DECAY MODEL FOR HANDLING MISSING DATA DUE TO INTERCURRENT EVENTS IN CLINICAL TRIALS

By

TAO SHENG

A dissertation submitted to the

School of Graduate Studies

Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

For the degree of

Doctor of Philosophy

Graduate Program in Public Health

Written under the direction of

Weichung Joe Shih, Ph.D.

And approved by

_____

_____

_____

_____

New Brunswick, New Jersey

January, 2019

ABSTRACT OF THE DISSERTATION

DECAY MODEL FOR HANDLING MISSING DATA DUE TO INTERCURRENT EVENTS IN CLINICAL TRIALS

by Tao Sheng

Dissertation Director: Weichung Joe Shih, Ph.D

US Food and Drug Association (FDA) presented a draft guidance of E9(R1) Statistical Principles for Clinical Trials Addendum: Estimands and Sensitivity Analysis in Clinical Trials in June 2017. This draft guidance has been widely referenced during recent research discussions. The aim of the draft guidance was to clarify the concept of estimand and to connect estimand with the concept of trial objective. An emphasized discussion about the impact and handling methods of missing data was also addressed. The draft guidance introduced the concept of 'intercurrent event' to describe all events that would cause either potential missing data or discontinuation from initial randomized treatment assignment. Five intercurrent event handling strategies were proposed with each strategy targeting a specific estimand which eventually represents a trial objective. Some of these strategies would result in missing data problem that required additional assumptions regarding the missing mechanism. In this dissertation, I will propose an alternative procedure in terms of connecting the

intercurrent event handling strategy with estimand specification. The proposed procedure can be considered as an event-type driven strategy that selects the desirable estimand based on not only primary trial objective but also potential intercurrent event types. In this dissertation, I will also discuss the importance of sensitivity analysis and the relationship between sensitivity analysis missing mechanism assumptions and primary analysis missing mechanism assumptions. Literature review will be focused on recent developments on the topic of sensitivity analysis methods, especially the reference based imputation (RBI) method and the δ-adjustment tipping point analysis method. The benefits and drawbacks of both methods will be discussed in detail. This dissertation will contain a proposal of a modified Mixed Model Repeated Measure (MMRM) model that targets the 'De Facto' estimand when rescue medication is offered in a randomized clinical trial. Primary estimator can be represented as a linear combination of this modified MMRM model parameters. Delta approximation method will be used to directly derive the inference of the estimator. The result and performance will be compared with the result using multiple imputation method. Secondly, I will propose an alternative sensitivity analysis method called 'decay model tipping point analysis method'. The highlights of this method are as follows, 1) It is capable of covering all possible sensitivity scenarios, including but not limited to the ones studied using RBI method. 2) The adjusted missing data effect is associated with dropout time. Patients who dropped out early will be adjusted with a greater value comparing to those who dropped out later. The adjustment decreases for time points that are further away from the patient dropout time point. This is a more reasonable approach comparing to δ-adjustment tipping point method which adjusts the effect at each time point with a same constant. 3) The range of the adjustment can be set within the a clinical meaningful boundary. This will avoid the over-adjustment problem in δ-adjustment tipping point method. 4) The decay rate parameter serves as a unified sensitivity parameter. It can be compared between

different studies as a measurement of robustness of primary analysis result in terms of the missing mechanism assumption. 5) The tipping point can be solved directly without iterative searching the total domain sensitivity parameter, therefore saving computing resource and power. Simulation studies will be conducted to verify the modified MMRM model. Inference derived based on delta approximation method will be verified using empirical inference result from simulation. A simulated study will be presented to verify the direct calculation for the tipping point and demonstrate the features of decay model tipping point method. In addition, a case study of using the decay model sensitivity analysis in a real world rare blood disease trial study will be presented. The possibility of extending the decay model beyond continuous endpoint will be briefly discussed and some technical issues occurred in the current research will be included in future research plan.

Acknowledgements

This work would not have been possible without the guidance of all the committee members and the support from my family.

I would like to express my deepest gratitude and appreciation for my committee chair Dr. Weichung Shih for the patient guidance, encouragement and advice he has provided throughout my time as his student. He has taught me more than I could give him credit for. By his example, I have learned what a good scientist and person should be. I would like to thank Dr. Yong Lin, who provided great suggestions on the modeling strategies and computing resources for my simulation work. I am also grateful to Dr. Kitaw Demissie and Dr. Hui Quan for being my committee members and providing me insightful suggestions.

I would like to thank my loving wife, Mengjue who believed in me and supported me during the years of my Ph.D study. I would also like to thank my parents for their constant encouragement.

Dedication

To my wife Mengjue, my parents and Pancake.

Table of Contents

List of Tables

List of Figures

Chapter 1
Introduction and Overview

1.1 Missing data and estimand problem

In a randomized clinical trial, patients are randomly assigned to different treatment arms. The difference of mean effects among treatment arms at the end of study period is usually of interest. In an ideal world, all patients will take the initial randomized assigned treatment following the protocol specifications. However, in a real world, events that break the initial randomization are usually inevitable. These events include but not limited to patients discontinued treatment due to the cure of the disease, patients discontinued treatment due to lack of efficacy, patients switched to rescue medication due to adverse events or patients deceased due to study or non-study related causes. In the E9 addendum Guidance [85] drafted by FDA in Jun 2017 (referred to as 'the guidance'), the terminology of 'intercurrent event' was introduced to describe these type of events. The existence of the intercurrent events would introduce ambiguity to the measurement of treatment effect. One motivating example to illustrate this problem is a diabetic drug trial sponsored by BMS and AstraZeneca[2]. The trial was designed to investigate the effect of studied drug (dapagliflozin) in terms of the glycated haemoglobin (HbA1C) change from baseline at week 24 comparing to reference arm. During the study, patients were allowed to take rescue medication if HbA1C measurement exceeded pre-specified safety threshold. During the statistical analysis step for this study, the measurements collected after patients discontinued from initial assignment and switched to rescue medication were considered as missing

and no longer included in the primary analysis. At the time of the study, the 'last observed value carried forward (LOCF)' method was a standard approach to handle missing data as sensitivity analysis. That is, the last measurement value before the rescue medication being offered would be carried forward to the end of study at week 24 as the end-of-trial value. The FDA reviewer questioned the sponsor's approach and pointed out that instead of ignoring the measurements collected during rescue medication period, a more preferred analysis would be to conduct the primary analysis using the observed value of patients who switched to rescue medication. FDA and the sponsor targeted two different scientific questions of interests regarding the treatment effect. From sponsor's perspective, the targeted question was 'What is the treatment effect when no rescue medication is offered'. From FDA's perspective, the targeted question was 'What is the treatment effect when rescue medication is allowed and in fact used'. The two different targeted questions resulted in different practices in terms of how to handle the measurements post intercurrent events. In order to clarify this, the concept of 'estimand' was introduced to mathematically describe the study scientific question of interest. Carpenter used the terms 'De Jure' estimand to address the 'what if no rescue medication offered' question and 'De Facto' estimand to address the 'what if rescue medication is offered' question which is the one that FDA was more concerned about [8]. The guidance went one step further and clarified the scientific questions of interest, the corresponding estimands as well as the intercurrent event strategies to address each individual estimand. Five intercurrent event strategies were introduced in the guidance and will be discussed in detail in this dissertation.

1.2 Sensitivity analysis

The intercurrent event strategies introduced in the guidance specified the post intercurrent event measurement handling methods during the design stage. However, theses strategies all depend on hypothetical assumptions regarding the mechanism of missing measurements. For example, in the previously mentioned dapagliflozin study, the primary analysis could have been conducted by including the switch to rescue medication patient's measurements and targeted the 'De Facto' estimand. However, for those who dropped out of the study without taking rescue medication, missing mechanism assumptions still needed to be made in order to include the dropout effect into account. If the primary analysis result was significant, one needed to bear in mind that the result was only significant in terms of the assumptions. Therefore, a sensitivity analysis should always be conducted regardless of the choice of primary analysis estimand as long as there were missing data in the study and the primary analysis result was significant. The purpose of the sensitivity analysis is to access the robustness of the statistical significance of the primary analysis result. Note that there are other assumptions besides the missing mechanism assumptions made during the primary analysis, for example, the normality assumption or independent and identically distributed assumption regarding the variable distribution. The sensitivity analysis which targets these assumptions is out of the scope of this dissertation. Usually, assumptions that would decrease the significance of the primary analysis will be made during sensitivity analysis. One well accepted sensitivity analysis approach is the 'reference based imputation' method, which borrows the reference arm observed measurements information to impute active arm missing measurements. If the primary scientific question of interest is the treatment effect difference between two arms, by using reference arm information to impute active arm missing measurements, one is making a relatively conservative assumption that there is no difference

between active arm and reference arm dropout patient treatment effects. However, the downside of this method is the limited number of discrete scenarios that can be checked by using different reference arm information borrowing methods. In addition, there is no continuous sensitivity parameter that unifies the scenarios in terms of the level of deviation from primary assumption. Another widely implemented method is 'δ-adjustment' tipping point sensitivity analysis method. A constant value of δ is adjusted from the hypothetical measurement under primary assumption. The δ is increased continuously until it reaches the point that the adjusted analysis result is no longer statistically significant. The advantage of this method is that δ as a sensitivity parameter covers a wide range of possible sensitivity assumptions. However, the magnitude of δ depends on the scale of the measurement. Therefore, the interpretation of δ is not clear by itself. One cannot determine the level of robustness of primary analysis by only looking at the tipping point δ but needs to compare with measurement scale. Another trade off is that one might be over adjusting the measurement to a clinically unreasonable range. For example, during the dapagliflozin study, the HbA1c measurement might be adjusted to negative, or to a value that is unrealistically high and biologically impossible, in order to reach the statistical significance tipping point.

1.3 Objective and dissertation outline

This dissertation has three main objectives. First is to clarify the concept of estimand and to clarify the study question of interest that the estimand is aiming. The second objective is to modify the existing 'Mixed Model Repeated Measure' (MMRM) model to include rescue medication arm measurements if rescue medication is allowed in the protocol. The third objective is to introduce a new sensitivity analysis method which is inspired by the 'δ-adjustment' tipping point method. The remaining chapters are

organized as follows. In chapter 2, I will review current literature within the topic of missing data handling methods and the concept of estimand. In chapter 3, I will introduce a modified MMRM modeling approach to derive the estimator for 'De Facto' estimand when rescue medication is allowed in a clinical trial. The reference based imputation sensitivity analysis estimators will also be represented as functions of the modified MMRM parameters. In chapter 4, the delta approximation method would be implemented to analytically derive the primary analysis point estimator as well as the reference based imputation sensitivity analysis point estimators and their inferences. Simulation study will be conducted to verify the derivation using delta approximation method as well as to compare with the performance of conventional multiple imputation approach. In chapter 5, a modified tipping point analysis which uses an exponential decay model to represent the deviate effect from primary analysis assumption will be explained in detail. I will also derive the representation of various reference based imputation scenarios as functions of decay parameter. In chapter 5, I will include a case study which uses the decay model sensitivity analysis method to access the robustness of the significant primary analysis result in a rare blood disease trial study. The dissertation will be summarized with a discussion and conclusion section in chapter 6. Further research topics would also be briefly mentioned.

Chapter 2

Literature review

2.1 Missing data classification

The missing data problem was first studied by Rubin in 1976 [64]. In his paper, Rubin introduced the cornerstone for future missing data research which is the classification of missing data by their missing mechanism. It has been widely recognized and implemented since then. Use $Y_{i,t}$ to denote measurement for patient i at time point t, use $\boldsymbol{Y}_{i,obs}$ to denote observed measurements for patient i prior to time point t and $M_{i,t}$ as missing indicator, $M_{i,t} = 1$ to denote measurement missing for patient i at time point t and $M_{i,t} = 0$ to denote measurement observed for patient i at time point t. The three categories that Rubin used can be presented as follow,

- Missing completely at random (MCAR): The probability of data being missing does not depend on the observed or unobserved data. $P(M_{i,t} = 1|\boldsymbol{y}_{i,obs}, y_{i,t}) = P(M_{i,t} = 1)$. For example, the data is missing because of a typo during the record input. Another example would be that a patient dropped out of the study due to relocation.

- Missing at random (MAR): The probability of data being missing does not depend on the unobserved data but condition on the previous observed data. $P(M_{i,t} = 1|\boldsymbol{y}_{i,obs}, y_{i,t}) = P(M_{i,t} = 1|\boldsymbol{y}_{i,obs})$. For example, a patient dropped out of the study because of lack of efficacy based on previous measurements. This assumption implies that the missing data can be predicted from the observed

variables.

- Missing not at random (MNAR): The probability of data being missing depends on unobserved as well as observed data. $P(M_{i,t} = 1|\boldsymbol{y}_{i,obs}, y_{i,t})$ does not simplify. For example, patients drops out of the study due to lack of efficacy. The missing data is supposed to be worse than the observed measurements. In this scenario, the value of the unobserved responses depends on information not available for the analysis. Missing observations cannot be predicted without further assumptions [92].

The missingness under MAR and MCAR are considered ignorable missing because the 'true' underlying effect can be estimated without bias using observed data. The missingness under MNAR is considered non-ignorable missing because it is not possible to derive an unbiased estimator under this scenario. The key point of the missing data classification is that the underlying missing mechanism is essentially untestable. Therefore, a conventional way is to use modeling approach to adjust for all possible covariates so that the adjusted response will be more likely to satisfy the MAR/MCAR assumption and becomes ignorable.

2.2 Imputation Methods

One way to handle missing data is to impute the missing data with hypothetical values. There are two types of imputation approaches, single imputation and multiple imputation. Single imputation methods such as 'Last Observation Carried Forward' (LOCF), 'Baseline Observation Carried Forward' (BOCF) have long been the default missing data handling methods. However, there is a potential disadvantage of single imputation method which is that these methods do not take the variance caused by missingness into account. Hence, the risk of underestimating the variance of

treatment effect when using single imputation has been criticized and alternative methods based on multiple imputation have been recommended. Multiple imputation methods generate multiple copies of original data set by replacing missing values based on appropriate imputation models. For each completed copy of data set, a pre-specified analysis model is implemented to provide final parameter estimates. Combine the different parameter estimates across the imputation copies to produce a unique point estimate and standard error taking into account the uncertainty of the imputation process. The three stages of multiple imputation are described in detail as below:

- Step 1: Generate multiple imputed complete datasets: For a single incomplete variable z, model z using the observed portion of dataset. Denote the estimated model coefficients $\hat{\boldsymbol{\beta}}$ and residual variance $\boldsymbol{V}$ have their corresponding variance covariance matrix. Draw a set of $\tilde{\boldsymbol{\beta}}$ from the posterior distribution approximated by $\tilde{\boldsymbol{\beta}} \sim MVN(\hat{\boldsymbol{\beta}}, \boldsymbol{\Sigma})$. Impute z by generating z^* based on $\tilde{\boldsymbol{\beta}}, \boldsymbol{V}$ and the appropriate probability distribution. This is referred to as the imputation model.

- Step 2: Analyze imputed data sets: Each imputed complete dataset is analyzed separately using analysis model. The desirable parameters that address the study objectives, denoted as $\hat{\boldsymbol{\theta}}_j$, are estimated from each imputed dataset, together with their variance covariance matrices, $\hat{\boldsymbol{W}}_j$, for the jth imputed copy of dataset.

- Step 3: Combine estimates from imputed datasets: Suppose total of m copies of imputed datasets are generated. Estimates are combined into an overall estimate and variance estimate using Rubin's rules. The combined estimate $\hat{\boldsymbol{\theta}}$ is the average of the individual estimates:

$$\hat{\boldsymbol{\theta}} = \frac{1}{m} \sum_{j=1}^{m} \hat{\boldsymbol{\theta}}_j \tag{2.1}$$

The total variance of $\hat{\boldsymbol{\theta}}$ is the sum of the within-imputation variance $\boldsymbol{W} = \frac{1}{m}\sum_{j=1}^{m}\hat{\boldsymbol{W}}_j$ and the between-imputation variance $\boldsymbol{B} = \left(\frac{1}{m-1}\right)\sum_{j=1}^{m}(\hat{\boldsymbol{\theta}}_j - \hat{\boldsymbol{\theta}})^2$;

$$var(\hat{\boldsymbol{\theta}}) = \boldsymbol{W} + (1 + \frac{1}{m})\boldsymbol{B}. \tag{2.2}$$

2.3 Mixed model repeated measure

One recommended model for imputation and analysis is the 'Mixed Model Repeated Measure' (MMRM) model introduced by Mallinckrodt [51]. The MMRM model is a modified mixed effect model. The general form of the model is as follow

$$Y = baseline + treatment + time + treatment \times time + baseline \times time + error \tag{2.3}$$

where Y is the observed measurement, *baseline* is the collection of baseline covariates, *treatment* is the initial treatment arm indicator, *time* is the indicator of the time point that Y is observed. This model has the advantage of directly estimating the treatment effect at each measurement time point adjusting for baseline covariates. The model uses unstructured covariance matrix to model within-subject errors. The model parameters are estimated using restricted maximum likelihood (ReML) estimation method.

2.4 Concept of Estimand

Most of the recent literature have been focusing on estimating methods when there are missing data involved, despite the fact that a more essential question should be regarding what to estimate. The guidance provided some valuable ideas in terms of how to systematically process the clinical trial with the missing data problem considered during the design stage. It proposed a framework of clinical study including design, conduct, analysis and interpretation (Figure 2.1). According to this flow chart, a suitable estimand corresponding to the key scientific question of interest should be

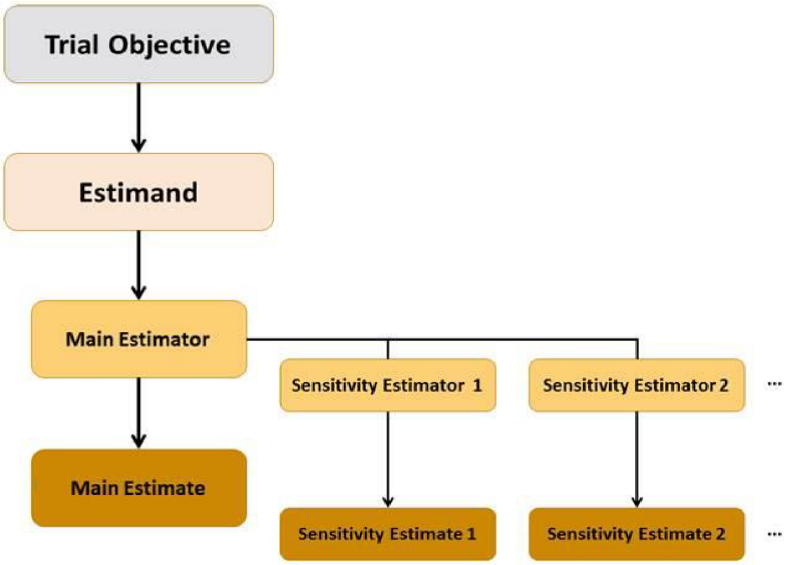

Figure 2.1: Flow chart for clinical trial design proposed by FDA in E9(R1) Statistical Principles for Clinical Trials Addendum

clearly defined at the beginning of a clinical trial. A main estimator targeting the estimand could be then selected based on certain assumptions. A sensitivity analysis should be conducted to investigate the main estimator robustness in terms of the assumptions. The important concept was to process this sequentially and 'not for the choice of an estimator to determine the estimand'[85]. The guidance described four attributes of an estimand including 1) The targeted population; 2) Patient variable to address the research question; 3) Strategies for addressing intercurrent events and 4) Population-level summary for the variable of interest. Among the four attributes, 2 and 3 were the ones that needed detail specification in order to reflect different scientific question of interest. For attribute 2, there had been a discussion whether the variable needed to be specifically defined prior to conduct any statistical analysis. However, in many cases, the specific variable that the estimand was targeting would be determined only when the statistical model was determined. A recommended practice was that the investigator identified the clinical issue that the study planned to target and formulated the clinical question using statistical model. After

the statistical estimand was clearly stated, the clinical question of interest of study could be updated with a more accurate description. For attribute 3, five strategies were introduced in detail by the guidance to address intercurrent events including 'Treatment policy strategy', 'Composite strategy', 'Hypothetical strategy', 'Principal stratum strategy' and 'While on treatment strategy'. Notice the intercurrent event definition in the guidance was broader than the missing data concept. Measurements after intercurrent events happened did not necessarily have to be missing. Thus the process of constructing estimand should be exclusive from the concept of missing data. The missing data problem should be considered during the step of developing estimator.

2.5 'De Jure' estimand and 'De Facto' estimand

Carpenter, Roger and Kenward introduced the concept of 'De Facto' estimand and 'De Jure' estimand in their paper [8]. The 'De Jure' estimand was defined as 'What would the expected treatment effect be in the target population of eligible patients (as defined by the trial inclusion criteria) if the treatment and control were taken as specified in the protocol'. The 'De Facto' estimand is described as 'What would be the effect seen in practice if this treatment were assigned to the target population of eligible patients, as defined by the trial inclusion criteria'. For example, one might want to conduct an efficacy trial which studies the drug effect in an ideal world that all patients would follow initial randomization. This would be corresponding to the 'De Jure' estimand. On the other hand, an effectiveness trial also might be of interest to study the drug effect in a real world clinical setting that patients might drop from the study or switch to rescue medication. The effect is therefore a mixture of initial assignment effect and rescue medication effect. This is referred to as the 'De Facto' estimand. The estimands corresponding to the five strategies described in the

guidance can be viewed as detailed specifications within the concept of 'De Jure' and 'De Facto'.

2.6 Five strategies for addressing intercurrent events

- **Principle stratum strategy** Under this strategy, the scientific question of interest is 'What would the drug effect be if patients stay on treatments assigned by initial randomization'. In order to fully implement this strategy, only those patients who do not have intercurrent events would be included in the data analysis. In other word, this strategy constrains the scientific question of interest to be related with only a subset of the population who would be expected not to experience an intercurrent event. Although this question is straightforward, the identification of the subset population is almost impossible before starting a randomized clinical trial. The estimand that this strategy addresses is a 'De Jure' estimand since it is targeting a scenario that happens in an ideal world.

- **While on treatment strategy** The scientific question under such strategy is that 'What would the drug effect be if the last on treatment effect can be maintained to the end of study after the patient discontinues the treatment'. This strategy is similar to the 'Last observation carried forward (LOCF)' approach. The last measurement prior to the occurrence of intercurrent event is included in the final data analysis. This estimand is a 'De Jure' estimand due to the fact that it does not take the possible measurements that would have been collected post intercurrent event into account.

- **Treatment-policy strategy** Under this strategy, the scientific question is that 'What is the treatment effect regardless of the occurrence of intercurrent events'. This is similar to the 'ITT principle (Intent to treat)'. This estimand is generally acceptable to support regulatory decision making. It can be viewed as a 'De

Facto' estimand.

- **Composite strategy** The scientific question under such strategy is that 'What is the treatment effect if the intercurrent event is part of the endpoint itself'. The occurrence of intercurrent event itself is more interested than the missing measurements post event. In fact, consider the case that the intercurrent event is death and the measurement is blood pressure, the measurements post event is not missing but rather not exist at all. I am not going to include this strategy in the discussion because this requires another set of analysis methods and is beyond the scope of missing data problem.

- **Hypothetical strategy** This strategy is of most interest. The scientific question of interest is 'What would the treatment effect be if patients drop out from the treatment group would switch to a specific hypothetical scenario'. The 'specific hypothetical scenario' has the flexibility of covering a wide range of different possible hypothetical situations. One might assume the patients would be offered another reference drug, i.e, an active control or rescue medication. Or one might assume the patients would maintain the drug effect or the drug effect would totally vanish. Depending on the different hypothesis, this strategy may address either a 'De Jure' estimand or a 'De Facto' estimand.

2.7 Propose an alternative approach to determine the estimand

The 11th Annual Conference on Statistical Issues in Clinical Trials was held at University of Pennsylvania in April 2018. During the meeting, experts from industry and academia were invited to discuss the draft guidance. Major comments were that investigators considered the five strategies introduced in the guidance not detailed enough to fit all possible situations in real world studies. The ambiguity of the concept of

'Intercurrent event strategy' was criticized by Daniel Scharfstein. He proposed the alternative definition using 'outcome' instead of 'intercurrent event strategy' attribute

- ITT ('De Facto'): outcome regardless of adherence to the treatment strategy
- Composite: defined to include the occurrence of key post-randomization event(s)
- Counterfactual ('De Jure'): outcome under full adherence to the treatment strategy
- While Adherent (While on treatment): outcome during adherence to the treatment strategy

Taking the discussion during the annual conference into consideration, I propose an alternative approach to understand and determine an appropriate intercurrent event strategy. In Figure 2.2, a flow chart of detail steps of choosing estimand and corresponding intercurrent event handling methods is illustrated. The first decision to be made in terms of intercurrent events is whether the post intercurrent event measurements are indeed collectable or not. The post intercurrent event measurements should only be considered as missing if they are actually observable. For example, a patient's biomarkers are observable if the patient is not deceased during study period. Any post intercurrent event measurement is not observable if death is the event itself. Assumptions for unobservable post intercurrent event measurements should not be made. Instead, the event can be considered as a binary response and use time-to-event model such as survival model to integrate the intercurrent event information into the primary analysis. This is referred to as the 'Composite estimand'. Another approach is that use only the measurements collected on initial assignment period to target the 'While on treatment estimand'.

Suppose that the post intercurrent event measurements do exist. However, due to the reasons that can not be controlled that cause the measurements on initial randomized

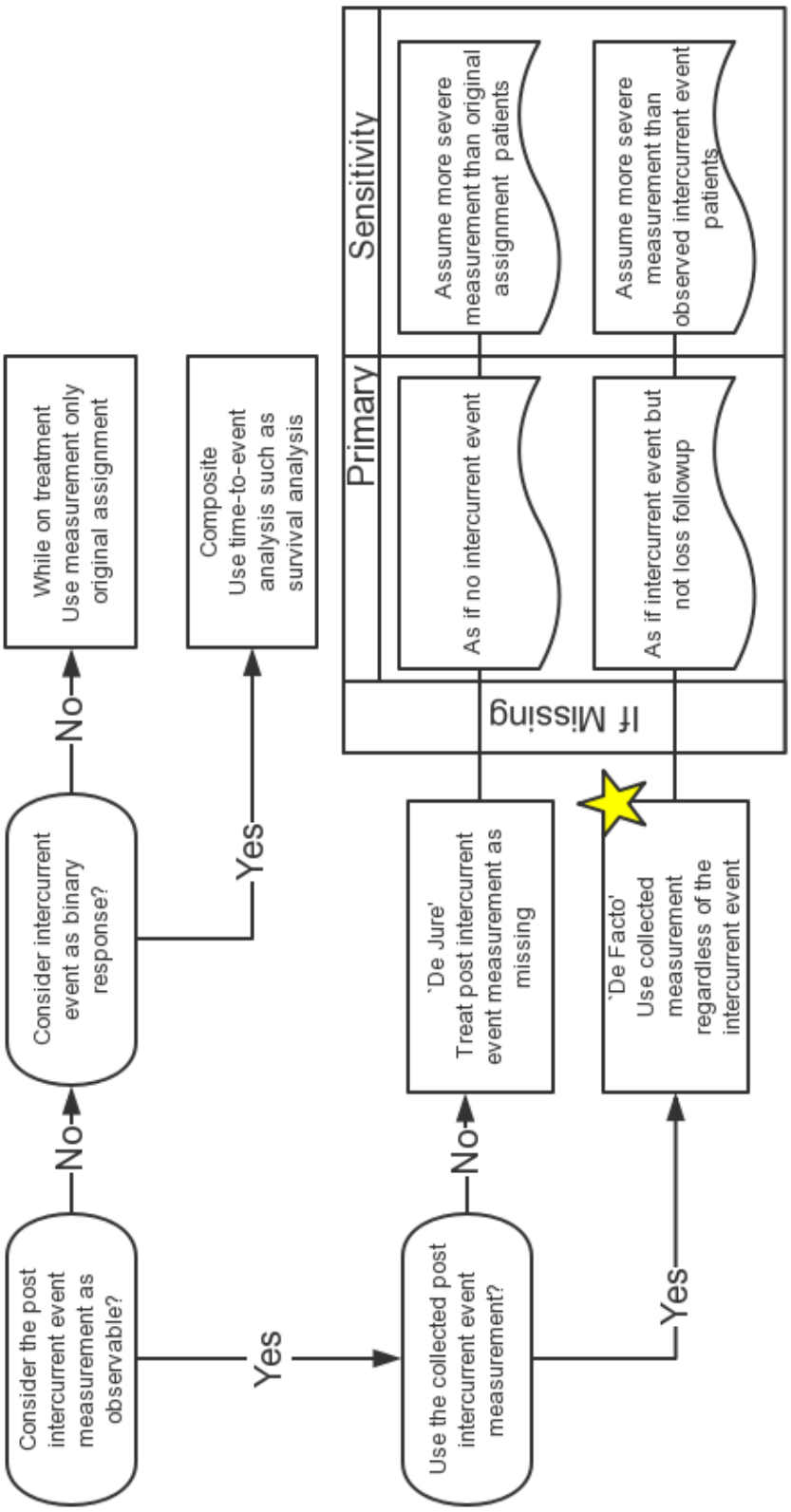

Figure 2.2: Propose alternative approach to determine an appropriate estimand

assignment not observed. For example, patients might discontinue taking initial assigned treatment due to lack of efficacy or adverse event. Under these circumstances, it is unethical to force patients to continue on initial assigned treatment. Alternatively, rescue medication might be allowed. Efforts should be made to collect the measurements on rescue medication period. If the investigator's scientist question of interest is to study the treatment effect based on initial randomized treatment only, then regardless the rescue medication period measurements are collected or not, the post intercurrent event measurements are considered as missing. This is referred to as the 'De Jure' estimand. The primary analysis assumption for the 'De Jure' estimand is to assume the missing mechanism is MAR. The sensitivity analysis should therefore target the MAR assumption. If the investigator's scientist question of interest is to study the effect where rescue medication will either be offered as a substitution of initial assigned treatment or as a combination with the initial assigned treatment, the 'De Facto estimand' is preferred. Post intercurrent event measurements would be included as-is during the primary analysis. If there are still patients measurements missing, assumptions might be made that they would have been offered rescue medication. Thus the primary analysis assumption for missing data under 'De Facto' estimand is that the dropout patients are MAR from the rescue medication arm. Thus the sensitivity analysis should be conducted to target the MAR from rescue medication assumption.

2.8 Reference based imputation (RBI)

Carpenter introduced a set of different sensitivity analysis scenarios which assumes reference arm patients are MAR and borrows reference arm mean treatment effect information to impute active arm missing values. I will demonstrate three borrowing

methods in detail, namely the 'Jump to reference' (J2R) method, the 'Copy increment in reference' (CIR) method and the 'Copy reference' (CR) method through a diabetic trial example. In all three methods, the reference arm missing patients are imputed under MAR assumption.

