Treatment Effect Estimation for Randomized Clinical Trials Subject to Noncompliance and Missing Outcomes

by

Shan Kang

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Biostatistics)
in The University of Michigan
2014

Doctoral Committee:

Professor Roderick J. Little, Co-Chair
Professor Thomas M. Braun, Co-Chair
Professor Susan A. Murphy
Professor Jeremy M.G. Taylor

To Shengchun and my parents

ACKNOWLEDGEMENTS

First and foremost, I would like to express my deepest appreciation to my co-advisors Dr. Roderick Little and Dr. Thomas Braun, who have always been insightful, patient, encouraging, and inspiring. Without their support and guidance, the dissertation would not be completed. Besides the improvement on my statistical knowledge, the structured creative thinking, scientific writing, and many professional skills that I learned from them will benefit tremendously on my career and life. I am also deeply grateful to my committee members, Dr. Susan Murphy and Dr. Jeremy Taylor for serving on my dissertation committee and providing valuable and inspiring suggestions on my research.

I am also thankful for all the knowledgeable faculty members of the Department of Biostatistics at the University of Michigan for their brilliant lectures and ideas. My thanks also go out to the staffs for their support, and all the smart students for making a wonderful environment. It is impossible to mention everybody who had impact on this work. I would like to thank all of my colleagues who had supported me throughout my graduate studies.

Finally my special thanks to my parents, Xingang Kang and Ying Liu, for their unconditional love and absolute faith in me. Thanks to my wife Shengchun, who makes every day of my life meaningful.

TABLE OF CONTENTS

LIST OF TABLES

Table

ABSTRACT

Treatment Effect Estimation for Randomized Clinical Trials Subject to Noncompliance and Missing Outcomes

by

Shan Kang

Chair: Roderick Little and Thomas Braun

Noncompliance and missing outcomes are common in randomized clinical trials. In this dissertation, we explore treatment arm switching issues for survival data and nonrandom dropout issues for masked clinical trials.

In Chapter 2, we consider noncompliance in phase III clinical trials in oncology. Although patients are randomized to their treatment assignments, the option of treatment switching may be offered to patients who experience disease recurrence for ethical considerations. Standard statistical methods that ignore this nonrandom noncompliance can lead to biased estimations. Although methods do exist to account for the effect of treatment arm switching, several of these methods focus on quantifying an overall switching effect, which can still lead to biased results if the benefit derived from switching varies among patients. We propose a new parametric method to address this limitation that factorizes the likelihood into two parts in order to evaluate the individual benefit of switching. A more robust latent event time approach is also proposed for the possible assumption violation. In simulation studies, our proposed methods outperform the existing methods.

In Chapter 3, we consider missing outcome problems in masked (blinded) clinical trials. Most standard models for analyzing the data make the missing at random (MAR) assumption, but in practice, there are often situations where missingness is likely to depend on the outcome, so MAR is not valid. For masked trials, we propose a specific missing not at random (MNAR) assumption, which we call masked MNAR (MMNAR): since the specific treatment received is unknown, masking justifies the assumption that missingness does not depend on treatment assignment after conditioning on outcomes and side effects. We suggest that methods based on the MMNAR assumption are useful for masked clinical trials, either in their own right or to provide a form of sensitivity analysis for deviations from MAR. We formulate models for categorical and continuous outcomes under this assumption. Simulations show that our proposed methods outperform other methods when MAR is violated and the efficiency of treatment effect estimates is similar to that of MAR methods when MAR is true. We apply our methods to the TRial Of Preventing HYpertension (TROPHY) study (Julius et al., 2006).

In Chapter 4, we develop regression-based multiple imputation models that exploit the MMNAR assumption proposed in Chapter 3 for longitudinal data. Simulation studies are conducted to compare the performance of the proposed method with other methods. The idea is also illustrated with the TROPHY study.

CHAPTER I

Introduction

Randomized controlled trials (RCTs) are the gold standard for investigating the effectiveness of a medical intervention, because the randomization of patients to the experimental treatment arm or comparative control protects the comparison between two arms from selection bias and confounders of treatment effect. For an ideal clinical experiment, after participants are randomly assigned to treatment arms, they actually receive the interventions exactly as specified in the protocols, and provide measures of outcome. In this case, randomization, complete compliance, and no loss of follow up enable us to draw valid statistical inferences for the treatment effect. In this cases, the 'double blind' experiments are preferred, because this approach conceals the treatment assignment from the subject and the clinician to prevent the potential bias caused by knowing this information.

In practice, such conditions are difficult to achieve because of a variety of human behaviors. As a result, missing data are common in many clinical trials. For example, subjects may (partially) fail to take their assigned treatments, or receive an alternative treatment as a rescue therapy. This treatment discontinuation can be conceptualized as a form of noncompliance (Rubin, 1987) and must be accounted for properly. When participants discontinue their assigned treatment, their outcomes can still be recorded, and those who remain on their assigned treatment can have

missing outcomes. For example, when subjects miss their clinic visits, outcomes are not recorded but they still receive assigned treatment. Or the study is terminated for administrative reasons, all outcomes are missing but there is no evidence to show that the participants would discontinue their treatment if the study was still going on. It is important to distinguish between treatment discontinuation and missing outcomes (Meinert, 1980), although they often appear together, as when the subjects are lost to follow-up, the underlying reason could be noncompliance and the outcomes are missing.

There are many possible reasons for noncompliance, but as they are often not random, the benefits of randomization are undermined. The most widely accepted approach to handle noncompliance is the 'intention-to-treat' (ITT) analysis. In ITT analysis, all randomized patients are analyzed according to the treatment they were originally assigned, regardless of the actual treatment received or subsequent withdrawal from treatment (Hill, 1961). In ITT analysis, the entire treatment regimen including treatment discontinuation is evaluated as a whole. This approach tends to estimate a diluted treatment effect, sometimes called treatment effectiveness, which is generally a conservative estimate comparing with treatment efficacy, the effectiveness of a treatment when it is in fact taken (Little et al., 2009).

An alternative to ITT analysis is 'per-protocol' (PP) analysis, in which only the patients who comply with the assigned treatment are counted towards the final result. In PP analysis, the treatment efficacy is estimated. However, with this approach, selection bias may be a problem, because people who experience better outcomes are more likely to remain in the data set, and PP analysis may provide overoptimistic estimates of the efficacy of the treatment resulting from the removal of non-compliers.

Several existing methods for analyzing complete data sets are briefly reviewed here. Most of the missing data approaches introduced below are based on them. Maximum likelihood (ML) methods maximize the likelihood function, which is based

on the specified joint density function for all observations. ML has many good properties including consistency, asymptotic normality, functional invariance, asymptotic efficiency, and many others under some regularity conditions. Bayesian inference method is another approach. This approach uses Bayes' rule to update the probability estimate when additional evidence is acquired. For longitudinal data, mixed effect models can be used to introduce correlation between observations. Another approach is to specify the variance covariance matrix directly. Generalized estimating equations (GEE) (Liang and Zeger, 1986) are also popular in estimating the parameters of a model with a possible unknown correlation between outcomes, which is because parameter estimates from the GEE are consistent even when the covariance structure is misspecified.

When outcomes are missing, one common approach is complete case (CC) analysis, in which only subjects who have all variables observed are used. CC analysis results are unbiased when the missingness is independent of both observed and unobserved variables, which is called missing completely at random (MCAR). Although CC analysis is the default option in many statistical software, it is criticized because MCAR is generally unrealistic and even if MCAR is assumed, it discards the information in the incomplete cases. CC analysis is also valid when the missingness does not depend on outcomes in a regression, in which no distribution assumption is made for the covariates.

Alternative missing data adjustment methods often assume that the missing data are missing at random (MAR), in the sense that missingness does not depend on the missing values after conditioning on the observed data (Rubin, 1976). Approaches based on MAR include ignorable likelihood-based methods such as parametric multiple imputation (MI) (Rubin, 1987) and maximum likelihood (ML) estimation, and inverse probability-weighted methods (Robins et al., 1995). Ignorable likelihood-based methods have the advantage of retaining all the data, and are fully efficient

as long as MAR is true. Furthermore, they are robust, performing reasonably well even when the MAR assumption is slightly violated (Little and Zhang, 2011). The robustness is because the efficiency gain by using more cases outweighs the bias resulting from incorrectly ignoring the missing data mechanism. In multiple imputation, missing values are replaced by D (=5 or 10) imputed versions. For inference, each of the imputed complete data sets is analyzed by standard methods, and those results are combined to calculate estimates and confidence intervals which can incorporate uncertainty from missing data.

When data on outcomes are missing because of adverse events or lack of treatment efficacy, missingness depends on the missing values and the MAR assumption is violated. Such mechanisms are called missing not at random (MNAR), or non-ignorable (Rubin, 1976; Little and Rubin, 2002). Joint models for the outcomes and missing indicators are usually required for non-ignorable missing data mechanisms. Little (1993) considered two classes of joint models: selection models and pattern-mixture models. Selection models factorize the joint model of outcomes and missing indicators into a model for the marginal outcomes and another model of missing indicators conditional on the possibly unobserved outcomes. Pattern-mixture models factorize the joint model into a model of missing data patterns and another model for outcomes given missing data patterns. In either approach, unverifiable restrictions or assumptions are required to identify the parameters (Fitzmaurice et al., 2005). Researchers have proposed sensitivity analyses to address this issue (Little and Rubin, 2002; Scharfstein et al., 1999; National Research Council, 2010), but in practice only a limited set of MNAR models can be assessed.

In this dissertation, we explore two missing data problems in clinical trials: treatment arm switching issues for survival data and nonrandom dropout issues for masked clinical trials.

In Chapter 2, we consider noncompliance problem in phase III clinical trials in

oncology, which evaluate the survival benefit for a new treatment relative to an existing treatment or standard-of-care. Although patients are randomized to their treatment assignments, ethical motivations dictate that patients who experience disease recurrence or other event indicating increased likelihood of death may be offered the option to switch to the other treatment arm and continue to be followed for survival. Standard statistical methods that ignore this non-random noncompliance can lead to biased estimation of the survival benefit attributed to the new treatment. Although methods do exist to account for the effect of treatment arm switching, several of these methods focus on quantifying an overall switching effect, which can still lead to biased results if the benefit derived from switching varies among patients.

We propose a new parametric method to address this limitation that factorizes the likelihood into two parts in order to evaluate the individual benefit of switching. For the cases when assumptions in the above parametric method may be violated, we propose another more robust latent time method inspired by iterative parameter estimation (IPE) procedure (Branson and Whitehead, 2002). Via simulation, we examine the performance of our methods and compare the performance with existing methods including (1) ITT analysis; (2) PP analysis in which patients who switch treatment arms are either omitted entirely from analysis or are treated in the analysis as censored at the time of switching; (3) Cox proportional hazards model (PH) with treatment arm as a time-varying covariate; (4) IPE procedure which assumes a parametric accelerated failure-time (AFT) model with a multiplicative treatment effect of $\exp(\eta)$, and iteratively calculate the value of $\hat{\eta}$ and latent event times until values for $\hat{\eta}$ converge.

In Chapter 3, we consider missing outcome problems in masked (blinded) clinical trials. Most standard models for analyzing the data make the MAR assumption, but in practice, there are often situations where missingness is likely to depend on the outcome, so MAR is not valid. For masked trials, we propose a specific MNAR

assumption, which we call masked MNAR (MMNAR): since the specific treatment received is unknown, masking justifies the assumption that missingness does not depend on treatment assignment after conditioning on outcomes and side effects. We suggest that methods based on the MMNAR assumption are useful for masked clinical trials, either in their own right or to provide a form of sensitivity analysis for deviations from MAR. MAR analysis might be favored on grounds of efficiency if the estimates based on MMNAR and MAR are similar, but if the estimates are substantially different, the MMNAR estimates might be preferred because the mechanism is more plausible. We formulate models for categorical and continuous outcomes under this assumption. Simulations are conducted to examine the finite sample performance of ML methods assuming MMNAR and compare them with other methods such as CC analysis and ML assuming MAR. We also applied our methods to the TRial Of Preventing HYpertension (TROPHY) study (Julius et al., 2006).

In Chapter 4, we further extend the MMNAR assumption proposed in Chapter 3 to longitudinal data models. Since there are many possible longitudinal models, we propose a strategy to develop a longitudinal MMNAR model based on the preferred longitudinal model for complete data. The estimation procedure including a regression step and an imputation step is also presented. Simulation studies are conducted to compare the performance of the proposed method with other methods. The idea is also illustrated with the TROPHY study.

CHAPTER II

Estimating Survival Benefit in Randomized Clinical Trials with Treatment Arm Switching After Disease Progression

2.1 Introduction

Randomized controlled trials (RCTs) are often considered the gold standard for evaluating the efficacy of an experimental treatment, as the randomization of patients to the treatment arm or the comparative control arm limits the impact of selection bias and possible confounding of the potential treatment effect. After randomization, each patient is followed for a pre-specified period of time, often several years, during which the occurrence of a primary event of interest, usually death from any cause, may occur. Patients who do not die before the end of their follow-up are considered censored for death, necessitating the use of censored data methods, i.e. log-rank tests and Cox regression, to compare the difference between the overall survival (OS) rates of the treatment arms.

In many diseases, such as cancer and acquired immune deficiency syndrome (AIDS), death may be preceded by a serious event, such as recurrence of disease with cancer or continued elevation of CD34+ cells with AIDS, that is associated with death and would suggest that the treatment assignment of the patient be switched for compas-

sionate and ethical reasons. For example, Slamon et al. (2001) described an RCT comparing the use of standard chemotherapy (anthracycline plus cyclophosphmide or paclitaxel) alone or in combination with the recombinant monoclonal antibody trastuzimab (Herceptin) for the treatment of metastatic breast cancer. At the time of disease progression, all patients had the option of enrolling in a follow-up (non-randomized) study of Herceptin alone or in combination with chemotherapy in hopes of prolonging overall survival. Of the 234 patients randomized to chemotherapy alone, 66% elected to receive Herceptin as part of this follow-up study. The combination of the data from the the RCT and the follow-up study were then used to assess the effect of Herceptin on overall survival (Lewis et al., 2002).

A second example is the RECORD-1 trial, which studied everolimus, which alters the mammalian target of rapamyacin (mTOR) pathway, a known pathway of the pathogenesis of renal cell carcinoma (RCC) (Motzer et al., 2008). RECORD-1 enrolled 416 patients with metastatic RCC, each of whom was randomized to receive either everolimus plus best standard of care (BSC) or placebo plus BSC. The primary endpoint was progression-free survival (PFS), which is the earlier of disease-progression or death, because the treatment assignment of patients with disease progression were unblinded, and those randomized to placebo plus BSC were allowed to switch to everolimus plus BSC at the time of disease progression. However, the published PFS results were accompanied by an editorial that questioned the use of PFS as an endpoint for RCC (Knox, 2008), a concern that has now pervaded through the design of RCTs in cancer treatment, with most RCTs now using OS as the primary endpoint. In response to the editorial, the OS results of the RECORD-1 trial, based upon the rank-preserving structural failure time model of Robins and Tsiatis (1991) were published later (Motzer et al., 2010). Other examples of treatment switching and the resulting statistical issues include the use of sunitinib for the treatment of gastrointestinal stromal tumors (Blay, 2010), the study of ganitumab in the treatment

of breast cancer (Robertson et al., 2013), and the study of zidovudine in patients with HIV infection (Concorde Coordinating Committee, 1994; White et al., 1999; Hernan et al., 2000).

It is obvious that when OS is the primary endpoint for a RCT that allows patients to switch from their randomized assignments, comparison of the two arms will be biased unless the switching is accounted for properly. When using standard statistical methods to estimate the difference in OS between the two arms, the simplest and most-commonly used approach is an intent-to-treat (ITT) analysis, whereby the follow-up of each patient is assigned completely to their original arm assignment, regardless if they later switched to the other arm. However, if the treatment is truly effective, then an ITT analysis will only serve to give a diluted treatment effect estimate. Furthermore, ITT produces an estimate of the effect of the entire treatment regimen that includes switching, whereas usually investigators are interested solely in the effect in the treatment when switching is not an option in practice, a parameter referred to as "biological efficacy" by Sommer and Zeger (2011).

Another method is termed a "per-protocol" approach, in which patients who switch treatment arms are either omitted entirely from analysis or are treated in the analysis as censored at the time of switching. Neither of these approaches is satisfactory, as both approaches fail to use all information available in the data, thereby reducing the power of the study. Moreover, both per-protocol approaches lead to a biased analysis of a selected subset of patients who are no longer balanced with respect to all confounding factors. The bias is further compounded when the probability of switching is related to disease progression, as it is in all the RCTs cited earlier. We note that Law and Kaldor (1996) proposed a modified ITT analysis approach that seems to have received little use in application, although it has been shown to lead to valid estimation under a weaker set of restrictions than a crude ITT analysis. A more detailed examination of the weaknesses of ITT and per-protocol approaches can

be found in Morden et al. (2011).

Several statistical approaches have been developed that incorporate the complete follow-up of all patients and give an unbiased estimate of treatment effect when certain assumptions are met. As stated earlier, data from the RECORD-1 trial were analyzed using the methods of Robins and Tsiatis (1991), which assume that treatment has a multiplicative effect (e^{η}; the model with η will be described later in detail) on each patient's overall survival. Each patient is assigned a latent event time that would have been observed had no treatment been received, and a rank-based statistic, given a value of η, is computed using the latent values. A grid search among a range of possible values for η is used until one identifies the value that leads to a rank-based statistic of zero; this value of η is the estimated treatment effect $\hat{\eta}$. However, this grid search can be computationally expensive, as no systematic or iterative process exists for identifying $\hat{\eta}$.

In response to this limitation, Branson and Whitehead (2002) proposed an iterative parameter estimation (IPE) procedure. A parametric (e.g. Weibull or exponential) accelerated failure-time model is assumed with a multiplicative treatment effect of e^{η}, and $\hat{\eta}$ is computed from the data as randomized. Based upon this value of $\hat{\eta}$, latent event times are then formulated for patients who switched treatment and these latent event times replace the corresponding observed event times in the data. A updated value of $\hat{\eta}$ is then computed, from which one formulates updated latent event times, and this iterative process continues until values for $\hat{\eta}$ converge. Thus, as stated by White (2006), the methods of Robins and Tsiatis (1991) and Branson and Whitehead (2002) assume the same estimand and differ only by whether a rank or parametric test is used to determine $\hat{\eta}$. However, White (2006) does emphasize that the methods of Branson and Whitehead (2002) as published do not adequately deal with censored subjects because they do not correctly recensor subjects as outlined by White et al. (1999). Thus, in a study with a large amount of censoring, the value of

$\hat{\eta}$ can be biased with IPE.

Shao et al. (2005) also extended the IPE procedure by proposing to use a Cox regression model, rather than a fully parametric model, for survival times, and thereby have a likelihood-based approach for estimating η which is computationally simpler than the iterative approach of Branson and Whitehead (2002). Shao et al. (2005) also introduce a weight function that allows for a differential treatment effect for patients randomized initially to the treatment arm and patients randomized initially to control and later switch to the treatment arm. Nonethless, White (2006) demonstrated that these methods are based upon a likelihood that is conditional upon the switching times and therefore can be biased when switching is not ignorable, i.e. is correlated with prognosis. Most recently, Zheng et al. (2012) developed a computationally-intensive method based on a semi-parametric hazards model. Instead of calculating latent event times, Zheng et al. (2012) only model the observed event times and instead use separate models for each of: (a) the time of the event causing a switch to another arm, (b) the time from switching to death, and (c) the time of death for those who did not switch from their original arm assignment.

We present a parametric approach as an alternative to the methods just discussed. In Section 2.2, we propose our new parametric model and Section 2.3 contains the details of parameter estimation. In Section 2.4, we examine the performance of our methods via simulation. In Section 2.5, we compare the performance of our methods to existing approaches and also present two approaches for limiting the bias of the IPE algorithm applied to multivariate normal data. Concluding remarks are in Section 2.6.

2.2 Parametric Model

We have a clinical trial designed to assess the difference in overall survival between patients receiving a control and patients receiving an experimental agent (treatment).

Patients are initially randomized equally to either arm ($j = 0$ for control; $j = 1$ for treatment) and each is followed until the earliest of (a) censoring, (b) progression of disease, or (c) death. Patients on the control (treatment) arm with progression of disease are allowed to switch to the treatment (control) arm at the time their disease progresses. We let $X_i, i = 1, 2, \ldots n$ denote the arm to which patient i was assigned and Y_i denotes the arm assignment for patient i after disease progression in patients whose disease progresses. We emphasize that disease progression does not imply a patient will switch to the other arm with certainty, as some patients most likely will remain on their original arm even after disease progression. To that end, M_{ij} is the binary indicator of whether or not subject i experienced disease progression after randomization to arm j, and R_{ij} denotes if patient i switched treatment, i.e $X_i = j$ and $X_i \neq Y_i$. Thus, M_{i,X_i} is the observed indicator of progression for subject i, while $M_{i,(1-X_i)}$ is a latent indicator of progression for subject i had they been randomized to the other arm. We let $p_{1j} = Prob(M_{ij} = 1)$ and $p_{0j} = (1 - p_{1j}) = Prob(M_{ij} = 0)$.

We let T_{i0} and G_{i0} be the respective time to progression and time to death after progression in patients initially randomized to control, and TG_{i0} is the time from randomization to death in patients randomized to control without disease progression. We have similar definitions of T_{i1}, G_{i1}, and TG_{i1} for patients randomized to the treatment arm. Note that under this model, G_{ij}, the time to death for patient i in arm j after disease progression, is independent of the original treatment assignment of patient i. Specifically, if patient i were originally assigned to the control arm and did not switch to the treatment arm after disease progression, a period of time G_{i0} would be observed for this patient. By our assumption, if that patient had instead been assigned to the treatment arm and switched to the control arm after progression, the same length of time G_{i0} would be observed after progression. Last, we let C_i denote the censoring time for patient i. We further assume that the decision of treatment switching is independent of the potential outcomes G_{i0} and G_{i1} given the

progression time T_{ij}, and the censoring time is also assumed to be independent of each of T_{ij}, G_{ij}, and TG_{ij}. In summary, the observed data for each patient are one of twelve possibilities outlined in Table 2.1.

