-Delta method, there are three basic considerations:

First, choose a metric/unit for the delta-shift, which can be, from simple (STD_1, RSD_1) to more involved ones (STD_j, RSD_j). The (STD_1, RSD_1) metric is constant over time points, while (STD_j, RSD_j) changes over time points.

Second, choose the shift-parameter *f* to determine the amount of the delta-shift as in the following expression (see Chapter 3):

$$\Delta_j = f_j \times STD_j \text{ (or } RSD_j, \text{ or } STD_1, \text{ or } RSD_1) \times Unif \qquad (4.7)$$

where f_j = f for all time-point j = 1, .., 5 and Unif is a random variable from the standard uniform distribution U(0,1).

Third, consider the non-sequential, sequential, or the single-adjustment method to place the delta-shift, as discussed in Chapter 3.

4.3 Simulation Results

4.3.1 Correctness and Robustness of the MMRM and MI

I start with MMRM and MI because these likelihood based methods are known to be appropriate, i.e., giving the correct alpha levels, when the MDM is MAR. I first verify this fact in Table 4.4.1 to show correctness of these methods. As shown in Table 4.4.1, based on 5,000 simulations runs, MMRM is slightly more conservative than MI when the dropout proportions are unequal between treatment groups. MI, as a likelihood-based method, provides unbiased estimation of $\Delta\mu_B - \Delta\mu_A$ at Week5.

Table 4.4.1 Correctness of the Alpha Level for MMRM and MI When MDM is MAR (nominal alpha=0.025 one-sided test)

Treatment Group	a	b	Week5 Dropout	MMRM Type=UN	MI nimpute=5

			Rates		
A	-6.91	0.58	21.1		
B	-6.91	0.58	21.1		
$\Delta\mu_B - \Delta\mu_A$ (Week 5)	0.01			0.023	0.022
A	-6.21	0.58	34.0		
B	-5.81	0.58	43.0		
$\Delta\mu_B - \Delta\mu_A$ (Week 5)	-0.08			0.023	0.024
A	-6.91	0.58	21.1		
B	-6.91	0.68	34.1		
$\Delta\mu_B - \Delta\mu_A$ (Week 5)	-0.11			0.018	0.021
A	-6.21	0.58	34.0		
B	-5.81	0.68	58.4		
$\Delta\mu_B - \Delta\mu_A$ (Week 5)	-0.23			0.017	0.022
	Estimate of $\Delta\mu_B - \Delta\mu_A$ at Week 5: 2.5-97.5 percentile range over simulation runs:			0.042 (-0.052,0.060)	0.003; (-0.573, 0.575)

Next in Table 4.4.2, I show that MMRM and MI are not necessarily robust when the underlying MDM is actually NFD-NMAR. (This is what I suspected, see Chapter 2.) Hence we need to find a right method to correct the alpha level under NFD-NMAR, see next sub-section.

Table 4.4.2 Robustness of the Alpha Level for MMRM and MI When MDM is NFD-NMAR (nominal alpha=0.025 one-sided test)

Treatment Group	a	b	Week5 Dropout Rates	MMRM Type=UN	MI nimpute=5
A	-6.91	0.68	29.1		
B	-6.91	0.68	29.1		
$\Delta\mu_B - \Delta\mu_A$ (Week 5)	0.00			0.024	0.024
A	-6.21	0.68	42.0		
B	-5.81	0.68	50.4		

$\Delta\mu_A - \Delta\mu_B$ (Week 5)	-0.13			0.046	0.051
A	-6.91	0.58	18.1		
B	-6.91	0.68	29.0		
$\Delta\mu_B - \Delta\mu_A$ (Week 5)	-0.19			0.060	0.062
A	-6.21	0.58	29.4		
B	-5.81	0.68	50.3		
$\Delta\mu_B - \Delta\mu_A$ (Week 5)	-0.37			0.136	0.149
	Estimate of $\Delta\mu_B - \Delta\mu_A$ at Week 5: 2.5-97.5 percentile range over simulation runs:			-0.231 (-0.738,0.252)	-0.265 (-0.779, 0.224)

Table 4.4.2 shows:

- Inflated type-I error rate when use MMRM or MI under NFD-NMAR in the situation that dropout rates are different between treatment groups;

- The degree of inflation of type-I error rate is more dependent on the degree of the imbalance of the dropout rates between treatment groups than the dropout rate itself in each treatment group. When the dropout rates are the same between treatment groups, no type-I error inflation occurs.

- The point estimate of MI $\Delta\mu_B - \Delta\mu_A$ at Week 5 is biased.

Next, we consider the situation of mixture of MAR and NFD-NMAR in Table 4.4.3 for MI and MMAR.

Table 4.4.3 Robustness of the Alpha Level for MMRM and MI When MDM is Mixture of MAR and NFD-NMAR

Treatment Group		a	b	Week 5 Dropout Rates	MMRM Type=UN	MI nimpute=5
A	MAR	-7.25	0.50	9.8		
	NFD-NMAR	-6.11	0.57	25.3		
$\Delta\mu_A$ (Week 5)	-2.45					
B	MAR	-7.25	0.50	9.8		
	NFD-NMAR	-6.11	0.57	25.3		
$\Delta\mu_B$ (Week 5)	-2.44					
$\Delta\mu_B - \Delta\mu_A$ (Week 5)	0.01				0.024	0.024
A	MAR	-7.25	0.50	9.8		
	NFD-NMAR	-6.91	0.58	15.0		
$\Delta\mu_A$ (Week 5)	-2.29					
B	MAR	-7.25	0.50	9.8		
	NFD-NMAR	-6.91	0.68	24.2		
$\Delta\mu_B$ (Week 5)	-2.46					
$\Delta\mu_B - \Delta\mu_A$ (Week 5)	-0.17				0.046	0.047
A	MAR	-7.25	0.50	9.8		
	NFD-NMAR	-7.91	0.58	6.4		
$\Delta\mu_A$ (Week 5)	-2.15					
B	MAR	-7.25	0.50	9.8		
	NFD-NMAR	-6.91	0.68	24.2		
$\Delta\mu_B$ (Week 5)	-2.46					
$\Delta\mu_B - \Delta\mu_A$ (Week 5)	-0.31				0.068	0.074
	Estimate of $\Delta\mu_B - \Delta\mu_A$ at Week 5: 2.5-97.5 percentile range over simulation runs:				-0.123 (-0.593,0.347)	-0.142 (-0.608,0.325)

Table 4.4.3 show that, with mixed MAR and NFD-NMAR,

- No inflation on type-I error rate occurs when drop out distributions are the same between treatment groups (same MAR and same NFD-NMAR proportion);

- Type-I error rate is inflated when the NFD-NMAR dropout rate is imbalanced between treatment groups. Here, we assume, as more likely the case in practice, MAR dropouts rates are similar between treatment groups. The greater imbalance of the NFD-NMAR dropout rates, the higher inflation of the type-I error rate.

4.3.2 Correctness and Robustness of the Delta Mean-Shift (NFD-Delta) Method

I continue to investigate the NFD-Delta method (as described in Chapter 3). The goal is to find an appropriate shift parameter so that the inflated alpha level under the NFD-NMAR by MI and MMRM (seen in Table 4.4.2) may be adjusted at least to some extent close to the nominal level by the NFD-Delta method. I anticipated that the shift parameter will depend on the unit (STD_1 RSD_1, STD_k, or RSD_k) and the algorithm (sequential, non-sequential, or single-adjustment). However, we also need to pay attention how robust, when applying the delta mean-shift, the NFD-Delta method is under the MAR situation, since in practice, we do not really know the true MDM.

4.3.2.1 Delta Mean-shift (NFD-Delta) Method with Unit of STD_1

STD_1 is the pooled standard deviation (over treatment groups) of Y_1 derived from the complete data after the MI (under MAR). The algorithms using STD_1 as the (constant) scale of the delta shift for all the time-points with missing data are either sequential or

non-sequential. Since STD_1 is applied to all time-points, the "single adjustment" method is not applicable here.

Correctness

Simulations shown in Table 4.5.1 are to demonstrate the correctness of the alpha levels for the mean-shift (NFD-Delta) method under the null hypothesis and when MDM is NFD-NMAR using the STD_1 (i.e., pooled STD at Week 1 of the original Y value) as the constant unit over the time-points for the delta adjustment (see expression (4.6)). Table 4.5.1 also shows the sensitivity analyses by varying the shift parameter f with different algorithm (non-sequential and sequential).

Table 4.5.1 Correctness of the Alpha Level for Mean-shift Method (NFD-Delta) When MDM is NFD-NMAR; Delta unit = STD_1

Treatment Group	a	b	Week 5 Dropout Rates	f=0.25 Non-seq/ Sequential	f=0.5 Non-seq/ Sequential	f=1.0 Non-seq/ Sequential	f=1.5 Non-seq/ Sequential
A	-6.91	0.68	29.1				
B	-6.91	0.68	29.1				
$\Delta\mu_B - \Delta\mu_A$ (Week 5)	0.00 (STD_1=1.79)			0.024/ 0.022	0.023/ 0.021	0.023/ NN	0.021/ NN
	Estimate of $\Delta\mu_B - \Delta\mu_A$ at Week 5: 2.5-97.5 percentile range over simulation runs:			0.002/ 0.000 (-0.49,0.47)/ (-0.50,0.49)	0.002/ 0.001 (-0.50,0.48)/ (-0.52,0.51)	0.002/ 0.003 (-0.51,0.49)/ (-0.59,0.59)	0.002/ NN (-0.53,0.50)/ NN
A	-6.21	0.68	42.0				
B	-5.81	0.68	50.4				
$\Delta\mu_B - \Delta\mu_A$ (Week 5)	-0.13 (STD_1=1.75)			0.045/ 0.023	0.037/ 0.012	0.027/ 0.003	0.016/ NN
A	-6.91	0.58	18.1				
B	-6.91	0.68	29.0				
$\Delta\mu_B - \Delta\mu_A$ (Week5)	-0.19 (STD_1=1.81)			0.051/ 0.029	0.040/ 0.013	0.024/ 0.003	0.014/ NN
A	-6.21	0.58	29.4				

B	-5.81	0.68	50.3				
$\Delta\mu_B - \Delta\mu_A$ (Week 5)	-0.37 (STD_1=1.78)			0.108/ 0.037	0.084/ 0.008	0.041/ 0.001	0.017/ 0.000*
	Estimate of $\Delta\mu_B - \Delta\mu_A$ at Week 5: 2.5-97.5 percentile range over simulation runs:			-0.219/ -0.089 (-0.73,0.27)/ (-0.60,0.42)	-0.172/ 0.087 (-0.70,0.33)/ (-0.45,0.63)	-0.08/ 0.441 (-0.62,0.44)/ (-0.18,1.06)	0.016/ 0.795 (-0.54, 0.55)/ (0.068, 1.52)

*NN: "Not Needed", since the value is already known to be less than 0.025.

Table 4.5.1 shows that, with STD_1 as the metric for delta:

- If dropout rates are the same between treatment groups, the alpha level holds at nominal level even under NFD-NMAR, no adjustment is needed.

- If dropouts rates are different between treatment groups:

    - The greater imbalance of dropout rates, the higher inflation of the type-I error rate by MI is, thus the larger *f* is needed for adjustment;

    - The sequential method corrects the type-I error inflation rapidly and could easily over-correct and leads to losing power. So, when use the sequential method, we should start with a small *f* and gradually to make the correction. The over-correction of the sequential algorithm is demonstrated in the last case with f =1.5, where the point estimate of $\Delta\mu_B - \Delta\mu_A$ (Week 5) and the 2.5-97.5% range becomes positive.