Consider a randomized clinical trial with two arms, active arm and reference arm. Suppose the repeated measurements $\boldsymbol{y} = (y_1, ..., y_t)'$ follow a multivariate normal distribution, $\boldsymbol{y} \sim MVN(\boldsymbol{\mu}, \Sigma)$ where Σ is the same variance covariance matrix for both active and reference groups. Let $\boldsymbol{\mu}^d = (\mu_1^d, \mu_2^d, ..., \mu_t^d)$ and $\boldsymbol{\mu}^p = (\mu_1^p, \mu_2^p, ..., \mu_t^p)$ represent mean vectors for active arm and reference arm respectively. Use $\boldsymbol{\mu^p} = [\boldsymbol{\mu_o^p}, \boldsymbol{\mu_m^p}]$ to distinguish the observed portion and missing portion reference arm mean effect. Use $\boldsymbol{\mu^d} = [\boldsymbol{\mu_o^d}, \boldsymbol{\mu_m^d}]$ to distinguish the observed portion and missing portion active arm mean effect. Figure 2.3 is a plot of two arm means HbA1c change values from baseline (Week 0) to Week 24. Assume the observations from different time points follow multivariate normal distribution. The dotted line represents the mean vector of the multivariate normal distribution. The solid triangles represent observed HbA1c values for one specific patient at baseline (Week 0), Week 4 and Week 8. The observations at Week 16 and Week 24 are missing. Use $\boldsymbol{y_o}$ to denote observed measurements and $\boldsymbol{y_m}$ to denote post intercurrent event measurements. The observed measurements together with missing measurements jointly follow multivariate normal distribution $\begin{bmatrix} \boldsymbol{y_o} \\ \boldsymbol{y_m} \end{bmatrix} \sim N\left(\begin{bmatrix} \boldsymbol{\mu_o^p} \\ \boldsymbol{\mu_m^p} \end{bmatrix}, \boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{oo} & \boldsymbol{\Sigma}_{om} \\ \boldsymbol{\Sigma}_{mo} & \boldsymbol{\Sigma}_{mm} \end{bmatrix} \right)$ for reference arm patients; $\begin{bmatrix} \boldsymbol{y_o} \\ \boldsymbol{y_m} \end{bmatrix} \sim N\left(\begin{bmatrix} \boldsymbol{\mu_o^d} \\ \boldsymbol{\mu_m^d} \end{bmatrix}, \boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{oo} & \boldsymbol{\Sigma}_{om} \\ \boldsymbol{\Sigma}_{mo} & \boldsymbol{\Sigma}_{mm} \end{bmatrix} \right)$ for active arm patients. The open triangles are the imputed HbA1c values drawn from conditional multivariate normal distribution $E(\boldsymbol{y_m}|\boldsymbol{y_o}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \boldsymbol{\mu_m^d} + \boldsymbol{\Sigma}_{mo}\boldsymbol{\Sigma}_{oo}^{-1}(\boldsymbol{y_o} - \boldsymbol{\mu_o^d})$. Suppose the primary analysis is under MAR assumption, the mean vector is assumed to be the same as active arm mean, that is, the dotted line overlaps with active arm mean curve. However, since

the patient's observed 3 measurements are lower than the population mean curve, the missing data will be imputed from a normal distribution with a conditional mean that is also lower than population mean.

- Jump to reference: Patient's post deviation mean response distribution is same as a reference group. This might be used as a worst-case scenario in terms of reducing any treatment effect since withdrawn patients on active will lose the effect of their period on treatment. Figure 2.4 indicates that missing values are imputed with the reference arm means. The conditional multivariate normal distribution mean is $E(\boldsymbol{y_m}|\boldsymbol{y_o}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \boldsymbol{\mu_m^p} + \boldsymbol{\Sigma}_{mo}\boldsymbol{\Sigma}_{oo}^{-1}(\boldsymbol{y_o} - \boldsymbol{\mu_o^d})$.

- Copy increments in reference: Patient's post deviation mean increments are the same as those from the reference group. In Figure 2.5, the increment of reference arm mean effect between Week 16 and Week 8 as well as the increment of reference arm mean effect between Week 24 and Week 16 are borrowed to determine the active arm population mean at Week 16 and Week 24. Therefore, the population mean curve segments post dropout time point are parallel with the reference arm mean segments. The conditional multivariate normal distribution mean is $E(\boldsymbol{y_m}|\boldsymbol{y_o}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \boldsymbol{\mu_m^p} + (\boldsymbol{\mu_j^d} - \boldsymbol{\mu}_j^p) + \boldsymbol{\Sigma}_{mo}\boldsymbol{\Sigma}_{oo}^{-1}(\boldsymbol{y_o} - \boldsymbol{\mu_o^d})$.

- Copy reference: Patient's measurement distributions for all time points, both pre and post deviation, are assumed to be the same as the reference group. Figure 2.6 demonstrated the idea of 'Copy Reference'. The population mean for the observed patient is assumed to be as the reference arm from baseline to Week 24. Since the observed measurements are higher than reference arm mean curve, the conditional mean for the missing time points are higher than population mean. The conditional multivariate normal distribution mean is $E(\boldsymbol{y_m}|\boldsymbol{y_o}, \boldsymbol{\mu}, \hat{\boldsymbol{\Sigma}}) = \boldsymbol{\mu_m^p} + \boldsymbol{\Sigma}_{mo}\boldsymbol{\Sigma}_{oo}^{-1}(\boldsymbol{\mu_o^d} - \boldsymbol{\mu_o^p}) + \boldsymbol{\Sigma}_{mo}\boldsymbol{\Sigma}_{oo}^{-1}(\boldsymbol{y_o} - \boldsymbol{\mu_o^d})$.

Liu and Pang extended the idea in more detail. Using MMRM, assume the repeated measurements $\boldsymbol{y} = (y_1, ..., y_t)'$ follow a multivariate normal distribution, $\boldsymbol{y} \sim MVN(\boldsymbol{\mu}, \Sigma)$ where Σ is the same for both active and reference groups. Let $\boldsymbol{\mu}^d = (\mu_1^d, \mu_2^d, ..., \mu_t^d)$ and $\boldsymbol{\mu}^p = (\mu_1^p, \mu_2^p, ..., \mu_t^p)$ represent mean vectors for active and reference, then under MMRM the mean for the jth measurement of patient i is:

$$E(Y_{ij}) = \mu_j = \begin{cases} \boldsymbol{X}_i\boldsymbol{\beta}_j + \theta_j^p = \mu_j^p, & \text{for reference} \\ \boldsymbol{X}_i\boldsymbol{\beta}_j + \theta_j^d = \mu_j^d, & \text{for active} \end{cases} \tag{2.4}$$

where $\boldsymbol{x}_i$ is a collection of baseline covariates to be adjusted in the analysis model and $\boldsymbol{\beta}_j$ is a vector of corresponding coefficients for time point j. θ_j^p and θ_j^d are the adjusted treatment effects for each arm at each measurement time point. All the parameters can be estimated from a Bayesian method, i.e, Markov Chain Monte Carlo (MCMC). Assume non-informative prior distribution for parameters $\beta_j, \theta_j^p, \theta_j^d$ and an inverse Wishart prior for Σ, MCMC may sample parameters and the missing data iteratively. For a specific patient who dropped out at time $j+1$, the missing data vector will be sampled from a conditional distribution:

$$\boldsymbol{y}_{mis}|\boldsymbol{y}_{obs}, \boldsymbol{X}, \boldsymbol{\mu}, \Sigma \sim N(\boldsymbol{\mu} + \Sigma_{mo}\Sigma_{oo}^{-1}(\boldsymbol{y}_{obs} - \boldsymbol{\mu}_o), \Sigma_{mm} - \Sigma_{mo}\Sigma_{oo}^{-1}\Sigma_{om}) \tag{2.5}$$

Under RBI analysis, the conditional mean is different. Using the PMM approach, for the jth pattern, $\boldsymbol{y}_{obs} = (y_1, ..., y_j)'$, the missing data for treatment group will be specified by the RBI:

$$\boldsymbol{\mu}_m^d = \begin{cases} \boldsymbol{\mu}_m^p + \mu_j^d - \mu_j^p & \text{for CIR} \\ \boldsymbol{\mu}_m^p & \text{for JR} \\ \boldsymbol{\mu}_m^p + \Sigma_{mo}\Sigma_{oo}^{-1}(\boldsymbol{\mu}_o^d - \boldsymbol{\mu}_o^p) & \text{for CIR} \end{cases} \tag{2.6}$$

where $\boldsymbol{\mu}_m^d$ is the mean vector used in imputation for missing data at time point j. The overall treatment difference at the last time point under RBI is $\theta^{RBI} = \sum_{j=1}^t \pi_j \mu_{tj}^d - \mu_t^p$ where μ_{tj}^d is the conditional mean at the last time point t under missing pattern j

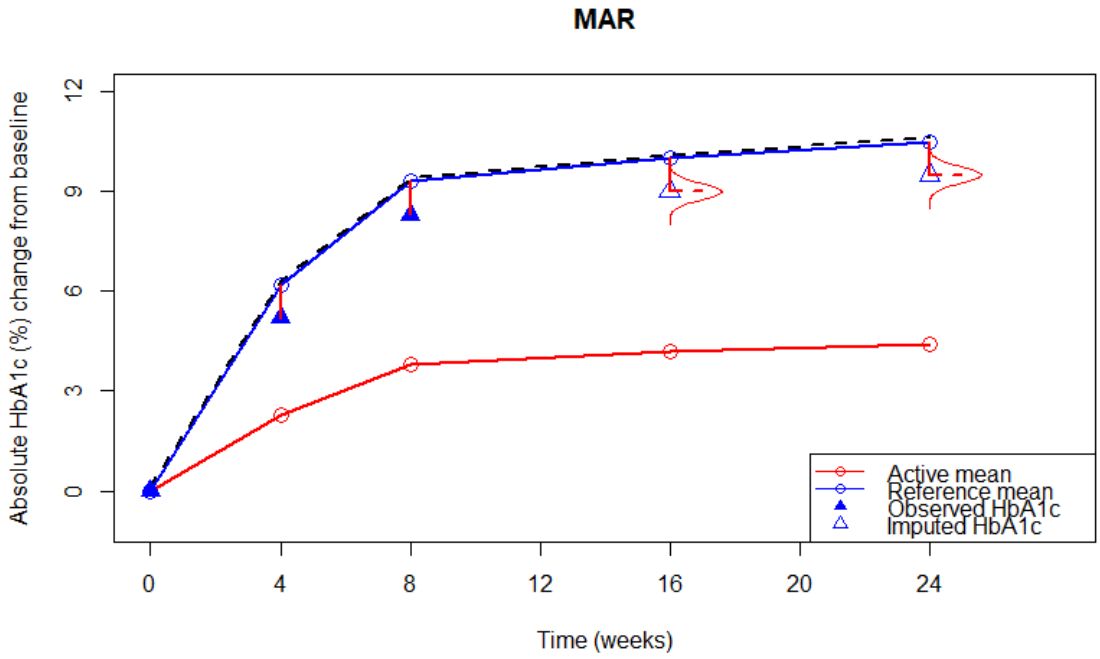

Figure 2.3: Multiple imputation model under primary analysis MAR assumption

and π_j is the proportion of patients in the missing data pattern j for the drug group. Based on (2.6), the difference can be written in the following format for different imputation methods

$$\theta^{RBI} = \begin{cases} \sum_{j=1}^{t} \pi_j(\mu_j^d - \mu_j^p) & \text{for CIR} \\ \pi_t(\mu_t^d - \mu_t^p) & \text{for JR} \\ \sum_{j=1}^{t} \pi_j(\mu_{tj}^d - \mu_t^p) & \text{for CIR} \end{cases} \tag{2.7}$$

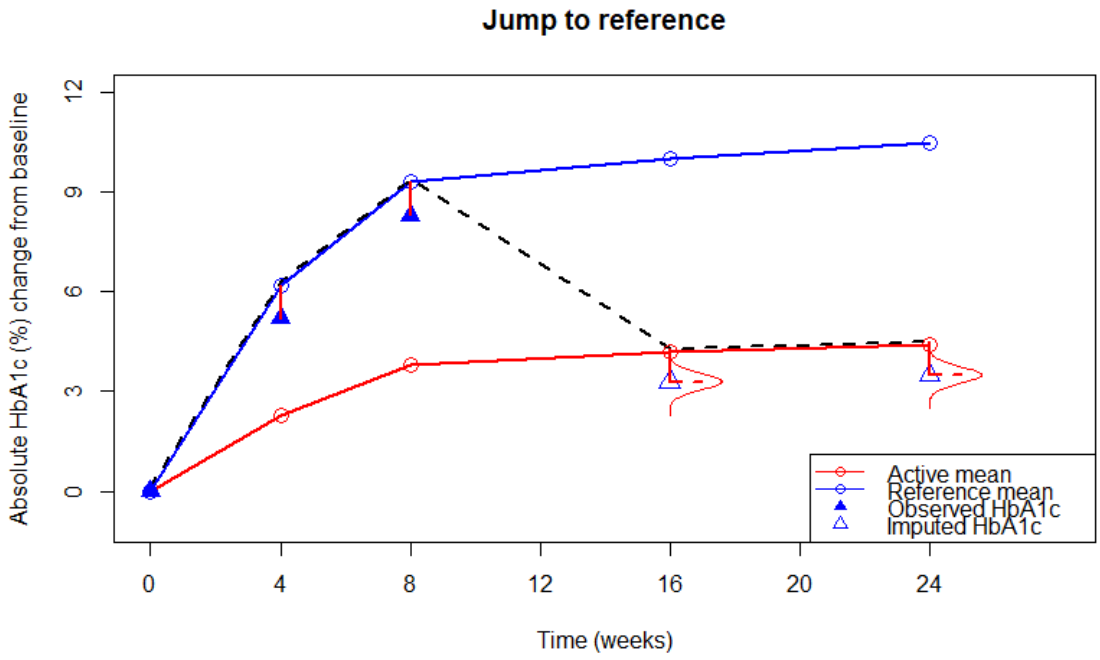

Figure 2.4: Multiple imputation model under sensitivity analysis J2R assumption

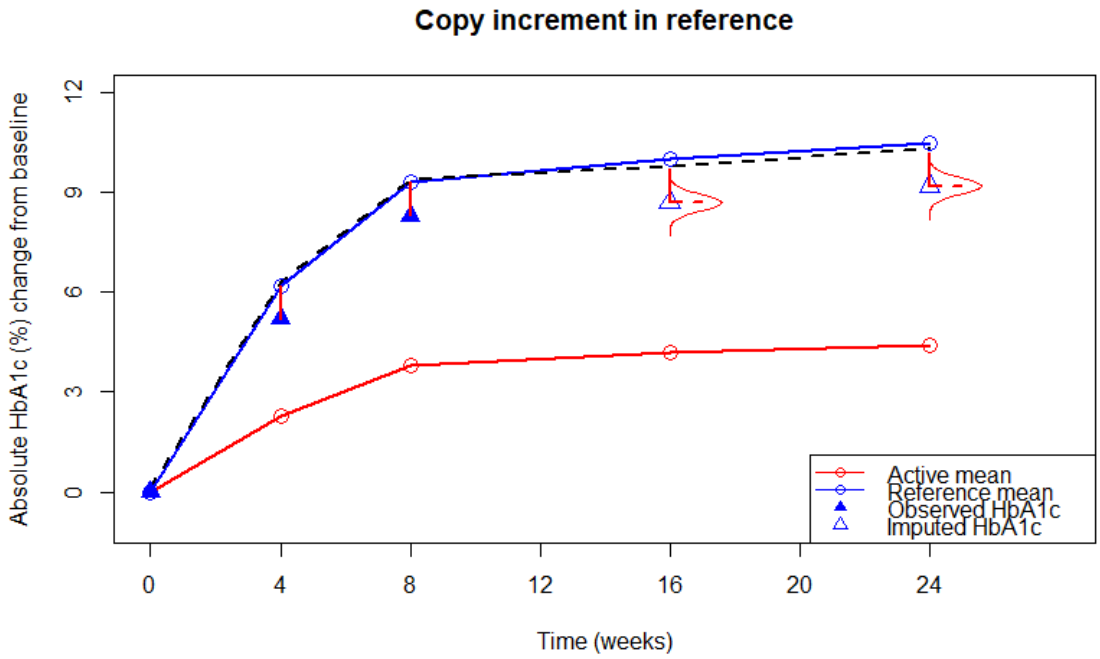

Figure 2.5: Multiple imputation model under sensitivity analysis CIR assumption

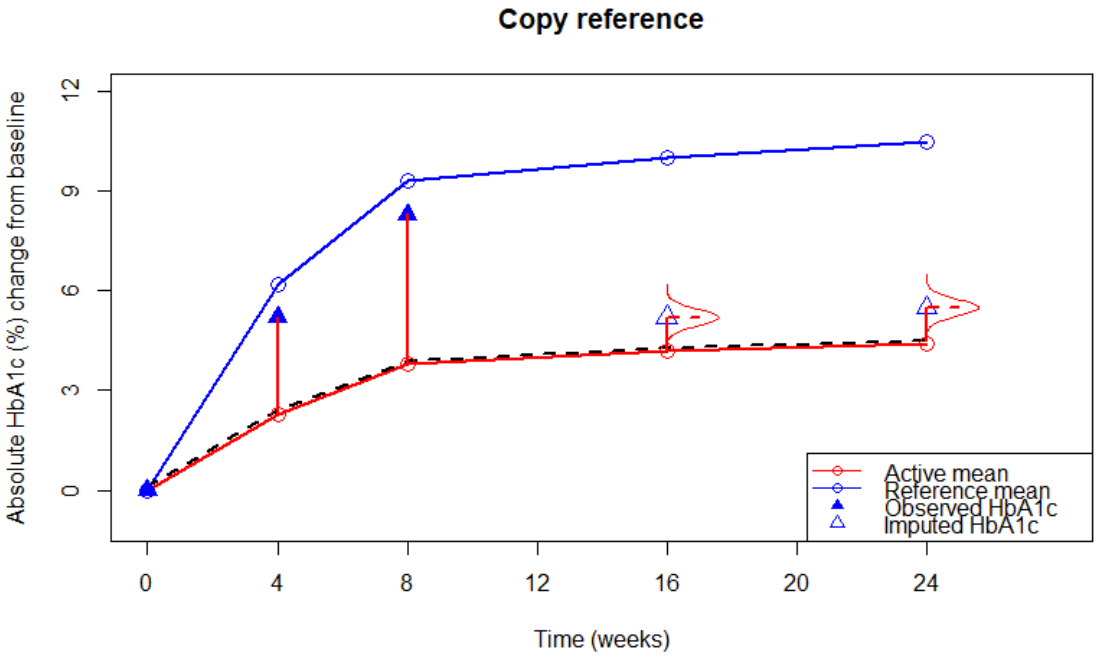

Figure 2.6: Multiple imputation model under sensitivity analysis CR assumption

Chapter 3

MMRM model with rescue medication allowed

3.1 Prelimilary

In a modern well designed clinical trial, rescue medications are usually available for patients who discontinued from initial assigned treatment. An effect difference of the mixture of the initial assigned treatments and rescue medication at the end of study period is often of interest. Suppose a clinical study is conducted with two arms and T post baseline measurement time points. Rescue medication is allowed after the first measurement time point t_1. Patients might either choose to take the rescue medication and stay in the study or drop out of the study completely. Assume the study question of interest is the effect between a studied treatment and a reference arm account for the effect when the rescue medication is offered, this can be considered as a 'De Facto' estimand. To statistically construct the 'De Facto' estimand, the following notation is used for the data structure.

3.2 Data structure and notation

In Table 3.1, the data structure for the treatment arm is presented. Total of T scheduled visit measurements collected at time point $t_1, ..., t_T$. Denote the measurement for the ith patient at t_j as $y_{i,j}$. Denote G_i the initial assignment indicator for patient i. $G_i = 1$ if the patient is randomized to treatment arm and $G_i = 0$ if the patient is randomized to reference arm. Denote S_i the intercurrent event indicator for patient i.

$S_i = 1$ if the patient stayed on the initial assignment to the end of the study. $S_i = 2$ if the patient discontinued initial assignment and switched to rescue medication prior to the end of the study. $S_i = 3$ if the patient discontinued initial assignment and dropped out of the study with no follow up measurement. E_i is used to denote the event happening time point. If $S_i = 1$, then $E_i = T$ indicates patient i stayed for the whole study period. If $S_i = 2$, $E_i = j$ indicates patient i switched to rescue medication at time point t_j. If $S_i = 3$, $E_i = j$ indicates patient i dropped out from the study at time point t_j and measurements afterwards are missing. Assume that for treatment arm, r patients remained on initial assignment, q_j patients switched to rescue medication at time point t_j with a total of $q = \sum_{j=2}^{T} q_j$ patients switched to rescue medication from treatment any time during the study. p_j patients dropped out of the study at time point t_j and missed the following measurements with a total of $p = \sum_{j=2}^{T} p_j$ patients dropped out any time during the study. For reference arm, l patients remained on initial assignment, m_j patients switched to rescue medication at time point t_j with a total of $m = \sum_{j=2}^{T} m_j$ patients switched to rescue medication from reference arm any time during the study. n_j patients dropped out of the study at time point t_j and missed the following measurements with a total of $n = \sum_{j=2}^{T} n_j$ patients dropped out any time during the study. The probability of an intercurrent event happened is by random. Denote the probability of staying on initial assignment for treatment arm patient as π_O (ω_O for reference arm. All notations in parentheses indicate reference arm probabilities). Denote the probability of discontinuation from initial assignment at t_j as π_{Sj} (ω_{Sj}). Denote the overall probability of discontinued from initial assignment as π_S (ω_S). Notice that $\pi_O + \pi_S = 1$ ($\omega_O + \omega_S = 1$). Condition on the patient being discontinued at t_j, denote the probability of dropout at t_j as π_{Mj} (ω_{Mj}).The marginal dropout probability at t_j is $\pi_{Sj}\pi_{Mj}$ ($\omega_{Sj}\omega_{Mj}$). Denote the overall probability of dropout as π_M (ω_M). The marginal dropout probability is $\pi_S\pi_M$ ($\omega_S\omega_M$). The number of patients of each portion can thus be considered as following

multinomial distribution. For treatment arm, $[r, q_2, ..., q_T, p_2, ..., p_T] \sim MVN(N, \boldsymbol{\pi})$ where $\boldsymbol{\pi} = [\pi_O, (1-\pi_{M2})\pi_{S2}, ..., (1-\pi_{MT})\pi_{ST}, \pi_{M2}\pi_{S2}, ..., \pi_{MT}\pi_{ST}]$. For reference arm:$[l, m_2, ..., m_T, n_2, ..., n_T] \sim MVN(N, \boldsymbol{\omega})$ where $\boldsymbol{\omega} = [\omega_O, (1-\omega_{M2})\omega_{S2}, ..., (1-\omega_{MT})\omega_{ST}, \omega_{M2}\omega_{S2}, ..., \omega_{MT}\omega_{ST}]$

3.3 Construction of the statistical estimand

In order to specifically construct the statistical estimand that corresponding to the scientific question of interest, which is the mean effect difference at the last time point with the effect of switch to rescue medication taken into account, the well accepted MMRM model is first used to analyze the data. It has the advantage of directly estimating the treatment effect at each measurement time point. The following MMRM model is used to fit the observed data,

$$\begin{aligned} &Y_{i,j}|G_i = g \\ &= \alpha_0 + \alpha_1 g + \alpha_2 I_{j=2} + ... + \alpha_T I_{j=T} + \alpha_{T+1} g I_{j=2} + ... + \alpha_{2T-1} g I_{j=T} + \epsilon_{i,j} \end{aligned} \tag{3.1}$$

where $[\epsilon_{i,1}, ..., \epsilon_{i,T}]' \sim N(\mathbf{0}_{T\times 1}, \boldsymbol{\Sigma}_\epsilon)$. $g = 1$ indicates treatment arm patients and $g = 0$ indicates reference arm patients.

To present the model in matrix format, denote $\mathbf{Y}_i = [Y_{i,1}, Y_{i,2}, ..., Y_{i,T}]'$ and $\mathbf{U}$ the design matrix for patient i,

$$\mathbf{U} = \begin{bmatrix} 1 & g & 0 & ... & 0 & 0 & ... & 0 \\ 1 & g & 1 & ... & 0 & g & ... & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & g & 0 & ... & 1 & 0 & ... & g \end{bmatrix}$$

$$\boldsymbol{\alpha} = [\alpha_0, \alpha_1, \alpha_2, \alpha_3, ..., \alpha_{T+1}, \alpha_{T+2}, ..., \alpha_{2T-1}]'$$

Therefore, $E(\mathbf{Y}_i|G_i = g) = \mathbf{U}\boldsymbol{\alpha}$. The 'De Facto' estimand can be defined as

$$\theta = \alpha_1 + \alpha_{2T-1} \tag{3.2}$$

Table 3.1: Data structure for active arm when rescue medication is allowed for patients who discontinued from initial randomized treatment assignment

Subject	Time 1 (t_1)	...	Time T (t_T)	S_i(Event)	E_i(Evt Time)	G_i(Group)
1	$y_{1,1}$	$\cdots$	$y_{1,T}$	1	T	1
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
r	$y_{r,1}$	$\cdots$	$y_{r,T}$	1	T	1
$r+1$	$y_{r+1,1}$	$\cdots$	$y_{r+1,T}$	2	T	1
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$r+q_T$	$y_{q_1,1}$	$\cdots$	$y_{q_1,T}$	2	T	1
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$r+q_T+\cdots+q_3$	$y_{r+q_T+\cdots+q_3,1}$	$\cdots$	$y_{r+q_T+\cdots+q_3,T}$	2	3	1
$r+q_T+\cdots+q_3+1$	$y_{r+q_T+\cdots+q_3+1,1}$	$\cdots$	$y_{r+q_T+\cdots+q_3+1,T}$	2	2	1
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$r+q_T+\cdots+q_3+q_2$	$y_{r+q_T+\cdots+q_3+q_2,1}$	$\cdots$	$y_{r+q_T+\cdots+q_3+q_2,T}$	2	2	1
$r+q+1$	$y_{r+q_T+\cdots+q_3+q_2+1,1}$	$\cdots$	? $(y_{r+q_T+\cdots+q_3+q_2+1,T})$	3	T	1
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$r+q+p_T$	$y_{r+q+p_T,1}$	$\cdots$	? $(y_{r+q+p_T,T})$	3	T	1
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$r+q+p_T+\cdots+p_3+1$	$y_{r+q+p_T+\cdots+p_3+1,1}$	$\cdots$	? $(y_{r+q+p_T+\cdots+p_3+1,T})$	3	2	1
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$r+q+p_T+\cdots+p_3+p_2$	$y_{r+q+p_T+\cdots+p_3+p_2,1}$	$\cdots$	? $(y_{r+q+p_T+\cdots+p_3+p_2,T})$	3	2	1

3.4 Conditional MMRM model

In the previous section, the MMRM is constructed regardless of the intercurrent event. Nonetheless, the missing data would be ignored without further investigation. The previous model is referred to as a marginal MMRM model. In order to investigate the missing data effect, a conditional MMRM model can be introduced as,

$$
\begin{aligned}
&Y_{i,j}|S_i, E_i, G_i \\
&= \beta_0 + \beta_1 G_i + \beta_2 I_{j=2} + ... + \beta_T I_{j=T} + \beta_{T+1} I_{j=2} G_i + ... + \beta_{2T-1} I_{j=T} G_i \\
&+ \gamma_{2,1} I_{j=2,S_i=2,E_i=2,G_i=0} + \gamma_{2,2} I_{j=2,S_i=2,E_i=2,G_i=1} + \gamma_{2,3} I_{j=2,S_i=3,E_i=2,G_i=0} \\
&+ \gamma_{2,4} I_{j=2,S_i=3,E_i=2,G_i=1} \\
&+ \gamma_{3,1} I_{j=3,S_i=2,E_i=2,G_i=0} + \gamma_{3,2} I_{j=3,S_i=2,E_i=3,G_i=0} + \gamma_{3,3} I_{j=3,S_i=2,E_i=2,G_i=1} \\
&+ \gamma_{3,4} I_{j=3,S_i=2,E_i=3,G_i=1} \\
&+ \gamma_{3,5} I_{j=3,S_i=3,E_i=2,G_i=0} + \gamma_{3,6} I_{j=3,S_i=3,E_i=3,G_i=0} + \gamma_{3,7} I_{j=3,S_i=3,E_i=2,G_i=1} \\
&+ \gamma_{3,8} I_{j=3,S_i=3,E_i=3,G_i=1} \\
&+ ... \\
&+ \gamma_{a,1} I_{j=a,S_i=2,E_i=2,G_i=0} + ... + \gamma_{a,a-1} I_{j=a,S_i=2,E_i=a,G_i=0} \\
&+ \gamma_{a,a} I_{j=a,S_i=2,E_i=2,G_i=1} + ... + \gamma_{a,2a-2} I_{j=a,S_i=2,E_i=a,G_i=1} \\
&+ \gamma_{a,2a-1} I_{j=a,S_i=3,E_i=2,G_i=0} + ... + \gamma_{a,3a-3} I_{j=a,S_i=3,E_i=a,G_i=0} \\
&+ \gamma_{a,3a-2} I_{j=a,S_i=3,E_i=2,G_i=1} + ... + \gamma_{a,4a-4} I_{j=a,S_i=3,E_i=a,G_i=1} \\
&+ ... \\
&+ \gamma_{T,1} I_{j=T,S_i=2,E_i=2,G_i=0} + ... + \gamma_{T,T-1} I_{j=T,S_i=2,E_i=T,G_i=0} \\
&+ \gamma_{T,T} I_{j=T,S_i=2,E_i=2,G_i=1} + ... + \gamma_{T,2T-2} I_{j=T,S_i=2,E_i=T,G_i=1} \\
&+ \gamma_{T,2T-1} I_{j=T,S_i=3,E_i=2,G_i=0} + ... + \gamma_{T,3T-3} I_{j=T,S_i=3,E_i=T,G_i=0} \\
&+ \gamma_{T,3T-2} I_{j=T,S_i=3,E_i=2,G_i=1} + ... + \gamma_{T,4T-4} I_{j=T,S_i=3,E_i=T,G_i=1} \\
&+ \epsilon_{i,j}
\end{aligned}
\tag{3.3}
$$

where $[\epsilon_{i,1}, ..., \epsilon_{i,T}] \sim N(\mathbf{0}_{T\times 1}, \mathbf{\Sigma}_\epsilon)$. The model can be viewed as two parts. The β part is the initial effect if no intercurrent event happened, that is, neither switch nor dropout happened. If an intercurrent event happened, the event effect is considered to be deviated from the initial effect, which is denoted as γ. The effect varies at different measurement time points t_j as well as the intercurrent happening time E_i. For example, for time point t_2, $\boldsymbol{\gamma}_2 = [\gamma_{2,1}, \gamma_{2,2}, \gamma_{2,3}, \gamma_{2,4}]$ represents the different deviate effect due to intercurrent event happened on t_2. $\gamma_{2,1}$ represents the deviate effect if reference arm patients switched to rescue medication at t_2, $\gamma_{2,2}$ represents the deviate effect if treatment arm patients switched to rescue medication at t_2, $\gamma_{2,3}$ represents the deviate effect if reference arm patients dropped out at t_2 and $\gamma_{2,4}$ represents the deviate effect if treatment arm patients dropped out at t_2. For t_2 measurements, only events that happened before t_2, which is at t_1 will have an deviate effect on the t_2 measurements. In general, only events that happened prior to the measurement time point t_j ($E_i \leq j$) will have a deviate effect for measurements at time point t_j. Notice that $\gamma_{2,3}, \gamma_{2,4}$ are related to the dropout measurements that are unobserved. Thus these two parameters are inestimable. The technique to handle this problem will be discussed later in estimator construction section.

