Evaluating Treatment Efficacy in Randomized Controlled Trials with Treatment Noncompliance and Multivariate Outcomes

by

Lulu Guo

M.Sc., Shandong University, 2018
B.Sc., Zhongnan University of Economics and Law, 2015

Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of
Doctor of Philosophy

in the
Department of Statistics and Actuarial Science
Faculty of Science

SIMON FRASER UNIVERSITY
Summer 2024

Declaration of Committee

Name:	**Lulu Guo**
Degree:	**Doctor of Philosophy**
Thesis title:	**Evaluating Treatment Efficacy in Randomized Controlled Trials with Treatment Noncompliance and Multivariate Outcomes**
Committee:	**Chair:** Liangliang Wang Associate Professor, Statistics and Actuarial Science

Hui Xie
Co-Supervisor
Professor, Health Sciences

X. Joan Hu
Co-Supervisor
Professor, Statistics and Actuarial Science

Yi Qian
Committee Member
Associate Professor, Business
University of British Columbia

Miremad Soleymanian
Committee Member
Assistant Professor, Business

Rachel MacKay Altman
Examiner
Associate Professor, Statistics and Actuarial Science

Paul Gustafson
External Examiner
Professor, Statistics
University of British Columbia

Abstract

In many real-world randomized controlled trials (RCTs), noncompliance behaviour often occurs and can greatly complicate assessing the intervention efficacy. Furthermore, multiple outcomes are usually employed to measure underlying complex traits when evaluating the performance of multifaceted behaviour interventions for chronic disease (e.g., arthritis). Statistical procedures ignoring treatment noncompliance and the correlations among multiple outcomes can lead to biased estimates of treatment efficacy and a significant loss of power to detect treatment efficacy. This dissertation aims at developing novel statistical methodologies to evaluate the efficacy of multifaceted behaviour interventions while addressing noncompliance issues and correlated multiple outcomes simultaneously.

To deal with noncompliance issues, a principal stratification approach is employed to estimate complier average causal effects. To address the correlated multiple trial outcomes, this dissertation proposed novel methodologies based on mixed-effects regression models and the latent-factor approach. The first work proposes a multivariate longitudinal potential outcome model based on a hierarchical random-effects approach stratified on latent compliance types under all-or-none compliance. The second work proposes a latent-factor multivariate complier average causal effects (MCACE) model for multidimensional longitudinal outcomes with principal strata of compliance types. Under the model, high dimensional outcomes are reduced to low dimensional latent factors, leading to a parsimonious and efficient test of overall CACEs on multiple endpoints, mitigating the multiple testing issues associated with multidimensional endpoints. The third work considers partial compliance and extends to multivariate CACE estimation under the framework of partial compliance. Comprehensive simulation studies demonstrate the validity of the proposal methods and large gains in the estimation efficiency (several-fold increase in statistical power to detect CACEs compared to existing methods). The application of these proposed methodologies to assess the efficacy of a multifaceted behaviour intervention (Arthritis Health Journal) in a longitudinal trial conducted at Arthritis Research Canada yields novel findings not discovered previously.

Keywords: Causal inference; Principal stratification; Potential outcome model; Multi-level model; Treatment effects estimation

Dedication

To my beloved parents and younger brother, for their unconditional love and constant support.

Acknowledgements

Firstly, I would like to express my deep appreciation to my senior supervisor Dr. Hui Xie for his patient guidance and generous support throughout my PhD studies. An important thing I learnt from him is that persistence is a precious quality in academia. During my job search, I would not have had the courage to try many great opportunities without his encouragement and support. He saw my potential and encouraged me never to give up until I tried it. These invaluable qualities would definitely inspire me to explore the unknown in my future endeavours.

I would also like to extend my gratitude to my supervisor Dr. X. Joan Hu. I am deeply grateful for being invited by her to join the group meeting where I made so many friends, received suggestions and comments regarding my research, improved my skills in presenting original work, and had the chance to learn from others' research. I appreciate her support in providing me with the valuable chance to give an invited talk, and precious feedback and suggestions about my work and presentation skills during our discussions.

I am also grateful to my examining committees, Dr. Yi Qian, Dr. Miremad Soleymanian, Dr. Rachel MacKay Altman and Dr. Paul Gustafson for taking the time to review my thesis and attending my defence during their busy schedules. Their constructive questions and insightful comments definitely inspire me to think critically and strengthen my thesis. Many thanks to Dr. Liangliang Wang for chairing my defence. I sincerely appreciate their support in making my thesis defence go so well.

I truly appreciate all faculty members and staff in the Department of Statistics and Actuarial Science for creating a wonderful environment where I greatly enjoyed my academic journey. A warm thank you to my lovely friends and fellow graduate students who brought me joy and support over the years.

Last but not least, special thanks go to Arthritis Research Canada (ARC) where I obtained the data motivating my PhD research, and suggestions about my research and job talk from ARC members. I always enjoyed and cherished the time we spent together at the ARC retreats.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Background

When evaluating causal effects of new interventions, randomized controlled trials (RCTs) are widely used as a gold standard. However, in real-world RCTs, noncompliance behaviour usually occurs since participants may not comply with their initial assignments due to inconvenience or possible side effects. Noncompliance behaviour can greatly challenge assessing the treatment efficacy of the new intervention. There are several approaches commonly employed in real-world RCTs. Intention-to-treat (ITT) analysis compares the outcomes of participants who were assigned to the intervention group with the outcomes of participants assigned to the control group. Thus, ITT analysis focuses on the use/program effectiveness, the effect of treatment assigned. Although the policymakers are interested in the use/program effectiveness, patients and health care providers are more interested in the method effectiveness/intervention efficacy, the effect of treatment received. Use effectiveness depends on patients' compliance behaviour and may change when patients have more knowledge about the intervention later, while method effectiveness is independent of patients' compliance behaviour and informs what can be expected when patients adhere to the treatment strictly. Typically, ITT analysis tends to give conservative estimates of method effectiveness under noncompliance (Sheiner & Rubin 1995). Alternatively, the other commonly used approach, as-treated (AT) analysis, compares outcomes of participants based on the actual receipt of the treatment. Under noncompliance, AT analysis violates the randomization assumption and also provides biased estimates for method effectiveness (Sheiner & Rubin 1995, Hernán & Hernández-Díaz 2012).

To overcome the noncompliance issue and obtain unbiased estimates of intervention efficacy, a complier-average causal effect (CACE) approach has been proposed and deals with the noncompliance issue by analyzing the outcomes of a subgroup consisting of compliers who always comply with their initial assignments (Sommer & Zeger 1991, Angrist et al. 1996, Imbens & Rubin 1997, Hirano et al. 2000). Under CACE analysis, the entire population is divided into four groups: compliers, never-takers, always-takers and defiers. Compliers

always comply with their initial assignments. Never-takers never take the treatment regardless of their assignments. Always-takers always take the treatment no matter which group they are assigned to. Defiers always take the opposite action of their initial assignments. Participants are categorized into these groups based on their compliance behaviour under any possible assignments. The definitions for the four groups are discussed in detail in the following chapters. CACE analysis discusses all-or-none compliance where patients either take the treatment or do not take the treatment. In reality, participants may take a portion of the treatment which is referred to as partial compliance. In addition to discussing the all-or-none compliance, partial compliance is also discussed in this dissertation.

For chronic diseases, multifaceted behaviour interventions are usually developed to help patients manage their diseases. To assess the efficacy of multifaceted tools, multiple outcomes are often employed to measure complex traits of such behaviour interventions. A common approach to dealing with correlated multiple outcomes is to conduct univariate analyses which analyze these outcomes individually. However, ignoring the correlations among multiple outcomes could cause a loss of power to detect the treatment efficacy of the multifaceted tool. Furthermore, conducting hypothesis testing for each outcome individually can lead to multiple testing issues. These limitations raise the other challenge which is how to incorporate correlations among outcome measures across all time points when estimating intervention efficacy. In this dissertation, we aim to develop new methodologies to improve the efficiency of estimating the intervention efficacy while addressing the noncompliance issue and correlated outcome measures simultaneously.

1.2 Motivating Example

1.2.1 Arthritis Health Journal Study

To treat chronic diseases, long-term treatment and management are required. Especially, self-management is critical for patients to achieve clinical remission of chronic diseases. Self-management refers to the ability to participate in various activities, such as complying with treatment regimes, and seeking doctors' help when the target is not met. Although self-management is important when treating chronic diseases, it's hard for patients to keep engaged in self-management activities due to busy daily schedules or fluctuating health conditions. In practice, for rheumatoid arthritis (RA) patients starting RA treatment, the adherence rates for using disease-modifying anti-rheumatic drugs (DMARDs) are only 30% (Van Den Bemt et al. 2012) despite the DMARDs has been proven to be efficient to treat RA (Choi et al. 2002). Due to the difficulty of managing chronic disease efficiently, researchers developed a multifaceted tool to help patients enhance their self-management abilities.

Arthritis Health Journal (AHJ), a multifaceted behaviour intervention, is developed by Dr. Diane Lacaille to improve rheumatoid arthritis (RA) patients' self-management ability. AHJ is an online tool and consists of six components: symptom and exercise log, disease

activity, mood assessment, medical information, goals and action plan and health reports. It is designed to help patients monitor and manage their diseases. Specifically, the AHJ tool benefits RA patients in terms of the "Treat to Target" approach. By actively monitoring their disease and communicating with doctors when the target is not achieved, the treatment plan can be adjusted accordingly by patients themselves or doctors.

To assess the use of AHJ online tool, a longitudinal clinical trial was conducted at Arthritis Research Canada. Researchers recruited participants from the Mary Pack Arthritis Program. Eligible participants must be eighteen years old or older, are diagnosed with rheumatoid arthritis, are currently under the medical care of rheumatologists, and have the ability to get access to the Internet. In the AHJ study, 94 participants were assigned to two groups randomly. 45 patients were in the first group provided with the access to AHJ tool immediately, while 49 participants were assigned to the second group and had to wait for six months before getting access to the tool. In this dissertation, we focused on the datasets collected during the first six months. Therefore, the first group is treated as the treatment group, and the second group is considered as the control group receiving usual care.

In the clinical trial, patients were evaluated by online questionnaires every three months. At baseline, online questionnaires collected information about demographics and disease information.

1. **Disease duration**: having RA for more than two years is recorded as the late disease. Otherwise, the disease duration is reported as early disease.

2. **Disease activity**: high disease activity is recorded if there is high RAPID4 value. low disease activity is observed if there are remission, and moderate/low RAPID4 value.

3. **Gender**: male or female.

4. **Age**: patients are classified into two categories based on whether they are older than or equal to 54.5 years old.

The follow-up questionnaires collected information on six quantities, as well as the frequency of using the tool at each evaluation time. Six quantities are average scores calculated based on several items.

1. **Effective consumer 17 scale**: the average score of 17 items about how participants manage their disease on a 0 to 100 scale with 100 indicating “most confident”.

2. **Manage symptoms scale**: the average score of 5 items about how patients manage their symptoms on a 0 to 10 scale with 10 indicating “totally confident”.

3. **Manage disease in general scale**: the average score of 5 items about how patients manage their disease in general on 0 to 10 scale with 10 indicating “totally confident”.

4. **Communicate with physician scale**: the average score of 3 items about patients' confidence in communicating with their rheumatologists on a 0 to 10 scale with 10 indicating "totally confident".

5. **Partners in health scale**: the average score of 11 items about patients' knowledge of disease and treatment on a 0 to 80 scale with 80 indicating "poor self-management".

6. **Satisfaction with various aspects of medical care**: the average score of 8 items about their satisfaction with various aspects of medical care on a 0 to 10 scale with 10 indicating "completely satisfied".

Six quantities were rescaled on the 0 to 100 scale for better comparisons. When conducting data analysis, we also adjusted the direction of the fifth outcome so that higher values represent better outcomes for all six outcomes.

1.2.2 Research Aims

It's worth noting that six quantities are collected longitudinally to measure complex underlying constructs of the AHJ online tool. Six quantities are designed to jointly evaluate the performance of the tool in terms of enhancing patients' self-management ability. Additionally, these outcomes can also be used to measure several underlying constructs (e.g., self-efficacy in disease management and the effectiveness in shared decision-making). In addition to determining whether the AHJ online tool enhances patients' overall self-management ability, making statistical inference on these underlying constructs is also of great interest.

Furthermore, we are interested in predicting participants' compliance behaviour based on their baseline characteristics. Since the follow-up questionnaires provide information on the frequency of using the online tool at each evaluation time, compliance behaviour can be studied under all-or-none compliance and partial compliance. When handling the noncompliance under all-or-none compliance, dichotomization is employed through a chosen cut-off value to convert continuously-measured compliance to binary compliance. Under the framework of partial compliance, we determine the compliance behaviour by calculating a ratio of the times of using the tool to the number of days during the follow-up period. Details about determining compliance behaviour can be found in the following chapters.

Our goals in this dissertation are summarized as follows:

1. Capturing the correlations among multivariate longitudinal outcomes to improve the estimation efficiency when evaluating the treatment efficacy of the new intervention.

2. Making statistical inference on complex underlying constructs (e.g., self-efficacy and interaction with caregivers).

3. Handling noncompliance behaviour under the framework of principal stratification (all-or-none compliance and partial compliance).

1.3 Literature Review

Noncompliance issue often occurs in randomized controlled trials involving human subjects. Early literature includes work by Imbens & Angrist (1994) who employed the instrumental variable (IV) approach to estimate local average treatment effects (LATEs) and discussed its application in an RCT with noncompliance. Baker & Lindeman (1994) derived a likelihood-based approach to estimate the treatment effect within a subgroup in a paired availability design. Angrist et al. (1996) showed that instrumental variables estimand can be employed to estimate causal effects for compliers under some assumptions within the Rubin Causal Model (Holland 1986) when imperfect compliance is presented. Imbens & Rubin (1997) developed a Bayesian framework to estimate causal effects in RCTs with noncompliance. Under some assumptions, they showed that their approach outperformed standard IV estimand and other existing methods previously in terms of estimating complier-average causal effects.

Following the work by Angrist et al. (1996) and Imbens & Rubin (1997), Hirano et al. (2000) developed an extended framework to study the causal effects in encouragement designs with noncompliance. Their approach allows for considering the pretreatment covariates, and conducting sensitivity analysis when exclusion restrictions are violated. Yau & Little (2001) extended the work by Imbens & Rubin (1997) to estimate complier-average causal effects for longitudinal data with noncompliance and missing data. Additionally, some works focus on developing new methodologies to deal with time-varying noncompliance. Lin et al. (2008) proposed a nested latent class model to deal with time-varying noncompliance by formulating time-invariant superclasses based on longitudinal compliance patterns. Gao et al. (2014) extended the work by Lin et al. (2008) and proposed a Markovian approach to estimate complier-average causal effects at each follow-up time point. However, these works mentioned above are interested in estimating causal effects from univariate outcomes in RCTs with noncompliance.

In real-world RCTs evaluating multifaceted behaviour interventions, multiple outcomes are often employed. In practice, there are some approaches to deal with multiple outcomes. One method to handle multiple outcomes is to estimate causal effects for each outcome individually. However, univariate analysis fails to provide an overall causal effect of the new intervention in an RCT. Moreover, univariate analysis leads to multiple testing issues and a loss of power to detect treatment efficacy due to the ignorance of the correlations among multiple outcomes. The other method is the dimensionality reduction approach. The dimensionality of multivariate outcomes can be reduced by choosing a function (e.g., sum or mean) of responses. Furthermore, principal component analysis (Everitt & Hothorn 2011) can be another way to achieve the reduction of dimensionality. However, principal component analysis may lead to the difficulty of interpretation of the results since one outcome may contribute to two or more components. A more elaborate way to deal with multiple outcomes is conducting factor analysis (Everitt & Hothorn 2011), which leads to a cleaner

interpretation by applying factor rotation techniques. Similar to the factor analysis, Roy & Lin (2000) proposed a latent variable model and introduced a latent variable representing an underlying outcome of major interest characterized by multivariate outcomes measured repeatedly. An et al. (2013) generalized the framework in Roy & Lin (2000) and developed a model allowing for more than one factor.

However, there is limited literature studying causal effects from multivariate outcomes in RCTs with noncompliance. Specifically, most work estimated CACE by jointly considering two outcomes in a cross-sectional setting. Jo & Muthén (2001) jointly modeled two major outcomes to increase the precision of identifying compliance type and the power to detect CACEs on the outcome of primary interest in a cross-sectional setting. Mealli & Pacini (2013) employed a secondary outcome to tighten nonparametric bounds for CACEs on the primary outcome on which the exclusion restriction may not hold. Mattei et al. (2013) also showed that a secondary outcome can improve the inference on the primary outcome under a Bayesian approach in a cross-sectional setting. In this dissertation, we are interested in developing novel methodologies to estimate CACEs from longitudinal multivariate outcomes in RCTs with noncompliance.

Instead of considering binary compliance (all-or-none compliance), partial compliance is common in RCTs where participants may take a portion of the assigned treatment. Jin & Rubin (2008) extended the principal stratification framework (Frangakis & Rubin 2002) to formulate new principal stratum defined on combinations of continuous potential intermediate variables ranging from 0 to 1. Bartolucci & Grilli (2011) generalized the work of Jin & Rubin (2008) by allowing for more flexibility in specifying the conditional distribution of potential outcomes given compliances. Meanwhile, they modeled the drug compliance and placebo compliance by a Plackett copula. However, the recent literature focuses on partial compliance in RCTs in a cross-sectional setting. In the following chapters, we are interested in extending the methods regarding partial compliance mentioned above to a more general RCT setting where multivariate longitudinal outcomes are measured.

1.4 Outline

To assess the treatment efficacy efficiently for longitudinal RCTs where multivariate outcomes and noncompliance are presented, we developed novel approaches in the remainder of the dissertation.

In Chapter 2, we proposed a multivariate longitudinal potential outcome model with stratification on latent compliance types under all-or-none compliance. The model was proposed based on a hierarchical random-effects approach, and assessed an overall complier-average causal effect (CACE) by combining all information among multiple outcomes across all visits. The proposed methodology and its application in the AHJ study in this chapter led to one publication in the journal *Statistics in Medicine* (Guo et al. 2022). In Chapter 3,

a latent-factor multivariate complier-average causal effect (MCACE) model with principal strata of compliance types is developed to study underlying constructs measured by multiple outcomes. In addition to assessing an overall CACE by considering only one latent factor, the proposed model allows us to make statistical inference on more than one underlying constructs through corresponding latent factors. Instead of dealing with all-or-none compliance, chapter 4 extended the hierarchical random-effects approach to a partial compliance setting where a continuously-measured compliance behaviour is considered. Throughout the dissertation, simulation studies are conducted to demonstrate the validity and the improvement in the estimation efficiency of the proposed methodologies. We also applied these novel approaches to the AHJ data and produced some interesting results not discovered previously. Finally, a discussion is given in chapter 5.

Chapter 2

Assessing Complier Average Causal Effects from Longitudinal Trials with Multiple Endpoints and Treatment Noncompliance

2.1 Introduction

Randomized controlled trials (RCTs) are the gold standard for evaluating treatment effects. However, noncompliance often occurs and can greatly complicate assessing treatment effects. Intention-to-treat (ITT) analysis preserves randomization and is the main method to report trial results (Lee et al. 1991, Meier 1991). However, ITT analysis estimates the effect of assigning treatment and targets program effectiveness. While program effectiveness is often of interest to policymakers, patients and their health decision-makers are more interested in intervention efficacy that informs what to expect when patients comply with treatment (Sheiner & Rubin 1995, Steele et al. 2015). Generally, the ITT gives conservative estimates of intervention efficacy (Sheiner & Rubin 1995). The alternative as-treated (AT) approach compares outcomes based on actually received treatments. The AT estimate violates the randomization assumption and can be confounded by unobserved factors correlated with compliance behaviors. Thus AT is subject to selection bias as an estimate of intervention efficacy and should not be used without first evaluating the size of potential bias (Xie & Heitjan 2004).

An appealing alternative is to estimate the complier average causal effect (CACE) via the method of latent class instrumental variables (Baker et al. 2016) that can properly adjust for post-randomization compliance status when estimating treatment effects. Under noncompliance, the principal strata are (partially) unobserved compliance types (compliers, always-takers, never-takers and defiers) determined by the joint potential compliance behaviors under both control and treatment groups (Imbens & Angrist 1994, Baker & Lindeman 1994). These compliance types are predetermined before randomization, which

permits one to define causal effects within subpopulations partitioned by compliance types. Imbens & Angrist (1994) and Baker & Lindeman (1994) showed how to estimate the average treatment effects for the subpopulation who would comply regardless of treatment assignment. Patients and their treatment decision makers are typically interested in knowing the expected treatment effect when taking the treatment. The CACE is considered as more relevant for such patient-oriented treatment effects, as compared with the program effectiveness targeted by ITT (Steele et al. 2015).

In this chapter, we propose a multivariate longitudinal potential outcome model with principal strata for latent compliance types to efficiently assess multivariate CACEs (MCACE) in longitudinal RCTs with multiple endpoints. Real-world RCTs evaluating multifaceted interventions often employ longitudinal designs and multiple endpoints. In such trials, limited sample sizes, low compliance rates or small to moderate effect sizes on endpoints can significantly reduce the power to detect CACE. This work is motivated by a longitudinal RCT conducted at Arthritis Research Canada to evaluate the effectiveness of a behavioral intervention, the Arthritis Health Journal (AHJ) (Lacaille et al. 2015). AHJ is an online tool that enables patients with rheumatoid arthritis to monitor their disease activity. Six health endpoints were collected longitudinally, measuring multifaceted aspects of managing disease, symptoms, knowledge etc. A preliminary evaluation using ITT analysis for each endpoint separately reported no significant treatment effects on all endpoints (Lacaille et al. 2015) with full results presented in Table 2.4 in Supplemental Information (section 2.6). However, a substantial number of participants did not use the intervention, or used it rarely, which can render ITT estimates too conservative for evaluating the patient-oriented intervention effect. The low compliance rate combined with the limited sample size ($n = 94$) and moderate effect size expected from using AHJ motivated us to seek the most efficient analysis to estimate CACEs by combining data across multiple endpoints and visits.

The literature demonstrating the benefits of pooling information across multiple endpoints in RCTs has largely focused on perfect compliance. One exception is Jo & Muthén (2001) who considered CACE estimation with multiple correlated endpoints to increase the power to detect intervention effects for cross-sectional data. Mealli & Pacini (2013) and Mattei et al. (2013) showed a secondary outcome can be exploited to sharpen the nonparametric bound and Bayesian inference of CACE of the primary endpoint in the cross-sectional setting. Yau & Little (2001) and Jo & Muthén (2001) developed methods for CACE estimation with longitudinal measurements of single endpoint subject to noncompliance and attrition, demonstrating the benefits of longitudinal data for CACE estimation, including increased power, better handling of missing data and estimation of growth trends. We extend these prior works to longitudinal RCTs with multiple endpoints and treatment noncompliance.

Our MCACE model consists of a sub-model for the unobserved compliance types and a hierarchical random-effects potential outcome sub-model for longitudinal measurements of multiple endpoints within each compliance type. Unlike univariate CACE (UCACE)

analysis that analyzes each endpoint separately, under MCACE each subject has a single estimate of class membership of compliance type which permits more accurate estimation of the unobserved subject-specific compliance types. By combining data from all endpoints and longitudinal trajectories, MCACE model maximizes the information used to estimate CACE. A global likelihood ratio test is used to test the null hypothesis of no treatment effect on all endpoints. We compared MCACE analysis with UCACE analysis in simulation studies. Results show a significant increase in the estimation efficiency with the MCACE model, including up to 50% reduction in standard errors of CACE estimates and a 1-fold increase in the power to detect CACE. Finally, we apply the proposed MCACE model to the AHJ data. With MCACE model, we detect a significant overall treatment effect using the global likelihood ratio test and identify statistically significant CACEs on two out of six endpoints. In contrast, the less powerful UCACE analysis cannot detect any significant treatment effects.

Next in Section 2.2, we describe the proposed model and its estimation and inference. Section 2.3 describes a simulation study that compares the MCACE method with the UCACE method in terms of the point estimate, nominal rate and width of confidence intervals and power of hypothesis testing. Finally, we apply the proposed MCACE model and UCACE model to the AHJ data in Section 2.4, followed by a discussion in Section 2.5.

2.2 Notation and Model

Let A_i indicate the i^{th} subject's treatment assignment, where $i = 1, \cdots, N$. $A_i = 1$ (or 0) if subject i is assigned to the treatment (or placebo). Let D_i be the indicator of the receipt of the treatment. D_i equals to 1 (or 0) if subject i receives the treatment (or placebo). Let $\boldsymbol{A}$ and $\boldsymbol{D}$ be N-dimensional vectors with the i^{th} elements equal to A_i and D_i. We consider K endpoints measured over time on each of N participants. For CACE analysis, we define two types of potential outcomes, the secondary potential outcome $D_i(\boldsymbol{A})$ and the primary potential outcome $Y_{ijk}(\boldsymbol{A}, \boldsymbol{D(A)})$. $D_i(\boldsymbol{A})$ is the potential treatment received by subject i when subjects are randomized to $\boldsymbol{A}$. $Y_{ijk}(\boldsymbol{A}, \boldsymbol{D(A)})$ is the potential outcome value for the k^{th} outcome at occasion j for subject i under treatment assignment $\boldsymbol{A}$ and treatment received $\boldsymbol{D(A)}$, where $i = 1, \cdots, N,\ j = 0, \cdots, J$ and $k = 1, \cdots, K$. Let $\boldsymbol{Y_i(A, D)}$ be the vector of $K * (J + 1)$ potential outcomes for subject i given $\boldsymbol{A}$ and $\boldsymbol{D(A)}$. Let $\boldsymbol{Y(A, D)}$ denote the vector of potential outcomes collecting all the $Y_{ijk}(\boldsymbol{A}, \boldsymbol{D(A)})$ over i, j, k.

2.2.1 Assumptions for Complier Average Causal Effect Analysis

Our analysis makes two assumptions which are often invoked for causal inference in RCTs. The stable unit treatment value assumption (SUTVA (Rubin 1978, 1980, 1990)) requires no interference between subjects and no multiple versions of treatments. In AHJ study, SUTVA is plausible because the online tool was the same for all participants who independently

accessed and used the online tool with minimum expected interference. SUTVA allows us to write $\boldsymbol{Y_i(A, D)}$ and $D_i(\boldsymbol{A})$ as $\boldsymbol{Y_i}(A_i, D_i)$ and $D_i(A_i)$. The second assumption is random assignment, which assumes that given observed baseline variables, A_i is independent of potential outcomes $\boldsymbol{Y_i(A, D)}$ and $D_i(\boldsymbol{A})$. This assumption is satisfied in RCTs since treatments were randomly assigned to study participants.

In our analysis of CACEs in the AHJ data, we make two additional assumptions. The third assumption is no access to the treatment in the control group. This assumption holds in many placebo-controlled trials, including the AHJ study in which participants in the control group would not have access to the online tool during the 6 months after randomization. Following Imbens & Rubin (1997), we denote C_i, the compliance behavior of participant i, as:

$$C_i = \begin{cases} c \ \text{(complier)}, & \text{if } \ D_i(a) = a, \ \text{ for } \ a = 0, 1, \\ n \ \text{(never-taker)}, & \text{if } \ D_i(a) = 0, \ \text{ for } \ a = 0, 1, \\ a \ \text{(always-taker)}, & \text{if } \ D_i(a) = 1, \ \text{ for } \ a = 0, 1, \\ d \ \text{(defier)}, & \text{if } \ D_i(a) = 1 - a, \ \text{ for } \ a = 0, 1. \end{cases}$$

When held, the third assumption excludes defiers and always-takers. As a result, the compliance status (complier .vs. never-taker) is known for participants in the treatment group but remains unknown for those in the control group. The fourth assumption is exclusion restriction (Imbens & Angrist 1994, Baker & Lindeman 1994). Under the assumption, $\boldsymbol{Y(A, D)} = \boldsymbol{Y(A', D)}$ $\forall \boldsymbol{A}, \boldsymbol{A'}$ and $\forall \boldsymbol{D}$, and thus $\boldsymbol{Y(A, D)}$ can be written as $\boldsymbol{Y(D)}$. In the AHJ study, this implies never-takers' outcomes were the same regardless of treatment assignment. Because compliance behavior C_i is based on the potential outcomes of $D_i(A_i)$, C_i is unaffected by the treatment assignment and thus behaves like a baseline variable. One can perform treatment effect evaluation within each compliance strata if C_i was observed for each participant. However, because C_i is only partially observed, the fourth assumption can be exploited to sharpen the estimation of the compliance-strata-specific treatment effects (Imbens & Rubin 1997). Under the above set of assumptions, CACE is shown to be identifiable with likelihood-based inference (Baker & Lindeman 1994, Imbens & Rubin 1997).