Table 2.1: Summary of the data to be observed in the clinical trial, dependent upon disease progression, treatment arm switching, and censoring.

Initial Arm	Censored	Progression	Switch	Observed Data for Patient i
Control	no	no	n/a	$X_i = 0$, $M_{i0} = 0$, TG_{i0}
		yes	no	$X_i = 0$, $M_{i0} = 1$, $R_{i0} = 0$, T_{i0}, G_{i0}
		yes	yes	$X_i = 0$, $M_{i0} = 1$, $R_{i0} = 1$, T_{i0}, G_{i1}
Treatment	no	no	n/a	$X_i = 1$, $M_{i1} = 0$, TG_{i1}
		yes	no	$X_i = 1$, $M_{i1} = 1$, $R_{i1} = 0$, T_{i1}, G_{i1}
		yes	yes	$X_i = 1$, $M_{i1} = 1$, $R_{i1} = 1$, T_{i1}, G_{i0}
Control	yes	no	n/a	$X_i = 0$, C_i
		yes	no	$X_i = 0$, $M_{i0} = 1$, $R_{i0} = 0$, T_{i0}, C_i
		yes	yes	$X_i = 0$, $M_{i0} = 1$, $R_{i0} = 1$, T_{i0}, C_i
Treatment	yes	no	n/a	$X_i = 1$, C_i
		yes	no	$X_i = 1$, $M_{i1} = 1$, $R_{i1} = 0$, T_{i1}, C_i
		yes	yes	$X_i = 1$, $M_{i1} = 1$, $R_{i1} = 1$, T_{i1}, C_i

Thus, in the absence of censoring, we have a total of six duration of times for subject i: $T_{i0}, G_{i0}, TG_{i0}, T_{i1}, G_{i1}$, and TG_{i1}, some of which are observed and some of which are latent depending on the original arm assignment and whether or not disease progression occurs. We will directly model the joint distribution of T_{i0}, G_{i0}, TG_{i0}, and T_{i1}, and then make assumptions that allow us to determine distributions for G_{i1} and TG_{i1}. Specifically, the vector $(\log T_{i0}, \log T_{i1}, \log G_{i0}, \log TG_{i0})$ has a multivariate normal distribution with mean vector $(\mu_0, \mu_1, \mu_2, \mu_{0E})$ and covariance matrix

$$\begin{pmatrix} \sigma_0^2 & \rho_1\sigma_0\sigma_1 & \rho_2\sigma_0\sigma_2 & \rho_4\sigma_0\sigma_{0E} \\ \rho_1\sigma_0\sigma_1 & \sigma_1^2 & \rho_3\sigma_1\sigma_2 & \rho_5\sigma_1\sigma_{0E} \\ \rho_2\sigma_0\sigma_2 & \rho_3\sigma_1\sigma_2 & \sigma_2^2 & \rho_6\sigma_2\sigma_{0E} \\ \rho_4\sigma_0\sigma_{0E} & \rho_5\sigma_1\sigma_{0E} & \rho_6\sigma_2\sigma_{0E} & \sigma_{0E}^2 \end{pmatrix} \tag{2.1}$$

We define $D_{Ti} = \log T_{i1} - \log T_{i0}$, which is also normally distributed and has mean $\Delta = \mu_1 - \mu_0$ and variance $\sigma_D^2 = \sigma_1^2 + \sigma_0^2 - 2\rho_1\sigma_0\sigma_1$. The parameter Δ quantifies the treatment effect of interest, and, as defined, holds only for the time before progression in patients with disease progression. We therefore assume that this treatment effect carries over to the time from progression to death, as well as the time to death in patients who do not experience disease progression. Specifically, we assume that $D_{Gi} = \log G_{i1} - \log G_{i0}$ and $D_{TGi} = \log TG_{i1} - \log TG_{i0}$ are both equal to D_{Ti}, i.e. $D_{Ti} \equiv D_{Gi} \equiv D_{TGi}$, and this assumption leads to distributions for G_{i1} and TG_{i1} and establishes a fully parametric model.

Although our assumption that the treatment effect is the same both before and after disease progression may seem strong, it actually leads to intuitive results. In patients randomized to control who experience disease progression, we would expect that the effect of the treatment in these patients would be related to when their disease progressed. Specifically, the benefit of the treatment is expected to be greater in those whose disease progressed earlier than average as compared to those whose disease progressed later than average, a concept referred to by White (2006) as "individual benefit to be gained from treatment," which we denote as B_i.

This concept can be explained by our proposed model. If patient i is assigned to the control arm and experiences disease progression at time t_{i0}, the expected conditional treatment effect after progression is

$$\begin{aligned}
E(\log G_{i1} - \log G_{i0} | \log T_{i0} = \log t_{i0}) &= E(\log T_{i1} - \log T_{i0} | \log T_{i0} = \log t_{i0}) \\
&= E(\log T_{i1} | \log T_{i0} = \log t_{i0}) - \log t_{i0} \\
&= \mu_1 + \rho_1\sigma_0/\sigma_1(\log t_{i0} - \mu_0) - \log t_{i0} \\
&= \mu_1 - \mu_0 + \rho_1\sigma_0/\sigma_1(\log t_{i0} - \mu_0) - \log t_{i0} + \mu_0 \\
&= \Delta + (\rho_1\sigma_0/\sigma_1 - 1)(\log t_{i0} - \mu_0)
\end{aligned}$$

with $B_i = (\rho_1\sigma_0/\sigma_1 - 1)(\log t_{i0} - \mu_0)$. Since we expect σ_0 and σ_1 to be relatively similar to each other and ρ will not be large enough to dominate the ratio of σ_0 and σ_1, it is very likely in practice that $(\rho_1\sigma_0/\sigma_1 - 1) < 0$. Thus, $B_i > 0$ when $\log t_{i0} < \mu_0$ and $B_i < 0$ when $\log t_{i0} > \mu_0$, meaning that the individual treatment benefit from switching is large for patients with early progression relative to patients with average time to progression. In fact, our model presumes that those who progress later than average will receive less benefit from switching to the treatment arm than those with average time to progression. Therefore, the concept of a positive benefit attributed to switching is supported by our model independent of the sign of $\mu_1 - \mu_0$, and switching a patient's treatment arm assignment based on their observed time to progression is justified.

We note that our methods define the treatment effect as $\Delta = E\{\log(T_{i1}/T_{i0})\}$, so that $e^{\Delta} \approx E\{T_{i1}/T_{i0}\}$. We contrast this definition of treatment effect with the following two alternatives

$$\Delta^* = \log\left(\frac{E[M_{i1}(T_{i1} + G_{i1})] + E[(1 - M_{i1})TG_{i1}]}{E[M_{i0}(T_{i0} + G_{i0})] + E[(1 - M_{i0})TG_{i0}]}\right) \tag{2.2}$$

$$\Delta^{**} = E\left[\log\left(\frac{M_{i1}(T_{i1} + G_{i1}) + (1 - M_{i1})TG_{i1}}{M_{i0}(T_{i0} + G_{i0}) + (1 - M_{i0})TG_{i0}}\right)\right] \tag{2.3}$$

Although Δ, Δ^*, and Δ^{**} are not identical quantities, it can be shown that $\Delta = \Delta^{**}$ if either $p_{10} = p_{11}$, i.e. the probability of progression is the same in both arms, or $T_{i0} + G_{i0} = TG_{i0}$, i.e. the sum of the time to progression and time from progression to death for a patient with progression is the same as the time to death for a patient without progression. However, both of these equalities have little biological justification. Although Δ^{**} is related more closely to Δ, Δ^* is the most commonly defined treatment effect because it can be directly obtained from the data in the absence of treatment switching and censoring. In Section 2.4, estimates of Δ,

Δ^*, and Δ^{**} will all be presented as summaries of the performance of our proposed method and estimates of Δ^* will be presented when comparing our proposed methods to existing methods.

2.3 Likelihood and Parameter Estimation

Since we have a fully parametric model, we can use maximum-likelihood methods to estimate parameters, although some of the parameters are not estimable and will be discussed in further detail. As outlined in Table 2.1, the contribution of each subject to the likelihood is a function of their original treatment assignment, whether or not they experience disease progression, whether they switch treatment arms after disease progression, and when they are censored. We let $\theta = \{p_{10}, p_{11}, \mu_0, \mu_1, \mu_2, \mu_{0E}, \sigma_0^2, \sigma_1^2, \sigma_2^2, \sigma_{0E}^2, \sigma_{1E}^2, \rho_1, \rho_2, \rho_3\}$ be the vector of the 14 parameters to be estimated. The parameter σ_{1E}^2 will be described in the outline below, and is used in the likelihood in place of ρ_4, ρ_5, and ρ_6, which are not estimable from the data. We let $K_i = 1, 2, \ldots 12$ be the indicator as to which of the twelve possibilities subject i resides and their corresponding likelihood contribution to be L_{K_i}, leading to a joint likelihood function $\prod_{i=1}^{n} L_{K_i}$. We now outline the twelve possible likelihood contributions of each subject in more specific terms. In the following outline, $\phi(x \mid \mu, \sigma^2)$ and $\Phi(x \mid \mu, \sigma^2)$ denote the density and cumulative probability function, respectively, of a normal distribution with mean μ and variance σ^2. We also organize our outline by first describing the six possibilities when patients are not censored, which are then followed by the corresponding possibilities when patients are censored.

- $K_i = 1$: patient i is assigned to the control arm and does not experience disease progression. We observe their time to death TG_{i0} and $L_1 = \phi(\log TG_{i0} \mid \mu_{0E}, \sigma_{0E}^2)$;
- $K_i = 2$: patient i is assigned to the control arm, experiences disease progression, and does not switch to the treatment arm. We observe their time to disease

progression, T_{i0}, and their time from disease progression to death, G_{i0}. Since (T_{i0}, G_{i0}) is not a random sample, their likelihood contribution cannot be based upon their joint distribution. Instead, we use the product of the conditional distribution of G_{i0} given T_{i0} and the marginal distribution of T_{i0}. Since $\log G_{i0}$ given $T_{i0} = t_{i0}$ has a normal distribution with mean $\tilde{\mu}_{i2} = \mu_2 + (\rho_2\sigma_2/\sigma_0)(\log t_{i0} - \mu_0)$ and variance $\tilde{\sigma}_2^2 = \sigma_2^2(1-\rho_2^2)$, we have $L_2 = \phi(\log T_{i0} \mid \mu_0, \sigma_0^2)\phi(\log G_{i0} \mid \tilde{\mu}_{i2}, \tilde{\sigma}_2^2)$;

- $K_i = 3$: patient i is assigned to the control arm, experiences disease progression, and switches to the treatment arm. We observe their time to disease progression, T_{i0}, and time from disease progression to death, G_{i1}. Through our assumption that $\log G_{i1} - \log G_{i0} = \log T_{i1} - \log T_{i0}$, we can derive the mean and variance of G_{i1} conditional on $T_{i0} = t_{i0}$, whose respective values are $\tilde{\mu}_{i3} = \mu_1 + \mu_2 - \log t_{i0} + (\rho_1\sigma_1/\sigma_0 + \rho_2\sigma_2/\sigma_0)(\log t_{i0} - \mu_0)$ and $\tilde{\sigma}_3^2 = \sigma_1^2(1-\rho_1^2) + \sigma_2^2(1-\rho_2^2) + 2\sigma_1\sigma_2(\rho_3 - \rho_1\rho_2)$. Thus, $L_3 = \phi(\log T_{i0} \mid \mu_0, \sigma_0^2)\phi(\log G_{i0} \mid \tilde{\mu}_{i3}, \tilde{\sigma}_3^2)$;

- $K_i = 4$: patient i is assigned to the treatment arm and does not experience disease progression. We observe their time to death, $\log TG_{i1}$ and from our assumption that $\log TG_{i1} - \log TG_{i0} = \log T_{i1} - \log T_{i0}$, we know that $\log TG_{i1}$ has a normal distribution whose variance is a function of the inestimable parameters ρ_4, ρ_5, and ρ_6. Thus, we introduce σ_{1E}^2 to denote the variance of $\log TG_{i1}$ and use this parameter directly in the likelihood. Thus, $L_4 = \phi(\log TG_{i1} \mid \tilde{\mu}_4, \sigma_{1E}^2)$, in which $\tilde{\mu}_4 = \mu_{0E} - \mu_0 + \mu_1$.

- $K_i = 5$: patient i is assigned to the treatment arm, experiences progression, and does not switch to the control arm. We observe their time to disease progression, T_{i1}, and time from disease progression to death, G_{i1}. Analogous to the likelihood when $K_i = 3$, we first derive the conditional mean and variance of $\log G_{i1} | T_{i1} = t_{i1}$, which are $\tilde{\mu}_{i5} = \mu_2 - \mu_0 + \log t_{i1} + (\rho_3\sigma_2/\sigma_1 - \rho_1\sigma_0/\sigma_1)(\log t_{i1} - \mu_1)$ and $\tilde{\sigma}_5^2 = \sigma_0^2(1-\rho_1^2) + \sigma_2^2(1-\rho_3^2) - 2\sigma_0\sigma_2(\rho_2 - \rho_1\rho_3)$.

Thus, $L_5 = \phi(\log T_{i1} \mid \mu_1, \sigma_1^2)\phi(\log G_{i1} \mid \tilde{\mu}_5, \tilde{\sigma}_5^2)$.

- $K_i = 6$: patient i is assigned to the treatment arm, experiences progression, and switches to the control arm. We observe their time to progression, T_{i1}, and time from disease progression to death, G_{i0}. Analogous the likelihood when $K_i = 2$, we see that $\log G_{i0}|T_{i1} = t_1$ has a normal distribution with mean of $\tilde{\mu}_{i6} = \mu_2 + (\rho_3\sigma_2/\sigma_1)(\log t_{i1} - \mu_1)$ and variance $\tilde{\sigma}_6^2 = \sigma_2^2(1 - \rho_3^2)$. Thus, $L_6 = \phi(\log T_{i1} \mid \mu_1, \sigma_1^2)\phi(\log G_{i0} \mid \tilde{\mu}_{i6}, \tilde{\sigma}_6^2)$.

- $K_i = 7$: patient i is assigned to the control arm and is censored before either progression or death occurs. We observe their time of censoring, C_i, and the likelihood contribution is a mixture of patients who would have otherwise died in the absence of disease progression and patients who would have otherwise experienced disease progression prior to death. Thus, $L_7 = (1 - p_{10})[1 - \Phi(\log C_i \mid \mu_{0E}, \sigma_{0E}^2)] + p_{10}[1 - \Phi(\log C_i \mid \mu_0, \sigma_0^2)]$.

- $K_i = 8$: patient i is assigned to the control arm, does not switch arms at time of disease progression, and is censored before death. We observe their time to disease progression, T_{i0}, and time from disease progression to censoring, $U_i = C_i - T_{i0}$. Thus, $L_8 = \phi(\log T_{i0} \mid \mu_0, \sigma_0^2)[1 - \Phi(\log U_i \mid \tilde{\mu}_2, \tilde{\sigma}_2^2)]$.

- $K_i = 9$: patient i is assigned to the control arm, switches arm at time of disease progression, and is censored before death. We observe their time to disease progression, T_{i0}, and time from disease progression to censoring, $U_i = C_i - T_{i0}$. Thus, $L_9 = \phi(\log T_{i0} \mid \mu_0, \sigma_0^2)[1 - \Phi(\log U_i \mid \tilde{\mu}_3, \tilde{\sigma}_3^2)]$.

- $K_i = 10$: patient i is assigned to the treatment arm and is censored before either disease progression or death occurs. We observe their time of censoring, C_i. Analogous to the likelihood when $K_i = 7$, we have $L_{10} = p_{11}[1 - \Phi(\log C_i \mid \mu_1, \sigma_1^2)] + (1 - p_{11})[1 - \Phi(\log C_i \mid \tilde{\mu}_4, \sigma_{1E}^2)]$.

- $K_i = 11$: patient i is assigned to the treatment arm, does not switch arms at time of disease progression, and is censored before death. We observe their time to disease progression T_{i1}, and time from disease progression to censoring $U_i = C_i - T_{i1}$. Thus, $L_{11} = \phi(\log T_{i1} \mid \mu_1, \sigma_1^2)[1 - \Phi(\log U_i \mid \tilde{\mu}_5, \tilde{\sigma}_5^2)]$.

- $K_i = 12$: patient i is assigned to the treatment arm, switches arms at time of disease progression, and is censored before dead. We observe their time to disease progression, T_{i1}, and time from disease progression to censoring, $U_i = C_i - T_{i1}$. Thus, $L_{12} = \phi(\log T_{i1} \mid \mu_1, \sigma_1^2)[1 - \Phi(\log U_i \mid \tilde{\mu}_6, \tilde{\sigma}_6^2)]$.

2.4 Simulation Studies

We examine the finite sample properties of our proposed methods via simulation. We have a hypothetical clinical trial comparing an experimental agent and a control. 400 patients are randomized so that each arm has 200 patients. We first generate all the possible event times (progression, death after progression, and death in absence of progression) for each subject. We first draw the vector $(\log T_{i0}, \log T_{i1}, \log G_{i0}, \log TG_{i0})$ from a multivariate normal distribution with mean vector $\mu = (-1.3, -1.1, -1.2, -0.5)$ and variance matrix with diagonal elements $\sigma_0^2 = 2.0$, $\sigma_1^2 = 1.8$, $\sigma_2^2 = 1.3$, and $\sigma_{0E}^2 = 2.0$ and off-diagonal elements such that $\rho_1 = 0.58, \rho_2 = 0.50, \rho_3 = 0.10, \rho_4 = 0.40, \rho_5 = 0.39$, and $\rho_6 = 0.30$. From the given values of $\log T_{i0}$, $\log T_{i1}$, $\log G_{i0}$, and $\log TG_{i0}$, we compute $\log G_{i1} = \log G_{i0} + \log T_{i1} - \log T_{i0}$ and $\log TG_{i1} = \log TG_{i0} + \log T_{i1} - \log T_{i0}$. We also draw independent censoring times C_i for each subject from a exponential distribution with mean 2. We then simulate a binary indicator of progression for each subject using the probabilities $p_{11} = 0.5$ and $p_{10} = 0.6$. If a subject experiences progression, their data is comprised of either (i) their censoring time, (ii) their separate times to disease progression and from disease progression to censoring, or (iii) their separate times to disease progression and from disease progression to death. The data

for subjects without disease progression are either (i) their censoring time or (ii) the time to death.

Recall that when a patient has disease progression, it is possible for them to switch to the arm to which they were not randomized if their expected individual benefit is large enough to warrant switching. Thus, a rule must be formulated that determines if a patient will switch arms at disease progression. We first examine the properties of our methods when switching occurs if a patient has disease progression that occurs earlier than the median progression time for their arm, which is $\exp(-1.1) = 0.33$ for the treatment arm and $\exp(-1.3) = 0.27$ for the control arm. Since the actual median may not be known and will have to be estimated, we will examine the properties of our methods using a variety of decision rules.

Based upon the parameter values and switching rules stated above, the original control arm consisted of 27% of patients who were censored before progression or death, 12% of patients who were censored for death after progression (61% of whom switched treatment arm), 26% of patients who died in the absence of disease progression, and 35% of patients whose death was observed after progression (60% of whom switched treatment arm). The original treatment arm consisted of 33% of patients who were censored before progression or death, 9% of patients who were censored for death after progression (50% of whom switched treatment arms), 29% of patients who died in the absence of disease progression, and 29% of patients whose death was observed after progression (65% of whom switched treatment arms).

We use a Newton-Raphson algorithm to maximize the likelihood presented in Section 2.3, with the marginal means and variances computed from the observed data as initial parameter estimates. The data can also be used to generate a starting value for ρ_2. However, since ρ_1 and ρ_3 cannot be estimated, we use a starting value of 0.15 for each. Theoretical expressions of the first and second derivatives are used in the algorithm. The properties of our methods based on 1,000 simulations are

presented in Table 2.2. The table contains the true value for each parameter, the empirical mean of the estimates among the 1,000 simulations, the empirical MSE of the estimates among 1,000 simulations, and the coverage probability of the 95% confidence interval computed from the inverse of the information matrix and a normal approximation, i.e. estimate $\pm$ 1.96 standard deviations. The true value of Δ is simply $\mu_1 - \mu_0 = 0.2$, and true values of Δ^* and Δ^{**} are the average over 1,000 simulations assuming all information is known for each subject. Since Δ, Δ^*, and Δ^{**} are functions of parameters, the corresponding variance estimates of their estimates are calculated with the delta method.

Based on Table 2.2, it can be seen that all parameter estimates are unbiased and the coverage of the 95% confidence intervals are very close to 0.95. Table 2.3 summarizes the finite sample properties of our methods with different rules for switching treatment arms. The last six columns contain the empirical mean and MSE of each parameter with switching rules that differ from the ones used with the results in Table 2.2. Switching rule 2 (SR2) shows the performance when both of thresholds for switching ($\exp(-0.7) = 0.50$ for the treatment arm and $\exp(-1.0) = 0.37$ for the control arm) are greater than the true medians. Switching rule 3 (SR3) illustrates the result when both of the switching thresholds ($\exp(-1.4) = 0.25$ for the treatment arm and $\exp(-1.7) = 0.18$ for the control arm) are less than the true medians. In switching rule 4 (SR4), threshold for the treatment arm ($\exp(-1.3) = 0.27$) is less than the true median ($\exp(-1.1) = 0.33$) when the threshold for the control arm ($\exp(-1.1) = 0.33$) is greater than the true median ($\exp(-1.3) = 0.27$). The results presented in Table 2.3 demonstrate that estimation is relatively unaffected by the actual switching rule that is used.

Since our methods are fully parametric, we also examine the performance of our methods when some of the assumptions are violated. Table 2.4 summarizes simulation results based upon data drawn from a log-Gamma distribution rather than a

Table 2.2: Finite sample properties of estimating parameters based upon 1,000 simulations of a clinical trial with 200 patients in each arm. Patients are allowed to switch at disease progression if their progression time is less than $\exp(-1.1) = 0.33$ (treatment arm) or $\exp(-1.3) = 0.27$ (control arm); MSE = mean squared error; CP = coverage probability for 95% confidence interval.