The conditional MMRM can as well be written in a matrix format.

$$\mathbf{Y}_i | S_i = s, E_i = e, G_i = g = \mathbf{U}\boldsymbol{\beta} + \boldsymbol{\delta}_{s,e,g}\boldsymbol{\gamma} + \boldsymbol{\epsilon}_i \tag{3.4}$$

where $\mathbf{U}$ is the same design matrix as the marginal MMRM model for patient i, $\boldsymbol{\delta}_{s,e,g}$ is an indicator matrix denoting the deviate effects caused by the intercurrent event for the patient with no intercurrent event ($s = 1$), switch to rescue medication ($s = 2$) or dropout ($s = 3$) at time point e for group g. For no intercurrent event patients, $s = 1$,

$$\boldsymbol{\delta}_{1,T,0} = \boldsymbol{\delta}_{1,T,1} = 0$$

For patients with intercurrent event $s > 1$,

$$\boldsymbol{\delta}_{s,e,g} = \begin{bmatrix} \mathbf{0} & \mathbf{0} & ... & \mathbf{0} \\ \mathbf{d}_{2,seg} & \mathbf{0} & ... & \mathbf{0} \\ \mathbf{0} & \mathbf{d}_{3,seg} & ... & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & ... & \mathbf{d}_{T,seg} \end{bmatrix} \quad \text{(Note: } 2 \leq e \leq j\text{)}$$

where

$$\begin{aligned} \mathbf{d}_{j,2e0} &= [0...1...0, 0...0...0, 0...0...0, 0...0...0] \\ \mathbf{d}_{j,2e1} &= [0...0...0, 0...1...0, 0...0...0, 0...0...0] \\ \mathbf{d}_{j,3e0} &= [0...0...0, 0...0...0, 0...1...0, 0...0...0] \\ \mathbf{d}_{j,3e1} &= [0...0...0, 0...0...0, 0...0...0, 0...1...0] \end{aligned} \tag{3.5}$$

$$\begin{aligned} \boldsymbol{\beta} &= [\beta_0, ..., \beta_{2T-1}]' \\ \boldsymbol{\gamma} &= [\boldsymbol{\gamma}_2, ..., \boldsymbol{\gamma}_T]' \end{aligned} \tag{3.6}$$

where

$$\begin{aligned} \boldsymbol{\gamma}_j = [&\gamma_{j,1}, ..., \gamma_{j,e-1}, ...\gamma_{j,j-1}, \gamma_{j,j}, ..., \gamma_{j,j-2+e}, ..., \gamma_{j,2j-2}, \\ &\gamma_{j,2j-1}, ..., \gamma_{j,2j-3+e}, ...\gamma_{j,3j-3}, \gamma_{j,3j-2}, ..., \gamma_{j,3j-4+e}, ...\gamma_{j,4j-4}] \end{aligned} \tag{3.7}$$

3.5 Convert between marginal and conditional MMRM models

The 'De Facto' estimand defined from the marginal MMRM can also be represented by the parameters from the conditional MMRM model. The conditional and marginal MMRM model parameters can be connected through the mean of each time point. From marginal MMRM model, the mean at each time point is $E(\mathbf{Y}_i|G_i) = E(\mathbf{U}\boldsymbol{\alpha} +$

$\boldsymbol{\epsilon}_i) = \mathbf{U}\boldsymbol{\alpha}$. To write in detail,

$$E\left(\begin{bmatrix} y_{i,1} \\ y_{i,2} \\ \vdots \\ y_{i,T} \end{bmatrix} |G_i = 0 \right) = \begin{bmatrix} \alpha_0 \\ \alpha_0 + \alpha_1 \\ \vdots \\ \alpha_0 + \alpha_{T-1} \end{bmatrix} \tag{3.8}$$

$$E\left(\begin{bmatrix} y_{i,1} \\ y_{i,2} \\ \vdots \\ y_{i,T} \end{bmatrix} |G_i = 1 \right) = \begin{bmatrix} (\alpha_0 + \alpha_1) \\ (\alpha_0 + \alpha_1) + (\alpha_2 + \alpha_{T+1}) \\ \vdots \\ (\alpha_0 + \alpha_1) + (\alpha_T + \alpha_{2T-1}) \end{bmatrix} \tag{3.9}$$

The expectation can also be obtained from the weighted sum of conditional means, that is

$$E(\mathbf{Y}_i|G_i = g) = \sum_{s=1}^{3}\sum_{e=2}^{T} E(\mathbf{Y}_i|S_i = s, E_i = e, G_i = g)p_{seg} = \sum_{s=1}^{3}\sum_{e=2}^{T}(\mathbf{U}\boldsymbol{\beta} + \boldsymbol{\delta}_{seg}\boldsymbol{\gamma})p_{seg} \tag{3.10}$$

where p_{seg} denotes the probability of event type $S_i = s$ and event time $E_i = e$ for arm $G_i = g$. To be more specific, for each arm,

$$\begin{aligned} E(\mathbf{Y}_i|G_i = 1) = {} & E(\mathbf{Y}_i|G_i = 1, S_i = 1, E_i = T)\pi_O \\ & + \sum_{e=2}^{T} E(\mathbf{Y}_i|G_i = 1, S_i = 2, E_i = e)(1 - \pi_{Me})\pi_{Se} \\ & + \sum_{e=2}^{T} E(\mathbf{Y}_i|G_i = 1, S_i = 3, E_i = e)\pi_{Me}\pi_{Se} \end{aligned} \tag{3.11}$$

$$\begin{aligned} E(\mathbf{Y}_i|G_i = 0) = {} & E(\mathbf{Y}_i|G_i = 0, S_i = 1, E_i = T)\omega_O \\ & + \sum_{e=2}^{T} E(\mathbf{Y}_i|G_i = 0, S_i = 2, E_i = e)(1 - \omega_{Me})\omega_{Se} \\ & + \sum_{e=2}^{T} E(\mathbf{Y}_i|G_i = 0, S_i = 3, E_i = e)\omega_{Me}\omega_{Se} \end{aligned} \tag{3.12}$$

To expand into details, denote $\boldsymbol{y_i} = \Big[y_{i,1}y_{i,2}...y_{i,T}\Big]'$

$$
\begin{aligned}
&E\Big(\boldsymbol{y_i}|G_i=1\Big)\\
&=E\Big(\boldsymbol{y_i}|G_i=1,S_i=1,E_i=T\Big)\pi_O+\sum_{e=2}^{T}E\Big(\boldsymbol{y_i}|G_i=1,S_i=2,E_i=e\Big)(1-\pi_{Me})\pi_{Se}\\
&+\sum_{e=2}^{T}E\Big(\boldsymbol{y_i}|G_i=1,S_i=3,E_i=e\Big)\pi_{Me}\pi_{Se}\\
&=\begin{bmatrix}(\beta_0+\beta_1)\\(\beta_0+\beta_1)+(\beta_2+\beta_{T+1})\\\vdots\\(\beta_0+\beta_1)+(\beta_T+\beta_{2T-1})\end{bmatrix}\pi_O\\
&+\begin{bmatrix}(\beta_0+\beta_1)(1-\pi_M)\pi_S\\\big((\beta_0+\beta_1)+(\beta_2+\beta_{T+1})\big)(1-\pi_M)\pi_S+(1-\pi_{M2})\pi_{S2}\gamma_{2,2}\\\vdots\\\big((\beta_0+\beta_1)+(\beta_T+\beta_{2T-1})\big)(1-\pi_M)\pi_S+\sum_{e=2}^{T}(1-\pi_{Me})\pi_{Se}\gamma_{T,T-2+e}\end{bmatrix}\\
&+\begin{bmatrix}(\beta_0+\beta_1)\pi_M\pi_S\\\big((\beta_0+\beta_1)+(\beta_2+\beta_{T+1})\big)\pi_M\pi_S+\pi_{M2}\pi_{S2}\gamma_{2,4}\\\vdots\\\big((\beta_0+\beta_1)+(\beta_T+\beta_{2T-1})\big)\pi_M\pi_S+\sum_{e=2}^{T}\pi_{Me}\pi_{Se}\gamma_{T,3T-4+e}\end{bmatrix}\\
&=\begin{bmatrix}(\beta_0+\beta_1)\\(\beta_0+\beta_1)+(\beta_2+\beta_{T+1})+(1-\pi_{M2})\pi_{S2}\gamma_{2,2}+\pi_{M2}\pi_{S2}\gamma_{2,4}\\\vdots\\(\beta_0+\beta_1)+(\beta_T+\beta_{2T-1})+\sum_{e=2}^{T}(1-\pi_{Me})\pi_{Se}\gamma_{T,T-2+e}+\sum_{e=2}^{T}\pi_{Me}\pi_{Se}\gamma_{T,3T-4+e}\end{bmatrix}
\end{aligned}
\tag{3.13}
$$

$$
\begin{aligned}
&E\left(\boldsymbol{y_i}|G_i=0\right)\\
&=E\left(\boldsymbol{y_i}|G_i=0,S_i=1,E_i=T\right)\omega_O+\sum_{e=2}^{T}E\left(\boldsymbol{y_i}|G_i=0,S_i=2,E_i=e\right)(1-\omega_{Me})\omega_{Se}\\
&+\sum_{e=2}^{T}E\left(\boldsymbol{y_i}|G_i=0,S_i=3,E_i=e\right)\omega_{Me}\omega_{Se}\\
&=\begin{bmatrix}\beta_0\\ \beta_0+\beta_2\\ \vdots\\ \beta_0+\beta_T\end{bmatrix}\omega_O+\begin{bmatrix}\beta_0(1-\omega_M)\omega_S\\ (\beta_0+\beta_2)(1-\omega_M)\omega_S+(1-\omega_{M2})\omega_{S2}\gamma_{2,1}\\ \vdots\\ (\beta_0+\beta_T)(1-\omega_M)\omega_S+\sum_{e=2}^{T}(1-\omega_{Me})\omega_{Se}\gamma_{T,e-1}\end{bmatrix}\\
&+\begin{bmatrix}\beta_0\omega_M\omega_S\\ (\beta_0+\beta_2)\omega_M\omega_S+\omega_{M2}\omega_{S2}\gamma_{2,3}\\ \vdots\\ (\beta_0+\beta_T)\omega_M\omega_S+\sum_{e=2}^{T}\omega_{Me}\omega_{Se}\gamma_{T,2T-3+e}\end{bmatrix}\\
&=\begin{bmatrix}\beta_0\\ \beta_0+\beta_2+(1-\omega_{M2})\omega_{S2}\gamma_{2,1}+\omega_{M2}\omega_{S2}\gamma_{2,3}\\ \vdots\\ \beta_0+\beta_T+\sum_{e=2}^{T}(1-\omega_{Me})\omega_{Se}\gamma_{T,e-1}+\sum_{e=2}^{T}\omega_{Me}\omega_{Se}\gamma_{T,2T-3+e}\end{bmatrix}
\end{aligned}
\tag{3.14}
$$

From the two different modeling approaches, the following equation can be derived,

$$
\mathbf{U}\boldsymbol{\alpha}=\sum_{s=1}^{3}\sum_{e=2}^{T}(\mathbf{U}\boldsymbol{\beta}+\boldsymbol{\delta}_{s,e,g}\boldsymbol{\gamma})p_{s,e,g} \tag{3.15}
$$

First I investigate the reference arm marginal means $G_i = 0$

$$\begin{bmatrix} \alpha_0 \\ \alpha_0 + \alpha_2 \\ \vdots \\ \alpha_0 + \alpha_T \end{bmatrix} = \begin{bmatrix} \beta_0 \\ \beta_0 + \beta_2 + (1 - \omega_{M2})\omega_{S2}\gamma_{2,1} + \omega_{M2}\omega_{S2}\gamma_{2,3} \\ \vdots \\ \beta_0 + \beta_T + \sum_{e=2}^{T}(1 - \omega_{Me})\omega_{Se}\gamma_{T,e-1} + \sum_{e=2}^{T}\omega_{Me}\omega_{Se}\gamma_{T,2T-3+e} \end{bmatrix} \tag{3.16}$$

Next I investigate the treatment arm $G_i = 1$

$$\begin{bmatrix} (\alpha_0 + \alpha_1) \\ (\alpha_0 + \alpha_1) + (\alpha_2 + \alpha_{T+1}) \\ \vdots \\ (\alpha_0 + \alpha_1) + (\alpha_T + \alpha_{2T-1}) \end{bmatrix}$$

$$\| \tag{3.17}$$

$$\begin{bmatrix} (\beta_0 + \beta_1) \\ (\beta_0 + \beta_1) + (\beta_2 + \beta_{T+1}) + (1 - \pi_{M2})\pi_{S2}\gamma_{2,2} + \pi_{M2}\pi_{S2}\gamma_{2,4} \\ \vdots \\ (\beta_0 + \beta_1) + (\beta_T + \beta_{2T-1}) + \sum_{e=2}^{T}(1 - \pi_{Me})\pi_{Se}\gamma_{T,T-2+e} + \sum_{e=2}^{T}\pi_{Me}\pi_{Se}\gamma_{T,3T-4+e} \end{bmatrix}$$

From previous two equations, the following equation can be derived

$$\alpha_1 = \beta_1 \tag{3.18}$$

$$\begin{aligned} \alpha_{2T-1} = \beta_{2T-1} + \sum_{e=2}^{T} \Big(& \pi_{Se}\gamma_{T,T-2+e} - \omega_{Se}\gamma_{T,e-1} + \pi_{Me}\pi_{Se}(\gamma_{T,3T-4+e} - \gamma_{T,T-2+e}) \\ & - \omega_{Me}\omega_{Se}(\gamma_{T,2T-3+e} - \gamma_{T,e-1}) \Big) \end{aligned} \tag{3.19}$$

Thus the targeted 'De Facto' estimand can be written as

$$\begin{aligned} \theta = \beta_1 + \beta_{2T-1} + \sum_{e=2}^{T} \Big(& \pi_{Se}\gamma_{T,T-2+e} - \omega_{Se}\gamma_{T,e-1} + \pi_{Me}\pi_{Se}(\gamma_{T,3T-4+e} - \gamma_{T,T-2+e}) \\ & - \omega_{Me}\omega_{Se}(\gamma_{T,2T-3+e} - \gamma_{T,e-1}) \Big) \end{aligned} \tag{3.20}$$

Chapter 4

Primary and sensitivity analysis point estimator and inference

4.1 General form for 'De Facto' estimator and inference

Denote the estimator for the 'De Facto' estimand as

$$\hat{\theta} = \hat{\beta}_1 + \hat{\beta}_{2T-1} + \sum_{e=2}^{T} \Big(\hat{\pi}_{Se}\hat{\gamma}_{T,T-2+e} - \hat{\omega}_{Se}\hat{\gamma}_{T,e-1} + \hat{\pi}_{Me}\hat{\pi}_{Se}(\hat{\gamma}_{T,3T-4+e} - \hat{\gamma}_{T,T-2+e}) - \hat{\omega}_{Me}\hat{\omega}_{Se}(\hat{\gamma}_{T,2T-3+e} - \hat{\gamma}_{T,e-1}) \Big) \tag{4.1}$$

which is a linear combination of parameters from the conditional MMRM model. Regardless what method is used to obtain the parameter estimates, which will be discussed in the later chapter, as I mentioned briefly before, some of the parameters are inestimable. Those are the parameters related to the missing measurements after dropout, namely $\gamma_{T,2T-1}, ..., \gamma_{T,4T-4}$. One way to handle these inestimable parameters is to borrow information from estimable parameters based on intuitive assumptions. For example, assumption can be made that patients who dropped out from the study might be still on treatment arm if not missing. However, this is rarely the case in the real world because patients are more often dropped out due to lack of efficacy or safety issue. It will lead to a biased result if assume the dropout patients would have same effect as those who remained on treatment. It is more likely that the patient would be offered a rescue medication after discontinued from the initial assignment and thus they would behave as those patients who switched to rescue and remained in

the study. The estimator constructed based on this assumption would be the primary estimator in this dissertation, denoted as $\hat{\theta}^{RES}$. If the hypothesis testing result based on the primary assumption is statistical significant, investigators might want to know how robust the result is regarding this assumption Therefore one or more sensitivity analysis is necessary. If the hypothesis testing result is not significant, there is no need to conduct any sensitivity analysis. One might want to modify the estimand to target other clinical relevant question of interest. In this chapter, I am going to construct point estimator for primary analysis as well as different types of sensitivity analysis methods. Regardless of the assumptions made to estimate the inestimable parameters $\gamma_{T,2T-1}, ..., \gamma_{T,4T-4}$, delta approximation method can be used to derive the variance for the point estimator $\hat{\theta}$ since it is a linear combination of MMRM model parameters. Thus a general form of the variance estimator can be derived.

$$
\begin{aligned}
& V(\hat{\theta}) \\
=& V\Big(E\big(\hat{\theta}|\mathbf{\Omega}\big)\Big) + E\Big(V\big(\hat{\theta}|\mathbf{\Omega}\big)\Big) \\
=& V\Bigg(E\Big(\hat{\beta}_1 + \hat{\beta}_{2T-1} + \sum_{e=2}^{T}\big(\hat{\pi}_{Se}\hat{\gamma}_{T,T-2+e} - \hat{\omega}_{Se}\hat{\gamma}_{T,e-1} + \hat{\pi}_{Me}\hat{\pi}_{Se}(\hat{\gamma}_{T,3T-4+e} - \hat{\gamma}_{T,T-2+e}) \\
& - \hat{\omega}_{Me}\hat{\omega}_{Se}(\hat{\gamma}_{T,2T-3+e} - \hat{\gamma}_{T,e-1})\big)\Big)\Bigg) \\
& + E\Bigg(V\Big(\hat{\beta}_1 + \hat{\beta}_{2T-1} + \sum_{e=2}^{T}\big(\hat{\pi}_{Se}\hat{\gamma}_{T,T-2+e} - \hat{\omega}_{Se}\hat{\gamma}_{T,e-1} + \hat{\pi}_{Me}\hat{\pi}_{Se}(\hat{\gamma}_{T,3T-4+e} - \hat{\gamma}_{T,T-2+e}) \\
& - \hat{\omega}_{Me}\hat{\omega}_{Se}(\hat{\gamma}_{T,2T-3+e} - \hat{\gamma}_{T,e-1})\big)\Big)\Bigg) \\
=& V\Bigg(\beta_1 + \beta_{2T-1} + \sum_{e=2}^{T}\Big(\hat{\pi}_{Se}\gamma_{T,T-2+e} - \hat{\omega}_{Se}\gamma_{T,e-1} + \hat{\pi}_{Me}\hat{\pi}_{Se}(\gamma_{T,3T-4+e} - \gamma_{T,T-2+e}) \\
& - \hat{\omega}_{Me}\hat{\omega}_{Se}(\gamma_{T,2T-3+e} - \gamma_{T,e-1})\Big)\Bigg) \\
& + E\Bigg(V\Big(\hat{\beta}_1 + \hat{\beta}_{2T-1} + \sum_{e=2}^{T}\big(\hat{\pi}_{Se}\hat{\gamma}_{T,T-2+e} - \hat{\omega}_{Se}\hat{\gamma}_{T,e-1} + \hat{\pi}_{Me}\hat{\pi}_{Se}(\hat{\gamma}_{T,3T-4+e} - \hat{\gamma}_{T,T-2+e}) \\
& - \hat{\omega}_{Me}\hat{\omega}_{Se}(\hat{\gamma}_{T,2T-3+e} - \hat{\gamma}_{T,e-1}))|\mathbf{\Omega}\Big)\Bigg)
\end{aligned}
\tag{4.2}
$$

where $\mathbf{\Omega} = [q_2, ..., q_T, p_2, ..., p_T, m_2, ..., m_T, n_2, ..., n_T]$.

The first part of the variance can be written in the following matrix format

$$\begin{aligned} & V\Bigg(\beta_1 + \beta_{2T-1} + \sum_{e=2}^{T}\Big(\hat{\pi}_{Se}\gamma_{T,T-2+e} - \hat{\omega}_{Se}\gamma_{T,e-1} + \hat{\pi}_{Me}\hat{\pi}_{Se}(\gamma_{T,3T-4+e} - \gamma_{T,T-2+e}) \\ & - \hat{\omega}_{Me}\hat{\omega}_{Se}(\gamma_{T,2T-3+e} - \gamma_{T,e-1})\Big)\Bigg) \\ & = P\Sigma_P P' + O\Sigma_O O' \end{aligned} \tag{4.3}$$

where

$$P = \begin{bmatrix} \gamma_{T,3T-2} & \gamma_{T,3T-1} & \cdots & \gamma_{T,4T-4} & \gamma_{T,T} & \gamma_{T,T+1} & \cdots & \gamma_{T,2T-2} \end{bmatrix}$$

$$O = \begin{bmatrix} \gamma_{T,1} & \gamma_{T,2} & \cdots & \gamma_{T,2T-1} \end{bmatrix}$$

$$\Sigma_P = \begin{bmatrix} \Sigma_{P11} & \Sigma_{P12} \\ \Sigma_{P12} & \Sigma_{P22} \end{bmatrix}$$

and

$$\Sigma_{P11} = \begin{bmatrix} \pi_{M2}\pi_{S2}(1-\pi_{M2}\pi_{S2}) & \pi_{M2}\pi_{S2}\pi_{M3}\pi_{S3} & \cdots & \pi_{M2}\pi_{S2}\pi_{MT}\pi_{ST} \\ \pi_{M2}\pi_{S2}\pi_{M3}\pi_{S3} & \pi_{M3}\pi_{S3}(1-\pi_{M3}\pi_{S3}) & \cdots & \pi_{M3}\pi_{S3}\pi_{MT}\pi_{ST} \\ \vdots & \vdots & \ddots & \vdots \\ \pi_{M2}\pi_{S2}\pi_{MT}\pi_{ST} & \pi_{M3}\pi_{S3}\pi_{MT}\pi_{ST} & \cdots & \pi_{MT}\pi_{ST}(1-\pi_{MT}\pi_{ST}) \end{bmatrix}$$

$$\Sigma_{P12} = \begin{bmatrix} \pi_{M2}\pi_{S2}(1-\pi_{M2})\pi_{S2} & \pi_{M2}\pi_{S2}(1-\pi_{M3})\pi_{S3} & \cdots & \pi_{M2}\pi_{S2}(1-\pi_{MT})\pi_{ST} \\ \pi_{M3}\pi_{S3}(1-\pi_{M2})\pi_{S2} & \pi_{M3}\pi_{S3}(1-\pi_{M3})\pi_{S3} & \cdots & \pi_{M3}\pi_{S3}(1-\pi_{MT})\pi_{ST} \\ \vdots & \vdots & \ddots & \vdots \\ \pi_{M2}\pi_{S2}(1-\pi_{MT})\pi_{ST} & \pi_{M3}\pi_{S3}(1-\pi_{MT})\pi_{ST} & \cdots & \pi_{MT}\pi_{ST}(1-\pi_{MT})\pi_{ST} \end{bmatrix}$$

$$\Sigma_{P22} = \begin{bmatrix} (1-(1-\pi_{M2})\pi_{S2})(1-\pi_{M2})\pi_{S2} & (1-\pi_{M2})\pi_{S2}(1-\pi_{M3})\pi_{S3} & \cdots & (1-\pi_{M2})\pi_{S2}(1-\pi_{MT})\pi_{ST} \\ (1-\pi_{M2})\pi_{S2}(1-\pi_{M3})\pi_{S3} & (1-(1-\pi_{M3})\pi_{S3})(1-\pi_{M3})\pi_{S3} & \cdots & (1-\pi_{M3})\pi_{S3}(1-\pi_{MT})\pi_{ST} \\ \vdots & \vdots & \ddots & \vdots \\ (1-\pi_{M2})\pi_{S2}(1-\pi_{MT})\pi_{ST} & (1-\pi_{M3})\pi_{S3}(1-\pi_{MT})\pi_{ST} & \cdots & (1-(1-\pi_{MT})\pi_{ST})(1-\pi_{MT})\pi_{ST} \end{bmatrix}$$

$$\Sigma_O = \begin{bmatrix} \Sigma_{O11} & \Sigma_{O12} \\ \Sigma_{O12} & \Sigma_{O22} \end{bmatrix}$$

where

$$\Sigma_{O11} = \begin{bmatrix} \omega_{M2}\omega_{S2}(1-\omega_{M2}\omega_{S2}) & \omega_{M2}\omega_{S2}\omega_{M3}\omega_{S3} & ... & \omega_{M2}\omega_{S2}\omega_{MT}\omega_{ST} \\ \omega_{M2}\omega_{S2}\omega_{M3}\omega_{S3} & \omega_{M3}\omega_{S3}(1-\omega_{M3}\omega_{S3}) & ... & \omega_{M3}\omega_{S3}\omega_{MT}\omega_{ST} \\ \vdots & \vdots & \ddots & \vdots \\ \omega_{M2}\omega_{S2}\omega_{MT}\omega_{ST} & \omega_{M3}\omega_{S3}\omega_{MT}\omega_{ST} & ... & \omega_{MT}\omega_{ST}(1-\omega_{MT}\omega_{ST}) \end{bmatrix}$$

$$\Sigma_{P12} = \begin{bmatrix} \omega_{M2}\omega_{S2}(1-\omega_{M2})\omega_{S2} & \omega_{M2}\omega_{S2}(1-\omega_{M3})\omega_{S3} & ... & \omega_{M2}\omega_{S2}(1-\omega_{MT})\omega_{ST} \\ \omega_{M3}\omega_{S3}(1-\omega_{M2})\omega_{S2} & \omega_{M3}\omega_{S3}(1-\omega_{M3})\omega_{S3} & ... & \omega_{M3}\omega_{S3}(1-\omega_{MT})\omega_{ST} \\ \vdots & \vdots & \ddots & \vdots \\ \omega_{M2}\omega_{S2}(1-\omega_{MT})\omega_{ST} & \omega_{M3}\omega_{S3}(1-\omega_{MT})\omega_{ST} & ... & \omega_{MT}\omega_{ST}(1-\omega_{MT})\omega_{ST} \end{bmatrix}$$

$$\Sigma_{P22} = \begin{bmatrix} (1-(1-\omega_{M2})\omega_{S2})(1-\omega_{M2})\omega_{S2} & (1-\omega_{M2})\omega_{S2}(1-\omega_{M3})\omega_{S3} & ... & (1-\omega_{M2})\omega_{S2}(1-\omega_{MT})\omega_{ST} \\ (1-\omega_{M2})\omega_{S2}(1-\omega_{M3})\omega_{S3} & (1-(1-\omega_{M3})\omega_{S3})(1-\omega_{M3})\omega_{S3} & ... & (1-\omega_{M3})\omega_{S3}(1-\omega_{MT})\omega_{ST} \\ \vdots & \vdots & \ddots & \vdots \\ (1-\omega_{M2})\omega_{S2}(1-\omega_{MT})\omega_{ST} & (1-\omega_{M3})\omega_{S3}(1-\omega_{MT})\omega_{ST} & ... & (1-(1-\omega_{MT})\omega_{ST})(1-\omega_{MT})\omega_{ST} \end{bmatrix}$$

The second part of the variance can also been written in matrix format,

$$E\Big(V\Big(\hat{\beta}_1 + \hat{\beta}_{2T-1} + \sum_{e=2}^{T}\big(\hat{\pi}_{Se}\hat{\gamma}_{T,T-2+e} - \hat{\omega}_{Se}\hat{\gamma}_{T,e-1} + \hat{\pi}_{Me}\hat{\pi}_{Se}(\hat{\gamma}_{T,3T-4+e} - \hat{\gamma}_{T,T-2+e}) - \hat{\omega}_{Me}\hat{\omega}_{Se}(\hat{\gamma}_{T,2T-3+e} - \hat{\gamma}_{T,e-1})\big)|\boldsymbol{\Omega}\Big) = B\Sigma_B B' \tag{4.4}$$

where $B = \begin{bmatrix} \boldsymbol{D}_\beta & \boldsymbol{D}_\gamma \end{bmatrix}$

$\boldsymbol{D}_\beta = \begin{bmatrix} 0 & 1 & ...(2\text{T-1 0's})... & 1 \end{bmatrix}$

$\boldsymbol{D}_\gamma = \begin{bmatrix} -(1-\hat{\omega}_{M2})\hat{\omega}_{S2} & ... & -(1-\hat{\pi}_{MT})\hat{\pi}_{ST} & (1-\hat{\pi}_{M2})\hat{\pi}_{S2} & ... & (1-\hat{\pi}_{MT})\hat{\pi}_{ST} & -\hat{\omega}_{M2}\hat{\omega}_{S2} & ... & \hat{\pi}_{MT}\hat{\pi}_{ST} & \hat{\pi}_{M2}\hat{\pi}_{S2}... & \hat{\pi}_{MT}\hat{\pi}_{ST} \end{bmatrix}$

4.2 Primary estimator

The primary estimator is constructed based on the assumption that the dropout patients would have behaved like the patients who switched to rescue medication,

that is,

$$\begin{aligned}\hat{E}(Y_{i,T}|S_i=3,E_i=e,G_i=0)&=\hat{E}(Y_{i,T}|S_i=2,E_i=e,G_i=0),\quad e=2,...,T\\ \hat{E}(Y_{i,T}|S_i=3,E_i=e,G_i=1)&=\hat{E}(Y_{i,T}|S_i=2,E_i=e,G_i=1),\quad e=2,...,T\end{aligned}\tag{4.5}$$

which leads to

$$\begin{bmatrix}\hat{\gamma}_{T,2T-1}\\ \vdots\\ \hat{\gamma}_{T,2T-3+e}\\ \vdots\\ \hat{\gamma}_{T,3T-3}\\ \hat{\gamma}_{T,3T-2}\\ \vdots\\ \hat{\gamma}_{T,3T-4+e}\\ \vdots\\ \hat{\gamma}_{T,4T-4}\end{bmatrix}=\begin{bmatrix}\hat{\gamma}_{T,1}\\ \vdots\\ \hat{\gamma}_{T,e-1}\\ \vdots\\ \hat{\gamma}_{T,T-1}\\ \hat{\gamma}_{T,T}\\ \vdots\\ \hat{\gamma}_{T,T-2+e}\\ \vdots\\ \hat{\gamma}_{T,2T-2}\end{bmatrix}\tag{4.6}$$

The left matrix is the matrix of inestimable parameters and the right matrix is the matrix of estimable parameters. The primary estimator can be written as

$$\begin{aligned}\hat{\theta}^{RES}&=\hat{\beta}_1+\hat{\beta}_{2T-1}+\sum_{e=2}^{T}\Big(\hat{\pi}_{Se}\hat{\gamma}_{T,T-2+e}-\hat{\omega}_{Se}\hat{\gamma}_{T,e-1}+\hat{\pi}_{Me}\hat{\pi}_{Se}(\hat{\gamma}_{T,T-2+e}-\hat{\gamma}_{T,T-2+e})\\ &\quad-\hat{\omega}_{Me}\hat{\omega}_{Se}(\hat{\gamma}_{T,e-1}-\hat{\gamma}_{T,e-1})\Big)\\ &=\hat{\beta}_1+\hat{\beta}_{2T-1}+\sum_{e=2}^{T}\Big(\hat{\pi}_{Se}\hat{\gamma}_{T,T-2+e}-\hat{\omega}_{Se}\hat{\gamma}_{T,e-1}\Big)\end{aligned}\tag{4.7}$$

4.3 Sensitivity analysis estimator

The assumptions made in terms of the inestimable parameters are eventually untestable. The purpose of the sensitivity analysis is to test the robustness of the result under different assumptions, usually more extreme assumptions because the question of interest is the difference between treatment and reference arms. Thus a more 'extreme'

assumption would be an assumption that shrinks the difference between two arms.