2.2.2 Models for Outcomes and Compliance

We consider modeling the joint distribution of two types of potential outcomes: the potential values of multiple endpoints after being assigned to control and treatment $(\boldsymbol{Y(D(0))}, \boldsymbol{Y(D(1))})$ and the potential treatment received $(\boldsymbol{D(0)}, \boldsymbol{D(1)})$, given the randomization $\boldsymbol{A}$ and covariates $\boldsymbol{W}$. For notation convenience, we denote $\boldsymbol{Y(D(l))}$ by $\boldsymbol{Y}^l$, $\boldsymbol{Y_i(D_i(l))}$ by $\boldsymbol{Y_i}^l$ and $Y_{ijk}(D_i(l))$ by Y_{ijk}^l, $l \in \{0, 1\}$ denoting potential assignment to control and treatment, respectively. Because the potential outcomes $(\boldsymbol{D(0)}, \boldsymbol{D(1)})$ are one-to-one functions of the compliance behavior $\boldsymbol{C}$, we can equivalently model the joint distribution of $(\boldsymbol{Y^0}, \boldsymbol{Y^1}, \boldsymbol{C})$, which is then expressed as the product of the conditional distribution $(\boldsymbol{Y^0}, \boldsymbol{Y^1})$ given the

(latent) compliance type $\boldsymbol{C}$ and the distribution of $\boldsymbol{C}$. When modeling the conditional distribution $(\boldsymbol{Y^0}, \boldsymbol{Y^1})$ given the compliance type $\boldsymbol{C}$, we employ a hierarchical random-effects potential outcome model for longitudinal measurements of multiple endpoints within each compliance type. This hierarchical model has two levels with the first level specifying a within-subjects model for potential outcomes given subject- and endpoint-specific random effects $\boldsymbol{b^l_{mik}}$, and the second level specifying a between-subjects model for $\boldsymbol{b^l_{mik}}$. The distribution of $\boldsymbol{C}$ is specified based on a logistic regression model. To guide model development, we depict the main structure of the proposed MCACE model in Figure 2.1 with nodes to be fully defined in the following two subsections. We first describe the two-level sub-model for potential outcomes $\boldsymbol{Y^l}$ given partially observed compliance type $\boldsymbol{C}$. Then we introduce the sub-model for the compliance type $\boldsymbol{C}$ given baseline covariates $\boldsymbol{W}$.

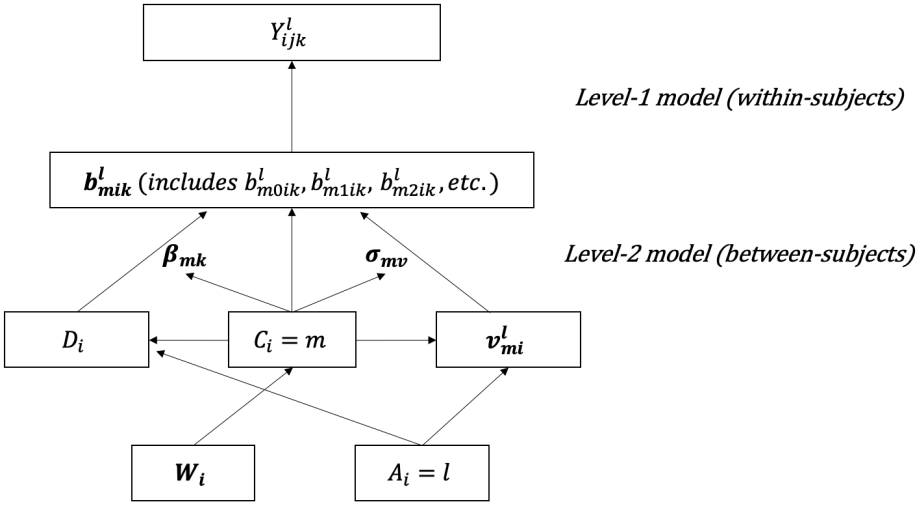

Figure 2.1: Illustration of the structure of the MCACE model.

The sub-model for $\boldsymbol{Y^l} \mid \boldsymbol{C}$

Let m denote the unique value of compliance type, $m \in \{c, n\}$. We assume the following potential outcome model for multiple endpoints within the compliance type m:

$$\boldsymbol{Y^l_i} \mid (C_i = m, \boldsymbol{X_i}) \sim \mathcal{N}(\boldsymbol{\mu^l_{mi}}, \Sigma^l_{mi}) \tag{2.1}$$

where $\boldsymbol{Y^l_i} = (Y^l_{i01}, \cdots, Y^l_{i0K}, Y^l_{i11}, \cdots, Y^l_{i1K}, \cdots, Y^l_{iJ1}, \cdots, Y^l_{iJK})$, and $\boldsymbol{X_i}$ is the vector of explanatory variables for the potential outcomes. The above model assumes $\boldsymbol{Y_i}^0$ and $\boldsymbol{Y_i}^1$ are independent conditional on compliance type and covariates. This is justified because $\boldsymbol{Y_{obs,i}} = A_i\boldsymbol{Y^1_i} + (1-A_i)\boldsymbol{Y^0_i}$ and thus $\boldsymbol{Y_{obs,i}}|C_i, \boldsymbol{X_i} \sim \mathcal{N}(A_i\boldsymbol{\mu^1_{mi}} + (1-A_i)\boldsymbol{\mu^0_{mi}}, A_i\Sigma^1_{mi} + (1-A_i)\Sigma^0_{mi})$, meaning that the likelihood function of the observed data does not depend on the correla-

tion between potential outcomes $\boldsymbol{Y_i^0}$ and $\boldsymbol{Y_i^1}$. Therefore, the correlation between potential outcomes $\boldsymbol{Y_i^0}$ and $\boldsymbol{Y_i^1}$ becomes unimportant under likelihood-based methods (Page 181 in Chapter 8 in Imbens & Rubin (2015), Hirano et al. (2000)). Even with the modeling simplification, for multivariate longitudinal data, it is still very challenging to specify a sensible structure for Σ_{mi}^l. For a general unstructured covariance matrix, Σ_{mi}^l is a $(J+1)K \times (J+1)K$ covariance matrix with $\begin{pmatrix} (J+1)K+1 \\ 2 \end{pmatrix}$ parameters. When $K = 6$ and $J = 2$, the number of unique parameters needed to be estimated in Σ_{mi}^l will be 171. Such parameter proliferation can be a severe issue because the limited sample sizes compounded by the low compliance rates in many practical RCTs lead to an insufficient number of compliers that do not afford enough degree of freedom to estimate a general covariance matrix. To reduce the number of nuisance parameters in Σ_{mi}^l, we employ the hierarchical random-effects modeling approach (Laird & Ware 1982) that captures the potentially complex variance structure by explicitly modeling individual heterogeneity in longitudinal trajectories of multiple outcomes.

The hierarchical random-effects model is also known as the multi-level model. The level-1 part of our multi-level model specifies the following within-subjects model for the potential outcome for the k^{th} endpoint at occasion j for individual i under treatment assignment l, given the participant i's compliance type m and random effects $\boldsymbol{b_{mik}^l}$:

$$Y_{ijk}^l | (C_i = m, \boldsymbol{b_{mik}^l}, \boldsymbol{Z_{ij}}) \quad = \quad \boldsymbol{Z_{ij}}^T \boldsymbol{b_{mik}^l} + \epsilon_{mijk}^l. \tag{2.2}$$

In Eqn 2.2, $\boldsymbol{Z_{ij}}$ contains an intercept and time-varying covariates, such as the time t_{ij} and higher-order terms of t_{ij} to capture potentially non-linear time trends in the potential outcomes. Let $\boldsymbol{\epsilon_{mij}^l} = (\epsilon_{mij1}^l, \cdots, \epsilon_{mijK}^l)^T$ and $\boldsymbol{\epsilon_{mij}^l} \overset{iid}{\sim} \mathcal{N}(\boldsymbol{0}, \Phi_m)$, where $\Phi_m = diag(\sigma_{m1}^2, \cdots, \sigma_{mK}^2)$. We assume $\boldsymbol{\epsilon_{mij}^1}$ is independent of $\boldsymbol{\epsilon_{mij}^0}$. The level-2 model specifies the between-subjects model for the individual-specific random effects $\boldsymbol{b_{mik}^l}$.

$$\boldsymbol{b_{mik}^l} \quad = \quad \boldsymbol{\beta_{m0k}} + \boldsymbol{\beta_{m1k}} D_i(l) + \boldsymbol{v_{mi}^l}, \tag{2.3}$$

where $\boldsymbol{\beta_{m0k}}$ is the vector containing population average regression coefficients for subjects with compliance type m, assigned to treatment l and actually received the control; $\boldsymbol{\beta_{m1k}}$ represents the population average changes in these regression coefficients when these subjects actually received the treatment; $\boldsymbol{v_{mi}^l}$ is the deviation of subject i's coefficients from the population mean. Here we assume that $\boldsymbol{v_{mi}^l}$ is a mean zero Gaussian variable with variance Σ_{mv}. We also assume that $\boldsymbol{v_{mi}^1}$ is independent of $\boldsymbol{v_{mi}^0}$, $\boldsymbol{v_{mi}^l}$ and $\boldsymbol{\epsilon_{mi}^l}$ are independent, where $\boldsymbol{\epsilon_{mi}^l} = (\boldsymbol{\epsilon_{mi0}^l}^T, \cdots, \boldsymbol{\epsilon_{miJ}^l}^T)^T$.

Combining the above two-level models for all k endpoints at all time points, we can obtain one overall model for the potential outcomes for individual i with compliance type

m as

$$\boldsymbol{Y_i^l}|(C_i = m, \boldsymbol{v_{mi}}, \boldsymbol{X_i}) \quad = \quad (X_{i,l} \otimes I_K)\boldsymbol{\beta_m} + \left(Z_i \boldsymbol{v_{mi}^l}\right) \otimes 1_K + \boldsymbol{\epsilon_{mi}^l},$$

where $\boldsymbol{Y_i^l} = \{Y_{ijk}^l : j = 0, \cdots, J; k = 1, \cdots, K\}$, $\boldsymbol{\beta_m} = \{\beta_{mpqk} : p = 0, \cdots, P; q = 0, \cdots, Q; k = 1, \cdots, K\}$, where P and Q depend on the forms of Eqns 2.2 and 2.3: $(P+1)$ equals the dimension of random effects in level-1 model; Q equals the number of predictors in level-2 model and could be greater than 1 if more predictors are included in the level-2 model. $X_{i,l}$ and Z_i are design matrices for fixed effects and random effects respectively. Besides, $X_{i,l}$ is a $(J+1)$ by R matrix where R equals the number of fixed effects coefficients in Eqn 2.3. Z_i is a $(J+1)$ by H matrix where H equals the dimension of random effects in Eqn 2.3.

By combining random effects and residual error terms, we obtain the marginal distribution for $\{\boldsymbol{Y_i^l}|C_i = m\}$ in Eqn 2.1 as

$$\boldsymbol{Y_i}^l|C_i = m, \boldsymbol{X_i} \sim MVN_{\boldsymbol{\beta_m}, \boldsymbol{\psi_m}}(\boldsymbol{\mu_{mi}^l}, \Sigma_{mi}), \tag{2.4}$$

where $\boldsymbol{\mu_{mi}^l} = (X_{i,l} \otimes I_K)\boldsymbol{\beta_m}$, $\Sigma_{mi} = (Z_i \Sigma_{mv} Z_i^T) \otimes (1_K 1_K^T) + V_m$; $V_m = var(\boldsymbol{\epsilon_{mi}^l}) = diag(\Phi_{m0}, \Phi_{m1}, \cdots, \Phi_{mJ})$, $\Phi_{m0} = \Phi_{m1} = \cdots = \Phi_{mJ} = \Phi_m$; $\boldsymbol{\psi_m} = (\boldsymbol{\sigma_{mv}}^T, \sigma_{m1}^2, \cdots, \sigma_{mK}^2)^T$, $\boldsymbol{\sigma_{mv}}$ is the vector of unique parameters in Σ_{mv}, the variance-covariance matrix of random effects $\boldsymbol{v_{mi}^l}$.

An illustrative example For illustration purpose, consider the following level-1 model with a quadratic function of time since baseline:

$$Y_{ijk}^l|(C_i = m, \boldsymbol{b_{mik}^l}, Z_{ij}) = b_{m0ik}^l + b_{m1ik}^l t_{ij} + b_{m2ik}^l t_{ij}^2 + \epsilon_{mijk}^l, \tag{2.5}$$

with the following level-2 (between-subjects) models are as below,

$$\begin{aligned} b_{m0ik}^l &= \beta_{m00k} + \beta_{m01k} D_i(l) + v_{m0i}^l, \\ b_{m1ik}^l &= \beta_{m10k} + \beta_{m11k} D_i(l) + v_{m1i}^l, \\ b_{m2ik}^l &= \beta_{m20k} + \beta_{m21k} D_i(l) + v_{m2i}^l, \end{aligned} \tag{2.6}$$

where $\boldsymbol{v_{mi}^l} = (v_{m0i}^l, v_{m1i}^l, v_{m2i}^l)^T \overset{iid}{\sim} \mathcal{N}(\mathbf{0}, \Sigma_{mv})$, $\quad \Sigma_{mv} = \begin{bmatrix} \sigma_{v_{m0}}^2 & \sigma_{v_{m0}v_{m1}} & \sigma_{v_{m0}v_{m2}} \\ \sigma_{v_{m0}v_{m1}} & \sigma_{v_{m1}}^2 & \sigma_{v_{m1}v_{m2}} \\ \sigma_{v_{m0}v_{m2}} & \sigma_{v_{m1}v_{m2}} & \sigma_{v_{m2}}^2 \end{bmatrix}$.

The matrix form of the model for the k^{th} outcome of individual i is

$$\boldsymbol{Y_{ik}^l} \mid (C_i = m) = X_{i,l} * \boldsymbol{\beta_{mk}} + Z_i * \boldsymbol{v_{mi}^l} + \boldsymbol{\epsilon_{mik}^l}, \tag{2.7}$$

where $\boldsymbol{y_{ik}^l} = (y_{i0k}^l, y_{i1k}^l, \cdots, y_{iJk}^l)^T$, $\boldsymbol{\beta_{mk}} = (\beta_{m00k}, \beta_{m10k}, \beta_{m20k}, \beta_{m01k}, \beta_{m11k}, \beta_{m21k})^T$, $\boldsymbol{v_{mi}^l} = (v_{m0i}^l, v_{m1i}^l, v_{m2i}^l)^T$, $\boldsymbol{\epsilon_{mik}^l} = (\epsilon_{mi0k}^l, \epsilon_{mi1k}^l, \cdots, \epsilon_{miJk}^l)^T$. Specifically,

$$X_{i,l} = \begin{pmatrix} 1 & 0 & 0 & D_i(l) & 0 & 0 \\ 1 & t_{i1} & t_{i1}^2 & D_i(l) & t_{i1} * D_i(l) & t_{i1}^2 * D_i(l) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & t_{iJ} & t_{iJ}^2 & D_i(l) & t_{iJ} * D_i(l) & t_{iJ}^2 * D_i(l) \end{pmatrix}, \quad Z_i = Z = \begin{pmatrix} 1 & 0 & 0 \\ 1 & t_{i1} & t_{i1}^2 \\ \vdots & \vdots & \vdots \\ 1 & t_{iJ} & t_{iJ}^2 \end{pmatrix}.$$

When the errors ϵ_{mijk}^l are independent of each other over i, j, k, the correlations among endpoints at the same or different times are induced by the shared random effects $\boldsymbol{v_{mi}^l}$. Correlations among longitudinal repeated measurements for the same endpoint are induced by the random effects $\boldsymbol{v_{mi}^l}$ as shown in Eqn 2.7. In Eqn 2.6, $\boldsymbol{b_{mik}^l}$ for different endpoints contains the same $\boldsymbol{v_{mi}^l}$, which induces the correlations among different endpoints. The correlations for all endpoints at all time points within the same subject induced by the random effects $\boldsymbol{v_{mi}^l}$ could also be seen by examining the specific form of Σ_{mi} in Eqn 2.4, in which $\Sigma_{mi} = (Z_i \Sigma_{mv} Z_i^T) \otimes (1_K 1_K^T) + V_m$. A combination of flexible specifications of design matrix Z_i and variance-covariance matrix Σ_{mv} for random-effects $\boldsymbol{v_{mi}^l}$ and V_m for residuals can generate complex forms of Σ_{mi}. For example, variance-covariance matrix and correlation matrix of repeated measures are allowed to differ across endpoints in Σ_{mi}.

Based on the above $X_{i,l}$ and Z_i, we can obtain the marginal distribution using Eqn 2.4. Compared with specifying a general variance matrix for Σ_{mi}, this model reduces the number of nuisance parameters from $\begin{pmatrix} (J+1)K+1 \\ 2 \end{pmatrix}$ down to $K+6$. If $J = 2$ and $K = 6$, then the number of nuisance covariance parameters dropped from 171 to 12.

Principal causal effects Principal causal effects (PCE) are defined as the ITT effects of the treatment within subpopulations defined by compliance type. In many RCTs including the AHJ study, subjects were followed up at equal time intervals ($t_{ij} = t_j$). Our interest is the visit-specific PCEs for compliers, which are:

$$ITT_{c,jk} = E(Y_{ijk}^1 - Y_{ijk}^0 | C_i = c) \tag{2.8}$$

For the above illustrative example with the level-1 and level-2 model specified in Eqns 2.5 and 2.6, respectively, we have $y_{ijk}^l|(C_i = m) = \beta_{m00k} + \beta_{m10k} t_j + \beta_{m20k} t_j^2 + \beta_{m01k} D_i(l) + \beta_{m11k} D_i(l) t_j + \beta_{m21k} D_i(l) t_j^2 + v_{m0i}^l + v_{m1i}^l t_j + v_{m2i}^l t_j^2 + \epsilon_{mijk}^l$. For participants, the mean response for the k^{th} outcome at visit j is:

$$\begin{aligned} E(y_{ijk}^1 | C_i = c) &= \beta_{c00k} + \beta_{c10k} t_j + \beta_{c20k} t_j^2 + \beta_{c01k} + \beta_{c11k} t_j + \beta_{c21k} t_j^2, \\ E(y_{ijk}^0 | C_i = c) &= \beta_{c00k} + \beta_{c10k} t_j + \beta_{c20k} t_j^2, \\ E(y_{ijk}^1 | C_i = n) &= E(y_{ijk}^0 | C_i = n) = \beta_{n00k} + \beta_{n10k} t_j + \beta_{n20k} t_j^2. \end{aligned}$$

Therefore, $ITT_{c,jk} = E(y^1_{ijk}|C_i = c) - E(y^0_{ijk}|C_i = c) = \beta_{c01k} + \beta_{c11k}t_j + \beta_{c21k}t_j^2$, $ITT_{n,jk} = E(y^1_{ijk}|C_i = n) - E(y^0_{ijk}|C_i = n) = 0$. $ITT_{c,jk}$ and $ITT_{n,jk}$ estimate the effect of treatment assignment for compliers and never-takers respectively. For compliers, the treatment received is the same as the treatment assigned. Thus, $ITT_{c,jk}$ also estimates the complier average causal effect of the treatment received. On the other hand, $ITT_{n,jk}$ compares potential outcomes which result from actually receiving the control regardless of treatment assignment. Thus $ITT_{n,jk}$ always equals zero under the exclusion restriction assumption. Under the randomization assumption, there should be no difference in outcomes at baseline for compliers between two groups. Thus, we expect that β_{c01k} be zero, and β_{c11k} and β_{c21k} jointly determine the CACE for the k^{th} outcome. The CACE is null for the k^{th} outcome if both β_{c11k} and β_{c21k} equal zero.

Model for compliance status C

Given the baseline covariates $\boldsymbol{W}$, we can model the probability of being a complier using a logistic regression model as

$$p_{ci} = Pr(C_i = c|\boldsymbol{W_i} = \boldsymbol{w_i}, \boldsymbol{\alpha}) = \frac{exp(\boldsymbol{w_i}'\boldsymbol{\alpha})}{1 + exp(\boldsymbol{w_i}'\boldsymbol{\alpha})}. \tag{2.9}$$

As noted before, compliance status (complier or never-taker) is observed for participants assigned to the treatment group, but is unobserved for those in the control group. Therefore, the compliance model Eqn 2.9 can not be estimated directly on the entire sample.

Finally, when $K = 1$, the multivariate CACE (MCACE) model proposed above reduces to the univariate CACE (UCACE) model, which is akin to Yau & Little (2001). One can apply UCACE to analyze each endpoint separately. Unlike MCACE, the UCACE method ignores and does not pool the information from multiple correlated endpoints to sharpen the estimation of compliance status and CACE.

2.2.3 Estimation and Inference

Let $\boldsymbol{Y_{obs,i}} = a_i\boldsymbol{Y_i^1} + (1 - a_i)\boldsymbol{Y_i^0}$, $d_i = D_{obs,i} = D_i(a_i)$. We denote $\boldsymbol{Y_{obs}}$ as the vector of observed outcomes collecting all the $\boldsymbol{Y_{obs,i}}$ over i and $\boldsymbol{D_{obs}}$ as a $N \times 1$ vector with the i^{th} element equal to d_i. In our case, there are three combinations for (a_i, d_i): (1,1), (1,0) and (0,0). We use $\mathcal{S}(1,1)$, $\mathcal{S}(1,0)$ and $\mathcal{S}(0,0)$ to indicate the subsets of units exhibiting each pattern of (a_i, d_i). In our case where the population only includes compliers and never-takers, $\mathcal{S}(1,1)$ and $\mathcal{S}(1,0)$ include the compliers and never-takers, respectively, in the treatment group and $\mathcal{S}(0,0)$ represents a mixture of compliers and never takers in the control group. Let $\boldsymbol{\pi} = (\boldsymbol{\beta_c}, \boldsymbol{\beta_n}, \boldsymbol{\psi_c}, \boldsymbol{\psi_n}, \boldsymbol{\alpha})$ denote the vector collecting all model parameters and $\boldsymbol{X}$ denote the vector collecting all $\boldsymbol{X_i}$ over i, the likelihood function based on observed data for all

participants in the study is

$$\mathcal{L}(\boldsymbol{\pi}; \boldsymbol{Y_{obs}}, \boldsymbol{D_{obs}}, \boldsymbol{A_{obs}} \mid \boldsymbol{W}, \boldsymbol{X}) = L_{11} \times L_{10} \times L_{00}$$

where

$$\begin{aligned}
L_{11} &= \prod_{\{i \in S(1,1)\}} p_{ci} \frac{1}{(2\pi)^{\frac{(J+1)K}{2}} |\Sigma_c|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}\left(y_{obs,i} - \mu^1_{ci}\right)^T \Sigma_c^{-1} \left(y_{obs,i} - \mu^1_{c_i}\right)\right\}, \\
L_{10} &= \prod_{\{i \in S(1,0)\}} (1 - p_{ci}) \frac{1}{(2\pi)^{\frac{(J+1)K}{2}} |\Sigma_n|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}\left(y_{obs,i} - \mu^1_{ni}\right)^T \Sigma_n^{-1} \left(y_{obs,i} - \mu^1_{ni}\right)\right\}, \\
L_{00} &= \prod_{\{i \in S(0,0)\}} [p_{ci} \frac{1}{(2\pi)^{\frac{(J+1)K}{2}} |\Sigma_c|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}\left(y_{obs,i} - \mu^0_{c_i}\right)^T \Sigma_c^{-1} \left(y_{obs,i} - \mu^0_{c_i}\right)\right\} \\
&\quad + (1 - p_{ci}) \frac{1}{(2\pi)^{\frac{(J+1)K}{2}} |\Sigma_n|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}\left(y_{obs,i} - \mu^0_{ni}\right)^T \Sigma_n^{-1} \left(y_{obs,i} - \mu^0_{ni}\right)\right\}].
\end{aligned}$$

The maximum likelihood estimates (MLEs) of model parameters can be obtained by maximizing the observed data log-likelihood function via the Quasi-Newton algorithm for function optimization. The starting values are chosen based on the result from univariate ITT analysis conducted for six outcomes separately. Multiple different starting values are tried to ensure the algorithm converges to the same result. The variance of the estimates can be obtained via the inverse Hessian matrix of the log-likelihood function evaluated at the MLEs. The estimates of PCEs in Eqn 2.8 can be obtained by plugging in the MLE model parameter estimates with the standard errors of PCE estimates obtained using the delta method.

2.3 Simulation Study

We compare the MCACE model proposed above with the alternative approach of fitting separate univariate CACE (UCACE) models in their performance of treatment effect estimation and inference with longitudinal observations of multiple study endpoints. Our comparison evaluates the consistency and variability of CACE estimates, width and coverage rate of confidence intervals as well as the power of hypothesis testing for CACE.

2.3.1 Description of Data Generation

We simulated data from the MCACE model as specified in Eqns 2.5, 2.6 and 2.9 with six endpoints ($K = 6$) at three time points ($J = 2$) for n individuals, where $n = 100, 200$ or 500. To simplify the simulation setting, the model is set as a random-intercept and fixed-slope model, which means $b^l_{m2ik} = 0$ in Eqn 2.5, $v_{c1i} = 0$ in Eqn 2.6 and the level-2 model only include the first two rows in Eqn 2.6. We further set $\beta_{c01k} = 0$ (i.e., no baseline

difference between two randomized arms). The compliance status model (Eqn 2.9) includes only intercept, i.e., $p_{ci} = p_c$.

Under the above linear trend model, β_{c11k} informs the complier average causal effect on the k^{th} endpoint, which is of primary interest. Recall $\boldsymbol{\beta_{c11}} = (\beta_{c111}, \cdots, \beta_{c116})^T$. For ease in result presentation, we set $\boldsymbol{\beta_{c11}}$ as a vector of the same value in the simulation. When evaluating the consistency and efficiency of the MLEs and confidence intervals of $\boldsymbol{\beta_{c11}}$, we choose the common value as 0, 1.5 and 3. However, the treatment effect size for each endpoint also depends on the variance of the endpoint, thus differs across endpoints because variances vary over endpoints (section 2.6.1 in Supplemental Information (section 2.6)). True values of other parameters in the multi-level MCACE model are informed by the AHJ data and can be found in section 2.6.1 in Supplemental Information (section 2.6). Last, we simulated compliance status from a Bernoulli distribution with $p_c = 0.3$. A detailed description of the data generating process can be found in section 2.6.1 in Supplemental Information (section 2.6).

2.3.2 Simulation Results

Because the complier average causal effect is of our primary interest, we focus on the results for the CACE effects captured by $\{\beta_{c11k}\}$.

Point estimate

Figure 2.2 shows the sample means, sample standard deviations of estimates and asymptotic standard error estimates for β_{c11k}'s from fitting the MCACE model (red lines) and multiple UCACE models (green lines) under different sample sizes (listed as column headings) and different effect sizes (listed as row headings). The dashed lines indicate the true values in different settings. These results are obtained based on 500 repetitions. Table 2.5 in Supplemental Information (section 2.6) presents the same result in a tabular format.

For both MCACE and multiple UCACE models, the sample means of estimates are close to the corresponding true values. This verifies the consistency of MLEs. Figure 2.2 also shows that the sample standard deviations of the estimates from the MCACE model are almost half of that yielded by multiple UCACE models. Therefore, we conclude that the MCACE model significantly improves the estimation efficiency compared with UCACE analyses. Furthermore, the means of standard error estimates (red narrow bars) produced by the Fisher information are almost identical to the true standard deviations (red wide bars) of the MCACE estimates. When performing the UCACE analysis of each endpoint separately, most means of standard error estimates (green narrow bars) produced by the Fisher information are noticeably smaller than their true values (green wide bars) when the sample size equals 100. However, as sample size increases, the means of standard error estimates produced by the Fisher information become closer to their true values. This is

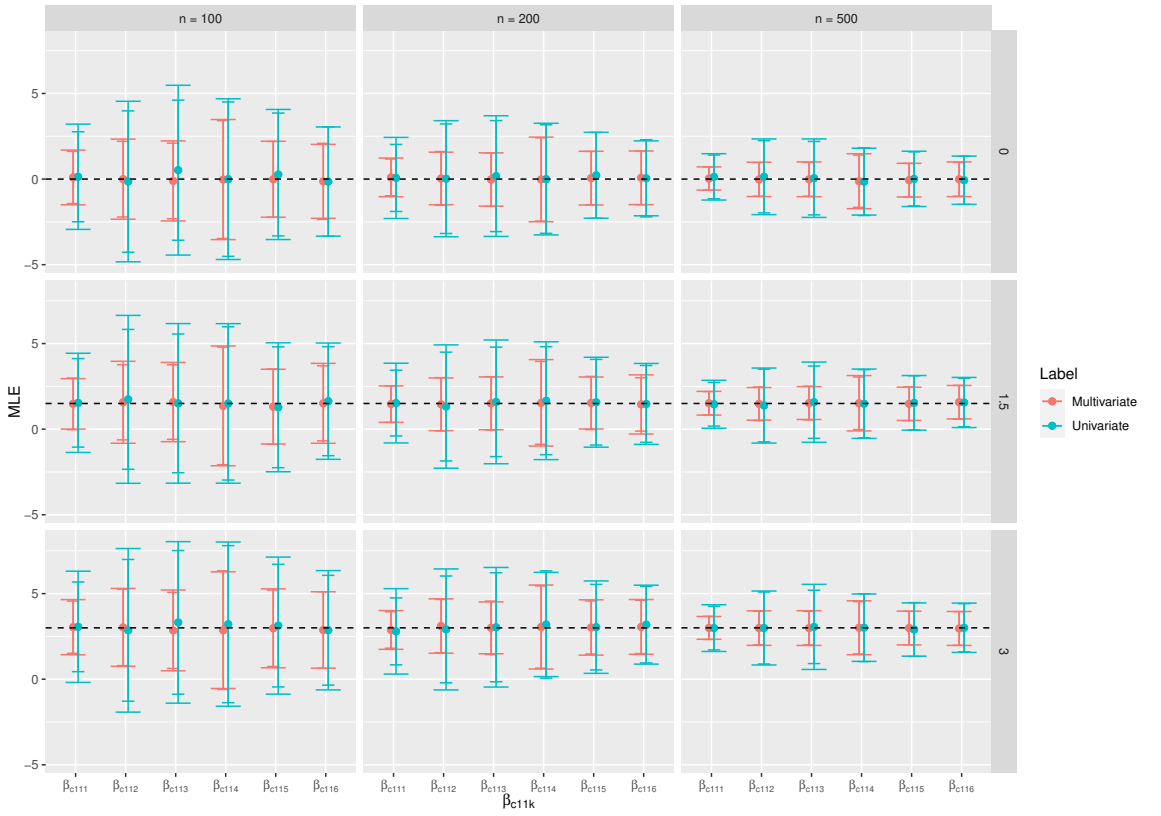

Figure 2.2: Means ± standard deviations (wide bar) of MCACE and UCACE estimates of β_{c11k} as well as ± mean of standard error estimates (narrow bar) computed from the Fisher information matrix of the MCACE model and UCACE model over 500 replications.

expected because the standard error estimator produced by the Fisher information approximates the true standard deviation well when the sample size is large enough, and may perform poorly when the sample size is small. We do not observe inaccurate estimation of standard errors from the MCACE model because the MCACE model analyzes six outcomes simultaneously and gets estimates based on larger datasets.