Parameter	True Value	Mean	MSE	CP
μ_0	-1.30	-1.30	0.03	0.91
μ_1	-1.10	-1.11	0.03	0.92
μ_2	-1.20	-1.20	0.02	0.95
μ_{0E}	-0.50	-0.51	0.03	0.95
σ_0^2	2.00	1.99	0.13	0.92
σ_1^2	1.80	1.78	0.13	0.89
σ_2^2	1.30	1.32	0.07	0.93
σ_{0E}^2	2.00	2.00	0.21	0.92
σ_{1E}^2	3.49	3.44	0.47	0.95
ρ_1	0.58	0.57	0.05	0.89
ρ_2	0.50	0.50	0.02	0.92
ρ_3	0.10	0.11	0.03	0.93
p_{10}	0.60	0.60	0.00	0.94
p_{11}	0.50	0.50	0.00	0.95
Δ	0.20	0.20	0.02	0.96
Δ^*	0.69	0.68	0.11	0.95
Δ^{**}	0.19	0.18	0.02	0.95

Table 2.3: Comparison of finite sample properties of estimating parameters based upon 1,000 simulations of a clinical trial with 200 patients in each arm for a variety of stopping rules. SR2 = Patients are allowed to switch at disease progression if their progression time is less than $\exp(-0.7) = 0.50$ (treatment arm) or $\exp(-1.0) = 0.37$ (control arm); SR3 = Patients are allowed to switch at disease progression if their progression time is less than $\exp(-1.4) = 0.25$ (treatment arm) or $\exp(-1.7) = 0.18$ (control arm); SR4 = Patients are allowed to switch at disease progression if their progression time is less than $\exp(-1.3) = 0.27$ (treatment arm) or $\exp(-1.1) = 0.33$ (control arm); MSE = mean squared error; CP = coverage probability for 95% confidence interval.

	True	SR2		SR3		SR4	
Parameter	Value	Mean	MSE	Mean	MSE	Mean	MSE
μ_0	-1.30	-1.30	0.03	-1.30	0.02	-1.31	0.03
μ_1	-1.10	-1.10	0.03	-1.11	0.03	-1.11	0.03
μ_2	-1.20	-1.20	0.02	-1.20	0.02	-1.20	0.03
μ_{0E}	-0.50	-0.51	0.03	-0.51	0.03	-0.51	0.03
σ_0^2	2.00	2.00	0.13	1.99	0.12	1.99	0.13
σ_1^2	1.80	1.78	0.13	1.78	0.13	1.78	0.13
σ_2^2	1.30	1.31	0.06	1.32	0.07	1.32	0.08
σ_{0E}^2	2.00	1.99	0.21	2.00	0.21	1.99	0.21
σ_{1E}^2	3.49	3.44	0.47	3.44	0.48	3.42	0.47
ρ_1	0.58	0.57	0.04	0.57	0.05	0.57	0.05
ρ_2	0.50	0.50	0.02	0.50	0.02	0.49	0.03
ρ_3	0.10	0.11	0.02	0.11	0.03	0.11	0.03
p_{10}	0.60	0.60	0.00	0.60	0.00	0.60	0.00
p_{11}	0.50	0.50	0.00	0.50	0.00	0.50	0.00
Δ	0.20	0.20	0.02	0.20	0.02	0.20	0.02
Δ^*	0.69	0.68	0.11	0.68	0.12	0.68	0.12
Δ^{**}	0.19	0.19	0.02	0.18	0.02	0.18	0.02

lognormal distribution. The following approach is used to generate draws from a multivariate Gamma distribution. First, a vector of $\{Q_{i1}, Q_{i2}, Q_{i3}, Q_{i4}\}$ is generated for patient i from a mean-zero multivariate normal distribution that has variance matrix

$$\begin{pmatrix} 1 & \rho_1 & \rho_2 & \rho_4 \\ \rho_1 & 1 & \rho_3 & \rho_5 \\ \rho_2 & \rho_3 & 1 & \rho_6 \\ \rho_4 & \rho_5 & \rho_6 & 1 \end{pmatrix} \tag{2.4}$$

Then T_{i0} is calculated with $\{T_{i0} = \mathcal{B}(\Phi(Q_{i1}); \kappa_1, \theta_1)$, where $\Phi(x)$ is the CDF of the standard normal distribution and $\mathcal{B}(w; \kappa_1, \theta_1)$ is the inverse CDF of a Gamma distribution with shape parameter κ_1 and scale parameter θ_1. κ_1 and θ_1 are chosen so that logarithm of a variable with this Gamma distribution has a mean of μ_0 and a variance of σ_0^2. Similarly, T_{i1}, G_{i0}, and TG_{i0} are calculated with $\mathcal{B}(\Phi(Q_{i2}); \kappa_2, \theta_2)$, $\mathcal{B}(\Phi(Q_{i3}); \kappa_3, \theta_3)$, and $\mathcal{B}(\Phi(Q_{i4}); \kappa_4, \theta_4)$. $\{\kappa_2, \theta_2\}$, $\{\kappa_3, \theta_3\}$, and $\{\kappa_4, \theta_4\}$ are chosen based on the values of $\{\mu_1, \sigma_1^2\}$, $\{\mu_2, \sigma_2^2\}$, and $\{\mu_{0E}, \sigma_{0E}^2\}$ accordingly. Specifically, in the 1,000 simulations summarized in Table 2.4, we have $\{\kappa_1, \theta_1\} = \{0.88, 0.60\}$, $\{\kappa_2, \theta_2\} = \{0.94, 0.66\}$, $\{\kappa_3, \theta_3\} = \{1.18, 0.41\}$, and $\{\kappa_4, \theta_4\} = \{0.88, 1.33\}$. We note that although this approach does provide a vector of correlated Gamma random variables, the actual correlation matrix for the resulting vector is not exactly the same as the matrix in (2.4).

It can be seen that although the parameter estimates are biased, the treatment effects Δ, Δ^* and Δ^{**} estimates are close to their true values. Note that in practice, the proposed parametric model can be easily extended to other distributions such as Gamma if the observed distribution is not likely to be normal distributed. We relegate comparing our methods to existing methods in the next section after we first develop methods to reduce the inherent bias in the IPE algorithm of Branson and Whitehead with multivariate normal data.

Table 2.4: Comparison of finite sample properties of estimating parameters based upon 1,000 simulations of a clinical trial with 200 patients in each arm when distributions of event times are gamma rather than normal; Patients are allowed to switch at disease progression if their progression time is less than $\exp(-1.1) = 0.33$ (treatment arm) or $\exp(-1.3) = 0.27$ (control arm); MSE = mean squared error; CP = coverage probability for 95% confidence interval.

Parameter	True Value	Mean	MSE	CP
μ_0	-1.30	-1.14	0.04	0.81
μ_1	-1.10	-0.92	0.05	0.81
μ_2	-1.20	-1.22	0.02	0.95
μ_{0E}	-0.50	-0.53	0.03	0.98
σ_0^2	2.00	2.40	0.43	0.77
σ_1^2	1.80	2.37	0.63	0.71
σ_2^2	1.30	1.55	0.25	0.87
σ_{0E}^2	2.00	2.36	0.73	0.83
σ_{1E}^2	3.49	4.03	1.55	0.86
ρ_1	0.58	0.36	0.09	0.74
ρ_2	0.50	0.65	0.04	0.51
ρ_3	0.10	0.08	0.02	0.92
p_{10}	0.60	0.62	0.00	0.91
p_{11}	0.50	0.53	0.00	0.89
Δ	0.20	0.21	0.02	0.98
Δ^*	0.74	0.80	0.23	0.94
Δ^{**}	0.20	0.20	0.02	0.97

2.5 Modifying IPE to Accomodate Multivariate Normality

In the IPE algorithm of Branson and Whitehead (2002), if a patient in the control arm switches to the treatment arm when their disease progresses, their latent time from randomization to death that would have been observed had the patient not switched is calculated as $T_{i0} + \exp(-\hat{\eta})G_{i1}$. In this formulation, $\exp(\eta)$ is assumed to be both (i) the ratio $E(G_{i1})/E(G_{i0})$, and (ii) the treatment effect $\exp(\Delta^*)$ defined in Section 2.2. However, quantities (i) and (ii) are very different under the multivariate normal model we have used to generate our data. Another reason for the difference between (i) and (ii) is that the decision of whether a patient switches their treatment arm depends on the disease progression time T_{i0}, so that even when quantities (i) and (ii) are the same, the calculated latent time $T_{i0} + \exp(-\hat{\eta})G_{i1}$ is still biased.

Recall that we assume $\log(T_{i0})$, $\log(T_{i1})$, and $\log(G_{i0})$ follow a multivariate normal distribution defined in (2.1), with the distribution of $\log(G_{i1})$ defined by $\log(G_{i1}) = \log(G_{i0}) + \log(T_{i1}) - \log(T_{i0})$. A patient in the control arm will switch to the treatment group after disease progression if their time to disease progression for this patient is earlier than a certain value. This patient's latent time to death (the one that would have been observed in the absence of switching) would be $T_{i0} + G_{i0}$, which based upon the definition of treatment effect in the IPE algorithm should equal $T_{i0} + \exp(-\eta)G_{i1}$. Now, conditional upon $T_{i0} = t_{i0}$, we have:

$$\begin{aligned} E(T_{i0} + G_{i0} \mid T_{i0} = t_{i0}) &= E(T_{i0} + \exp(-\eta)G_{i1} \mid T_{i0} = t_{i0}) \\ &= E(T_{i0} + [G_{i1} * E(G_{i0})/E(G_{i1})] \mid T_{i0} = t_{i0}) \end{aligned}$$

so that

$$E(G_{i0} \mid T_{i0} = t_{i0})E(G_{i1}) = E(G_{i1} \mid T_{i0} = t_{i0})E(G_{i0})$$

resulting in the requirement that

$$\frac{E(G_{i1} \mid T_{i0} = t_{i0})}{E(G_{i0} \mid T_{i0} = t_{i0})} \frac{E(G_{i0})}{E(G_{i1})} = 1 \tag{2.5}$$

Using earlier notation, we know

$$\begin{aligned} \frac{E(G_{i1} \mid T_{i0} = t_{i0})}{E(G_{i0} \mid T_{i0} = t_{i0})} &= \frac{exp(\tilde{\mu}_{i3} + \tilde{\sigma}_3^2/2)}{exp(\tilde{\mu}_{i2} + \tilde{\sigma}_2^2/2)} \\ &= exp([\tilde{\mu}_{i3} - \tilde{\mu}_{i2}] + [\tilde{\sigma}_3^2 - \tilde{\sigma}_2^2]/2) \\ &= exp\{\mu_1 - \log t_{i0} + \rho_1\sigma_1/\sigma_0(\log t_{i0} - \mu_0) \\ &\quad +\sigma_1^2(1-\rho_1^2)/2 + \sigma_1\sigma_2(\rho_3 - \rho_1\rho_2)\} \end{aligned}$$

and

$$\begin{aligned} \frac{E(G_{i0})}{E(G_{i1})} &= \frac{exp(\mu_2 + \sigma_2^2/2)}{exp\{\mu_1 + \mu_2 - \mu_0 + \sigma_0^2/2 + \sigma_1^2/2 + \sigma_2^2/2 - \rho_1\sigma_0\sigma_1 - \rho_2\sigma_0\sigma_2 + \rho_3\sigma_1\sigma_2\}} \\ &= exp\{-\mu_1 + \mu_0 - \sigma_0^2/2 - \sigma_1^2/2 + \rho_1\sigma_0\sigma_1 + \rho_2\sigma_0\sigma_2 - \rho_3\sigma_1\sigma_2\} \end{aligned}$$

Thus, we can express the ratio in Equation (2.5) as

$$\begin{aligned} \log\left\{\frac{E(G_{i1} \mid T_{i0} = t_{i0})}{E(G_{i0} \mid T_{i0} = t_{i0})} \frac{E(G_{i0})}{E(G_{i1})}\right\} &= (\rho_1\sigma_1/\sigma_0 - 1)(\log t_{i0} - \mu_0) - \rho_1^2\sigma_1^2/2 \\ &- \sigma_0^2/2 + \rho_1\sigma_0\sigma_1 + \rho_2\sigma_0\sigma_2 - \rho_1\rho_2\sigma_1\sigma_2. \end{aligned} \tag{2.6}$$

In order to better examine the magnitude to which the equality in Equation (2.5) is violated, we make two further assumptions to simplify Equation (2.6). The first assumption is that variance of $\log(G_{i0}) = \sigma_2^2/2$, is equal to variance of $\log(G_{i1}) = \sigma_0^2/2 + \sigma_1^2/2 + \sigma_2^2/2 - \rho_1\sigma_0\sigma_1 - \rho_2\sigma_0\sigma_2 + \rho_3\sigma_1\sigma_2$. Such an assumption is biologically plausible and is often used by many models, such as the accelerated failure time (AFT) model. Under this assumption, the logarithm of the ratio becomes $(\rho_1\sigma_1/\sigma_0 - 1)(\log t_{i0} - \mu_0) + (1-\rho_1^2)\sigma_1^2/2 + (\rho_3 - \rho_1\rho_2)\sigma_1\sigma_2$. Our second assumption is that T_{i1}

and G_{i0} are independent given $T_{i0} = t_{i0}$ so that we have $\rho_3 = \rho_1\rho_2$, which further simplifies Equation (2.6) to be $(\rho_1\sigma_1/\sigma_0 - 1)(\log t_{i0} - \mu_0) + (1-\rho_1^2)\sigma_1^2/2$. As mentioned in Section 2.2, it is reasonable to assume that $\rho_1\sigma_1/\sigma_0 - 1$ is negative, so that Equation (2.6) is strictly greater than zero when $\log t_{i0} < \mu_0$. Therefore, the latent event time will be biased for patients in the control arm who switch to the treatment arm after a relatively early recurrence of disease. As a result, application of the IPE algorithm of Branson and Whitehead (2002) leads to potentially biased treatment effect estimates.

To remedy this bias in the IPE algorithm, the latent event times for control arm patients who switch to the treatment arm should not be computed as $T_{i0} + G_{i1}/\Delta^*$, in which Δ^* is constant for all values of T_{i0}, but should instead be $T_{i0} + G_{i1}/[\Delta^*\tau(t_{i0})]$, in which $\tau(t_{i0}) = \exp[(\rho_1\sigma_1/\sigma_0 - 1)(\log t_{i0} - \mu_0)]$ and varies by when a subject had disease progression. Notice that $\tau(t_{i0})$ does not contain all quantities in Equation (2.6), but only the parameters related to the mean change of $\log T_{i0}$ after treatment arm switching. The remaining parameters in Equation (2.6) may vary significantly for different assumed distributions for T_{i0}, T_{i1}, G_{i0}, and G_{i1}; thus we ignore these parameters in order to create an approach fairly robust to distributional assumptions.

Nonetheless, computation of $\tau(t_{i0})$ is less straightforward than computation of Δ^*. Thus, we make a further simplifying assumption that the variance of $\log(T_{i0})$ is equal to variance of $\log(T_{i1})$, i.e. $\sigma_1^2 = \sigma_0^2$, so that computation of $\tau(t_{i0})$ only requires estimates of μ_0 and ρ_1. Direct estimation of μ_0 (denoted as $\hat{\mu}_0$) is possible because T_{i0} is fully observed, but ρ_1 cannot be directly estimated as it expresses the within-patient association of T_{i0} and T_{i1}, one of which will be latent for each patient. Instead, one must assign a suitable value ρ_1^* to ρ_1 and perform a sensitivity analysis across a range of values for ρ_1 regarding the overall performance of the IPE algorithm. We define our approach of replacing Δ^* with $\Delta^*\hat{\tau}(t_{i0})$ as IPE Adjustment 1, where $\hat{\tau}(t_{i0}) = \exp[(\rho_1^* - 1)(\log t_{i0} - \hat{\mu}_0)]$.

Because G_{i1} and G_{i0} are only observed after treatment switching, $E(G_{i1})/E(G_{i0})$

cannot be correctly estimated from the data, making it impossible to confirm if $E(G_{i1})/E(G_{i0})$ is equal to Δ^*, which is one of the assumptions of the IPE algorithm. However, because both T_{i1} and T_{i0} are observed prior to treatment switching, we can estimate $E(T_{i1})/E(T_{i0})$, which is similar to the inestimable quantity $E(G_{i1})/E(G_{i0})$. If we take Equation (2.6) and substitute $E(G_{i1})/E(G_{i0})$ with $E(T_{i1})/E(T_{i0}) = exp\{(\mu_0 + \sigma_0^2/2) - (\mu_1 + \sigma_1^2/2)\}$, we have

$$\begin{aligned} \log\left\{\frac{E(G_{i1} \mid T_{i0} = t_{i0})}{E(G_{i0} \mid T_{i0} = t_{i0})}\frac{E(T_{i0})}{E(T_{i1})}\right\} &= (\rho_1\sigma_1/\sigma_0 - 1)(\log t_{i0} - \mu_0) \\ &- \rho_1^2\sigma_1^2/2 + \sigma_0^2/2 + (\rho_3 - \rho_1\rho_2)\sigma_1\sigma_2. \quad (2.7) \end{aligned}$$

The minor difference between Equations (2.6) and (2.7) is explained by variance parameters that were omitted in IPE Adjustment 1. Thus, we can take IPE Adjustment 1, $\Delta^*\hat{\tau}(t_{i0})$, and replace Δ^* with $E(T_{i1})/E(T_{i0})$, giving us IPE Adjustment 2, $[\widehat{E(T_{i1})}/\widehat{E(T_{i0})}]\hat{\tau}(t_{i0})$, which attempts to limit dependence on the unverifiable assumption that $\Delta^* = E(G_{i1})/E(G_{i0})$. Note that because Δ^* is no longer used in the estimating procedure, iteration is avoided in IPE Adjustment 2.

Now that we have developed methods for reducing the bias of the IPE algorithm, we present simulation results in Table 2.5 that compare the performance of our proposed methods with other existing methods for estimation of Δ^*. The intent-to-treat and per-protocol analyses used an AFT model assuming a log-normal distribution, as did the standard IPE, and IPE with Adjustments 1 and 2. Data were simulated similar to the procedure described in Section 2.4, except that no censoring was incorporated into the data. The true treatment effect is $\Delta^* = 0.71$ and 1,000 data sets were simulated. The effects of Adjustments 1 and 2 to the IPE algorithm were examined using three different values of $\rho_1^* = \{0.2, 0.5, 1.0\}$. Table 2.5 displays the bias and MSE across the 1,000 simulations. We acknowledge that most of the competing methods are semiparametric or nonparametric, and thus our proposed method has

the smallest MSE. Therefore another comparison is illustrated under the last column labeled as MSE*. If we suppose that the unobserved G_{i0} for control arm patients who switch and the unobserved G_{i1} for treatment arm patients who switch can actually be observed, then we consider the ratio of the sample means of overall survival time between the two arms as a gold standard nonparametric estimate of Δ^* and we call this the "true treatment effect" for each simulation. For each method, MSE* is defined to be the averaged squared difference between the estimate from the method and the "true treatment effect" and provides a measure of how well a method can recover the information lost due to switching treatment arms.

With regard to bias, we see in Table 2.5 that our proposed method is unbiased and the IPE algorithm has the least bias among all other methods examined. The intent-to-treat analysis underestimates the treatment effect as expected, and downward bias also exists for the per-protocol methods. Exclusion of patients who switch treatment leads to bias because exclusion is a function of when subjects experience disease progression, and treating treatment switching as a censoring event also creates bias because the censoring is informative. The bias occurring in a Cox proportional hazards model with treatment arm as a time-varying covariate occurs because the assumption of proportional hazards (PH) is violated. This last approach has the largest bias because the PH assumption holds for the entire sample of data, while the bias of the other approaches only related to assumption made about patients who switch treatment arms.

The bias of the treatment effect estimate from the IPE algorithm using Adjustment 1 is small and appears to be insensitive to the value of ρ_1^*; note that when $\rho_1^* = 1$, $\hat{\tau}(t_{i0})$ is actually one, and using IPE with Adjustment 1 is equivalent to the original IPE algorithm. The bias of the treatment effect estimate for the IPE algorithm with Adjustment 2 appears to be more sensitive to the value of ρ_1^* which indications of non-negligible bias with values of $\rho_1^* > 0.50$. Although bias of Adjustment 2 is slightly

Table 2.5: Finite sample properties of different methods for estimating a known treatment effect of $\Delta^* = 0.71$. There were 200 patients in each arm and 1,000 simulations were conducted. Proposed: parametric MLE assuming multivariate normality; ITT: Intent-to-treat analysis; PPES: per-protocol excluding data from those who switch treatment arms; PPCS: per-protocol treating disease progression as a censoring event for future death; CoxTV: Cox proportional hazards model with treatment arm as a time-varying covariate; IPE: iterative parameter estimation algorithm by Branson and Whitehead (2002); A1 and A2: Adjustment methods 1 and 2 for IPE as described in Section 2.5; MSE: mean-squared error; MSE*: MSE with recovery of latent event times.

Method	Assumed Value for ρ_1^*	Bias	MSE	MSE*
Proposed	·	0.00	0.05	0.064
ITT	·	-0.12	0.12	0.011
PPES	·	-0.05	0.12	0.003
PPCS	·	-0.04	0.12	0.003
CoxTV	·	-0.25	0.12	0.104
IPE	·	0.01	0.14	0.005
IPE A1	0.2	0.02	0.13	0.003
	0.5	0.02	0.13	0.003
	1.0	0.01	0.14	0.005
IPE A2	0.2	-0.03	0.11	0.001
	0.5	-0.04	0.11	0.001
	1.0	-0.08	0.11	0.005

larger than bias of Adjustment 1, we should point out that it is because there is a small difference between the value of 0.71 for Δ^* calculated from the true parameters and the average of the "true treatment effect" over the 1,000 simulation, which is 0.68. If the latter value had been assumed to be the true value of Δ^*, the bias of Adjustment 2 is much less.