4.3.1 Jump to reference

The first sensitivity analysis estimator is constructed under the assumption that the treatment arm dropout patients would behave like reference patients while reference arm dropout patients still assumed to be as if switched to rescue medication. The estimator is denoted as $\hat{\theta}^{J2R}$ where the 'J2R' indicates that the treatment effect is assumed to 'Jump to reference'. To be more specific, for the reference arm, the assumption is the same as the primary analysis assumption,

$$\hat{\gamma}_{T,2T-3+e} = \hat{\gamma}_{T,e-1} \tag{4.8}$$

For the treatment arm,

$$\begin{aligned} \hat{E}(Y_{i,T}|S_i = 3, E_i = e, G_i = 1) &= \hat{E}(Y_{i,T}|S_i = 1, E_i = e, G_i = 0), \quad e = 2, ..., T \\ \hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_T + \hat{\beta}_{2T-1} + \hat{\gamma}_{3T-4+e} &= \hat{\beta}_0 + \hat{\beta}_T \\ \hat{\gamma}_{3T-4+e} &= -\hat{\beta}_1 - \hat{\beta}_{2T-1} \end{aligned} \tag{4.9}$$

Thus the 'J2R' estimator can be written as

$$\hat{\theta}^{J2R} = \hat{\beta}_1 + \hat{\beta}_{2T-1} + \sum_{e=2}^{T} \Big(\hat{\pi}_{Se}\hat{\gamma}_{T,T-2+e} - \hat{\omega}_{Se}\hat{\gamma}_{T,e-1} + \hat{\pi}_{Me}\hat{\pi}_{Se}(-\hat{\beta}_1 - \hat{\beta}_{2T-1} - \hat{\gamma}_{T,T-2+e}) \Big) \tag{4.10}$$

4.3.2 Jump to zero measurement

An even more extreme sensitivity analysis assumption is that the treatment arm dropout patients measurements would return to zero while reference arm dropout patients still assumed to be as if switched to rescue medication. The estimator is denoted as $\hat{\theta}^{J2Z}$ where the 'J2Z' indicates that the treatment effect is assumed to 'Jump to Zero'. Therefore, for the treatment arm,

$$\begin{aligned}
&\hat{E}(Y_{i,T}|S_i = 3, E_i = e, G_i = 1) = 0, \quad e = 2, ..., T \\
&\hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_T + \hat{\beta}_{2T-1} + \hat{\gamma}_{3T-4+e} = 0 \\
&\hat{\gamma}_{3T-4+e} = -(\hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_T + \hat{\beta}_{2T-1})
\end{aligned} \tag{4.11}$$

In another word, if the measured effect after dropout is assumed to become most extreme, which is zero (think of measurements such as heart rate, blood pressure or a score system which zero indicates the worst score), then the deviat effect is a constant for patients who dropped out at any time point. The 'J2Z' estimator can be written as

$$\begin{aligned}
\hat{\theta}^{J2Z} =& \hat{\beta}_1 + \hat{\beta}_{2T-1} + \sum_{e=2}^{T} \Big(\hat{\pi}_{Se} \hat{\gamma}_{T,T-2+e} - \hat{\omega}_{Se} \hat{\gamma}_{T,e-1} + \hat{\pi}_{Me} \hat{\pi}_{Se} (\hat{\gamma}_{T,3T-4+e} - \hat{\gamma}_{T,T-2+e}) \Big) \\
=& \hat{\beta}_1 + \hat{\beta}_{2T-1} + \sum_{e=2}^{T} \Big(\hat{\pi}_{Se} \hat{\gamma}_{T,T-2+e} - \hat{\omega}_{Se} \hat{\gamma}_{T,e-1} \\
& + \hat{\pi}_{Me} \hat{\pi}_{Se} (-(\hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_T + \hat{\beta}_{2T-1}) - \hat{\gamma}_{T,T-2+e}) \Big) \\
=& (1 - \hat{\pi}_M \hat{\pi}_S)(\hat{\beta}_1 + \hat{\beta}_{2T-1}) - \hat{\pi}_M \hat{\pi}_S (\hat{\beta}_0 + \hat{\beta}_T) \\
& + \sum_{e=2}^{T} \Big((1 - \hat{\pi}_{Me}) \hat{\pi}_{Se} \hat{\gamma}_{T,T-2+e} - \hat{\omega}_{Se} \hat{\gamma}_{T,e-1} \Big)
\end{aligned} \tag{4.12}$$

4.4 Estimates from MMRM model

The point estimator is straight forward. The data will be fitted based on the modified conditional MMRM model. Based on normality assumption, restricted maximum likelihood (ReML) will be construct and 'EM' algorithm will be used to calculate the estimate. The measurements are assumed to jointly follow a multivariate normal distribution. Denote $\hat{\boldsymbol{\eta}} = [\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}] = [\hat{\beta}_0, ..., \hat{\beta}_{3T-1}, \hat{\boldsymbol{\gamma}}_2, ..., \hat{\boldsymbol{\gamma}}_T]$, $\hat{\boldsymbol{\Sigma}}_\epsilon$ and the covariance matrix for the parameter

$$\hat{\Sigma}_\eta = \begin{bmatrix} \hat{\Sigma}_1 & \hat{\Sigma}_{12} \\ \hat{\Sigma}_{12} & \hat{\Sigma}_2 \end{bmatrix}$$

as the 'ReML' estimators where $\hat{\Sigma}_1$ is the variance matrix for $\hat{\boldsymbol{\beta}}$, $\hat{\Sigma}_2$ is the variance matrix for $\hat{\boldsymbol{\gamma}}$ and $\hat{\Sigma}_{12}$ is the covariance matrix for $[\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}]$. The intercurrent event probabilities are estimated using the observed proportion of patient in each pattern.

$$
\begin{aligned}
&\hat{\pi}_O = \frac{r}{N} \\
&\hat{\pi}_{Se} = \frac{q_e + p_e}{N} \\
&\hat{\pi}_{Me} = \frac{p_e}{q_e + p_e} \\
&\hat{\pi}_S = \frac{q+p}{N} \quad \text{where } q = \sum_{e=2}^{T} q_e, p = \sum_{e=2}^{T} p_e \\
&\hat{\pi}_M = \frac{p}{q+p} \\
&\hat{\omega}_O = \frac{l}{N} \\
&\hat{\omega}_{Se} = \frac{m_e + n_e}{N} \\
&\hat{\omega}_{Me} = \frac{n_e}{m_e + n_e} \\
&\hat{\omega}_S = \frac{m+n}{N} \quad \text{where } m = \sum_{e=2}^{T} m_e, n = \sum_{e=2}^{T} n_e \\
&\hat{\omega}_M = \frac{n}{m+n}
\end{aligned}
\tag{4.13}
$$

Note that $r+q+p = l+m+n = N$. The primary point estimator can be written as

$$
\hat{\theta}^{RES} = \hat{\beta}_1 + \hat{\beta}_{2T-1} + \sum_{e=2}^{T} \left(\frac{q_e + p_e}{N} \hat{\gamma}_{T,T-2+e} - \frac{m_e + n_e}{N} \hat{\gamma}_{T,e-1} \right) \tag{4.14}
$$

4.5 Approximate the inference for point estimator using delta method

The inference for primary point estimator as well as the sensitivity analysis estimator can be approximated by using delta method. For the primary point estimator,

$$\begin{aligned}
V(\hat{\theta}^{RES}) =& V\Big(E\big(\hat{\theta}^{RES}|\mathbf{\Omega}\big)\Big) + E\Big(V\big(\hat{\theta}^{RES}|\mathbf{\Omega}\big)\Big) \\
=& V\bigg(E\Big(\hat{\beta}_1 + \hat{\beta}_{2T-1} + \sum_{e=2}^{T}\big(\frac{q_e+p_e}{N}\hat{\gamma}_{T,T-2+e} - \frac{m_e+n_e}{N}\hat{\gamma}_{T,e-1}\big)|\mathbf{\Omega}\Big)\bigg) \\
&+ E\bigg(V\Big(\hat{\beta}_1 + \hat{\beta}_{2T-1} + \sum_{e=2}^{T}\big(\frac{q_e+p_e}{N}\hat{\gamma}_{T,T-2+e} - \frac{m_e+n_e}{N}\hat{\gamma}_{T,e-1}\big)|\mathbf{\Omega}\Big)\bigg) \\
=& V\bigg(\beta_1 + \beta_{2T-1} + \sum_{e=2}^{T}\big(\frac{q_e+p_e}{N}\gamma_{T,T-2+e} - \frac{m_e+n_e}{N}\gamma_{T,e-1}\big)\bigg) \\
&+ E\bigg(V\Big(\hat{\beta}_1 + \hat{\beta}_{2T-1} + \sum_{e=2}^{T}\big(\frac{q_e+p_e}{N}\hat{\gamma}_{T,T-2+e} - \frac{m_e+n_e}{N}\hat{\gamma}_{T,e-1}\big)|\mathbf{\Omega}\Big)\bigg)
\end{aligned} \tag{4.15}$$

where $\mathbf{\Omega} = [q_2, ..., q_T, p_2, ..., p_T, m_2, ..., m_T, n_2, ..., n_T]$. The first term can be written as

$$\begin{aligned}
& V\bigg(\beta_1 + \beta_{2T-1} + \sum_{e=2}^{T}\big(\frac{q_e+p_e}{N}\gamma_{T,T-2+e} - \frac{m_e+n_e}{N}\gamma_{T,e-1}\big)\bigg) \\
=& V\bigg(\sum_{e=2}^{T}\big(\frac{q_e+p_e}{N}\gamma_{T,T-2+e} - \frac{m_e+n_e}{N}\gamma_{T,e-1}\big)\bigg) \\
=& \frac{1}{N^2}\bigg(V\Big(\sum_{e=2}^{T}\big((q_e+p_e)\gamma_{T,T-2+e}\big)\Big) + V\Big(\sum_{e=2}^{T}\big((m_e+n_e)\gamma_{T,e-1}\big)\Big)\bigg) \\
=& \frac{1}{N^2}\bigg(\sum_{e=2}^{T}\Big(\gamma_{T,T-2+e}^2 N\pi_{Se}(1-\pi_{Se})\Big) - \sum_{f\neq q=2}^{T}\Big(N\gamma_{T,T-2+f}\gamma_{T,T-2+g}\pi_{Sf}\pi_{Sg}\Big) \\
&+ \sum_{e=2}^{T}\Big(\gamma_{T,e-1}^2 N\omega_{Se}(1-\omega_{Se})\Big) - \sum_{f\neq q=2}^{T}\Big(N\gamma_{T,f-1}\gamma_{T,g-1}\omega_{Sf}\omega_{Sg}\Big)\bigg) \\
=& \frac{1}{N}\bigg(\sum_{e=2}^{T}\Big(\gamma_{T,T-2+e}^2\pi_{Se}(1-\pi_{Se}) + \gamma_{T,e-1}^2\omega_{Se}(1-\omega_{Se})\Big) \\
&- \sum_{f\neq q=2}^{T}\Big(\gamma_{T,T-2+f}\gamma_{T,T-2+g}\pi_{Sf}\pi_{Sg} + \gamma_{T,f-1}\gamma_{T,g-1}\omega_{Sf}\omega_{Sg}\Big)\bigg)
\end{aligned} \tag{4.16}$$

It can also be represented using matrix format.

Denote $P1 = [\gamma_{T,T}, \gamma_{T,T+1}, ..., \gamma_{T,2T-2}]$, $P2 = [\gamma_{T,1}, \gamma_{T,2}, ..., \gamma_{T,T-1}]$,

$$\Sigma_{P1} = \begin{bmatrix} \pi_{S2}(1-\pi_{S2}) & -\pi_{S2}\pi_{S3} & ... & -\pi_{S2}\pi_{ST} \\ -\pi_{S3}\pi_{S2} & \pi_{S3}(1-\pi_{S3}) & ... & -\pi_{S3}\pi_{ST} \\ \vdots & \vdots & \ddots & \vdots \\ -\pi_{ST}\pi_{S2} & -\pi_{ST}\pi_{S3} & ... & \pi_{ST}(1-\pi_{ST}) \end{bmatrix}$$

$$\Sigma_{P2} = \begin{bmatrix} \omega_{S2}(1-\omega_{S2}) & -\omega_{S2}\omega_{S3} & ... & -\omega_{S2}\omega_{ST} \\ -\omega_{S3}\omega_{S2} & \omega_{S3}(1-\omega_{S3}) & ... & -\omega_{S3}\omega_{ST} \\ \vdots & \vdots & \ddots & \vdots \\ -\omega_{ST}\omega_{S2} & -\omega_{ST}\omega_{S3} & ... & \omega_{ST}(1-\omega_{ST}) \end{bmatrix}$$

The first term can thus be written using matrix as

$$V\Big(\beta_1+\beta_{2T-1}+\sum_{e=2}^{T}\big(\frac{q_e+p_e}{N}\gamma_{T,T-2+e}-\frac{m_e+n_e}{N}\gamma_{T,e-1}\big)\Big) = \frac{1}{N}\big(P1\Sigma_{P1}P1'+P2\Sigma_{P2}P2'\big) \tag{4.17}$$

For the second part, use the delta method to approximate the inference,

$$\begin{aligned} & V\Big(\hat{\beta}_1 + \hat{\beta}_{2T-1} + \sum_{e=2}^{T}\big(\frac{q_e+p_e}{N}\hat{\gamma}_{T,T-2+e} - \frac{m_e+n_e}{N}\hat{\gamma}_{T,e-1}\big)|\mathbf{\Omega}\Big) \\ \approx & \hat{\mathbf{B}}\mathbf{\Sigma}_{\eta'}\hat{\mathbf{B}}' \end{aligned} \tag{4.18}$$

where $\hat{\mathbf{B}}$ is the design vector $\hat{\mathbf{B}} = [1, 1, -\frac{m_2+n_2}{N}, ..., -\frac{m_T+n_T}{N}, \frac{q_2+p_2}{N}, ..., \frac{q_T+p_T}{N}]$ and $\mathbf{\Sigma}_{\eta'}$ is the partitioned covariance matrix from the 'ReML' parameter estimator covariance matrix. And

$$\begin{aligned} & E\Big(V\Big(\hat{\beta}_1 + \hat{\beta}_{2T-1} + \sum_{e=2}^{T}\big(\frac{q_e+p_e}{N}\hat{\gamma}_{T,T-2+e} - \frac{m_e+n_e}{N}\hat{\gamma}_{T,e-1}\big)|\mathbf{\Omega}\Big)\Big) \\ & \approx E\Big(\hat{\mathbf{B}}\mathbf{\Sigma}_{\eta'}\hat{\mathbf{B}}'\Big) \\ & \approx \mathbf{B}\mathbf{\Sigma}_{\eta'}\mathbf{B}' \end{aligned} \tag{4.19}$$

where $\mathbf{B} = [1, 1, -\omega_{S2}, ..., -\omega_{ST}, \pi_{S2}, ..., \pi_{ST}]$. The variance of the primary analysis estimator is therefore

$$V(\hat{\theta}^{RES}) \approx \frac{1}{N}\big(P1\Sigma_{P1}P1' + P2\Sigma_{P2}P2'\big) + \mathbf{B}\mathbf{\Sigma}_{\eta'}\mathbf{B}' \tag{4.20}$$

Plug in the point estimator to construct the variance estimator. That is for the primary analysis,

$$\hat{V}(\hat{\theta}^{RES}) \approx \frac{1}{N}\big(P1\hat{\Sigma}_{P1}P1' + P2\hat{\Sigma}_{P2}P2'\big) + \hat{\mathbf{B}}\hat{\mathbf{\Sigma}}_{\eta'}\hat{\mathbf{B}}' \tag{4.21}$$

where $\hat{\Sigma}_{P1} = \begin{bmatrix} \frac{(q_2+p_2)(N-q_2-p_2)}{N^2} & -\frac{(q_2+p_2)(q_3+p_3)}{N^2} & \cdots & -\frac{(q_2+p_2)(q_T+p_T)}{N^2} \\ -\frac{(q_2+p_2)(q_3+p_3)}{N^2} & \frac{(q_3+p_3)(N-q_3-p_3)}{N^2} & \cdots & -\frac{(q_3+p_3)(q_T+p_T)}{N^2} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{(q_2+p_2)(q_T+p_T)}{N^2} & -\frac{(q_3+p_3)(q_T+p_T)}{N^2} & \cdots & \frac{(q_T+p_T)(N-q_T-p_T)}{N^2} \end{bmatrix}$

$\hat{\Sigma}_{P1} = \begin{bmatrix} \frac{(m_2+n_2)(N-m_2-n_2)}{N^2} & -\frac{(m_2+n_2)(m_3+n_3)}{N^2} & \cdots & -\frac{(m_2+n_2)(m_T+n_T)}{N^2} \\ -\frac{(m_2+n_2)(m_3+n_3)}{N^2} & \frac{(m_3+n_3)(N-m_3-n_3)}{N^2} & \cdots & -\frac{(m_3+n_3)(m_T+n_T)}{N^2} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{(m_2+n_2)(m_T+n_T)}{N^2} & -\frac{(m_3+n_3)(m_T+n_T)}{N^2} & \cdots & \frac{(m_T+n_T)(N-m_T-n_T)}{N^2} \end{bmatrix}$

$\hat{\mathbf{B}} = [1, 1, -\frac{m_2+n_2}{N}, ..., -\frac{m_T+n_T}{N}, \frac{q_2+p_2}{N}, \frac{q_T+p_T}{N}]$ and $\hat{\mathbf{\Sigma}}_{\eta'}$ is from MMRM.

The point estimator and inference estimator can be derived for the RBI estimators using the same approach above. For J2Z,

$$\hat{\theta}^{J2Z} = (1 - \frac{p}{N})(\hat{\beta}_1 + \hat{\beta}_{2T-1}) - \frac{p}{N}(\hat{\beta}_0 + \hat{\beta}_T) + \sum_{e=2}^{T}\Big(\frac{q_e}{N}\hat{\gamma}_{T,T-2+e} - \frac{m_e + n_e}{N}\hat{\gamma}_{T,e-1}\Big) \tag{4.22}$$

$$\hat{V}(\hat{\theta}^{J2Z}) \approx \frac{1}{N}\big(P1^{J2Z}\hat{\Sigma}_{P1}^{J2Z}P1^{J2Z'} + P2\hat{\Sigma}_{P2}P2\big) + \hat{\mathbf{B}}^{J2Z}\hat{\mathbf{\Sigma}}_{\eta'}\hat{\mathbf{B}}^{J2Z'} \tag{4.23}$$

where $P1^{J2Z'} = \begin{bmatrix} -\hat{\beta}_1 - \hat{\beta}_{2T-1} - \hat{\beta}_0 - \hat{\beta}_T \\ +\hat{\gamma}_{T,T} \\ +\hat{\gamma}_{T,T+1} \\ \vdots \\ \hat{\gamma}_{T,2T-2} \end{bmatrix}$

$$\hat{\Sigma}_{P1}^{J2Z} = \begin{bmatrix} \frac{p(N-p)}{N^2} & -\frac{pq_2}{N^2} & -\frac{pq_3}{N^2} & \cdots & -\frac{pq_T}{N^2} \\ -\frac{pq_2}{N^2} & \frac{q_2(N-q_2)}{N^2} & -\frac{q_2q_3}{N^2} & \cdots & -\frac{q_2q_T}{N^2} \\ -\frac{pq_3}{N^2} & -\frac{q_2q_3}{N^2} & \frac{q_3(N-q_3)}{N^2} & \cdots & -\frac{q_3q_T}{N^2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -\frac{pq_T}{N^2} & -\frac{q_2q_T}{N^2} & -\frac{q_3q_T}{N^2} & \cdots & \frac{q_T(N-q_T)}{N^2} \end{bmatrix}$$

$\hat{\mathbf{B}}^{J2Z'} = [-\frac{p}{N}, 1-\frac{p}{N}, ...(\text{T 0's})..., -\frac{p}{N}, ...(\text{T 0's})..., 1-\frac{p}{N}-\frac{m_2+n_2}{N}, ..., -\frac{m_T+n_T}{N}, -\frac{q_2}{N}, ..., -\frac{q_T}{N}]$

For J2R,

$$\hat{\theta}^{J2R} = \hat{\beta}_1 + \hat{\beta}_{2T-1} + \sum_{e=2}^{T} \Big(\frac{q_e + p_e}{N} \hat{\gamma}_{T,T-2+e} - \frac{m_e + n_e}{N} \hat{\gamma}_{T,e-1} + \frac{p_e}{N} (-\hat{\beta}_1 - \hat{\beta}_{2T-1} - \hat{\gamma}_{T,T-2+e}) \Big) \tag{4.24}$$

$$\hat{V}(\hat{\theta}^{J2R}) \approx \frac{1}{N} \big(P1^{J2R} \hat{\Sigma}_{P1}^{J2R} P1^{J2R'} + P2 \hat{\Sigma}_{P2} P2 \big) + \hat{\mathbf{B}}^{J2R} \hat{\mathbf{\Sigma}}_{\eta'} \hat{\mathbf{B}}^{J2R'} \tag{4.25}$$

where $P1^{J2R} = \begin{bmatrix} -\hat{\beta}_1 - \hat{\beta}_{2T-1} & -\hat{\beta}_1 - \hat{\beta}_{2T-1} & \cdots & -\hat{\beta}_1 - \hat{\beta}_{2T-1} & \hat{\gamma}_{T,T} & \hat{\gamma}_{T,T+1} & \cdots & \hat{\gamma}_{T,2T-2} \end{bmatrix}$

$$\hat{\Sigma}_{P1}^{J2R} = \begin{bmatrix} \frac{p_2(1-p_2)}{N^2} & -\frac{p_2p_3}{N^2} & \cdots & -\frac{p_2p_T}{N^2} & -\frac{p_2q_2}{N^2} & -\frac{p_2q_3}{N^2} & \cdots & -\frac{p_2q_T}{N^2} \\ -\frac{p_2p_3}{N^2} & \frac{p_3(1-p_3)}{N^2} & \cdots & -\frac{p_3p_T}{N^2} & -\frac{p_3q_2}{N^2} & -\frac{p_3q_3}{N^2} & \cdots & -\frac{p_3q_T}{N^2} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ -\frac{p_2p_T}{N^2} & -\frac{p_3p_T}{N^2} & \cdots & \frac{p_T(1-p_T)}{N^2} & -\frac{p_Tq_2}{N^2} & -\frac{p_Tq_3}{N^2} & \cdots & -\frac{p_Tq_T}{N^2} \\ -\frac{p_2q_2}{N^2} & -\frac{p_2q_3}{N^2} & \cdots & -\frac{p_2q_T}{N^2} & \frac{q_2(1-q_2)}{N^2} & -\frac{q_2q_3}{N^2} & \cdots & -\frac{q_2q_T}{N^2} \\ -\frac{p_3q_2}{N^2} & -\frac{p_3q_3}{N^2} & \cdots & -\frac{p_3q_T}{N^2} & \frac{q_2q_2}{N^2} & -\frac{q_3(1-q_3)}{N^2} & \cdots & -\frac{q_2q_T}{N^2} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ -\frac{p_Tq_2}{N^2} & -\frac{p_Tq_3}{N^2} & \cdots & -\frac{p_Tq_T}{N^2} & -\frac{q_2q_T}{N^2} & -\frac{q_3q_T}{N^2} & \cdots & \frac{q_T(1-q_T)}{N^2} \end{bmatrix}$$

$\hat{\mathbf{B}}^{J2R} = [0, 1 - \frac{\sum p_e}{N}, ...(\text{2T 0's})..., 1 - \frac{\sum p_e}{N}, -\frac{m_2+n_2}{N}, ..., -\frac{m_T+n_T}{N}, \frac{q_2}{N}, ..., \frac{q_T}{N}]$

4.6 Simulation study using parallel computing to verify delta approximate method

In order to verify the validity of point estimator based on modified MMRM model and delta approximation approach for the inference proposed in this chapter, 5000 simulation studies were conducted with the setup borrowed from Liu and Pang [43]. For the purpose of demonstration, primary analysis and sensitivity analysis with 'Jump to zero' method and 'Jump to reference' were performed. 'Jump to zero' method is studied under both null scenario and alternative scenario. 'Jump to reference' method is studied only under alternative scenario because under null scenario, reference arm has same effect as treatment arm.

4.6.1 Basic Setup

- Number of patients N=100 for each arm.
- Total 4 measurement points, t_1, t_2, t_3, t_4
- Reference arm effect mean: (1.3, 2.3, 3.2, 4)
- Treatment arm effect mean: (1.3, 2.3, 3.2, 4) (assumed to be the same as reference arm under null hypothesis)
- Reference arm switch to rescue medication effect mean: (1.3, 1.15, 1.6, 2)
- Treatment arm switch to rescue medication effect mean: (1.3, 1.15, 1.6, 2) (assumed to be the same as reference group under null hypothesis)
- Standard deviations identical among two groups: (1.8, 2.0, 2.1, 2.2)

- Correlation matrix $= \begin{bmatrix} 1 & 0.7 & 0.5 & 0.2 \\ 0.7 & 1 & 0.6 & 0.4 \\ 0.5 & 0.6 & 1 & 0.5 \\ 0.2 & 0.4 & 0.5 & 1 \end{bmatrix}$

Denote $\psi_{s,g,k}$ the kth parameter for event type s (s=2 represents observed patients who switch to rescue medication; s=3 represents missing patients) and group g. The probability of intercurrent event at t_j for treatment arm patients ($G_i = 1$) is modeled as

$$logit(\pi_{Sj}) = \psi_{2,1,0} + \psi_{2,1,1} y_{i,j-1} + \psi_{2,1,2} y'_{i,j} \tag{4.26}$$

and the probability of missing within these event patients is modeled as

$$logit(\pi_{Mj}) = \psi_{3,1,0} + \psi_{3,1,1} y_{i,j-1} + \psi_{3,1,2} y_{i,j} \tag{4.27}$$

Denote $y'_{i,j} = y_{i,j}$ if the patient did not switch to rescue medication. Otherwise it would be a hypothetical value that is not observable. The probability of intercurrent event at t_j for treatment arm patients ($G_i = 0$) is modeled as

$$logit(\omega_{Sj}) = \psi_{2,0,0} + \psi_{2,0,1} y_{i,j-1} + \psi_{2,0,2} y'_{i,j} \tag{4.28}$$

and the probability of missing within these event patients is modeled as

$$logit(\omega_{Mj}) = \psi_{3,0,0} + \psi_{3,0,1} y_{i,j-1} + \psi_{3,0,2} y_{i,j} \tag{4.29}$$

Notice that based on the definition of different missing mechanisms, if $\psi_{2,1,1}$, $\psi_{2,1,2}$, $\psi_{3,1,1}$, $\psi_{3,1,2}$, $\psi_{2,0,1}$, $\psi_{2,0,2}$, $\psi_{3,0,1}$, $\psi_{3,0,2}$ all equal zero, then the missing mechanism is MCAR. If $\psi_{2,1,1}$, $\psi_{3,1,1}$, $\psi_{2,0,1}$, $\psi_{3,0,1}$ not equal 0 and $\psi_{2,1,2}$, $\psi_{3,1,2}$, $\psi_{2,0,2}$, $\psi_{3,0,2}$ equal zero, then the missing mechanism is MAR. If $\psi_{2,1,1}$, $\psi_{2,1,2}$, $\psi_{3,1,1}$, $\psi_{3,1,2}$, $\psi_{2,0,1}$, $\psi_{2,0,2}$, $\psi_{3,0,1}$, $\psi_{3,0,2}$ all not equal zero, then the missing mechanism is MNAR.

4.6.2 Switch and dropout scenarios

The null hypothesis is two arm mean are the same at t_4, regardless of switching to rescue medication. I will investigate following null and alternative scenarios:

1. Null scenario: Treatment and reference arm means are the same ($\mu_t = \mu_p$), probability of switch to rescue medication are the same ($\pi_{S4} = \omega_{S4}$).