Confidence interval

In the MCACE model, under appropriate regularity conditions, the MLE of $\boldsymbol{\beta_{c11}}$ has asymptotic normality, for $n \to \infty$, $\sqrt{n}(\hat{\boldsymbol{\beta}}_{\boldsymbol{c11}} - \boldsymbol{\beta_{c11}}) \xrightarrow{d} MVN(0, [I(\boldsymbol{\beta}_{c11})]^{-1})$, where $I(\boldsymbol{\beta_{c11}})$ is the Fisher information. Based on the asymptotic normality, we are able to calculate simultaneous confidence intervals and use Bonferroni correction to ensure that the probability of all confidence intervals containing their true values is no less than $1-\alpha$. For β_{c11k}, we could get a confidence interval as $t_k'\hat{\boldsymbol{\beta}}_{\boldsymbol{c11}} \pm c\sqrt{t_k'([I_n(\hat{\boldsymbol{\beta}}_{\boldsymbol{c11}})]^{-1})t_k}$, where t_k is a six-dimensional column vector with the k^{th} element being 1 and all other elements being 0, and $I_n(\hat{\boldsymbol{\beta}}_{\boldsymbol{c11}})$ is Fisher information for n independent units and is estimated by the inverse Hessian matrix of the log-likelihood for the sample of n units. When conducting UCACE analysis, the confidence

intervals are constructed as $\hat{\beta}_{c11k} \pm c\sqrt{1/I_n(\hat{\beta}_{c11k})}$, where $\hat{\beta}_{c11k}$ and $I_n(\hat{\beta}_{c11k})$ are obtained from UCACE analysis on the kth endpoint only and are generally different from those calculated from MCACE. For both MCACE model and UCACE model, c is the critical value and is set as $Z_{1-\alpha/(2k)}$ based on Bonferroni correction.

Figure 2.3.a shows the distribution of the length of 95% confidence intervals based on 500 simulated datasets. For both MCACE model and UCACE models, the average length of the 95% confidence intervals decreases and the distributions of the length become more concentrated as sample size increases. With the same sample size, the confidence intervals from MCACE model are shorter than those from UCACE models.

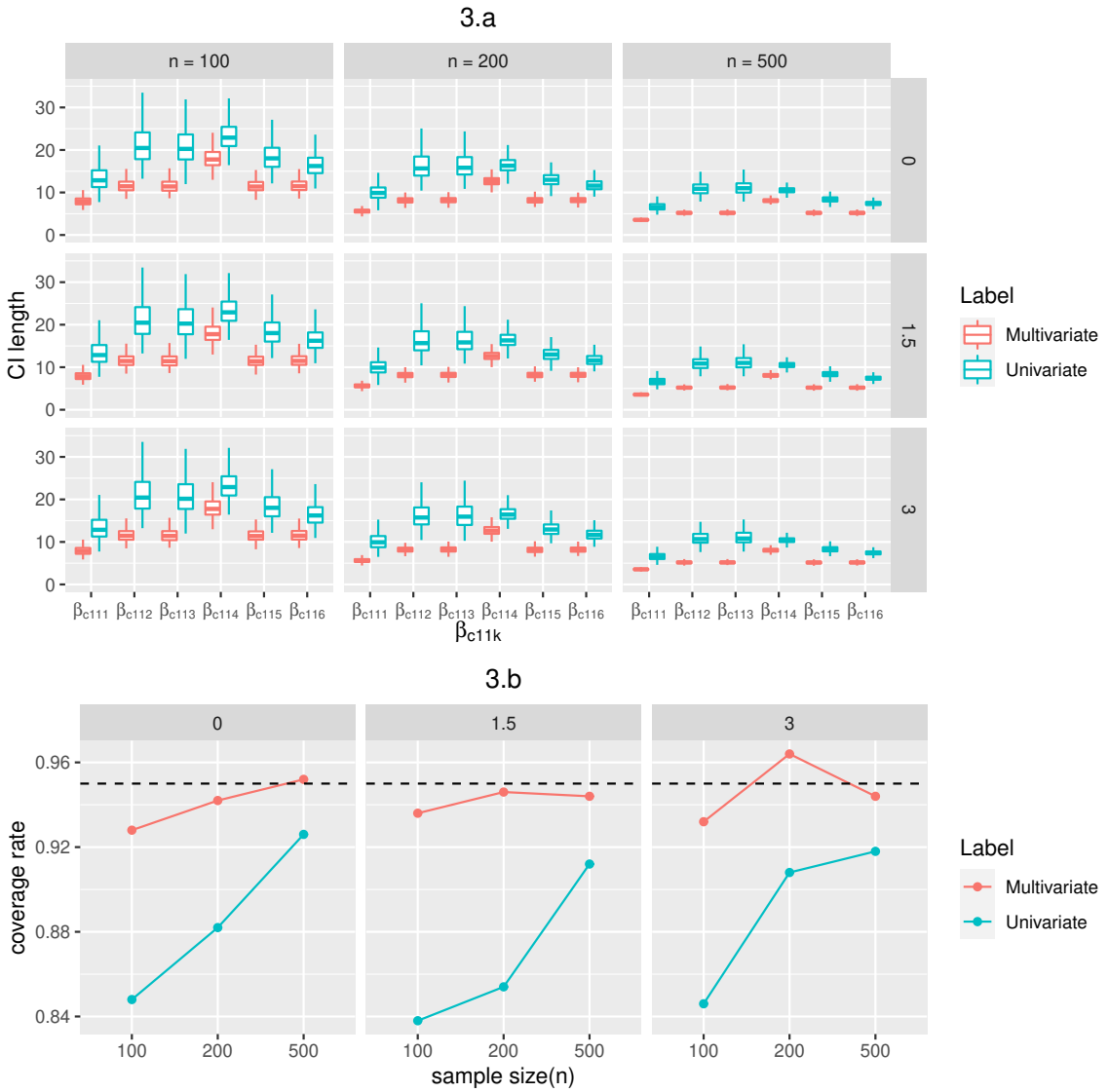

Figure 2.3: Plot 2.3.a describes the distribution of the length of 95% confidence intervals; plot 2.3.b shows coverage rate of 95% confidence intervals, and the dashed line corresponds to the value of 0.95. The results from both figures are based on 500 repetitions.

Figure 2.3.b plots the coverage rates of confidence intervals from the MCACE model and multiple UCACE models. The coverage rate is calculated as the proportion of times that all six confidence intervals include their corresponding true values simultaneously. We observe that the coverage rates calculated from the MCACE model are higher than those from the multiple UCACE models and are closer to the nominal 95% rate. However, as sample size increases, the coverage rates from UCACE models improves and get closer to the 95% nominal rate. Similar with the point estimate, the improvement in UCACE performance as sample size increases can be explained by the asymptotic property of the MLE which requires large sample sizes for MLEs to perform well. MCACE model makes inference by analyzing six outcomes at one time, which makes use of more information from a larger number of observations. Thus the coverage rates from MCACE are closer to the nominal 95% rate and perform well for the number of subjects as small as 100.

Statistical power

We also compare the statistical power of the MCACE model with that of multiple UCACE models. Power is the probability of rejecting the null hypothesis when the null hypothesis is false. When conducting power analysis for MCACE model, we consider the global null hypothesis as $H_0 : \beta_{c11k} = 0$ for all k, and calculate the proportion of times of rejecting the null hypothesis among 500 simulated data sets. The likelihood ratio statistic is

$$\lambda = -2(l_{reduced}|_{\hat{\boldsymbol{\pi}}_{\boldsymbol{r}}} - l_{full}|_{\hat{\boldsymbol{\pi}}_{\boldsymbol{f}}}),$$

where l is the log-likelihood, $\hat{\boldsymbol{\pi}}_{\boldsymbol{r}}$ and $\hat{\boldsymbol{\pi}}_{\boldsymbol{f}}$ are the MLEs of model parameters obtained from the reduced model and the full model, respectively. The full model consists of all parameters and reduced model sets $\beta_{c11k} = 0$ for all k. Under H_0, the test statistic λ follows asymptotically a chi-square distribution with a degree of freedom K. Because UCACE model analyzes each endpoint separately, UCACE analysis does not offer a single global test for $H_0 : \beta_{c11k} = 0$ for all k, as MCACE analysis does. Thus, when conducting the power analysis of multiple UCACE models, we analyze and conduct a likelihood-ratio test of zero complier average causal effects for each endpoint separately with Bonferroni's adjustment for multiple tests to control the overall Type-I error rate at 0.05, which means if any of these K hypotheses is rejected at the significance level $0.05/6 \approx 0.008$, we conclude the CACE is present for at least one endpoint. Because of high efficiency of MCACE estimation and conservativeness of Bonferonni's adjustment, we expect the Bonferroni's adjustment for multiple testing employed in UCACE analysis can have substantially inflated Type-II error rates (i.e., low power), as compared with the global likelihood test available in MCACE analysis.

Figure 2.4 plots the power curves under different sample sizes. To ease the result presentation, we set the values of β_{c11k} to be the same across k and vary from 0 to 15 when simulating data. When β_{c11k}'s values are equal to 0, the value of the power function is the

Type I error rate. We observe that Type I errors for MCACE from the three plots are around 0.05, which means the global likelihood ratio test provided by MCACE controls Type I error rate well in our simulation settings. It's worth noticing that the Type I error under UCACE models when sample size equals 100 equals 0.06, which is a little higher than 0.05. This is consistent with our earlier results where standard error estimator tends to give inaccurate estimates under UCACE models when sample size is small. As the sample size increases, the power reaches 1 for both the MCACE model and multiple UCACE models. However, the power curves of the MCACE model consistently have a steeper slope than the corresponding power curves of multiple UCACE models and reach to 1 sooner as the effect size increases. The increase in study power can be substantial. For example, when $\beta_{c11k} = 5$ and $n = 100$, the power can be increased from 0.46 when using multiple UCACE analysis to 0.90 when conducting the MCACE analysis. Thus, MCACE model can lead to a 100% increase in the power to reject the null compared with the separate UCACE analysis.

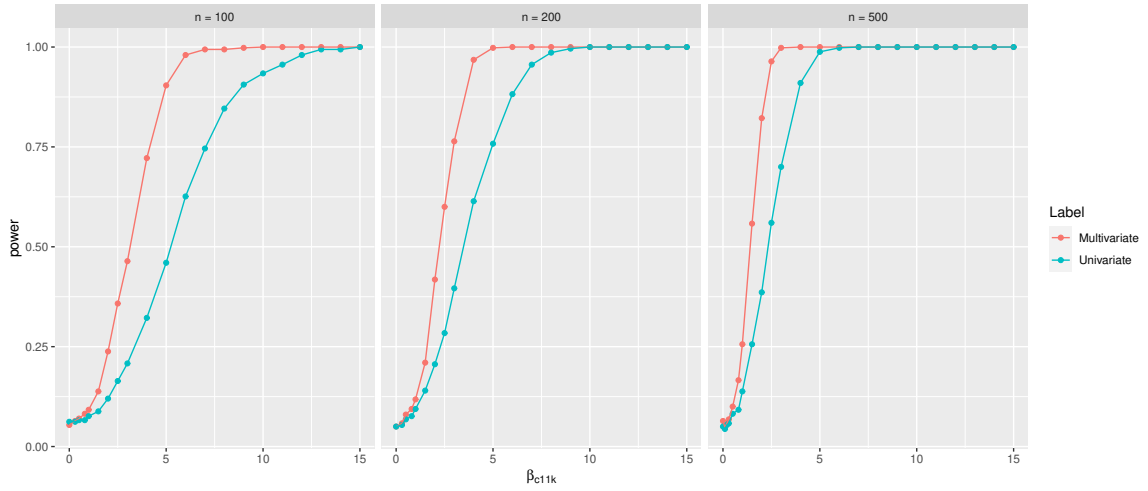

Figure 2.4: Power analysis, based on 500 simulated datasets.

Overall, the simulation results demonstrate that MCACE model outperforms multiple UCACE models in terms of the efficiency of point estimates for CACE, nominal rate and width of confidence intervals and the power of hypothesis testing.

2.4 Application

2.4.1 Study Description and Preliminary Analysis

In this section, we apply the proposed model to estimate the CACE of Arthritis Health Journal (AHJ). The study is a randomized controlled trial comparing the AHJ with the usual care in managing rheumatoid arthritis (RA). AHJ is a patient-centered online tool to help patients track symptoms, monitor disease activity and develop action plans (Lacaille et al. 2015). By helping RA patients better monitor their disease activity, this tool aims

to facilitate the treat to target approach by providing early signs when the disease is not controlled.

A total of 94 patients were randomly assigned to two groups. Patients in the first group ($n = 45$) were provided with online access to AHJ (the intervention) immediately; patients in the second group ($n = 49$) received usual care (control) for 6 months at which time point they were provided with online access to AHJ. We illustrate the proposed methodology using 6-month data of the study, during which period the second group served as the control to the intervention. When they began the intervention, participants were provided with online access to the AHJ and were asked to use it for 6 months. They were evaluated every three months using a self-administered questionnaire. The baseline questionnaires collected information about the demographics and disease information. The follow-up questionnaires evaluated the frequency of using the tool, satisfaction with care, self-management, consumer effectiveness and health status. The study has the following 6 endpoints on which to evaluate the treatment effects of using AHJ : **effective consumer 17 scale**, the overall score of questions about how patients manage their disease on a 0 to 100 scale with 100 indicating "most confident"; **manage symptoms scale**, the overall score of questions about how patients manage their symptoms on a 0 to 10 scale with 10 indicates "totally confident"; **manage disease in general scale**, the overall score of questions about how patients manage their disease in general on 0 to 10 scale with 10 indicates "totally confident"; **communicate with physician scale**, the overall score of patients' confidence in communicating with their rheumatologists on a 0 to 10 scale with 10 indicates "totally confident"; **partners in health scale**, the overall score of patients' knowledge of disease and treatment on a 0 to 80 scale with 80 indicates "poor self-management"; **satisfaction with various aspects of medical care**, the overall score of their satisfaction with the content and format of the tool on a 0 to 10 scale with 10 indicates "completely satisfied". Because these six endpoints are of different scales, we rescaled them all on the 0 to 100 scale. For the fifth endpoint, a higher value represents a worse outcome. We thus redefine the endpoint as 100 minus the original value so that a higher value represents a better outcome for all endpoints.

Figure 2.5 plots the means and standard errors of means for six endpoints by treatment arm and visit. We observe that the AHJ intervention group (green lines) had comparable baseline values for all endpoints as the control group (red lines) and that the AHJ group appeared to have higher average values than the control group consistently for all endpoints at the two post-intervention visits. However, the standard error bars are wide. Analyzing these endpoints separately showed none of the group differences at the sixth month was statistically significant (full results are available in Table 2.4 in Supplemental Information (Lacaille et al. 2015)). This suggests examining multiple endpoints simultaneously in order to pool similar treatment effects across endpoints to increase study power. Furthermore, many in the intervention group rarely used the AHJ for a variety of reasons. Consequently,

the program effectiveness estimated by the ITT analysis can be substantially smaller than the treatment efficacy, the latter of which is often of more interest to patients and caregivers. Figure 2.5 also plots the raw statistics for the compliers in the treatment group. These compliers consist of patients who used the AHJ at least one time per month on average within six months after randomized to the intervention group. We observe that the upward trends in endpoint measurements for compliers in the treatment group (blue lines) appear to be larger than those for the overall treatment group, especially for the fourth, fifth and sixth endpoints. Overall, the moderate sample size, low compliance rate and moderate beneficial treatment effect sizes across multiple endpoints motivated us to perform a multivariate longitudinal analysis of treatment efficacy. Such analysis aims to maximize the power to detect the overall treatment efficacy by pooling CACE estimation across all endpoints over longitudinal measurement occasions.

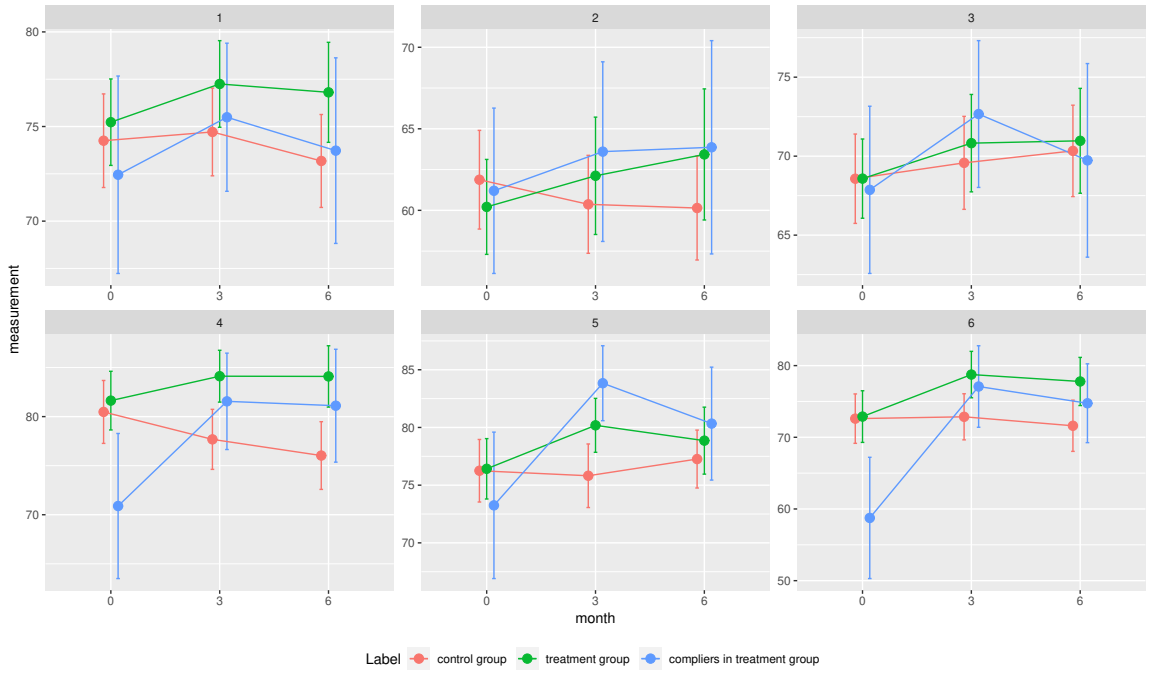

Figure 2.5: Means and standard errors for means in the treatment group, control group and the compliers in the treatment group for each of the six endpoints.

Our CACE analysis also considers the following baseline covariates: **Disease Duration:** an indicator variable for early disease (having RA for no more than two years); **Disease Activity:** an indicator variable for high disease activity (high RAPID4 values) with the reference level including remission and moderate/low RAPID4 values; **Gender:** an indicator variable for male; **Age:** an indicator variable for older than the median age (54.5). These baseline variables are well balanced between the intervention and control groups (Table 2.1). In contrast, the compliers in the intervention group had longer RA duration, higher disease activity, and were younger and all-female (Table 2.1). Table 2.1 also reports the

missing data patterns. There were a moderate number of dropouts and a small amount of intermittent missingness in both treatment arms. Our CACE analysis employs the likelihood approach, which has the benefit of yielding valid inference under the more general missing data mechanism (missing at random) than missing completely at random. Interestingly, there were no dropouts or intermittent missingness for the compliers, another indication of inherent differences between the subgroup of compliers and the overall study population. Thus the conventional AT analysis that directly compares the compliers in the treatment group with those untreated will be confounded by these inherent differences and be biased for treatment efficacy. The CACE analysis overcomes this limitation of the AT analysis.

2.4.2 MCACE Analysis

We analyzed the AHJ data using the method proposed in Section 2.2. Figure 2.5 suggests the possibility of quadratic time trends for the study endpoints for the compliers in treatment, which motivated us to start with a quadratic time trend in the submodel (Eqns 2.5 and 2.6) for our MCACE analysis. This submodel corresponds to a saturated time effect model for three visits. For the submodel of compliance status (Eqn 2.9), disease duration, disease activity, gender and age were included in $\boldsymbol{W}$ to estimate the probability of being a complier. Using likelihood ratio tests and AIC statistics, we conducted model selection to select a parsimonious and reasonable model (see Table 2.7 in Supplemental Information (section 2.6)) and chose MCACE.M7 as our best model for MCACE analysis. The estimation results from the model MCACE.M7 are presented in Table 2.8 in Supplemental Information (section 2.6). In model MCACE.M7, the fixed effects parameters on quadratic trends for compliers (β_{c20k} and β_{c21k}) and never takers (β_{n20k}) were all no different from zero and were thus dropped while keeping the random effects of quadratic trends for both two compliance strata ($\sigma^2_{v_{c2}} \neq 0$ and $\sigma^2_{v_{n2}} \neq 0$). Besides, based on the likelihood ratio test, the set of β_{c01k} parameters were no different from zero, which is expected in an RCT.

As a comparison, we also conducted the UCACE analysis by performing CACE estimation for the endpoints one by one. The model specification for each endpoint was the same as that in the model MCACE.M7, but unlike MCACE, the UCACE analysis ignored the correlations among endpoints. Thus, the UCACE analysis did not borrow information across multiple endpoints as MCACE did, when attempting to identify compliers from the never-takers in the control group which consists of a mixture distribution of these two subgroups of patients. Consequently, we expect a reduced power to detect the presence of treatment efficacy for UCACE as compared with MCACE. The estimates of the 6-month treatment efficacy ($2\hat{\beta}_{c11k}$) from both UCACE and MCACE analysis are reported in Table 2.2, where β_{c11k} represents average treatment difference at the 3rd month in the k^{th} outcome for compliers in treatment group. We observe the treatment efficacy estimates from the MCACE model are different from those from multiple UCACE models. The UCACE analysis shows that half of the estimates point to a harmful treatment effect in UCACE

models, although none of them is statistically significant. However, in the MCACE model, all estimates except the one for the 3rd endpoint point to beneficial treatment effects. This observation that MCACE model and UCACE models give different directions of treatment effects is possible because of the large variability of these estimates. Consistent with the results from simulation studies, the standard errors from MCACE analysis are smaller than those from UCACE analysis except for the first two endpoints. As noted in the simulation study, the standard error estimator via the Fisher information gives inaccurate estimates in UCACE when sample size is as small as 100. Therefore, it is likely that the true standard deviations from the MCACE analysis are all no greater than those from the UCACE analysis for all endpoints.

Table 2.1: Baseline characteristics and missingness patterns during follow-up visits

Covariates*	Control group (N = 49)		Treatment group (N = 45)		Compliers (Treatment) (N = 15)	
	n	%	n	%	n	%
Disease Duration (Early)	6	12.2	5	11.1	1	6.7
Disease Activity (High)	38	77.6	33	73.3	13	86.7
Gender (Male)	5	10.2	6	13.3	0	0
Age (>54.5)	25	51.0	22	48.9	6	40.0
Missing data pattern†	0 (1): presence (absent)					
000	41	83.7	33	73.3	15	100
011	2	4.1	9	20.0	0	0
001	5	10.2	1	2.2	0	0
010	1	2.0	2	4.4	0	0

∗ The covariates are binary and are reported as n and Percentage for the category indicated in the parenthesis.

† The 3-digit indicator represents the missing data pattern for outcomes at baseline, months 3 and 6.

Table 2.2: Estimates and standard errors for causal treatment effects at six months

Outcome	$CACE^{\dagger}$						$ITT^{\ddagger}$			$AT^{\ddagger}$		
	MCACE			UCACE								
	est	se	p-value	est	se	p-value	est	se	p-value	est	se	p-value
1	0.324	2.275	0.887	-0.410	1.857	0.825	1.262	2.096	0.547	0.232	3.178	0.942
2	1.706	4.379	0.697	-1.180	3.568	0.741	0.711	3.198	0.824	1.628	5.720	0.776
3	-4.005	4.379	0.360	7.810	8.213	0.342	-1.625	2.738	0.553	-1.283	4.724	0.786
4	17.754	6.329	0.005	16.383	7.173	0.022	6.413	3.189	0.044	12.858	5.681	0.024
5	1.834	4.375	0.675	-2.125	5.830	0.716	0.408	2.540	0.872	5.609	4.223	0.184
6	15.647	5.457	0.004	14.965	5.885	0.011	4.634	3.196	0.147	16.013	5.613	0.004
overall p-value			0.008			-			0.294			0.028

† The estimates from CACE models are for parameters $2 * \beta_{c11k}$.

‡ The estimates from ITT analysis and AT analysis are for $2 * \beta_{11k}$.

Table 2.2 also reports estimation results from ITT analysis and AT analysis. Both ITT and AT analyses employ hierarchical random-effects models with fixed-effects linear time trends for all longitudinal endpoints that pool information across endpoints. However, unlike the MCACE model, they do not model the partially observed compliance behavior. Instead, they compare the outcome trajectories between either the treatment assigned for ITT analysis or treatment received in AT analysis (section 2.6.2 in Supplemental Information (section 2.6)). Thus ITT and AT analyses generally do not yield consistent estimates for treatment efficacy as MCACE and UCACE do. Table 2.2 shows appreciable differences between estimates from MCACE and ITT, especially for endpoints 4 and 6 while AT estimates are relatively closer to those from MCACE.

We next move to the hypothesis testing of treatment effects on the six endpoints in the AHJ study. A hypothesis-testing strategy to control an inflated Type I error rate in RCTs with multiple endpoints is to first conduct a global test of no treatment effects for all endpoints and proceed to examine the individual endpoint if the global test rejects the null hypothesis of no treatment effects for all endpoints. For this purpose, we conduct multivariate Wald global tests of treatment differences for MCACE, ITT and AT. The UCACE analysis does not provide such a global test since it analyzes each endpoint separately. The null hypothesis for the global test in MCACE analysis is that population mean differences between treatment received among compliers are zeros for all six endpoints simultaneously (i.e., $\beta_{c11k} = 0$ for $k = 1, \cdots, 6$) whereas the global null hypothesis for ITT and AT is the population mean differences between treatment assigned for ITT and treatment received for AT are zeros for all endpoints (i.e., $\beta_{11k} = 0$ for $k = 1, \cdots, 6$), respectively. The last row in Table 2.2 reports the p-values from the global test for MCACE, ITT and AT. Both MCACE and AT analysis rejected the global null hypothesis (p-value < 0.05) while ITT failed to reject the global null hypothesis.

Given that MCACE rejected the global null hypothesis, we conclude that there were non-zero CACEs for at least one endpoint and move to examine which endpoints have non-zero CACEs. We apply a Wald test for each endpoint separately with Bonferroni correction that sets the threshold value for statistical significance at 0.05/6=0.0083 for each test. We observe that MCACE analysis found statistically significant beneficial CACEs of using AHJ on the fourth endpoint (communication with a physician) and the sixth endpoint (satisfaction with medical care). In comparison, UCACE analysis failed to detect a treatment effect for any endpoint with a threshold value of 0.0083. We attribute the lack of power to detect treatment effects in UCACE to its loss of estimation efficiency because of its ignoring correlations among endpoints. Although AT analysis also rejects the global null hypothesis and finds statistical significance for endpoint 6 at the level of 0.0083, we note that its test result and the AT estimates for individual endpoints are confounded by subjects' nonrandom compliance behavior and thus are generally biased for treatment efficacy.

We now turn to the estimation results of the compliance model. In the compliance model (Eqn 2.9), $\boldsymbol{W}$ includes disease duration, disease activity, gender and age at baseline with corresponding regression coefficients reported as α_1 to α_4 in Table 2.8 in Supplemental Information (section 2.6). All these baseline variables are binary variables. The intercept estimate $\hat{\alpha}_0 = -0.970$ indicates that a female patient under 54.5 years old with more than two years' disease and low disease activity had a probability of 0.27 to be a complier. Recall that there is no male in compliers in the treatment group (Table 2.1). Thus we expect a large coefficient estimate for gender: indeed $\hat{\alpha}_3 = -16.744$ in Table 2.8 in Supplemental Information (section 2.6). In this case, the coefficients of other baseline variables represent the independent effects of these variables on the probability of being a complier in females only. For example, $\hat{\alpha}_2 = 1.053$ (p-value=0.140) implies that a female patient with high disease activity is more likely to be a complier than a female patient with low disease activity, holding other predictors constant. Overall these coefficient estimates seem to suggest that, within female participants, patients who are younger with longer disease duration and high disease activity were more likely to be a complier. Although only the coefficient estimate for disease activity approaches statistical significance, such analysis could be useful for understanding the characteristics of compliers and for predicting the compliance of participants.

Based on the compliance model, we calculate the probability of being compliers for participants in the control group. Table 2.3 reports the average probability of being compliers for participants in the control group from MCACE and UCACE models. UCACE analysis yields six different fitted compliance models and the average probability of being compliers in the control group ranges from 0.29 to 0.36. However, MCACE model is able to pool the information across all endpoints to provide one fitted compliance model. The improved accuracy in identifying compliers helps MCACE achieve higher accuracy in CACE estimation.