- The non-sequential correction method is milder than the sequential method and may need larger *f* to adjust.

A more realistic scenario is MDM is mixture of MAR and NFD-NMAR. This mixture MDM is a representation of different reasons of withdrawal: NFD-NMAR represents withdrawal for AE (adverse event) or LOE (lack of efficacy), and MAR represents withdrawal for LOF (loss of follow-up). Table 4.5.2 is the same as Table 4.5.1 except that the missing data is generated using this mixture MDM, and the NFD-Delta method, as described in Chapter 3, is used by incorporating the corresponding shifting parameters f=(0, v) for (MAR, NFD-NMAR) where v varies at: 0.25, 0.5 and 1. Three scenarios are here: First scenario that the two treatment groups have the same MAR (~10%) and NMAR (~25%); the second scenario that group A has MAR (~ 10%) and NMAR (15%), while group B has MAR (~ 10%) and NMAR (24%). In the NMAR model, the intercept (a) is the same, but the slopes (b) are different between the two treatment groups. For the third scenario, group A has MAR (~ 10%) and NMAR (6%), while group B has MAR (~ 10%) and NMAR (24%). In the NFD-NMAR model, the intercepts and the slopes are both different between the two treatment groups.

Table 4.5.2 Correctness of the Alpha Level for Mean-shift Method (NFD-Delta) When MDM is Mixture of MAR and NFD-NMAR; Delta unit = STD_1

Treatment Group		a	b	Week 5 Dropout Rates	f=0.25 Non-seq/ Sequential	f=0.50 Non-seq/ Sequential	f=1.0 Non-seq/ Sequential
A	MAR	-7.25	0.50	9.8			
	NFD-NMAR	-6.11	0.57	25.3			
$\Delta\mu_A$ (Week 5)	-2.45						
B	MAR	-7.25	0.50	9.8			
	NFD-	-6.11	0.57	25.3			

	NMAR						
$\Delta\mu_B$ (Week5)	-2.44						
$\Delta\mu_B - \Delta\mu_A$ (Week 5)	0.01				0.024/0.023	0.022/0.020	0.023/NN
A	MAR	-7.25	0.50	9.8			
	NFD-NMAR	-6.91	0.58	15.0			
$\Delta\mu_A$ (Week 5)	-2.29						
B	MAR	-7.25	0.50	9.8			
	NFD-NMAR	-6.91	0.68	24.2			
$\Delta\mu_B$ (Week 5)	-2.46						
$\Delta\mu_B - \Delta\mu_A$ (Week 5)	-0.17				0.040/0.024	0.032/0.012	0.022/NN
A	MAR	-7.25	0.50	9.8			
	NFD-NMAR	-7.91	0.58	6.4			
$\Delta\mu_A$ (Week 5)	-2.15						
B	MAR	-7.25	0.50	9.8			
	NFD-NMAR	-6.91	0.68	24.2			
$\Delta\mu_B$ (Week 5)	-2.46						
$\Delta\mu_B - \Delta\mu_A$ (Week 5)	-0.31				0.051/0.020	0.037/0.004	0.017/NN
	Estimate of $\Delta\mu_B - \Delta\mu_A$ at Week 5: 2.5-97.5 percentile range over simulation runs:				-0.101/-0.008 (-0.57,0.37)/ (-0.48,0.47)	-0.060/0.127 (-0.53,0.41)/ (-0.36, 0.62) (two-sided alpha=0.053)	0.022/NN (-0.45,0.50)/ NN

Similar to the pure NFD-NMAR case, Table 4.5.2 shows that:

- If dropout rates are the same between treatment groups, the alpha level holds at the nominal level even under mixture of MAR and NFD-NMAR, and no adjustment is needed.

- If dropout rates are different between treatment groups:

    - The greater imbalance of the dropout rate of NFD-NMAR, the higher the inflation of the type-I error rate by MI, thus a larger *f* is needed for adjustment;

    - The sequential method corrects the type-I error inflation rapidly and could easily over-correct and leads to power loss. So, when use sequential method, we should start with a small *f* to gradually make the correction;

    - The over-correction of the sequential method can be seen from the last case with f=0.5 where the point estimate of $\Delta\mu_B - \Delta\mu_A$ (Week5) becomes positive;

    - The non-sequential correction method is milder than the sequential method and may need a larger *f* to adjust.

Next, we also consider the case where the NFD-Delta method is used when the missing data is actually MAR. This is rarely the case that MAR occurs but not recognized in a clinical trial. Nevertheless, I report here the over-correction by the shift of the mean from MI could lead to severe loss of power in this rare situation.

Robustness

Table 4.5.3 Robustness of the Alpha Level for Mean-shift Method (NFD-Delta) When MDM is MAR; Delta unit = STD_1

Treatment Group	a	b	Week 5 Dropout Rates	f=0.25 Non-seq/ Sequential	f=0.5 Non-seq/ Sequential	f=1.5 Non-seq/ Sequential
A	-6.91	0.58	21.1			
B	-6.91	0.58	21.1			
$\Delta\mu_B - \Delta\mu_A$ (Week 5)	0.01			0.023/0.023	0.023/0.022	0.021/NN
A	-6.21	0.58	34.0			
B	-5.81	0.58	43.0			
$\Delta\mu_B - \Delta\mu_A$ (Week 5)	-0.08			0.019/0.01	0.015/0.005	0.006/NN
A	-6.91	0.58	21.1			
B	-6.91	0.68	34.1			
$\Delta\mu_B - \Delta\mu_A$ (Week 5)	-0.11			0.015/0.006	0.010/0.001	0.003/NN
A	-6.21	0.58	34.0			
B	-5.81	0.68	58.4			
$\Delta\mu_B - \Delta\mu_A$ (Week 5)	-0.23			0.018/0.002	0.011/0.000	0.001/NN
	Estimate of $\Delta\mu_B - \Delta\mu_A$ at Week 5: 2.5-97.5 percentile range over simulation runs:			0.060/0.233 (-0.52,0.64)/ (-0.36,0.82)	0.118/0.466 (-0.47,0.71)/ (-0.14,1.10)	0.347/NN (-0.27,0.97)/ NN

As shown in Table 4.5.3,

- When MDM is MAR, alpha level should be at or close to the nominal level if MI is used. But if mean shift method (NFD-delta) is used instead, the alpha level will be deflated, leading to losing power to some extent even when small *f* is used. When large *f* is used, the test could lose much of the power.

- The sequential acts more wildly than the non-sequential method.

- Mean-shift method (NFD-Delta) is robust under MAR only when the dropout distribution is the same between treatment groups.

4.3.2.2 Delta Mean-shift (NFD-Delta) Method with Unit of RSD_1

RSD_1 is the residual standard deviation of Y_1 adjusted for treatment group and Y_0 based on the complete data after the MI (under MAR). The algorithms using RSD_1 as the (constant) scale of the delta shift for all the time-points with missing data are either sequential or non-sequential. Since RSD_1 is applied to all time-points, the "single adjustment" method is not applicable here.

Correctness

Simulations shown in Tables 4.6.1 to 4.6.3 are to demonstrate the same as Table 4.5.1 to 4.5.3 except for using the constant RSD_1 (i.e., regression RSD at Week 1, adjusted for treatment group and baseline) as the scale/unit for the delta adjustments over time-points. These tables also show the sensitivity analyses by varying the shift parameter *f* with different algorithms (sequential and non-sequential).

Table 4.6.1 Correctness of the Alpha Level for Mean-shift Method (NFD-Delta) When MDM is NFD-NMAR; Delta unit = RSD_1

Treatment Group	a	b	Week 5 Dropout Rates	f=0.25 Non-seq/ Sequential	f=0.5 Non-seq/ Sequential	f=1.5 Non-seq/ Sequential
A	-6.91	0.68	29.1			
B	-6.91	0.68	29.1			
$\Delta\mu_B - \Delta\mu_A$ (Week 5)	0.00			0.024/0.022	0.024/0.022	0.023/NN
A	-6.21	0.68	42.0			
B	-5.81	0.68	50.4			
$\Delta\mu_B - \Delta\mu_A$ (Week 5)	-0.13			0.045/0.026	0.040/0.017	0.024/NN

A	-6.91	0.58	18.1			
B	-6.91	0.68	29.0			
$\Delta\mu_B - \Delta\mu_A$ (Week 5)	-0.19			0.053/0.034	0.046/0.019	0.021/NN
A	-6.21	0.58	29.4			
B	-5.81	0.68	50.3			
$\Delta\mu_B - \Delta\mu_A$ (Week 5)	-0.37 (RSD_1=1.36)			0.116/0.053	0.094/0.016	0.035/NN
	Estimate of $\Delta\mu_B - \Delta\mu_A$ at Week 5: 2.5-97.5 percentile range over simulation runs:			-0.230/-0.131 (-0.75,0.26)/ (-0.64,0.38)	-0.194/0.004 (-0.71,0.30)/ (-0.53,0.53)	-0.051/NN (-0.60,0.47)/ NN

*NN: "Not Needed", since the value is already known to be less than 0.025.

Compared between Tables 4.5.1 (metric is STD_1) and 4.6.1 (metric is RSD_1):

- At same level of *f*, using RSD may be easier than using STD to control the degree of the correction and to avoid over-correction that sequential method might cause;
- When the dropout rates are quite different between treatment groups, larger *f* may be needed to correct type-I error inflation by using the RSD metric than that by using the STD metric since RSD values are smaller and decreasing toward the end of the time-point.

Next, we consider the situation of mixture of MAR and NMAR in Table 4.6.2.

Table 4.6.2 Correctness of the Alpha Level for Mean-shift Method (NFD-Delta) When MDM is Mixture of MAR and NFD-NMAR; Delta unit = RSD_1

Treatment Group		a	b	Week 5 Dropout Rates	f=0.25 Non-seq/ Sequential	f=0.50 Non-seq/ Sequential	f=1.0 Non-seq/ Sequential
A	MAR	-7.25	0.50	9.8			
	NFD-NMAR	-6.11	0.57	25.3			
$\Delta\mu_A$	-2.45						

(Week 5)							
B	MAR	-7.25	0.50	9.8			
	NFD-NMAR	-6.11	0.57	25.3			
$\Delta\mu_B$ (Week 5)	-2.44						
$\Delta\mu_B - \Delta\mu_A$ (Week 5)	0.01				0.024/0.022	0.023/0.020	0.023/0.019
A	MAR	-7.25	0.50	9.8			
	NFD-NMAR	-6.91	0.58	15.0			
$\Delta\mu_A$ (Week 5)	-2.29						
B	MAR	-7.25	0.50	9.8			
	NFD-NMAR	-6.91	0.68	24.2			
$\Delta\mu_B$ (Week 5)	-2.46						
$\Delta\mu_B - \Delta\mu_A$ (Week 5)	-0.17				0.042/0.028	0.036/0.016	0.027/NN
A	MAR	-7.25	0.50	9.8			
	NFD-NMAR	-7.91	0.58	6.4			
$\Delta\mu_A$ (Week 5)	-2.15						
B	MAR	-7.25	0.50	9.8			
	NFD-NMAR	-6.91	0.68	24.2			
$\Delta\mu_B$ (Week 5)	-2.46						
$\Delta\mu_B - \Delta\mu_A$ (Week 5)	-0.31				0.056/0.026	0.044/0.009*	0.024/NN
	Estimate of $\Delta\mu_B - \Delta\mu_A$ at Week 5: 2.5-97.5 percentile range over simulation runs:				-0.111/-0.041 (-0.58,0.36)/ (-0.51,0.43)	-0.080/0.061 (-0.55,0.39)/ (-0.42, 0.54) (two-sided alpha=0.039)	-0.018/NN (-0.49, 0.45)/ NN

Similar to the pure NMAR case, and comparing between Tables 4.5.2 (STD metric) and 4.6.2 (RSD metric), for the mixture of MAR and NMAR case:

- At same level of *f*, using RSD metric may be easier to control the degree of the correction and avoid over-correction that sequential method might cause than using STD metric;

- When the dropout rates are quite different between treatment groups, larger *f* may be needed to correct type-I error inflation by using RSD than by using STD since RSD values are smaller and decreasing toward the end of the time-point.