We acknowledge that our method has the lowest MSE among the methods because the data are simulated from our assumed model. However, the corresponding MSE* for our method is the second largest among all approaches because of the parametric estimation relative to the other semi- and non-parametric methods. Although the true value of ρ_1 in our simulation is 0.58, we observe that the MSE and MSE* for the IPE algorithm with Adjustment 2 are very small when ρ_1^* is 0.2 or 0.5, demonstrating that much of the lost information can be recovered by the IPE algorithm when using Adjustment 2 even if the specified ρ_1^* is less than ρ_1. Generally, we suggest that Adjustment 2 with ρ_1^* between 0.0 and 0.5 should be used in practice because: (i) ρ_1 is unlikely to be very large in realistic settings, and (ii) the recommended value for ρ_1^* should be less than that for ρ_1 if D_{Ti} is not equal to but positively correlated with D_{Gi}.

We chose to simulate data without censoring because White (2006) pointed out that the recensoring approach proposed by Branson and Whitehead (2002) for the IPE algorithm was potentially biased and White (2006) provided a corrected version of recensoring to use with the IPE algorithm. Unfortunately, this modified recensoring approach cannot be applied to our approach because the censoring times are not observed for uncensored patients. A potential approach to recensoring for our approach is still under investigation.

2.6 Discussion

Our work has shown that when treatment arm switching depends on the observed progression time of patients, the IPE adjustment of Branson and Whitehead (2002) may be biased. This is because the causal effect can be altered after treatment switching selection. For the same reason, if treatment switching depends on something that is not measured in the data set (for example, an investigator's judgment and/or prognosis), unverifiable assumptions are required to identify the effect of these unmeasured factors. In this chapter, we propose a robust approach to address the issue when treatment switching depends on the observed progression time, and we assume independence between treatment switching and the future event time given observed progression time. This assumption, although weaker than the assumptions required by IPE, is still questionable in practice, so it is highly recommended to collect more variables associated with both treatment switching and the future event time to weaken the unverified assumption.

Our assumption that $D_{Ti} \equiv D_{Gi} \equiv D_{TGi}$ may appear strong but is intuitive. It provides an idea to understand the association between the causal effect of treatment with survival both before progression and after progression and leads to our proposed adjustments for treatment arm switching. The assumption can be relaxed by assuming D_{Gi} is positively correlated with D_{Ti} instead of equal to D_{Ti}. Note that although the correlation between D_{Gi} and D_{Ti} (denoted as ρ_{TG}) may be unidentifiable, the robust approaches proposed in Section 2.5 with ρ_1 replaced by $\rho_1\rho_{TG}$ could be applied, although a sensitivity analysis for the assumed value of $\rho_1\rho_{TG}$ would still be required.

CHAPTER III

Missing Not at Random Models for Masked Clinical Trials with Dropouts

3.1 Introduction

Randomized controlled trials (RCTs) are the gold standard for assessing the effectiveness of health interventions. However, missing data are common in many experimental studies and undermine the benefits of randomization. Although missing data can be reduced by preventive measures in the design and conduct of clinical trials (Little et al., 2012), some missing values are often unavoidable.

Simple methods such as complete case analysis and last observation carried forward imputation can yield severe bias or loss of efficiency (National Research Council, 2010). Alternative missing data adjustment methods often assume that the missing data are missing at random (MAR), in the sense that missingness does not depend on the missing values after conditioning on the observed data (Rubin, 1976). Approaches based on MAR include ignorable likelihood-based methods such as parametric multiple imputation (Rubin, 1987) and maximum likelihood (ML) estimation, and inverse probability-weighted methods (Robins et al., 1995). However, in many settings, such as when data on outcomes are missing because of adverse events or lack of treatment efficacy, missingness depends on the missing values, in which case the MAR assump-

tion is violated. Such mechanisms are called missing not at random (MNAR), or non-ignorable (Little and Rubin, 2002).

Joint models for the outcomes and missing indicators are usually required for non-ignorable missing data mechanisms. Little (1993) considered two classes of joint models: selection models and pattern-mixture models. Selection models factorize the joint model of outcomes and missing indicators into a model for the marginal outcomes and another model of missing indicators conditional on the possibly unobserved outcomes. Pattern-mixture models factorize the joint model into a model of missing data patterns and another model for outcomes given missing data patterns. In either approach, unverifiable restrictions or assumptions are required to identify the parameters (Fitzmaurice et al., 2005). Researchers have proposed sensitivity analyses to address this issue (Little and Rubin, 2002; Scharfstein et al., 1999; National Research Council, 2010), but in practice only a limited set of MNAR models can be assessed.

Masked (or blinded) experiments conceal the treatment assignment from the subject (and perhaps also from the clinician) to prevent the potential bias caused by knowing this information. We suggest that, since the specific treatment received is unknown, masking justifies the assumption that missingness does not depend on treatment assignment after conditioning on outcomes and side effects. That is, any treatment effect on missingness is fully mediated through outcome and side effects. Since missingness of outcomes is allowed to depend on the outcomes values, this assumption is MNAR – we call the assumption masked missing not at random (MMNAR). Even if some participants guess the treatment group to which they are assigned, their judgement on whether to stay in the study still depends primarily on outcomes and side effects, so our proposed assumption may be more plausible than MAR. Unlike MAR, MMNAR is not always sufficient to identify the parameters, as discussed for categorical and continuous data models in Section 3.2.

Our motivating example is the TRial Of Preventing HYpertension (TROPHY)

study (Julius et al., 2006), a randomized, double-blind clinical trial that examined whether early treatment of prehypertensive patients might prevent or delay the development of hypertension. After the entry criteria were met, the participants were randomized to treatment or placebo. In the first 2-year phase, subjects in the treatment group received candesartan at a dose level of 16mg daily. This was followed by the second 2-year phase, in which all study patients received placebo. For subjects in both groups, return visits were scheduled every 3 months, with two additional visits at month 1 and month 25. At each clinic visit, sitting and resting blood pressures were recorded along with adverse effects. Throughout the 4-year period, subjects and study investigators remained masked to the original treatment assignment. Our analysis was to determine whether for patients with prehypertension, 2 years of treatment (candesartan) would reduce the incidence of hypertension for up to 2 years after active treatment was discontinued. Among 772 randomized patients, 109 of them discontinued participation before reaching the trial end point. This missing data problem was originally addressed with the last observation carried forward (LOCF) method, in which the blood pressure recorded at the last clinic visit was carried forward. We apply our proposed MMNAR model to this study.

The remainder of the chapter is organized as follows. In Section 3.2, we formalize our assumption about the missing data mechanism and incorporate it in two models for categorical and continuous outcomes, along with associated estimation procedures. Section 3.2.4 describes simulation studies that compare our methods with alternatives. In Section 3.3, we analyze the TROPHY data. We conclude with a discussion in Section 3.4.

3.2 Masked Missing Not at Random (MMNAR) Models

3.2.1 The MMNAR Assumption

We consider a randomized clinical trial with two or more arms and n subjects. The full data for ith individual are given by X_i, Z_i, S_i, Y_i, where (a) X_i denotes the assigned treatment arm, (b) Z_i denotes a set of baseline covariates, (c) S_i denotes side effects recorded in the course of the study, and (d) Y_i denotes an outcome of interest at the end of the study. The variables X_i, Z_i, and S_i are assumed to be fully recorded, but some values of the outcome Y_i are missing. We define a missing data indicator M_i, which equals 1 if Y_i is missing and 0 if Y_i is observed. The data for ith individual are modeled by the joint distribution of the outcome, side effects and missing data indicator given the covariates and treatment indicators, with density $f(Y_i, S_i, M_i|Z_i, X_i, \theta)$, where θ is the set of model parameters.

We assume observations are independent. Then the missing data mechanism is missing at random (MAR, (Rubin, 1976)) if

$$\Pr(M_i = 1|Y_i, S_i, Z_i, X_i, \theta) = \Pr(M_i = 1|S_i, Z_i, X_i, \theta), \tag{3.1}$$

that is, probability of being missing can depend on observed side effects, covariates, and treatment indicators, but does not depend on the outcomes Y_i after conditioning on these variables. An equivalent assumption is

$$f(Y_i|S_i, Z_i, X_i, M_i = 1, \theta) = f(Y_i|S_i, Z_i, X_i, M_i = 0, \theta), \tag{3.2}$$

that is, the distribution of Y_i given S_i, Z_i, X_i is the same for cases with Y missing as for cases with Y observed. Under the MAR assumption of Eqs. (3.1) or (3.2), inference for the parameters θ does not require a model for the missing data mechanism (National Research Council, 2010).

Most standard models for analyzing the data make the MAR assumption, but in practice, there are often situations where missingness is likely to depend on the outcome Y, so MAR is not valid. In masked trials, it seems plausible that after conditioning on relevant observed information, namely side effects and outcomes, missingness does not depend on the treatment assigned. This motivates the following alternative to MAR:

$$\Pr(M_i = 1|X_i, Y_i, S_i, Z_i, \theta) = \Pr(M_i = 1|Y_i, S_i, Z_i, \theta), \tag{3.3}$$

that is, the probability that Y is missing is allowed to depend on the side effects, covariates and outcome, but does not depend on the assigned treatment. We call this the masked MNAR (MMNAR) assumption. Like MAR, Eq. (3.3) cannot be verified from the observed data, but we suggest that it can be justified by the nature of masked experiments. It is easy to show (see Appendix 3.5.1 for details) that Eq. (3.3) is equivalent to

$$f(X_i|Y_i, S_i, Z_i, M_i = 1) = f(X_i|Y_i, S_i, Z_i, M_i = 0) \tag{3.4}$$

so the MMNAR assumption implies that the conditional probability of being in a treatment group given outcomes, side effects, and covariates, is the same for individuals with Y_i observed and individuals with Y_i missing.

3.2.2 Some MMNAR models

We describe some MMNAR models that assume X_i and M_i are categorical variables, but consider various models for the distribution of S_i and Y_i.

Example 1. A categorical MMNAR model. Suppose there are no covariates Z_i, and X_i, Y_i, S_i, and M_i are all categorical; X_i has J categories, corresponding to J treatment arms, Y_i has K categories corresponding to K possible outcomes,

and S_i has L categories, corresponding to no side effect or $L-1$ different types of side effects. These, together with the binary missing data indicator M_i, form a 4-way $J \times K \times L \times 2$ contingency table, which is incompletely observed since Y_i is missing for cases with $M_i = 1$. We assume that conditional on the sample size n the cell counts follow a multinomial distribution with parameters $\gamma = \{\gamma_{jklm})$ where $\gamma_{jklm} = \Pr\{X_i = j, Y_i = k, S_i = l, M_i = m\}$. We can express the cell probabilities $\{\gamma_{jklm}\}$ as a saturated loglinear model

$$\begin{aligned} \log \gamma_{jklm} &= \alpha^{(0)} + \alpha_j^{(1)} + \alpha_k^{(2)} + \alpha_l^{(3)} + \alpha_m^{(4)} + \alpha_{jk}^{(12)} + \alpha_{jl}^{(13)} + \alpha_{jm}^{(14)} + \alpha_{kl}^{(23)} \\ &\quad + \alpha_{km}^{(24)} + \alpha_{lm}^{(34)} + \alpha_{jkl}^{(123)} + \alpha_{jkm}^{(124)} + \alpha_{jlm}^{(134)} + \alpha_{klm}^{(234)} + \alpha_{jklm}^{(1234)}, \end{aligned} \tag{3.5}$$

where the α terms are constrained to sum to zero over any of their subscripts. This model has $2JKL - 1$ parameters. It is not identified without restrictions, because the observed data only provides estimates for $JKL + JL - 1$ parameters – there are JKL cell counts for cases with Y observed, JL cell counts for cases with Y missing, less one for the contraint that the probabilities sum to one. However, under the MMNAR assumption, there are no associations between X_i and M_i given Y_i and S_i, so $\alpha_{jm}^{(14)} = \alpha_{jkm}^{(124)} = \alpha_{jlm}^{(134)} = \alpha_{jklm}^{(1234)} = 0$. In total, $(J-1)KL$ parameters are set to zero, so this MMNAR model has $2JKL-1-(J-1)KL = JKL+KL-1$ unknown parameters. This model is identified if and only if $JKL + KL - 1 \leq JKL + JL - 1 \iff K \leq J$, that is, the number of treatment groups is greater than or equal to the number of categories of outcomes (See Appendix 3.5.2 for more details). ML estimates can be readily computed using an EM algorithm.

This saturated categorical MMNAR model can be written as (XYS, YSM), using the notation in Agresti (2002) which lists the highest-order associations for each variable. It may require a large sample size to yield estimates with sufficient precision, so unsaturated models that further constrain the parameters are also of interest. Table

3.1 lists all hierarchical log-linear models that (a) include the YM association (not MAR models), and (b) include at least one of the XS and XY associations (the treatment assignment is associated with either side effects or outcomes). The table summarizes the degrees of freedom for these models and the conditions required for them to be identified. For example, if we assume there is no three-way association YSM, as in model 5 in Table 3.1, there are $(K-1)(L-1)$ less parameters in the model. This model is identified if $K \leq JL-L+1$, a much less stringent condition than $K \leq J$. These models can be further extended by including categorical covariates Z.

Table 3.1: All categorical models assuming MMNAR that include YM and at least one of XS and XY.

Model	Symbol	Model DF	Identified if
1	XYS, YSM	$(J+1)KL-1$	$K \leq J$
2	XY, XS, YSM	$2KL+JL+JK-J-L-K$	$K \leq J$
3	XY, YSM	$2KL+(J-1)K-1$	$K \leq J$
4	XS, YSM	$2KL+(J-1)L-1$	Never
5	XYS, SM, YM	$JKL+K+L-2$	$K \leq JL-L+1$
6	XY, XS, YS, SM, YM	$KL+JL+JK-J-1$	$K \leq JL-L+1$
7	XY, YS, SM, YM	$KL+JK+L-2$	$K \leq JL-L+1$
8	XS, YS, SM, YM	$KL+JL+K-2$	Never
9	XY, XS, SM, YM	$JL+JK-J+K+L-2$	$K \leq J$
10	XY, SM, YM	$JK+K+2L-3$	$K \leq J$
11	XS, SM, YM	$JL+2K+L-3$	Never
12	XYS, YM	$JKL+K-1$	$K \leq JL$
13	XY, XS, YS, YM	$KL+JL+JK-J-L$	$K \leq JL$
14	XY, YS, YM	$KL+JK-1$	$K \leq JL$
15	XS, YS, YM	$KL+JL+K-L-1$	$K \leq L$
16	XY, XS, YM	$JL+JK-J+K-1$	$K \leq J$
17	XY, S, YM	$JK+K+L-2$	$K \leq J$
18	XY, YM	$JK+K+L-3$	$K \leq J$
19	XS, YM	$JL+2K-2$	Never

Example 2. A normal MMNAR model. We now suppose X_i remains categorical, with two treatment arms, a control arm ($X_i = 0$) and an experimental arm ($X_i = 1$), but Y_i is a continuous variable that (possibly after transformation) can

be modeled as normal. No distributional assumption is made for S_i and Z_i.

Specifically, we assume the model of M_i given X_i, S_i, and Z_i is logit(Pr($M_i = 1|X_i = x_i, S_i = s_i, Z_i = z_i$))=$\beta_{M0} + \beta_{MX}x_i + \beta_{MS}s_i + \beta_{MZ}z_i$. The conditional distribution of Y_i given X_i, M_i, S_i, and Z_i is assumed to be

$$[Y_i|X_i = x_i, M_i = m_i, S_i = s_i, Z_i = z_i] \sim N(\beta_{Y0}+\beta_{YX}x_i+\beta_{YM}m_i+\beta_{YS}s_i+\beta_{YZ}z_i, \sigma^2)$$

where $N(\mu, \sigma^2)$ denotes the normal distribution with mean μ and variance σ^2, assumed to be the same for different x_i, m_i, s_i, and z_i. Interactions are not included for simplicity, but other more complex models can be considered and a table similar to Table 3.1 can be generated. The set of parameters β is

$$\{\beta_{M0}, \beta_{MX}, \beta_{MS}, \beta_{MZ}, \beta_{Y0}, \beta_{YX}, \beta_{YM}, \beta_{YS}, \beta_{YZ}, \sigma^2\}$$

This model can be shown with some algebra to imply that logit(Pr($M_i = 1|Y_i = y_i, X_i = x_i, S_i = s_i, Z_i = z_i, \beta$))= $\alpha_{M0} + \alpha_{MY}y_i + \alpha_{MX}x_i + \alpha_{MS}s_i + \alpha_{MZ}z_i$, where α_{M0}, α_{MY}, α_{MX}, α_{MS}, and α_{MZ} are known functions of β. The MMNAR assumption implies one restriction $\alpha_{MX} = \beta_{MX} - \beta_{YM}\beta_{YX}/\sigma^2 = 0$. Since there is only one inestimable parameter β_{YM} (which is caused by missingness of Y_i when $M_i = 1$) in the unconstrained model, the constrained model is just identified. The loglikelihood is maximized by (a) calculating the ML estimates of β_{M0}, β_{MX}, β_{MS}, β_{MZ} with all observations; (b) calculating the ML estimates of β_{Y0}, β_{YX}, β_{YS}, β_{YZ} with all observations with observed Y_i; and (c) calculating the estimated value of $\beta_{YM} = \beta_{MX}\sigma^2/\beta_{YX}$. Once the ML estimates are obtained, an multiple imputation procedure (details in Appendix 3.5.3) (Rubin, 1987) can be used to draw inference. We label this model the normal MMNAR model.

3.2.3 Efficiency of MMNAR model estimates

The intersection of the MAR and MMNAR model (when both MAR and MMNAR are true) is a restricted MAR model where missingness can depend on the side effects and the covariates, but does not depend on the outcome or the treatment. That is:

$$\Pr(M_i = 1|X_i, Y_i, S_i, Z_i, \theta) = \Pr(M_i = 1|S_i, Z_i, \theta),$$

We label this the MARSZ model. If MARSZ is the correct model, ML assuming MAR is prefered to ML assuming MMNAR, since estimates of treatment effects (say θ_1) under MAR are more or equally efficient – a sketch of the proof is given in the Appendix . Some simulation comparisons of the efficiency of these methods when MARSZ is true are given in Section 3.2.4. The MMNAR model might be fitted as a form of sensitivity analysis for deviations from MAR. If the estimates based on the MMNAR model are similar to those based on MAR, the MAR analysis might be favored on grounds of efficiency. But if the estimates are substantially different, the MMNAR estimates might be preferred because the mechanism is more plausible. More formally, a likelihood ratio test could be performed to check if the data are consistent with MARSZ given MMNAR, but the power of this test may be limited.

3.2.4 Simulation Studies

Scenario 1&2. The categorical MMNAR model. Finite sample properties of the categorical MMNAR model in Section 3.2.2 are explored via simulation. We assume two treatment arms ($J = 2$), a binary outcome ($K = 2$), and two categories of side effects ($L = 2$). X_i, Y_i, M_i, S_i for 400 subjects are drawn based on factorization $\gamma_{jklm} = \gamma_{(j)kl+}\gamma_{+kl(m)}\gamma_{+kl+}$. The specific parameter settings are shown in Table 3.2, Scenario 1, and they imply that $\Pr(X_i = 1) = 0.5$. After X_i, Y_i, M_i, and S_i are generated, values of Y_i for cases with $M_i = 1$ are deleted to make 40% of outcome

values missing.

Table 3.2: Parameters used to generate categorical and normal models

Scenario 1: the categorical model when MMNAR is true.				
$\{Y_i, S_i\}$	{1, 1}	{1, 2}	{2, 1}	{2, 2}
$\Pr(Y_i, S_i)$	0.17	0.20	0.40	0.23
$\Pr(X_i = 1 \mid Y_i, S_i)$	0.94	0.24	0.57	0.28
$\Pr(M_i = 1 \mid Y_i, S_i)$	0.30	0.31	0.54	0.33
Scenario 2: the categorical model when both MMNAR and MAR are true.				
$\{S_i, M_i\}$	{1, 0}	{1, 1}	{2, 0}	{2, 1}
$\Pr(S_i, M_i)$	0.02	0.09	0.71	0.18
$\Pr(X_i, Y_i \mid S_i = 1)$	0.08	0.26	0.52	0.14
$\Pr(X_i, Y_i \mid S_i = 2)$	0.45	0.07	0.08	0.40
Scenario 3: the normal model when MMNAR is true.				
Z_{1i}	$\Pr(Z_{1i} = 1) = 0.4$			
$Z_{2i} \mid Z_{1i}$	$[Z_{2i} \mid Z_{1i}] \sim N(1 + 0.2Z_{1i}, 1)$			
$S_i \mid X_i, Z_{1i}, Z_{2i}$	$\text{logit}(S_i = 1 \mid X_i, Z_{1i}, Z_{2i}) = 1 - 2X_i + Z_{1i} - 0.5Z_{2i}$			
$M_i \mid S_i, X_i, Z_{1i}, Z_{2i}$	$\{\beta_{M0}, \beta_{MX}, \beta_{MZ1}, \beta_{MZ2}, \beta_{MS}\}$=$\{-0.5, -1, 2, -0.5, 1\}$			
$Y_i \mid M_i, S_i, X_i, Z_{1i}, Z_{2i}$	$\{\beta_{Y0}, \beta_{YX}, \beta_{YZ1}, \beta_{YZ2}, \beta_{YS}, \beta_{YM}, \sigma^2\}$=$\{-1, 1, 2, 0.5, -1, -1, 1\}$			
Scenario 4: the normal model when both MMNAR and MAR are true.				
Z_{1i}	$\Pr(Z_{1i} = 1) = 0.4$			
$Z_{2i} \mid Z_{1i}$	$[Z_{2i} \mid Z_{1i}] \sim N(1 + 0.2Z_{1i}, 1)$			
$S_i \mid X_i, Z_{1i}, Z_{2i}$	$\text{logit}(S_i = 1 \mid X_i, Z_{1i}, Z_{2i}) = 1 - 2X_i + Z_{1i} - 0.5Z_{2i}$			
$M_i \mid S_i, X_i, Z_{1i}, Z_{2i}$	$\{\beta_{M0}, \beta_{MX}, \beta_{MZ1}, \beta_{MZ2}, \beta_{MS}\}$=$\{-0.5, 0, 2, -0.5, 1\}$			
$Y_i \mid M_i, S_i, X_i, Z_{1i}, Z_{2i}$	$\{\beta_{Y0}, \beta_{YX}, \beta_{YZ1}, \beta_{YZ2}, \beta_{YS}, \beta_{YM}, \sigma^2\}$=$\{-1, 1, 2, 0.5, -1, 0, 1\}$			

Two other methods, complete case analysis and ML assuming MAR, are also applied to the generated samples for comparison purposes. The "before deletion" data analysis, which assumes that we have access to all Y_i, is also included as a benchmark for comparison. Differences of Y_i between groups of X_i, $\Pr(Y_i = 1 \mid X_i = 2)$-$\Pr(Y_i = 1 \mid X_i = 1)$, along with $\Pr(Y_i = 1 \mid X_i = 1)$ are of interest. We compare estimates for each method and their estimated variances. Variance estimates for the ML methods are computed from the inverse of the information matrix by the delta method.