Table 2.3: Average probability of being a complier for control group

Treatment group	Control group						
Proportion of	Aver.prob. from UCACE						Aver.prob. from MCACE
compliers	1	2	3	4	5	6	
0.333	0.329	0.336	0.289	0.357	0.352	0.360	0.311

2.4.3 Alternative Analysis

One issue different from the CACE analysis of multiple endpoints (our primary focus here) is the definition of compliers, which may not be clear-cut in all RCTs with treatment non-compliance. In the AHJ study, the expected benefit of AHJ is mainly the increased patient general self-awareness. Use of the AHJ tool is expected to lead to patients' increased realization of uncontrolled symptoms, increased understanding of the patterns in how their disease worked or the connection between symptoms (e.g., pain) and day-to-day life events (e.g.,

sleep and medications), as well as more efficient and effective rheumatology consultation during visits to doctors. Regular use of the AHJ is needed to achieve the anticipated benefits of the tool, and it is believed that this requires a minimum of monthly usage of AHJ. Hence the analysis so far defines the compliers as those who would use the AHJ at least once per month on average during the study period if assigned to the treatment group. The definition implies never-takers in intervention group includes patients who used the AHJ too rarely (less than once per month on average) to experience treatment effect. Although the implication seems to be reasonable, the difference between never-takers in treatment group and never-takers in control group reduces the plausibility of the exclusion restriction assumption.

One approach to increasing the plausibility of exclusion restriction is to relax the definition of compliers. We conduct the following analyses using two alternative definitions of compliers. The first alternative definition (A1) defines compliers as patients who would use the AHJ at least once within six months if assigned to treatment group. With this definition, never-takers in both groups did not use the AHJ and the assumption of exclusion restriction is more plausible. The tradeoff is to mis-classify those patients who used AHJ rarely as compliers and consequently dilute the CACEs. The second alternative (A2) defines compliers as patients who would use the AHJ at least three times in the 6-month period after being assigned to treatment group. The estimation results using the two alternative definitions of compliers are reported in Table 2.9 in Supplemental Information (section 2.6). The last row in Table 2.9 in Supplemental Information (section 2.6) reports the p-values from the multivariate Wald global tests of overall treatment differences across all six endpoints from MCACE analysis. The p-values for the global tests are 0.119 using definition A1, 0.010 using definition A2 and 0.008 as reported in Table 2.2 using the original definition of compliers. The finding is consistent with the expectation that the CACEs could be diluted by the less strict definition of compliers, albeit with increased plausibility of exclusion restriction. However, regardless of the definition used to classify compliers, we find that the p-values for the global test from MCACE analysis are all smaller than the overall p-value of 0.294 from the ITT analysis reported in Table 2.2.

2.5 Discussion

CACE is considered as more relevant for patient-oriented treatment effects of interest for RCTs under noncompliance. We propose a multivariate longitudinal potential outcome model with principal strata for latent compliance types to make inferences for CACE in longitudinal studies with multiple endpoints and treatment noncompliance. The method combines all data from correlated endpoints and over all longitudinal visits, and can substantially improve the estimation efficiency in RCTs with low compliance rate and moderate effect sizes on correlated endpoints. Simulation studies show significantly higher estimation

efficiency for MCACE as compared with the UCACE analysis, including up to 50% smaller standard errors of CACE estimates. In the power analysis, we evaluate a single overall test of the null hypothesis of no treatment effect under the MCACE model, which produces a 1-fold increase in the power of rejecting the null hypothesis compared with the UCACE analysis. These results demonstrate the potential of the proposed MCACE method to improve the efficiency and accuracy of evaluating comparative effectiveness and to reduce financial and time costs of conducting patient-oriented research.

We apply the proposed MCACE model, multiple UCACE models, multivariate ITT and AT models to the study of Arthritis Health Journal. Examining the overall p-value in Table 2.2, both MCACE analysis and AT analysis show the presence of a significant overall treatment effect while ITT analysis does not. However, AT analysis violates the randomization assumption and its p-value is not reliable. Besides, under Bonferroni correction, none of the p-values for the CACE estimates of individual endpoints from the UCACE analysis exhibits statistically significant treatment effects, whereas the MCACE finds significant CACEs for two out of six endpoints. These findings demonstrate the impact that the efficient MCACE procedure can make in real-world RCTs.

In our level-1 model, we assume diagonal matrices for V_c and V_n, the variance-covariance matrix for the residuals given the random effects and compliance type, which means the correlations among six outcomes and all time points are attributed to random effects and compliance types. This assumption could be relaxed by specifying a structure for Φ_m (e.g., compound symmetric or auto-regressive). Our MCACE assumes the potential outcomes given the compliance type follow multivariate normal distributions. The parametric distributional assumption permits efficient estimation of CACE estimates at the expense of potential model misspecifications. Future work can relax this assumption. One approach is to consider more flexible distributions, such as the multivariate t-distribution.

The assumption of exclusion restriction is often invoked to sharpen the CACE estimation. The definition of compliers may not be clear in all RCTs with treatment noncompliance and may involve a trade-off between the plausibility of exclusion restriction and the accuracy in classifying compliers. Instead of considering all-or-none compliance, extending the proposed methodology to continuously-measured partial compliance could be considered, which avoids the need to define a dichotomized compliance measure. A major challenge in the partial compliance approach is to find reasonable assumptions for model identification (Baker et al. 2016). Besides, with multiple endpoints, the exclusion restriction assumption may be more plausible for some of these endpoints than the remaining ones. Although it is not the focus of this chapter, the proposed method can be extended to relax the assumption of exclusion restriction for all endpoints.

2.6 Supplemental Information

2.6.1 Simulation Setting

True values for the parameters of compliers and never-takers are set as below:

$$\begin{aligned}
\boldsymbol{\beta_{c00}} &= (\beta_{c001}, \ldots, \beta_{c006})^T = (67, 56, 64, 66, -30, 54)^T, \\
\boldsymbol{\beta_{c10}} &= (\beta_{c101}, \cdots, \beta_{c106})^T = (1, 1, 3, -1, 3, 3)^T, \\
\boldsymbol{\beta_{n00}} &= (\beta_{n001}, \ldots, \beta_{n006})^T = (78, 63, 70, 88, -20, 82)^T, \\
\boldsymbol{\beta_{n10}} &= (\beta_{n101}, \cdots, \beta_{n106})^T = (0.2, 0.4, 0.8, -1.4, -0.4, -0.9)^T, \\
\boldsymbol{\psi_c} &= (\sigma^2_{v_{c0}}, \sigma^2_{c1}, \cdots, \sigma^2_{c6})^T = (e^{5.5}, e^4, e^5, e^5, e^6, e^5, e^5)^T, \\
\boldsymbol{\psi_n} &= (\sigma^2_{v_{n0}}, \sigma_{v_{n0}v_{n1}}, \sigma^2_{v_{n1}}, \sigma^2_{n1}, \cdots, \sigma^2_{n6})^T = (124, 8, 8, 12, e^4, e^5, e^5, e^4, e^4, e^4).
\end{aligned}$$

We assigned half individuals to the treatment group and the other half to the control group. The assignments of individuals were recorded in the vector $\boldsymbol{A_{obs}}$ where $A_{obs,i} = 1$ when i^{th} participant was assigned to the treatment group and $A_{obs,i} = 0$ when i^{th} subject was assigned to the control group. We used $\boldsymbol{D_{obs}}$ to record the actual receipt of treatment and set $D_{obs,i} = 0$ when $A_{obs,i} = 0$. $D_{obs,i} \sim Bin(1, p_c)$ when $A_{obs,i} = 1$.

Based on the three combinations of $(A_{obs,i}, D_{obs,i})$, we could divide the population into three groups : $\mathcal{S}(1,1), \mathcal{S}(1,0), \mathcal{S}(0,0)$. Then the response $\boldsymbol{y_i}$ was generated as below,

1. if $i \in \mathcal{S}(1,1)$, then generated $\boldsymbol{y_i}$ from normal distribution with mean $\boldsymbol{\mu_{ci}}$ and variance Σ_{ci}. Given the true value for parameters, $\boldsymbol{\mu_{ci}}$ and Σ_{ci} could be calculated based on Eqn 2.4.

2. if $i \in \mathcal{S}(1,0)$, then generated $\boldsymbol{y_i}$ from normal distribution with mean $\boldsymbol{\mu_{ni}}$ and variance Σ_{ni}. $\boldsymbol{\mu_{ni}}$ and Σ_{ni} could be calculated based on Eqn 2.4.

3. if $i \in \mathcal{S}(0,0)$, we generated u_i from Bernoulli distribution, $u_i \sim Bin(1, p_c)$. If $u_i \leq p_c$, then generated $\boldsymbol{y_i}$ from normal distribution with mean $\boldsymbol{\mu_{ci}}$ and variance Σ_{ci}. Otherwise, generated $\boldsymbol{y_i}$ from normal distribution with mean $\boldsymbol{\mu_{ni}}$ and variance Σ_{ni}.

In this simulation, we recorded time $\boldsymbol{T}$ as $0, 1, 2$ and the dataset was formed as $\{(\boldsymbol{A_{obs}}, \boldsymbol{D_{obs}}, \boldsymbol{T}, \boldsymbol{y_i})\}$.

2.6.2 Intention-to-treat Analysis & As-treated Analysis

When conducting the ITT analysis, we consider level-1 model as

$$Y_{ijk} = g(t_{ij}; \boldsymbol{b_{ik}}) + \epsilon_{ijk}.$$

For exposition simplicity, we set $g(t_{ij}; \boldsymbol{b_{ik}})$ as a linear function and $g(t_{ij}; \boldsymbol{b_{ik}}) = b_{0ik} + b_{1ik} t_{ij}$. Let $\boldsymbol{\epsilon_i} = (\boldsymbol{\epsilon_{i0}}^T, \boldsymbol{\epsilon_{i1}}^T, \cdots, \boldsymbol{\epsilon_{iJ}}^T)^T, \boldsymbol{\epsilon_{ij}} = (\epsilon_{ij1}, \epsilon_{ij2}, \cdots, \epsilon_{ijK})^T$. We assume $\boldsymbol{\epsilon_{ij}} \overset{iid}{\sim} \mathcal{N}(\boldsymbol{0}, \Phi)$ in

level 1 model, where $\Phi = diag(\sigma_1^2, \cdots, \sigma_K^2)$. And the level-2 models are considered as

$$\begin{aligned} b_{0ik} &= \beta_{00k} + \beta_{01k} A_{obs,i} + v_{0i}, \\ b_{1ik} &= \beta_{10k} + \beta_{11k} A_{obs,i} + v_{1i}, \end{aligned}$$

where $\boldsymbol{v_i} = \begin{pmatrix} v_{0i} \\ v_{1i} \end{pmatrix} \overset{iid}{\sim} \mathcal{N}(\mathbf{0}, \Sigma_v), \Sigma_v = \begin{bmatrix} \sigma_{v_0}^2 & \sigma_{v_0 v_1} \\ \sigma_{v_0 v_1} & \sigma_{v_1}^2 \end{bmatrix}$.

And we assume $\boldsymbol{\epsilon_i}$ and $\boldsymbol{v_i}$ are independent of each other. In this model, β_{00k} represents the average baseline measurement in the k^{th} outcome for individuals in control group. β_{10k} implies the average improvement in the k^{th} outcome for individuals in the control group. β_{01k} expresses the average baseline difference in the k^{th} outcome for individuals in the treatment group. β_{11k} expresses the average improvement difference in the k^{th} outcome for individuals in the treatment group. The random effects v_{0i} indicate individual deviation from average intercept and v_{1i} represents individual deviation from average improvement. Among these parameters, β_{11k} is of our primary interest, which is the intention to treat effect.

The level-1 model and level-2 models can be combined as

$$Y_{ijk} = \beta_{00k} + \beta_{10k} t_{ij} + \beta_{01k} A_{obs,i} + \beta_{11k} A_{obs,i} t_{ij} + v_{0i} + v_{1i} t_{ij} + \epsilon_{ijk}.$$

It's easy to derive the matrix form of the model for k^{th} outcome of individual i as below,

$$\begin{pmatrix} Y_{i0k} \\ Y_{i1k} \\ \vdots \\ Y_{iJk} \end{pmatrix} = X_i^\star * \begin{pmatrix} \beta_{00k} \\ \beta_{10k} \\ \beta_{01k} \\ \beta_{11k} \end{pmatrix} + Z_1 * \begin{pmatrix} v_{0i} \\ v_{1i} \end{pmatrix} + \begin{pmatrix} \epsilon_{i0k} \\ \epsilon_{i1k} \\ \vdots \\ \epsilon_{iJk} \end{pmatrix},$$

where

$$X_i^\star = \begin{pmatrix} 1 & 0 & A_{obs,i} & 0 \\ 1 & 1 & A_{obs,i} & A_{obs,i} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & J & A_{obs,i} & J * A_{obs,i} \end{pmatrix}, \qquad Z_1 = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ \vdots & \vdots \\ 1 & J \end{pmatrix}.$$

By using $X_i^\star$ and Z_1 to denote the design matrix of fixed effects and random effects respectively, we could rewrite the model as

$$\boldsymbol{Y_i} = (X_i^\star \otimes I_K)\boldsymbol{\beta} + \left(Z_1 \begin{pmatrix} v_{0i} \\ v_{1i} \end{pmatrix} \right) \otimes 1_K + \boldsymbol{\epsilon_i},$$

where $\boldsymbol{\beta} = (\beta_{001}, \cdots, \beta_{00K}, \beta_{101}, \cdots, \beta_{10K}, \beta_{011}, \cdots, \beta_{01K}, \beta_{111}, \cdots, \beta_{11K})^T$, $\boldsymbol{Y_i} = (Y_{i01}, \cdots, Y_{i0K}$, $Y_{i11}, \cdots, Y_{i1K}, \cdots, Y_{iJ1}, \cdots, Y_{iJK})^T$, $\boldsymbol{\epsilon_i} = (\boldsymbol{\epsilon_{i0}}^T, \cdots, \boldsymbol{\epsilon_{iJ}}^T)^T$.

Therefore, the marginal distribution for $\boldsymbol{Y_i}$ can be expressed as

$$f(\boldsymbol{y_i}|\boldsymbol{A_{obs}}, \boldsymbol{T}; \boldsymbol{\beta}, \boldsymbol{\psi}) \sim MVN(\boldsymbol{\mu_i}, \Sigma),$$

where $\boldsymbol{\mu_i} = (X_i^\star \otimes I_K)\boldsymbol{\beta}, \Sigma_i = \Sigma = (Z_1 \Sigma_v Z_1^T) \otimes (1_K 1_K^T) + V, V = var(\boldsymbol{\epsilon_i}) = diag(\Phi_0, \Phi_1, \cdots, \Phi_J)$, and $\Phi_0 = \Phi_1 = \cdots = \Phi_J = \Phi = diag(\sigma_1^2, \cdots, \sigma_K^2)$.

In this model, we denote $\boldsymbol{\psi} = (\sigma_{v_0}^2, \sigma_{v_0 v_1}, \sigma_{v_1}^2, \sigma_1^2, \cdots, \sigma_K^2)^T$. It's easy to know that the likelihood function is

$$L(\boldsymbol{\beta}, \boldsymbol{\psi}; \boldsymbol{y_{obs}}|\boldsymbol{A_{obs}}, \boldsymbol{T}) = \prod_{i=1}^{N} (2\pi)^{-\frac{(J+1)K}{2}} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\boldsymbol{y_i} - \boldsymbol{\mu_i})^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} (\boldsymbol{y_i} - \boldsymbol{\mu_i})\right).$$

Unlike ITT analysis, AT analysis compares the outcome based on the actual receipt of treatment ignoring the initial assignment of the treatment. Therefore, when conducting as-treated analysis, we only need to replace $A_{obs,i}$ with $D_{obs,i}$ in level-2 models and all other steps remain the same.

2.6.3 Tables

Table 2.4: A preliminary ITT analysis for each endpoint separately

Endpoint	est	se	p-value
1*	1.949	1.921	0.312
2†	2.305	3.099	0.459
3*	-1.511	2.590	0.560
4*	6.479	3.335	0.054
5*	-0.434	2.614	0.868
6*	4.268	3.388	0.210

∗ The results are based on linear mixed-effects models with linear time trends and random intercepts. Based on the likelihood ratio test, there is no group difference at baseline.

† The results for the second endpoint are based on a mixed-effects regression model with linear time trend, random intercept and trend. Based on the likelihood ratio test, there is no group difference at baseline.

Table 2.5: Estimation of CACE parameters with simulated data from Multivariate CACE model and univariate CACE models, based on 500 repetitions

Sample Size (n)		β_{c111}	β_{c112}	β_{c113}	β_{c114}	β_{c115}	β_{c116}
		True value : $\beta_{c11k} \equiv 0$					
100	$sm^{\dagger}$	0.096(0.135)*	-0.003(-0.144)	-0.107(0.520)	-0.028(-0.004)	-0.010(0.268)	-0.132(-0.144)
	$sse^{\ddagger}$	1.599(3.074)	2.339(4.688)	2.337(4.960)	3.508(4.694)	2.218(3.800)	2.160(3.188)
	$sm_{se}^{\ddagger}$	1.525(2.629)	2.210(4.129)	2.208(4.092)	3.444(4.510)	2.211(3.588)	2.218(3.197)
200	sm	0.099(0.066)	0.038(0.021)	-0.025(0.175)	-0.020(0.001)	0.056(0.221)	0.074(0.045)
	sse	1.129(2.370)	1.534(3.390)	1.558(3.525)	2.471(3.260)	1.568(2.508)	1.566(2.189)
	sm_{se}	1.075(1.959)	1.569(3.200)	1.562(3.243)	2.429(3.170)	1.560(2.506)	1.569(2.253)
500	sm	0.035(0.129)	-0.017(0.134)	-0.014(0.053)	-0.123(-0.150)	-0.063(0.011)	-0.012(-0.067)
	sse	0.677(1.353)	1.002(2.211)	1.017(2.292)	1.603(1.952)	0.983(1.615)	1.013(1.411)
	sm_{se}	0.676(1.276)	0.982(2.108)	0.985(2.145)	1.525(1.978)	0.980(1.578)	0.984(1.410)
		True value : $\beta_{c11k} \equiv 1.5$					
100	sm	1.477(1.536)	1.570(1.741)	1.582(1.506)	1.359(1.503)	1.316(1.277)	1.509(1.634)
	sse	1.472(2.898)	2.394(4.904)	2.313(4.661)	3.498(4.659)	2.183(3.766)	2.333(3.397)
	sm_{se}	1.510(2.584)	2.198(4.085)	2.179(4.051)	3.434(4.476)	2.198(3.529)	2.190(3.186)
200	sm	1.462(1.523)	1.452(1.320)	1.509(1.593)	1.537(1.661)	1.529(1.573)	1.445(1.474)
	sse	1.063(2.328)	1.539(3.603)	1.543(3.614)	2.525(3.441)	1.519(2.626)	1.726(2.363)
	sm_{se}	1.069(1.920)	1.558(3.177)	1.556(3.198)	2.419(3.154)	1.553(2.502)	1.559(2.245)
500	sm	1.515(1.453)	1.480(1.379)	1.524(1.575)	1.513(1.484)	1.480(1.537)	1.573(1.557)
	sse	0.693(1.400)	0.953(2.191)	0.959(2.344)	1.616(2.026)	0.971(1.595)	0.975(1.464)
	sm_{se}	0.678(1.274)	0.984(2.122)	0.985(2.111)	1.534(1.991)	0.986(1.580)	0.986(1.411)
		True value : $\beta_{c11k} \equiv 3$					
100	sm	3.039(3.061)	3.028(2.856)	2.852(3.319)	2.866(3.217)	2.974(3.132)	2.877(2.860)
	sse	1.611(3.250)	2.275(4.779)	2.360(4.717)	3.408(4.798)	2.306(4.002)	2.232(3.483)
	sm_{se}	1.526(2.619)	2.234(4.140)	2.222(4.197)	3.465(4.587)	2.232(3.580)	2.233(3.206)
200	sm	2.878(2.796)	3.108(2.909)	3.001(3.035)	3.049(3.197)	3.023(3.041)	3.056(3.188)
	sse	1.131(2.495)	1.588(3.533)	1.515(3.493)	2.454(3.044)	1.616(2.698)	1.599(2.305)
	sm_{se}	1.065(1.955)	1.555(3.120)	1.561(3.184)	2.403(3.135)	1.550(2.499)	1.558(2.234)
500	sm	2.998(2.986)	2.987(2.992)	2.987(3.056)	3.006(3.011)	2.990(2.903)	2.965(3.006)
	sse	0.665(1.365)	1.009(2.161)	1.015(2.485)	1.574(1.968)	0.985(1.556)	0.992(1.441)
	sm_{se}	0.676(1.272)	0.981(2.103)	0.982(2.143)	1.526(1.980)	0.981(1.571)	0.982(1.407)

$*$ Values in brackets are from multiple univariate models

$\dagger sm$: sample mean of estimates

$\ddagger sse$: sample standard deviation of estimates

$\ddagger sm_{se}$: sample mean of standard error estimates by the Fisher information

Table 2.6: Estimates of compliance rate from the multivariate CACE model and univariate CACE under different sample size

Sample Size(n)	p_c	Univariate CACE							Multivariate CACE
		1	2	3	4	5	6	Mean	
100	0.3	0.2987	0.2993	0.2986	0.2985	0.2975	0.2993	0.2987	0.2982
200	0.3	0.2986	0.2980	0.2979	0.2978	0.2984	0.2980	0.2981	0.2973
500	0.3	0.2996	0.2998	0.2998	0.3002	0.2996	0.2999	0.2998	0.2999

- Effect size $\beta_{c3k} \equiv 0$

Table 2.7: Model selection for MCACE analysis

Model	#(parameter)	v_{c2i}	v_{n2i}	β_{n20k}	β_{c01k}	β_{c21k}	β_{c20k}	$logL$	AIC	Model comparison	p-value
MCACE.M1	83	✓	✓	✓	✓	✓	✓	-5873.55	11841.1		
MCACE.M2	80	✓	×	✓	✓	✓	✓	-5873.55	11841.1	vs. M1	$< .0001$
MCACE.M3	80	×	✓	✓	✓	✓	✓	-5845.869	11851.74	vs. M1	0.0004
MCACE.M4	77	✓	✓	×	✓	✓	✓	-5839.893	11833.79	vs. M1	0.585
MCACE.M5	71	✓	✓	×	×	✓	✓	-5841.705	11825.41	vs. M4	0.727
MCACE.M6	65	✓	✓	×	×	×	✓	-5843.845	11817.69	vs. M5	0.639
MCACE.M7	59	✓	✓	×	×	×	×	-5848.05	**11814.10**	vs. M6	0.210
MCACE.M8	77	✓	✓	✓	×	✓	✓	-5839.364	11832.73	vs. M1	0.727
MCACE.M9	71	✓	✓	✓	×	×	✓	-5841.467	11824.93	vs. M8	0.649
MCACE.M10	65	✓	✓	✓	×	×	×	-5845.694	11821.39	vs. M9	0.207
										vs. M7	0.581

Note: P-values are calculated based on the likelihood ratio test and are divided by 2 when dealing with M1 vs. M2, M3.

Table 2.8: Estimation results from the selected MCACE model (MCACE.M7), which dropped β_{n20k}, β_{c20k},β_{c01k}, β_{c21k} from Eqn 2.6

parameter	estimates	standard error	p-value	test	p-value	parameter	estimates	standard error	p-value
Outcome model									
β_{c001}	72.175	3.237	$< .0001$			β_{n001}	78.133	1.761	$< .0001$
β_{c002}	60.424	3.682	$< .0001$			β_{n002}	63.090	2.308	$< .0001$
β_{c003}	69.013	3.740	$< .0001$			β_{n003}	70.390	1.962	$< .0001$
β_{c004}	70.305	4.317	$< .0001$			β_{n004}	87.842	1.805	$< .0001$
β_{c005}	74.379	3.837	$< .0001$			β_{n005}	79.268	1.809	$< .0001$
β_{c006}	58.256	4.058	$< .0001$			β_{n006}	82.036	1.881	$< .0001$
β_{c101}	0.433	0.914	0.635			β_{n101}	0.696	0.832	0.403
β_{c102}	0.472	1.996	0.813			β_{n102}	0.610	1.495	0.683
β_{c103}	2.334	1.979	0.238			β_{n103}	1.575	1.100	0.152
β_{c104}	-2.992	2.908	0.303			β_{n104}	-0.700	0.927	0.450
β_{c105}	2.668	1.999	0.182			β_{n105}	0.334	0.854	0.696
β_{c106}	1.852	2.519	0.462			β_{n106}	-0.430	0.992	0.665
$\boldsymbol{\beta_{c111}}$	0.162	1.138	0.887						
$\boldsymbol{\beta_{c112}}$	0.853	2.190	0.697						
$\boldsymbol{\beta_{c113}}$	-2.002	2.190	0.360	$\beta_{c11k} \equiv 0$	0.008				
$\boldsymbol{\beta_{c114}}$	8.877	3.164	0.005						
$\boldsymbol{\beta_{c115}}$	0.917	2.188	0.675						
$\boldsymbol{\beta_{c116}}$	7.823	2.728	0.004						
$\sigma^2_{v_{c0}}$	341.694	97.758				$\sigma^2_{v_{n0}}$	139.366	27.324	
$\sigma_{v_{c0}v_{c1}}$	-171.396	72.372				$\sigma_{v_{n0}v_{n1}}$	19.144	28.298	
$\sigma_{v_{c0}v_{c2}}$	80.965	34.253				$\sigma_{v_{n0}v_{n2}}$	-10.082	12.462	
$\sigma^2_{v_{c1}}$	185.752	73.594				$\sigma^2_{v_{n1}}$	157.379	48.094	
$\sigma_{v_{c1}v_{c2}}$	-89.273	35.055				$\sigma_{v_{n1}v_{n2}}$	-61.735	20.373	
$\sigma^2_{v_{c2}}$	42.915	16.956				$\sigma^2_{v_{n2}}$	25.718	8.955	
σ^2_{c1}	21.716	7.001				σ^2_{n1}	48.846	7.020	
σ^2_{c2}	140.149	25.671				σ^2_{n2}	219.260	25.840	
σ^2_{c3}	139.854	24.370				σ^2_{n3}	106.465	13.253	
σ^2_{c4}	311.166	49.846				σ^2_{n4}	66.773	8.822	
σ^2_{c5}	140.818	23.675				σ^2_{n5}	51.095	7.267	
σ^2_{c6}	233.662	37.813				σ^2_{n6}	80.106	10.439	
Compliance model*									
α^*_0	-0.970	0.714	0.175						
α^*_1	-1.907	1.118	0.088						
α^*_2	1.053	0.714	0.140						
α^*_3	-16.744	2190.414	0.994						
α^*_4	-0.670	0.509	0.188						
$-logL$(59 parameter)	5848.05								

$*$ These α parameters in the compliance model correspond to intercept, coefficients for disease duration, disease activity, gender and age.

Table 2.9: Alternative analysis: estimates and standard errors for causal treatment effects at six months

Outcome	MCACE.A1*			$MCACE.A2^{\dagger}$		
	est	se	p-value	est	se	p-value
1	-0.030	2.093	0.988	-2.556	2.495	0.306
2	3.134	3.643	0.390	-1.428	3.800	0.707
3	0.199	3.246	0.951	-6.187	3.446	0.073
4	8.167	3.783	0.031	9.011	4.534	0.047
5	-0.199	2.816	0.944	-2.693	3.282	0.412
6	8.423	3.751	0.025	9.231	4.263	0.030
overall p-value			0.119			0.010

$*$ The results under MCACE.A1 were obtained when compliers are patients who would use the AHJ at least once in 6-month period if assigned to treatment group.

† The results under MCACE.A2 were obtained when compliers are patients who would use the tool at least three times in 6-month period if assigned to treatment group. Estimates from both two CACE analysis are for parameters $2 * \beta_{c11k}$.

Chapter 3

A Latent-factor MCACE Model for Multidimensional Outcomes and Treatment Noncompliance

3.1 Introduction

Randomized controlled trials (RCTs) are the preferred study design to assess intervention effects for healthcare policy decision makings. In many real-world RCTs, however, individuals randomized to an intervention group often do not comply with the assigned treatment. This is especially the case while evaluating complex interventions such as behavioural interventions. With treatment noncompliance, standard intention-to-treat (ITT) analysis typically provides conservative estimates of intervention efficacy (Sheiner & Rubin 1995). To overcome the limitation of ITT analysis, the complier average causal effect (CACE), the principal causal effect (PCE) within the stratum of compliers, has been developed to estimate the intervention efficacy for the subpopulation who would comply regardless of assigned treatment (Baker & Lindeman 1994, Imbens & Angrist 1994, Imbens & Rubin 1997). CACE has been considered as patient-oriented intervention effects of interest under treatment noncompliance (Steele et al. 2015).

Furthermore, real-world RCTs evaluating multifaceted interventions often employ multiple study outcomes (also known as endpoints) to measure a limited set of underlying latent constructs, such as psychological traits, mental health status, quality of life, self-efficacy, knowledge, and attitudes. Multifaceted interventions contain multiple components designed to impact a set of underlying constructs (e.g., self-efficacy in disease management and effectiveness in shared decision-making), each of which is measured by a number of study outcomes. Frequently, these multiple study outcomes are collected longitudinally on study participants, yielding multidimensional longitudinal outcomes. Evaluating CACEs for each outcome separately can suffer from the problems of multiple testing, a significant loss of statistical power, lacking of the ability to directly answer main scientific questions of treatment efficacy on the underlying constructs, and the difficulty in interpreting potentially

conflicting results among individual outcomes. To overcome these limitations, this chapter develops a latent factor model for parsimonious CACE estimation of intervention effects in RCTs with multidimensional longitudinal outcomes and treatment noncompliance.