Again, we also consider the case where the NFD-Delta method is used when the missing data is actually MAR. As indicated previously, this is rarely the case that MAR occurs but not recognized in a clinical trial. Nevertheless, I report here the over-correction by the shift of the mean from MI could lead to severe loss of power in this rare situation, as shown in Table 4.6.3, when the dropouts are different between treatment groups.

Table 4.6.3 Robustness of the Alpha Level for Mean-shift Method (NFD-Delta) When the MDM is MAR; Delta unit = RSD_1

Treatment Group	a	b	Week 5 Dropout Rates	f=0.25 Non-seq/ Sequential	f=0.5 Non-seq/ Sequential	f=1.5 Non-seq/ Sequential
A	-6.91	0.58	21.1			
B	-6.91	0.58	21.1			
$\Delta\mu_B - \Delta\mu_A$ (Week 5)	0.01			0.022/0.023	0.023/0.023	0.021/NN
A	-6.21	0.58	34.0			
B	-5.81	0.58	43.0			
$\Delta\mu_B - \Delta\mu_A$ (Week 5)	-0.08			0.021/0.012	0.018/0.007	0.009/NN
A	-6.91	0.58	21.1			
B	-6.91	0.68	34.1			
$\Delta\mu_B - \Delta\mu_A$ (Week 5)	-0.11			0.017/0.009	0.012/0.003	0.005/NN
A	-6.21	0.58	34.0			
B	-5.81	0.68	58.4			

$\Delta\mu_B - \Delta\mu_A$ (Week 5)	-0.23	0.019/0.004	0.014/0.000	0.002/NN
	Estimate of $\Delta\mu_B - \Delta\mu_A$ at Week 5: 2.5-97.5 percentile range over simulation runs:	0.047/0.176 (-0.54,0.63)/ (-0.41,0.76)	0.090/0.352 (-0.50,0.67)/ (-0.25,0.96)	0.262/NN (-0.34,0.87)/ NN

4.3.2.3 Delta Mean-shift (NFD-Delta) Method with Unit of STD_k

STD_k is the pooled standard deviation (over treatment groups) of Y_k based on the complete data after the MI (under MAR). The algorithms using STD_k as the (time-varying) scale of the delta shift for each time-points with missing data are either sequential or non-sequential. For the single adjustment method, $STD_{k\text{-first}}$ is used, which is the STD of the time-point where the missing data first occurred for that subject. The single adjustment algorithm applies the delta shift only once at the time-point where the missing data first occurred, then the rest adjustments are taken care of automatically through the correlations. STD_k increases as k increases from 1 to T.

Correctness

Table 4.7.1 Correctness of the Alpha Level for Mean-shift Method (NFD-Delta) When MDM is NFD-NMAR; Delta unit = STD_k

Treatment Group	a	b	Week 5 Dropout Rates	f=0.25 Single-adjust/ Non-seq/ Sequential	f=0.5 Single-adjust/ Non-seq/ Sequential	f=1.0 Single-adjust/ Non-seq/ Sequential	f=1.5 Single-adjust/ Non-seq/ Sequential
A	-6.91	0.68	29.1				
B	-6.91	0.68	29.1				
$\Delta\mu_B - \Delta\mu_A$	0.00			0.023/0.024/ 0.022	0.022/0.023/ 0.021	0.022/0.023/ 0.019	0.021/0.021/ NN

(Week 5)	Estimate of $\Delta\mu_B - \Delta\mu_A$ at Week 5: 2.5-97.5 percentile range over simulation runs:			-0.000/ 0.002/ 0.000 (-0.49,0.47)/ (-0.49,0.47)/ (-0.51,0.49)	-0.000/ 0.002/ 0.001 (-0.50,0.48)/ (-0.50,0.48)/ (-0.54,0.53)	-0.000/ 0.002/ 0.003 (-0.52,0.49)/ (-0.52,0.50)/ (-0.67,0.67)	-0.000/ 0.003/ NN (-0.54,0.52)/ (-0.55,0.52)/ NN
A	-6.21	0.68	42.0				
B	-5.81	0.68	50.4				
$\Delta\mu_B - \Delta\mu_A$ (Week 5)	-0.13			0.037/0.042/ 0.020	0.033/0.034/ 0.008	0.025/0.021/ 0.002	0.016/0.013/ NN
A	-6.91	0.58	18.1				
B	-6.91	0.68	29.0				
$\Delta\mu_B - \Delta\mu_A$ (Week 5)	-0.19			0.048/0.047/ 0.024	0.039/0.035/ 0.008	0.024/0.019/ 0.001	0.014/0.009/ NN
A	-6.21	0.58	29.4				
B	-5.81	0.68	50.3				
$\Delta\mu_B - \Delta\mu_A$ (Week 5)	-0.37 (STD_k; k=1 to 5) = (1.78,2.03,2.18,2.25,2.34)			0.100/0.101/ 0.025	0.073/0.070/ 0.003	0.039/0.027/ 0.000	0.017/0.007* /NN
	Estimate of $\Delta\mu_B - \Delta\mu_A$ at Week 5: 2.5-97.5 percentile range over simulation runs:			-0.223/ -0.204/ -0.046 (-0.73,0.28)/ (-0.72,0.29)/ (-0.57,0.48)	-0.180/ -0.14/ 0.189 (-0.69,0.33)/ (-0.67,0.36)/ (-0.37,0.75)	-0.092/ -0.020/ 0.737 (-0.62,0.44)/ (-0.57, 0.51)/ (0.03,1.46)	-0.000/ 0.103/ NN (-0.55,0.55)/ (-0.46, 0.66)/ NN

As in the case of previous metric STD_1 or RSD_1, Table 4.7.1 shows that when using STD_k:

- When the dropout rates are the same between treatment groups, the alpha level holds at nominal level.
- When the dropout rates are different between treatment groups:
    - STD_k increases as time point increases. The single adjustment method only adjusts at the first time point with missing data, the extent of its correction is lowest among the three methods. The graduate control of the type-I error rate by increasing f is smother than the sequential method. It is a good choice for low to intermediate imbalance of dropout rates when using STD_k. In addition,

the single adjustment is easier to implement than sequential and non-sequential methods.

- Non-sequential method is another good choice for low to intermediate imbalance of dropout rate between treatment groups when using STD_k. It may need a little larger *f* than the single adjustment needs.
- More caution is needed for the sequential method with STD_k. It can easily and unexpectedly over-correct and cause the unstable results when large imbalance in dropout rates is present. So, I recommend that we should avoid this approach. It is only acceptable if the difference in dropout rates is very large, and meanwhile, use small f, say 0.25.

Next, we consider the case of mixture of MAR and NMAR as shown in Table 4.7.2.

Table 4.7.2 Correctness of the Alpha Level for Mean-shift Method (NFD-Delta) When MDM is Mixture of MAR and NFD-NMAR; Delta unit = STD_k

Treatment Group		a	b	Week 5 Dropout rates	f=0.25 Single-adjust/ Non-seq/ Sequential	f=0.50 Single-adjust/ Non-seq/ Sequential	f=1.0 Single-adjust/ Non-seq/ Sequential
A	MAR	-7.25	0.50	9.8			
	NFD-NMAR	-6.11	0.57	25.3			
$\Delta\mu_A$ (Week 5)	-2.45						
B	MAR	-7.25	0.50	9.8			
	NFD-NMAR	-6.11	0.57	25.3			
$\Delta\mu_B$ (Week 5)	-2.44						
$\Delta\mu_B - \Delta\mu_A$ (Week 5)	0.01				0.021/ 0.023/ 0.021	0.022/ 0.023/ 0.020	0.021/ 0.022/ NN
A	MAR	-7.25	0.50	9.8			
	NFD-NMAR	-6.91	0.58	15.0			

$\Delta\mu_A$ (Week 5)	-2.29						
B	MAR	-7.25	0.50	9.8			
	NFD-NMAR	-6.91	0.68	24.2			
$\Delta\mu_B$ (Week 5)	-2.46						
$\Delta\mu_B$ - $\Delta\mu_A$ (Week 5)	-0.17				0.037/ 0.036/ 0.018	0.031/ 0.028/ 0.006	0.020/ 0017/ NN
A	MAR	-7.25	0.50	9.8			
	NFD-NMAR	-7.91	0.58	6.4			
$\Delta\mu_A$ (Week 5)	-2.15						
B	MAR	-7.25	0.50	9.8			
	NFD-NMAR	-6.91	0.68	24.2			
$\Delta\mu_B$ (Week 5)	-2.46						
$\Delta\mu_B$ - $\Delta\mu_A$ (Week 5)	-0.31				0.045/ 0.047/ 0.012	0.030/ 0.029/ 0.001	0.014/ 0.009/ NN
	Estimate of $\Delta\mu_B$ - $\Delta\mu_A$ at Week 5:				-0.101/ -0.087/ 0.028	-0.059/ -0.033/ 0.207	0.027/ 0.076/ NN
	2.5-97.5 percentile range over simulation runs:				(-0.57,0.37)/ (-0.56,0.38)/ (-0.44,0.51)	(-0.53,0.42)/ (-0.51, 0.44)/ (-0.30,0.72)	(-0.46,0.52)/ (-0.41,0.56)/ NN

The observations are similar to the case of pure NFD-NMAR in Table 4.7.1.

The case of using NFD-Delta method when the missing data is actually MAR is shown in Table 4.7.3.

Table 4.7.3 Robustness of the Alpha Level for Mean-shift Method (NFD-Delta) When the MDM is MAR; Delta unit = STD_k

Treatment Group	a	b	Week 5 Dropout Rates	f=0.25 Single-adjust/ Non-seq/ Sequential	f=0.5 Single-adjust/ Non-seq/ Sequential	f=1.0 Single-adjust/ Non-seq/ Sequential
A	-6.91	0.58	21.1			
B	-6.91	0.58	21.1			

$\Delta\mu_B - \Delta\mu_A$ (Week 5)	0.01			0.021/0.023/ 0.023	0.022/0.023/ 0.022	NN
A	-6.21	0.58	34.0			
B	-5.81	0.58	43.0			
$\Delta\mu_B - \Delta\mu_A$ (Week 5)	-0.08			0.016/0.018/ 0.008	0.013/0.013/ 0.003	NN
A	-6.91	0.58	21.1			
B	-6.91	0.68	34.1			
$\Delta\mu_B - \Delta\mu_A$ (Week 5)	-0.11			0.013/0.013/ 0.004	0.011/0.008/ 0.001	NN
A	-6.21	0.58	34.0			
B	-5.81	0.68	58.4			
$\Delta\mu_B - \Delta\mu_A$ (Week 5)	-0.23			0.012/0.014/ 0.001	0.009/0.008/ 0.000	NN
	Estimate of $\Delta\mu_B - \Delta\mu_A$ at Week 5:			0.055/0.081/ 0.297	0.111/0.158/ 0.621	NN
	2.5-97.5 percentile range over simulation runs:			(-0.53,0.63)/ (-0.50,0.67)/ (-0.30,0.90)	(-0.49,0.69)/ (-0.43,0.76)/ (-0.02,1.28)	

Also, similar to the previously examined cases with metrics STD_1 or RSD_1, we would lose power if the mean-shift method is used under MAR, as the alpha is lower than the nominal level, unless the dropout rates are the same between treatment groups.