2. Alternative scenario: Treatment and reference arm means are different ($\mu_t \neq \mu_p$), probability of switch to rescue medication are different ($\pi_{S4} \neq \omega_{S4}$). And the mixture means are also different ($\pi_{S4}\mu_t \neq \omega_{S4}\mu_p$). Under this scenario, the treatment arm mean effect is set as (1.0, 1.0, 1.0, 1.0) which indicates no effect at all and the treatment arm switch to rescue medication effect mean is set as (1.3, 1.1, 1.3, 1.5)

For simplicity purpose, patients who completed on initial treatment will be mentioned as the 'complete patients', patients who switched to rescue medication but stayed in the study will be mentioned as the 'switch patients' and patients who discontinued from the initial treatment will be mentioned as the 'dropout patients', the switching rate is controlled at 50% and missing rate for treatment arm patients at 20% vs. missing rate for reference arm patients at 30%. Both MAR and MNAR scenarios are considered. Within each missing mechanism scenario, both high or low correlation between missing mechanism and the measurements will be investigated. The detail settings are as below,

Table 4.1: Parameter specification for different underlying dropout simulation scenarios under null hypothesis

Parameter		1. MAR	2. MAR	3. MNAR	4. MNAR
Reference	$\psi_{2,0,0}$	-0.6	-0.6	-0.6	-0.6
	$\psi_{2,0,1}$	-0.4	-0.4	-0.4	-0.4
	$\psi_{2,0,2}$	0	0	0	0
	$\psi_{3,0,0}$	0.2	0.3	0.3	0.3
	$\psi_{3,0,1}$	0.4	0.2	0	0
	$\psi_{3,0,2}$	0	0	0.4	0.2
Treatment	$\psi_{2,1,0}$	-0.6	-0.6	-0.6	-0.6
	$\psi_{2,1,1}$	-0.4	-0.4	-0.4	-0.4
	$\psi_{2,1,2}$	0	0	0	0
	$\psi_{3,1,0}$	-0.1	-0.2	-0.1	-0.2
	$\psi_{3,1,1}$	-0.4	-0.2	0	0
	$\psi_{3,1,2}$	0	0	-0.4	-0.2

Simulations are also conducted under the following alternative scenarios

Table 4.2: Parameter specification for different underlying dropout simulation scenarios under alternative hypothesis

Parameter		1. MAR	2. MAR	3. MNAR	4. MNAR
Reference	$\psi_{2,0,0}$	-1.0	-1.0	-1.0	-1.0
	$\psi_{2,0,1}$	-0.4	-0.4	-0.4	-0.4
	$\psi_{2,0,2}$	0	0	0	0
	$\psi_{3,0,0}$	0.3	0.4	0.3	0.2
	$\psi_{3,0,1}$	0.4	0.2	0	0
	$\psi_{3,0,2}$	0	0	0.4	0.2
Treatment	$\psi_{2,1,0}$	-0.6	-0.6	-0.6	-0.6
	$\psi_{2,1,1}$	-0.4	-0.4	-0.4	-0.4
	$\psi_{2,1,2}$	0	0	0	0
	$\psi_{3,1,0}$	-0.1	-0.2	-0.1	-0.2
	$\psi_{3,1,1}$	-0.4	-0.2	0	0
	$\psi_{3,1,2}$	0	0	-0.4	-0.2

4.6.3 Point estimators and corresponding inference from MMRM model

The model for this specific setup can be written as

$$\begin{aligned}
&Y_{i,j}|S_i, E_i, G_i \\
&= \beta_0 + \beta_1 G_i + \beta_2 I_{j=2} + \beta_3 I_{j=3} + \beta_4 I_{j=4} + \beta_5 I_{j=2} G_i + \beta_6 I_{j=3} G_i + \beta_7 I_{j=4} G_i \\
&+ \gamma_{2,1} I_{j=2,S_i=2,E_i=2,G_i=0} + \gamma_{2,2} I_{j=2,S_i=2,E_i=2,G_i=1} \\
&+ \gamma_{3,1} I_{j=3,S_i=2,E_i=2,G_i=0} + \gamma_{3,2} I_{j=3,S_i=2,E_i=3,G_i=0} \\
&+ \gamma_{3,3} I_{j=3,S_i=2,E_i=2,G_i=1} + \gamma_{3,4} I_{j=3,S_i=2,E_i=3,G_i=1} \\
&+ \gamma_{4,1} I_{j=4,S_i=2,E_i=2,G_i=0} + \gamma_{4,2} I_{j=4,S_i=2,E_i=3,G_i=0} + \gamma_{4,3} I_{j=4,S_i=2,E_i=4,G_i=0} \\
&+ \gamma_{4,4} I_{j=4,S_i=2,E_i=2,G_i=1} + \gamma_{4,5} I_{j=4,S_i=2,E_i=3,G_i=1} + \gamma_{4,6} I_{j=4,S_i=2,E_i=4,G_i=1} \\
&+ \epsilon_{i,j}
\end{aligned} \tag{4.30}$$

The primary analysis point estimators and corresponding variance is

$$\hat{\theta}^{RES} = \hat{\beta}_1 + \hat{\beta}_7 + \sum_{e=2}^{4} \left(\frac{q_e + p_e}{N} \hat{\gamma}_{4,2+e} - \frac{m_e + n_e}{N} \hat{\gamma}_{4,e-1} \right) \tag{4.31}$$

$$\begin{aligned}
&V(\hat{\theta}^{RES}) \\
\approx &\frac{1}{N} \left(\sum_{e=2}^{4} \left(\gamma_{4,2+e}^2 \pi_{Se}(1 - \pi_{Se}) + \gamma_{4,e-1}^2 N \omega_{Se}(1 - \omega_{Se}) \right) \right. \\
&\left. - \sum_{f \neq q=2}^{4} \left(\gamma_{4,2+f} \gamma_{4,2+g} \pi_{Sf} \pi_{Sg} + \gamma_{4,f-1} \gamma_{4,g-1} \omega_{Sf} \omega_{Sg} \right) \right) + \mathbf{B} \mathbf{\Sigma}_{\eta'} \mathbf{B}'
\end{aligned} \tag{4.32}$$

where $\hat{\mathbf{B}} = [1, 1, -\frac{m_2+n_2}{N}, -\frac{m_3+n_3}{N}, -\frac{m_4+n_4}{N}, \frac{q_2+p_2}{N}, \frac{q_3+p_3}{N}, \frac{q_4+p_4}{N}]$ and $\mathbf{\Sigma}_{\eta'}$ is the covariance matrix corresponding to the $\hat{\beta}_1, \hat{\beta}_7, \hat{\gamma}_{4,1}, \hat{\gamma}_{4,2}, \hat{\gamma}_{4,3}, \hat{\gamma}_{4,4}, \hat{\gamma}_{4,5}, \hat{\gamma}_{4,6}$

For J2R,

$$\hat{\theta}^{J2R} = \hat{\beta}_1 + \hat{\beta}_7 + \sum_{e=2}^{4} \left(\frac{q_e + p_e}{N} \hat{\gamma}_{4,2+e} - \frac{m_e + n_e}{N} \hat{\gamma}_{4,e-1} + \frac{p_e}{N} (-\hat{\beta}_1 - \hat{\beta}_7 - \hat{\gamma}_{4,2+e}) \right) \tag{4.33}$$

$$\hat{V}(\hat{\theta}^{J2R}) \approx \frac{1}{N} \left(P1^{J2R} \hat{\Sigma}_{P1}^{J2R} P1^{J2R'} + P2 \hat{\Sigma}_{P2} P2 \right) + \hat{\mathbf{B}}^{J2R} \hat{\mathbf{\Sigma}}_{\eta'} \hat{\mathbf{B}}^{J2R'} \tag{4.34}$$

For J2Z,

$$\hat{\theta}^{J2Z} =(1-\frac{p}{N})(\hat{\beta}_1+\hat{\beta}_{2T-1})-\frac{p}{N}(\hat{\beta}_0+\hat{\beta}_T)+\sum_{e=2}^{T}\Big(\frac{q_e}{N}\hat{\gamma}_{T,T-2+e}-\frac{m_e+n_e}{N}\hat{\gamma}_{T,e-1}\Big) \tag{4.35}$$

$$\hat{V}(\hat{\theta}^{J2Z})\approx\frac{1}{N}\big(P1^{J2Z}\hat{\Sigma}_{P1}^{J2Z}P1^{J2Z'}+P2\hat{\Sigma}_{P2}P2\big)+\hat{\mathbf{B}}^{J2Z}\hat{\mathbf{\Sigma}}_{\eta'}\hat{\mathbf{B}}^{J2Z'} \tag{4.36}$$

4.7 The primary and RBI estimator using MI approach

MI approach can also be used to estimate the primary and RBI estimators. Suppose there is

$$[Y_{i,e},Y_{i,e+1},...,Y_{i,T}|y_{i,1},...,y_{i,e-1},G_i=0,S_i=3,E_i=e]\sim MVN\Bigg(\hat{\boldsymbol{\mu}},\mathbf{\Sigma}\Bigg) \tag{4.37}$$

where $\hat{\boldsymbol{\mu}}=\begin{bmatrix}\hat{\mu}_{i,e,0}\\ \hat{\mu}_{i,e+1,0}\\ \vdots\\ \hat{\mu}_{i,T,0}\end{bmatrix}+\hat{\mathbf{\Sigma}}_{e:T,1:e-1}\hat{\mathbf{\Sigma}}_{1:e-1,1:e-1}^{-1}\begin{bmatrix}y_{i,1}-\hat{\mu}_{i,1,0}\\ y_{i,2}-\hat{\mu}_{i,2,0}\\ \vdots\\ y_{i,e-1}-\hat{\mu}_{i,e-1,0}\end{bmatrix}$

$\mathbf{\Sigma}=\hat{\mathbf{\Sigma}}_{1:e-1,1:e-1}-\hat{\mathbf{\Sigma}}'_{e:T,1:e-1}\hat{\mathbf{\Sigma}}_{e:T,e:T}^{-1}\hat{\mathbf{\Sigma}}_{1:e-1,e:T}$. For reference arm, the conditional mean estimators are

$$\begin{aligned}
&\hat{\mu}_{i,1,0}=\hat{\beta}_0\\
&\hat{\mu}_{i,2,0}=\hat{\beta}_0+\hat{\beta}_2\\
&\vdots\\
&\hat{\mu}_{i,e-1,0}=\hat{\beta}_0+\hat{\beta}_{e-1}\\
&\hat{\mu}_{i,e,0}=\hat{\beta}_0+\hat{\beta}_e+\hat{\gamma}_{e,e-1}\\
&\vdots\\
&\hat{\mu}_{i,T,0}=\hat{\beta}_0+\hat{\beta}_T+\hat{\gamma}_{T,T-1}
\end{aligned} \tag{4.38}$$

For active treatment arm patients, for primary estimator the conditional mean estimators are

$$\begin{aligned}
&\hat{\mu}_{i,1,1} = \hat{\beta}_0 + \hat{\beta}_1 \\
&\hat{\mu}_{i,2,1} = \hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_2 + \hat{\beta}_{T+1} \\
&\vdots \\
&\hat{\mu}_{i,e-1,1} = \hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_{e-1} + \hat{\beta}_{T+e-2} \\
&\hat{\mu}_{i,e,1} = \hat{\beta}_0 + \hat{\beta}_e + \hat{\beta}_{T+e-1} + \hat{\gamma}_{e,2e-2} \\
&\vdots \\
&\hat{\mu}_{i,T,1} = \hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_T + \hat{\beta}_{2T-1} + \hat{\gamma}_{T,2T-2}
\end{aligned} \tag{4.39}$$

The following conditional mean estimators are used for active arm under 'Jump to zero' method.

$$\begin{aligned}
&\hat{\mu}_{i,1,1} = \hat{\beta}_0 + \hat{\beta}_1 \\
&\hat{\mu}_{i,2,1} = \hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_2 + \hat{\beta}_{T+1} \\
&\vdots \\
&\hat{\mu}_{i,e-1,1} = \hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_{e-1} + \hat{\beta}_{T+e-2} \\
&\hat{\mu}_{i,e,1} = 0 \\
&\vdots \\
&\hat{\mu}_{i,T,1} = 0
\end{aligned} \tag{4.40}$$

The following conditional mean estimators are used for active arm under 'Jump to reference' method.

$$\begin{aligned}
\hat{\mu}_{i,1,1} &= \hat{\beta}_0 + \hat{\beta}_1 \\
\hat{\mu}_{i,2,1} &= \hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_2 + \hat{\beta}_{T+1} \\
&\vdots \\
\hat{\mu}_{i,e-1,1} &= \hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_{e-1} + \hat{\beta}_{T+e-2} \\
\hat{\mu}_{i,e,1} &= \hat{\beta}_0 + \hat{\beta}_e + \hat{\gamma}_{e,e-1} \\
&\vdots \\
\hat{\mu}_{i,T,1} &= \hat{\beta}_0 + \hat{\beta}_T + \hat{\gamma}_{T,T-1}
\end{aligned} \tag{4.41}$$

Use the MMRM covariance estimate $\hat{\mathbf{\Sigma}}_y$ as the variance covariance matrix. Thus, the detail steps for MI based on joint multivariate modeling approach are

1. Fit the imputation conditional MMRM model using observed data.
2. Impute missing data based on the conditional multivariate normal model constructed using adjusted decay sensitivity analysis conditional mean.
3. Fit the analysis MMRM model using the imputed complete dataset.
4. Record the decay estimator and corresponding inference as $\hat{Q}^{(w)}, \hat{V}^{(w)}$ for the wth imputation.
5. Repeat step 1-4 for W times. W is suggested to be around 10 times.
6. Combine the W estimates using Rubin's combining rule.

$$\begin{aligned}
\bar{Q} &= W^{-1} \sum \hat{Q}^{(w)} \\
T &= (1 + W^{-1})(W - 1)^{-1} \sum (\hat{Q}^{(w)} - \bar{Q})^2 + \bar{V}
\end{aligned} \tag{4.42}$$

4.7.1 Result summary based on 5000 simulations

Table 4.3 and Table 4.4 are the summaries of the 5000 simulated study results under null hypothesis and alternative hypothesis respectively. For the reported parameter statistics, $mean(\hat{\theta})$ is the mean of the sample of 5000 simulated study point estimators. $s.e$ is the empirical standard error of $mean(\hat{\theta})$ derived from the 5000 simulation point estimators. $mean\ se(\hat{\theta})$ is the mean of the analytic estimated standard error for $\hat{\theta}$ from the simulations. $s.e$ and $mean\ se(\hat{\theta})$ should agree with each other. *coverage* is the proportion of the confidence intervals from each simulation that covers the 'true' underlying treatment difference which is equal to 0 under null hypothesis. *Type* 1 *error* is the proportion of the simulations that falsely reject the null hypothesis when they are simulated under null hypothesis. *Power* is the proportion of the simulations that correctly reject the null hypothesis when it is simulated under alternative hypothesis. The simulation results suggest that both methods return the identical point estimator as expected. Under the null scenario with missing at random (#1 MAR), the point estimators estimate the effect difference correctly. This verifies the modified MMRM model point estimator construction. The inference estimator provided by the delta approximation method agrees with the true underlying value. This is the proof that delta approximate estimates the standard error correctly. However, the multiple imputation over estimates the standard error while using the 'J2Z' as the sensitivity analysis estimator. This problem is well known as the uncongenial problem mentioned by Xiao-Li Meng [53], Kaifeng Lu [45] and Liu and Pang [43]. Under the null scenario with missing not at random (#1 MNAR), the point estimator is biased because the primary (RES) estimator is based on the wrong assumption that patients who dropped out would have the same mean effect as those who switched to rescue medication. Under MNAR, the mean effects are actually different. The standard errors are still correctly estimated using the delta approximate method.

Table 4.3: Summary result based on 5000 simulation studies under null hypothesis

Scenario	Parameter	$\hat{\theta}^{RES}$		$\hat{\theta}^{J2Z}$	
		Approx	MI	Approx	MI
#1 MAR High	$mean(\hat{\theta})$	-0.02	-0.03	-1.43	-1.43
	s.e.	0.44	0.44	0.53	0.54
	mean $se(\hat{\theta})$	0.42	0.45	0.51	0.57
	coverage	94%	95%	21%	28%
	Type I error	2.20%	2.00%	0.00%	0.00%
#2 MAR Low	$mean(\hat{\theta})$	0.02	0.02	-1.37	-1.37
	s.e.	0.43	0.44	0.52	0.52
	mean $se(\hat{\theta})$	0.42	0.44	0.50	0.56
	coverage	94%	95%	22%	30%
	Type I error	3.50%	3.00%	0.00%	0.00%
#3 MNAR High	$mean(\hat{\theta})$	0.24	0.23	-1.19	-1.19
	s.e.	0.45	0.46	0.54	0.55
	mean $se(\hat{\theta})$	0.43	0.45	0.51	0.57
	coverage	91%	92%	37%	46%
	Type I error	8.80%	6.80%	0.00%	0.00%
#4 MNAR Low	$mean(\hat{\theta})$	0.13	0.14	-1.25	-1.25
	s.e.	0.43	0.43	0.51	0.51
	mean $se(\hat{\theta})$	0.42	0.44	0.50	0.56
	coverage	95%	95%	31%	39%
	Type I error	4.60%	4.20%	0.00%	0.00%

Table 4.4: Summary result based on 5000 simulation studies under alternative hypothesis

Scenario	Parameter	$\hat{\theta}^{RES}$		$\hat{\theta}^{J2R}$		$\hat{\theta}^{J2Z}$	
		Approx	MI	Approx	MI	Approx	MI
#1 MAR High	$mean(\hat{\theta})$	1.72	1.72	1.62	1.62	0.30	0.30
	s.e.	0.39	0.40	0.31	0.33	0.49	0.50
	$mean\ se(\hat{\theta})$	0.41	0.42	0.41	0.43	0.50	0.55
	coverage	96%	95%	98%	98%	17%	15%
	Power	99%	99%	99%	99%	9.6%	7.3%
#2 MAR Low	$mean(\hat{\theta})$	1.75	1.75	1.64	1.64	0.36	0.36
	s.e.	0.41	0.42	0.34	0.34	0.52	0.53
	$mean\ se(\hat{\theta})$	0.41	0.42	0.41	0.43	0.49	0.55
	coverage	95%	95%	98%	98%	25%	30%
	Power	99%	99%	99%	99%	13.1%	10.4%
#3 MNAR High	$mean(\hat{\theta})$	2.04	2.04	1.84	1.84	0.63	0.63
	s.e.	0.41	0.42	0.34	0.35	0.50	0.51
	$mean\ se(\hat{\theta})$	0.41	0.42	0.41	0.43	0.50	0.55
	coverage	88%	89%	97%	98%	41%	49%
	Power	100%	100%	100%	100%	24.4%	20.2%
#4 MNAR Low	$mean(\hat{\theta})$	1.88	1.87	1.72	1.72	0.41	0.42
	s.e.	0.41	0.41	0.32	0.34	0.49	0.49
	$mean\ se(\hat{\theta})$	0.40	0.42	0.41	0.42	0.49	0.55
	coverage	93%	95%	99%	99%	22%	31%
	Power	100%	100%	100%	100%	12.8%	9.8%

Chapter 5
Decay model tipping point sensitivity analysis

5.1 Decay model: a flexible sensitivity analysis estimator

In a real world study, the investigators might consider the primary analysis assumption under MAR to be too optimistic while the 'Jump to zero' assumption to be too conservative. Furthermore, the limited scenarios of RBI might not be sufficient enough to investigate the robustness of the primary analysis assumption. The choice of a sensitivity analysis with an option of a continuously adjustable sensitivity parameters is narrowed down to only one available method, which is the δ-adjustment tipping point analysis. However, as briefly mentioned in the introduction chapter, there are several drawbacks of this method. To name a few, a constant adjustment is implemented at the dropout time point regardless of the dropout time point. There is no distinction between a patient who dropped out in the early stage of the study versus a patient who dropped out close to the end of the study. There is also no unified magnitude for the δ because it depends on the magnitude of the measurement itself. Therefore, a tipping point δ can not be directly used to determine the robustness of the assumption. It needs further investigation case by case. Because of the ambiguity of the magnitude of the δ, one might accidentally over adjust the effect to a unrealistic range and comes up with a tipping point with clinically impossible value. In order to address the problems addressed above, I introduce a method that uses an exponential function $\exp\big(-\phi \cdot \Delta t\big)$ to model the decay effect from the patient's dropout time point to the end of study period. The ϕ here serves as a sensitivity parameter that

represents the decay rate between two measurement time points. Denote Δt_e as the time period from event time point e to endpoint T, $\Delta t_e = T - e + 1$. The deviate effect for the last time point $\hat{\gamma}_{T,3T-2}, ..., \hat{\gamma}_{T,3T-4+e}, ..., \hat{\gamma}_{T,4T-4}$ can be written as

$$\begin{aligned}
\hat{E}(Y_{i,T}|S_i = 3, E_i = e, G_i = 1) =& \Big(\hat{E}(Y_{i,T}|S_i = 2, E_i = e, G_i = 1)\Big) \exp\left(-\phi \cdot \Delta t_e\right) \\
\hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_T + \hat{\beta}_{2T-1} + \hat{\gamma}_{3T-4+e} =& \big(\hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_T + \hat{\beta}_{2T-1} + \hat{\gamma}_{T,T-2+e}\big) \exp\left(-\phi \cdot \Delta t_e\right) \\
\hat{\gamma}_{3T-4+e} =& \big(\hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_T + \hat{\beta}_{2T-1}\big)\big(\exp\left(-\phi \cdot \Delta t_e\right) - 1\big) + \\
& \hat{\gamma}_{T,T-2+e} \exp\left(-\phi \cdot \Delta t_e\right)
\end{aligned} \tag{5.1}$$

where $e = 2, ..., T$. Comparing to the δ-adjustment tipping point method, by using the decay model, the adjustment from the primary analysis assumption changes with dropout time. Patients who dropped out in the early stage of the study will have a larger decay from the primary analysis assumption comparing to those who dropped out at the later stage of the study. Although the absolute value of the decay effect is different among different dropout time point patients, the rate of the decay is assumed to be the same ϕ. It can be interpreted as a proportion of the effect under primary assumption and not affected by the magnitude of the measurement. The decay sensitivity analysis parameter has a unified scale and thus can be compared among different studies and the sense of robustness would be more straight forward.

5.2 Relationship between decay model tipping point and various RBI scenarios

Another practical feature of decay model tipping point method is that all of the assumption scenarios including primary analysis, jump to reference scenario and jump to zero scenario can all be represented by decay model with corresponding ϕ's. To show in detail, denote the decay model sensitivity analysis estimator (abbreviated as

decay estimator) as $\hat{\theta}^{DCY}$.

$$\begin{aligned}\hat{\theta}^{DCY} =&\hat{\beta}_1 + \hat{\beta}_{2T-1} + \sum_{e=2}^{T}\Bigg((1-\hat{\pi}_{Me})\hat{\pi}_{Se}\hat{\gamma}_{T,T-2+e} - \hat{\omega}_{Se}\hat{\gamma}_{T,e-1} \\ &+ \hat{\pi}_{Me}\hat{\pi}_{Se}\Big(\big(\hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_T + \hat{\beta}_{2T-1}\big)\Big(\exp\left(-\phi\cdot\Delta t_e\right) - 1\Big) \\ &+ \hat{\gamma}_{T,T-2+e}\exp\left(-\phi\cdot\Delta t_e\right)\Big)\Bigg)\end{aligned} \tag{5.2}$$

When $\phi = +\infty$,

$$\begin{aligned}\hat{\gamma}_{T,3T-4+e} =&\big(\hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_T + \hat{\beta}_{2T-1}\big)\Big(\exp\big(-\infty\cdot\Delta t_e\big) - 1\Big) + \hat{\gamma}_{T,T-2+e}\exp\big(-\infty\cdot\Delta t_e\big) \\ =&-\big(\hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_T + \hat{\beta}_{2T-1}\big)\end{aligned} \tag{5.3}$$

which is equivalent to the 'J2Z' scenario.

When $\phi = 0$,

$$\begin{aligned}\hat{\gamma}_{T,3T-4+e} =& \big(\hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_T + \hat{\beta}_{2T-1}\big)\Big(\exp\big(-0\cdot\Delta t_e\big) - 1\Big) + \hat{\gamma}_{T,T-2+e}\exp\big(-0\cdot\Delta t_e\big) \\ =& \hat{\gamma}_{T,T-2+e}\end{aligned} \tag{5.4}$$

which is equivalent to the primary analysis scenario. For 'J2R' scenario, the relationship is not straight forward. To find the ϕ that returns an equivalent result as the 'J2R' scenario, that is to set $\hat{\theta}^{J2R} = \hat{\theta}^{DCY}$, the following equation is used

$$\begin{aligned}&\sum_{e=2}^{T}\Big(\hat{\pi}_{Se}\hat{\gamma}_{T,T-2+e} + \hat{\pi}_{Me}\hat{\pi}_{Se}(-\hat{\beta}_1 - \hat{\beta}_{2T-1} - \hat{\gamma}_{T,T-2+e})\Big) \\ =&\sum_{e=2}^{T}\Bigg((1-\hat{\pi}_{Me})\hat{\pi}_{Se}\hat{\gamma}_{T,T-2+e} + \hat{\pi}_{Me}\hat{\pi}_{Se}\Big(\big(\hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_T + \hat{\beta}_{2T-1}\big)\Big(\exp\big(-\phi\cdot\Delta t_e\big) - 1\Big) \\ &+ \hat{\gamma}_{T,T-2+e}\exp\big(-\phi\cdot\Delta t_e\big)\Big)\Bigg)\end{aligned} \tag{5.5}$$

The decay parameter ϕ can be solved using numerical method such as Newton-Raphson algorithm. Figure 5.1 is used as a brief demonstration of the idea of the

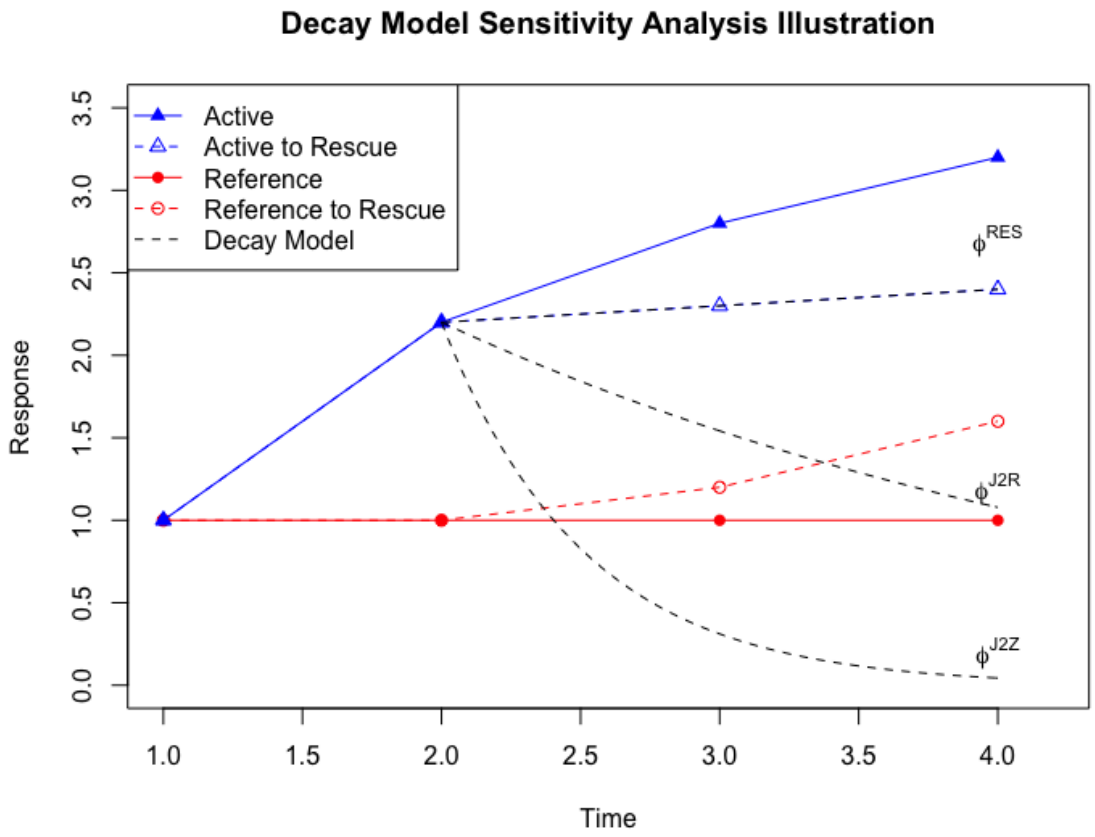

Figure 5.1: Demonstration of the decay model tipping point sensitivity analysis method and its relationship to RBI sensitivity analysis method

decay model and how it covers various of RBI method scenarios. Suppose the blue line represents the active treatment arm mean effect and the red line represents the reference arm mean effect. The dotted blue and red curves represent the mean effect for patients who switched to rescue medication after discontinued at time point 2 and time point 3 from original assignment, treatment arm and reference arm respectively. Under the primary analysis assumption, patients who dropped out from the study would be assumed to have the same effect as the switch to rescue medication patients, which are the dotted lines in the plot. Since the hypothesis testing is aiming on the difference between two arms, investigators would want to conduct a sensitivity analysis to be more conservative towards null assumption. Assume the reference arm dropout patients still behave as switch to rescue medication patients, but the active treatment arm dropout patients will have worse effects. The black line represents the hypothetical mean effect under the decay model in terms of different choice of decay parameter ϕ. In general, patients who dropped out at time 2 has a larger decay at the end of study time point 4 comparing to the patients who dropped out at time 4. By adjusting ϕ, the decay effect at the last time point continuously change from the most optimistic scenario, which is primary analysis scenario, to the most extreme case, which is 'J2Z'. The 'J2R' is one discrete scenario among the all possibles cases which will otherwise not be investigated using RBI method.