The application motivating this work is the Arthritis Health Journal (AHJ) study, an RCT comparing the AHJ (intervention group) with the usual care (control group) in managing rheumatoid arthritis (RA) (Lacaille et al. 2015, Guo et al. 2022). As there has been no cure for RA, to achieve optimal health outcomes, people with RA need to engage in efficient self-management and effective collaboration with their healthcare providers (Barlow et al. 2002, Tam et al. 2019). AHJ is a patient-centered online tool designed to improve self-efficacy in disease management and shared decision-making for RA patients. By helping RA patients better monitor their disease activity and collaborate with their healthcare providers, this tool aims to facilitate the treat-to-target approach by providing early signs when the disease is not controlled. In the RCT, RA patients randomized to the AHJ were provided with online access to the AHJ after randomization and were asked to use it for 6 months, while those randomized to the control group did not have access to the AHJ during the 6-month period.

The primary objective of this study is to evaluate the treatment efficacy of the AHJ tool on underlying constructs (e.g., patients' self-efficacy in disease management and the effectiveness in shared decision-making). To effectively capture these complex underlying constructs, the AHJ study employed a total of six study endpoints (effective consumer 17 scale, manage symptom scale, manage disease in general scale, partners in health scale, communicate with physician scale, and satisfaction with medical care) measured using self-reported questionnaires administered at baseline, 3 months and 6 months after baseline. Besides study endpoints, the baseline questionnaires collected demographic and disease information, and the follow-up questionnaires administered at 3 and 6 months also evaluated the frequency of using the AHJ online tool. Like many other real-world RCTs, treatment noncompliance occurred in the AHJ study, and a significant number of study participants randomized to the AHJ did not use, or used it rarely, during the study period. A secondary objective of this study is to investigate predictors of compliance behaviour of study participants.

In the presence of such treatment noncompliance, the traditional ITT analysis or the as-treated analysis can yield biased estimates of the intervention effects of interest. To achieve our primary objective, an attractive alternative is evaluating the effect of an intervention on the outcomes, adjusting properly for the treatment noncompliance using the principal stratification (PS) approach. In the context of treatment noncompliance, proper analysis adjusts for the principal strata corresponding to compliance types formed by the joint potential compliance behaviours under both control and intervention. As such defined, the values of principal strata are unaffected by the treatment assignment and behave like a baseline categorical variable. Thus, one can define the causal effects within each subpopula-

tion determined by the compliance types. An intervention effect estimand of great interest is the CACE, the intervention effects in the subpopulation of compliers who would comply regardless of treatment assigned (Baker & Lindeman 1994, Imbens & Rubin 1997).

Estimating CACEs for AHJ requires handling multidimensional longitudinal outcomes, which raises several statistical issues. One issue is the multiple endpoints. These endpoints measure different aspects about a small set of underlying constructs. Estimating CACE demands proper statistical methods that can jointly consider all the endpoints and properly combine information from these correlated multiple endpoints for efficient and interpretable treatment effect estimation. Furthermore, statistical methods permitting parsimonious testing of CACEs and mitigating the multiple testing issues in the presence of multiple endpoints are desired. Finally, these methods must account for the longitudinal correlations among repeated measures and cross-sectional correlations among multiple endpoints.

One approach is to estimate CACE for each endpoint separately. This approach is straightforward to apply. However, without considering the correlations across multiple outcomes, the method is inefficient and can significantly reduce the study power to detect intervention effects. Besides the multiple testing issues associated with analyzing these endpoints individually, the results can be difficult to interpret because of lacking direct connection with the underlying constructs of main interest and potentially conflicting results among individual endpoints. To improve the efficiency of CACE estimation in the presence of treatment noncompliance, most works focus on modelling two outcomes jointly in a cross-sectional setting. Jo & Muthén (2001) employed a secondary outcome to increase the precision of identifying compliance class and the power to detect intervention effects on the primary outcome. Mattei et al. (2013) proposed a Bayesian approach to exploit bivariate outcomes to sharpen inferences for weakly identified models within principal strata in the cross-sectional setting. Mealli & Pacini (2013) showed a secondary outcome helps tighten the nonparametric bounds of CACE. An exception is Guo et al. (2022) who considered the CACE estimation for multidimensional longitudinal outcomes in RCTs with treatment noncompliance and showed that jointly modeling of multiple study endpoints significantly improves the precision of estimating CACE and the power to detect CACE for individual endpoints. However, none of these methods is designed to exploit the underlying constructs targeted in RCTs of multifaceted interventions, including the AHJ study. Therefore, these methods can suffer from multiple testing issues in the presence of multiple endpoints and may yield less interpretable results.

In this chapter, we introduce a latent-factor multivariate CACE (MCACE) model that exploits underlying constructs for parsimonious CACE estimation in RCTs with multidimensional longitudinal outcomes and treatment noncompliance. Within each (potentially unobserved) compliance type, a latent-factor hierarchical regression model is used to connect longitudinally measured multiple endpoints with latent factors representing underlying constructs. Then, separate linear mixed-effects models are used to model these latent fac-

tors and the principal causal effects on the latent factors. Compared with analyzing the causal effects of multiple endpoints individually, latent factors can capture the correlations across multiple outcomes. Inference based on latent factors makes more efficient use of information from all sources than making inference using data on each individual endpoint separately. Thus, efficiency and power will be increased by introducing latent factors. As compared with alternative joint CACE estimation approaches (Jo & Muthén 2001, Mattei et al. 2013, Mealli & Pacini 2013, Guo et al. 2022), our proposed approach exploits underlying constructs, thereby increasing the interpretability of results and reducing multiple testing issues with multiple endpoints. Unlike data reduction techniques using pre-specified functions of individual endpoints (e.g., sum or weighted average) or variable reduction methods such as principal components analysis, our approach derives the underlying constructs and factor loading using all data on multidimensional outcomes at all time points, yielding results that can be more interpretable and efficient.

We apply the proposed approach to evaluate the treatment efficacy of the AHJ online tool. Model comparison selects two underlying constructs (patients' self-efficacy and interactions with their caregivers), permitting parsimonious and more powerful tests of CACEs on the low-dimensional latent constructs as compared with the CACE analysis on the six outcomes separately. In particular, using the proposed model, we can detect significant and beneficial CACEs of AHJ, adjusting for multiple testing issues, on both latent constructs that are scientifically relevant. The findings differ importantly from those using the alternative joint CACE modeling approach that jointly models all six study endpoints directly without considering the underlying constructs. Specifically, in the joint CACE model proposed in Guo et al. (2022), after rejecting the null hypothesis of no CACE for all six endpoints using a global test, one has to examine CACEs for the six endpoints separately to determine the location (which endpoint?) and direction (beneficial or harmful?) of intervention effects. Statistically significant CACEs were found in only two out of six study endpoints after multiple testing adjustments, leading to less efficient and less clear interpretation of the RCT data compared with using the method proposed in this chapter.

We describe the methodology in Section 3.2. Section 3.3 describes the simulation studies, which demonstrate the performance and advantages of the proposed model. We then apply the method to the AHJ data in Section 3.4. Finally, a discussion is given in Section 3.5.

3.2 Methodology

3.2.1 Notation and Assumptions

Let A_i indicate the i^{th} subject's group assignment, $i = 1, \cdots, N$. Participants are randomly assigned to the intervention group ($A_i = 1$) or the control group ($A_i = 0$). Let $D_i(A_i)$ indicate the receipt of the treatment if the i^{th} individual was assigned to group A_i. In the AHJ study, we define $D_i(A_i = 1) = 1$ when subject i would use the tool at least one time

per month on average within six months after being assigned to the intervention group, and define $D_i(A_i = 1) = 0$ if otherwise (Guo et al. 2022). Let $\boldsymbol{A}$ and $\boldsymbol{D}$ denote N-dimensional vectors of A_i and D_i, respectively. Two types of potential outcomes can be defined. $D_i(\boldsymbol{A})$ is the potential treatment received by subject i when subjects are randomized to $\boldsymbol{A}$. Assuming K endpoints were observed over time for each individual in the study, $Y_{ijk}(\boldsymbol{A}, \boldsymbol{D})$ is the potential outcome for the k^{th} endpoint collected at the j^{th} time point for individual i under treatment assignment $\boldsymbol{A}$ and treatment receipt $\boldsymbol{D}$, where $j = 0, 1, \cdots, J$ and $k = 1, \cdots, K$. In the AHJ study, $j = \{0, 1, 2\}$ for the baseline, third month and sixth month, respectively, and $K = 6$ for the six endpoints collected every three month for each participant. Let $\boldsymbol{Y_i(A, D)}$ denote the vector of $K * (J + 1)$ potential outcomes for subject i under treatment assignment $\boldsymbol{A}$ and treatment receipt $\boldsymbol{D}$.

Table 3.1 summarizes the key assumptions made for the identification of the latent-factor MCACE model in the study. Below we first discuss Assumptions 1 to 4 with the remaining ones to be described later in the development of the latent-factor MCACE model.

Assumption 1. *Stable Unit Treatment Value Assumption (SUTVA, Rubin 1978, 1980, 1990).*

The SUTVA assumption assumes no interference and no multiple versions of treatment. The former one implies the potential outcomes of an individual are not influenced by possible treatment assignments of others. The latter one assumes each individual receives exactly the same version of the treatment. The SUTVA assumption helps define the unit-level causal effect and allows us to simplify $\boldsymbol{Y_i(A, D)}$ and $D_i(\boldsymbol{A})$ as $\boldsymbol{Y_i}(A_i, D_i)$ and $D_i(A_i)$. SUTVA is satisfied in the AHJ study because participants get access to exactly the same AHJ tool independently.

Assumption 2. *Random assignment.*

This assumption implies the assignment A_i is independent of potential outcomes $\boldsymbol{Y_i}(A_i, D_i)$ and $D_i(A_i)$ given all observed baseline variables. The randomization assumption is satisfied in RCTs because RCTs randomly assign participants to the intervention and control groups.

Assumption 3. *Patients in the control group do not have access to the treatment.*

In the AHJ study, participants assigned to the control group had no access to the AHJ online tool during the first six-month period, and Assumption 3 is satisfied. Since $D_i(A_i)$ is a binary variable, the combination of potential outcomes $(D_i(1), D_i(0))$ defines four possible compliance patterns: compliers $(1, 0)$, never-takers $(0, 0)$, always-takers $(1, 1)$ and defiers $(0, 1)$. Under Assumption 3, $D_i(0) = 0$ for all individuals which ruled out defiers and always-takers. Compliers $(1, 0)$ and never-takers $(0, 0)$ can be distinguished in the treatment group because $D_i(1)$ is observable in the treatment arm. However, compliers and never-takers can not be distinguished in the control group since $D_i(1)$ is not observable for the individuals

assigned to the control group. Let C_i be the compliance type of the participant i, $C_i \in \{c, n\}$ where c denotes compliers and n denotes never-takers. C_i is observable in the treatment group and unknown in the control group.

In other RCTs where controls can take the treatment, Assumption 3 does not hold, and the monotonicity assumption is often invoked to rule out certain compliance types. For example, Hirano et al. (2000) employs the monotonicity assumption ($D_i(1) \geq D_i(0)$) to rule out defiers for model identification and makes inference based on the population consisting of compliers, never-takers and always-takers. Our proposed method is general and can be extended to these situations when Assumption 3 is relaxed.

Assumption 4. *Exclusion restriction (Imbens & Angrist 1994, Baker & Lindeman 1994).*

Assumption 4 implies $\boldsymbol{Y(A, D)} = \boldsymbol{Y(A', D)}$ $\forall \boldsymbol{A}, \boldsymbol{A'}$ and $\forall \boldsymbol{D}$, which implies $\boldsymbol{Y(A, D)} = \boldsymbol{Y(D)}$. This means that there is no difference in potential outcomes between treatment and control groups for never-takers.

Table 3.1: Assumptions for the identification of latent-factor MCACE model

Assumptions	Statements
1: Stable Unit Treatment Value Assumption	No interference and no multiple versions of treatment.
2: Random assignment	Assignments are independent of potential outcomes given all observed baseline variables.
3: No access to the treatment in the control group	Rule out defiers and always-takers.
4: Exclusion restriction	$\boldsymbol{Y(A, D)} = \boldsymbol{Y(D)}$
5: Conditional independence	Potential outcomes are conditionally independent given latent factors $\boldsymbol{U^a_{mij}}$.
6: Stability of factor loading matrix Λ	Λ is constant over time and across different compliance patterns.
7: Rotation restriction	Restrictions are imposed on matrix Λ to fix its rotation under both confirmatory and exploratory analyses.
8: Scale restriction	$\boldsymbol{x^a_{qij}}$ and $\boldsymbol{z_{qij}}$ do not include intercepts and $\epsilon^a_{qmij} \sim N(0, 1)$.

3.2.2 Models

In this section, we propose a latent factor model with principal strata for partially observed compliance types under the potential outcome framework. In the AHJ study, follow-up questionnaires employ multiple endpoints to measure the underlying constructs (factors): patients' self-efficacy and satisfaction with health professional care and communication. These latent factors are of primary interest in the study, but cannot be directly observed. Instead, multiple endpoints are used to measure these latent factors, which capture the interdependence of those multiple endpoints. Thus, it is of scientific interest to directly estimate the causal effect of the intervention on the latent factors.

To achieve the above goal, we introduce the latent factor $\boldsymbol{U}_i^a$ when modeling the joint distribution of potential treatment received $(D_i(0), D_i(1))$ and potential outcomes of multiple endpoints $(\boldsymbol{Y}_i^0, \boldsymbol{Y}_i^1)$, where $\boldsymbol{Y}_i^a$ denotes the potential outcome $\boldsymbol{Y}_i(D_i(A_i = a)), a = 0, 1$. Specifically, we model $(\boldsymbol{Y}_i^0, \boldsymbol{Y}_i^1) \mid C_i, \boldsymbol{U}_i^a$ in which the compliance type C_i has one-to-one correspondence to $(D_i(0), D_i(1))$, and the latent factor $\boldsymbol{U}_i^a$ captures the interdependence of multiple endpoints induced by sharing a common set of latent factors in $\boldsymbol{U}_i^a$ within the compliance type C_i. We then model the conditional distribution $\boldsymbol{U}_i^a \mid C_i$, which captures the CACEs on the latent factors in $\boldsymbol{U}_i^a$. Finally, the compliance behavior C_i is modelled by using a logistic regression model. The diagram in Figure 3.1 outlines the model structure with modeling details described below.

Model for $\boldsymbol{Y_i^a} \mid C_i, \boldsymbol{U_i^a}$

Let $\boldsymbol{Y_{ij}^a} = (Y_{ij1}^a, \cdots, Y_{ijK}^a)^T$. The level-1 part of our multi-level model specifies the relationship between $\boldsymbol{Y_{ij}^a}$ and the latent factors $\boldsymbol{U_{mij}^a}$ given the compliance type $C_i = m$ as

$$\boldsymbol{Y_{ij}^a} \;\Big|\; (C_i = m, \boldsymbol{U_{mij}^a}, \boldsymbol{b_{mi}^a}) = \boldsymbol{\lambda_{m0}} + \Lambda \boldsymbol{U_{mij}^a} + \boldsymbol{b_{mi}^a} + \boldsymbol{e_{mij}^a}, \tag{3.1}$$

where $m \in \{c, n\}$ denotes the unique value of compliance type; $\boldsymbol{U_{mij}^a} = (U_{1mij}^a, \cdots, U_{Qmij}^a)^T$ are Q ($Q \ll K$) latent factors for subject i at time j with group assignment a and compliance type m; Λ is a K by Q matrix of regression coefficients with λ_{kq} at the k^{th} row and q^{th} column, which does not change over time and across different compliance patterns; $\boldsymbol{\lambda_{m0}} = (\lambda_{m01}, \cdots, \lambda_{m0K})^T$ is the average baseline measurements for subjects in control group under the compliance type m; $\boldsymbol{b_{mi}^a} = (b_{mi1}^a, \cdots, b_{miK}^a)^T$ with b_{mik}^a representing the k^{th} outcome's random intercept. In the above level-1 model, the latent factors $\boldsymbol{U_{mij}^a} = (U_{1mij}^a, \cdots, U_{Qmij}^a)^T$ capture the variability and interdependence among the K responses at each time point j. For the k^{th} endpoint, b_{mik}^a captures the correlation across longitudinal measurements of y_{ijk}^a over time. We assume b_{mik}^a follows a normal distribution with mean 0 and variance ξ_{mk} and $b_{mik}^a \perp\!\!\!\perp b_{mih}^a$, $k \neq h$. Finally, the error term $\boldsymbol{e_{mij}^a} = (e_{mij1}^a, \cdots, e_{mijK}^a)^T$ with e_{mijk}^a distributed independently as $N(0, \tau_{mk}^2)$, $e_{mijk}^a \perp\!\!\!\perp e_{mijh}^a$ for $k \neq h$. We further assume $e_{mijk}^0 \perp\!\!\!\perp e_{mijk}^1$, $b_{mik}^0 \perp\!\!\!\perp b_{mik}^1$ and $e_{mijk}^a \perp\!\!\!\perp b_{mik}^a$.

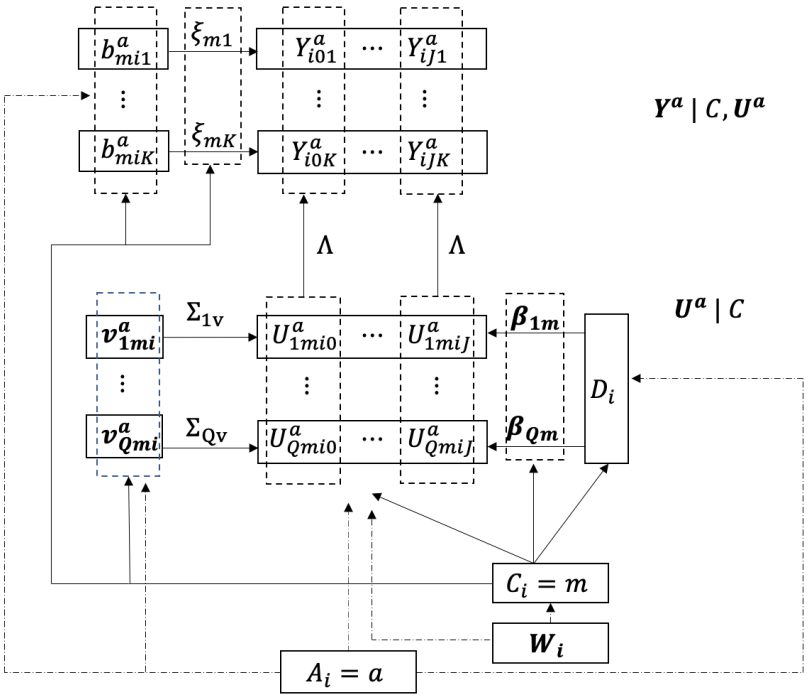

Figure 3.1: Illustration of the structure of the latent-factor MCACE model with principal strata for latent compliance types.

The level-1 model makes the following key assumptions for model identification. First, the potential outcomes are conditionally independent given the latent factors $\boldsymbol{U^a_{mij}}$ (Assumption 5 in Table 3.1), which means b^a_{mik} $(k = 1, \cdots, K)$ are independent. At time j, the cross-sectional correlation among the potential outcomes y^a_{ijk} $(k = 1, \cdots, K)$ is induced by the common latent factors $\boldsymbol{U^a_{mij}}$. Second, the correlation between the potential outcomes and latent factors, Λ, remains the same over time and across different compliance patterns (Assumption 6 in Table 3.1), which is required for the sake of identifiability. For any orthogonal matrix T that satisfies $TT' = T'T = I$, $\Lambda \boldsymbol{U^a_{mij}} = \Lambda TT' \boldsymbol{U^a_{mij}} = \Lambda^* \boldsymbol{U^{a*}_{mij}}$ where $\Lambda^* = \Lambda T$ and $\boldsymbol{U^{a*}_{mij}} = T' \boldsymbol{U^a_{mij}}$. Since there are infinite possible orthogonal matrices, Λ can be rotated to Λ^* in infinite ways. Therefore, we need to enforce some restrictions to fix the rotation of matrix Λ (Assumption 7 in Table 3.1). In confirmatory analysis, the structure of matrix Λ is specified based on the scientific relationships between potential outcomes and latent factors, which imposes sufficient constraints to fix the rotation of the matrix Λ. In exploratory analysis, we do not make any assumptions about the latent structure of potential outcomes except that we set $\lambda_{kq} = 0$, $q > k$ (An et al. 2013).

Model for $\boldsymbol{U^a_i} \mid C_i$.

In the level-2 model, we assume a linear mixed-effects model to study the longitudinal latent factors U^a_{qmij}, $q = 1, \cdots, Q$, for individuals with compliance type m,

$$U^a_{qmij} \;\Big|\; (C_i = m, \boldsymbol{v^a_{qmi}}) = \boldsymbol{x^a_{qij}\beta_{qm}} + \boldsymbol{z_{qij}v^a_{qmi}} + \epsilon^a_{qmij}, \tag{3.2}$$

where $\boldsymbol{x^a_{qij}}$ and $\boldsymbol{z_{qij}}$ are vectors of covariates for fixed effects and random effects for the q^{th} latent factor respectively. The vector of covariates ($\boldsymbol{x^a_{qij}}$) could include time trends, the receipt of the treatment, baseline characteristics (demographic information, disease severity, etc), while $\boldsymbol{z_{qij}}$ is a subset of $\boldsymbol{x^a_{qij}}$. $\boldsymbol{\beta_{qm}}$ and $\boldsymbol{v^a_{qmi}}$ are vectors of fixed-effect parameters and random-effect parameters for the q^{th} latent factor, respectively. Within the latent principal stratum ($C_i = m$), the treatment received is either deterministic (for never-takers) or randomized (for compliers) (Imbens & Rubin 2015). This means the treatment receipt is uncorrelated with the error term ϵ^a_{qmij} in Equation 3.2, permitting CACE estimation using standard regression methods if C_i is fully observed. The challenge is that the principal stratum is unobserved for individuals randomized to the control group. Therefore, although compliers are observed in the treatment group ($D_i(A_i = 1) = 1$), these compliers alone do not permit CACE estimation. Furthermore, comparing compliers and noncompliers in the treatment group typically yields biased treatment effect estimation since they come from different principal strata and so are not comparable because of the self-selection nature of compliance.

In the level-2 model, the random effects $\boldsymbol{v^a_{qmi}}$ are used to model the correlation of repeated measurements of the q^{th} latent factor U^a_{qmij}. We assume $\boldsymbol{v^a_{qmi}} \sim N(0, \Sigma_{qv})$ and $\boldsymbol{v^a_{qmi}} \perp\!\!\!\perp \boldsymbol{v^a_{hmi}}, q \neq h$. Furthermore, $\epsilon^1_{qmij} \perp\!\!\!\perp \epsilon^0_{qmij}$, $\boldsymbol{v^1_{qmi}} \perp\!\!\!\perp \boldsymbol{v^0_{qmi}}$ and $\epsilon^a_{qmij} \perp\!\!\!\perp \boldsymbol{v^a_{qmi}}$. To ensure the model identifiability, we make the following assumptions (Assumption 8 in Table 3.1). Since the level-1 part of our multi-level model already includes intercepts ($\boldsymbol{\lambda_{m0}} = (\lambda_{m01}, \cdots, \lambda_{m0K})^T$) and individual-specific random intercepts ($\boldsymbol{b^a_{mi}} = (b^a_{mi1}, \cdots, b^a_{miK})^T$), $\boldsymbol{x^a_{qij}}$ and $\boldsymbol{z_{qij}}$ do not include intercepts so that the model can be identified. Furthermore, we also assume $\epsilon^a_{qmij} \sim N(0, 1)$ which fixes the scale of the latent factor U^a_{qmij} for the sake of identifiability.

We are interested in the principal causal effects (PCEs) on the latent factors $U^a_{qmij} (q = 1, \cdots, Q)$ because these latent factors correspond to the underlying constructs. The PCE on the q^{th} latent factor U^a_{qmij} within the compliance pattern m is defined as

$$E(U^1_{qmij}|C_i = m) - E(U^0_{qmij}|C_i = m) = \boldsymbol{x^1_{qij}\beta_{qm}} - \boldsymbol{x^0_{qij}\beta_{qm}} \tag{3.3}$$

In Eqn 3.3, the terms including random effects $\boldsymbol{v^a_{qmi}}$ and error term ϵ^a_{qmij} disappear because the expectations of these terms equal 0 as explained above. To study the causal effect of assignment A_i on latent factors among compliers, the PCE is obtained by comparing the expectation of q^{th} latent factor U^1_{qcij} for compliers in the treatment group with the expectation of q^{th} latent factor U^0_{qcij} for compliers in the control group.

Here we note that treatment assignment A_i does not influence potential outcomes $\boldsymbol{Y_i(D)}$ directly under Assumption 4. Since the latent factor $\boldsymbol{U_i^a}$ captures the characteristics of potential outcomes $\boldsymbol{Y_i(D)}$, treatment assignment A_i does not influence latent factor $\boldsymbol{U_i^a}$ directly either.

Using the matrix notation, models (3.1) and (3.2) can be succinctly written as

$$\begin{aligned}\boldsymbol{Y_i^a} \;\Big|\; (C_i = m, \boldsymbol{U_{mi}^a}) &= \boldsymbol{\lambda_{m0}} \otimes 1_{J+1} + (\Lambda \otimes I_{J+1})\boldsymbol{U_{mi}^a} + \boldsymbol{b_{mi}^a} \otimes 1_{J+1} + \boldsymbol{e_{mi}^a}, \\ \boldsymbol{U_{mi}^a} &= X_{i,a}\boldsymbol{\beta_m} + Z_i\boldsymbol{v_{mi}^a} + \boldsymbol{\epsilon_{mi}^a} \end{aligned} \tag{3.4}$$

where $\boldsymbol{Y_i^a} = \{Y_{ijk}^a : j = 0, \cdots, J; k = 1, \cdots, K\}$, $\boldsymbol{e_{mi}^a} = \{e_{mijk}^a : j = 0, \cdots, J; k = 1, \cdots, K\}$, $\boldsymbol{U_{mi}^a} = \{U_{qmij}^a : q = 1, \cdots, Q; j = 0, \cdots, J\}$, $\boldsymbol{\beta_m} = \{\beta_{qmpr} : p = 0, \cdots, P; r = 0, \cdots, R\}$, where P and R depend on the forms of model (3.2). $\boldsymbol{v_{mi}^a} = (\boldsymbol{v_{1mi}^a}^T, \cdots, \boldsymbol{v_{Qmi}^a}^T)^T$. $X_{i,a}$ and Z_i are design matrices for fixed effects and random effects in the model for $\boldsymbol{U_{mi}^a}$. $\boldsymbol{\epsilon_{mi}^a} = \{\epsilon_{qmij}^a : q = 1, \cdots, Q; j = 0, \cdots, J\}$.

After combining the above level-1 model and level-2 model, an overall model for the potential outcomes for individual i with compliance type m can be obtained as

$$\begin{aligned}\boldsymbol{Y_i^a} \;\Big|\; (C_i = m, \boldsymbol{v_{mi}^a}, \boldsymbol{b_{mi}^a}) &= \boldsymbol{\lambda_{m0}} \otimes 1_{J+1} + (\Lambda \otimes I_{J+1})(X_{i,a}\boldsymbol{\beta_m}) \\ &\quad + (\Lambda \otimes I_{J+1})(Z_i\boldsymbol{v_{mi}^a} + \boldsymbol{\epsilon_{mi}^a}) + \boldsymbol{b_{mi}^a} \otimes 1_{J+1} + \boldsymbol{e_{mi}^a}\end{aligned}$$

By combining random effects and error terms in both the level-1 model and level-2 model, the marginal distribution for $\{\boldsymbol{Y_i^a} \big| C_i = m\}$ can be obtained as

$$\boldsymbol{Y_i}^a \Big| C_i = m, \boldsymbol{X_i} \sim MVN_{\boldsymbol{\beta_m}, \boldsymbol{\lambda_{m0}}, \boldsymbol{\lambda}, \boldsymbol{\sigma_v}, \boldsymbol{\psi_m}}(\boldsymbol{\mu_{mi}^a}, \Sigma_{mi}) \tag{3.5}$$

where $\boldsymbol{\mu_{mi}^a} = \boldsymbol{\lambda_{m0}} \otimes 1_{J+1} + (\Lambda \otimes I_{J+1})(X_{i,a}\boldsymbol{\beta_m})$, $\Sigma_{mi} = (\Lambda \otimes I_{J+1})Z_i\Sigma_v[(\Lambda \otimes I_{J+1})Z_i]^T + (\Lambda \otimes I_{J+1})(\Lambda \otimes I_{J+1})^T + diag(\xi_{m1}, \cdots, \xi_{mK}) \otimes (1_{J+1}1_{J+1}^T) + diag(\tau_{m1}^2, \cdots, \tau_{mK}^2) \otimes I_{J+1}$. Σ_v is the variance-covariance matrix of random effects $\boldsymbol{v_{mi}^a}$ and $\boldsymbol{\sigma_v}$ is the collection of unique parameters in Σ_v. $\boldsymbol{\lambda}$ is the collection of all the elements in matrix Λ. $\boldsymbol{\psi_m} = (\xi_{m1}, \cdots, \xi_{mK}, \tau_{m1}^2, \cdots, \tau_{mK}^2)^T$.