4.3.2.4 Delta Mean-shift (NFD-Delta) Method with Unit of RSD_k

RSD_k is the residual standard deviation of Y_k adjusted for treatment group and Y_0 to Y_{k-1} based on the complete data after the MI (under MAR). The algorithms using RSD_k as the (time-varying) scale of the delta shift for each time-points with missing data are either sequential or non-sequential. For the single adjustment method, $RSD_{k\text{-first}}$ is used, which is the RSD of the time-point where the missing data first occurred for that subject. The single adjustment algorithm applies the delta shift only once at the time-point where the missing data first occurred, then the rest adjustments are taken care of automatically through the correlations. RSD_k decreases as k increases from 1 to T.

Correctness

Table 4.8.1 Correctness of the Alpha Level for Mean-shift Method (NFD-Delta) When MDM is NFD-NMAR; Delta unit = RSD_k

Treatment Group	a	b	Week 5 Dropout rates	f=0.25 Single-adjust/ Non-seq/ Sequential	f=1.5 Single-adjust/ Non-seq/ Sequential	f=2.0 Single-adjust/ Non-seq/ Sequential	f=2.5 Single-adjust/ Non-seq/ Sequential
A	-6.91	0.68	29.1				
B	-6.91	0.68	29.1				
$\Delta\mu_B - \Delta\mu_A$ (Week 5)	0.00			0.023/0.024/ 0.022	0.022/0.023/ 0.020	0.021/0.024 /NN	0.020/0.023/ NN
A	-6.21	0.68	42.0				
B	-5.81	0.68	50.4				
$\Delta\mu_B - \Delta\mu_A$ (Week 5)	-0.13			0.038/0.048/ 0.030	0.024/0.031/ 0.004	0.020/0.027/ NN	0.016/0.022/ NN
A	-6.91	0.58	18.1				
B	-6.91	0.68	29.0				
$\Delta\mu_B - \Delta\mu_A$ (Week 5)	-0.19			0.052/0.057/ 0.039	0.030/0.033/ 0.004	0.022/ 0.024/ NN	0.016/0.020/ NN
A	-6.21	0.58	29.4				
B	-5.81	0.68	50.3				
$\Delta\mu_B - \Delta\mu_A$ (Week 5)	-0.37 (RSD_k; k=1 to 5) = (1.36, 1.12, 1.08, 0.98, 0.89)			0.111/0.126/ 0.067	0.042/0.060/ 0.001	0.026/0.044/ NN	0.016/0.030/ NN
	Estimate of $\Delta\mu_B - \Delta\mu_A$ at Week 5: 2.5-97.5 percentile range over simulation runs:			-0.239/-0.242/ -0.161 (-0.74,0.26)/ (-0.76,0.25)/ (-0.67,0.34)	-0.102/ -0.125/ 0.362 (-0.62,0.42)/ (-0.66,0.38)/ (-0.34,0.95)	-0.047/ -0.078/ NN (-0.57, 0.48)/ (-0.61,0.44)/ NN	0.008/-0.031/ NN (-0.53,0.54)/ (-0.58,0.50)/ NN

Table 4.8.1 shows that, using RSD_k metric, under NMAR

- When dropout rates are the same between treatment groups, the alpha level holds at the nominal level.
- When dropout rates are different between treatment groups:

- The single adjustment method only adjusts at the first time-point where missing data occurs, and the adjustments for the rest time-points will be taken care by the correlations through the MI process. Since RSD_k decreases as time point increases, the extent of correction by the single adjustment is larger than non-sequential method at Week 5, but less than the (cumulative) sequential method.
- The non-sequential correction is the weakest, hence requires larger shift *f.*
- As shown, with the last simulation layout, the single adjustment method could correct the alpha level to the nominal level with the shift parameter f =2.0 using the $RSD_{k\text{-first}}$ metric, and provides a consistent estimate of $\Delta\mu_B$ - $\Delta\mu_A$ at Week 5.
- We should always be cautious when using the sequential method, as it easily over-correct and results in losing power.
- Compared to STD_k, we need larger shift *f* when using RSD_k.

Table 4.8.2 Correctness of the Alpha Level for Mean-shift Method (NFD-Delta) When MDM is Mixture of MAR and NFD-NMAR; Delta unit = RSD_k

Treatment Group		a	b	Week 5 Dropout Rates	f=0.25 Single-adjust/ Non-seq/ Sequential	f=0.50 Single-adjust/ Non-seq/ Sequential	f=1.0 Single-adjust/ Non-seq/ Sequential
A	MAR	-7.25	0.50	9.8			
	NFD-NMAR	-6.11	0.57	25.3			
$\Delta\mu_A$ (Week 5)	-2.45						
B	MAR	-7.25	0.50	9.8			

	NFD-NMAR	-6.11	0.57	25.3			
$\Delta\mu_B$ (Week 5)	-2.44						
$\Delta\mu_B - \Delta\mu_A$ (Week 5)	0.01				0.022/0.024/ 0.022	0.022/0.024/ 0.022	0.022/0.023/ NN
A	MAR	-7.25	0.50	9.8			
	NFD-NMAR	-6.91	0.58	15.0			
$\Delta\mu_A$ (Week 5)	-2.29						
B	MAR	-7.25	0.50	9.8			
	NFD-NMAR	-6.91	0.68	24.2			
$\Delta\mu_B$ (Week 5)	-2.46						
$\Delta\mu_B - \Delta\mu_A$ (Week 5)	-0.17				0.040/0.044 /0.031	0.036/0.040/ 0.020	0.029/0.032/ NN
A	MAR	-7.25	0.50	9.8			
	NFD-NMAR	-7.91	0.58	6.4			
$\Delta\mu_A$ (Week 5)	-2.15						
B	MAR	-7.25	0.50	9.8			
	NFD-NMAR	-6.91	0.68	24.2			
$\Delta\mu_B$ (Week 5)	-2.46						
$\Delta\mu_B - \Delta\mu_A$ (Week 5)	-0.31				0.055/0.064/ 0.034	0.044/0.052/ 0.016	0.028/0.037/ NN
	Estimate of $\Delta\mu_B - \Delta\mu_A$ at Week 5:				-0.120/-0.122/ -0.067	-0.097/-0.102/ 0.010	-0.050/- 0.063/NN
	2.5-97.5 percentile range over simulation runs:				(-0.58,0.34)/ (-0.59,0.35)/ (-0.53,0.40)	(-0.56,0.37)/ (-0.57,0.37)/ (-0.46,0.48)	(-0.52, 0.42)/ (-0.53,0.40)/ NN

Table 4.8.2 shows the similar observation for the mixture MAR and NMAR data as in Table 4.8.1 for the purely NMAR data. As shown, with the last simulation layout, the single adjustment method could correct the alpha level to the nominal level with the shift parameter f =1.0 using the $RSD_{k\text{-first}}$ metric, and provides a consistent estimate of $\Delta\mu_B$ - $\Delta\mu_A$ at Week 5.

Table 4.8.3 Robustness of the Alpha Level for Mean-shift Method (NFD-Delta) When the MDM is MAR; Delta unit = RSD_k

Treatment Group	a	b	Week 5 Dropout Rates	f=0.25 Single-adjust/ Non-seq/ Sequential	f=1.5 Single-adjust/ Non-seq/ Sequential	f=2.5 Single-adjust/ Non-seq/ Sequential
A	-6.91	0.58	21.1			
B	-6.91	0.58	21.1			
$\Delta\mu_B$ - $\Delta\mu_A$ (Week 5)	0.01			0.021/0.023/0.022	0.022/0.023/0.021	NN
A	-6.21	0.58	34.0			
B	-5.81	0.58	43.0			
$\Delta\mu_B$ - $\Delta\mu_A$ (Week 5)	-0.08			0.018/0.021/0.014	0.009/0.013/0.003	NN
A	-6.91	0.58	21.1			
B	-6.91	0.68	34.1			
$\Delta\mu_B$ - $\Delta\mu_A$ (Week 5)	-0.11			0.016/0.018/0.011	0.006/0.008/0.000	NN
A	-6.21	0.58	34.0			
B	-5.81	0.68	58.4			
$\Delta\mu_B$ - $\Delta\mu_A$ (Week 5)	-0.23			0.014/0.020/0.007	0.003/0.007/0.000	NN
	Estimate of $\Delta\mu_B$ - $\Delta\mu_A$ at Week 5:			0.036/0.031/0.135	0.216/0.168/0.807	NN
	2.5-97.5 percentile range over simulation runs:			(-0.55,0.61)/ (-0.55,0.61)/ (-0.45,0.71)	(-0.38,0.81)/ (-0.42,0.77)/ (0.14,1.49)	

Table 4.8.3 shows that, consistent to the previously examined cases with metrics STD_1, RSD_1, and STD_k, we would lose power if the mean-shift method is used under MAR, unless the dropout rates are the same between treatment groups.

4.4 Robustness of NFD-Delta Method for Dependent Dropouts

Let D = log(L) denote the log of the follow up time L. For dependent dropouts, I follow the approach in Wei and Shih (2001) by generating (Y_0, Y_1, Y_2, Y_3, Y_4, Y_5, D) from multivariate normal distribution with mean

(6.88, 6.15, 5.52, 5.25, 5.07, 4.92, 1.75) (4.8)

for treatment group A and

(6.88, 6.15, 5.52, 5.25, 5.07, 4.92, 2.3) (4.9)

for treatment group B.

a. Low Dependence Case

For the low dependence case, the correlations between D and Y are between 0.10 to 0.17 for group A and 0.08 to 0.13 for Group B. The variance-covariance (correlation) matrix is

	Y_0	Y_1	Y_2	Y_3	Y_4	Y_5	D
Y_0	2.27	1.85	1.70	1.70	1.69	1.60	0.175
	(1.00)	(0.66)	(0.52)	(0.48)	(0.46)	(0.42)	(0.17)
Y_1	1.85	3.50	3.45	3.45	3.28	3.13	0.175
	(0.66)	(1.00)	(0.85)	(0.78)	(0.71)	(0.65)	(0.14)
Y_2	1.70	3.45	4.70	4.48	4.34	4.13	0.175
	(0.52)	(0.85)	(1.00)	(0.89)	(0.83)	(0.76)	(0.12)
Y_3	1.70	3.45	4.48	5.48	5.17	5.14	0.175
	(0.48)	(0.78)	(0.89)	(1.00)	(0.91)	(0.88)	(0.11)

Y_4	1.69	3.28	4.34	5.17	5.87	5.65	0.175
	(0.46)	(0.71)	(0.83)	(0.91)	(1.00)	(0.92)	(0.11)
Y_5	1.60	3.13	4.13	5.14	5.64	6.29	0.175
	(0.42)	(0.65)	(0.76)	(0.88)	(0.92)	(1.00)	(0.10)
D	0.175	0.175	0.175	0.175	0.175	0.175	0.45
	(0.17)	(0.14)	(0.12)	(0.11)	(0.11)	(0.10)	(1.0)

for treatment group A, and the variance-covariance (correlation) matrix is

	Y_0	**Y_1**	**Y_2**	**Y_3**	**Y_4**	**Y_5**	**D**
Y_0	2.27	1.85	1.70	1.70	1.69	1.60	0.175
	(1.00)	(0.66)	(0.52)	(0.48)	(0.46)	(0.42)	(0.13)
Y_1	1.85	3.50	3.45	3.45	3.28	3.13	0.175
	(0.66)	(1.00)	(0.85)	(0.78)	(0.71)	(0.65)	(0.10)
Y_2	1.70	3.45	4.70	4.48	4.34	4.13	0.175
	(0.52)	(0.85)	(1.00)	(0.89)	(0.83)	(0.76)	(0.09)
Y_3	1.70	3.45	4.48	5.48	5.17	5.14	0.175
	(0.48)	(0.78)	(0.89)	(1.00)	(0.91)	(0.88)	(0.08)
Y_4	1.69	3.28	4.34	5.17	5.87	5.65	0.175
	(0.46)	(0.71)	(0.83)	(0.91)	(1.00)	(0.92)	(0.08)
Y_5	1.60	3.13	4.13	5.14	5.64	6.29	0.175
	(0.42)	(0.65)	(0.76)	(0.88)	(0.92)	(1.00)	(0.08)
D	0.175	0.175	0.175	0.175	0.175	0.175	0.85
	(0.13)	(0.10)	(0.09)	(0.08)	(0.08)	(0.08)	(1.00)

for treatment group B.