Table 3.3 presents the true parameter values, and for each method the empirical

mean and empirical MSE of the estimates, and the coverage proportion (CP) of the 95% confidence interval (CI), based on 1,000 simulated data sets. The 95% confidence interval is calculated with a normal approximation, i.e. point estimate $\pm$ 1.96 standard deviations.

Table 3.3: Finite sample properties of four methods based upon 1,000 simulations of a clinical trial with 400 patients. All of X_i, Y_i, and S_i are categorical. In Scenario 1, MMNAR is true (M_i and X_i are independent given Y_i and S_i). In Scenario 2, both MMNAR and MAR are true (M_i are independent with X_i and Y_i given S_i). MSE = $10^4 \times$ mean squared error. CP = $10^2 \times$ coverage proportion for 95% confidence interval.

Method	Parameter	True	Mean	MSE	CP
Scenario 1: the categorical model when MMNAR is true.					
Before deletion	$\Pr(Y_i = 1 \mid X_i = 1)$	0.42	0.42	12	95.3
data analysis	$\Pr(Y_i = 1 \mid X_i = 2)$-$\Pr(Y_i = 1 \mid X_i = 1)$	-0.09	-0.09	23	95.0
Complete case	$\Pr(Y_i = 1 \mid X_i = 1)$	0.42	0.49	83	60.0
analysis	$\Pr(Y_i = 1 \mid X_i = 2)$-$\Pr(Y_i = 1 \mid X_i = 1)$	-0.09	-0.12	49	92.4
ML assuming	$\Pr(Y_i = 1 \mid X_i = 1)$	0.42	0.50	88	57.8
MAR	$\Pr(Y_i = 1 \mid X_i = 2)$-$\Pr(Y_i = 1 \mid X_i = 1)$	-0.09	-0.16	80	82.0
ML assuming	$\Pr(Y_i = 1 \mid X_i = 1)$	0.42	0.42	72	98.0
MMNAR	$\Pr(Y_i = 1 \mid X_i = 2)$-$\Pr(Y_i = 1 \mid X_i = 1)$	-0.09	-0.09	101	97.6
Scenario 2: the categorical model when both MMNAR and MAR are true.					
Before deletion	$\Pr(Y_i = 1 \mid X_i = 1)$	0.82	0.82	7	96.1
data analysis	$\Pr(Y_i = 1 \mid X_i = 2)$-$\Pr(Y_i = 1 \mid X_i = 1)$	-0.56	-0.56	17	93.9
Complete case	$\Pr(Y_i = 1 \mid X_i = 1)$	0.82	0.85	20	74.1
analysis	$\Pr(Y_i = 1 \mid X_i = 2)$-$\Pr(Y_i = 1 \mid X_i = 1)$	-0.56	-0.66	118	37.8
ML assuming	$\Pr(Y_i = 1 \mid X_i = 1)$	0.82	0.82	12	97.9
MAR	$\Pr(Y_i = 1 \mid X_i = 2)$-$\Pr(Y_i = 1 \mid X_i = 1)$	-0.56	-0.56	29	98.5
ML assuming	$\Pr(Y_i = 1 \mid X_i = 1)$	0.82	0.82	14	96.1
MMNAR	$\Pr(Y_i = 1 \mid X_i = 2)$-$\Pr(Y_i = 1 \mid X_i = 1)$	-0.56	-0.57	33	96.4

In Table 3.3, treatment effects from both complete case analysis and ML assuming MAR are biased, showing the potential for bias when the missing data mechanism is MNAR. Our proposed method is essentially unbiased with small MSE and coverage proportion very close to 95%, confirming as expected that our proposed model outperforms the MAR model when MMNAR is true.

It is of interest to compare ML under MMNAR and ML under MAR when M_i does not depend on X_i or Y_i given S_i, so both methods are valid. X_i, Y_i, M_i, S_i are drawn based on factorization $\Pr(X_i, Y_i, S_i, M_i) = \Pr(S_i, M_i)\Pr(X_i, Y_i|S_i)$. More specifical parameter setting can be found in Table 3.2, Scenario 2, which leads to $\Pr(X_i = 1) = 0.5$. After all X_i, Y_i, M_i, and S_i are generated, values of Y_i for cases with $M_i = 1$ are deleted to make 27% of outcome values missing. Results for 1,000 data sets simulated in this way are shown in Table 3.3.

The estimates from complete case analysis are biased. Both ML under MMNAR and ML assuming MAR have small empirical bias, and confidence intervals close to nominal 95% coverage. This is expected since model assumptions in both models are satisfied. We noted above that ML assuming MAR is at least as efficient as ML assuming MMNAR. In this simulation, ML under MMNAR is very close to ML under MAR in terms of efficiency.

Scenario 3&4. The normal MMNAR model. Simulation studies are also conducted to examine the performance of the normal MMNAR model. We consider a design with an experimental group and a control group with 200 patients in each group. We generate one binary covariate Z_{1i}, one normal covariate Z_{2i}, one binary side effect S_i, the missing indicator M_i, and the normal outcome Y_i successively with the parameters specified in Table 3.2, Scenario 3. The rate of missing data for this scenario is 45%.

Before deletion data analysis, complete case analysis and the model assuming MAR are compared with our proposed model in estimating $E(Y_i|X_i = 0)$ and $E(Y_i|X_i = 1) - E(Y_i|X_i = 0)$. In complete case analysis, a general linear regression of Y_i on X_i is fitted for the subsample with $M_i = 0$. In ML estimates assuming MAR and MMNAR, the multiple imputation described in Section 3.2.2 is used to estimate $E(Y_i|X_i = 0)$ and $E(Y_i|X_i = 1) - E(Y_i|X_i = 0)$. Table 3.4 presents the true value, the empirical mean of the estimates, the empirical MSE of the estimates, and the coverage

proportion of the 95% confidence intervals (CI) based on 1,000 simulations.

Table 3.4: Finite sample properties of four methods based upon 1,000 simulations of a clinical trial with 400 patients. X_i, Z_{1i}, M_i and S_i are categorical. Y_i, and Z_{2i} are continuous. In Scenario 3, MMNAR is true (M_i and X_i are independent given Y_i, Z_{1i}, Z_{2i}, and S_i). In Scenario 4, both MMNAR and MAR are true (M_i are independent with X_i and Y_i given Z_{1i}, Z_{2i}, and S_i). MSE = 10^3 $\times$ mean squared error. CP = 10^2 $\times$ coverage proportion for 95% confidence interval.

Method	Parameter	True	Mean	MSE	CP
Scenario 3: the normal model when MMNAR is true.					
Before deletion	$E(Y_i\|X_i=0)$	-0.93	-0.92	13	95.1
data analysis	$E(Y_i\|X_i=1)-E(Y_i\|X_i=0)$	1.69	1.67	24	95.4
Complete case	$E(Y_i\|X_i=0)$	-0.93	-0.53	183	32.8
analysis	$E(Y_i\|X_i=1)-E(Y_i\|X_i=0)$	1.69	1.47	91	80.1
ML assuming	$E(Y_i\|X_i=0)$	-0.93	-0.34	361	2.2
MAR	$E(Y_i\|X_i=1)-E(Y_i\|X_i=0)$	1.69	1.42	107	68.2
ML assuming	$E(Y_i\|X_i=0)$	-0.93	-0.93	56	93.5
MMNAR	$E(Y_i\|X_i=1)-E(Y_i\|X_i=0)$	1.69	1.69	38	95.3
Scenario 4: the normal model when both MMNAR and MAR are true.					
Before deletion	$E(Y_i\|X_i=0)$	-0.35	-0.34	13	94.0
data analysis	$E(Y_i\|X_i=1)-E(Y_i\|X_i=0)$	1.43	1.43	26	93.9
Complete case	$E(Y_i\|X_i=0)$	-0.35	-0.53	61	77.4
analysis	$E(Y_i\|X_i=1)-E(Y_i\|X_i=0)$	1.43	1.41	48	94.7
ML assuming	$E(Y_i\|X_i=0)$	-0.35	-0.34	24	94.0
MAR	$E(Y_i\|X_i=1)-E(Y_i\|X_i=0)$	1.43	1.43	37	95.0
ML assuming	$E(Y_i\|X_i=0)$	-0.35	-0.35	48	94.9
MMNAR	$E(Y_i\|X_i=1)-E(Y_i\|X_i=0)$	1.43	1.44	37	95.3

Both complete case analysis and ML assuming MAR are biased, as expected. Our proposed method has small empirical bias with variance estimates that yield confidence intervals with close to nominal coverage. The comparisons between ML assuming MAR and the proposed method, when both assumptions are satisfied, are also presented in Table 3.4.

In this setting, M_i are independent with X_i and Y_i given Z_{1i}, Z_{2i}, and S_i. The parameters for generating the samples are specified in Table 3.2, Scenario 4. The rate of missing data for this scenario is 45%.

As seen in Table 3.4, complete case analysis is biased, while the other two methods have small empirical bias and have confidence intervals close to nominal 95% coverage. As for the categorical MMNAR model simulation, the MSE of the treatment effect estimates from the two methods are very similar. Here the loss of efficiency from the MMNAR model is very small, but further investigation is needed to establish this finding more generally.

3.3 Application

We illustrate the categorical MMNAR model described in Section 3.2.2 with the TROPHY study (Julius et al., 2006). The population consisted of 772 patients enrolled at 71 centers, randomly assigned to one of the two groups: placebo (381) and candesartan (391). Let x_i be the binary treatment indicator with $x_i = 0$ denoting the placebo arm and $x_i = 1$ denoting the candesartan arm. Let y_i be the indicator of hypertension being observed in the 4-year period for ith subject. Development of hypertension was defined as the first appearance of one of the following outcomes: systolic pressure of 140 mm Hg or higher or diastolic pressure of 90 mm Hg or higher, for any three visits; systolic pressure of 160 mm Hg or higher or diastolic pressure of 100 mm Hg or higher for any visit; initiation of pharmacologic treatment; or systolic pressure of 140 mm Hg or higher or diastolic pressure of 90 mm Hg or higher at the visit at month 48. Among those 772 patients, 109 (55 in candesartan group) dropped out before development of hypertension. Let m_i be the missing data indicator for y_i with $m_i = 1$ denoting subject i dropping out before y_i was observed. If patient i dropped out after development of hypertension, the value of m_i was 0 because $y_i = 1$ was already observed. Therefore m_i was likely to be related to y_i and the independence between y_i and m_i given other variables was questionable.

Two types of adverse effects were recorded. The first type contained severe adverse effects (SAE), such as hospitalization, disability/incapacity, and some others. 88 out

of the 772 subjects had at least one of SAE. The second type included other adverse effects (OAE), such as exacerbation of back pain and sinusitis. For each subject, OAE were recorded for most visits along with their severity (mild, moderate, or severe) and whether OAE were related to treatment (probably, possible, or unlikely). We summarize the adverse effects into two main categories, creating a binary adverse effect variable S_i for subject i. If subject i did not have SAE, and the recorded OAE were unlikely to be related to treatment and were also not severe, s_i is defined to be 0 (506 subjects), otherwise s_i is 1 (266 subjects).

As discussed in Section 3.2.2, other covariates related to outcomes or missing indicators could be included in the model to make the proposed assumption more plausible. There were many baseline covariates recorded in this study, so some dimension reduction is needed. We perform two logistic regressions of m_i and observed y_i on those baseline covariates for the control group patients and determine the associations based on Nagelkerke's pseudo R^2. The baseline covariates are more strongly associated with y_i (Nagelkerke's pseudo $R^2 = 0.343$) than with m_i (Nagelkerke's pseudo $R^2 = 0.135$). Therefore we select the variables significantly related to y_i (sex, race, weight, BMI, and the baseline blood pressures measured at home), perform a logistic regression of y_i on them for the control group, and calculate the predicted probability of $\Pr(y_i|\text{covariates})$ (denoted as z_i^*) for each subject with estimated coefficients from this regression. We define, z_i^* is transformed to be a binary variable z_i ($z_i = 0$ if z_i^* is less than the median and $z_i = 1$ otherwise), and then combined with its binary s_i. This leads to a new categorical variable s_i^* with four categories. The missing values of baseline covariates are imputed with simple mean imputation since the missing data rate is very low (less than 1%).

The EM algorithm is applied to the incomplete data to obtain ML estimates for the categorical MMNAR model. Complete case analysis is also conducted, including only the subjects who did not drop out from the study. ML assuming MAR is also

performed to compare with our proposed model. Table 3.5 shows estimates and associated 95% confidence intervals for conditional means of y_i given different x_i for the three methods. The estimates of the treatment effect given as absolute risk reduction (ARR)=Pr($y_i = 1|x_i = 0$)-Pr($y_i = 1|x_i = 1$) are also shown in Table 3.5.

Table 3.5: Estimates of the absolute risk reduction (ARR) and the incidence of hypertension under different treatment groups in TROPHY study. LCI = lower bound of 95% confidence interval. HCI = higher bound of 95% confidence interval.

Method	Parameter	Mean	LCI	HCI
Complete	Pr($y_i = 1|x_i = 0$)	0.73	0.69	0.78
case	Pr($y_i = 1|x_i = 1$)	0.62	0.57	0.67
analysis	ARR=Pr($y_i = 1|x_i = 0$)-Pr($y_i = 1|x_i = 1$)	0.11	0.04	0.19
ML	Pr($y_i = 1|x_i = 0$)	0.73	0.68	0.78
assuming	Pr($y_i = 1|x_i = 1$)	0.61	0.56	0.67
MAR	ARR=Pr($y_i = 1|x_i = 0$)-Pr($y_i = 1|x_i = 1$)	0.11	0.05	0.19
ML	Pr($y_i = 1|x_i = 0$)	0.71	0.45	0.96
assuming	Pr($y_i = 1|x_i = 1$)	0.59	0.34	0.83
MMNAR	ARR=Pr($y_i = 1|x_i = 0$)-Pr($y_i = 1|x_i = 1$)	0.12	0.05	0.19

In Table 3.5, the incidence of hypertension for different treatment groups, Pr($y_i = 1|x_i = 0$) and Pr($y_i = 1|x_i = 1$) have relatively wide confidence intervals for ML assuming MMNAR because MMNAR allows the missingness of y_i to depend on y_i itself and this can potentially weaken the confidence of association between y_i and x_i. However, the estimated ARR (or treatment effect) and its 95% confidence interval for the categorical MMNAR model are 12% and (5%, 19%), which are very similar to ML assuming MAR and the complete case analysis. Therefore the estimated treatment effect is robust to deviations from MAR of the form implied by the MMNAR model and we can conclude that there is a significant difference between the two groups. Since Eq. (3.3) is more plausible than MAR for masked trials, no sensitivity analysis is necessary.

There are several limitations related to this application. First, the transformation

from original adverse effects variables to S_i is post-hoc, and might be improved by specifying one clearly defined side effect variable in the protocol. Second, outcomes were recorded for every visit, and our simple analysis does not exploit the longitudinal aspect of the study. The extension of MMNAR to longitudinal studies is under investigation.

3.4 Discussion

The utility of masking in randomized clinical trials is widely recognized, since it removes potential distortions from placebo effects, differential application of entry criteria, or other behaviors that might bias treatment comparisons. However, we have not seen discussions about the implications of masking for the treatment of missing data in previous research. The potential for non-MAR missing data in clinical trials is well known, but the class of MNAR models is very broad, so any reasonable way of limiting the number of models to be considered is valuable. In masked trials, it seems reasonable that missingness does not depend on the masked treatment, after conditioning on side effect and outcome data. This motivates the class of MMNAR models which we explore in this chapter. We describe MMNAR models for continuous and categorical outcomes, and apply maximum likelihood and multiple imputation based on these models to estimate treatment effects. Clearly many other models could be constructed that incorporate the MMNAR assumption, depending on setting.

We note that for the MMNAR assumption to be plausible, important side effects need to be recorded and incorporated in the analysis. Clinically unimportant side effects do not need to be included, since to bias the treatment effect a side effect needs to be associated with the outcome as well as the likelihood of dropping out.

Here we have considered the simple case of missing data in a univariate outcome variable. In future work we plan to consider the class of MMNAR models for repeated measures data with more than one missing data pattern, and to allow for missing data

in covariates.

3.5 Appendix

3.5.1 Proof that Eq. (3.4) follows from Eq. (3.3)

$$\begin{aligned} f(X_i|Y_i, S_i, Z_i, M_i = m_i) &= f(X_i, Y_i, S_i, Z_i, M_i = m_i)/f(Y_i, S_i, Z_i, M_i = m_i) \\ &= f(M_i = m_i|X_i, Y_i, S_i, Z_i)f(X_i, Y_i, S_i, Z_i) \\ &\quad /[f(M_i = m_i|Y_i, S_i, Z_i)f(Y_i, S_i, Z_i)] \\ &= f(X_i, Y_i, S_i, Z_i)/f(Y_i, S_i, Z_i) \\ &= f(X_i|Y_i, S_i, Z_i) \end{aligned}$$

3.5.2 Conditions under which the categorical MMNAR model is identified

For the categorical MMNAR model, when $K > J$, the number of model parameters is larger than the number of cell counts, so the model is not identified. We show that the model is identified under some regular conditions if $K \leq J$. Follow (3.5), cell probabilities are

$$\begin{aligned} \log \gamma_{jklm} &= \alpha^{(0)} + \alpha_j^{(1)} + \alpha_k^{(2)} + \alpha_l^{(3)} + \alpha_m^{(4)} + \alpha_{jk}^{(12)} + \alpha_{jl}^{(13)} + \alpha_{kl}^{(23)} \\ &\quad + \alpha_{km}^{(24)} + \alpha_{lm}^{(34)} + \alpha_{jkl}^{(123)} + \alpha_{klm}^{(234)}, \end{aligned}$$

where the α terms are constrained to sum to zero over any of their subscripts. Define $\beta_j^{(1)} = \alpha_j^{(1)} - \alpha_1^{(1)}$, $\beta_{jk}^{(12)} = \alpha_{jk}^{(12)} - \alpha_{1k}^{(12)} - \alpha_{j1}^{(12)} + \alpha_{11}^{(12)}$, $\beta_{jkl}^{(123)} = \alpha_{jkl}^{(123)} - \alpha_{1kl}^{(123)} - \alpha_{j1l}^{(123)} - \alpha_{jk1}^{(123)} + \alpha_{j11}^{(123)} + \alpha_{1k1}^{(123)} + \alpha_{11l}^{(123)} - \alpha_{111}^{(123)}$ for $j = 1, \ldots, J$, $k = 1, \ldots, K$, $l = 1, \ldots, L$, and define other β terms analogously. Therefore all terms with at least one subscript of 1 are restricted to 0. $\beta^{(0)}$ is a function of other β terms such that the cell probabilities sum to 1.

Define $\tau^{(0)}$, $\tau_j^{(1)}$, $\tau_k^{(2)}$, $\tau_l^{(3)}$, $\tau_{jk}^{(12)}$, $\tau_{jl}^{(13)}$, $\tau_{kl}^{(23)}$, and $\tau_{jkl}^{(123)}$ to be the same as $\beta^{(0)}$, $\beta_j^{(1)}$, $\beta_k^{(2)}$, $\beta_l^{(3)}$, $\beta_{jk}^{(12)}$, $\beta_{jl}^{(13)}$, $\beta_{kl}^{(23)}$, and $\beta_{jkl}^{(123)}$, respectively. Define $\tau_1^{(4)} = \exp(\beta^{(0)} + \beta_1^{(4)})$, $\tau_{k1}^{(24)} = \tau_1^{(4)} \exp(\beta_k^{(2)} + \beta_{k1}^{(24)})$, $\tau_{l1}^{(34)} = \tau_1^{(4)} \exp(\beta_{l1}^{(34)})$ and $\tau_{kl1}^{(234)} = \tau_{l1}^{(34)} \tau_{k1}^{(24)} / \tau_1^{(4)} \exp(\beta_{kl}^{(23)} + \beta_{kl1}^{(234)})$, for $k = 2, \ldots, K$ and $l = 2, \ldots, L$.

First of all, $\tau^{(0)}$, $\tau_j^{(1)}$, $\tau_k^{(2)}$, $\tau_l^{(3)}$, $\tau_{jk}^{(12)}$, $\tau_{jl}^{(13)}$, $\tau_{kl}^{(23)}$, and $\tau_{jkl}^{(123)}$ are identified since they are corresponding to the cells that Y_i is observed.