Here we assume potential outcomes $\boldsymbol{Y_{ij}^1}$ and $\boldsymbol{Y_{ij}^0}$ are independent given compliance type, covariates and parameters. Because we never observe both $\boldsymbol{Y_{ij}^1}$ and $\boldsymbol{Y_{ij}^0}$ at the same time, the likelihood function of the observed data does not depend on the correlation between potential outcomes. Thus, the correlation between potential outcomes is unimportant under the likelihood-based approach (see Page 181 in Chapter 8 of Imbens & Rubin 2015, Hirano et al. 2000).

Model for compliance type C_i.

For the probability of being a complier, we use a logistic regression model

$$p_{ci} = Pr(C_i = c \;\Big|\; \boldsymbol{W_i} = \boldsymbol{w_i}, \boldsymbol{\eta}) = \frac{exp(\boldsymbol{w_i}'\boldsymbol{\eta})}{1 + exp(\boldsymbol{w_i}'\boldsymbol{\eta})}, \tag{3.6}$$

where $\boldsymbol{W_i}$ is the collection of baseline covariates for individual i and $\boldsymbol{\eta}$ is the collection of coefficients for corresponding covariates.

3.2.3 Inference

Based on $\boldsymbol{A_{obs}}$ and $\boldsymbol{D_{obs}}$, there are three possible observed patterns of (a_i, d_i): $(1,1)$, $(1,0)$, $(0,0)$. Let $\mathcal{S}(1,1)$, $\mathcal{S}(1,0)$ and $\mathcal{S}(0,0)$ denote the subsets of units exhibiting each pattern separately. This implies $\mathcal{S}(1,1)$ and $\mathcal{S}(1,0)$ include the compliers and never-takers, respectively, in the treatment group and $\mathcal{S}(0,0)$ represents a mixture of compliers and never-takers in the control group. Let $\boldsymbol{\pi} = (\boldsymbol{\beta_c}, \boldsymbol{\beta_n}, \boldsymbol{\lambda}, \boldsymbol{\lambda_{c0}}, \boldsymbol{\lambda_{n0}}, \boldsymbol{\sigma_v}, \boldsymbol{\psi_c}, \boldsymbol{\psi_n}, \boldsymbol{\eta})$, the likelihood function based on observed data for all participants in the study is

$$\begin{aligned}
&\mathcal{L}(\boldsymbol{\pi}; \boldsymbol{Y_{obs}}, \boldsymbol{D_{obs}}, \boldsymbol{A_{obs}} \;\Big|\; \boldsymbol{X}) \\
&= \prod_i \iiiint f(\boldsymbol{y_i^1}, \boldsymbol{y_i^0} \Big| \boldsymbol{U_{mi}^1}, \boldsymbol{U_{mi}^0}, D_i(1), D_i(0), \boldsymbol{X_i}; \boldsymbol{\lambda}, \boldsymbol{\lambda_{c0}}, \boldsymbol{\lambda_{n0}}, \boldsymbol{\psi_c}, \boldsymbol{\psi_n}) \\
&\quad f(\boldsymbol{U_{mi}^1}, \boldsymbol{U_{mi}^0} \Big| D_i(1), D_i(0), \boldsymbol{X_i}; \boldsymbol{\beta_c}, \boldsymbol{\beta_n}, \boldsymbol{\sigma_v}) f(D_i(1), D_i(0) \Big| \boldsymbol{W_i}; \boldsymbol{\eta}) d\, \boldsymbol{U_{mi}^1} d\, \boldsymbol{U_{mi}^0} d\, \boldsymbol{Y_i^{mis}} d\, D_i^{mis} \\
&= \prod_i \iint f(\boldsymbol{y_i^1}, \boldsymbol{y_i^0} \Big| D_i(1), D_i(0), \boldsymbol{X_i}; \boldsymbol{\beta_c}, \boldsymbol{\beta_n}, \boldsymbol{\lambda}, \boldsymbol{\lambda_{c0}}, \boldsymbol{\lambda_{n0}}, \boldsymbol{\sigma_v}, \boldsymbol{\psi_c}, \boldsymbol{\psi_n}) \\
&\quad f(D_i(1), D_i(0) \Big| \boldsymbol{W_i}; \boldsymbol{\eta}) d\, \boldsymbol{Y_i^{mis}} d\, D_i^{mis} \\
&= L_{11} \times L_{10} \times L_{00}
\end{aligned} \tag{3.7}$$

where

$$\begin{aligned}
L_{11} &= \prod_{\{i \in S(1,1)\}} p_{ci} \frac{1}{(2\pi)^{\frac{J(K+1)}{2}} |\Sigma_{ci}|^{\frac{1}{2}}} \exp\left\{ -\frac{1}{2} \left(\boldsymbol{y_{obs,i}} - \boldsymbol{\mu_{ci}^1} \right)^T \Sigma_{ci}^{-1} \left(\boldsymbol{y_{obs,i}} - \boldsymbol{\mu_{c_i}^1} \right) \right\} \\
L_{10} &= \prod_{\{i \in S(1,0)\}} (1 - p_{ci}) \frac{1}{(2\pi)^{\frac{J(K+1)}{2}} |\Sigma_{ni}|^{\frac{1}{2}}} \exp\left\{ -\frac{1}{2} \left(\boldsymbol{y_{obs,i}} - \boldsymbol{\mu_{ni}^1} \right)^T \Sigma_{ni}^{-1} \left(\boldsymbol{y_{obs,i}} - \boldsymbol{\mu_{ni}^1} \right) \right\} \\
L_{00} &= \prod_{\{i \in S(0,0)\}} [p_{ci} \frac{1}{(2\pi)^{\frac{J(K+1)}{2}} |\Sigma_{ci}|^{\frac{1}{2}}} \exp\left\{ -\frac{1}{2} \left(\boldsymbol{y_{obs,i}} - \boldsymbol{\mu_{c_i}^0} \right)^T \Sigma_{ci}^{-1} \left(\boldsymbol{y_{obs,i}} - \boldsymbol{\mu_{c_i}^0} \right) \right\} \\
&\qquad + (1 - p_{ci}) \frac{1}{(2\pi)^{\frac{J(K+1)}{2}} |\Sigma_{ni}|^{\frac{1}{2}}} \exp\left\{ -\frac{1}{2} \left(\boldsymbol{y_{obs,i}} - \boldsymbol{\mu_{ni}^0} \right)^T \Sigma_{ni}^{-1} \left(\boldsymbol{y_{obs,i}} - \boldsymbol{\mu_{ni}^0} \right) \right\}],
\end{aligned}$$

$\boldsymbol{\mu_{mi}^a}$ and Σ_{mi} are defined as shown in Eqn 3.5. By combing the level-1 model and level-2 model, we obtain the above closed-form simplified marginal likelihood that integrates out the latent factors $\boldsymbol{U_{mi}^a}$. That is, in the likelihood function (Eqn 3.7), the second equality holds by applying the conclusion shown in Eqn 3.5. The observed data log-likelihood function can

be maximized using the Quasi-Newton algorithm implemented in the R function `optim()`, which yields the maximum likelihood estimates (MLEs) of model parameters. The variance of the estimates can be obtained via the inverse Hessian matrix of the log-likelihood function evaluated at the MLEs.

3.3 Simulation Study

In this section, we conduct simulation studies to examine the performance of the model proposed in Section 3.2. When generating the simulated dataset, we consider the following level-1 model with two ($Q = 2$) latent factors (U^a_{1mij} and U^a_{2mij}):

$$y^a_{ijk} \Big| (C_i = m, U^a_{1mij}, U^a_{2mij}, b^a_{mik}) = \lambda_{m0k} + \lambda_{k1} U^a_{1mij} + \lambda_{k2} U^a_{2mij} + b^a_{mik} + e^a_{mijk}, \quad (3.8)$$

with the following level-2 model for the q^{th} latent factor U^a_{qmij}:

$$U^a_{qmij} = \beta_{qm10} t_{ij} + \beta_{qm20} t^2_{ij} + \beta_{qm01} D_i(a) + \beta_{qm11} D_i(a) t_{ij} + \beta_{qm21} D_i(a) t^2_{ij} + v^a_{qm1i} t_{ij} + \epsilon^a_{qmij} \quad (3.9)$$

where the index of latent factor $q = 1$ or 2. For these two latent factors, we assume that the covariates for fixed effects include the linear time trend (t_{ij}), quadratic time trend (t^2_{ij}), treatment receipt $D_i(a)$, and the interactions between the receipt of treatment and these time trends ($D_i(a)t_{ij}$ and $D_i(a)t^2_{ij}$). Specifically, the term $\beta_{qm01} D_i(a)$ captures the mean baseline differences in the q^{th} latent factor within compliance type m between the treatment group and the control group. For a randomized controlled trial, the baseline difference within compliance type m between treatment group and control group is expected to be negligible for two factors. Therefore, we set $\beta_{qm01} = 0$ for $q = 1$ or 2 in the simulation studies. To simplify the simulation setting and for the comparison convenience, we did not incorporate additional baseline covariates in the level-2 model. The covariates for random effects v^a_{qm1i} contain linear trend only. For the sake of identification, the intercept and random intercept are removed as these parameters already appear in the level-1 model (Assumption 8 in Table 3.1). When generating the compliance status C_i in Eqn 3.6, we do not include covariates and set $p_{ci} = p_c = 0.3$. True values of other parameters for fixed effects, random effects and the error terms are listed in section 3.6.1 in Supplemental Information.

In the simulation study, six outcomes were simulated at three time points for each individual based on the model in Eqns 3.8 and 3.9. For the factor loading matrix Λ, we set the first, second, third and fifth outcomes to load on the first factor and the remaining two outcomes to load on the second factor in the process of generating the data. This means the fourth element and the sixth element of the first column and the first three elements and the fifth element of the second column of the factor loading matrix (Λ) are fixed at 0

as shown below:

$$\Lambda = \begin{pmatrix} \lambda_{11} & 0 \\ \lambda_{21} & 0 \\ \lambda_{31} & 0 \\ 0 & \lambda_{42} \\ \lambda_{51} & 0 \\ 0 & \lambda_{62} \end{pmatrix}. \tag{3.10}$$

Under the setting, we conducted two simulation studies to examine the identification of the proposed model. In the first simulation study, we focus on confirmatory analysis, which means the structure of the factor loading matrix Λ is known before conducting the data analysis. This occurs when researchers have the knowledge of which outcomes load on a specific factor, although the factor loading values are unknown and need to be estimated.

In the second simulation study, we focus on exploratory analysis. Under exploratory analysis, we assume no prior knowledge of which outcomes load on each factor. To fix the rotation of the matrix Λ, we imposed the restriction of $\lambda_{kq} = 0$, $q > k$ as noted above. That is, the exploratory factor analysis fixes λ_{12} to be 0 as shown in Eqn 3.11.

$$\Lambda = \begin{pmatrix} \lambda_{11} & 0 \\ \lambda_{21} & \lambda_{22} \\ \lambda_{31} & \lambda_{32} \\ \lambda_{41} & \lambda_{42} \\ \lambda_{51} & \lambda_{52} \\ \lambda_{61} & \lambda_{62} \end{pmatrix}. \tag{3.11}$$

When estimating the parameters in the factor loading matrix Λ in Eqns 3.10 and 3.11, one element in each column should be restricted to being positive so that parameters can be completely identifiable. Here we assume λ_{11} and λ_{62} are positive. Otherwise, considering $-\Lambda$ and $-\boldsymbol{U^a_{mij}}$ gives the same value of the likelihood function.

Based on Eqn 3.9, the principal causal effects (PCEs) on latent factors U^a_{qmij} $(q = 1, 2)$ within compliance pattern $C_i = c$ are

$$E(U^1_{qcij}|C_i = c) - E(U^0_{qcij}|C_i = c) = \beta_{qc01} + \beta_{qc11} t_{ij} + \beta_{qc21} t^2_{ij}$$

Because $\beta_{qc01} = 0$ for $q = 1, 2$ in the simulation setting, β_{qc11} and β_{qc21} jointly capture the PCEs. Thus, we evaluate the performance of the proposed procedure for estimating β_{qc11} and β_{qc21}. In addition, we evaluate the estimation accuracy of the elements in the factor loading matrix (i.e., λ_{k1}, λ_{k2}). Table 3.2 and Table 3.3 report the results related to parameters β_{qc11} and β_{qc21} for the confirmatory analysis and exploratory analysis, respectively. In both tables, results are obtained based on 500 repetitions when sample size equals 500. We observe sample means of parameter estimates are close to their corresponding true values. Sample

means of standard error estimates obtained by using the Fisher information are close to their corresponding sample standard deviations of estimates. This means model parameters can be recovered very well under both confirmatory analysis and exploratory analysis.

Table 3.2: Estimation accuracy under confirmatory analysis based on 500 repetitions when sample size equals 500

Parameter	True value	Sample mean	$sse^{\dagger}$ ($sm_{se}^{\ddagger}$)
Treatment effect			
β_{1c11}	0.5	0.506	0.482 (0.480)
β_{1c21}	-1	-1.009	0.224 (0.214)
β_{2c11}	-1	-0.996	0.810 (0.765)
β_{2c21}	1	1.016	0.388 (0.363)
Factor loading matrix Λ			
λ_{11}	7	6.956	0.270 (0.269)
λ_{21}	3	2.988	0.135 (0.134)
λ_{31}	8	7.963	0.301 (0.302)
λ_{41}	0	-	-
λ_{51}	8	7.954	0.294 (0.293)
λ_{61}	0	-	-
λ_{12}	0	-	-
λ_{22}	0	-	-
λ_{32}	0	-	-
λ_{42}	6	5.926	0.531 (0.533)
λ_{52}	0	-	-
λ_{62}	5	4.930	0.423 (0.423)

$\dagger sse$: sample standard deviation of estimates
$\ddagger sm_{se}$: sample mean of standard error estimates obtained based on the Fisher information

Table 3.2 and Table 3.3 also show the estimation results for the 2-factor loading matrix. For both confirmatory and exploratory analysis, the sample means of the estimates for the entries in the loading matrix are close to their corresponding true values and sample means of their standard error estimates obtained by using the Fisher information approximate their corresponding sample standard deviations of estimates. In exploratory analysis, since we impose less restriction on Λ and only restrict $\lambda_{kq} = 0$ when $q > k$ to fix the rotation of the factor loading matrix Λ, the matrix Λ produced based on this restriction may be hard to interpret. In this case, factor rotation techniques are often used to improve interpretability of the factor loading matrix. There are two types of rotation: orthogonal rotation and oblique rotation (Everitt & Hothorn 2011). Orthogonal rotation creates uncorrelated rotated factors, while oblique rotation allows correlated factors. For orthogonal rotation, varimax rotation is commonly used. For oblique rotation, oblimin and promax rotation are commonly used. Usually, we wish to find a rotation matrix which ensures each endpoint has a high loading on only one latent factor. These rotation techniques can be applied to achieve a simple structure

Table 3.3: Estimation accuracy under exploratory analysis based on 500 repetitions when sample size equals 500

Parameter	True value	Sample mean	$sse^{\dagger}$ ($sm_{se}^{\ddagger}$)
Treatment effect			
β_{1c11}	0.5	0.505	0.481 (0.481)
β_{1c21}	-1	-1.008	0.224 (0.215)
β_{2c11}	-1	-0.997	0.808 (0.766)
β_{2c21}	1	1.017	0.388 (0.367)
Factor loading matrix Λ			
λ_{11}	7	6.959	0.270 (0.258)
λ_{21}	3	2.988	0.135 (0.133)
λ_{31}	8	7.963	0.303 (0.290)
λ_{41}	0	-0.003	0.293 (0.294)
λ_{51}	8	7.954	0.293 (0.285)
λ_{61}	0	0.006	0.259 (0.264)
λ_{12}	0	-	-
λ_{22}	0	0.002	0.108 (0.103)
λ_{32}	0	-0.004	0.152 (0.160)
λ_{42}	6	5.923	0.533 (0.506)
λ_{52}	0	-0.010	0.141 (0.149)
λ_{62}	5	4.927	0.423 (0.414)

$\dagger sse$: sample standard deviation of estimates
$\ddagger sm_{se}$: sample mean of standard error estimates obtained based on the Fisher information.

which allows the rotated matrix to be more interpretable. In our simulation study, we have applied varimax rotation of the estimated loading matrix in the exploratory analysis.

It's worth noting the benefits of confirmatory analysis where the structure of factor loading matrix is known. Compared to exploratory analysis (Table 3.3), the means of standard error estimates (sm_{se}) for non-zero components in Λ are noticeably closer to the corresponding sample standard deviations of estimates (sse) under confirmatory analysis (Table 3.2). Specifically, the differences between sm_{se} and sse are all around 0.001 in Table 3.2, while the differences are around 0.01 on average in Table 3.3. Interestingly, the variability of non-zero components in Λ (sse) are comparable between confirmatory analysis (Table 3.2) and exploratory analysis (Table 3.3). This suggests the restriction of $\lambda_{12} = 0$ imposed in exploratory analysis is adequate to identify the factor structure.

We further conducted a power analysis to compare the latent-factor multivariate CACE analysis to univariate CACE analysis in terms of statistical power to detect intervention effects on multiple study endpoints. Instead of analyzing multiple endpoints jointly as conducted in latent-factor multivariate CACE analysis, univariate CACE analysis is conducted for each endpoint separately. Thus, univariate CACE analysis ignores the potential correlation among multiple outcomes collected simultaneously. We describe in detail the

model specification for the univariate CACE analysis in Supplemental Information (section 3.6). The data is generated as described above except for comparison convenience, we set $\beta_{qc21} = 0$ for $q = 1, 2$ (Eqn 3.9). Both the latent-factor multivariate CACE analysis and the univariate CACE analysis include t_{ij}, t_{ij}^2 and $D_i(a)t_{ij}$ as predictors in the model. For compliers, the coefficient for $D_i(a)t_{ij}$ is the difference in outcomes between the treatment group and the control group, which captures the CACE in univariate CACE analysis. In the latent-factor MCACE analysis, one single global likelihood ratio test is conducted. The null hypothesis for the global test is $\beta_{qc11} = 0$ for $q = 1, 2$, which means there are no treatment effects for both latent factors among compliers. For univariate CACE analysis, the null hypothesis is that the coefficient of the predictor $D_i(a)t_{ij}$ equals 0 for a specific outcome, which implies no treatment effect is detected for this outcome. Therefore, in univariate CACE analysis, six likelihood ratio tests are conducted for each outcome because these outcomes are analyzed separately. Because power is defined as the probability of rejecting the null hypothesis when the null hypothesis is false, we calculate the proportion of rejecting the null hypothesis over 500 simulated datasets. The significance level for the single global test of the null hypothesis of no intervention effects on both latent factors is set as 0.05 for latent-factor multivariate CACE analysis. For univariate CACE analysis, the significance level for each individual test is adjusted as 0.05/K, where $K = 6$ because there are 6 tests. This adjustment is based on the Bonferroni correction so that the overall familywise Type I error rate can be well controlled at the 0.05 level with multiple testing. When calculating the power for univariate analysis, we reject the null hypothesis if treatment effect is detected for at least one outcome at the significance level 0.05/6.

Figure 3.2 shows power curves for the latent-factor multivariate CACE model and univariate CACE analysis when sample size equals 500. To ease the result presentation, we set the values of β_{qc11} across q to be the same and the common value varies from 0 to 5. When $\beta_{qc11} = 0$ for any q, the probability of rejecting the null hypotheses is the Type I error rate. We observe the Type I error rates are around 0.05 for both the latent-factor multivariate CACE model and the univariate CACE model (Figure 3.2), which means Type I error rates are controlled well under both models. As the value of β_{qc11} increases, we observe that the power for the latent-factor multivariate CACE model (the solid line in Figure 3.2) approaches 1 much faster than that for the univariate CACE model (the dashed line in Figure 3.2). The power gain by the latent-factor multivariate CACE model relative to the univariate CACE model can be substantial. For instance, the power increases from 0.25 under the univariate CACE analysis to 0.97 under the latent-factor multivariate CACE model when $\beta_{qc11} = 1$. Therefore, multivariate CACE analysis based on the latent-factor MCACE model outperforms univariate CACE analysis in terms of statistical power for hypothesis testing. This difference is due to correlations across multiple endpoints being ignored by univariate CACE analysis, while the latent-factor multivariate CACE analysis captures the

correlations across endpoints by latent factors. Therefore, under the latent-factor MCACE model, more information is used when making inference, which increases the study power.

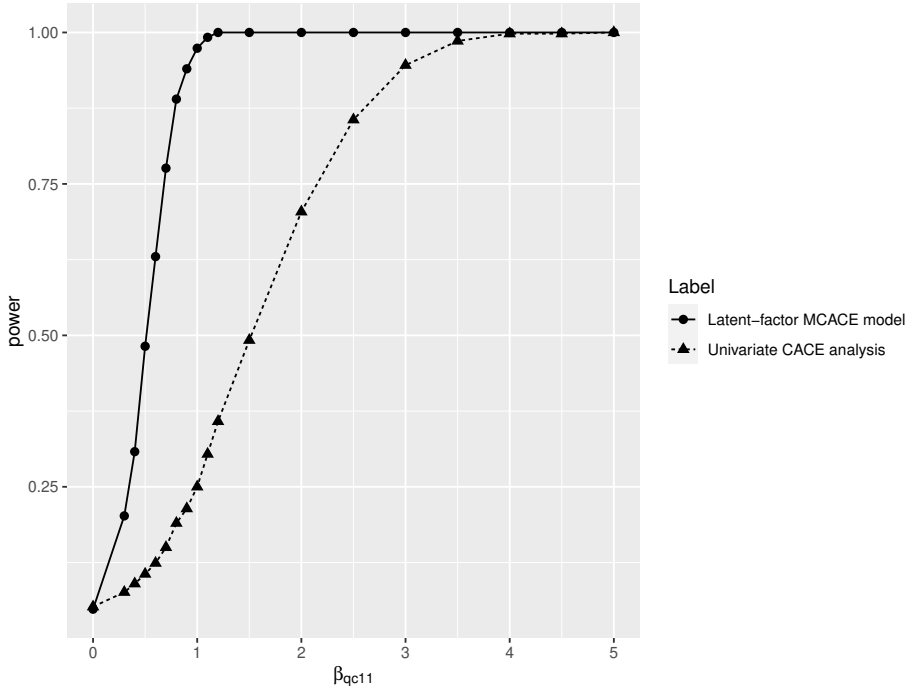

Figure 3.2: Power analysis, based on 500 simulated datasets.

3.4 Application

We applied the proposed model to the Arthritis Health Journal (AHJ) study. AHJ is a patient-centered online tool for rheumatoid arthritis (RA) patients to better manage their disease by actively monitoring their symptoms and tracking disease activity. A total of 94 participants were recruited and randomly assigned to the treatment group (n=45) and the control group (n=49). Patients in the treatment group received access to the AHJ tool immediately, while patients in the control group had to wait for six months before getting access to the AHJ tool. The primary analysis of the study focused on the data from the first six months. Some participants randomized to the treatment group may not use the tool or use it rarely which leads to noncompliance behavior. The discussions with doctors and patients suggest that using the tool once per month is necessary to produce effects. Therefore, compliers are defined as patients who would use the tool at least one time per month on average if assigned to the treatment group (Guo et al. 2022).

Participants are evaluated by online questionnaires every 3 months (baseline, the third month, the sixth month). Baseline questionnaires collected information about demographics (age, gender, education) and disease information. Follow-up questionnaires evaluated the

frequency of using the tool, consumer effectiveness attributes, self-efficacy and satisfaction with care. In this study, six endpoints are collected to evaluate the treatment efficacy of AHJ tool on underlying constructs (e.g., self-efficacy and satisfaction with care): **1. effective consumer 17 scale**, the average score of 17 items about how participants manage their disease on a 0 to 100 scale with 100 indicating "most confident"; **2. manage symptoms scale**, the average score of 5 items about how patients manage their symptoms on a 0 to 10 scale with 10 indicating "totally confident"; **3. manage disease in general scale**, the average score of 5 items about how patients manage their disease in general on a 0 to 10 scale with 10 indicating "totally confident"; **4. communicate with physician scale**, the average score of 3 items about patients' confidence in communicating with their rheumatologists on a 0 to 10 scale with 10 indicating "totally confident"; **5. partners in health scale**, the average score of 11 items about patients' knowledge of disease and treatment on a 0 to 80 scale with 80 indicating "poor self-management"; **6. satisfaction with various aspects of medical care**, the average score of 8 items about their satisfaction with various aspects of medical care on a 0 to 10 scale with 10 indicating "completely satisfied". All six endpoints are the averages of several individual items and as a result take continuous values. Six outcomes are rescaled to a 0 to 100 scale so that these endpoints have comparable variances. The direction of the fifth outcome is also adjusted so that a higher value represents a beneficial result for all these six endpoints. The rescaling of outcomes eases the interpretation of estimation results but does not influence statistical inference.

Furthermore, although the number of participants is only moderately large (n=94), six outcomes were collected at three time points for participants in this study. Therefore, there are eighteen individual-level outcomes in the study. Under the MCACE model, all these observations (six outcomes across three time points per person) are considered jointly. Therefore, the total number of observations used to estimate the model is much larger than the number of study participants.

The interest in this study is to (1) evaluate the effectiveness of the AHJ tool on underlying constructs (e.g., patients' self-efficacy in disease management and the effectiveness in shared decision-making); and (2) determine how covariates predict the compliance behavior. The characteristics of six endpoints suggest some possible latent structures. First, because the six endpoints capture different perspectives of the effect of using the tool, it can be a good starting point to consider one latent factor only. Under this assumption, all six endpoints load on only one latent factor, which captures the overall treatment effect. Thus, we can test whether the AHJ helps patients manage their disease or not based on the common latent factor.

Alternatively, the design of these six endpoints suggests the other more plausible structure which involves two latent factors. Substantively, in the AHJ study, effective consumer 17 scale (the 1st endpoint), manage symptoms scale (the 2nd endpoint), manage disease in general scale (the 3rd endpoint), and partners in health scale (the 5th endpoint) were

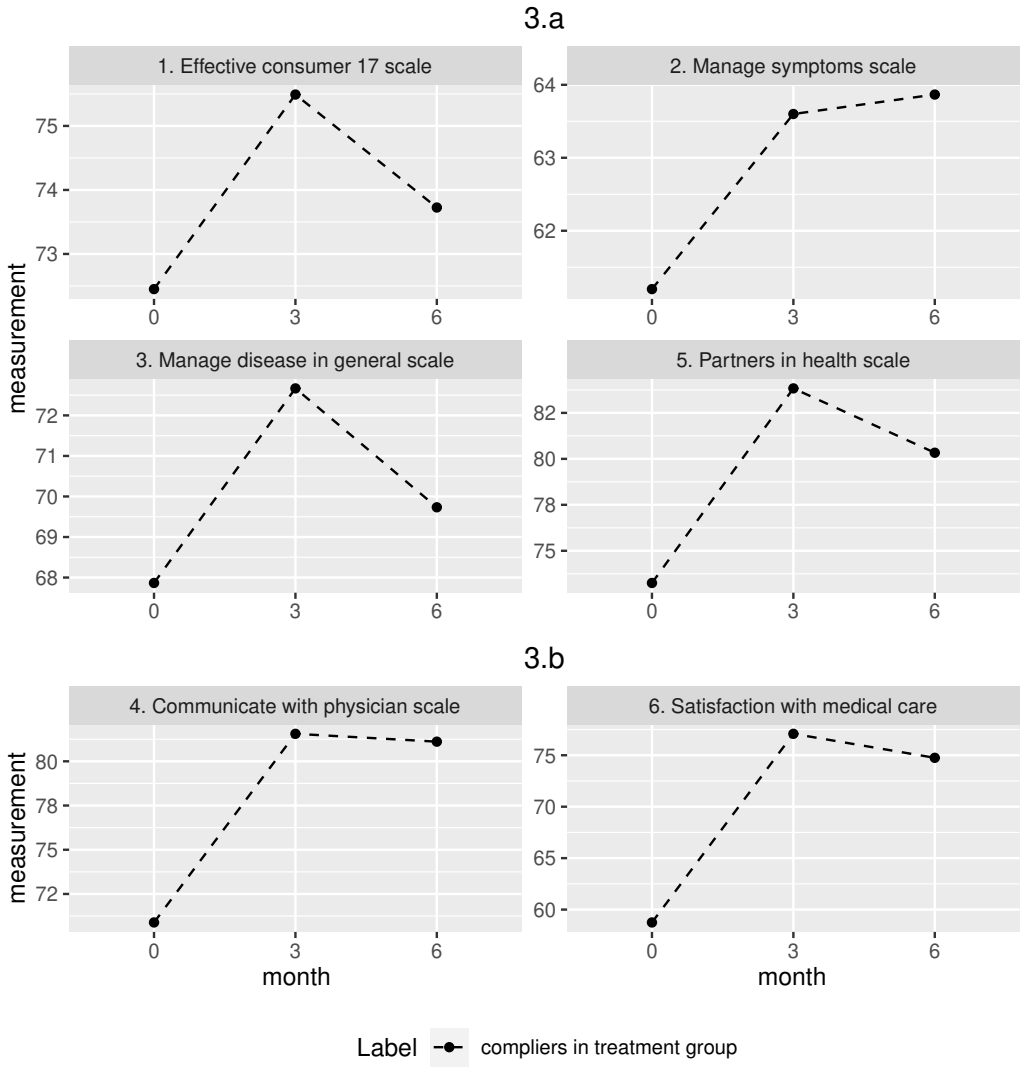

Figure 3.3: Plot 3.3.a shows outcome means of the compliers in the treatment group for the 1st, 2nd, 3rd and 5th endpoints; plot 3.3.b shows outcome means for the 4th and 6th endpoints.

used to measure self-efficacy. Interaction with health care providers was measured by the communicate with physician scale (the 4th endpoint) and satisfaction with various aspects of medical care (the 6th endpoint). The two-group structure of these endpoints based on their substantive meanings is consistent with the trends shown in Figure 3.3, which justifies a model with two latent constructs. Figure 3.3.a shows the outcome means in the compliers in the treatment group for the first three endpoints and the fifth endpoint. Figure 3.3.b shows the outcome means for the fourth and sixth endpoints. We observe that the outcome trajectories over time for the compliers in the treatment group are similar among endpoints within each of the two panels (3.3.a and 3.3.b) of Figure 3.3. It's worth noting that the

trends for the fourth and sixth endpoints are almost identical in Figure 3.3.b. In Figure 3.3.a, the trends are similar among outcomes except for the second one. However, the second outcome is still grouped into the Figure 3.3.a because it is closely related to self-efficacy conceptually.