The missing data is generated by comparing exp(D) = L to the intervals < 1, (1, 2), (2, 3), (4, 5), and > 5: if $j-1 < L < j$ then y_j to y_5 are set missing. Completers are those with $L > 5$. The baseline y_0 is always observed.

In this case, Table 4.9 shows that the dropouts do not affect the mean change from baseline in the treatment groups; the resulting $\Delta\mu_A$ and $\Delta\mu_B$ are basically the same as (4.2). Hence it seems that weak correlations between D and Y in the range of 0.10 to 0.17 do not materially differ from the MCAR case where D and Y are independent.

Table 4.9 Cumulative dropout rates at each time point (%) by treatment group and resulting mean change from baseline, and difference ($\Delta\mu_B - \Delta\mu_A$) – Dependent Dropouts with weak correlation

Treatment Group	Week1	Week2	Week3	Week4	Week5
A	0.4	5.8	16.6	29.4	41.7
$\Delta\mu_A$	-0.73	-1.36	-1.63	-1.81	-1.97
B	0.6	4.1	9.7	16.2	22.8
$\Delta\mu_B$	-0.74	-1.36	-1.63	-1.81	-1.96
$\Delta\mu_B - \Delta\mu_A$	0.00	0.00	0.01	0.00	0.01

Table 4.10 Robustness of the alpha level for MI and mean-shift method (NFD-Delta) when the MDM is Dependent Missing (Low Correlation) (nominal alpha=0.025); Delta unit = STD_k / RSD_k ; Algorithm = Non-sequential

Treatment Group	Week5 Dropout Rates	MI nimpute=5	NFD-Delta Non-Sequential		
			f = 0.25	f = 0.5	f = 1.0
A	41.7				
$\Delta\mu_A$ (Week5)	-1.97				
B	22.8				
$\Delta\mu_B$ (Week5)	-1.96				
$\Delta\mu_A - \Delta\mu_B$ (Week5)	0.01	*0.030*	*0.052/ 0.038*	*0.075/ 0.044*	*0.153/ 0.061*
	Estimate of $\Delta\mu_B - \Delta\mu_A$ at Week 5: 2.5-97.5 percentile range over simulation runs:	*-0.006* *(-0.51,0.51)*	*-0.066/-0.027* *(-0.57,0.44)/ (-0.53,0.48)*	*-0.125/-0.049* *(-0.64,0.38)/ (-0.55,0.46)*	*-0.244/-0.091* *(-0.76,0.28)/ (-0.60,0.42)*

Table 4.10 shows that:

- Type-I error rate will not inflate much if we use MI under future-dependent missing data mechanism with low (or weak) correlation between D and Y which is close to the case of MCAR in which D and Y are independent;
- Non-future dependent mean-shift model doesn't work robustly if the MDM is future-dependent even with low or weak correlation between D and Y: the type-I error rate is inflated and will not be corrected to the nominal level:
    - The bigger the magnitude, the bigger inflation of type-I error rate.
    - STD seems to cause more inflation than RSD.

b. Intermediate Dependence Case

For the intermediate dependence case, the correlations between D and Y are about 0.50 for group A and 0.36 for Group B. The variance-covariance (correlation) matrix is

	$\mathbf{Y_0}$	$\mathbf{Y_1}$	$\mathbf{Y_2}$	$\mathbf{Y_3}$	$\mathbf{Y_4}$	$\mathbf{Y_5}$	**D**
$\mathbf{Y_0}$	2.27	1.85	1.70	1.70	1.69	1.60	0.50
	(1.00)	(0.66)	(0.52)	(0.48)	(0.46)	(0.42)	(0.49)
$\mathbf{Y_1}$	1.85	3.50	3.45	3.45	3.28	3.13	0.625
	(0.66)	(1.00)	(0.85)	(0.78)	(0.71)	(0.65)	(0.50)
$\mathbf{Y_2}$	1.70	3.45	4.70	4.48	4.34	4.13	0.719
	(0.52)	(0.85)	(1.00)	(0.89)	(0.83)	(0.76)	(0.49)
$\mathbf{Y_3}$	1.70	3.45	4.48	5.48	5.17	5.14	0.777
	(0.48)	(0.78)	(0.89)	(1.00)	(0.91)	(0.88)	(0.49)
$\mathbf{Y_4}$	1.69	3.28	4.34	5.17	5.87	5.65	0.804
	(0.46)	(0.71)	(0.83)	(0.91)	(1.00)	(0.92)	(0.49)
$\mathbf{Y_5}$	1.60	3.13	4.13	5.14	5.64	6.29	0.830
	(0.42)	(0.65)	(0.76)	(0.88)	(0.92)	(1.00)	(0.49)
D	0.50	0.625	0.719	0.777	0.804	0.830	0.45
	(0.49)	(0.50)	(0.49)	(0.49)	(0.49)	(0.49)	(1.0)

for treatment group A, and the variance-covariance (correlation) matrix is

	$\mathbf{Y_0}$	$\mathbf{Y_1}$	$\mathbf{Y_2}$	$\mathbf{Y_3}$	$\mathbf{Y_4}$	$\mathbf{Y_5}$	**D**
$\mathbf{Y_0}$	2.27	1.85	1.70	1.70	1.69	1.60	0.50

	(1.00)	(0.66)	(0.52)	(0.48)	(0.46)	(0.42)	(0.36)
$\mathbf{Y_1}$	1.85	3.50	3.45	3.45	3.28	3.13	0.625
	(0.66)	(1.00)	(0.85)	(0.78)	(0.71)	(0.65)	(0.36)
$\mathbf{Y_2}$	1.70	3.45	4.70	4.48	4.34	4.13	0.719
	(0.52)	(0.85)	(1.00)	(0.89)	(0.83)	(0.76)	(0.36)
$\mathbf{Y_3}$	1.70	3.45	4.48	5.48	5.17	5.14	0.777
	(0.48)	(0.78)	(0.89)	(1.00)	(0.91)	(0.88)	(0.36)
$\mathbf{Y_4}$	1.69	3.28	4.34	5.17	5.87	5.65	0.804
	(0.46)	(0.71)	(0.83)	(0.91)	(1.00)	(0.92)	(0.37)
$\mathbf{Y_5}$	1.60	3.13	4.13	5.14	5.64	6.29	0.830
	(0.42)	(0.65)	(0.76)	(0.88)	(0.92)	(1.00)	(0.36)
D	0.50	0.625	0.719	0.777	0.804	0.830	0.85
	(0.36)	(0.36)	(0.36)	(0.36)	(0.37)	(0.36)	(1.00)

for treatment group B.

The missing data is generated by comparing exp(D) = L to the intervals < 1, (1, 2), (2, 3), (4, 5), and > 5: if j-1 < L < j then y_j to y_5 are set missing. Completers are those with L > 5. The baseline y_0 is always observed. Table 4.11 shows that the dependent dropout rates with intermediate correlations between D and Y affect the mean changes from baseline in each treatment group. The resulting $\Delta\mu_A$ and $\Delta\mu_B$ are not the same and the between-group differences in changes from baseline are not zero and this case differs from the case where D and Y are independent or with low correlations.

Table 4.11 Cumulative dropout rates at each time point (%) by treatment group, resulting mean change from baseline, and difference ($\Delta\mu_B - \Delta\mu_A$) departure from the null hypothesis – Dependent Dropouts (Intermediate Correlation)

Treatment Group	Week1	Week2	Week3	Week4	Week5
A	0.4	5.8	16.6	29.4	41.7
$\Delta\mu_A$	-0.73	-1.32	-1.50	-1.59	-1.63
B	0.6	4.1	9.7	16.1	22.7
$\Delta\mu_B$	-0.73	-1.34	-1.57	-1.71	-1.82

$\Delta\mu_B - \Delta\mu_A$	0.00	-0.02	-0.07	-0.12	-0.19

Table 4.12 Robustness of the alpha level for MI and mean-shift method (NFD-Delta) when the MDM is Dependent Missing (Intermediate Correlation) (nominal alpha=0.025); Delta unit = STD_k / RSD_k; Algorithm = Non-sequential

Treatment Group	Week5 Dropout Rates	MI nimpute=5	NFD-Delta Non-Sequential		
			f = 0.25	f = 0.5	f = 1.0
A	41.7				
$\Delta\mu_A$ (Week5)	-1.63				
B	22.7				
$\Delta\mu_B$ (Week5)	-1.82				
$\Delta\mu_A - \Delta\mu_B$ (Week5)	-0.19	0.060	0.092/ 0.069	0.131/ 0.084	0.240/ 0.109
	Estimate of $\Delta\mu_B - \Delta\mu_A$ at Week 5:	-0.087	-0.146/ -0.109	-0.205/ -0.130	-0.322/ -0.173
	2.5-97.5 percentile range over simulation runs:	(-0.58,0.43)	(-0.64,0.37)/ (-0.61,0.41)	(-0.69,0.30)/ (-0.63,0.38)	(-0.82,0.19)/ (-0.67,0.34)

Table 4.12 shows that:

- Type-I error rate will be inflated if we use MI under future-dependent missing data mechanism with intermediate correlation between D and Y;
- Non-future dependent mean-shift model doesn't work robustly if the MDM is future-dependent even with intermediate correlation between D and Y: the type-I error rate is inflated more than those with low or weak correlation between D and Y:
    - The bigger the magnitude, the bigger inflation of type-I error rate.

- STD seems to cause more inflation than RSD.

c. High Dependence Case

From the results of low and intermediate dependence cases, for the high dependence case, we expect that the difference of mean changes from baseline between treatment group will get bigger and type-I error rate will get more inflated than low or intermediate dependence cases. The simulation for the high dependence case was not done since we have already seen the trend from the low and intermediate dependence cases.

Non-future dependent mean-shift model doesn't work robustly if the MDM is future-dependent. Methodology for dealing with future-dependence missing data is an on-going research topic. Scharfstein, McDermott, Olson, Wiegand (2014) proposed a fully parametric methodology for global sensitivity analysis that can be extended to future-dependent missing data. The method can be implemented by the associated software built in R package. This topic is out of the scope of this dissertation and will not be discussed here in details.