We illustrate how $\tau_1^{(4)}$ and $\tau_{k1}^{(24)}$, $k = 2, \ldots, K$ are identified. Cell probabilities for $M = 1$ and $S = 1$ are

$$\begin{aligned}
\log \gamma_{1111} &= \beta^{(0)} + \beta_1^{(4)} = \log \tau_1^{(4)} \\
\log \gamma_{1k11} &= \beta^{(0)} + \beta_1^{(4)} + \beta_k^{(2)} + \beta_{k1}^{(24)} = \log \tau_{k1}^{(24)} \\
\log \gamma_{j111} &= \beta^{(0)} + \beta_1^{(4)} + \beta_j^{(1)} = \log \tau_1^{(4)} + \tau_j^{(1)} \\
\log \gamma_{jk11} &= \beta^{(0)} + \beta_1^{(4)} + \beta_k^{(2)} + \beta_{k1}^{(24)} + \beta_j^{(1)} + \beta_{jk}^{(12)} = \log \tau_{k1}^{(24)} + \tau_j^{(1)} + \tau_{jk}^{(12)},
\end{aligned}$$

for $j = 2, \ldots, J$ and $k = 2, \ldots, K$. Since Y is missing for $M = 1$, the observed combined cell probabilities for $M = 1$ and $S = 1$ are

$$\begin{aligned}
\sum_{k=1}^{K} \gamma_{1k11} &= \tau_1^{(4)} + \sum_{k=2}^{K} \tau_{k1}^{(24)} \\
\sum_{k=1}^{K} \gamma_{jk11} &= \exp(\tau_j^{(1)})(\tau_1^{(4)} + \sum_{k=2}^{K} \tau_{k1}^{(24)} \exp(\tau_{jk}^{(12)})),
\end{aligned}$$

for $j = 2, \ldots, J$. Note that $\tau_j^{(1)}$ and $\tau_{jk}^{(12)}$ are already identified. There are J equations and K unknown parameters with $K \leq J$, so $\tau_1^{(4)}$, $\tau_{k1}^{(24)}$, $k = 2, \ldots, K$ are identified if

matrix (3.6) has full rank.

$$\begin{pmatrix} 1 & 1 & \dots & 1 \\ 1 & \exp(\tau_{22}^{(12)}) & \dots & \exp(\tau_{2K}^{(12)}) \\ \dots & & & \\ 1 & \exp(\tau_{J2}^{(12)}) & \dots & \exp(\tau_{JK}^{(12)}) \end{pmatrix} \tag{3.6}$$

We then illustrate how $\tau_{l1}^{(34)}$, $\tau_{kl1}^{(234)}$, $k = 2, \dots, K$ are identified for each l. Cell probabilities for $M = 1$ and $S = l$ are

$$\begin{aligned}
\log \gamma_{11l1} &= \beta^{(0)} + \beta_1^{(4)} + \beta_l^{(3)} + \beta_{l1}^{(34)} = \log \tau_{l1}^{(34)} + \tau_l^{(3)} \\
\log \gamma_{1kl1} &= \beta^{(0)} + \beta_1^{(4)} + \beta_k^{(2)} + \beta_{k1}^{(24)} + \beta_l^{(3)} + \beta_{l1}^{(34)} + \beta_{kl}^{(23)} = \log \tau_{kl1}^{(234)} + \tau_l^{(3)} \\
\log \gamma_{j1l1} &= \beta^{(0)} + \beta_1^{(4)} + \beta_l^{(3)} + \beta_{l1}^{(34)} + \beta_j^{(1)} + \beta_{jl}^{(13)} = \log \tau_{l1}^{(34)} + \tau_l^{(3)} + \tau_j^{(1)} + \tau_{jl}^{(13)} \\
\log \gamma_{jkl1} &= \beta^{(0)} + \beta_1^{(4)} + \beta_k^{(2)} + \beta_{k1}^{(24)} + \beta_l^{(3)} + \beta_{l1}^{(34)} + \beta_{kl}^{(23)} + \beta_j^{(1)} + \beta_{jl}^{(13)} + \beta_{jk}^{(12)} + \beta_{jkl}^{(123)} \\
&= \log \tau_{kl1}^{(234)} + \tau_l^{(3)} + \tau_j^{(1)} + \tau_{jl}^{(13)} + \tau_{jk}^{(12)} + \tau_{jkl}^{(123)},
\end{aligned}$$

for $j = 2, \dots, J$, $k = 2, \dots, K$, and $l = 2, \dots, L$. Since Y is missing for $M = 1$, the observed combined cell probabilities for $M = 1$ and $S = l$ are

$$\begin{aligned}
\sum_{k=1}^{K} \gamma_{1kl1} &= \exp(\tau_l^{(3)})(\tau_{l1}^{(34)} + \sum_{k=2}^{K} \tau_{kl1}^{(234)}) \\
\sum_{k=1}^{K} \gamma_{jkl1} &= \exp(\tau_l^{(3)} + \tau_j^{(1)} + \tau_{jl}^{(13)})(\tau_{l1}^{(34)} + \sum_{k=2}^{K} \tau_{kl1}^{(234)} \exp(\tau_{jk}^{(12)} + \tau_{jkl}^{(123)})),
\end{aligned}$$

for $j = 2, \dots, J$, $k = 2, \dots, K$, and $l = 2, \dots, L$. For each l in $\{2, \dots, L\}$, there are J equations and K unknown parameters, so $\tau_{l1}^{(34)}$, $\tau_{kl1}^{(234)}$, $k = 2, \dots, K$ are identified if

matrix (3.7) has full rank.

$$\begin{pmatrix} 1 & 1 & \dots & 1 \\ 1 & \exp(\tau_{22}^{(12)} + \tau_{22l}^{(123)}) & \dots & \exp(\tau_{2K}^{(12)} + \tau_{2Kl}^{(123)}) \\ \dots & & & \\ 1 & \exp(\tau_{J2}^{(12)} + \tau_{J2l}^{(123)}) & \dots & \exp(\tau_{JK}^{(12)} + \tau_{JKl}^{(123)}) \end{pmatrix} \tag{3.7}$$

After all τ parameters are identified, $\beta^{(0)}$, $\beta_j^{(1)}$, $\beta_k^{(2)}$, $\beta_l^{(3)}$, $\beta_{jk}^{(12)}$, $\beta_{jl}^{(13)}$, $\beta_{kl}^{(23)}$, and $\beta_{jkl}^{(123)}$ are directly obtained. $\beta_1^{(4)}$) and $\beta_{k1}^{(24)}$ are calculated with $\beta_1^{(4)}) = \log \tau_1^{(4)} - \beta^{(0)}$ and $\beta_{k1}^{(24)}) = \log(\tau_{k1}^{(24)}/\tau_1^{(4)}) - \beta_k^{(2)}$, for $k = 2, \dots, K$. $\beta_{l1}^{(34)}$ and $\beta_{kl1}^{(234)}$ are calculated with $\beta_{l1}^{(34)} = \log(\tau_{l1}^{(34)}/\tau_1^{(4)})$ and $\beta_{kl1}^{(234)} = \log(\tau_{kl1}^{(234)}\tau_1^{(4)}/\tau_{l1}^{(34)}/\tau_{k1}^{(24)}) - \beta_{kl}^{(23)}$, for $k = 2, \dots, K$. Note that the calculated β terms may not be in the parameter space.

3.5.3 An multiple imputation procedure for the normal MMNAR models

After the ML estimates are obtained, the variance covariance matrix of the estimates are calculated by multivariate delta method. Based on the estimates and their variance covariance matrix, 20 samples of the parameter set are drawn. Then each of the parameter set samples can be used to draw missing Y_i and provide an imputed data set. All 20 imputed data sets are analyzed separately and the 20 results are combined with the Rubin's combination rule (Rubin, 1987).

We apply the multiple imputation procedure because in this way, the results from ML assuming MMNAR are more comparable with other methods.

3.5.4 Sketch proof that if the model MARSZ holds, the ML estimate for θ_1 under MMNAR is less efficient than ML under MARSZ.

The ML estimate of θ_1 under MMNAR is less efficient than the ML estimate of θ_1 under MARSZ, since MARSZ is a submodel of MMNAR obtained by setting parameters relating missingness to Y to zero. On the other hand, the ML estimate

of θ_1 under MAR is the same as the ML of θ_1 under MARSZ, since both lead to the same likelihood ignoring the mechanism. Hence ML for θ_1 under MAR is more efficient than ML under MMNAR.

CHAPTER IV

Estimating Treatment Effect under the MMNAR Assumption for Longitudinal Data with Dropouts

4.1 Introduction

Many methods based on MAR assumption have been proposed for longitudinal data and a detailed review can be found in Molenberghs and Kenward (2007). Generally these methods include likelihood-based methods and weighted generalized estimating equations (wGEE), which directly model missingness conditional on observed data. An alternative is multiple imputation, which uses observed data to impute missing values. Many of these methods (Hedeker and Gibbons, 1997) allowed for the missingness mechanism to be MNAR, but Molenberghs and Kenward (2007) provided a method that can reproduce the result from any MNAR model by an MAR counterpart. This fact shows that there is no formal data-based difference between MAR and MNAR, so unless the MNAR model is unquestionable, a sensitivity analysis is recommended to examine whether the result is sensitive to the unverifiable assumption (National Research Council, 2010; Molenberghs and Kenward, 2007).

Many MNAR models are proposed as a form of sensitivity analysis based on selection models (Diggle and Kenward, 1994; Molenberghs et al., 2003; Mallinckrodt et al., 2013). In this chapter, however, we focus on the pattern-mixture model frame-

work, under which the distinctions between different missing patterns are stated in a transparent and meaningful way. Similar to available case missing values (ACMV) (Molenberghs et al., 1998), the counterpart of MAR in the PMM context, MNAR assumptions including complete case missing values (CCMV) (Little, 1993) which borrows information from the completers, and neighboring case missing values (NCMV) (Molenberghs et al., 2003) which borrows information from the nearest identified pattern, have been developed. Ratitch et al. (2013) proposed a delta-adjusting analysis strategy, which can explicitly control the deviation from MAR to reflect worsened outcomes for the dropped out subjects.

In this chapter, we apply the MMNAR assumption to longitudinal data models. Instead of specifying just one longitudinal model, we propose a flexible strategy that can extend a preferred complete data model to a MMNAR model, after which ML estimates are obtained and a multiple imputation procedure is used to draw inference.

The remainder of the chapter is organized as follows. In Section 4.2, we introduce the notation and assumptions used in later sections. Section 4.3 describes the procedure to extend a complete data model to incorporate the proposed MMNAR assumption and draw inference. In Section 4.4, we conduct simulation studies that compare our methods with alternatives. Section 4.5 presents the result from analyzing the TROPHY data. A summary and some future work are discussed in Section 4.6.

4.2 Notation and Assumptions

In this section, we discuss the strategy to extend MMNAR assumption to longitudinal studies. We consider a randomized clinical trial with n subjects. For the ith subject, where $i = 1, ..., n$, let X_i and Z_i be the treatment assignment and baseline covariates (Z_i might be a vector). Let Y_{ij} be the response of interest designed to be measured at visit j, where $j = 1, ..., T$. Similarly, let S_{ij} be the side effect measure at

visit j. The original intention was to have T observations of Y_{ij} for each individual, but some of them might not be observed due to dropout. Then for subject i and visit j, let M_{ij} be the missing indicator for Y_{ij}, so $M_{ij} = 0$, if Y_{ij} is observed, and $M_{ij} = 1$, if not. We assume that X_i, Z_i, M_{ij}, and S_{ij} are fully observed.

The missing pattern is assumed to be monotonic. In other words, some subjects discontinued from the study and have a number of visits with missing Y_{ij} after their discontinuation. Let M_i denote the last visit with Y_{ij} observed, so $M_i = m$ is equivalent to $M_{ij} = 0$ for $j \leq m$ and $M_{ij} = 1$ for $j > m$, $m = 1, ..., T$. If the data is complete for subject i, $M_i = T$. We define if Y_{ij} is missing from $j = 1$, $M_i = 0$.

The goal is to provide a strategy to apply MMNAR assumption to parametric longitudinal models. If the missing mechanism is MNAR, the missing indicators are required to be included in a joint model with other variables and we make key two assumptions to identify the parameters in this model. The first assumption is the proposed MMNAR assumption but in a longitudinal setting: if a subject has not dropped out at visit $j - 1$, the probability that this subject drops out at visit j is independent with the treatment assignment given all the information until visit j including the potentially missing outcome at visit j. MMNAR assumption can provide restrictions to identify the distribution of the missing data at the time of dropout, but the missing data for future times require the second assumption, which assumes that dropout may depend on the current, possibly unobserved, measurement, but not on future measurements.

4.3 Fitting Procedure

The fitting procedure includes two steps: regression step and imputation step. In the regression step, a parametric model is assumed and ML estimates which maximize the likelihood with MMNAR restrictions are calculated with their variance covariance matrix. The imputation step is a multiple imputation procedure which (a) draws D

sets of parameters based on the ML estimates and their variance covariance matrix obtained in the regression step; (b) draws one set of missing Y_{ij} at the time of dropout ($j = M_i + 1$) for each drawn parameter set; (c) sequentially draws one set of missing Y_{ij} for visits after dropout ($j > M_i + 1$) for each drawn parameter set; (d) performs statistical analyses on each of the imputed data sets and combines the results with Rubin's combination rule. The details are provided as follows.

4.3.1 Step 1. Regression step.

Suppose for the complete data, the vector $Y_{i1}, ..., Y_{iT}$ given X_i, Z_i, and S_{ik} ($k = 1, ..., T$) are

$$\begin{aligned}&[Y_{i1}, ..., Y_{iT}|X_i = x_i, Z_i = z_i, S_{ik} = s_{ik}, k = 1, ..., T, \beta] \\ &\sim N_T\left(\begin{pmatrix} f(x_i, z_i, s_{ik}, k = 1, \beta, j = 1) \\ ... \\ f(x_i, z_i, s_{ik}, k = 1, ..., T, \beta, j = T) \end{pmatrix}, \Sigma_T\right)\end{aligned} \tag{4.1}$$

where $N_T(\mu, \Sigma)$ denotes the T-dimensional multivariate normal distribution with mean vector μ and variance-covariance matrix Σ. and Σ_j is

$$\begin{pmatrix} \sigma_{11}^2 & \sigma_{12}^2 & ... & \sigma_{1j}^2 \\ ... & & & \\ \sigma_{j1}^2 & \sigma_{j2}^2 & ... & \sigma_{jj}^2 \end{pmatrix}, \text{ for} j = 1, ..., T$$

We further assume that there is no interaction between x_i and other variables to keep the form of MMNAR assumption simple. More specifically, $f(x_i, z_i, s_{ik}, k = 1, ..., j, \beta, j) = \beta_{YX_j} x_i + f^*(z_i, s_{ik}, k = 1, ..., j, \beta, j)$.

We introduce the missing indicator within the pattern-mixture model (PMM) framework, under which, subjects are stratified according to their missingness patterns. We assume a logistic form for the probability of subject i dropping out at visit

j given all the observed information. More specifically, for $j = 1, ..., T$, we assume

$$\begin{aligned} & \text{logit}(\Pr(M_{ij} = 1 | M_{i(j-1)} = 0, X_i = x_i, Z_i = z_i, S_{ik} = s_{ik}, Y_{il} = y_{il}, \\ & \qquad k = 1, ..., j, l = 1, ..., j-1)) \\ = \; & h(x_i, z_i, s_{ik}, y_{il}, k = 1, ..., j, l = 1, ..., j-1, \beta, j). \end{aligned} \tag{4.2}$$

For simplicity consideration, we assume there is no interaction between x_i and other variables. More specifically, $h(x_i, z_i, s_{ik}, y_{il}, k = 1, ..., j, l = 1, ..., j-1, \beta, j) = \beta_{MX_j} x_i + h^*(z_i, s_{ik}, y_{il}, k = 1, ..., j, l = 1, ..., j-1, \beta, j)$.

Under the missing pattern $M_i = m$, $m = 0, ..., T-1$, we add a 'dropping effect' $\beta_{YM(m+1)}$ to the mean of $Y_{i(m+1)}$ and the vector $Y_{i1}, ..., Y_{i(m+1)}$ given X_i, Z_i, and S_{ik} $(k = 1, ..., m+1)$ are

$$\begin{aligned} & [Y_{i1}, ..., Y_{i(m+1)} | X_i = x_i, Z_i = z_i, M_i = m, S_{ik} = s_{ik}, k = 1, ..., m+1] \\ \sim \; & N_{m+1} \left(\begin{pmatrix} f(x_i, z_i, s_{ik}, k = 1, \beta, j = 1) \\ ... \\ f(x_i, z_i, s_{ik}, k = 1, ..., m, \beta, j = m) \\ f(x_i, z_i, s_{ik}, k = 1, ..., m+1, \beta, j = m+1) + \beta_{YM(m+1)} \end{pmatrix}, \Sigma_{m+1} \right). \end{aligned} \tag{4.3}$$

In total of T parameters, $\beta_{YM1}, ..., \beta_{YMT}$, quantify the difference between missing patterns, but they cannot be identified for the unrestricted model. MMNAR assumption implies that probability of $M_{ij} = 1$ does not depend on X_i given all the observed information including the potentially unobserved Y_{ij}. So $\forall j \in \{1, ..., T\}$

$$\begin{aligned} & \Pr(M_{ij} = 1 | M_{i(j-1)} = 0, X_i, Z_i, S_{ik}, Y_{il}, k = 1, ..., j, l = 1, ..., j, \beta) \\ = \; & \Pr(M_{ij} = 1 | M_{i(j-1)} = 0, Z_i, S_{ik}, Y_{il}, k = 1, ..., j, l = 1, ..., j, \beta). \end{aligned} \tag{4.4}$$

The model specified in Eq. (4.2) and Eq. (4.3) implied that the left side of Eq.

(4.4) follows a logistic form

$$\begin{aligned} & \text{logit}(\Pr(M_{ij}=1|M_{i(j-1)}=0, X_i=x_i, Z_i=z_i, S_{ik}=s_{ik}, Y_{il}=y_{il}, \\ & \qquad k=1,...,j, l=1,...,j, \beta)) \\ = \quad & \alpha_{M_jX}x_i + g(z_i, s_{ik}, y_{il}, k=1,...,j, l=1,...,j, \beta, j), \end{aligned} \tag{4.5}$$

where α_{M_jX} and $g(\cdot)$ are known functions of β, $f(\cdot)$, and $h(\cdot)$ (Details are in Appendix 4.7.1). Eq. (4.4) is equivalent to $\alpha_{M_jX}=0$, $j=1,...,T$, which provides T restrictions. Since there are T unidentified parameters from the unrestricted model and T restrictions are implied from the MMNAR assumption, the model is just identified. ML estimates and their variance covariance matrix can be obtained by maximizing the likelihood with MMNAR restrictions and applying the delta method. Detailed estimation procedure is described in Appendix 4.7.2 for the specific example in Section 4.4.

Note that when $M_i=m$, only Y_{ij} up to $j=m+1$ are defined in Eq. (4.3). Distribution of $Y_{ij}|M_{i(j-1)}=1$ is restricted with the second key assumption: dropout may depend on the current, possibly unobserved, measurement, but not on future measurements. Kenward et al. (2003) showed that this assumption is equivalent to a non-future dependent missing value restriction (NFMV) as follows:

$$\forall j > m+1 : \mathrm{d}(Y_{ij}|Y_{i1},...,Y_{i(j-1)}, M_i=m) = \mathrm{d}(Y_{ij}|Y_{i1},...,Y_{i(j-1)}, M_i \geq j-1), \tag{4.6}$$

where $\mathrm{d}(\cdot)$ denotes the probability density function. Eq. (4.6) becomes the following restriction with side effects and covariates included.

$$\begin{aligned} \forall j > m+1 : \quad & \mathrm{d}(Y_{ij}|Y_{i1},...,Y_{i(j-1)}, S_{i1},...,S_{ij}, X_i, Z_i, M_i=m) \\ = \quad & \mathrm{d}(Y_{ij}|Y_{i1},...,Y_{i(j-1)}, S_{i1},...,S_{ij}, X_i, Z_i, M_i \geq j-1) \end{aligned} \tag{4.7}$$

The right side of Eq. (4.7) is determined with the model in Eq. (4.3) and Eq. (4.2) because it only involves Y_{ij} up to $j = M_i + 1$. Then distribution of $Y_{ij}|M_{i(j-1)} = 1$ is fully defined by β, $f(\cdot)$, and $h(\cdot)$.

Although patients may discontinue the treatment assigned after dropout, this issue is not in the scope of this chapter and we focus on the missing outcome that would have been observed if the patient does not drop out and continues to receive the assigned treatment.

4.3.2 Step 2. Imputation step.

After the ML estimates and their variance covariance matrix are obtained, D samples of the parameter set are drawn. For each sample of the parameters, for subject i with $M_i = m < T$, $Y_{i(m+1)}$ is drawn based on conditional distribution of $Y_{i(m+1)}$ given X_i, Z_i, S_{ik}, $(k = 1, ..., m+1)$, and Y_{il} $(l = 1, ..., m)$ (Details in Appendix 4.7.4). Then Y_{ij} for $j = m + 2, ..., T$ are sequentially drawn based on the conditional distribution of Y_{ij} given X_i, Z_i, S_{ik}, $(k = 1, ..., j)$, and Y_{il} $(l = 1, ..., j - 1)$ (Details in Appendix 4.7.5).

After D imputed data sets are generated, the desired statistical analyses can be performed on each of the imputed data sets and the results are combined with Rubin's combination rule.

4.4 Simulation studies

In this section we investigate the behaviour of the proposed likelihood analysis through simulation studies. The sample size is set to 400 to reflect a typical applied scenario in clinical trials. The sample has two groups, with 200 subjects assigned to each group, and T=3 equally spaced repeated measurements. One baseline covariate Z_i is included and independent with the treatment assignment X_i. For $j = 1, ..., T$, S_{ij} is assumed to only depend on X_i, Z_i, and $S_{i(j-1)}$; $M_{ij}|M_{i(j-1)} = 0$ is assumed to

only depend on X_i, Z_i, S_{ij}, and $S_{i(j-1)}$; $Y_{ij}|M_{i(j-1)} = 0$ is assumed to only depend on X_i, Z_i, S_{ij}, and M_{ij}; σ_{ij} is assumed to be $\sigma^2\rho_{|i-j|}$, where $\rho_0 = 1$. Parameters are specified in Table 4.1. Note for subject i with $M_i = m$, $Y_{i1}, ..., Y_{i(m+1)}$ are generated with (4.3) directly, and $Y_{i(m+1)}, ..., Y_{iT}$ are generated with the NFMV restriction.