Specifically, in a confirmatory factor analysis, the fourth and sixth outcomes can be grouped together to indicate one latent construct that measures the participants' knowledge and the ability to interact with healthcare providers. The remaining endpoints are grouped to indicate the other latent construct that measures self-efficacy in disease management. Therefore, the first three endpoints and fifth endpoint load on the first latent factor, while the fourth and sixth endpoints load on the second latent factor. When considering two latent factors, the level-1 model is specified as Eqn 3.1 where $Q = 2$ and $K = 6$. The level-2 model is specified as Eqn 3.2 where quadratic time trends are included (shown in Eqn 3.12 below). Within the level-2 model, we set the baseline difference between compliers in the treatment group and compliers in the control group to be null ($\beta_{qc01} = 0$ for $q = 1$ or 2 in Eqn 3.12) since this is a randomized controlled trial. Furthermore, the random effects for the second latent factor are removed ($v^a_{2m1i} = 0$ and $v^a_{2m2i} = 0$ for $m = c$ or n in Eqn 3.12) based on the likelihood ratio test (Berkhof & Snijders 2001).

$$U^a_{qmij} = \beta_{qm10}t_{ij}+\beta_{qm20}t^2_{ij}+\beta_{qm01}D_i(a)+\beta_{qm11}D_i(a)t_{ij}+\beta_{qm21}D_i(a)t^2_{ij}+v^a_{qm1i}t_{ij}+v^a_{qm2i}t^2_{ij}+\epsilon^a_{qmij} \tag{3.12}$$

The compliance model is specified as stated in Eqn 3.6. The compliance model (Eqn 3.6) includes four binary baseline covariates: Early disease (1 indicates early disease (0-2 years) and 0 indicates late disease ($\geq$ 2 years)), High disease activity (1 indicates high disease activity (high RAPID4 values) and 0 indicates low disease activity (remission, moderate/low RAPID4 values)), Male (1 indicates male and 0 indicates female), Older age (1 indicates above the median age (54.5) and 0 otherwise).

Additionally, we also conducted a model selection to determine whether it is necessary to consider a model with three factors or not. Since the outcome trend for the 2nd outcome (manage symptoms scale) is different from mean trends for other outcomes in Figure 3.3.a, the 1st, 3rd and 5th outcomes are grouped together to represent the underlying construct, self-efficacy in disease management; the 2nd outcome is grouped separately to indicate a new underlying construct representing itself. Similar to the two-group structure, the remaining outcomes (4th and 6th outcomes) are grouped together to indicate the underlying construct, interaction with health care providers. Therefore, the first, third, and fifth outcomes load on the first latent factor, the second outcome loads on the second latent factor and the remaining outcomes load on the third factor. Table 3.4 lists the maximums of the log-likelihood functions and the AIC values for models with different number of factors. It turns out that the model with two latent factors gives the minimum AIC value. Therefore,

it's sufficient to incorporate two factors and unnecessary to consider an additional factor in our model.

Table 3.4: Model selection

Number of factors	$logL$	AIC
1	-5915.344	11942.69
2	-5897.497	**11918.99**
3	-5892.786	11921.57

Table 3.5 presents the estimation results obtained under the model assuming two latent factors. The potential outcomes are positively correlated with two latent factors based on the positive estimates of matrix Λ. Because two latent factors are involved, the complier average causal effect on each latent factor at the sixth month is

$$E(U^1_{qmi2}|m = c) - E(U^0_{qmi2}|m = c) = \beta_{qc11} * 2 + \beta_{qc21} * 4 \tag{3.13}$$

Eqn 3.13 is obtained based on Eqn 3.12 when $j = 2$ and $t_{ij} = 2$. According to Table 3.5, the CACEs on the first and second latent factors are 1.392 (SE = 0.632) and 1.745 (SE = 0.660), respectively. Corresponding P-values are 0.028 and 0.008, respectively, based on Wald test. Because the estimates of CACEs for these two latent factors are positive and P-values are significant, these results suggest the AHJ had beneficial causal effects on both self-efficacy and interaction with health care among RA patients who comply with the assigned treatments.

We are also interested in identifying RA patients who are more likely to be a complier. Table 3.6 reports estimation results of the compliance model under the scenario where two latent factors are considered. The coefficients for all covariates except high disease activity are negative. In addition, the coefficients for early disease and high disease activity are statistically significant. The P-value of the coefficient for older age approaches statistically significance. Overall, the estimation results suggest that younger female patients with longer disease duration and high disease activity were more likely to be compliers.

3.5 Discussion

The proposed model introduced latent factors which capture the underlying constructs driving multiple endpoints measured over time, and yield parsimonious estimation of CACEs on these latent constructs in the presence of treatment noncompliance. Results from simulation studies show the model proposed can be identified, and all true values of model parameters are recovered well. Besides, compared with univariate CACE analysis, the power analysis shows a substantial gain in the study power under the latent-factor multivariate CACE model. In the application section, we first conducted a model selection and fixed the num-

Table 3.5: Estimates and standard errors for causal treatment effects in the AHJ study

Two latent factors			
Fixed effects estimation*			
Parameter	Estimate	SE	P-value
β_{1c11}	2.772	1.204	0.021
β_{1c21}	-1.038	0.529	0.050
β_{2c11}	1.900	1.176	0.106
β_{2c21}	-0.514	0.582	0.377
Treatment effect estimation at the sixth month			
	Estimate	SE	P-value
$\text{CACE}_1^\dagger$	1.392	0.632	0.028
$\text{CACE}_2^\dagger$	1.745	0.660	0.008
Estimation of factor loading matrix Λ			
Parameter	Estimate	SE	
λ_{11}	7.918	0.683	
λ_{21}	2.698	0.557	
λ_{31}	9.067	0.674	
λ_{41}	-	-	
λ_{51}	8.553	0.722	
λ_{61}	-	-	
λ_{12}	-	-	
λ_{22}	-	-	
λ_{32}	-	-	
λ_{42}	6.524	1.202	
λ_{52}	-	-	
λ_{62}	6.638	1.236	

$*$: Fixed effects estimation section only shows parameters related to treatment effect estimation.
†: CACE_1 represents CACE for the first factor, self-efficacy; CACE_2 represents CACE for the second factor, interaction with health care providers.

ber of factors at 2. Then we employed the confirmatory analysis and the endpoints are partitioned into two groups based on their characteristics. The first latent factor represents self-efficacy and the second latent factor measures interaction with healthcare providers. For both latent factors, we detected beneficial CACEs. Compared with the findings of Guo et al. (2022), who detected a global CACE over six endpoints, one advantage of the model proposed here is that we are able to determine the direction of treatment effect (beneficial or not) on the latent factor. Specifically, the analysis here shows that the CACEs on two latent factors are beneficial in the application. Furthermore, compared with conducting hypothesis testing for each endpoint individually, our global testing approach is more efficient by combining information over correlated outcomes corresponding to certain underlying construct across all time points and avoiding multiple testing issues which occur if treatment effect

Table 3.6: Estimation results of compliance model

Two latent factors			
Covariates	Estimate	Standard error	P-value
Intercept	-0.825	0.614	0.179
Early disease	-2.389	1.121	0.033
High disease activity	1.420	0.643	0.027
Male	-0.888	0.877	0.311
Older age	-0.938	0.488	0.054

Note: Early disease = 1 if early disease (0-2 years) and = 0 if late disease ($\geq$ 2 years)), high disease activity = 1 if high disease activity (high RAPID4 values) and = 0 if low disease activity (remission, moderate/low RAPID4 values)), Male = 1 if male and = 0 if female, Older age = 1 if above the median age (54.5) and = 0 if otherwise.

is tested on each endpoint individually. Thus, our analysis is able to directly answer the main scientific questions addressed by this RCT and yields novel findings not discovered previously. The compliance behavior also can be predicted based on the logistic regression model. The analysis shows that younger female patients with longer disease duration and high disease activity were more likely to be a complier.

There are also some limitations. In the methodology proposed, we only discuss the continuous responses and assume the responses conditioned on compliance type and covariates follow the multivariate normal distribution. In the future, other distribution assumptions can be considered. Additionally, in real-world RCTs, we may encounter categorical or mixed-type responses. This is beyond the scope of this chapter but future investigation can consider extending the proposed method to discrete and mixed-type responses.

3.6 Supplemental Information

3.6.1 Simulation Setting

In simulation studies, true values for parameters in Eqns 3.8 and 3.9 are listed as below:

$\boldsymbol{\lambda_{c0}} = (58, 60, 61, 59, 60, 58)^T$,

$\boldsymbol{\lambda_{n0}} = (79, 88, 79, 87, 82, 88)^T$,

$$\Lambda = \begin{pmatrix} 7 & 0 \\ 3 & 0 \\ 8 & 0 \\ 0 & 6 \\ 8 & 0 \\ 0 & 5 \end{pmatrix},$$

$\boldsymbol{\beta_c} = (\beta_{1c10}, \beta_{1c20}, \beta_{1c11}, \beta_{1c21}, \beta_{2c10}, \beta_{2c20}, \beta_{2c11}, \beta_{2c21})^T = (-3, 1, 0.5, -1, 2, -1, -1, 1)^T$,

$\boldsymbol{\beta_n} = (\beta_{1n10}, \beta_{1n20}, \beta_{1n11}, \beta_{1n21}, \beta_{2n10}, \beta_{2n20}, \beta_{2n11}, \beta_{2n21})^T = (-3, 2, 0, 0, -2, 1, 0, 0)^T$,
$var(v^a_{1m1i}) = exp(1)$, $var(v^a_{2m1i}) = exp(1.5)$,
$\boldsymbol{\psi_c} = (\xi_{c1}, \cdots, \xi_{c6}, \tau^2_{c1}, \cdots, \tau^2_{c6})^T = (exp(3), exp(5), exp(3), exp(4), exp(3), exp(4), exp(4), exp(4), exp(4), exp(5), exp(4), exp(5))^T$,
$\boldsymbol{\psi_n} = (\xi_{n1}, \cdots, \xi_{n6}, \tau^2_{n1}, \cdots, \tau^2_{n6})^T = (exp(3), exp(4), exp(3), exp(4), exp(3.5), exp(3), exp(5), exp(4), exp(5), exp(4), exp(4), exp(5))^T$.

Univariate CACE analysis

Under univariate CACE analysis, covariance pattern models with an unstructured variance-covariance structure are applied on individual outcome. For individual endpoint, the model for Y^a_{ij}, i^{th} patient's measurement at time point j within compliance type m when assigned to group a, is specified as below:

$$Y^a_{ij}|(C_i = m) = \beta_{m0} + \beta_{m1}t_{ij} + \beta_{m2}t^2_{ij} + \beta_{m3}D_i(a)t_{ij} + \epsilon^a_{mij}$$

where $m = c$ or n. Since $D_i(a) = 0$ for never-takers, $\beta_{n3} \equiv 0$. Let $\boldsymbol{\epsilon^a_{mi}} = (\epsilon^a_{mi0}, \epsilon^a_{mi1}, \cdots, \epsilon^a_{miJ})^T$, we assume $\boldsymbol{\epsilon^1_{mi}} \perp\!\!\!\perp \boldsymbol{\epsilon^0_{mi}}$ and $\boldsymbol{\epsilon^a_{mi}} \sim N(0, \Sigma_m)$ where Σ_m has an unstructured variance-covariance structure. Because the variance-covariance matrix in the latent-factor MCACE model is complicated, here covariance pattern models with an unstructured variance-covariance matrix are employed to avoid model misspecification when conducting univariate CACE analysis.

$$\Sigma_m = \begin{pmatrix} \sigma_{m11} & \sigma_{m12} & \sigma_{m13} & \cdots & \sigma_{m1J} \\ \sigma_{m21} & \sigma_{m22} & \sigma_{m23} & \cdots & \sigma_{m2J} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \sigma_{mJ1} & \sigma_{mJ2} & \sigma_{mJ3} & \cdots & \sigma_{mJJ} \end{pmatrix} \tag{3.14}$$

In Eqn 3.14, $\sigma_{mjj'} = \sigma_{mj'j}$ because variance-covariance matrix is symmetric.

We also use the logistic regression to model compliance patterns as shown in Eqn 3.6. Similar to multivariate analysis, we did not consider covariates and $p_{ci} = p_c = 0.3$.

We assume individuals are followed up at the same fixed time points, which means $t_{ij} = t_j$. The complier average causal effect for individual outcome is

$$E(Y^1_{ij} - Y^0_{ij}|C_i = c) = \beta_{c3}t_j,$$

which is captured by the coefficient of $D_i(a)t_{ij}$.

3.6.2 Tables

Table 3.7 reports the estimates and standard errors when all outcomes load on the same latent factor. In this case, we only consider one latent factor ($q = Q = 1$) in Eqn 3.1 (level-1 model). Besides, β_{1c01} in Eqn 3.12 (level-2 model) is also set to be 0 because this is a

randomized controlled trial. The estimates of parameters in Λ are positive, which implies the potential outcomes y^a_{ijk} given compliance type m are positively correlated with the latent factor U^a_{1mij}. Because we are interested in the overall CACE captured by the latent factor, the complier average causal effect on the common latent factor at the sixth month is

$$E(U^1_{1mi2}|m=c) - E(U^0_{1mi2}|m=c) = \beta_{1c11} * 2 + \beta_{1c21} * 4. \tag{3.15}$$

The complier average causal effect on the common latent factor is estimated to be 1.447 (SE = 0.636, p-value = 0.023, Table 3.7). Therefore, the latent-factor MCACE model detects an overall significant beneficial CACE of using the AHJ based on the common latent factor.

Table 3.8 reports the estimation results of the compliance model when one latent factor is considered. Similar to the situation where two latent factors are considered, the coefficients for all covariates except high disease activity are negative. Additionally, the coefficients for early disease, high disease activity and older age are statistically significant. Finally, we concluded that the estimation results suggested that younger female patients with longer disease duration and high disease activity were more likely to be compliers.

Table 3.7: Estimates and standard errors for causal treatment effects in the AHJ study

One latent factor only			
Fixed effects estimation*			
Parameter	Estimate	SE	P-value
β_{1c11}	2.818	1.195	0.018
β_{1c21}	-1.047	0.525	0.046
Treatment effect estimation at the sixth month			
	Estimate	SE	P-value
$\text{CACE}^\dagger_1$	1.447	0.636	0.023
Estimation of factor loading matrix Λ			
Parameter	Estimate	SE	
λ_{11}	8.192	0.084	
λ_{21}	3.074	0.571	
λ_{31}	9.302	0.680	
λ_{41}	2.972	0.666	
λ_{51}	8.967	0.737	
λ_{61}	2.872	0.648	

$*$: Fixed effects estimation section only shows parameters related to treatment effect estimation.
†: CACE_1 represents overall CACE for six outcomes when considering one latent factor only.

Table 3.8: Estimation results of compliance model

One latent factor			
Covariates	Estimate	Standard error	P-value
Intercept	-0.799	0.614	0.193
Early disease	-2.362	1.121	0.035
High disease activity	1.365	0.641	0.033
Male	-1.009	0.871	0.246
Older age	-0.956	0.484	0.048

Note: Early disease = 1 if early disease (0-2 years) and = 0 if late disease ($\geq$ 2 years)), high disease activity = 1 if high disease activity (high RAPID4 values) and = 0 if low disease activity (remission, moderate/low RAPID4 values)), Male = 1 if male and = 0 if female, Older age = 1 if above the median age (54.5) and = 0 if otherwise.

Chapter 4

Estimating Causal Effects Under Partial Compliance and Multivariate Endpoints

4.1 Introduction

In real-world randomized controlled trials (RCTs), participants who are randomly assigned to the treatment group may not take the full amount of the assigned treatment. This noncompliance issue is common and can greatly complicate statistical inference when assessing the treatment efficacy of new interventions. The intention-to-treat (ITT) approach focuses on estimating the effect of treatment assignments, which typically provides conservative estimates of the treatment efficacy (Sheiner & Rubin 1995). An alternative approach, as-treated (AT) analysis compares the outcomes by treatment actually receipt. AT analysis violates the randomization assumption and can produce significant biases in estimating treatment efficacy in RCTs with noncompliance.

To deal with the noncompliance issue, we employed a principal stratification approach (Frangakis & Rubin 2002) within the potential outcome framework. The principal stratification approach works by evaluating the treatment effect within a certain stratum defined on the combination of potential values of partially observed intermediate outcome variables, which are usually measured after treatment assignments. Under all-or-none compliance where patients are assumed to either take the treatment or do not take the treatment, a bivariate indicator of the treatment receipt is treated as the intermediate variable. The potential values of the treatment receipt in the intervention group and the control group jointly determine the compliance status of participants. In the case of all-or-none compliance, the whole population can be classified into four groups: compliers, never-takers, always-takers and defiers. Compliers comply with their initial assignments. Never-takers never take the treatment no matter which group they are assigned to. Always-takers always take the treatment and defiers always do the opposite of their initial assignments. The treatment effects

between compliers in the treatment group and compliers in the control group are known as complier-average causal effects (CACEs).

However, in many RCTs, patients may take a fraction of the treatment assigned instead of taking the full amount due to the side effects or inconvenience. This is usually described as partial compliance. Under partial compliance, a continuous variable measuring the portion of the treatment taken by a patient is treated as the intermediate variable. Similar to the all-or-none compliance, this intermediate variable is partially observed and the combination of potential values of the variable in the treatment and control group jointly determines the compliance status of certain individual. Then principal causal effects can be evaluated within certain stratum stratified on subjects' compliance status.

This chapter is motivated by the Arthritis Health Journal (AHJ) study conducted at Arthritis Research Canada (PI: Diane Lacaille). The AHJ is a patient-centred online tool designed to enhance patients' self-management ability by helping patients monitor disease status and generate action plans. Participants are assigned to the intervention group and the usual care group randomly. They are evaluated every three months using a self-administered questionnaire. We focused on the data collected during the first six months. Since multiple outcomes are usually employed to evaluate the treatment efficacy of multifaceted interventions (e.g, AHJ tool), the follow-up questionnaires collect six health outcomes to evaluate the effectiveness of the AHJ tool.

Furthermore, the follow-up questionnaires also include questions about the frequency of using the tool, which can be used to define the compliance status. Guo et al. (2022) assumes all-or-none compliance and considers compliers and never-takers only. Always-takers and defiers are ruled out because patients in the control group do not have access to the intervention in the AHJ study. Compliers are defined as patients who use the tool at least once per month on average over six months when assigned to the treatment group. However, this dichotomization can cause a loss of information when modeling compliance behaviour. To overcome the limitation, we consider partial compliance in this chapter and define the compliance status based on the frequency rate. The frequency rate is calculated as the ratio of the times of using the tool to the number of days of the follow-up period. In the intervention group, the frequency rate of each participant is observable because patients have access to the tool. However, the frequency rate is unobservable for individuals in the control group since we never know how they will behave if they are assigned to the intervention group. Similar to all-or-none compliance, the frequency rate can be treated as the baseline characteristic of the participant and is used to form the principal stratum. Principal causal effects are evaluated within participants in a certain stratum based on the possible values of the frequency rate.

Much of existing research focuses on all-or-none compliance (Angrist et al. 1996, Imbens & Rubin 1997, Hirano et al. 2000, Yau & Little 2001, Gao et al. 2014) or categorical multilevel compliance (e.g., no compliance, partial compliance, and full compliance) (Sanders

et al. 2021, Shrier et al. 2018). In the RCTs involving partial adherence, continuous partial compliance is preferred since the discretization to discrete/categorical compliance can cause a loss of information. Jin & Rubin (2008) considered continuous partial compliance when analyzing the Lipid Research Clinics Coronary Primary Prevention Trial (LRC-CPPT), a placebo-controlled randomized controlled trial reported by Efron & Feldman (1991). Within the principal stratification approach, they proposed an extended partial compliance framework where drug compliance and placebo compliance are treated as continuous and jointly determine the principal stratum. The Beta distribution is chosen to model placebo compliance and drug compliance given placebo compliance separately. Linear regression models are employed to link the cross-sectional potential outcomes and drug and placebo compliance. Bartolucci & Grilli (2011) reanalyzed the data in Efron & Feldman (1991). Instead of specifying the marginal distribution of the continuous partial compliance directly, they modeled the joint distribution of the drug compliance and placebo compliance through a Plackett copula. Besides, they generalized the linear regression models for cross-sectional potential outcomes in Jin & Rubin (2008) by considering interaction terms between drug compliance and placebo compliance and heteroscedasticity for potential outcomes. Schwartz et al. (2011) utilized a Dirichlet process mixture (DPM) model to capture the possible complex structure of the joint distribution of drug compliance and placebo compliance. The DPM model permits a flexible nonparametric way to model the principal strata conditional on covariates and allows for better interpretation due to the byproduct of clustering. They also illustrated their approach in the data from Efron & Feldman (1991).

Our work is different from the research presented above in several ways. Firstly, multiple endpoints are collected at several time points in the AHJ study. To capture the correlations among multiple endpoints across all time points, a hierarchical random-effects approach is employed to model the primary potential outcomes given principal strata. Secondly, although Beta distribution is also considered to model continuously-measured partial compliance which is similar to Jin & Rubin (2008), we utilized the Beta regression model which allows for covariates so that compliance behavior can be predicted based on the baseline covariates. Thirdly, we only need to consider the compliance behaviour of using the new intervention (intervention compliance) in the AHJ study. Therefore, the principal strata are determined by the potential outcomes of intervention compliance only. We also conducted a simulation study to validate the performance of the model proposed. After applying the proposed model to the AHJ data, the findings are consistent with that in Guo et al. (2022) who analyzed the AHJ data by assuming binary compliance. However, this chapter detected a much smaller overall p-value when examining if there is an overall treatment effect by combining information from all outcomes. This is possibly caused by less information being lost under partial compliance.

This chapter is organized as follows. Section 4.2 introduces the motivating example. Section 4.3 describes the models for principal strata and primary potential outcomes given

principal strata. Section 4.4 presents the simulation study to demonstrate the performance of the model proposed. Then the proposed model is applied to the AHJ data in Section 4.5. Lastly, a brief discussion is given in Section 4.6.

4.2 Motivating Application

Self-management is crucial for rheumatoid arthritis (RA) patients. RA patients could benefit substantially by being actively engaged in disease management since closely monitoring symptoms and treatment use could help them determine their current disease status and seek medical attention timely. However, it's not easy to manage disease well for patients with chronic disease (e.g., rheumatoid arthritis). Unstable symptoms during a long period make it hard for patients to follow treatment regimes strictly. Therefore, there is a strong demand to develop a multifaceted tool to enhance RA patients' self-management ability.

Arthritis Health Journal (AHJ) is a multifaceted tool developed to foster the active involvement of RA patients in monitoring their disease activity so that early warnings can be provided when targets are not being achieved. A randomized controlled trial is designed to evaluate the effectiveness of the behavioral intervention (AHJ). Among 94 participants, 45 patients were randomized to the intervention group where they received the intervention immediately, whereas 49 patients were asked to wait for six months before being provided access to the AHJ tool. This chapter focuses on the first 6-month data of the study, which implies the second group received usual care and thus served as the control group. RA patients enrolled in the study were evaluated every 3 months using self-administered questionnaires. Demographics and disease information were collected in the baseline questionnaires. The three- and six-month questionnaires collected information about six endpoints including effective consumer 17 scale, manage symptoms scale, manage disease in general scale, communicate with physician scale, partners in health scale and satisfaction with various aspects of medical care. For ease of comparison, six outcomes were rescaled to the same scale (0 to 100) and higher values represented better results.

Follow-up questionnaires collected information on how frequently the tool was used by patients. Instead of considering compliers and noncompliers in the case of all-or-none compliance, continuously-measured partial compliance is considered here. Let A_i denote treatment assignment for patient i, where $i = 1, \cdots, N$. $A_i = T$ if subject i was randomized to the intervention group and $A_i = C$ if the subject was assigned to the control group. We define $D_i(\boldsymbol{A})$ to be the frequency rate of using the AHJ for patient i if subjects are randomized to group $\boldsymbol{A}$, where $\boldsymbol{A}$ represents the collection of all subjects' assignments. For the definition of $D_i(\boldsymbol{A})$, the frequency rate is calculated by dividing the number of times of using the tool over six months by 180 days. Therefore, $D_i(\boldsymbol{A})$ is a continuous variable with a value ranging from 0 to 1. Under the principal stratification framework (Frangakis & Rubin 2002), the population is classified into different strata based on the combination of

potential values of intermediate variable $D_i(\boldsymbol{A})$. We define potential outcomes combination $S_i = [D_i(\boldsymbol{T}), D_i(\boldsymbol{C})]$ as the principal stratum where patient i belongs. In the case of all-or-none compliance where $D_i(\boldsymbol{A}) = 1$ or 0, there are four principal strata: compliers, never-takers, always-takers and defiers. Then the principal causal effects are evaluated within each principal stratum. Similarly, when dealing with continuously-measured partial compliance, principal causal effects are assessed within the principal stratum $S_i = [D_i(\boldsymbol{T}), D_i(\boldsymbol{C})]$ when $D_i(\boldsymbol{A})$ is a continuous variable whose value ranges from 0 to 1. Specifically, patients in the control group don't have access to the AHJ, which implies $D_i(\boldsymbol{C}) = 0$. Therefore, the principal stratum can be simplified as $S_i = [D_i(\boldsymbol{T}), 0]$. Specifically, $D_i(\boldsymbol{T})$ is observable for participants assigned to the treatment group, but is unknown for patients in the control group.

Since six outcomes are collected over time for each individual, let $Y_{ijk}(\boldsymbol{A})$ denote the primary potential outcome value of k^{th} endpoint at j^{th} time point for i^{th} individual, where $k = 1, \cdots, K$, $j = 0, \cdots, J$. Under the setting of the AHJ study, $K = 6$ and $J = 2$.

4.3 Model

Let $\boldsymbol{Y}_i(\boldsymbol{A})$ denote the vector of all primary potential outcome values for individual i across all endpoints and time points. We use $\boldsymbol{D}$ and $\boldsymbol{Y(A)}$ to indicate vectors collecting potential values of $D_i(\boldsymbol{A})$ and $\boldsymbol{Y}_i(\boldsymbol{A})$ for all patients, respectively.

To specify the complete data distribution, we make the following assumptions. First, two standard assumptions are made in our analysis. Ignorable treatment assignment assumption (Rubin 1978) assumes that the assignment mechanism is independent of all potential outcomes, conditional on observable baseline covariates. This assumption is satisfied in our setting (RCTs) since patients are randomly assigned to two groups. Under the ignorable treatment assignment assumption, there is no need to model the assignment mechanism separately. Stable unit treatment value assumption (SUTVA) (Rubin 1980) consists of two components: no interference and no multiple versions of the treatment. SUTVA is reasonable in the AHJ study because each patient gets access to the same version of AHJ tool independently, which ensures that one patient's potential outcomes will not be affected by other patients' assignments. Thus, potential outcomes $D_i(\boldsymbol{A})$ and $\boldsymbol{Y}_i(\boldsymbol{A})$ can be written as $D_i(A_i)$ and $\boldsymbol{Y}_i(A_i)$, respectively. Considering randomization $\boldsymbol{A}$, potential treatment received $\boldsymbol{D}$ and primary potential outcomes $\boldsymbol{Y(A)}$, the complete data distribution is given

by

$$
\begin{aligned}
f(\boldsymbol{A}, \boldsymbol{D}, \boldsymbol{Y(T)}, \boldsymbol{Y(C)}) &= f(\boldsymbol{A}) f(\boldsymbol{D}, \boldsymbol{Y(T)}, \boldsymbol{Y(C)}) \\
&= \prod_{i=1}^{N} f(A_i) f(D_i(T), D_i(C), \boldsymbol{Y}_i(T), \boldsymbol{Y}_i(C)) \\
&= \prod_{i=1}^{N} f(A_i) f(S_i) f(\boldsymbol{Y_i}(T), \boldsymbol{Y_i}(C) | S_i) \\
&= \prod_{i=1}^{N} f(A_i) f(D_i(T)) f(\boldsymbol{Y_i}(T) | D_i(T)) f(\boldsymbol{Y_i}(C) | D_i(T))
\end{aligned}
\tag{4.1}
$$

Here $f()$ represents a density function generally. The equalities in Eqn 4.1 also imply two additional assumptions. As patients in the control group were asked to wait for six months before being provided access to the AHJ tool and we focused on the data collected in the first six months, the third assumption assumes patients in the control group do not have access to the treatment ($D_i(C) = 0$). Therefore, $S_i = [D_i(T), D_i(C)] = [D_i(T), 0]$. We model $D_i(T)$ directly instead of modeling S_i, the principal stratum for individual i. The fourth assumption is $\boldsymbol{Y}_i(T)$ is independent of $\boldsymbol{Y}_i(C)$ given S_i. We make this assumption since $\boldsymbol{Y}_i(T)$ and $\boldsymbol{Y}_i(C)$ are never jointly observed for patient i and the likelihood function based on observed data does not depend on the correlations between $\boldsymbol{Y}_i(T)$ and $\boldsymbol{Y}_i(C)$. Furthermore, we made an exclusion restriction assumption, which means the effect of the treatment assignment on primary potential outcomes must be passed through the treatment receipt.