4.5 Method for ACT Estimand – Control-based Imputation

ACT analysis has been used as another kind of sensitivity analysis. However, strictly speaking, ACT analysis should *not* be regarded as a sensitivity analysis; rather, it should be regarded as a secondary analysis since it is to address a different estimand to the AAT estimand (See Section 2.7). ACT analysis assumes that the dropouts are all NMAR and that those from the treatment group would behave like those in the control (e.g., placebo)

group, while AAT does not make this assumption. The discrepancy in the results, if it appears, would then be attributed to the interpretation of the different estimands. Note that in the literature (Ratitch et al., 2013; Lu, 2014), other authors regard placebo dropouts are all MAR.

I consider three scenarios here. The first scenario is the NFD-NMAR case, followed by the second scenario of the mixture of MAR and NFD-NMAR case. The third scenario is when the underlying MDM is really (the unlikely case of) MAR, thus the ACT analysis should not really be applied. The purpose is of the third scenario is to demonstrate the ACT analysis will lose power under MAR.

4.5.1 Alpha Level of ACT Analysis When MDM is NFD-NMAR

The layout of the NFD-NMAR case is as shown previously in Table 4.2.

Table 4.13 Alpha Level of the ACT Analysis When MDM is NFD-NMAR

Treatment Group	a	b	Week 5 Dropout Rates	ACT Analysis nimpute=5
A	-6.91	0.68	29.1	
B	-6.91	0.68	29.1	
$\Delta\mu_B - \Delta\mu_A$ (Week 5)	0.00			0.008
A	-6.21	0.68	42.0	
B	-5.81	0.68	50.4	
$\Delta\mu_B - \Delta\mu_A$ (Week 5)	-0.13			0.004
A	-6.91	0.58	18.1	
B	-6.91	0.68	29.0	
$\Delta\mu_B - \Delta\mu_A$ (Week 5)	-0.19			0.024
A	-6.21	0.58	29.4	
B	-5.81	0.68	50.3	
$\Delta\mu_B - \Delta\mu_A$ (Week 5)	-0.37			0.024
	Estimate of $\Delta\mu_B - \Delta\mu_A$ at Week 5:			-0.177

	2.5-97.5 percentile range over simulation runs:	(-0.50, 0.13)

Table 4.13 is very interesting. It shows that the control-based imputation, being very conservative works (i.e., preserve the nominal alpha level) only when there is a large imbalance in dropout rates between the treatment groups. If the differential dropouts rate are minor, e.g., less than 10%, this method could lose power.

Next, we examine the case of mixture of MAR and NMAR.

4.5.2 Alpha Level of ACT Analysis When MDM is Mixture of MAR and NFD-NMAR

The layout of the mixture of MAR and NFD-NMAR case is as shown previously in Table 4.2.

Table 4.14 Alpha Level of the ACT Analysis When MDM is Mixture of MAR and NFD-NMAR

Treatment Group		a	b	Week 5 Dropout Rates	ACT Analysis nimpute=5
A	MAR	-7.25	0.50	9.8	
	NFD-NMAR	-6.11	0.57	25.3	
$\Delta\mu_A$ (Week 5)	-2.45				
B	MAR	-7.25	0.50	9.8	
	NFD-NMAR	-6.11	0.57	25.3	
$\Delta\mu_B$ (Week 5)	-2.44				
$\Delta\mu_B - \Delta\mu_A$ (Week 5)	0.01				0.010
A	MAR	-7.25	0.50	9.8	
	NFD-NMAR	-6.91	0.58	15.0	
$\Delta\mu_A$ (Week 5)	-2.29				

B	MAR	-7.25	0.50	9.8	
	NFD-NMAR	-6.91	0.68	24.2	
$\Delta\mu_B$ (Week 5)	-2.46				
$\Delta\mu_B - \Delta\mu_A$ (Week 5)	-0.17				0.024
A	MAR	-7.25	0.50	9.8	
	NFD-NMAR	-7.91	0.58	6.4	
$\Delta\mu_A$ (Week 5)	-2.15				
B	MAR	-7.25	0.50	9.8	
	NFD-NMAR	-6.91	0.68	24.2	
$\Delta\mu_B$ (Week 5)	-2.46				
$\Delta\mu_B - \Delta\mu_A$ (Week 5)	-0.31				0.040
	Estimate of $\Delta\mu_B - \Delta\mu_A$ at Week 5:				-0.143
	2.5-97.5 percentile range over simulation runs:				(-0.53,0.24)

As show in Table 4.14, the conservative control-based imputation in the mixture of MAR and NMAR data loses power when the dropouts rates are similar (in each category of MAR and NMAR) between treatment groups, as the alpha is lower than the nominal level. On the other hand, the third scenario shows that it may be necessary to regard all MAR as NMAR (as in Table 4.13) in order to reach the nominal alpha level when the MAR/NMAR rates reversed in each treatment group. Therefore, I carry out the following simulation, as shown in Table 4.15, treating all MAR as if they were NMAR. It confirms that by regarding all dropouts are NMAR, the conservative control-based imputation would lose much power when the dropouts rates are similar (as the alpha level was much lower than the nominal level), but it preserves the nominal alpha level in other situations. This is an empirical support to the literature of Ratitch et al., 2013 and Lu, 2014.

Table 4.15 Alpha Level of the ACT Analysis When MDM is Mixture of MAR and NFD-NMAR (Treat MAR as NMAR)

Treatment Group		a	b	Week 5 Dropout Rates	ACT Analysis nimpute=5
A	MAR	-7.25	0.50	9.8	
	NFD-NMAR	-6.11	0.57	25.3	
$\Delta\mu_A$ (Week 5)	-2.45				
B	MAR	-7.25	0.50	9.8	
	NFD-NMAR	-6.11	0.57	25.3	
$\Delta\mu_B$ (Week 5)	-2.44				
$\Delta\mu_B - \Delta\mu_A$ (Week 5)	0.01				0.007
A	MAR	-7.25	0.50	9.8	
	NFD-NMAR	-6.91	0.58	15.0	
$\Delta\mu_A$ (Week 5)	-2.29				
B	MAR	-7.25	0.50	9.8	
	NFD-NMAR	-6.91	0.68	24.2	
$\Delta\mu_B$ (Week 5)	-2.46				
$\Delta\mu_B - \Delta\mu_A$ (Week 5)	-0.17				0.013
A	MAR	-7.25	0.50	9.8	
	NFD-NMAR	-7.91	0.58	6.4	
$\Delta\mu_A$ (Week 5)	-2.15				
B	MAR	-7.25	0.50	9.8	
	NFD-NMAR	-6.91	0.68	24.2	
$\Delta\mu_B$ (Week 5)	-2.46				
$\Delta\mu_B - \Delta\mu_A$ (Week 5)	-0.31				0.023
	Estimate of $\Delta\mu_B - \Delta\mu_A$ at Week 5:				-0.113
	2.5-97.5 percentile range over simulation runs:				(-0.47, 0.24)

4.5.3 Lack of Robustness of ACT Analysis When MDM is MAR

The layout of the MAR case is as shown previously in Table 4.1. As expected, if the missing data is actually all MAR, the control-based imputation is too conservative, results in losing almost all the power of the study, as shown in Table 4.16.

Table 4.16 Alpha Level of the ACT Analysis When MDM is MAR

Treatment Group	a	b	Week 5 Dropout Rates	ACT Analysis nimpute=5
A	-6.91	0.58	21.1	
B	-6.91	0.58	21.1	
$\Delta\mu_B - \Delta\mu_A$ (Week 5)	0.01			0.012
A	-6.21	0.58	34.0	
B	-5.81	0.58	43.0	
$\Delta\mu_B - \Delta\mu_A$ (Week 5)	-0.08			0.003
A	-6.91	0.58	21.1	
B	-6.91	0.68	34.1	
$\Delta\mu_B - \Delta\mu_A$ (Week 5)	-0.11			0.005
A	-6.21	0.58	34.0	
B	-5.81	0.68	58.4	
$\Delta\mu_B - \Delta\mu_A$ (Week 5)	-0.23			0.001
	Estimate of $\Delta\mu_B - \Delta\mu_A$ at Week 5:			-0.001
	2.5-97.5 percentile range over simulation runs:			(-0.32, 0.31)

4.6 Sample Size and Shift Parameter Selection

To examine whether the sample size affects the type-I error rate and the selection of shift parameter *f*, I performed additional simulations on the non-future dependent data with sample size of N=400 for each treatment group. Delta mean shift method was used and I compared the results to those results for the sample size of N=200 and examined whether or not the sample size affects the selection of the shift parameter. To serve this purpose,

non-sequential algorithm with RSD_k unit was selected to use since it's more stable and easier to control, which is also one of our recommendations of the algorithm and unit to use.

The following Table 4.17 displays the average dropout proportions over 5,000 simulation runs with sample size of N=400 for each treatment group.

Table 4.17 Cumulative dropout rates at each time point (%) by treatment group, resulting mean change from baseline, and difference ($\Delta\mu_A$ - $\Delta\mu_B$) – NFD-NMAR (N=400 for Each Treatment Group)

Treatment Group	a	b	Week1	Week2	Week3	Week4	Week5
A	-6.91	0.68	10.3	16.7	21.8	25.8	29.1
$\Delta\mu_A$			-0.83	-1.57	-1.95	-2.21	-2.44
B	-6.91	0.68	10.3	16.8	21.8	25.9	29.2
$\Delta\mu_B$			-0.83	-1.57	-1.95	-2.21	-2.44
$\Delta\mu_B$ - $\Delta\mu_A$			0.00	0.00	0.00	0.00	0.00
A	-6.21	0.68	17.1	26.6	33.2	38.2	42.2
$\Delta\mu_A$			-0.89	-1.69	-2.11	-2.39	-2.65
B	-5.81	0.68	22.2	33.4	40.8	46.4	50.6
$\Delta\mu_B$			-0.93	-1.76	-2.21	-2.52	-2.79
$\Delta\mu_B$ - $\Delta\mu_A$			-0.04	-0.08	-0.11	-0.12	-0.14
A	-6.91	0.58	5.4	9.4	12.8	15.7	18.3
$\Delta\mu_A$			-0.78	-1.47	-1.80	-2.04	-2.25
B	-6.91	0.68	10.3	16.8	21.8	25.8	29.1
$\Delta\mu_B$			-0.83	-1.57	-1.95	-2.21	-2.44
$\Delta\mu_B$ - $\Delta\mu_A$			-0.05	-0.10	-0.14	-0.17	-0.19
A	-6.21	0.58	9.8	16.4	21.6	25.9	29.5
$\Delta\mu_A$			-0.82	-1.55	-1.92	-2.18	-2.41
B	-5.81	0.68	22.2	33.5	40.9	46.4	50.5
$\Delta\mu_B$			-0.93	-1.77	-2.21	-2.52	-2.78
$\Delta\mu_B$ - $\Delta\mu_A$			-0.11	-0.22	-0.29	-0.34	-0.37

Compare this table to Table 4.2, we can see that the numbers in the two tables are almost the same. Therefore, it shows that the sample size does not affect the dropout rate, the resulting mean change from baseline, and difference of mean change from baseline.

The following Table 4.18 displays the results of the alpha level for mean-shift method under NFD-NMAR using non-sequential algorithm with RSD_k for N=200 and N=400 for each treatment group.