Table 4.1: Parameters used to generate longitudinal models

Scenario 5: the longitudinal model when MMNAR is true.	
Variable	Parameters
Z_i	$Z_i \sim N(1,1)$
$S_{ij}\|X_i, Z_i, S_{i(j-1)}$	$\text{logit}(S_{ij} = 1\|x_i, z_i, s_{i(j-1)}) = -1 + 0.15x_i + 0.3z_i + 0.1s_{i(j-1)}$
$M_{ij}\|M_{i(j-1)} = 0$	$h(x_i, z_i, s_{ik}, y_{il}, \beta, j) = -2 - x_i + 0.5z_i + s_{ij} + 0.5s_{i(j-1)}$
$Y_{ij}\|M_{i(j-1)} = 0$	$f(x_i, z_i, s_{ik}, \beta, j) = -1 + x_i + 0.5z_i + 0.2s_{ij}$
$Y_{ij}\|M_{i(j-1)} = 0$	$\{\beta_{YM1}, \beta_{YM2}, \beta_{YM3}\} = \{-1, -1.4, -1.6\}$
$Y_{ij}\|M_{i(j-1)} = 0$	$\{\sigma^2, \rho_1, \rho_2\} = \{1, 0.4, 0.3\}$
Scenario 6: the longitudinal model when both MMNAR and MAR are true.	
Variable	Parameters
Z_i	$Z_i \sim N(1,1)$
$S_{ij}\|X_i, Z_i, S_{i(j-1)}$	$\text{logit}(S_{ij} = 1\|x_i, z_i, s_{i(j-1)}) = -1 + 0.15x_i + 0.3z_i + 0.1s_{i(j-1)}$
$M_{ij}\|M_{i(j-1)} = 0$	$h(x_i, z_i, s_{ik}, y_{il}, \beta, j) = -2 + 0.5z_i + s_{ij} + 0.5s_{i(j-1)}$
$Y_{ij}\|M_{i(j-1)} = 0$	$f(x_i, z_i, s_{ik}, \beta, j) = -1 + x_i + 0.5z_i + 0.2s_{ij}$
$Y_{ij}\|M_{i(j-1)} = 0$	$\{\beta_{YM1}, \beta_{YM2}, \beta_{YM3}\} = \{0, 0, 0\}$
$Y_{ij}\|M_{i(j-1)} = 0$	$\{\sigma^2, \rho_1, \rho_2\} = \{1, 0.4, 0.3\}$

In scenario 5, MMNAR is true but MAR is violated because $\beta_{YM1}, \beta_{YM2}, \beta_{YM3}$ are nonzero. This scenario leads to the missing rate of $\Pr(M_i = 0) = 26.3\%$, $\Pr(M_i = 1) = 20.5\%$, and $\Pr(M_i = 2) = 14.0\%$. The mean of last measurement with $X_i = 0$ ($E(Y_{i3}|X_i = 0)$, considered to be the intercept) and the difference between two groups ($E(Y_{i3}|X_i = 1) - E(Y_{i3}|X_i = 0)$, considered to be the treatment effect) are of interest. 1000 sets of data are generated and for each set, complete case analysis, ML assuming MAR, ML assuming MMNAR, and before deletion data analysis are applied (details in Appendix 4.7.3). For each method, the empirical mean and MSE of the estimates, and the coverage proportion (CP) of the 95% confidence interval (CI) are presented in Table 4.2. The 95% CI is calculated with a normal approximation. The true values are also provided for comparison. In Table 4.2, treatment effect from both

Table 4.2: Finite sample properties of four methods based upon 1,000 simulations of a clinical trial with 400 patients. In scenario 5, MMNAR is true and in scenario 6, both MMNAR and MAR are true. MSE = $10^3 \times$ mean squared error. CP = $10^2 \times$ coverage proportion for 95% confidence interval. CP* is based on Bootstrap methods.

Method	Parameter	True	Mean	MSE	CP	CP*
Scenario 5: the longitudinal model when MMNAR is true.						
Before deletion	$E(Y_{i3}\|X_i = 0)$	-0.46	-0.45	11	94.4	NA
data analysis	$E(Y_{i3}\|X_i = 1) - E(Y_{i3}\|X_i = 0)$	1.27	1.27	21	94.9	NA
Complete case	$E(Y_{i3}\|X_i = 0)$	-0.46	-0.35	36	87.8	NA
analysis	$E(Y_{i3}\|X_i = 1) - E(Y_{i3}\|X_i = 0)$	1.27	1.20	43	93.6	NA
ML assuming	$E(Y_{i3}\|X_i = 0)$	-0.46	0.07	290	0.3	0.4
MAR	$E(Y_{i3}\|X_i = 1) - E(Y_{i3}\|X_i = 0)$	1.27	1.01	89	65.0	56.9
ML assuming	$E(Y_{i3}\|X_i = 0)$	-0.46	-0.47	28	97.1	95.8
MMNAR	$E(Y_{i3}\|X_i = 1) - E(Y_{i3}\|X_i = 0)$	1.27	1.29	28	96.7	94.3
Scenario 6: the longitudinal model when both MMNAR and MAR are true.						
Before deletion	$E(Y_{i3}\|X_i = 0)$	0.07	0.07	10	94.3	NA
data analysis	$E(Y_{i3}\|X_i = 1) - E(Y_{i3}\|X_i = 0)$	1.01	1.01	21	95.0	NA
Complete case	$E(Y_{i3}\|X_i = 0)$	0.07	-0.35	199	20.0	NA
analysis	$E(Y_{i3}\|X_i = 1) - E(Y_{i3}\|X_i = 0)$	1.01	1.00	49	94.9	NA
ML assuming	$E(Y_{i3}\|X_i = 0)$	0.07	0.07	11	97.5	94.1
MAR	$E(Y_{i3}\|X_i = 1) - E(Y_{i3}\|X_i = 0)$	1.01	1.01	22	97.7	95.0
ML assuming	$E(Y_{i3}\|X_i = 0)$	0.07	0.06	18	96.9	95.3
MMNAR	$E(Y_{i3}\|X_i = 1) - E(Y_{i3}\|X_i = 0)$	1.01	1.02	22	97.9	95.3

complete case analysis and ML assuming MAR are biased, showing the potential for bias when the missing data mechanism is MNAR. Our proposed method is essentially unbiased with small MSE, but the coverage proportions are larger than 95%. This is because when the imputation model (4.3) is correct and more complex than the analysis model which does not make any distribution assumption, the result from the multiple imputation combination rule may overestimate the variance (Meng, 1994). CP* are based on the variance estimated from bootstrap methods, and have coverage proportions close to 95%. Since before deletion analysis and CC analysis do not involve the MI procedure, bootstrap is not necessary.

In scenario 6, both MMNAR and MAR are true because all of β_{MX_j}, β_{YM1}, β_{YM2}, and β_{YM3} are 0. The parameters for generating the samples are specified in Table 4.1. They lead to the missing rate of $\Pr(M_i = 0) = 19.3\%$, $\Pr(M_i = 1) = 16.4\%$, and $\Pr(M_i = 2) = 12.1\%$. True values of $E(Y_{i3}|X_i = 0)$ and $E(Y_{i3}|X_i = 1) - E(Y_{i3}|X_i = 0)$, along with the empirical mean and MSE of the estimates, CP and CP* of the 95% CI for each method are presented in Table 4.2. Both ML assuming MAR and ML assuming MMNAR are unbiased with small MSE, and CP* is close to 95%. Note that MSE of ML assuming MAR and ML assuming MMNAR for $E(Y_{i3}|X_i = 1) - E(Y_{i3}|X_i = 0)$ are almost the same as MSE of before deletion analysis, which is because the information of model assumptions in (4.3) is used in the imputation steps of ML assuming MAR and ML assuming MMNAR, but the before deletion analysis does not make such assumptions.

4.5 Application

The data for the TROPHY study (Julius et al., 2006) consisted of 772 patients who were randomized to placebo group (381) and candesartan group (391). Let x_i be the binary treatment indicator with $x_i = 0$ denoting the placebo arm and $x_i = 1$ denoting the candesartan arm. Let z_i be the baseline covariates collected at the beginning of

the study. After the treatment was initiated, 9 visits were scheduled at month 1, 3, 6, and every 3 months until month 24. Let y_{ij} be the systolic blood pressures (SBP) measured at visit j for ith subject, $j = 1, ..., 9$. One reason to cause y_{ij} missing was development of hypertension. Development of hypertension was defined as the first appearance of one of the following outcomes: systolic pressure of 140 mm Hg or higher or diastolic pressure of 90 mm Hg or higher, for any three visits; systolic pressure of 160 mm Hg or higher or diastolic pressure of 100 mm Hg or higher for any visit; initiation of pharmacologic treatment; or systolic pressure of 140 mm Hg or higher or diastolic pressure of 90 mm Hg or higher at the visit at month 48. Once a patient was diagnosed of hypertension, a rescue therapy was initiated and the blood pressures recorded were no longer relevant. Among the 772 patients, 198 (47 in candesartan group) had missing values of y_{ij} because of developing hypertension. Since development of hypertension was fully determined by the observed information, the missing mechanism was MAR. Let $\Delta_i = 1$ denote that subject i developed hypertension and let $\Delta_i = 0$ denote that subject i completed the study. Let h_i denote the specific visit that hypertension was diagnosed. If $\Delta_i = 1$, y_{ih_i} was the last measurement of SBP.

Our particular interest here is the average last SBP before hypertension or the end of study for different treatment groups, or $E_{x_i} = E(y_{ih_i} * I(\Delta_i = 1) + y_{i9} * I(\Delta_i = 0) \mid x_i)$, where $I(\cdot)$ is the indicator function. Note that missing values of y_{ij} for $j > h_i$ are not related to our parameter of interest, so from this point, missing y_{ij} caused by development of hypertension ($\Delta_i = 1$) are not considered as missing data to avoid confusion.

The other reason to cause y_{ij} missing was dropout, which is denoted by $\Delta_i = 2$. Among those 772 patients, 84 (37 in candesartan group) dropped out from the study before development of hypertension. Let m_i denote the visit in which the last measurement of SBP was observed before dropout. Let m_{ij} be the missing data indicator for y_{ij} with $m_{ij} = 1$ denoting subject i dropping out before visit j. The

missing pattern is monotone so for subject i with $\Delta_i = 2$, $m_i = r \Leftrightarrow m_{ij} = 0$ for $j = 1, ..., r$ and $m_{ij} = 1$ for $j = r+1, ..., 9$. If $\Delta_i \neq 2$, $m_{ij} = 0$ for $j = 1, ..., 9$.

Since blood pressures could be measured at home, patients might drop out because of lack of efficacy and therefore MAR is questionable. The proposed MMNAR model in Section 4.4 is applied except the true parameters are unknown. One naive longitudinal analysis was performed to determine the baseline covariates to be included in the MMNAR model. Race, sex, BMI, age, creat, and baseline SBP were significantly related to outcomes and therefore included in z_i. Time trend for SBP was not observed and excluded.

Following the fitting procedure described above, ML estimates are obtained and imputed data sets are generated, but after all the missing values of y_{ij} are imputed, an extra imputation step is required to create the possibility to diagnose hypertension. In this step, we assume the probability to diagnose hypertension only depends on the last observed SBP by a logistic form: $\text{logit}(\Pr(h_{i(j+1)} = 1 | y_{ij}, h_{ij} = 0)) = \beta_{H0} + \beta_{HY} y_{ij}$, $j = 1, ..., 8$, where h_{ij} is the hypertension indicator, $h_i = w \Leftrightarrow h_{ij} = 0$ for $j = 1, ..., w$ and $h_{ij} = 1$ for $j = w+1, ..., 9$. The ML estimates of β_{H0} and β_{HY} and their variance covariance matrix are obtained by fitting a logistic regression with all observed y_{ij}. Then for each imputed data set, following the imputation step, one set of β_{H0} and β_{HY} is randomly drawn to generate h_{ij} sequentially until $h_{ij} = 1$ or $j = 9$.

The analysis for complete data does not make any distribution assumption. $\hat{E}_0$ and $\hat{E}_1$ (estimates of E_0 and E_1), and $\hat{V}(E_0)$ and $\hat{V}(E_1)$ (their variances) are the empirical means and variances for the two treatment groups. Estimate of the treatment effect $E_1 - E_0$ and its variance are simply $\hat{E}_1 - \hat{E}_0$ and $\hat{V}(E_0) + \hat{V}(E_1)$. The results from complete case analysis, ML assuming MAR, and ML assuming MMNAR are presented in Table 4.3.

The estimated treatment effect from ML assuming MMNAR is slightly larger than the one from ML assuming MAR. It indicates that allowing the missingness of

Table 4.3: Estimates of the systolic blood pressure (SBP) at endpoint under different treatment groups in TROPHY study. LCI = lower bound of 95% confidence interval. HCI = higher bound of 95% confidence interval. Width = width of 95% confidence interval.

Method	Parameter	Mean	LCI	HCI	Width
Complete case	E_0	129.0	127.8	130.2	2.4
analysis	E_1-E_0	-6.7	-8.4	-5.1	3.3
ML assuming	E_0	130.8	129.5	132.0	2.5
MAR	E_1-E_0	-7.9	-9.6	-6.2	3.5
ML assuming	E_0	132.4	130.4	134.5	4.2
MMNAR	E_1-E_0	-8.8	-10.7	-6.9	3.9

y_{ij} to depend on y_{ij} itself may alter the treatment effect. This also confirms that the treatment effect to decrease blood pressures is robust to deviation from MAR assumption. E_0 has relatively wide confidence interval for ML assuming MMNAR, but the width of $E_1 - E_0$ confidence interval for ML assuming MMNAR is very similar to the one for ML assuming MAR. This observation is very similar to the pattern in previous simulation studies.

4.6 Discussion

In this section we describe MMNAR models for repeated measure data. Since the longitudinal data models are flexible, rather than specifying one MMNAR model, we provide the strategy to extend a complete data model to a MMNAR model and estimate parameters of interest by applying maximum likelihood and multiple imputation. Since the imputation model may be more complex than the analysis model for the complete data, the variance estimates from the multiple imputation combination rule may be overestimated and a bootstrap method is suggested to correct the variance estimation. Applying longitudinal MMNAR models to real data is also illustrated with TROPHY study, and ML assuming MMNAR provides a result that

is slightly different from ML assuming MAR. This indicates that MAR may be mildly violated.

In the TROPHY study application, side effect variables are summarized to one single variable to be included in the model, and the way to summarize those side effect data can be improved for the future work, for example, by using a propensity score which models the propensity to drop out.

In the future work, we will also extend the MMNAR assumption to other parametric or semiparametric regression models, like generalized linear models and survival analysis models. For those models, restrictions implied by the MMNAR assumption may not be as well defined as the logistic form illustrated above and how to apply the MMNAR assumption properly requires further investigation.

We already considered the case when the outcome is missing, and the MMNAR assumption can be useful for missing covariates when MAR is questionable. If the missing probability does not depend on outcomes, the estimates based on CC analysis may be valid, but efficiency loss is often unavoidable. More sophisticated methods such as subsample ignorable likelihood (SSIL) method (Little and Zhang, 2011) often assume MAR, but they are robust to slight violation of ignorability. The comparison between ML assuming MMNAR and other methods for different settings are of interest.

In the longitudinal MMNAR model, we assume the missing pattern is monotone and adopt a sequential MI procedure for the inference. For future work, it is interesting to see how to apply chained equation MI methods to the longitudinal data with arbitrary missing pattern. This can greatly simplify the application of the method by using existing software include IVEware (Raghunathan et al., 2001) and MICE (VanBuuren and Groothuis-Oudshoorn, 2011).

4.7 Appendix

4.7.1 Coefficients in Eq. (4.5)

$$
\begin{aligned}
& \operatorname{logit}(\Pr(M_{ij}=1|X_i,Z_i,S_{ik},k=1,...,j,Y_{il},l=1,...,j,M_{i(j-1)}=0,\beta)) \quad (4.8)\\
= \; & \log\frac{\Pr(M_{ij}=1|X_i,Z_i,S_{ik},k=1,...,j,Y_{il},l=1,...,j,M_{i(j-1)}=0,\beta)}{\Pr(M_{ij}=0|X_i,Z_i,S_{ik},k=1,...,j,Y_{il},l=1,...,j,M_{i(j-1)}=0,\beta)}\\
= \; & \log\frac{\Pr(M_{ij}=1,X_i,Z_i,S_{ik},k=1,...,j,Y_{il},l=1,...,j,M_{i(j-1)}=0,\beta)}{\Pr(M_{ij}=0,X_i,Z_i,S_{ik},k=1,...,j,Y_{il},l=1,...,j,M_{i(j-1)}=0,\beta)}\\
= \; & \log\frac{\mathrm{d}(Y_{ij}|M_{ij}=1,X_i,Z_i,S_{ik},k=1,...,j,Y_{il},l=1,...,j-1,M_{i(j-1)}=0,\beta)}{\mathrm{d}(Y_{ij}|M_{ij}=0,X_i,Z_i,S_{ik},k=1,...,j,Y_{il},l=1,...,j-1,M_{i(j-1)}=0,\beta)} \quad (4.9)\\
+ \; & \log\frac{\Pr(M_{ij}=1|X_i,Z_i,S_{ik},k=1,...,j,Y_{il},l=1,...,j-1,M_{i(j-1)}=0,\beta)}{\Pr(M_{ij}=0|X_i,Z_i,S_{ik},k=1,...,j,Y_{il},l=1,...,j-1,M_{i(j-1)}=0,\beta)} \quad (4.10)
\end{aligned}
$$

Expression (4.10) is $\beta_{MX_j}X_i + h^*(Z_i,S_{ik},Y_{il},k=1,...,j,l=1,...,j-1,\beta,j)$ based on Eq. (4.2). Eq. (4.3) implies that $[Y_{ij}|M_{ij}=0,X_i,Z_i,S_{ik},k=1,...,j,Y_{il},l=1,...,j-1,M_{i(j-1)}=0,\beta]$ follows a normal distribution with mean of μ_j and variance of σ_j^2. μ_j is

$$
\begin{aligned}
& f(X_i,Z_i,S_{ik},k=1,...,j,\beta,j)\\
& +\Sigma'_{1(j-1)}\Sigma_{j-1}^{-1}\begin{pmatrix} Y_{i1}-f(X_i,Z_i,S_{ik},k=1,\beta,1)\\ ...\\ Y_{i(j-1)}-f(X_i,Z_i,S_{ik},k=1,...,j-1,\beta,j-1)\end{pmatrix} \quad (4.11)\\
= \; & f^*(Z_i,S_{ik},k=1,...,j,\beta,j)\\
& +\Sigma'_{1(j-1)}\Sigma_{j-1}^{-1}\begin{pmatrix} Y_{i1}-f^*(Z_i,S_{ik},k=1,\beta,1)\\ ...\\ Y_{i(j-1)}-f^*(Z_i,S_{ik},k=1,...,j-1,\beta,j-1)\end{pmatrix}\\
& +[\beta_{YX_j}-\Sigma'_{1(j-1)}\Sigma_{j-1}^{-1}(\beta_{YX_1},...,\beta_{YX_{j-1}})']X_i\\
= \; & \mu_j^*+[\beta_{YX_j}-\Sigma'_{1(j-1)}\Sigma_{j-1}^{-1}(\beta_{YX_1},...,\beta_{YX_{j-1}})']X_i,
\end{aligned}
$$

where $\Sigma_{1(j-1)}$ is the vector $(\sigma^2_{11}, ..., \sigma^2_{1(j-1)})$ and μ^*_j does not depend on X_i. Variance σ^2_j has the form of $\sigma^2_{jj} - \Sigma'_{1(j-1)}\Sigma^{-1}_{j-1}\Sigma_{1(j-1)}$. Similarly, $[Y_{ij}|M_{ij} = 1, X_i, Z_i, S_{ik}, k = 1, ..., j, Y_{il}, l = 1, ..., j-1, M_{i(j-1)} = 0, \beta]$ follows a normal distribution with mean of $\mu_j + \beta_{YMj}$ and variance of σ^2_j. Plug μ_j, σ^2_j, and β_{YMj} in (4.9), we get

$$\begin{aligned}
& -1/2 \cdot \log(2\pi\sigma^2_j) - \frac{(Y_{ij} - \mu_j - \beta_{YMj})^2}{2\sigma^2_j} - (-1/2 \cdot \log(2\pi\sigma^2_j)^{-1/2} - \frac{(Y_{ij} - \mu_j)^2}{2\sigma^2_j}) \\
= \; & (2Y_{ij} - 2\mu_j - \beta_{YMj})\beta_{YMj}/(2\sigma^2_j) \\
= \; & (2Y_{ij} - 2\mu^*_j - \beta_{YMj})\beta_{YMj}/(2\sigma^2_j) \\
& -[\beta_{YX_j} - \Sigma'_{1(j-1)}\Sigma^{-1}_{j-1}(\beta_{YX_1}, ..., \beta_{YX_{j-1}})']X_i\beta_{YMj}/\sigma^2_j
\end{aligned}$$

Plug in the expressions of (4.9) and (4.10), we have

$$\begin{aligned}
& \text{logit}(\Pr(M_{ij} = 1|X_i, Z_i, S_{ik}, k = 1, ..., j, Y_{il}, l = 1, ..., j, M_{i(j-1)} = 0, \beta)) \\
= \; & (2Y_{ij} - 2\mu^*_j - \beta_{YMj})\beta_{YMj}/(2\sigma^2_j) \\
& -[\beta_{YX_j} - \Sigma'_{1(j-1)}\Sigma^{-1}_{j-1}(\beta_{YX_1}, ..., \beta_{YX_{j-1}})']X_i\beta_{YMj}/\sigma^2_j \\
& +\beta_{MX_j}X_i + h^*(Z_i, S_{ik}, Y_{il}, k = 1, ..., j, l = 1, ..., j-1, \beta, j) \\
= \; & \{\beta_{MX_j} - [\beta_{YX_j} - \Sigma'_{1(j-1)}\Sigma^{-1}_{j-1}(\beta_{YX_1}, ..., \beta_{YX_{j-1}})']\beta_{YMj}/\sigma^2_j\}X_i \\
& +(2Y_{ij} - 2\mu^*_j - \beta_{YMj})\beta_{YMj}/(2\sigma^2_j) \\
& +h^*(Z_i, S_{ik}, Y_{il}, k = 1, ..., j, l = 1, ..., j-1, \beta, j)
\end{aligned}$$

α_{M_jX} in Eq. (4.5) is $\beta_{MX_j} - [\beta_{YX_j} - \Sigma'_{1(j-1)}\Sigma^{-1}_{j-1}(\beta_{YX_1}, ..., \beta_{YX_{j-1}})']\beta_{YMj}/\sigma^2_j$.