5.3 Decay model point estimator and reference using delta approximation method

Using the general format of estimator from MMRM model derived in previous chapter, the point estimator for decay estimator can be represented as follow,

$$\begin{aligned}\hat{\theta}^{DCY} =&\hat{\beta}_1 + \hat{\beta}_{2T-1} + \sum_{e=2}^{T}\Bigg((1-\hat{\pi}_{Me})\hat{\pi}_{Se}\hat{\gamma}_{T,T-2+e} - \hat{\omega}_{Se}\hat{\gamma}_{T,e-1} + \hat{\pi}_{Me}\hat{\pi}_{Se}\times \\ &\Big(\big(\hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_T + \hat{\beta}_{2T-1}\big)\big(\exp\left(-\phi\cdot\Delta t_e\right) - 1\big) + \hat{\gamma}_{T,T-2+e}\exp\left(-\phi\cdot\Delta t_e\right)\Big)\Bigg)\end{aligned} \tag{5.6}$$

The same delta approximation method as what was used in primary and RBI sensitivity analysis can be performed for the decay estimator

$$\begin{aligned}&V(\hat{\theta}^{DCY}) \\ =&V\Bigg(\hat{\beta}_1 + \hat{\beta}_{2T-1} + \sum_{e=2}^{T}\Bigg((1-\hat{\pi}_{Me})\hat{\pi}_{Se}\hat{\gamma}_{T,T-2+e} - \hat{\omega}_{Se}\hat{\gamma}_{T,e-1} + \hat{\pi}_{Me}\hat{\pi}_{Se}\times \\ &\Big(\big(\hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_T + \hat{\beta}_{2T-1}\big)\big(\exp\left(-\phi\cdot\Delta t_e\right) - 1\big) + \hat{\gamma}_{T,T-2+e}\exp\left(-\phi\cdot\Delta t_e\right)\Big)\Bigg)\Bigg) \\ =&V\Bigg(E\Bigg(\hat{\beta}_1 + \hat{\beta}_{2T-1} + \sum_{e=2}^{T}\Bigg((1-\hat{\pi}_{Me})\hat{\pi}_{Se}\hat{\gamma}_{T,T-2+e} - \hat{\omega}_{Se}\hat{\gamma}_{T,e-1} + \hat{\pi}_{Me}\hat{\pi}_{Se}\times \\ &\Big(\big(\hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_T + \hat{\beta}_{2T-1}\big)\big(\exp\left(-\phi\cdot\Delta t_e\right) - 1\big) + \hat{\gamma}_{T,T-2+e}\exp\left(-\phi\cdot\Delta t_e\right)\Big)\Bigg)|\boldsymbol{\Omega}\Bigg)\Bigg) + \\ &E\Bigg(V\Bigg(\hat{\beta}_1 + \hat{\beta}_{2T-1} + \sum_{e=2}^{T}\Bigg((1-\hat{\pi}_{Me})\hat{\pi}_{Se}\hat{\gamma}_{T,T-2+e} - \hat{\omega}_{Se}\hat{\gamma}_{T,e-1} + \hat{\pi}_{Me}\hat{\pi}_{Se}\times \\ &\Big(\big(\hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_T + \hat{\beta}_{2T-1}\big)\big(\exp\left(-\phi\cdot\Delta t_e\right) - 1\big) + \hat{\gamma}_{T,T-2+e}\exp\left(-\phi\cdot\Delta t_e\right)\Big)\Bigg)|\boldsymbol{\Omega}\Bigg)\Bigg)\end{aligned} \tag{5.7}$$

First part

$$
\begin{aligned}
&V\Bigg(E\Bigg(\hat{\beta}_1+\hat{\beta}_{2T-1}+\sum_{e=2}^{T}\Bigg((1-\hat{\pi}_{Me})\hat{\pi}_{Se}\hat{\gamma}_{T,T-2+e}-\hat{\omega}_{Se}\hat{\gamma}_{T,e-1}+\hat{\pi}_{Me}\hat{\pi}_{Se}\times\\
&\Big(\big(\hat{\beta}_0+\hat{\beta}_1+\hat{\beta}_T+\hat{\beta}_{2T-1}\big)\big(\exp\left(-\phi\cdot\Delta t_e\right)-1\big)+\hat{\gamma}_{T,T-2+e}\exp\left(-\phi\cdot\Delta t_e\right)\Big)\Bigg)|\mathbf{\Omega}\Bigg)\Bigg)\\
=&V\Bigg(\beta_1+\beta_{2T-1}+\sum_{e=2}^{T}\Bigg(\frac{q_e}{N}\gamma_{T,T-2+e}-\frac{m_e+n_e}{N}\gamma_{T,e-1}+\\
&\frac{p_e}{N}\Big(\big(\beta_0+\beta_1+\beta_T+\beta_{2T-1}\big)\big(\exp\left(-\phi\cdot\Delta t_e\right)-1\big)+\gamma_{T,T-2+e}\exp\left(-\phi\cdot\Delta t_e\right)\Big)\Bigg)\Bigg)\\
=&\frac{1}{N^2}V\Bigg(\sum_{e=2}^{T}\Bigg(q_e\gamma_{T,T-2+e}+p_e\Big(\big(\beta_0+\beta_1+\beta_T+\beta_{2T-1}\big)\big(\exp\left(-\phi\cdot\Delta t_e\right)-1\big)+\\
&\gamma_{T,T-2+e}\exp\left(-\phi\cdot\Delta t_e\right)\Big)\Bigg)+\sum_{e=2}^{T}(m_e+n_e)\gamma_{T,e-1}\Bigg)
\end{aligned}
\tag{5.8}
$$

Denote $\big(\beta_0+\beta_1+\beta_T+\beta_{2T-1}\big)\big(\exp\left(-\phi\cdot\Delta t_e\right)-1\big)+\gamma_{T,T-2+e}\exp\left(-\phi\cdot\Delta t_e\right)$ as τ_e, the first part becomes

$$
\begin{aligned}
&\frac{1}{N^2}V\Bigg(\sum_{e=2}^{T}\Big(q_e\gamma_{T,T-2+e}+p_e\tau_e\Big)+\sum_{e=2}^{T}(m_e+n_e)\gamma_{T,e-1}\Bigg)\\
=&\frac{1}{N}\Bigg(\sum_{e=2}^{T}\gamma_{T,T-2+e}^2(1-\pi_{Me})\pi_{Se}(1-(1-\pi_{Me})\pi_{Se})+\sum_{e=2}^{T}\tau_e^2\pi_{Me}\pi_{Se}(1-\pi_{Me}\pi_{Se})-\\
&\sum_{f\neq g=2}^{T}\gamma_{T,T-2+f}\gamma_{T,T-2+g}(1-\pi_{Mg})\pi_{Sg}(1-\pi_{Mf})\pi_{Sf}-\sum_{f\neq g=2}^{T}\tau_f\tau_g\pi_{Mg}\pi_{Sg}\pi_{Mf}\pi_{Sf}-\\
&\sum_{f,g=2}^{T}\gamma_{T,T-2+f}\tau_g(1-\pi_{Mf})\pi_{Sf}\pi_{Mg}\pi_{Sg}+\\
&\sum_{e=2}^{T}\gamma_{T,e-1}^2\omega_{Se}(1-\omega_{Se})-\sum_{f\neq q=2}^{T}\gamma_{T,f-1}\gamma_{T,g-1}\omega_{Sf}\omega_{Sg}\Bigg)
\end{aligned}
\tag{5.9}
$$

For the second part, first derive the inner conditional variance part

$$
\begin{aligned}
&V\Bigg(\hat{\beta}_1+\hat{\beta}_{2T-1}+\sum_{e=2}^{T}\Bigg((1-\hat{\pi}_{Me})\hat{\pi}_{Se}\hat{\gamma}_{T,T-2+e}-\hat{\omega}_{Se}\hat{\gamma}_{T,e-1}+\hat{\pi}_{Me}\hat{\pi}_{Se}\times\\
&\Big(\big(\hat{\beta}_0+\hat{\beta}_1+\hat{\beta}_T+\hat{\beta}_{2T-1}\big)\big(\exp\left(-\phi\cdot\Delta t_e\right)-1\big)+\gamma_{T,T-2+e}\exp\left(-\phi\cdot\Delta t_e\right)\Big)\Bigg)|\mathbf{\Omega}\Bigg)\\
=&V\Bigg(\hat{\beta}_1+\hat{\beta}_{2T-1}+\sum_{e=2}^{T}\Bigg((1-\hat{\pi}_{Me})\hat{\pi}_{Se}\hat{\gamma}_{T,T-2+e}-\hat{\omega}_{Se}\hat{\gamma}_{T,e-1}+\hat{\pi}_{Me}\hat{\pi}_{Se}\times\\
&\big(\exp\left(-\phi\cdot\Delta t_e\right)-1\big)(\hat{\beta}_0+\hat{\beta}_1+\hat{\beta}_T+\hat{\beta}_{2T-1})+\exp\left(-\phi\cdot\Delta t_e\right)\hat{\gamma}_{T,T-2+e}\Big)|\mathbf{\Omega}\Bigg)\\
\approx&\hat{\mathbf{B}}\mathbf{\Sigma}_{\eta'}\hat{\mathbf{B}}'
\end{aligned}
\tag{5.10}
$$

where

$$
\hat{\mathbf{B}}'=\begin{bmatrix}
\sum_{e=2}^{T}\Big(\frac{p_e}{N}\big(\exp\left(-\phi\cdot\Delta t_e\right)-1\big)\Big)\\
\sum_{e=2}^{T}\Big(\frac{p_e}{N}\big(\exp\left(-\phi\cdot\Delta t_e\right)-1\big)\Big)+1\\
\sum_{e=2}^{T}\Big(\frac{p_e}{N}\big(\exp\left(-\phi\cdot\Delta t_e\right)-1\big)\Big)\\
\sum_{e=2}^{T}\Big(\frac{p_e}{N}\big(\exp\left(-\phi\cdot\Delta t_e\right)-1\big)\Big)+1\\
-\frac{n_2+m_2}{N}\\
-\frac{n_3+m_3}{N}\\
\vdots\\
-\frac{n_T+m_T}{N}\\
\frac{q_2}{N}+\frac{p_2}{N}\exp\big(-\phi(T-1)\big)\\
\frac{q_2}{N}+\frac{p_2}{N}\exp\big(-\phi(T-2)\big)\\
\vdots\\
\frac{q_T}{N}+\frac{p_T}{N}\exp\big(-\phi\big)
\end{bmatrix}
$$

and $\mathbf{\Sigma}_{\eta'}$ is the partition of $\mathbf{\Sigma}_{\eta}$ corresponding to

$$
[\hat{\beta}_0,\hat{\beta}_1,\hat{\beta}_T,\hat{\beta}_{2T-1},\hat{\gamma}_{T,1},\hat{\gamma}_{T,2},...,\hat{\gamma}_{T,T-1},\hat{\gamma}_{T,T},\hat{\gamma}_{T,T+1},...,\hat{\gamma}_{T,2T-2}]
$$

The expectation of the conditional variance is $\mathbf{B}\boldsymbol{\Sigma}_{\eta'}\mathbf{B}'$ where

$$\mathbf{B}' = \begin{bmatrix} \sum_{e=2}^{T}\Big(\pi_{Me}\pi_{Se}\big(\exp\left(-\phi\cdot\Delta t_e\right)-1\big)\Big) \\ \sum_{e=2}^{T}\Big(\pi_{Me}\pi_{Se}\big(\exp\left(-\phi\cdot\Delta t_e\right)-1\big)\Big)+1 \\ \sum_{e=2}^{T}\Big(\pi_{Me}\pi_{Se}\big(\exp\left(-\phi\cdot\Delta t_e\right)-1\big)\Big) \\ \sum_{e=2}^{T}\Big(\pi_{Me}\pi_{Se}\big(\exp\left(-\phi\cdot\Delta t_e\right)-1\big)\Big)+1 \\ -\omega_{S2} \\ -\omega_{S3} \\ \vdots \\ -\omega_{ST} \\ (1-\pi_{M2})\pi_{S2}+\pi_{M2}\exp\big(-\phi(T-1)\big) \\ (1-\pi_{M3})\pi_{S3}+\pi_{M3}\exp\big(-\phi(T-2)\big) \\ \vdots \\ (1-\pi_{MT})\pi_{ST}+\pi_{MT}\exp\big(-\phi\big) \end{bmatrix}$$

$$\begin{aligned} &\hat{V}(\hat{\theta}^{DCY}) \\ \approx&\frac{1}{N}\Bigg(\sum_{e=2}^{T}\hat{\gamma}_{T,T-2+e}^2\frac{q_e(N-q_e)}{N^2}+\sum_{e=2}^{T}\hat{\tau}_e^2 p_e(N-p_e)N^2 \\ &-\sum_{f\neq g=2}^{T}\hat{\gamma}_{T,T-2+f}\hat{\gamma}_{T,T-2+g}\frac{q_g q_f}{N^2}-\sum_{f\neq g=2}^{T}\hat{\tau}_f\hat{\tau}_g\frac{p_g p_f}{N^2} \\ &-\sum_{f,g=2}^{T}\hat{\gamma}_{T,T-2+f}\hat{\tau}_g\frac{q_f p_g}{N}+\sum_{e=2}^{T}\hat{\gamma}_{T,e-1}^2\frac{(m_e+n_e)(N-m_e-n_e)}{N^2} \\ &-\sum_{f\neq q=2}^{T}\hat{\gamma}_{T,f-1}\hat{\gamma}_{T,g-1}\omega_{Sf}\omega_{Sg}\frac{(m_f+n_f)(m_g+n_g)}{N^2}\Bigg) \\ &+\hat{\mathbf{B}}\hat{\boldsymbol{\Sigma}}_{\eta'}\hat{\mathbf{B}}' \end{aligned} \tag{5.11}$$

where

$$\hat{\tau}_e=\big(\hat{\beta}_0+\hat{\beta}_1+\hat{\beta}_T+\hat{\beta}_{2T-1}\big)\big(\exp\big(-\phi(T-e+1)\big)-1\big)+\hat{\gamma}_{T,T-2+e}\exp\big(-\phi(T-e+1)\big)$$

5.4 The decay estimator using MI approach

MI approach to estimate the decay estimator can also be used

$$[Y_{i,e}, Y_{i,e+1}, ..., Y_{i,T}|y_{i,1}, ..., y_{i,e-1}, G_i = g, S_i = 3, E_i = e] \sim MVN\left(\hat{\boldsymbol{\mu}}, \boldsymbol{\Sigma}\right) \tag{5.12}$$

where $\hat{\boldsymbol{\mu}} = \begin{bmatrix} \hat{\mu}_{i,e,g} \\ \hat{\mu}_{i,e+1,g} \\ \vdots \\ \hat{\mu}_{i,T,g} \end{bmatrix} + \hat{\boldsymbol{\Sigma}}_{e:T,1:e-1}\hat{\boldsymbol{\Sigma}}^{-1}_{1:e-1,1:e-1} \begin{bmatrix} y_{i,1} - \hat{\mu}_{i,1,g} \\ y_{i,2} - \hat{\mu}_{i,2,g} \\ \vdots \\ y_{i,e-1} - \hat{\mu}_{i,e-1,g} \end{bmatrix}$

$\boldsymbol{\Sigma} = \hat{\boldsymbol{\Sigma}}_{1:e-1,1:e-1} - \hat{\boldsymbol{\Sigma}}'_{e:T,1:e-1}\hat{\boldsymbol{\Sigma}}^{-1}_{e:T,e:T}\hat{\boldsymbol{\Sigma}}_{1:e-1,e:T}$. For reference arm, the conditional mean estimators are

$$\begin{aligned}
&\hat{\mu}_{i,1,0} = \hat{\beta}_0 \\
&\hat{\mu}_{i,2,0} = \hat{\beta}_0 + \hat{\beta}_2 \\
&\vdots \\
&\hat{\mu}_{i,e-1,0} = \hat{\beta}_0 + \hat{\beta}_{e-1} \\
&\hat{\mu}_{i,e,0} = \hat{\beta}_0 + \hat{\beta}_e + \hat{\gamma}_{e,e-1} \\
&\vdots \\
&\hat{\mu}_{i,T,0} = \hat{\beta}_0 + \hat{\beta}_T + \hat{\gamma}_{T,T-1}
\end{aligned} \tag{5.13}$$

For active treatment arm patients using decay estimator, the conditional mean estimators are

$$\begin{aligned}
&\hat{\mu}_{i,1,1} = \hat{\beta}_0 + \hat{\beta}_1 \\
&\hat{\mu}_{i,2,1} = \hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_2 + \hat{\beta}_{T+1} \\
&\vdots \\
&\hat{\mu}_{i,e-1,1} = \hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_{e-1} + \hat{\beta}_{T+e-2} \\
&\hat{\mu}_{i,e,1} = \left(\hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_e + \hat{\beta}_{T+e-1} + \hat{\gamma}_{e,2e-2}\right) \exp\left(-\phi\right) \\
&\vdots \\
&\hat{\mu}_{i,T,1} = \left(\hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_T + \hat{\beta}_{2T-1} + \hat{\gamma}_{T,2T-2}\right) \exp\left(-\phi(T-e+1)\right)
\end{aligned} \tag{5.14}$$

Use the MMRM covariance estimate $\hat{\mathbf{\Sigma}}_y$ as the variance covariance matrix. Thus, the detail steps for MI based on joint multivariate modeling approach are

1. Fit the imputation conditional MMRM model using observed data.
2. Impute missing data based on the conditional multivariate normal model constructed using adjusted decay sensitivity analysis conditional mean.
3. Fit the analysis MMRM model using the imputed complete dataset.
4. Record the decay estimator and corresponding inference as $\hat{Q}^{(w)}, \hat{V}^{(w)}$ for the wth imputation.
5. Repeat step 1-4 for W times. W is suggested to be around 10 times.
6. Combine the W estimates using Rubin's combining rule.

$$\begin{aligned}
\bar{Q} &= W^{-1} \sum \hat{Q}^{(w)} \\
T &= (1 + W^{-1})(W-1)^{-1} \sum (\hat{Q}^{(w)} - \bar{Q})^2 + \bar{V}
\end{aligned} \tag{5.15}$$

5.5 Derive the tipping point based on decay model

Based on the analytically derived point estimate and its inference for decay estimator from previous chapter, the tipping point that turns a statistical significant result to insignificant can be accessed directly without iterative search. Suppose the following hypothesis test is performed

$$H_0 : \theta = 0$$

$$H_A : \theta > 0$$

using the test statistic

$$T = \frac{\hat{\theta}}{s.e(\hat{\theta})} \sim t_{df=2N} \tag{5.16}$$

To find the tipping point, suppose the null hypothesis at level $\alpha = 5\%$ is being tested, consider the test statistic as a function of ϕ and set $T(\phi) = t_{0.975,2N}$ and solve for ϕ^{TIP}. That is,

$$t_{0.975,2N} = \frac{\hat{\theta}(\phi)}{s.e(\hat{\theta}(\phi))} \tag{5.17}$$

where

$$\begin{aligned}\hat{\theta}(\phi) = \hat{\beta}_1 + \hat{\beta}_{2T-1} + \sum_{e=2}^{T} \Bigg(& \frac{q_e}{N}\hat{\gamma}_{T,T-2+e} - \frac{m_e + n_e}{N}\hat{\gamma}_{T,e-1} \\ & + \frac{p_e}{N}\Big(\big(\hat{\beta}_0 + \hat{\beta}_T\big)\big(\exp\left(-\phi \cdot \Delta t_e\right) - 1\big) + \hat{\gamma}_{T,T-2+e} \exp\left(-\phi \cdot \Delta t_e\right)\Big)\Bigg)\end{aligned} \tag{5.18}$$

$$
\begin{aligned}
&s.e(\hat{\theta}(\phi)) = \\
&\Bigg(\hat{\mathbf{B}}\boldsymbol{\Sigma}_{\eta'}\hat{\mathbf{B}}' + \frac{1}{N}\Bigg(\sum_{e=2}^{T}\gamma_{T,T-2+e}^2(1-\pi_{Me})\pi_{Se}(1-(1-\pi_{Me})\pi_{Se}) \\
&+\sum_{e=2}^{T}\tau_e^2\pi_{Me}\pi_{Se}(1-\pi_{Me}\pi_{Se}) - \sum_{f\neq g=2}^{T}\gamma_{T,T-2+f}\gamma_{T,T-2+g}(1-\pi_{Mg})\pi_{Sg}(1-\pi_{Mf})\pi_{Sf} \\
&-\sum_{f\neq g=2}^{T}\tau_f\tau_g\pi_{Mg}\pi_{Sg}\pi_{Mf}\pi_{Sf} - \sum_{f,g=2}^{T}\gamma_{T,T-2+f}\tau_g(1-\pi_{Mf})\pi_{Sf}\pi_{Mg}\pi_{Sg} \\
&+\sum_{e=2}^{T}\gamma_{T,e-1}^2\omega_{Se}(1-\omega_{Se}) - \sum_{f\neq q=2}^{T}\gamma_{T,f-1}\gamma_{T,g-1}\omega_{Sf}\omega_{Sg}\Bigg)\Bigg)^{\frac{1}{2}}
\end{aligned}
\tag{5.19}
$$

Note that unfortunately the nonlinear equation does not have a close analytic form. Numerical methods such as Newton-Raphson algorithm can be performed to find the solution for this nonlinear equation.

5.6 Single simulation to demonstrate the usage and verify the inference of decay model tipping point sensitivity analysis

In order to demonstrate the usage of decay model tipping point sensitivity analysis method, a single simulated study under the same setup as the previous chapter is investigated. The summary of the simulated study is as follow,

Table 5.1: Summary table for single simulated study for the purpose of demonstrating the usage of decay model

Arm	t_1		t_2		t_3		t_4	
	Mean	Miss	Mean	Miss	Mean	Miss	Mean	Miss
Treatment	7.16	0%	8.16	15%	8.67	20%	10.58	22%
Reference	7.26	0%	8.68	27%	9.12	32%	9.49	35%

Figure 5.2 is a plot of the estimate vs the decay rate. The red curve is derived based on the delta approximate method. The black curve is derived based on MI method.

The dot is calculated based on Newton-Raphson method solving the equation based on the delta approximation method. The plot indicates that the Newton-Raphson method correctly solves the delta approximation method based equation and finds the tipping point that switches a significant result study to become insignificant. For this specific simulated case, the tipping point is $\phi^{TIP} = 0.13$. Thus the conclusion can be made that under the primary analysis assumption which is that the dropout patients would behave the same as those who switched to rescue medication if they were not missing, the two arm effect difference is significantly greater than zero. This test result significance would hold unless the treatment arm dropout patients do not behave the same as those who switched to rescue medication and the effect after dropout gets worse with time at certain exponential rate. The tipping point is at $\phi = 0.13$ which means the post intercurrent event effect is $\exp(-0.13\Delta t)$ times rescue medication effect. For patients who dropped out at t_2, the remaining post intercurrent event effect at t_4 is $\exp\big(-0.13 \times (4-1)\big) = 67.7\%$ of the rescue medication effect. For patients who dropped out at t_3, the remaining effect is $\exp\big(-0.13 \times (4-2)\big) = 77.1\%$ of the rescue medication effect. For patients who dropped out at t_4, the remaining effect is $\exp\big(-0.13 \times (4-3)\big) = 87.8\%$ of the rescue medication effect. From the plot, the tipping point is on the left edge of the ϕ envelop. Minor deviation from the MAR assumption would switch the significant result to insignificant, which indicates the MAR assumption is not robust. However, if investigators performed only 'J2R' RBI sensitivity analysis, the result would still be significant and therefore conclude the primary analysis result was robust. The decay model shows its advantage by identifying the level of the robustness of primary analysis result in this simulation scenario.

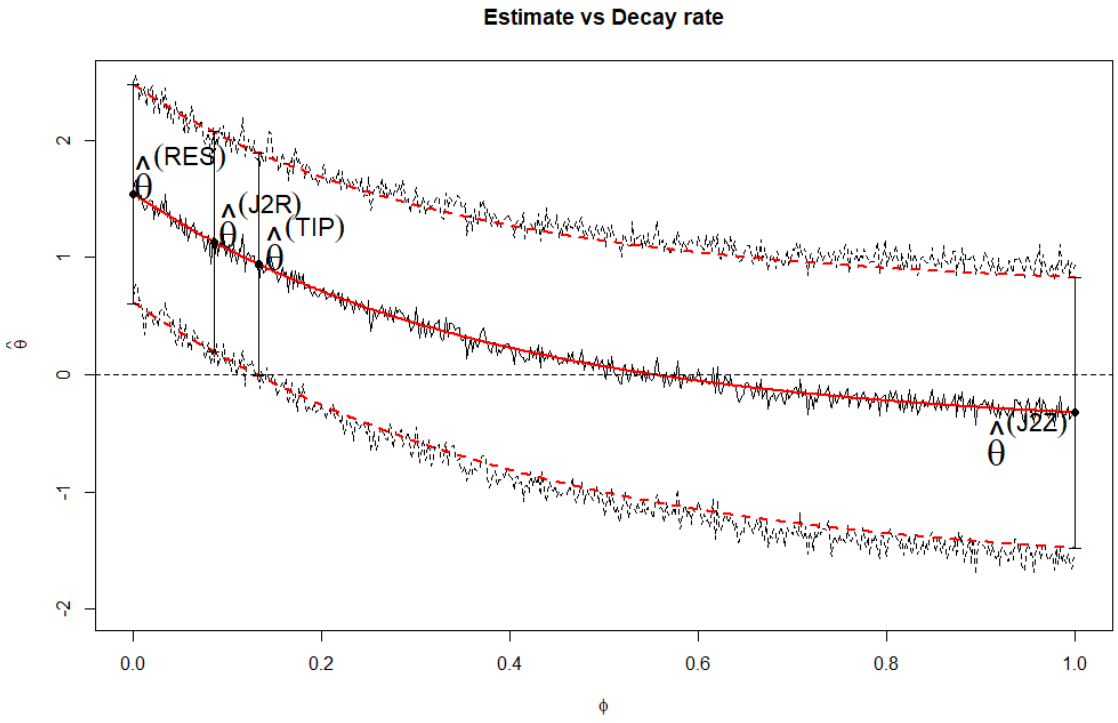

Figure 5.2: Simulated study to show the usage of decay model and compare with MI method

Chapter 6

Case study: A rare blood disease clinical trial study using decay model

6.1 Study background

In this chapter, I will use a real world ongoing clinical trial study to demonstrate the usefulness of decay model as sensitivity analysis. This study is conducted to investigate the effect of a new treatment for a rare slow progression blood disease. This trial is a randomized balanced two arm clinical trial. The active arm patients are treated using the studied drug. The reference arm patients are provided the standard of care. No rescue medication is allowed in this study. There are 127 patients in each arm. Total number of patients is 254. The standard of care for this disease has a rapid treatment effect. However, it also has a potential safety effect that will cause patients intolerable to the drug and cause high dropout rate. The studied drug is expected to cause less safety issue and more acceptable to patients. It takes longer period of time for the studied drug to reach equivalent treatment effect as the standard of care. However, the studied drug is expected to shows a continuous improvement of effect without a increase rate of dropout. The total study period is 24 months. It is divided into two stages. The first stage is from enrollment to Month 12. The second stage is from Month 12 to Month 24. There are three biomarkers that are the main determinants for disease diagnose. For the purpose of not revealing sensitive information of this ongoing study, the three biomarkers are referred to as H, W and P. The primary endpoint is the biomarker measurement difference between

active treatment arm and reference arm at Month 24. Additional endpoint is the proportion of events between two arms where the event is defined by the combination of H<45, W<10 and P < 400.

6.2 Data analysis

First I will investigate the data collected on biomarker H. Figure 6.1 is the longitudinal plot for biomarker H. Each line is a sequence of measurement for one individual patient. Red curve represents active arm and blue curve represents reference arm. Mean curves by groups are also plotted (thick lines). After Month 12, the plot visually shows less dense of the lines. This is a clear pattern showing that majority of patients missed measurement from Month 12. During the first stage, active arm mean effect started lower than the reference arm mean effect at Month 3. As the study goes on, the difference between active arm and reference arm decreases. During the second stage, the reference arm treatment effect diminished. The mean of biomarker H at the end of study at Month 24 for reference arm is worse than Month 3. For active arm, the treatment effect continued improvement until Month 18 and maintained until the end of study of Month 24.

Next, I will study the performance of biomarker W. Figure 6.2 is the longitudinal plot for biomarker B. The mean effect for active arm and reference arm almost overlap with each other during the first stage. During the second stage, active arm treatment effect keeps increase while reference arm effect stops improving and gets worse. Note that active arm patient measurements are more sparse than reference arm.

Figure 6.3 is the longitudinal plot for biomarker P. Active arm shows better mean effect at the end of the first stage and the effect is retained for the second stage while reference arm measurement becomes worse during the second stage.

Besides the mean effect, another major issue to be taken into consideration is the

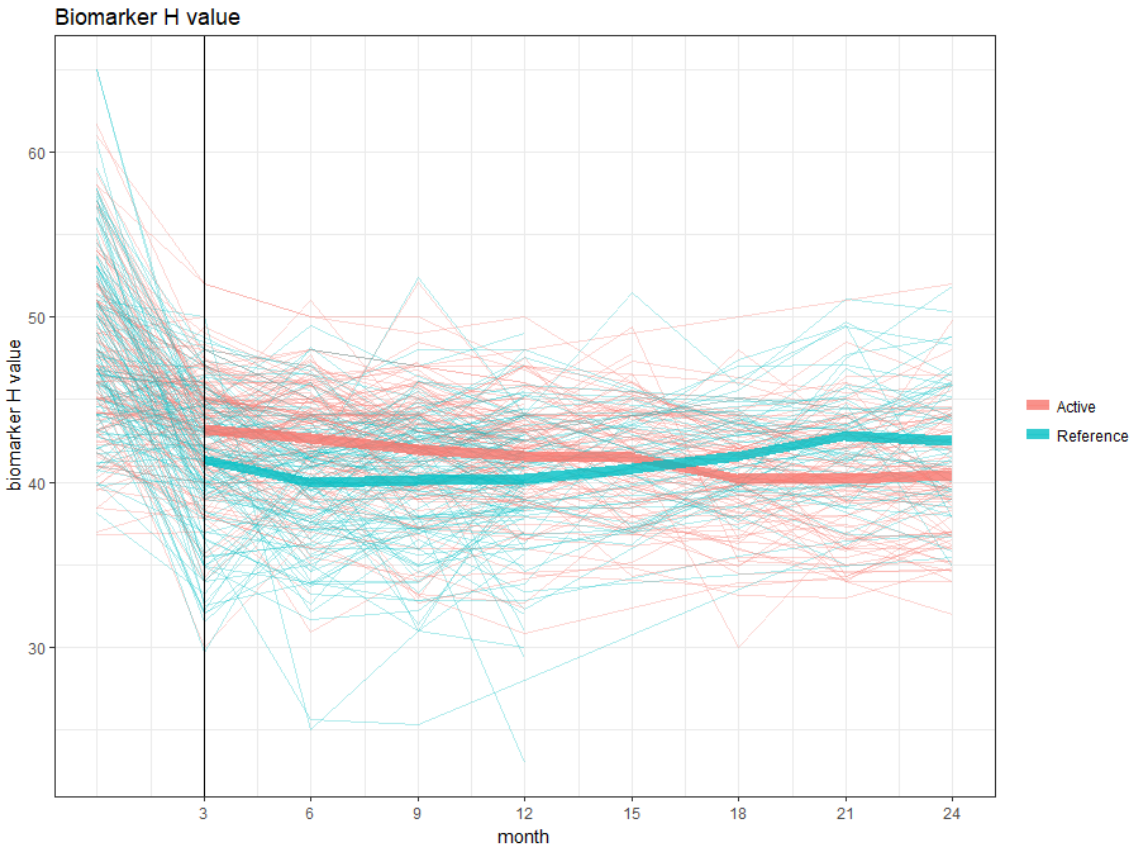

Figure 6.1: Longitudinal plot for biomarker H from baseline to Month 24

difference of dropout rate between two arms. Table 6.1 is the summary table for active arm and reference arm dropout rate during the first stage. The dropout rate is comparable between two arms. There is slightly higher dropout for active arm patients at Month 12. Since the measurements are taken every 3 months, use T4 to denote Month 12. Table 6.2 is the summary table for active arm and reference arm dropout rate during the second stage. Reference arm has much higher dropout rate comparing to active arm. Note that the dropout is not monotone. The greatest missing rate for active arm is at T6. The greatest missing rate for reference arm is at T5. One possible cause is the re-enrollment process after the end of first stage might cause some patients to miss measurements at T5 and T6.