Table 4.18 Comparison of the Alpha Levels for Mean-shift Method (NFD-Delta) When MDM is NFD-NMAR Using Non-sequential Algorithm; Delta unit =RSD_k; N=200 vs N=400

Treatment Group	a	b	Week 5 Dropout Rates	f=0.25 N=200/ N =400	f=1.5 N=200/ N =400	f=2.0 N=200/ N =400	f=2.5 N=200/ N =400
A	-6.91	0.68	29.1				
B	-6.91	0.68	29.2				
$\Delta\mu_B - \Delta\mu_A$ (Week 5)	0.00			0.024/0.024	0.023/0.023	0.024/0.024	0.023/0.023
A	-6.21	0.68	42.2				
B	-5.81	0.68	50.6				
$\Delta\mu_B - \Delta\mu_A$ (Week 5)	-0.14			0.048/0.048	0.031/0.031	0.027/0.027	0.022/0.022
A	-6.91	0.58	18.3				
B	-6.91	0.68	29.1				
$\Delta\mu_B - \Delta\mu_A$ (Week 5)	-0.19			0.057/0.057	0.033/0.033	0.024/0.024	0.020/0.020
A	-6.21	0.58	29.5				
B	-5.81	0.68	50.5				
$\Delta\mu_B - \Delta\mu_A$ (Week 5)	-0.37			0.126/0.126	0.060/0.060	0.044/0.044	0.030/0.030
	Estimate of $\Delta\mu_B - \Delta\mu_A$ at Week 5:			-0.242/ -0.242	-0.125/ -0.125	-0.078/ -0.078	-0.031/ -0.031
	2.5-97.5 percentile range over simulation runs:			(-0.76,0.25)/ (-0.76,0.25)	(-0.66,0.38)/ (-0.66,0.38)	(-0.61,0.44)/ (-0.61,0.44)	(-0.58,0.50)/ (-0.58,0.50)

Table 4.18 shows that the alpha level for N=200 and N=400 are exactly same for each scenario and each shift parameter *f*. The estimates of $\Delta\mu_B - \Delta\mu_A$ at Week 5 and its 2.5-97.5 percentile range are all the same as well. Therefore, we can say that the sample size would not affect the type-I error rate and the selection of shift parameter when use mean-shift method under the MDM of NFD-NMAR.

To double confirm this conclusion, I also performed a simulation to check the alpha levels using single-adjustment algorithm with RSD_k. The results are displayed in the following Table 4.19.

Table 4.19 Comparison of the Alpha Levels for Mean-shift Method (NFD-Delta) When MDM is NFD-NMAR Using Single-adjustment Algorithm with N=200 and N=400 for Each Treatment Group; Delta unit =RSD_k

Treatment Group	a	b	Week 5 Dropout Rates	f=0.25 N=200/ N =400	f=1.5 N=200/ N =400	f=2.0 N=200/ N =400	f=2.5 N=200/ N =400
A	-6.91	0.68	29.1				
B	-6.91	0.68	29.2				
$\Delta\mu_B - \Delta\mu_A$ (Week 5)	0.00			0.023/0.023	0.022/0.022	0.021/0.021	0.020/0.020
A	-6.21	0.68	42.2				
B	-5.81	0.68	50.6				
$\Delta\mu_B - \Delta\mu_A$ (Week 5)	-0.14			0.038/0.038	0.024/0.024	0.020/0.020	0.016/0.016
A	-6.91	0.58	18.3				
B	-6.91	0.68	29.1				
$\Delta\mu_B - \Delta\mu_A$ (Week 5)	-0.19			0.052/0.052	0.030/0.030	0.022/0.022	0.016/0.016
A	-6.21	0.58	29.5				
B	-5.81	0.68	50.5				
$\Delta\mu_B - \Delta\mu_A$ (Week 5)	-0.37			0.111/0.111	0.042/0.042	0.026/0.026	0.016/0.016
	Estimate of $\Delta\mu_B - \Delta\mu_A$ at Week 5:			-0.239/ -0.239	-0.102/ -0.102	-0.047/ -0.047	0.008/ 0.008
	2.5-97.5 percentile range over simulation runs:			(-0.74,0.26)/ (-0.74,0.26)	(-0.62,0.42)/ (-0.62,0.42)	(-0.57,0.48)/ (-0.57,0.48)	(-0.53,0.54)/ (-0.53,0.54)

Similar to the case of non-sequential algorithm, with single-adjustment approach, the type-I error rate for sample size N=200 and N=400 are exactly same. The estimate of $\Delta\mu_B - \Delta\mu_A$ at Week 5 and its 2.5-97.5 percentile range are all the same as well. However, an interesting case may arise when we use STD_k. The following table shows the results of the alpha level for N=200 and N=400 for mean-shift method under NFD-NMAR when we use STD_k with non-sequential algorithm.

Table 4.20 Comparisons of Alpha Levels for Mean-shift Method (NFD-Delta) When MDM is NFD-NMAR Using Non-sequential Algorithm with N=200 and N=400 for Each Treatment Group; Delta unit = STD_k

Treatment Group	a	b	Week 5 Dropout Rates	f=0.25 N=200/ N =400	f=0.5 N=200/ N =400	f=1.0 N=200/ N =400	f=1.5 N=200/ N =400
A	-6.91	0.68	29.1				
B	-6.91	0.68	29.2				
$\Delta\mu_B - \Delta\mu_A$ (Week 5)	0.00			0.024/0.025	0.023/0.024	0.023/0.022	0.021/0.020
A	-6.21	0.68	42.2				
B	-5.81	0.68	50.6				
$\Delta\mu_B - \Delta\mu_A$ (Week 5)	-0.14			0.042/0.056	0.034/0.043	0.021/0.025	0.013/0.013
A	-6.91	0.58	18.3				
B	-6.91	0.68	29.1				
$\Delta\mu_B - \Delta\mu_A$ (Week 5)	-0.19			0.047/0.067	0.035/0.046	0.019/0.018	0.014/0.007
A	-6.21	0.58	29.5				
B	-5.81	0.68	50.5				
$\Delta\mu_B - \Delta\mu_A$ (Week 5)	-0.37			0.101/0.173	0.070/0.106	0.027/0.027	0.007/0.006
	Estimate of $\Delta\mu_B - \Delta\mu_A$ at Week 5:			-0.204/ -0.205	-0.140/ -0.144	-0.020/ -0.021	0.103/ 0.102
	2.5-97.5 percentile range over simulation runs:			(-0.72,0.29)/ (-0.57,0.16)	(-0.67,0.36)/ (-0.51,0.23)	(-0.57,0.51)/ (-0.40,0.36)	(-0.46,0.66)/ (-0.30,0.51)

From the above Table 4.20, we can see that the alpha levels for N=200 and N=400 are little different for some shift parameters *f,* especially for f=0.25 and f=0.50. The estimates are not the same either. We could not exactly explain why this occurred, but these results further confirmed our conclusion that the alpha level is not stable and not easy to control when we use STD_k, compared to use RSD_k. Therefore, we would always recommend using RSD instead of STD whenever possible.

Chapter 5

Conclusion and Discussion

5.1 Summary

Missing data is a common problem in longitudinal clinical trials. Substantial missing data could introduce potential biases in the comparison of treatment groups and undermine the scientific credibility of causal conclusions from clinical trials. In recent NRC report, it notes that the missing data can be reduced and prevented to some extent by good study design and implementation, and by the vigilance and extent of data collection. Nevertheless, although good study design, implementation, and data collection could reduce the amount of missing data, missing data is probably unavoidable and still likely to be present in most clinical trials and would need to be handled during analysis on the basis of some assumptions regarding the missing outcomes.

To handle the missing data issue, it is always required by the regulatory agencies to pre-specify a primary analysis and sensitivity analysis in protocol or statistical analysis plan (SAP). Recent NRC report questioned the reasonableness of the MAR setting as the primary analysis since MAR is very special and doubtful assumption for the missing data mechanism, and the report encourages to use NMAR setting as the primary analysis. One of the NMAR mechanisms is non-future dependence missing data (NFD-NMAR). It is also one of the recommended methods in the NRC report. This dissertation addressed this issue and proposed a process to investigate the mean-shift model with NFD-NMAR mechanism (NFD-Delta method). The goal was to find an "appropriate" specification of the delta adjustment in the study protocol or SAP. The "appropriateness" is based on the

maintenance of the type-I error rate. The shift parameter of the NFD-Delta method then constitutes the sensitivity analysis.

The process of NFD-Delta method for specifying the primary (NMAR) analysis for a phase III study protocol has to be based on its phase II study experience in order to gather information on the set up for simulations. The set up information includes study duration and time-points of longitudinal data collection, expected dropout rates and reasons in each treatment group, and the expected outcomes and correlations of repeated measures for the control group under the null hypothesis. The attempt is, through the simulations, to find an appropriate shift parameter to specify the primary NMAR analysis in the phase III study protocol/SAP.

Several components were considered for the NFD shift parameter in this dissertation: the metric/unit, magnitude, and the algorithm to place the shift to examine the effect of these components on the type-I error rate (alpha) under the null hypothesis of no treatment effect.

In this dissertation, I conducted extensive simulations to investigate the type-I error rate for different methods. With NFD-Delta method, for the metric factor, four different metric units were considered: constant STD_1, constant RSD_1, STD_k, RSD_k; for the magnitude factor, different values of shift parameter *f* were considered from low to high, for example, *f* was chose from 0.25 to 2.5, etc. to investigate which *f* value is the appropriate shift parameter to control the type-I error rate to the nominal level; for the algorithm to implement the delta shift, three different method were used: sequential, non-sequential and single adjustment method. The evaluation of different method and the

selection of shift parameter is to control and maintain the type-I error rate at the nominal level.

MMRM and MI under NFD-NMAR assumption

First, I examined the MMRM and MI method under NFD-NMAR mechanism and the results showed that MMRM and MI inflated type-I error rate in the situation when dropout rates are different between treatment groups and the point estimate is biased. The degree of inflation of type-I error rate is more dependent on the degree of the imbalance of the dropout rates between treatment groups than the dropout rate itself in each treatment group. When the dropout rates are the same between treatment groups, no type-I error inflation occurs. It is similar for the case of the mixture of MAR and NFD-NMAR.

Correctness of NFD-Delta (mean-shift) method under NFD-NMAR assumption

I also examined the NFD-Delta (mean-shift) method under NFD-NMAR mechanism and different algorithm sequential, non-sequential and single adjustment to place the delta with different metric STD_1, RSD_1, STD_k and RSD_k. The simulation results show that if dropout rates are the same between treatment groups, the alpha level holds at nominal level under NFD-NMAR, no adjustment is needed. When the dropout rates are different between treatment groups, the greater imbalance of dropout rates, the higher inflation of the type-I error rate is, thus the larger *f* is needed for adjustment. The sequential method corrects the type-I error inflation rapidly and could easily over-correct and leads to losing power. So, when use the sequential method, we should start with a small *f* and gradually to make the correction. The non-sequential correction method is milder than the

sequential method and may need larger *f* to adjust. At same level of *f*, using RSD may be easier than using STD to control the degree of the correction and to avoid over-correction that sequential method might cause. When the dropout rates are quite different between treatment groups, larger *f* may be needed to correct type-I error inflation by using the RSD metric than that by using the STD metric since RSD values are smaller and decreasing toward the end of the time-point.

When STD_k is used:

Since the single adjustment method only adjusts at the first time point with missing data, the extent of its correction is lowest among the three algorithm methods. The graduate control of the type-I error rate by increasing *f* is smoother than the sequential method. It is a good choice for low to intermediate imbalance of dropout rates when using STD_k. In addition, the single adjustment is easier to implement than sequential and non-sequential methods. Non-sequential method is another good choice for low to intermediate imbalance of dropout rate between treatment groups when using STD_k. It may need a little larger *f* than the single adjustment does. More caution is needed for the sequential method with STD_k. It can easily and unexpectedly over-correct and cause the unstable results when large imbalance in dropout rates is present. So, I recommend that we should avoid this approach. It is only acceptable if the difference in dropout rates is very large, and meanwhile, use small f, say 0.25.