4.7.2 Obtain ML estimates for MMNAR models

Since there are T unidentified parameters from the unrestricted model and T restrictions are implied from the MMNAR assumption, the model is just identified. The loglikelihood is maximized by (a) calculating the ML estimates of parameters in $h(\cdot)$ with M_{ij} given X_i, Z_i, S_{ij}, and $S_{i(j-1)}$ for all $\{i, j\}$ that satisfy $M_{i(j-1)} = 0$;

(b) calculating the ML estimates of parameters in $f(\cdot)$ with $Y_{i1}, ..., Y_{im}$ given X_i, Z_i, $S_{i1}, ..., S_{im}$, for subjects with $M_i = m$; and (c) calculating the estimated value of $\beta_{YM1}, ..., \beta_{YM3}$ with restrictions $\alpha_{M_jX} = 0$, $j = 1, 2, 3$, implied by the MMNAR assumption (forms of α_{M_jX} in Appendix 4.7.1). Variance estimates in (a) and (b) are computed from the inverse of the information matrix and delta method is used to calculate the variance estimates in (c).

4.7.3 Estimation procedures for different methods

Let $\hat{\beta}_{\text{intercept}}$ and $\hat{V}_{\text{intercept}}$ be the estimate of $E(Y_{i3}|X_i = 0)$ and the associated variance estimate. Let $\hat{\beta}_{\text{trt.effect}}$ and $\hat{V}_{\text{trt.effect}}$ denote the estimate of $E(Y_{i3}|X_i = 1) - E(Y_{i3}|X_i = 0)$ and the associated variance estimate.

We first introduce the estimation method for complete data, in which no distribution assumption is made. We use the empirical mean and variance of Y_{i3} for treatment group X_i to estimate $E(Y_{i3}|X_i)$ and the associated variance, which are denoted by $\hat{\beta}_{X_i}$ and $\hat{V}_{X_i}$. Then we have $\hat{\beta}_{\text{intercept}} = \hat{\beta}_0$, $\hat{V}_{\text{intercept}} = \hat{V}_0$, $\hat{\beta}_{\text{trt.effect}} = \hat{\beta}_1 - \hat{\beta}_0$, and $\hat{V}_{\text{trt.effect}} = \hat{V}_1 + \hat{V}_0$.

Complete case analysis and before deletion data analysis are direct application of the estimation method for complete data. Note that this method is also the analysis performed after imputed data sets are generated in ML assuming MMNAR or ML assuming MAR. Estimation procedure for ML assuming MMNAR is provided in Section 4.3. ML assuming MAR has a very similar procedure, which sets estimates and the variances of $\beta_{YM1}, ..., \beta_{YM3}$ to 0 instead of calculating them with the restrictions (step (c) in Appendix 4.7.2).

4.7.4 Imputing $Y_{i(m+1)}$ for $M_i = m < T$

Since $M_i = m \Leftrightarrow M_{im} = 0$ & $M_{i(m+1)=1}$, conditional distribution of $Y_{i(m+1)}$ given X_i, Z_i, S_{ik}, $(k = 1, ..., m+1)$, and Y_{il} $(l = 1, ..., m)$ is $\mu_j + \beta_{YMj}$, where μ_j is given in

(4.11).

4.7.5 Imputing Y_{ij} for $j = m+2, ..., T$, $M_i = m < T-1$

Because Eq. (4.7), for $M_i = m < T-1$, $\forall j > m+1$,

$$
\begin{aligned}
& \mathrm{d}(Y_{ij}|Y_{i1}, ..., Y_{i(j-1)}, S_{i1}, ..., S_{ij}, X_i, Z_i, M_i = m) \\
= \; & \mathrm{d}(Y_{ij}|Y_{i1}, ..., Y_{i(j-1)}, S_{i1}, ..., S_{ij}, X_i, Z_i, M_i \geq j-1) \\
= \; & \sum_{l=j-1}^{T} \mathrm{d}(Y_{ij}|Y_{i1}, ..., Y_{i(j-1)}, S_{i1}, ..., S_{ij}, X_i, Z_i, M_i = l) \\
& \cdot \Pr(M_i = l|Y_{i1}, ..., Y_{i(j-1)}, S_{i1}, ..., S_{ij}, X_i, Z_i, M_i \geq j-1) \\
= \; & \sum_{l=j-2}^{T} \mathrm{d}(Y_{ij}|Y_{i1}, ..., Y_{i(j-1)}, S_{i1}, ..., S_{ij}, X_i, Z_i, M_i = l) \\
& \cdot \Pr(M_i = l|Y_{i1}, ..., Y_{i(j-1)}, S_{i1}, ..., S_{ij}, X_i, Z_i, M_i \geq j-1) \\
& + \mathrm{d}(Y_{ij}|Y_{i1}, ..., Y_{i(j-1)}, S_{i1}, ..., S_{ij}, X_i, Z_i, M_i = j-1) \\
& \cdot \Pr(M_i = j-1|Y_{i1}, ..., Y_{i(j-1)}, S_{i1}, ..., S_{ij}, X_i, Z_i, M_i \geq j-1)
\end{aligned}
$$

Note that since $[Y_{i1}, ..., Y_{i(j-1)}, Y_{ij}|S_{i1}, ..., S_{ij}, X_i, Z_i, M_i = l]$ is the same for all $l \geq j-2$, $[Y_{ij}|Y_{i1}, ..., Y_{i(j-1)}, S_{i1}, ..., S_{ij}, X_i, Z_i, M_i = l]$ is the same for all $l \geq j-2$. We

have

$$
\begin{aligned}
& \mathrm{d}(Y_{ij}|Y_{i1},...,Y_{i(j-1)},S_{i1},...,S_{ij},X_i,Z_i,M_i=m) \\
= & \sum_{l=j-2}^{T}\Pr(M_i=l|Y_{i1},...,Y_{i(j-1)},S_{i1},...,S_{ij},X_i,Z_i,M_i\geq j-1) \\
& \cdot\mathrm{d}(Y_{ij}|Y_{i1},...,Y_{i(j-1)},S_{i1},...,S_{ij},X_i,Z_i,M_i=T) \\
& +\mathrm{d}(Y_{ij}|Y_{i1},...,Y_{i(j-1)},S_{i1},...,S_{ij},X_i,Z_i,M_{ij}=1,M_{i(j-1)}=0) \\
& \cdot\Pr(M_{ij}=1|Y_{i1},...,Y_{i(j-1)},S_{i1},...,S_{ij},X_i,Z_i,M_{i(j-1)}=0) \\
= & \mathrm{d}(Y_{ij}|Y_{i1},...,Y_{i(j-1)},S_{i1},...,S_{ij},X_i,Z_i,M_{ij}=0,M_{i(j-1)}=0) \\
& \cdot\Pr(M_{ij}=0|Y_{i1},...,Y_{i(j-1)},S_{i1},...,S_{ij},X_i,Z_i,M_{i(j-1)}=0) \\
& +\mathrm{d}(Y_{ij}|Y_{i1},...,Y_{i(j-1)},S_{i1},...,S_{ij},X_i,Z_i,M_{ij}=1,M_{i(j-1)}=0) \\
& \cdot\Pr(M_{ij}=1|Y_{i1},...,Y_{i(j-1)},S_{i1},...,S_{ij},X_i,Z_i,M_{i(j-1)}=0) \\
= & \phi(Y_{ij},\mu_j,\sigma_j^2)\frac{1}{\exp\{1+h(Z_i,S_{ik},Y_{il},k=1,...,j,l=1,...,j-1,\beta,j)\}} \\
& +\phi(Y_{ij},\mu_j+\beta_{YMj},\sigma_j^2)\frac{\exp\{1+h(Z_i,S_{ik},Y_{il},k=1,...,j,l=1,...,j-1,\beta,j)\}}{1+\exp\{1+h(Z_i,S_{ik},Y_{il},k=1,...,j,l=1,...,j-1,\beta,j)\}},
\end{aligned}
$$

where $\phi(x,\mu,\sigma^2)$ denotes the normal density function with mean μ and variance σ^2.

CHAPTER V

Conclusions and Future Work

We consider missing data problems in randomized controlled trials. Noncompliance with the assigned treatment and missing outcomes can undermine the randomization, which is a key feature for drawing valid statistical inferences for the comparison between treatments. For the noncompliance issue, we focus on the treatment switching for phase III clinical trials in oncology and propose a likelihood-based method and a latent event time method. The proposed methods outperform existing methods in the simulation studies conducted. For the missing outcomes issue, a specific MNAR assumption is proposed and applied to a variety of models. The TROPHY study provides a real data application.

In Chapter 2, we propose a new parametric method to address the treatment arm switching issue. This method evaluates the individual benefit of switching based on observed progression time so the switching effect differs among patients. We also propose a latent event time method based on the iterative parameter estimation procedure. This method is more robust to violations of the distribution assumptions in the parametric method.

Via simulations, we show that our proposed methods are unbiased with small MSEs, but the existing methods, ITT analysis, PP analysis, Cox PH model with time-varying covariate, and the IPE procedure may be biased when treatment switching

depends on the subjects' progression time. We analyze the reasons for bias, in the case of the IPE procedure we suggest that bias arises because the selection of switching can change the causal effect.

For the inference to be valid, the decision of treatment switching must not depend on the future information (for example, investigator's judgement for the event time) conditional on the observed information at progression time. This assumption is unverifiable from the data being analyzed and questionable in practice, so we recommend weakening it by collecting variables associated with both treatment switching and the future event time.

Our key assumption that the treatment effect is the same both before and after disease progression, that is $D_{Ti} \equiv D_{Gi} \equiv D_{TGi}$ using the notation of Chapter 2, may appear strong but is intuitive. It provides an idea to understand the association between the causal effect of treatment with survival both before progression and after progression and leads to our proposed adjustments for treatment arm switching. For the future work, the assumption may be relaxed by assuming D_{Gi} and D_{Ti} are positively correlated instead of equal to each other. Note that although the correlation between D_{Gi} and D_{Ti} (denoted as ρ_{TG}) may be unidentifiable, the robust approaches proposed in Section 2.5 with ρ_1 replaced by $\rho_1\rho_{TG}$ could be applied, although a sensitivity analysis for the assumed value of $\rho_1\rho_{TG}$ would still be required.

A potential approach to recensoring for our latent event time approach is under investigation. We will also apply the proposed idea to semiparametric methods, such as a Cox PH model, in which the hazard ratio between treatment groups before progression is assumed to be equal to the hazard ratio after progression and the switching effect can be quantified individually.

In Chapter 3, we propose the masked missing not at random (MMNAR) assumption for masked clinical trials, where missingness does not depend on the masked treatment, after conditioning on side effect and outcome data. The MMNAR as-

sumption can be considered as an alternative to MAR or a way to limit the number of possible MNAR models for the sensitivity analysis.

We describe MMNAR models for continuous and categorical outcomes, and apply maximum likelihood and multiple imputation based on these models to estimate treatment effects. Simulation studies show that when MAR is violated, assuming MAR can lead to severe bias. Therefore if the estimates are substantially different, which indicates a deviation from MAR, we might prefer the estimates from MMNAR because the mechanism is more plausible but if the estimates based on MMNAR and MAR are similar, MAR analysis might be preferred for efficiency considerations. However, the mean squared errors of treatment effect estimates for MMNAR are very similar to those for MAR in almost all the simulations conducted. How to generalize this observation with theoretical approaches is still under investigation.

Clinically unimportant side effects do not need to be included because to bias the treatment effect a side effect needs to be associated with the outcome as well as the likelihood of dropping out. However, with extensive data on side effects methods for summarizing them are of interest, and better methods than those used in the TROPHY study application is a topic for future research. One potential approach is to use a propensity score that models the propensity to drop out.

We will also extend the MMNAR assumption to other parametric or semiparametric regression models, like generalized linear models and survival analysis models. In those models, the MMNAR assumption may imply some restrictions that are not in a well recognized form such as logistic. How to utilize those restrictions to identify MMNAR models requires further investigation.

In Chapter 4, we extend our proposed MMNAR assumption to repeated measure data. Instead of specifying some MMNAR model examples in Chapter 3, we provide the strategy to extend a complete data model to a MMNAR model and estimate parameters of interest by applying maximum likelihood and multiple imputation.

We notice the variance estimates from the multiple imputation combination rule may be overestimated because the difference between imputation model and analysis model, and the bootstrap approach can provide a valid estimate of the variance. Finite sample properties of the proposed method are examined by simulation studies, the result shows the same pattern as observed in the univariate outcome variable simulations in Chapter 3. When MAR is violated, The method assuming MAR can be biased when MAR is violated but it is preferred when MAR is correct. The application to TROPHY study shows a slightly difference between assuming MAR and assuming MMNAR.

We have considered the simple case when the missing pattern is monotone. However, for other missing patterns, or situations with missing covariates, the sequential MI procedure introduced cannot be applied and other approaches such as chained equation MI methods need to be developed. It would also be worthwhile to develop weighting methods for a monotone pattern with masked data, and compare them with our likelihood based methods.

BIBLIOGRAPHY

BIBLIOGRAPHY

Agresti, A. (2002). Categorical data analysis. *New York: Wiley*.

Blay, J.Y. (2010). Pharmacological management of gastrointestinal stromal tumours: an update on the role of sunitinib. *Annals of Oncology* **21**, 208-215.

Branson, M. and Whitehead, J. (2002). Estimating a treatment effect in survival studies in which patients switch treatment. *Statistics in Medicine* **21**, 2449-2463.

Concorde Coordinating Committee (1994). Concorde: MRC/ANRS randomised double-blind controlled trial of immediate and deferred zidovudine in symptom-free HIV infection. *The Lancet* **343**, 871-881.

Diggle, P.J. and Kenward, M.G. (1994). Informative dropout in longitudinal data analysis. *Applied Statistics* **43**, 49-93.

Fitzmaurice, G.M., Lipsitz, S.R., Molenberghs, G., and Ibrahim, J.G. (2005). A protective estimator for longitudinal binary data with non-ignorable non-monotone missingness. *Journal of the Royal Statistical Society A* **168**, 723-735.

Hedeker, D. and Gibbons, R. (1997). Application of random-effects pattern-mixture models for missing data in longitudinal studies. *Psychological Methods* **2**, 64-78.

Hernan, M.A., Brumback, B., and Robins, J.M. (2000). Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men. *Epidemiology* **11**, 561-570.

Hill, AB. (1961). Principles of Medical Statistics. 7th edn. *The Lancet: London*.

Julius, S., Nesbitt, S.M., Egan, B.D., Weber, M.A., Michelson, E.L., Kaciroti, N., Black, H.R., Grimm, R.H., Messerli, F.H., Oparil, S., and Schork, M.A., for the Trial of Preventing Hypertension (TROPHY) Study Investigators (2006). Feasibility of treating prehypertension with an angiotensin-receptor blocker. *New England Journal of Medicine* **354**, 1685-1697.

Kenward, M.G., Molenberghs, G., and Thijs, H. (2003). Pattern-mixture models with proper time dependence. *Biometrika* **90**, 53C71.

Knox, J.J. (2008). Progression-free survival as endpoint in metastatic RCC? *The Lancet* **372**, 427-429.

Law, M.G. and Kaldor, J.M. (1996). Survival analyses of randomized clinical trials adjusted for patients who switch treatments. *Statistics in Medicine* **15**, 2069-2076.

Lewis, R., Bagnall, A.M., Forbes, C., Shirran, E., Duffy, S., Kleijnen, J., Riesma, R., and ter Riet, G. (2002). The clinical effectiveness of trastuzumab for breast cancer: a systematic review. *Health Technology Assessment* **6**.

Liang, K.Y. and Zeger, S.L. (1986). Longitudinal Data Analysis Using Generalized Linear Models. *Biometrika* **73**, 13C22.

Little, R.J. (1993). Pattern-mixture models for multivariate incomplete data. *Journal of the American Statistical Association* **88**, 125-234.

Little, R.J., Cohen, M.L., Dickersin, K., Emerson, S.S., Farrar, J.T., Neaton, J.D., Shih, W., Siegel, J.P., and Stern, H. (2012). The design and conduct of clinical trials to limit missing data. *Statistics in Medicine* **31**, 3433-3443.

Little, R.J., Long, Q., and Lin, X. (2009). A comparison of methods for estimating the causal effect of a treatment in randomized clinical trials subject to noncompliance. *Biometrics* **65**, 640-649.

Little, R.J. and Rubin, D.B. (2002). Statistical analysis with missing data. *New York: Wiley*

Little, R.J. and Zhang, N. (2011). Subsample ignorable likelihood for regression with missing data. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **60**, 591-605.

Mallinckrodt, C.H., Lin, Q., and Molenberghs, M. (2013). A structured framework for assessing sensitivity tomissing data assumptions in longitudinal clinical trials. *Pharmaceutical Statistics* **12**, 1-6.

Meinert, C.L. (1980). Toward more definitive clinical trials. *Controlled Clinical Trials* **1**, 249-61.

Meng, X. (1994). Multiple-imputation inferences with uncongenial sources of input. *Statistical Science* **9**, 538-573.

Molenberghs, G. and Kenward. M.G. (2007). Missing Data in Clinical Studies. *John Wiley: Chichester, UK*

Molenberghs, G., Michiels, B., Kenward, M.G., and Diggle, P.J. (1998). Missing data mechanisms and pattern-mixture models. *Statistica Neerlandica* **52**, 153C161.

Molenberghs, G., Thijs, H., Kenward, M.G., and Verbeke, G. (2003). Sensitivity Analysis of Continuous Incomplete Longitudinal Outcomes. *Statistica Neerlandica* **57**, 112C135.

Morden, J.P., Lambert, P.C., Latimer, N., Abrams, K.R., and Wailoo, A.J. (2011). Assessing methods for dealing with treatment switching in randomised controlled trials: a simulation study. *BMC Medical Research Methodology* **11:4**.

Motzer, R.J., Escudier, B., Oudard, S., Hutson, T.E., Porta, C., Bracarda, S., Grunwald, V., Thompson, J.A., Figlin, R.A., Hollaender, N., Kay, A., and Ravaud, A. (2010). Phase 3 Trial of Everolimus for Metastatic Renal Cell Carcinoma: Final Results and Analysis of Prognostic Factors. *Cancer* **116**, 4256-4265.

Motzer, R.J., Escudier, B., Oudard, S., Hutson, T.E., Porta, C., Bracarda, S., Grunwald, V., Thompson, J.A., Figlin, R.A., Hollaender, N., Urbanowitz, G., Berg, W.J., Kay, A., Lebwohl, D., and Ravaud, A. (2008). Efficacy of everolimus in advanced renal cell carcinoma: a double-blind, randomised, placebo-controlled phase III trial. *The Lancet* **372**, 449-456.

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

Raghunathan, T.E., Lepkowski, J.M., VanHoewyk, J., and Solenberger, P. (2001). A multivariate technique for multiply imputing missing values using a sequence of regression models. *Survey Methodology* **27**, 85-95.

Ratitch, B., OKelly, M., and Tosielloc, R. (2013). Missing data in clinical trials: from clinical assumptions to statistical analysis using pattern-mixture models. *Pharmaceutical Statistics* **12**, 337-347.

Robertson, J.F.R., Ferrero., J.M., Bourgeois, H., Kennecke, H., de Boer, R.H., Jacot, W., McGreivy, J., Suzuki, S., Zhu, M., McCaffery, I., Loh, E., Gansert, J.L., and Kaufman, P.A. (2013). Ganitumab with either exemestane or fulvestrant for postmenopausal women with advanced, hormone- receptor-positive breast cancer: a randomised, controlled, double-blind, phase 2 trial. *The Lancet Oncology* **14**, 228-235.

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

Robins, J.M. and Tsiatis, A.A. (1991). Correcting for non-compliance in randomized trials using rank preserving structural failure time models. *Communications in Statistics - Theory and Methods* **20**, 2609-2631.

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

Rubin, D.B. (1987). Multiple imputation for nonresponse in surveys. *New York: Wiley*.

Scharfstein, D., Rotnizky, A., and Robins, J.M. (1999). Adjusting for non-ignorable drop-out using semiparametric non-response models (with discussion). *Journal of the American Statistical Association* **94**, 1096-1146.

Shao, J., Chang, M., and Chow, S.C. (2005). Statistical inference for cancer trials with treatment switching. *Statistics in Medicine* **24**, 1783-1790.

Slamon, D.J., Leyland-Jones, B., Shak, S., Fuchs, H., Paton, V., Bajamonde, A., Fleming, T., Eiermann, W., Wolter, J., Pegram, M., Baselga, J., and Norton, L. (2001). Use of Chemotherapy Plus a Monocolonal Antibody Against HER2 for Metstatic Breast Cancer that Overexpresses HER2. *The New England Journal of Medicine* **344**, 783-792.

Sommer, A. and Zeger, S.L. (1991). On estimating efficacy from clinical trials. *Statistics in Medicine* **10**, 45-52.

VanBuuren, S. and Groothuis-Oudshoorn, K. (2011). Mice: multivariate imputation by chained equations in R. *Journal of Statistical Software* **45**, 1-67.

White, I.R. (2006). Estimating treatment effects in randomized trials with treatment switching. *Statistics in Medicine* **25**, 1619-1622.

White, I.R., Babiker, A.G., Walker, S., and Barbyshire, J.H. (1999). Randomization-based Methods for Correcting for Treatment Changes: Examples from the Concorde Trial. *Statistics in Medicine* **18**, 2617-2634.

Zheng, D., Chen, Q., Chen., M.H., Ibrahim., J.G., and Amgen Research Group (2012). Estimating treatment effects with treatment switching via semi-competing risks models: An application to a colorectal cancer study. *Biometrika* **99**, 167-84.