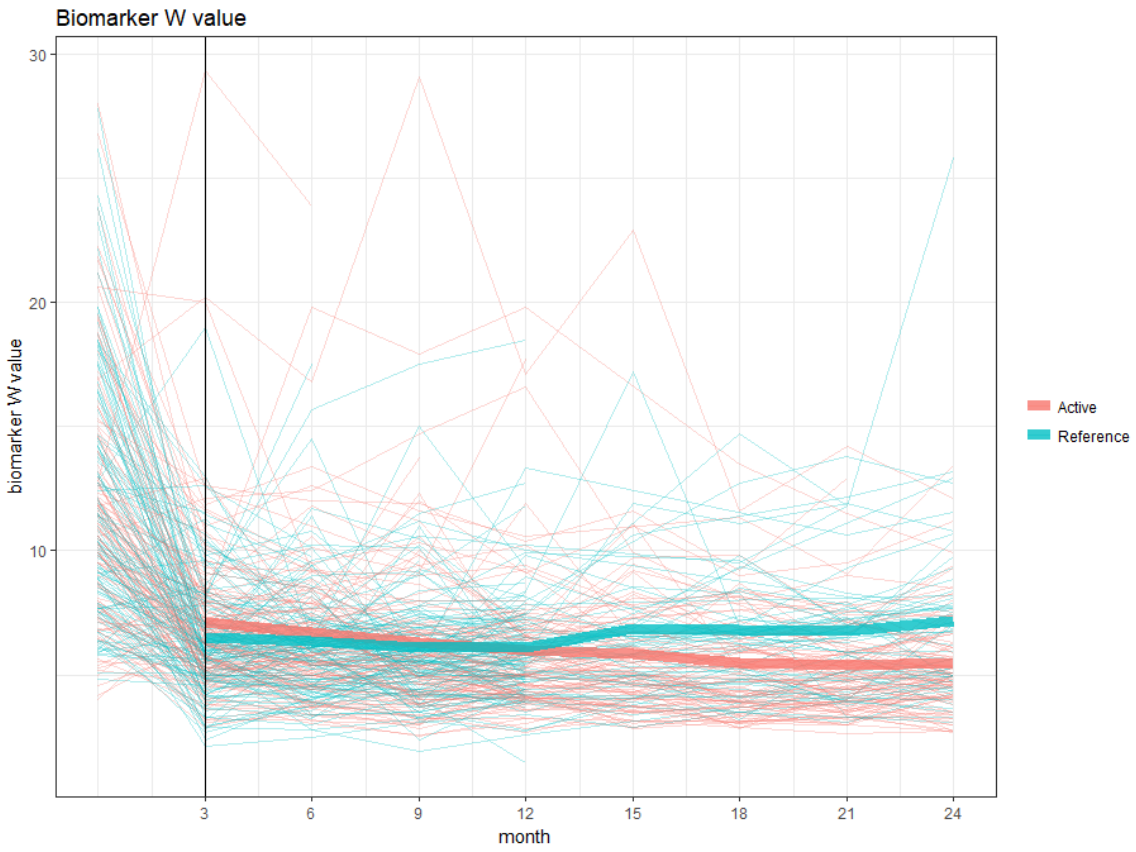

Figure 6.2: Longitudinal plot for biomarker B from baseline to Month 24

Table 6.1: Dropout probability table for first stage

Arm	First Stage			
	Month 3 (T1)	Month 6 (T2)	Month 9 (T3)	Month 12 (T4)
Active	0.79%	4.76%	9.52%	14.29%
Reference	0.81%	4.89%	8.13%	9.76%

Table 6.2: Dropout probability table for second stage

Arm	Second Stage			
	Month 15 (T5)	Month 18 (T6)	Month 21 (T7)	Month 24 (T8)
Active	38.89%	42.06%	32.54%	30.16%
Reference	82.93%	70.73%	63.41%	57.52%

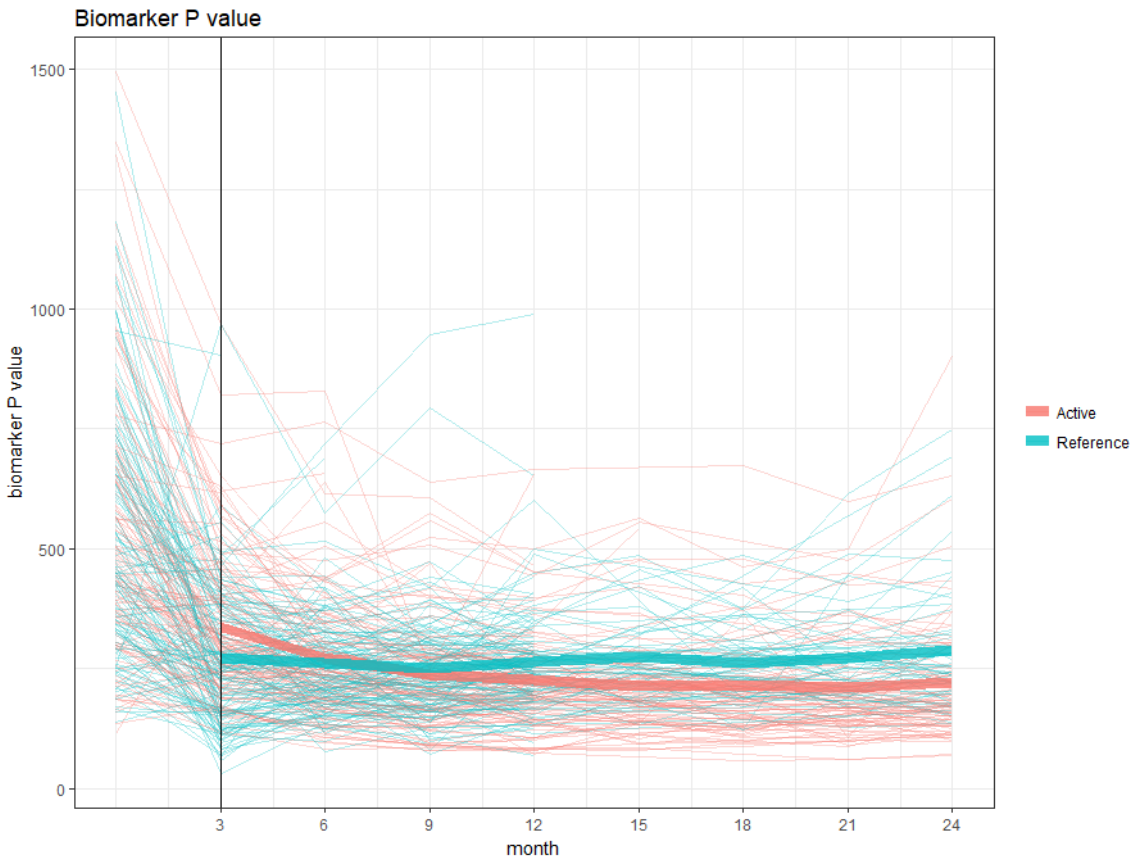

Figure 6.3: Longitudinal plot for biomarker P from baseline to Month 24

6.3 MMRM model and primary estimator for biomarker H

Biomarker H is studied as an example. Since there is no rescue medication allowed in the study, the MMRM model can be simplified as below,

$$\begin{aligned} & Y_{i,j}|G_i = \beta_0 + \beta_1 G_i \\ & + \beta_2 I_{j=2} + \beta_3 I_{j=3} + \beta_4 I_{j=4} + \beta_5 I_{j=5} + \beta_6 I_{j=6} + \beta_7 I_{j=7} + \beta_8 I_{j=8} \\ & + \beta_9 I_{j=2} G_i + \beta_{10} I_{j=3} G_i + \beta_{11} I_{j=4} G_i + \beta_{12} I_{j=5} G_i \\ & + \beta_{13} I_{j=6} G_i + \beta_{14} I_{j=7} G_i + \beta_{15} I_{j=8} G_i \end{aligned} \tag{6.1}$$

where $Y_{i,j}$ is the measurement of H at time point Tj for patient i. G_i is arm indicator. $G_i = 0$ represents reference arm and $G_i = 1$ represents active arm. The non-monotone part of the missing data is assumed to be missing at random. The primary estimator is under MAR assumption. The point estimate can be written as a linear combination of MMRM parameter estimates. $\hat{\theta} = \hat{\beta}_1 + \hat{\beta}_{15}$.

6.4 Decay model implementation

In the blood disease study, the lower value of biomarker H indicates a better result. Therefore, instead of setting the worst case as 0 in the genereal model in previous chapter, the greatest observed baseline measurements among all patients is used as the worst possible value, denote as Max. The decay effect for active arm can be represented as

$$E(Y_{i,8}|G_i = 1, E_i = e) = Max - (Max - E(Y_{i,8}|G_i = 1)) \exp\big(-\phi(8 - e + 1)\big) \tag{6.2}$$

For example, for active patients who dropped out at Month 15 (T5) the expected mean effect can be written as,

$$\begin{aligned} & E(Y_{i,8}|G_i = 1, E_i = 5) \\ & = Max - (Max - E(Y_{i,8}|G_i = 1)) \exp\big(-\phi(8 - 5 + 1)\big) \\ & = Max - (Max - (\beta_0 + \beta_1 + \beta_8 + \beta_{15})) \exp(-4\phi) \end{aligned} \tag{6.3}$$

The decay sensitivity analysis estimator can be written as,

$$\begin{aligned}\hat{\theta}^{DCY} &= \hat{\pi}_5\big(Max - (Max - (\hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_8 + \hat{\beta}_{15}))\exp(-4\phi)\big) \\ &+ \hat{\pi}_6\big(Max - (Max - (\hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_8 + \hat{\beta}_{15}))\exp(-3\phi)\big) \\ &+ \hat{\pi}_7\big(Max - (Max - (\hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_8 + \hat{\beta}_{15}))\exp(-2\phi)\big) \\ &+ \hat{\pi}_8\big(Max - (Max - (\hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_8 + \hat{\beta}_{15}))\exp(-\phi)\big) \\ &+ \hat{\pi}_{obs}\big(\hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_8 + \hat{\beta}_{15}\big)\end{aligned} \tag{6.4}$$

where $\hat{\pi}_5, \hat{\pi}_6, \hat{\pi}_7$ and $\hat{\pi}_8$ are the dropout proportions at $T5, T4, T3$ and $T2$ respectively. $\hat{\pi}_{obs}$ is the proportion of active arm patients that have measurement at T8.

The variance of the decay estimator can be approximated using delta method, $\hat{V}(\hat{\theta}^{DCY}) = \hat{P}^T\hat{\Sigma}_\pi\hat{P} + \hat{B}^T\hat{\Sigma}_\beta\hat{B}$ where

$$\hat{P} = \begin{bmatrix} Max - (Max - (\hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_8 + \hat{\beta}_{15}))\exp(-4\phi) \\ Max - (Max - (\hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_8 + \hat{\beta}_{15}))\exp(-3\phi) \\ Max - (Max - (\hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_8 + \hat{\beta}_{15}))\exp(-2\phi) \\ Max - (Max - (\hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_8 + \hat{\beta}_{15}))\exp(-\phi) \\ \hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_8 + \hat{\beta}_{15} \end{bmatrix}$$

$$\hat{B} = \begin{bmatrix} \hat{\pi}_5\exp(-4\phi) + \hat{\pi}_6\exp(-3\phi) + \hat{\pi}_7\exp(-2\phi) + \hat{\pi}_8\exp(-\phi) + \hat{\pi}_{obs} - 1 \\ \hat{\pi}_5\exp(-4\phi) + \hat{\pi}_6\exp(-3\phi) + \hat{\pi}_7\exp(-2\phi) + \hat{\pi}_8\exp(-\phi) + \hat{\pi}_{obs} \\ \hat{\pi}_5\exp(-4\phi) + \hat{\pi}_6\exp(-3\phi) + \hat{\pi}_7\exp(-2\phi) + \hat{\pi}_8\exp(-\phi) + \hat{\pi}_{obs} - 1 \\ \hat{\pi}_5\exp(-4\phi) + \hat{\pi}_6\exp(-3\phi) + \hat{\pi}_7\exp(-2\phi) + \hat{\pi}_8\exp(-\phi) + \hat{\pi}_{obs} \end{bmatrix}$$

$\hat{\Sigma}_\pi$ is the variance covariance matrix for $\hat{\pi}'s$ where the diagonal terms are $\hat{\pi}_i(1-\hat{\pi}_i)/N$ and off diagonal terms are $-\hat{\pi}_i\hat{\pi}_j/N$.

$\hat{\Sigma}_\beta$ is the variance covariance matrix for $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\beta}_8$ and $\hat{\beta}_{15}$ from MMRM model;

6.5 Decay model tipping point sensitivity analysis result

Figure 6.4 shows the result of primary as well as sensitivity analysis result for biomarker H. The red solid smooth curve is the mean curve for decay parameter ϕ versus the corresponding estimator $\hat{\theta}$. The red dotted smooth curves are the upper and lower 95% confidence interval derived using delta approximate method. The black dotted lines are the result from multiple imputation to verify the derived method. $\hat{\theta}^{(MAR)}$ is the primary analysis result under MAR. The 95% CI does not cover 0 which means active arm measurement is significantly lower than reference arm. $\hat{\theta}^{(J2Z)}$ is the sensitivity analysis result under most extreme case which makes the assumption that at month 24, all dropped out patients would have an effect equal to the worst measurement at baseline. The reason to set $\phi = 1$ for 'J2Z' is that when $\phi = 1$, the remaining effect is $\exp(-1 * 4) = 1.8\%$ of the original measurement for those who dropped out at time point T5, $\exp(-1 * 3) = 5.0\%$ of the original measurement for those who dropped out at time point T6, $\exp(-1 * 2) = 13.5\%$ of the original measurement for those who dropped out at time point T7 and $\exp(-1) = 36.8\%$ of the original measurement for those who dropped out at time point T8. Considering the fact that most of the patients dropped out from T5, and a relatively flat increasing trend of the $\hat{\theta}$ around $\phi = 1$, a $1 - 1.8\% = 98.2\%$ decay of effect can be considered as 'J2Z' scenario. The tipping point decay parameter $\hat{\theta}^{(TIP)}$ is calculated directly using Newton-Ralphson method. The plot indicates that the corresponding upper bound of 95% CI touches 0 which verifies the Newton-Raphson result. The decay rate parameter is $\phi = 0.10$ which indicates that when the active treatment effect is decayed by $1 - \exp(-0.10) = 9.5\%$ between every two measurement time points (every 3 months) will change the primary analysis result from significant to non-significant. Similar approaches are implemented for biomarker W and P (Figure 6.5 and Figure 6.6 respectively).

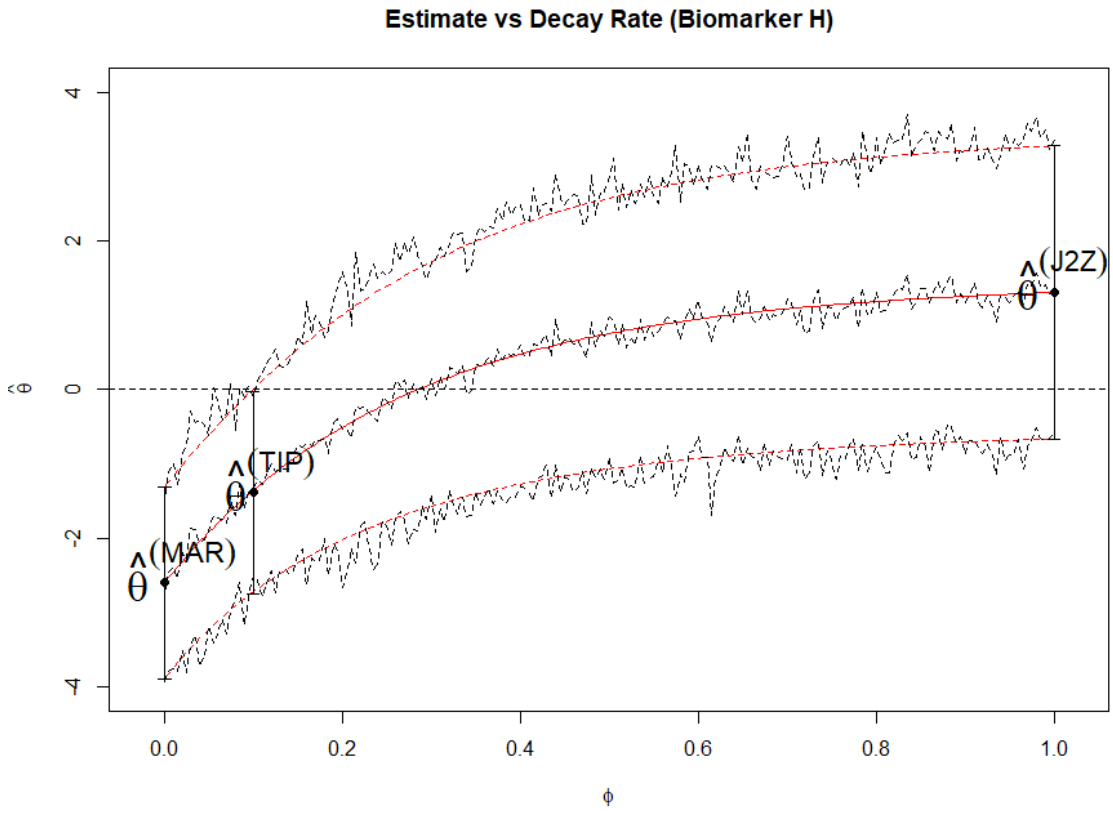

Figure 6.4: Blood disease study biomarker H compare delta approximation method versus MI method for primary and decay sensitivity analysis

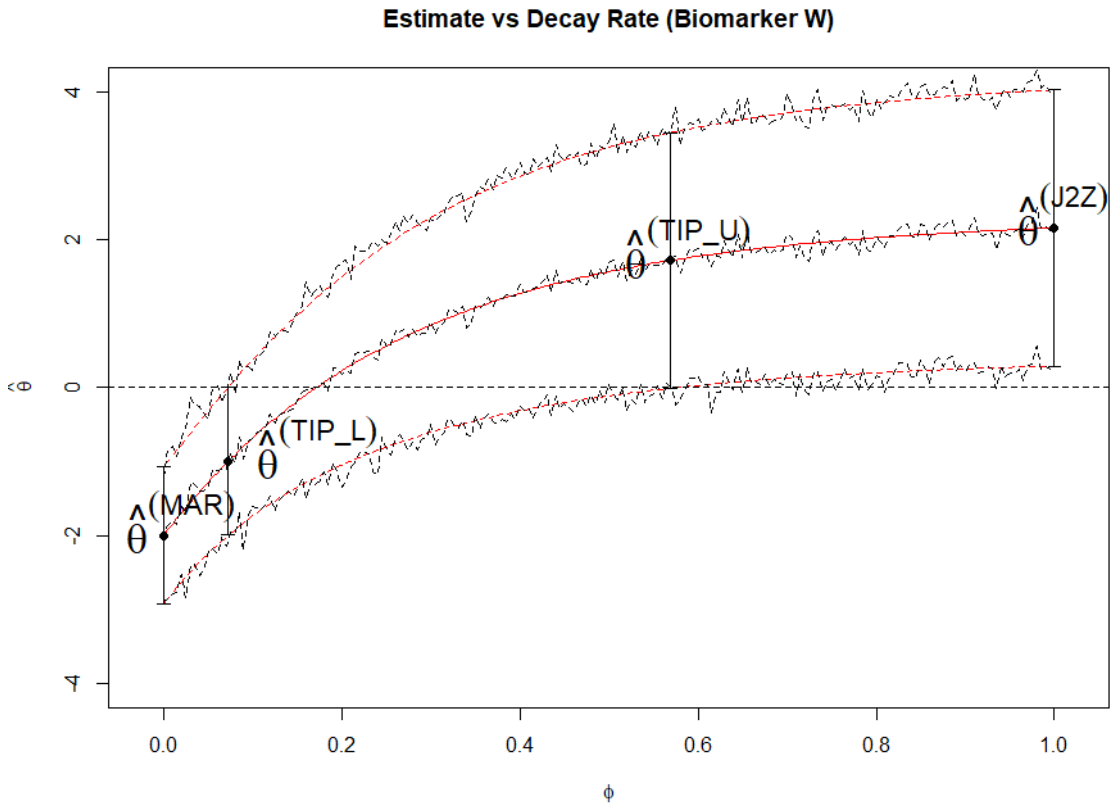

Figure 6.5: Blood disease study biomarker W compare delta approximation method versus MI method for primary and decay sensitivity analysis

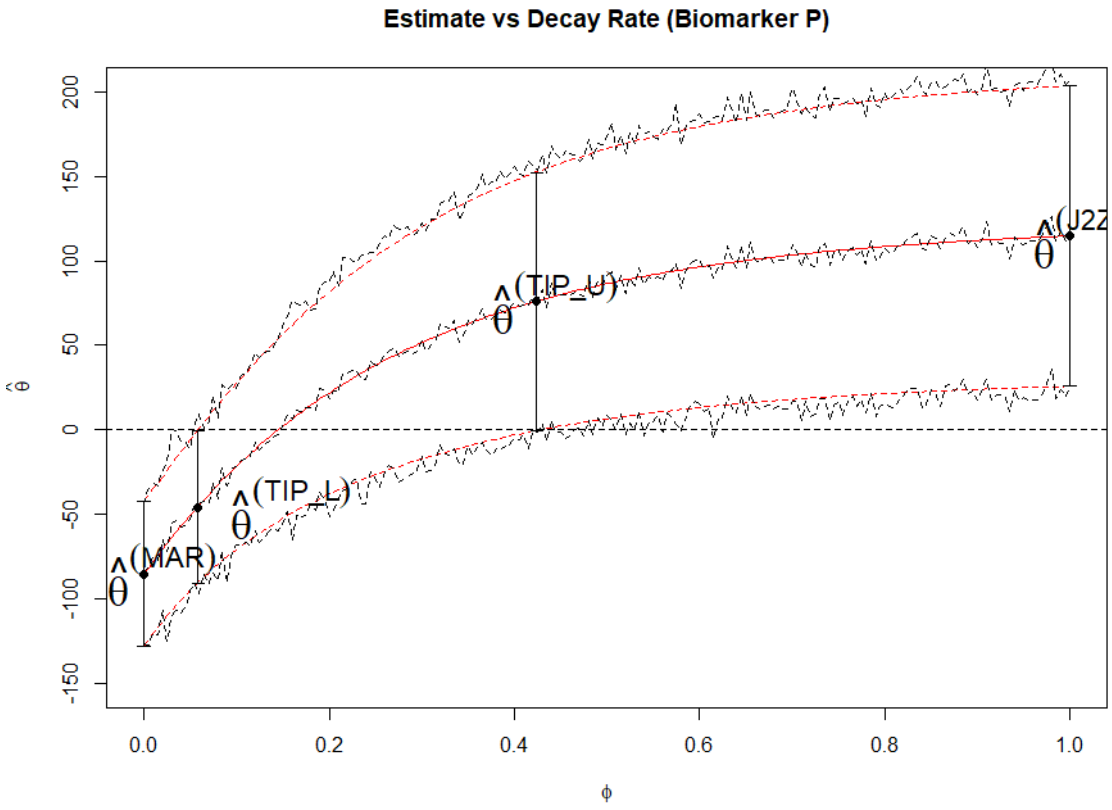

Figure 6.6: Blood disease study biomarker P compare delta approximation method versus MI method for primary and decay sensitivity analysis

6.6 Primary and sensitivity analysis result summary for biomarker H, W and P

The primary and sensitivity analysis result for all 3 biomarker H, W and P is summarized in Table 6.6. The remaining effect listed in the table is when patients dropped out at Month 15. Under MAR, all 3 biomarkers show significant better reduction for active treatment when comparing with standard of care at the end of second stage. For W and P, the sensitivity analysis under extreme case ($J2Z$) changes the result direction. Standard of care becomes significantly better than active treatment. This sets an alert for the investigator to be extra cautious when making decision based on the collected measurement because the missing data assumption might alter the conclusion to a totally different direction. This example shows one of the advantage of decay model which is although three biomarkers are on different scales, the derived tipping point are close to each other. The three decay rate parameters can be compared and the conclusion can be drawn that the statistical significance for biomarker H is most robust in terms of the primary MAR assumption. Even for the most extreme case, the active arm treatment effect will not be significantly worse than reference arm. The statistical significance for biomarker P is least robust in terms of the primary MAR assumption. For a decay rate of 21% at Month 24 if patient dropped out at Month 15, the active arm treatment effect is no longer significantly different from reference arm. For a decay rate of 81.4% at Month 24 if patient dropped out at Month 15, which is a relatively extreme assumption, the active arm treatment effect becomes significantly worse than reference arm.

Table 6.3: Primary and sensitivity analysis result summary for blood disease study biomarker H, W and P

	Endpoint Estimator $\hat{\theta}$		Decay Parameter ϕ			
	Primary MAR ($\hat{\theta}^{MAR}$)	Sensitivity J2Z ($\hat{\theta}^{J2Z}$)	Tipping Lower ϕ_L	Remain Effect [1]	Tipping Upper ϕ_U	Remain Effect [2]
H	-2.60 (-3.90, -1.31)	1.3 (-0.66, 3.28)	0.10	33.0%	NA	NA
W	-2.00 (-2.93, -1.08)	2.16 (0.29, 4.02)	0.072	25.0%	0.57	90.0%
P	-85.15 (-127.83, -42.46)	114.72 (25.83, 203.60)	0.059	21.0%	0.42	81.4%

[1] Remain effect for lower tipping point: $1 - \exp(-4\phi_L)$
[2] Remain effect for upper tipping point: $1 - \exp(-4\phi_U)$

6.7 Discussion: An extension to binary endpoint (Unsolved problem)

Recall that previously mentioned the disease diagnose is based on the comparison of Biomarker H, W and P with their corresponding thresholds, denote as cut_H, cut_W and cut_P. Denote $event_H = 1$ if $H < cut_H$, $event_H = 0$ otherwise. Denote $event_W = 1$ if $W < cut_W$, $event_W = 0$ otherwise. Denote $event_P = 1$ if $P < cut_P$, $event_P = 0$ otherwise. The binary estimator can be derived from sample estimated mean and variance together with normal assumption for the distribution of biomarkers. For reference arm, assume measurements at Month 24 (T8) follow a normal distribution $Y_{i,8}$ $N(\mu^r, \sigma^2)$. For active arm, assume measurements at Month 24 (T8) follow a mixture of normal distributions with different means due to the decay effect. $Y_{i,8} \sim \pi_5 N(\mu_5^a, \sigma^2) + \pi_6 N(\mu_6^a, \sigma^2) + \pi_7 N(\mu_7^a, \sigma^2) + \pi_8 N(\mu_8^a, \sigma^2) + \pi_{obs} N(\mu^a, \sigma^2)$ $\hat{\mu}^r = \hat{\beta}_0 + \hat{\beta}_8$ where $\hat{\mu}^a = \hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_8 + \hat{\beta}_{15}$, $\hat{\mu}_e^a = Max - (Max - \hat{\mu}^a)\exp\big(-\phi(8 - e + 1)\big)$ for $e = 5, 6, 7, 8$ and $\hat{\pi}_e$ is the proportion of the patients who dropout at Te. Observed reference and active arm pooled sample variance is used as the variance estimator $\hat{\sigma}^2$. Consider biomarker H first. The event probability estimator is written as follow. For reference arm, denote $\hat{p}^r = \hat{p}(Y_{i,8} < cut_H | G_i = 0) = \Phi(\frac{cut_H - \hat{\mu}^r}{\hat{\sigma}})$ where Φ is the CDF of

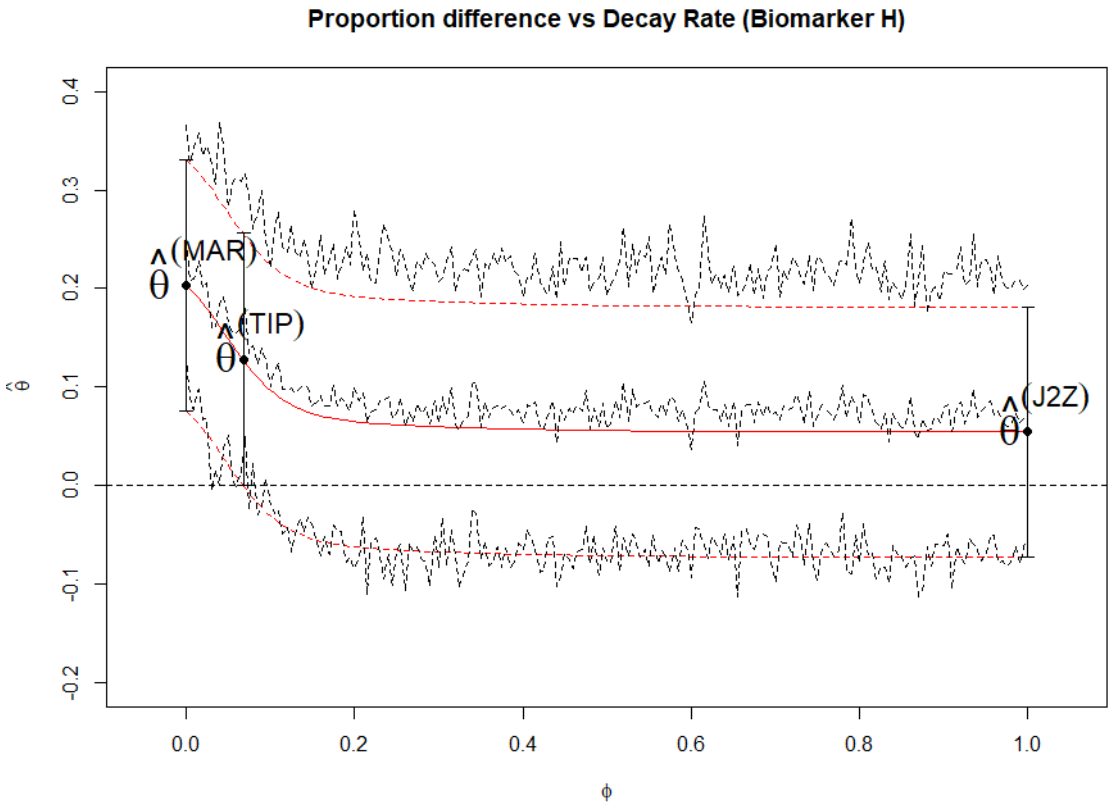

Figure 6.7: Biomarker H as binary endpoint compare delta approximation method versus MI method for primary and decay sensitivity analysis

$Normal(0,1)$. For active arm, denote $\hat{p}^a = \hat{p}(Y_{i,8} < cut_H | G_i = 1) = \hat{\pi}_5 \Phi(\frac{cut_H - \hat{\mu}_5^a}{\hat{\sigma}}) + \hat{\pi}_6 \Phi(\frac{cut_H - \hat{\mu}_6^a}{\hat{\sigma}}) + \hat{\pi}_7 \Phi(\frac{cut - \hat{\mu}_7^a}{\hat{\sigma}}) + \hat{\pi}_8 \Phi(\frac{cut_H - \hat{\mu}_8^a}{\hat{\sigma}}) + \hat{\pi}_{obs} \Phi(\frac{cut_H - \hat{\mu}^a}{\hat{\sigma}})$. The variance of the two estimators does not have close form and need to be solved numerically. Denote the inference estimators as $\hat{V}(\hat{p}^a)$ and $\hat{V}(\hat{p}^r)$. Z test is used to test the difference between two proportions. The test statistic is $Z = \frac{\hat{p}^a - \hat{p}^r}{\sqrt{\hat{V}(\hat{p}^a) + \hat{V}(\hat{p}^r)}}$.