When RSD_k is used:

The single adjustment method only adjusts at the first time-point where missing data occurs, and the adjustments for the rest time-points will be taken care by the correlations

through the MI process. Since RSD_k decreases as time point increases, the extent of correction by the single adjustment is larger than non-sequential method, but less than the (cumulative) sequential method. The non-sequential correction is the weakest, hence requires larger shift *f*. We should always be cautious when using the sequential method, as it easily over-correct and results in losing power. Compared to STD_k, we need larger shift *f* when using RSD_k.

Similar for the case of mixture of MAR and NFD-NMAR.

5.2 Discussion

5.2.1 Final Recommendations and Limitations with the NFD-Delta Method

Through the simulations conducted in this dissertation based on the diabetic pain study results, the non-sequential or the single adjustment algorithm using the metric of RSD seem to work better than the sequential algorithm or using the STD metric, thus are my recommendations. Looking back to Table 2.1, only one time point is unidentifiable and needs to make the assumption to link the distribution of dropouts and that of the observed by the shift. This might be one reason that contributions to the good performance of single adjustments of mean-shift model. However, since each trial with its own design (including endpoint and study duration) and data, it is not possible to establish a universal shift parameter that is appropriate for all types of studies when the primary analysis has to be based on a NMAR scenario. Rather, this dissertation provided, via the investigation process, a method of finding an appropriate shift parameter for any late phase trial by simulations based on either early phase data or information from interim data of the current trial.

The limitation of the NFD-Delta method is on its assumption of the non-future dependence missing data mechanism.

The simulation results show that we would lose power if the mean-shift method is used under MAR, unless the dropout rates are the same between treatment groups. This is a minor limitation since MAR is very unlikely. Another scenario is when the MDM is future-dependent. Results show that the NFD-Delta (mean-shift) doesn't work robustly for future-dependent missing data. Methodology for dealing with future-dependence missing data is an on-going research topic.

5.2.2 Regarding the ACT Estimand - Control-based Imputation

The simulation shows that the control-based imputation, being very conservative (i.e., preserve the nominal alpha level) works only when there is a large imbalance in dropout rates between the treatment groups. If the differential dropouts rate are minor, e.g., less than 10%, this method could lose power.

5.2.3 Power Issue

Power is always another issue. From the simulation results, we can see the mean-shift method could lose power if we over-adjust the delta aggressively. To investigate the power, we need first to fix the type-I error rate. Based on the simulation results and my recommendations, to investigate the power, we need to use single-adjustment (or non-sequential adjustment) method, use RSD and fix the shift parameter f that corrects the alpha level to the nominal level and then use simulations to investigate the power under a set of specified alternative hypotheses.

5.3 Future Work

In this dissertation, for the NFD-delta method, I applied the delta-adjustment to both control group and experimental group. A future work is to apply delta-adjustment to the subjects in experimental group only.

Reference

Birmingham, J., Rotnitzky, A., Fitzmaurice, G. M. (2003). Pattern-mixture and selection models for analyzing longitudinal data with monotone missing patterns. *Journal of the Royal Statistical Society. B* 65:275–297.

Cook, R. D. (1986). Assessment of local influence. *Journal of the Royal Statistical Society. B* 48:133–169.

Daniels, M. and Hogan, J. (2008). *Missing Data in Longitudinal Studies: Strategies for Bayesian Modeling and Sensitivity Analysis*. CRC Press/Chapman and Hall, New York.

Diggle, P. and Kenward, M. G. (1994). Informative drop-out in longitudinal data analysis. *Applied Statistics* 43, 49-73.

European Medicines Agency. *Guideline on missing data in confirmatory clinical trials*, EMA/CPMP/EWP/1776/99 Rev. 1, 2 July 2010. http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2010/09/WC500096793.pdf (accessed 21December 2012).

Heckman, J. (1976). The common structure of statistical models of truncation, sample selection and limited dependent variables, and a simple estimator for such models. *Annals of Economic and Social Measurement, 5*, 475-492.

Horton, N. J. and Lipsitz, S. R. (2001), Multiple Imputation in Practice: Comparison of Software Packages for Regression Models with Missing Variables. *The American Statistician*, 55, 244–254.

Kenward, M. G. (1998). Selection Models for Repeated Measurements with Nonrandom Dropout: An Illustration of Sensitivity. *Statistics in Medicine*, 17, 2, 723-2732.

Kenward, M., Molenberghs, G. and Thijs, H. (2003). Pattern mixture models with proper time dependence. *Biometrika*, 90, 53–71.

Kenward, M.G., and Carpenter, J.R. (2008). Multiple Imputation In Longitudinal Data Analysis, New York: CRC Press.

Lachin J. M. (2000). Statistical Considerations in the Intent-to-Treat Principle. *Controlled Clinical Trials* 21 (3): 167–189.

Laird, N. M. and Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics*, 38, 963–974.

Liang, K. and Zeger, S. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, 73:13–22.

Little, R. J. A. (1988). A Test of Missing Completely at Random for Multivariate Data with Missing Values. *Journal of the American Statistical Association*, 83 (404) 1198-1202.

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

Little, R. J. A. (1994). A class of pattern-mixture models for normal incomplete data. *Biometrika*, 81, 471–483.

Little, R. (1995). Modeling the drop-out mechanism in repeated-measures studies. *Journal of the American Statistical Association*, 90, 1112–1121.
Little, R. J. and Wang, Y. (1996). Pattern-mixture models for multivariate incomplete data with covariates. *Biometrics*, 52, 98–111.

Little, R. J. and Rubin, D. B. (2002). *Statistical Analysis with Missing Data*. Wiley, New York.

Little, R. J., D'Agostino R., Cohen, M. L., Dickersin, K., Emerson, S. S., Farrar, J. T., Neaton, J. D., Shih, W. J., Siegel, J. P. and Stern, H. (2012). The prevention and treatment of missing data in clinical trials. *The New England Journal of Medicine*. October 2012.

Little, R. J., Cohen, M. L., Dickersin, K., Emerson, S. S., Farrar, J. T., Neaton, J. D., Shih, W. J., Siegel, J. P. and Stern, H. (2012). The design and conduct of clinical trials to limit missing data. *Statistics in Medicine*. July 2012. doi: 10.1002/sim.5519.

Lu K. (2014). An analytic method for the placebo-based pattern-mixture model. *Statistics in Medicine* 2014, 33 1134–1145.

Ma, G., Toxel, A., and Heitjan, D. (2005). An Index of Local Sensitivity to Nonignorable Drop-out in Longitudinal Modelling. *Statistics in Medicine*, 24, 2129 - 2150.

Merante, D., Sharma, U., Feins, K., Hsu, C. and Vinik, A. *Efficacy of mirogabalin (DS-5565) in diabetic peripheral neuropathic pain: pain assessment and correlations with a numeric rating scale and visual analog scale.* 74th Scientific Sessions of the American Diabetes Association, San Francisco, CA.

Molenberghs, G., Michiels, B. and Kenward, M. (1998a). Pseudo-likelihood for combined selection and pattern-mixture models for missing data problems. *Biometrical Journal*, 40, 557–572.

Molenberghs, G., Michiels, B., Kenward, M. and Diggle, P. J. (1998b). Missing data mechanism and pattern mixture models. *Statistica Neerlandica*, 52, 135–161.
Molenberghs, G., Thijs, H., Kenward, M. G. and Verbeke, G. (2003). Sensitivity analysis of continuous incomplete longitudinal outcomes. *Statistics Neerlandica*, 57, 112–135

Morris T. P., Kahan B. C., and White I. R. (2014). Choosing sensitivity analyses for randomized trials: principles. *BMC Medical Research Methodology* 2014, 14:11

National Research Council (2010). Panel on Handling Missing Data in Clinical Trials. Committee on National Statistics, Division of Behavioral and Social Sciences and Education. *The prevention and treatment of missing data in clinical trials*. The National Academies Press: Washington, DC.

Paik M. C. (1997) The Generalized Estimating Equation Approach When Data Are Not Missing Completely at Random. *Journal of the American Statistical Association* 92(440):1320-1329.

Ratitch B., O'Kelly M., and Tosiello R. (2013). Missing data in clinical trials: from clinical assumptions to statistical analysis using pattern mixture models. *Pharmaceutical Statistics*. 12(6):337-47.

Robins, J. M., Rotnitzky, A. and Zhao, L. P. (1995). Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *Journal of the American Statistical Association* 90:106–121.

Rotnitzky, A., Robins, J. M., and Scharfstein, D. O. (1998). Semiparametric regression for repeated outcomes with nonignorable nonresponse. *Journal of the American Statistical Association* 93, 1321-1339.

Rubin, D. (1974). Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies, *Journal of Educational Psychology*, 66 (5), 688–701.

Rubin, D. B. (1976). Inference and missing data. *Biometrika*, 63, 581–592.

Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*. John Wiley & Sons, New York.

SAS Institute Inc (2013). *SAS/STAT User's Guide - Procedures.* SAS Institute Inc., Cary, NC. URL http://support.sas.com/documentation/onlinedoc/stat/131/mi.pdf.

SAS (mixed effect model for repeated measures by Proc Mixed in SAS) version 9.3. SAS Institute, North Carolina.

SAS (Proc MI and Proc MIANALYZE) version 9.3. SAS Institute, North Carolina.

Schafer, J. L. (1997). *Analysis of incomplete multivariate data*. Chapman and Hall, New York.

Scharfstein D., McDermott A., Olson W., Wiegand F. (2014) Global Sensitivity Analysis for Repeated Measures Studies with Informative Drop-out: A Fully Parametric Approach. Unpublished manuscript.

Shih W. J. and Aisner J. (2014). *Statistical Design and Analysis of Clinical Trials – Principles and Methods.* CRC Press/Chapman and Hall, New York.

Shih, W. J. and Quan, H. (1997). Testing for Treatment Difference With Dropouts Present in Clinical Trials - A Composite Approach. *Statistics in Medicine,* Vol. 16, 1225-1239.

Troxel, A., Ma, G., and Heitjan, D. (2004). An Index of Local Sensitivity to Nonignorability, *Statistica Sinica*, 14, 1221-1237.

U.S. Department of Health and Human Services, Food and Drug Administration Center for Drug Evaluation and Research (CDER). *Guidance for Industry Analgesic Indications: Developing Drug and Biological Products*. February 2014

Verbeke, G., Molenberghs, G., Thijs, H., Lesaffre, E., Kenward, M.G. (2001). Sensitivity analysis for non-random dropout: a local influence approach. *Biometrics* 57:43–50.

Vinik, A., Sharma, U., Feins, K., Hsu, C. and Merante D. *A randomized, double-blind, placebo-and active comparator (pregabalin)-controlled adaptive phase 2 study of mirogabalin (DS-5565) for the treatment of diabetic peripheral neuropathic pain*. 74th Scientific Sessions of the American Diabetes Association, San Francisco, CA

Wei, L. and Shih, W. J. (2001). Partial Imputation Approach to Analysis of Repeated Measurements With Dependent Dropouts. *Statistics in Medicine,* Vol 20, 1197-1214.

Yuan Y. (2011). Multiple Imputation Using SAS Software. *Journal of Statistical Software*. December 2011.