The primary and sensitivity analysis results are presented in Figure 6.7, Figure 6.8 and Figure 6.9. Notice that there is a minor disagreement between delta approximation method and MI method. The exact reason is still under investigation. One possible explanation is that the normal assumption for the missing portion of the data is not valid. I suspect there might be data truncation presented.

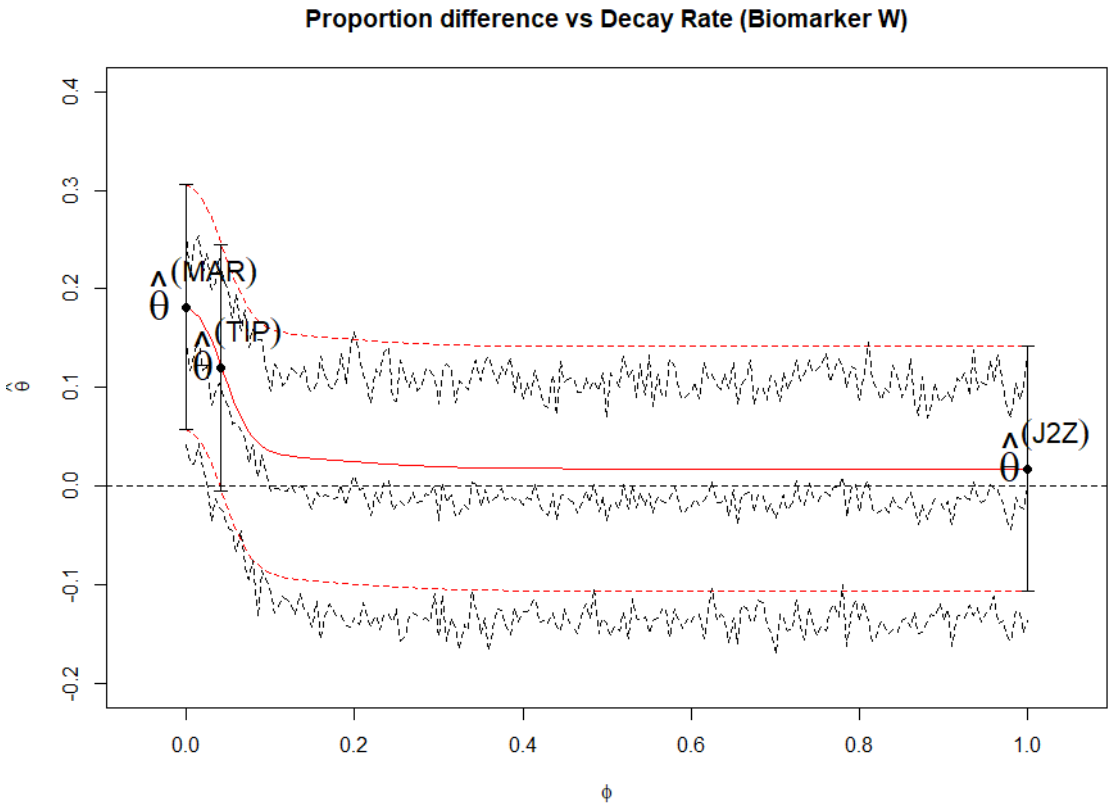

Figure 6.8: Biomarker W as binary endpoint compare delta approximation method versus MI method for primary and decay sensitivity analysis

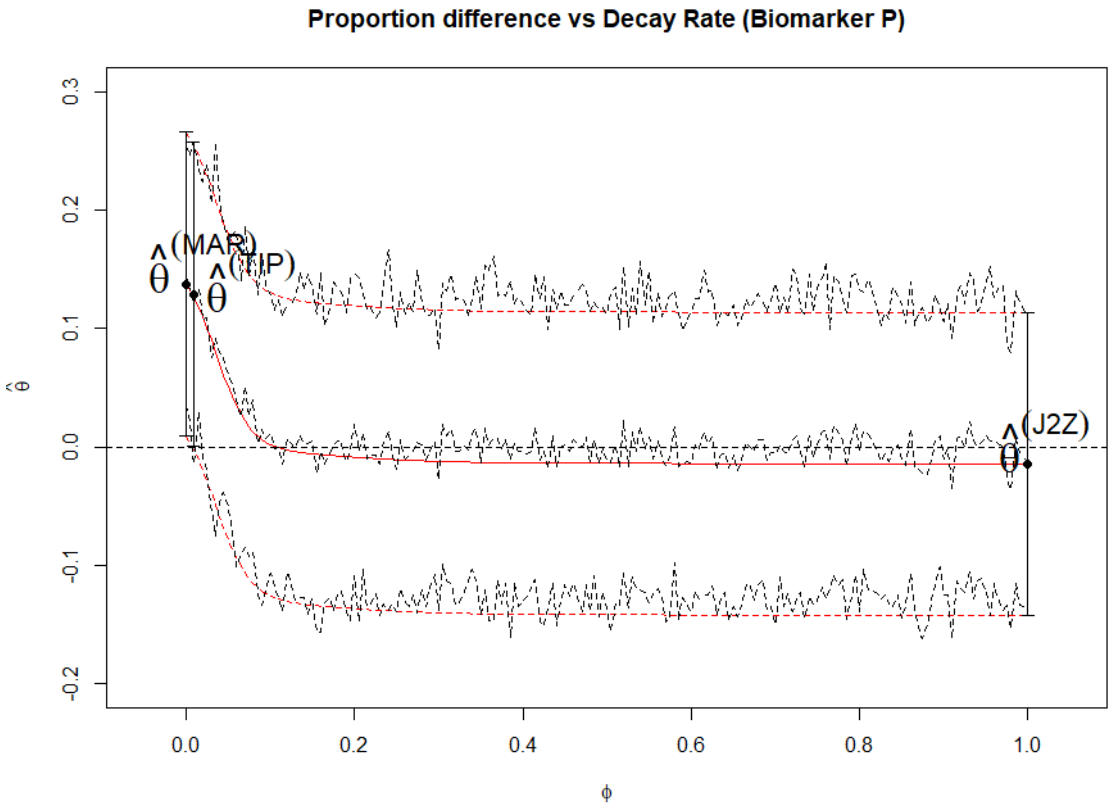

Figure 6.9: Biomarker P as binary endpoint compare delta approximation method versus MI method for primary and decay sensitivity analysis

Chapter 7
Conclusion and Discussion

7.1 Summary

FDA issued a draft guidance of E9 addendum focusing on the concept of estimand and sensitivity analysis in clinical trials. A framework regarding the process of constructing an appropriate estimand to target clinical objective was proposed. Estimand construction contains four attributes, which are a) targeted population, b) endpoint variable, c) intercurrent event strategy and d) population-level summary statistic. The terminology called 'intercurrent event' is a new concept introduced in the draft guidance to describe the events that cause discontinuation from initial randomized treatment. Five intercurrent event strategies were recommended in the draft guidance. The aim of the strategies is to identify potential intercurrent event and predetermine the handling methods during trial design stage. The strategies should be linked with the clinical objective, that is, the scientific question of interest that the study wants to answer. A lot of feedback has been given after the proposal of the draft version. Researchers raised the concern that five strategies might not be able to cover all intercurrent event scenarios. Also some strategy definitions, such as the 'treatment strategy' was not clearly explained. The concept of intercurrent event strategies and the estimand determination mingled together without a clear outline of the relationship. Thus I combined the literature discussion and the draft guidance and proposed an interuccernt event type oriented flowchart to help determine appropriate estimand. In this dissertation, I focused on the 'De Facto' estimand which includes the patient

who discontinued initial assigned treatment and switched to rescue medication measurements in the primary analysis.

After I conceptually clarified the relationship between estimand and intercurrent event and determined the focus point on the 'De Facto' estimand, new methodologies were developed to construct the primary point estimator and inference. First, the primary estimator for 'De Facto' estimand when rescue medication is allowed in a clinical trial was represented by a linear combination of modified MMRM model parameters. Next, instead of using multiple imputation and Rubin's Rule to estimate the inference for the point estimator, I proposed to use delta approximation method to derive the inference. The advantage of doing so was to decrease computing time that the MI would use. Another advantage was that a close form of the inference could be derived and it would be useful later on when deriving the tipping point in the sensitivity analysis. Simulations under different missing mechanisms under both null and alternative hypothesis were conducted to verify the modified MMRM model and delta approximation method.

The third part of the dissertation focused on sensitivity analysis regarding primary analysis estimator. An exponential function was used to model the deviation effect from primary analysis missing mechanism assumptions. A decay rate parameter integrated in the exponential function was used as a sensitivity parameter to cover all possible scenarios including the ones studied using RBI methods. The main purpose of the decay model effect was to study the robustness of primary assumption by completely exploring all possible sensitivity analysis scenarios. The delta approximation method that had been verified in previous chapters was implemented in decay model method to derive inference directly. Since the whole domain of possible decay rate is searched continuously, the avoidance of MI process saved tremendous amount of computing power and time. Nonetheless, using the close form of point estimator and corresponding inference, a tipping point that would change the primary analysis

from statistical significant to not significant was calculated directly without an iterative search. A single simulated study was used to show the implementation of the method.

This dissertation also demonstrated the implementation of the proposed methods through a case study using a real world rare blood disease clinical trial. The studied drug as active arm was compared to the standard of care treatment as an reference arm. The studied drug showed slower but continuous improvement of treatment effect with better tolerance comparing to a rapid effect standard of care which was hard to tolerant for patients and caused higher dropout rate. Sensitivity analysis was especially meaningful in this study due to the fact that most patients dropped out due to adverse events, which was clearly non-ignorable missing. Decay model sensitivity analysis found the tipping point and showed that a 33% reduction would cause the endpoint biomarker H to be insignificant, a 25% would cause the endpoint biomarker W to be insignificant and a 21% would cause the endpoint biomarker P to be insignificant. Thus among three endpoints, biomarker P is most sensitive to primary missing mechanism assumption. The decay model method also discovered that with an extreme case of reduction, which is 90.0% for biomarker W or 81.4% for biomarker P, the result would be reversed, which means under extreme assumption that if the drug effects decay rapidly, the active arm treatment effect might even become significantly worse than reference arm, which suggested investigators to be extra cautious when presenting the primary analysis result.

7.2 Discussion

There are some limitation regarding the methods proposed in the dissertation. Comparing the delta approximation method with the MI method, the delta approximation method requires extra derivation for each individual estimator. The MI method is

more general in terms of the extra modification needs to be done for different estimators. This is a trade off for the lesser computing source and time.

Note that the decay model sensitivity analysis method can choose different starting and ending values for the decay parameter. The starting value depends on the choice of estimand. In the dissertation, the single simulation study presented the case when the 'De Facto' estimand was considered. Thus the decay model started with the assumption that dropout patients would have behaved as if switched to rescue medication. In the case study, where the rescue medication is not of an option, the decay model started with the 'De Jure' estimand which assumed the dropout patients would have behaved as if still on initial randomized arm. Therefore, the left edge of the decay sensitivity plot depends on the choice of primary estimator assumption. In the dissertation, I picked $\phi = 1$ as the ending plot point for decay rate because the majority of patients dropped out 4 time points prior to the study endpoint. Thus $\phi = 1$ means more than 98% reduction which is a very extreme case. Estimate with greater ϕ can be plotted but the change will be minor. However, when comparing decay rate among different studies, a clarification of start and end point of decay rate search should be specified.

An attempt was made to extend the usage of delta approximation for decay model from continuous endpoint to binary endpoint. Technical issues were found as the missing might violate some of the distribution normality assumptions required during the derivation. Therefore only the MI result is presented. In addition, the three biomarker endpoints in the case studies are correlated. Imputation on correlated endpoints has yet been studied in literature. This is out of the scope of this dissertation but worth to be considered in future research.

References

[1] Kobi Abayomi, Andrew Gelman, and Marc Levy. Diagnostics for multivariate imputations. *Journal of the Royal Statistical Society. Series C: Applied Statistics*, 57(3):273–291, 2008.

[2] Mouna Akacha, Frank Bretz, David Ohlssen, Gerd Rosenkranz, and Heinz Schmidli. Estimands and their role in clinical trials. *Statistics in Biopharmaceutical Research*, 9(3):268–271, 2017.

[3] Mouna Akacha, Frank Bretz, and Stephen Ruberg. Estimands in clinical trials – broadening the perspective. *Statistics in Medicine*, 36(1):5–19, 2017.

[4] Mouna Akacha and Wolfgang Kothny. Estimands: A more strategic approach to study design and analysis. *Clinical Pharmacology & Therapeutics*, 102(6):894–896, 2017.

[5] Birhanu Teshome Ayele, Ilya Lipkovich, Geert Molenberghs, and Craig H. Mallinckrodt. A multiple-imputation-based approach to sensitivity analyses and effectiveness assessments in longitudinal clinical trials. *Journal of Biopharmaceutical Statistics*, 24(2):211–228, 2014.

[6] C. Hendricks Brown. Protecting against nonrandomly missing data in longitudinal studies. *Biometrics*, 46(1):143, 3 1990.

[7] James R. Carpenter, Michael G. Kenward, and Ian R. White. Sensitivity analysis after multiple imputation under missing at random: a weighting approach. *Statistical Methods in Medical Research*, 16(3):259–275, 2007.

[8] James R Carpenter, James H Roger, and Michael G Kenward. Analysis of longitudinal trials with protocol deviation: a framework for relevant, accessible assumptions, and inference via multiple imputation. *Journal of Biopharmaceutical Statistics*, 23(3):1352–71, 2013.

[9] Jr Carpenter and Mg Kenward. Missing data in randomised controlled trials–a practical guide. *AL: National Institute for Health Research*, pages 29–36, 2007.

[10] Fang Chen. Missing no more: using the MCMC procedure to model missing data. *SAS Global Forum*, pages 1–23, 2013.

[11] W Scorn Clark and R David. Type I error rates from mixed effects model repeated measures versus fixed effects ANOVA with missing values imputed via last observation carried forward. *Drug Information Journal*, 35:1215–1225, 2016.

[12] Avital Cnaan, NAN M. Laird, and Peter Slasor. Using the general linear mixed model to analyse unbalanced repeated measures and longitudinal data. *Stat.Med.*, 16(20):2349–2380, 1997.

[13] Committee for Medicinal Products for Human Use (CHMP) and Roderick J. A. Little. Guideline on missing data in confirmatory clinical trials. *London: European Medicines Agency*, 44(July 2010):1–12, 12 2011.

[14] An Creemers, Niel Hens, Marc Aerts, Geert Molenberghs, Geert Verbeke, and Michael G. Kenward. A sensitivity analysis for shared-parameter models for incomplete longitudinal outcomes. *Biometrical Journal*, pages NA–NA, 11 2009.

[15] An Creemers, Niel Hens, Marc Aerts, Geert Molenberghs, Geert Verbeke, and Michael G Kenward. Generalized shared-parameter models and missingness at random. *Statistical Modelling*, 11(4):279–310, 2011.

[16] Suzie Cro, Michael G. Kenward, and James R. Carpenter. Reference based sensitivity analysis for clinical trials with missing data via multiple imputation. *The Stata Journal*, 2016.

[17] Donald B. Rubin. *Multiple imputation for nonresponse in surveys*. Wiley, 1987.

[18] M Ichael G K Enward, Geert Molenberghs, Herbert Thijs, Bart Michiels, Geert Verbeke, and Michael G Kenward. Pattern-mixture models. In *Missing data in clinical studies*, pages 215–247. John Wiley & Sons, Ltd, Chichester, UK, 2004.

[19] C. Fletcher, S. Tsuchiya, and D.V. Mehrotra. Current practices in choosing estimands and sensitivity analyses in clinical trials. *Therapeutic Innovation & Regulatory Science*, 51(1):69–76, 2017.

[20] Jacques Emmanuel Galimard, Sylvie Chevret, Camelia Protopopescu, and Matthieu Resche-Rigon. A multiple imputation approach for MNAR mechanisms compatible with Heckman's model. *Statistics in Medicine*, 35(17):2907–2920, 2016.

[21] W. R. Gilks, C. C. Wang, B. Yvonnet, and P. Coursaget. Random-effects models for longitudinal data using Gibbs sampling. *Biometrics*, 49(2):441, 6 1993.

[22] Masahiko Gosho, Akihiro Hirakawa, Hisashi Noma, Kazushi Maruo, and Yasunori Sato. Comparison of bias-corrected covariance estimators for MMRM analysis in longitudinal data with dropouts. *Statistical Methods in Medical Research*, 26(5):2389–2406, 2017.

[23] Matthew J. Gurka, Lloyd J. Edwards, and Keith E. Muller. Avoiding bias in mixed model inference for fixed effects. *Statistics in Medicine*, 30(22):2696–2707, 2011.

[24] James J. Heckman. 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(4):475–492, 1976.

[25] Daniel F. Heitjan. Ignorability in general incomplete-data models. *Biometrika*, 81(4):701, 12 1994.

[26] Jg Ibrahim and G Molenberghs. Missing data methods in longitudinal studies: a review. *Test(Madr)*, 18(1):1–41, 2009.

[27] Joseph G Ibrahim, Ming-Hui Chen, Stuart R Lipsitz, and Amy H Herring. Missing-data methods for generalized linear models. *Journal of the American Statistical Association*, 100(469):332–346, 2005.

[28] Joseph G Ibrahim, Ming-hui Chen, Stuart R Lipsitz, Amy H Herring, Joseph G Ibrahim, Ming-hui Chen, Stuart R Lipsitz, Amy H Herring, Joseph G I Brahim, Ming-hui C Hen, Stuart R L Ipsitz, and Amy H H Erring. Missing-data methods for generalized linear models: A comparative review. *Journal of the American Statistical Association*, 1459(July), 2017.

[29] Michael G Kenward. The handling of missing data in clinical trials. *Clinical Investigation*, 3(3):241–250, 2013.

[30] Michael G Kenward and James Carpenter. Multiple imputation: current perspectives. *Statistical Methods in Medical Research*, 16(3):199–218, 2007.

[31] Gary G. Koch and Laura Elizabeth Wiener. Commentary for the Missing Data Working Group's perspective for regulatory clinical trials, estimands, and sensitivity analyses. *Statistics in Medicine*, 35(17):2887–2893, 2016.

[32] Nan M. Laird and James H Ware. Random-effects models for longitudinal data. *Biometrics*, 38(4):963, 12 1982.

[33] Lisa M Lavange. Missing data issues in regulatory clinical trials. *JPMA Symposium*, 2015.

[34] Lisa M. Lavange and Thomas Permutt. A regulatory perspective on missing data in the aftermath of the NRC report. *Statistics in Medicine*, (November 2015), 2015.

[35] Ann-Kristin Leuchs, Jörg Zinserling, Andreas Brandt, Dorothee Wirtz, and Norbert Benda. Choosing appropriate estimands in clinical trials. *Therapeutic Innovation & Regulatory Science*, 49(4):584–592, 2015.

[36] Roderick Little. *Selection and pattern-mixture models*. CRC Press, 2008.

[37] Roderick Little and Linda Yau. Intent-to-treat analysis for longitudinal studies with drop-outs. *Biometrics*, 52(4):1324, 12 1996.

[38] Roderick J. A. Little. A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association*, 83(404):1198, 12 1988.

[39] Roderick J. A. Little. Pattern-mixture models for multivariate incomplete data. *Journal of the American Statistical Association*, 88(421):125, 3 1993.

[40] Roderick J. A. Little. Modeling the drop-out mechanism in repeated-measures studies. *Journal of the American Statistical Association*, 90(431):1112, 9 1995.

[41] Roderick J A Little. Pattern-mixture models for multivariate incomplete data. *Journal of the American Statistical Association*, 88(421):125–134, 2017.

[42] Roderick J A Little and Yongxiao Wang. Pattern-mixture models for multivariate incomplete data with covariates. *Biometrics*, 52(1):98, 3 1996.

[43] G. Frank Liu and Lei Pang. On analysis of longitudinal clinical trials with missing data using reference-based imputation. *Journal of Biopharmaceutical Statistics*, (2013):150929112029004, 2015.

[44] G. Frank Liu and Lei Pang. Control-based imputation and delta-adjustment stress test for missing data analysis in longitudinal clinical trials. *Statistics in Biopharmaceutical Research*, 9(2):186–194, 2017.

[45] Kaifeng Lu. An analytic method for the placebo-based pattern-mixture model. *Statistics in Medicine*, 33(7):1134–1145, 2014.

[46] Kaifeng Lu. An extension of the placebo-based pattern-mixture model. *Pharmaceutical Statistics*, 13(2):103–109, 2014.

[47] C. Mallinckrodt, J. Roger, C. Chuang-Stein, G. Molenberghs, M. O'Kelly, B. Ratitch, M. Janssens, and P. Bunouf. Recent developments in the prevention and treatment of missing data. *Therapeutic Innovation & Regulatory Science*, 48(1):68–80, 2014.

[48] C. H. Mallinckrodt, P. W. Lane, D. Schnell, Y. Peng, and J. P. Mancuso. Recommendations for the primary analysis of continuous endpoints in longitudinal clinical trials. *Drug Information Journal*, 42(4):303–319, 2008.

[49] Craig Mallinckrodt, Geert Molenberghs, and Suchitrita Rathmann. Choosing estimands in clinical trials with missing data. *Pharmaceutical Statistics*, 2016.

[50] Craig Mallinckrodt, James Roger, Christy Chuang-stein, Geert Molenberghs, Peter W. Lane, Michael O'kelly, Bohdana Ratitch, Lei Xu, Steve Gilbert, Devan V. Mehrotra, Russ Wolfinger, and Herbert Thijs. Missing data: Turning guidance into action. *Statistics in Biopharmaceutical Research*, 5(4):369–382, 2013.

[51] Craig H. Mallinckrodt, W. Scott Clark, and Stacy R. David. Accounting for dropout bias using mixed-effects models. *Journal of Biopharmaceutical Statistics*, 11(1-2):9–21, 2001.

[52] Devan V. Mehrotra, Fang Liu, and Thomas Permutt. Missing data in clinical trials: Control-based mean imputation and sensitivity analysis. *Pharmaceutical Statistics*, (September 2016):378–392, 2017.

[53] Xiao-Li Meng. Multiple-imputation inference with uncogenial sources of input. *Statistical Science*, 9(4):538–573, 1994.

[54] By Geert Molenberghs, Caroline Beunckens, and Cristina Sotto. Every missing not at random model has got a missing at random bodyguard. *Journal of the Royal Statistical Society*, 70(2):371–388, 2008.

[55] E. Barry Moser. Repeated measures modeling with PROC MIXED. *SAS Users Group International 29*, (1):1–19, 2004.

[56] Panel on Handling Missing Data in Clinical Trials and Committee on National Statistics Division of Behavioral and Social Sciences and Education. The prevention and treatment of missing data in clinical trials panel on handling missing data in clinical trials; National research. *The National Academy of Sciences*, 2010.

[57] Lixian Peng. 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. *Department of Biostatistics, School of Public Health, Rutgers University*, 2014.

[58] Thomas Permutt. A taxonomy of estimands for regulatory clinical trials with discontinuations. *Statistics in Medicine*, (November 2015), 2015.

[59] Thomas Permutt. Sensitivity analysis for missing data in regulatory submissions. *Statistics in Medicine*, (November 2015):2014–2017, 2015.

[60] Alan Phillips, Juan Abellan-Andres, Andersen Soren, Frank Bretz, Chrissie Fletcher, Lesley France, Andrew Garrett, Raymond Harris, Magnus Kjaer, Oliver Keene, David Morgan, Michael O'Kelly, and James Roger. Estimands: discussion points from the PSI estimands and sensitivity expert group. *Pharmaceutical Statistics*, 16(1):6–11, 2017.

[61] Hui Quan, Daowen Zhang, Ji Zhang, and Laure Devlamynck. Analysis of a binary composite endpoint with missing data in components. *Statistics in Medicine*, 26(26):4703–4718, 11 2007.

[62] Bohdana Ratitch, Michael O'Kelly, and Robert Tosiello. Missing data in clinical trials: From clinical assumptions to statistical analysis using pattern mixture models. *Pharmaceutical Statistics*, 12(6):337–347, 2013.

[63] Dimitris Rizopoulos, Geert Verbeke, and Geert Molenberghs. Shared parameter models under random effects misspecification. *Biometrika*, 95(1):63–74, 1 2008.

[64] Donald B. Rubin. Inference and missing data. *Biometrika*, 63(3):581–592, 1976.

[65] Donald B Rubin. Multiple imputation after 18+ years. *Journal of the American Statistical Association*, 91(434):473, 6 1996.

[66] Alejandro Salazar, Begoña Ojeda, María Dueñas, Fernando Fernández, and Inmaculada Failde. Simple generalized estimating equations (GEEs) and weighted generalized estimating equations (WGEEs) in longitudinal studies with dropouts: guidelines and implementation in R. *Statistics in Medicine*, 35(19):3424–3448, 8 2016.

[67] A. Satty and H. Mwambi. Selection and pattern mixture models for modelling longitudinal data with dropout: An application study. *Sort*, 37(2):131–152, 2013.

[68] Joseph L. Schafer. Multiple imputation: A primer. *Statistical Methods in Medical Research*, 8(1):3–15, 1999.

[69] Joseph L. Schafer. Multiple imputation in multivariate problems when the imputation and analysis models differ. *Statistica Neerlandica*, 57(1):19–35, 2003.

[70] Shaun Seaman, John Galati, Dan Jackson, and John Carlin. What is meant by "Missing at random"? *Statistical Science*, 28(2):257–268, 2013.

[71] Shaun R. Seaman, Ian R. White, Andrew J. Copas, and Leah Li. Combining multiple imputation and inverse-probability weighting. *Biometrics*, 68(1):129–137, 2012.

[72] Weichung J. Shih, Hui Quan, and Myron N. Chang. Estimation of the mean when data contain non-ignorable missing values from a random effects model. *Statistics and Probability Letters*, 19(3):249–257, 1994.

[73] Weichung Joe Shih. Comments on the three papers by the FDA/CDER research team on the regulatory perspective of the missing data problem. *Statistics in Medicine*, 35(17):2880–2886, 2016.

[74] Weichung Joseph Shih. Problems in dealing with missing data and informative censoring in clinical trials. *Current Controlled Trials in Cardiovascular Medicine*, 3(1):4, 2002.

[75] Ohidul Siddiqui. MMRM versus MI in dealing with missing data - A comparison based on 25 NDA data sets. *Journal of Biopharmaceutical Statistics*, 21(3):423–436, 2011.

[76] Ohidul Siddiqui, H M James Hung, and Robert O Neill. MMRM VS. LOCF : A comprehensive comprehensive comparison based on simulation study and 25 NDA datasets. *Journal of Biopharmaceutical Statistics*, pages 227–246, 2009.

[77] Tao Song, Qunming Dong, Abdul J Sankoh, and Geert Molenberghs. Evaluation of overall treatment effect in MMRM. *Journal of Biopharmaceutical Statistics*, pages 1281–1293, 2013.

[78] Yongqiang Tang. An efficient monotone data augmentation algorithm for Bayesian analysis of incomplete longitudinal data. *Statistics and Probability Letters*, 104:146–152, 2015.

[79] Yongqiang Tang. Short notes on maximum likelihood inference for control-based pattern-mixture models. *Pharmaceutical Statistics*, 14(5):395–399, 2015.

[80] Yongqiang Tang. An efficient monotone data augmentation algorithm for multiple imputation in a class of pattern mixture models. *Journal of Biopharmaceutical Statistics*, 27(4):620–638, 2017.

[81] Yongqiang Tang. An efficient multiple imputation algorithm for control-based and delta-adjusted pattern mixture models using SAS. *Statistics in Biopharmaceutical Research*, 9(1):116–125, 1 2017.

[82] Yongqiang Tang. Closed-form REML estimators and sample size determination for mixed effects models for repeated measures under monotone missingness. *Statistics in Medicine*, 36(13):2135–2147, 2017.

[83] Yongqiang Tang. On the multiple imputation variance estimator for control-based and delta-adjusted pattern mixture models. *Biometrics*, 73(4):1379–1387, 2017.

[84] Herbert Thijs, Geert Molenberghs, Bart Michiels, Geert Verbeke, and Desmond Curran. Strategies to fit pattern-mixture models. *Biostatistics*, 3(2):245–265, 2002.

[85] INTERNATIONALCOUNCIL FOR HARMONISATION OF TECHNICAL REQUIREMENTS FOR PHARMACEUTICALS FOR HUMAN USE. Estimands and sensitivity analysis in clinical trials. 9(June), 2017.

[86] van Buren. *Flexible imputation of missing data*. CRC Interdisciplinary Statistics, 2012.

[87] Joanna J.J. Wang, Mark Bartlett, and Louise Ryan. Non-ignorable missingness in logistic regression. *Statistics in Medicine*, 36(19):3005–3021, 2017.

[88] Lynn Wei. An alternative way to classify missing data mechanism in clinical trials-A dialogue on missing data. *Journal of Biopharmaceutical Statistics*, 21(2):355–361, 2011.

[89] Lynn Wei and Weichung J. Shih. Partial imputation approach to analysis of repeated measurements with dependent drop-outs. *Statistics in Medicine*, 20(8):1197–1214, 2001.

[90] Ian R. White, Abdel G. Babiker, Sarah Walker, and Janet H. Darbyshire. Randomization-based methods for correcting for treatment changes: Examples from the Concorde trial. *Statistics in Medicine*, 18(19):2617–2634, 1999.

[91] Ian R. White and John B. Carlin. Bias and efficiency of multiple imputation compared with complete-case analysis for missing covariate values. *Statistics in Medicine*, 29(28):2920–2931, 2010.

[92] Ian R. White, Patrick Royston, and Angela M. Wood. Multiple imputation using chained equations: Issues and guidance for practice. *Statistics in Medicine*, 30(4):377–399, 2011.

[93] YC Yuan. Multiple imputation for missing data: Concepts and new development. *SAS Institute Inc, Rockville, MD*, pages 1–13, 2010.