4.3.1 Models

Modeling principal stratum S_i

As mentioned earlier, the principal stratum S_i is completely determined by the potential intervention compliance $D_i(T)$ with a value between 0 and 1. We employed a Beta regression model to model $D_i(T)$. The density of the Beta distribution can be rewritten as Eqn 4.2 in terms of intervention compliance $D_i(T)$'s expectation μ_{Di} and precision parameter ϕ.

$$
f(D_i(T); \mu_{Di}, \phi) = \frac{\Gamma(\phi)}{\Gamma(\mu_{Di}\phi)\Gamma((1-\mu_{Di})\phi)} D_i(T)^{\mu_{Di}\phi - 1} (1 - D_i(T))^{(1-\mu_{Di})\phi - 1}, \quad 0 < D_i(T) < 1,
\tag{4.2}
$$

where $\mu_{Di} \in (0, 1)$ and $\phi > 0$. Then a logit link function is chosen to link i^{th} subject's covariates $\boldsymbol{W_i}$ and mean μ_{Di}:

$$
log(\frac{\mu_{Di}}{1 - \mu_{Di}}) = \boldsymbol{W_i}^T \boldsymbol{\alpha}
\tag{4.3}
$$

where $\boldsymbol{\alpha}$ is the vector collecting all coefficients of corresponding covariates $\boldsymbol{W_i}$. Equivalently, $D_i(T)$ follows a Beta distribution with parameters p_{Di} and q_{Di}, where $p_{Di} = \mu_{Di}\phi$ and

$q_{Di} = (1-\mu_{Di})\phi$. If covariates $\boldsymbol{W_i}$ only includes an intercept, then $\mu_{Di} \equiv \mu_D$, which implies $p_{Di} = p_D$ and $q_{Di} = q_D$.

Modeling primary potential outcomes $\boldsymbol{Y}_i(A_i)$ given principal stratum S_i

We employ a hierarchical random-effects approach to model the primary potential outcomes $\boldsymbol{Y}_i(A_i)$ given principal stratum S_i. Since the principal stratum S_i is uniquely determined by the secondary potential outcome $D_i(T)$, equivalently, we model the primary potential outcomes $\boldsymbol{Y}_i(A_i)$ given the secondary potential outcome $D_i(T)$. The potential outcome for intervention group $\boldsymbol{Y}_i(T)$ and potential outcome for control group $\boldsymbol{Y}_i(C)$ are modelled separately and assumed to be independent. For notational convenience, we use $\boldsymbol{Y}_{ai}$ and Y_{aijk} to denote $\boldsymbol{Y}_i(A_i)$ and $Y_{ijk}(A_i)$ where $A_i = a$ ($a = T$ or C). The hierarchical random-effects approach includes two stages. Stage 1 specifies a within-subjects model for Y_{aijk} representing the i^{th} patient's primary potential outcome of k^{th} endpoint at time point j when randomized to group a.

$$Y_{aijk}|(D_i(T), \boldsymbol{b_{aik}}, \boldsymbol{Z_{ij}}) = \boldsymbol{Z_{ij}}^T \boldsymbol{b_{aik}} + \epsilon_{aijk}, \tag{4.4}$$

where $\boldsymbol{Z_{ij}}$ is a vector incorporating time-varying covariates. $\boldsymbol{b_{aik}}$ is a vector collecting all coefficients of time-varying covariates $\boldsymbol{Z_{ij}}$ for k^{th} endpoint of i^{th} individual under group a. We assume $\epsilon_{aijk} \sim N(0, \sigma^2_{ak})$ and $\epsilon_{Tijk} \perp\!\!\!\perp \epsilon_{Cijk}$. Besides, ϵ_{aijk}'s are assumed to be independent over time and across multiple endpoints.

Stage 2 specifies between-subjects models for parameters $\boldsymbol{b_{aik}}$ related to a specific individual i when randomized to group a.

$$\boldsymbol{b_{aik}} = \boldsymbol{\beta}_{0k} + \boldsymbol{\beta}_{a1k} D_i(T) + \boldsymbol{v}_{ai}, \tag{4.5}$$

$\boldsymbol{\beta}_{0k}$ consists of population average regression coefficients for subjects who would not receive any level of treatment if assigned to the intervention group. Under the exclusion restriction assumption, we assume patients who would not receive any level of the treatment in both the intervention group and control group share the same set of population average regression coefficients, $\boldsymbol{\beta}_{0k}$. $\boldsymbol{\beta}_{a1k}$ represents the population average change in these regression coefficients for subjects who belong to stratum $S_i = [D_i(T), 0]$ if assigned to group a. $\boldsymbol{v}_{ai}$ is a vector of random effects representing individual i's deviation from population average coefficients in group a. $\boldsymbol{v}_{ai}$ is assumed to follow a multivariate normal distribution with mean $\mathbf{0}$ and variance-covariance matrix Σ_{va}. We also assume $\boldsymbol{v}_{Ti} \perp\!\!\!\perp \boldsymbol{v}_{Ci}$ and random effects in $\boldsymbol{v}_{ai}$ are independent from error terms ϵ_{aijk}'s in stage 1 model for $a = T$ or C.

4.3.2 Principal Causal Effect in Stratum S

We extend the definition of principal causal effect (PCE) in Jin & Rubin (2008) to a longitudinal setting where multivariate outcomes are collected at each time point. The PCE for

k^{th} outcome at j^{th} time point in stratum S is defined as

$$PCE_{jk}(S) = E[Y_{Tijk} - Y_{Cijk} \mid S], \tag{4.6}$$

which is the average causal effect for patients in stratum S. To obtain the PCE in stratum S, $E(Y_{Tijk})$ and $E(Y_{Cijk})$ are required and can be derived from Eqns 4.4 and 4.5. By combining two stages (Eqns 4.4 and 4.5), $\boldsymbol{Y}_{ai}|S_i$, the potential outcomes for individual i within principal stratum S_i ($S_i = [D_i(T), 0]$) can be expressed as

$$\boldsymbol{Y}_{ai}|(S_i, \boldsymbol{v_{ai}}) = (X_i \otimes I_K)\boldsymbol{\beta_a} + \left(Z_i \boldsymbol{v_{ai}}\right) \otimes 1_K + \boldsymbol{\epsilon}_{ai} \tag{4.7}$$

where $\boldsymbol{Y}_{ai} = \{Y_{aijk} : i = 1, \cdots, N; j = 0, \cdots, J; k = 1, \cdots, K\}$, $\boldsymbol{\beta}_a = \{\beta_{apqk} : p = 0, \cdots, P; q = 0, \cdots, Q; k = 1, \cdots, K\}$, where P and Q depend on the forms of two stages. Here we assume $\beta_{Tp0k} = \beta_{Cp0k} = \beta_{p0k}$ since individuals in both the intervention group and control group share the same set of coefficients $\boldsymbol{\beta}_{0k}$ in Eqn 4.5 under the exclusion restriction assumption. $\boldsymbol{\epsilon}_{ai} = (\boldsymbol{\epsilon_{ai0}}^T, \cdots, \boldsymbol{\epsilon_{aiJ}}^T)^T$, where $\boldsymbol{\epsilon}_{aij} = (\epsilon_{aij1}, \cdots, \epsilon_{aijK})^T$. Design matrices for fixed effects and random effects are represented by X_i and Z_i, respectively.

Based on Eqn 4.7, the conditional distribution of potential outcomes $\boldsymbol{Y}_{ai}$ given principal strata S_i is

$$\boldsymbol{Y}_{ai}|S_i \sim MVN_{\boldsymbol{\beta}_a, \boldsymbol{\psi}_a}(\boldsymbol{\mu_{ai}}, \Sigma_{ai}), \tag{4.8}$$

where the expectation $\boldsymbol{\mu_{ai}} = (X_i \otimes I_K)\boldsymbol{\beta_a}$ and variance $\Sigma_{ai} = (Z_i \Sigma_{va} Z_i^T) \otimes (1_K 1_K^T) + diag(\Phi_{a0}, \Phi_{a1}, \cdots, \Phi_{aJ})$, where $\Phi_{a0} = \Phi_{a1} = \cdots = \Phi_{aJ} = \Phi_a = diag(\sigma^2_{a1}, \cdots, \sigma^2_{aK})$. In Eqn 4.8, $\boldsymbol{\psi}_a$ is a vector collecting all unique parameters in Σ_{va} and Φ_a.

Therefore, the PCE for patients within stratum S can be expressed as

$$\begin{aligned} PCE(S) &= E(\boldsymbol{Y}_{Ti} - \boldsymbol{Y}_{Ci}|S) \\ &= (X_i \otimes I_K)\boldsymbol{\beta_T} - (X_i \otimes I_K)\boldsymbol{\beta_C}. \end{aligned} \tag{4.9}$$

Next, we use an example to illustrate the principal causal effect for k^{th} outcome at j^{th} time point in stratum S. Specifically, stage 1 of the hierarchical random-effects model in Eqn 4.4 is specified as

$$Y_{aijk} \mid D_i(T) = b_{a0ik} + b_{a1ik} t_{ij} + \epsilon_{aijk}. \tag{4.10}$$

Stage 2 of the model proposed in Eqn 4.5 is given by

$$\begin{aligned} b_{aoik} &= \beta_{00k} + \beta_{a01k} D_i(T) + v_{a0i}, \\ b_{a1ik} &= \beta_{10k} + \beta_{a11k} D_i(T) + v_{a1i}. \end{aligned} \tag{4.11}$$

Based on Eqns 4.10 and 4.11, we obtain

$$Y_{aijk} \mid D_i(T) = \beta_{00k} + \beta_{10k}t_{ij} + \beta_{a01k}D_i(T) + \beta_{a11k}D_i(T)t_{ij} + v_{a0i} + v_{a1i}t_{ij} + \epsilon_{aijk}. \quad (4.12)$$

Thus, the conditional expectation of the primary potential outcome Y_{aijk} given $D_i(T)$ can be written as a linear function of $D_i(T)$ in the illustrative example:

$$E(Y_{aijk} \mid D_i(T)) = \beta_{00k} + \beta_{10k}t_{ij} + \beta_{a01k}D_i(T) + \beta_{a11k}D_i(T)t_{ij}. \quad (4.13)$$

Since the expectations of error terms in stage 1 and random effects in stage 2 are 0, the conditional expectation of Y_{aijk} given $D_i(T)$ can be expressed as a linear combination of fixed effects only. In many real-world randomized controlled trials, participants are followed up at the same time points, which implies $t_{ij} = t_j$. Then the principal causal effect for k^{th} outcome at time point j for patients within stratum S is

$$\begin{aligned} PCE_{jk}(S) &= E(Y_{Tijk} \mid D_i(T)) - E(Y_{Cijk} \mid D_i(T)) \\ &= (\beta_{T01k} - \beta_{C01k})D_i(T) + (\beta_{T11k} - \beta_{C11k})D_i(T)t_j \end{aligned} \quad (4.14)$$

Because participants are assigned to two groups randomly, it's reasonable to assume that there is no difference at baseline between the intervention group and the control group for patients who would take the level of active treatment $D_i(T)$ if randomized to the treatment group. Since the difference between parameters β_{T01k} and β_{C01k} captures the baseline difference, normally $\beta_{T01k} = \beta_{C01k}$ under randomization and the difference between parameters β_{T11k} and β_{C11k} characterizes the principal causal effect of our primary interest.

Besides, a matrix form of potential outcomes for k^{th} endpoint can be derived from Eqn 4.12. Let $\boldsymbol{Y}_{aik} = (Y_{ai0k}, Y_{ai1k}, \cdots, Y_{aiJk})^T$,

$$\boldsymbol{Y}_{aik} = X_i \times \begin{pmatrix} \beta_{00k} \\ \beta_{10k} \\ \beta_{a01k} \\ \beta_{a11k} \end{pmatrix} + Z_i \times \begin{pmatrix} v_{a0i} \\ v_{a1i} \end{pmatrix} + \boldsymbol{\epsilon}_{aik} \quad (4.15)$$

where $\boldsymbol{\epsilon}_{aik} = (\epsilon_{ai0k}, \epsilon_{ai1k}, \cdots, \epsilon_{aiJk})^T$. X_i and Z_i are specified as

$$X_i = \begin{pmatrix} 1 & 0 & D_i(T) & 0 \\ 1 & t_{i1} & D_i(T) & t_{i1} * D_i(T) \\ \vdots & \vdots & \vdots & \vdots \\ 1 & t_{iJ} & D_i(T) & t_{iJ} * D_i(T) \end{pmatrix}, \quad Z_i = Z = \begin{pmatrix} 1 & 0 \\ 1 & t_{i1} \\ \vdots & \vdots \\ 1 & t_{iJ} \end{pmatrix}, \quad (4.16)$$

which are the exact forms of X_i and Z_i in Eqn 4.7 when the model is given by Eqns 4.10 and 4.11.

4.3.3 Estimation

For a specific individual in group $A_{obs,i}$, $\boldsymbol{Y}_{obs,i} = \boldsymbol{Y}_i(A_{obs,i})$ and $\boldsymbol{Y}_{mis,i} = \boldsymbol{Y}_i(A_{mis,i})$ where $A_{mis,i}$ indicates the group where this individual was not assigned. $D_i(T)$ is observable for participants assigned to group T but is missing for participants assigned to group C. Here we use $D_{mis,i}$ to denote missing $D_i(T)$. Let $\boldsymbol{\pi} = (\boldsymbol{\beta_T}, \boldsymbol{\beta_C}, \boldsymbol{\psi_T}, \boldsymbol{\psi_C}, \boldsymbol{\alpha})$, the likelihood function based on observed data for all participants in the study is

$$\begin{aligned}
\mathcal{L}(\boldsymbol{\pi}; \boldsymbol{Y_{obs}}, \boldsymbol{D_{obs}}, \boldsymbol{A_{obs}} \mid \boldsymbol{W}) &= \prod_{i=1}^{N} \iint f(A_i) f(D_i(T)) f(\boldsymbol{Y_i}(T)|D_i(T)) f(\boldsymbol{Y_i}(C)|D_i(T)) d\boldsymbol{Y}_{mis,i} dD_{mis,i} \\
&= \prod_{i \in \{A_i = T\}} f(D_i(T)) f(\boldsymbol{Y_i}(T)|D_i(T)) \\
&\quad \times \prod_{i \in \{A_i = C\}} \int f(D_i(T)) f(\boldsymbol{Y_i}(C)|D_i(T)) dD_i(T) \\
&= L_1 \times L_0
\end{aligned} \tag{4.17}$$

where

$$\begin{aligned}
L_1 = \prod_{i \in \{A_i = T\}} & \frac{\Gamma(\phi)}{\Gamma(\mu_{Di}\phi)\Gamma((1-\mu_{Di})\phi)} D_i(T)^{\mu_{Di}\phi - 1} (1 - D_i(T))^{(1-\mu_{Di})\phi - 1} \\
& \frac{1}{(2\pi)^{\frac{(J+1)K}{2}} |\Sigma_{Ti}|^{\frac{1}{2}}} \exp\left\{ -\frac{1}{2} \left(y_{obs,i} - \mu_{Ti}\right)^T \left(\Sigma_{Ti}\right)^{-1} \left(y_{obs,i} - \mu_{Ti}\right) \right\} \\
L_0 = \prod_{i \in \{A_i = C\}} & \int_0^1 \frac{\Gamma(\phi)}{\Gamma(\mu_{Di}\phi)\Gamma((1-\mu_{Di})\phi)} D_i(T)^{\mu_{Di}\phi - 1} (1 - D_i(T))^{(1-\mu_{Di})\phi - 1} \\
& \frac{1}{(2\pi)^{\frac{(J+1)K}{2}} |\Sigma_{Ci}|^{\frac{1}{2}}} \exp\left\{ -\frac{1}{2} \left(y_{obs,i} - \mu_{Ci}\right)^T \left(\Sigma_{Ci}\right)^{-1} \left(y_{obs,i} - \mu_{Ci}\right) \right\} dD_i
\end{aligned} \tag{4.18}$$

The likelihood function in Eqn 4.17 includes two parts. L_1 is the likelihood function for the treatment group where both $D_i(T)$ and $\boldsymbol{Y}_i(T)$ are observed. L_0 is the likelihood function for the control group where $\boldsymbol{Y}_i(C)$ is observed while $D_i(T)$ is missing. Therefore, $D_i(T)$ needs to be integrated out in L_0. Since Eqns 4.2 and 4.8 give the distribution of $D_i(T)$ and the conditional distribution of $\boldsymbol{Y}_i(A_i = a)$ given principal stratum S_i, we could obtain the specific form of the likelihood functions L_0 and L_1 as shown in Eqn 4.18.

The integral in the likelihood function L_0 can be computed by using the "integrate" function in R. We obtained the maximum likelihood estimates (MLEs) by maximizing the whole log-likelihood function of observed data via “optim" function. The corresponding standard errors can be obtained by evaluating the inverse Hessian matrix at the MLEs.

4.4 Simulation Study

We conducted a simulation study to validate the performance of the model proposed. Six endpoints ($K = 6$) are simulated at three time points ($J = 2$) for 500 individuals. The intervention compliance $D_i(T)$ is modelled under the Beta regression model as specified in Eqns 4.2 and 4.3. To simplify the setting of our simulation, we assume $\boldsymbol{W_i}$ includes an intercept only in Eqn 4.3 and $D_i(T) \sim Beta(p_D, q_D)$. The model for $\boldsymbol{Y}_i(A_i)$ given the principal stratum S_i includes a linear trend in stage 1 (Eqn 4.10) and a random intercept only in stage 2, which means $v_{a1i} = 0$ for $a = T$ or C. Besides, under the randomization assumption, we also assume there is no baseline difference between the treatment group and control group for individuals within principal stratum $S_i = [D_i(T), 0]$, which implies $\beta_{T01k} = \beta_{C01k}$. Therefore, $PCE_{jk}(S) = (\beta_{T11k} - \beta_{C11k})D_i(T)t_j$ because $\beta_{T01k} - \beta_{C01k} = 0$ in Eqn 4.14.

Table 4.1 summarizes the estimation accuracy for the parameters (β_{T11k} and β_{C11k}) related to PCEs and parameters (p_D and q_D) for the Beta distribution. The results in Table 4.1 are obtained based on 1000 simulated datasets. Table 4.1 presents sample means, sample standard deviations of estimates and sample means of standard error estimates calculated using the Fisher information for parameters of our primary interest. The sample means are nearly equal to their corresponding true values. Besides, the sample means of standard error estimates obtained using the Fisher information are also close to their corresponding sample standard deviations of estimates.

4.5 Application to the AHJ Study

In this section, we apply the proposed model to the AHJ data. As defined in Section 4.2, $D_i(A)$ is the frequency rate of using the AHJ during a six-month period for individual i when randomized to group A. In the AHJ study, $D_i(C)$ always equals 0 as patients in the control group do not have access to the AHJ tool. The potential outcome $D_i(T)$ is observable for patient i randomized to the intervention group, but is latent for participant i randomized to the control group. The frequency rate of using the AHJ tool is obtained based on the information collected in the follow-up questionnaires and is calculated as the ratio of the times of using the tool to the total number of days they were followed up. Intervention compliance $D_i(T)$ is modeled under the Beta regression model specified in Eqns 4.2 and 4.3. Four binary covariates are included in the Beta regression: disease duration (1 = early disease, 0 = late disease), disease activity (1 = high RAPID4 values, 0 = moderate/low RAPID4 values), gender (1 = male, 0 = female) and age (1 = above 54.50 years old, 0 = below 54.50 years old).

Table 4.2 provides the summary statistics for baseline variables and missing data patterns for six outcomes during the follow-up period. One outlier has been excluded from the treatment group due to extremely large $D_i(T)$. Similar to Table 1 in Guo et al. (2022), the

Table 4.1: Estimation accuracy under the model proposed (based on 1000 repetitions when sample size equals 500)

Parameter	True value	Sample mean	$sse^{\dagger}$ ($sm_{se}^{\ddagger}$)
Parameters related to Treatment effect			
β_{T111}	2	1.952	2.132 (2.035)
β_{T112}	3	2.947	2.059 (1.977)
β_{T113}	2	1.984	2.115 (2.040)
β_{T114}	2	1.959	3.258 (3.038)
β_{T115}	3	3.015	2.056 (1.987)
β_{T116}	5	4.948	3.327 (3.248)
β_{C111}	1	0.909	2.490 (2.451)
β_{C112}	2	1.928	2.273 (2.190)
β_{C113}	1	1.009	2.443 (2.456)
β_{C114}	1	0.908	3.523 (3.204)
β_{C115}	2	1.942	2.313 (2.203)
β_{C116}	3	2.928	3.688 (3.584)
Beta distribution			
p_D	0.693	0.703	0.085 (0.084)
q_D	1.609	1.618	0.090 (0.090)

$\dagger sse$: sample standard deviation of estimates

$\ddagger sm_{se}$: sample mean of standard error estimates calculated based on the Fisher information

baseline covariates remain favourably balanced between the treatment group and the control group. In the intervention group, 9 patients dropped out of the study after participating in the baseline questionnaire. Since no follow-up information on these nine dropouts is observed, their compliance behavior $D_i(T)$ is also latent. When dealing with the likelihood function in terms of nine dropouts in the treatment group, $D_i(T)$ needs to be integrated out, just as we did in the likelihood function of participants in the control group. We note that there is a slight amount of intermittent missingness (001 and 010) in the intervention group. For intermittent missing participants in the treatment group, $D_i(T)$ is calculated as the ratio of the times of using the tool to 90 days since they were observed for three months. We assume data is missing at random and employed the likelihood-based approach which can provide valid inference under the less strict missing mechanism (missing at random) than missing completely at random if the analysis is done appropriately.

Next, we need to determine the relationship (e.g., linear or quadratic) between individual coefficients $\boldsymbol{b_{aik}}$ for the k^{th} endpoint and intervention compliance $D_i(T)$ in Eqn 4.5. Since $D_i(T)$ is only observable in the treatment group, we focused on the measurements for individuals in the treatment group and conducted descriptive statistics. For individual

Table 4.2: Summary statistics of baseline covariates and missing data patterns

Covariates	Usual care group (N = 49) $n(\%)$	Intervention group (N = 44) $n(\%)$
Disease Duration (Early)	6 (12.2)	5 (11.4)
Disease Activity (High)	38 (77.6)	32 (72.7)
Gender (Male)	5 (10.2)	6 (13.6)
Age (> 54.5)	25 (51.0)	22 (50.0)
Missing data pattern†		
000	41 (83.7)	32 (72.7)
011	2 (4.1)	9 (20.5)
001	5 (10.2)	1 (2.3)
010	1 (2.0)	2 (4.5)

† 0 and 1 indicate presence and absence, respectively. The 3-digit indicator displays the missing status for endpoints at baseline, the third and sixth month.

outcomes in the treatment group, a mixed-effects regression model with a linear time trend, random intercepts and slopes is employed. By using "PROC MIXED" in SAS, we obtained empirical Bayes estimates of subjects' intercepts and slopes. Figure 4.1 shows the plot of individual intercepts versus intervention compliance $D_i(T)$ for each endpoint. Figure 4.2 presents the plot of individual slopes versus $D_i(T)$. For each outcome within each plot, the data is fit under generalized additive models (GAM) which provide tests to see if the trends can be captured by linear functions. The null hypothesis is there is no non-linearity. Using the R function `gam()`, p-values are calculated and provided. In Figure 4.1, we did not observe any significant p-values which means the relationships between individual intercepts and $D_i(T)$ are linear. We can also assume linear relationships between individual slopes and $D_i(T)$ for all six endpoints since no significant p-values were observed in Figure 4.2. Therefore, in stage 2 (Eqn 4.11), linear relationships are specified between subject-specific parameters $\boldsymbol{b_{aik}}$ and $D_i(T)$. We also assume $\boldsymbol{Z_{ij}}$ includes an intercept term and a linear time trend t_{ij} in stage 1 (Eqn 4.10) when modelling primary potential outcomes $\boldsymbol{Y}_i(A_i)$ given the principal stratum S_i. Under Eqns 4.10 and 4.11, the principal causal effect for k^{th} endpoint within individuals who would use the AHJ tool $D_i(T)$ of the six-month period is expressed as Eqn 4.14. Under the randomization assumption, we assume $\beta_{T01k} = \beta_{C01k}$ and the principal causal effect for k^{th} endpoint at time point j given intervention compliance $D_i(T)$ is $(\beta_{T11k} - \beta_{C11k})D_i(T)t_j$.

Table 4.3 presents the estimates, standard errors and corresponding p-values for $\beta_{T11k} - \beta_{C11k}$ $(k = 1, \cdots, 6)$, which captures principal causal effects. Firstly, we examine the P-value for each outcome individually. The null hypothesis is there is no treatment difference for a specific outcome. P-values (4^{th} column) for six endpoints are obtained based on the Wald test. Since there are six hypothesis tests for six outcomes, Type I error will inflate if the

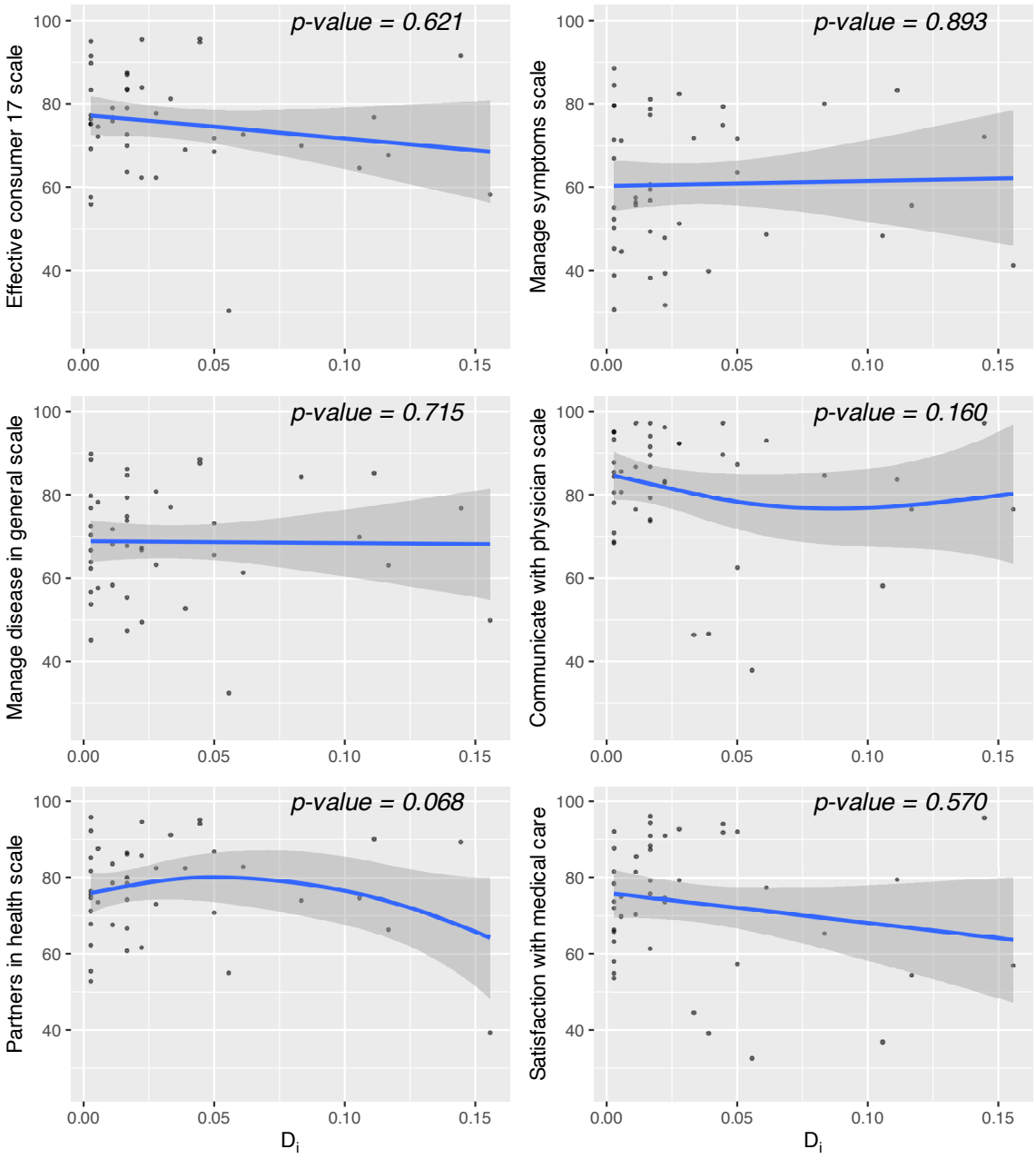

Figure 4.1: Plot of individual intercepts versus the intervention compliance $D_i(T)$. Estimates for individual intercepts are obtained based on the empirical Bayes approach. If the P-value is larger than 0.05, then there is not non-linearity.

significance cut-off value is not adjusted. To avoid this multiple testing issue, we apply the Bonferroni correction and the significance threshold is set as $0.05/6 = 0.008$. By comparing p-values (column 4) for six outcomes in Table 4.3 with 0.008, we noticed the difference between β_{T11k} and β_{C11k} is significant for the fourth endpoint (communicate with physician scale) and sixth endpoint (satisfaction with various aspects of medical care). In addition,

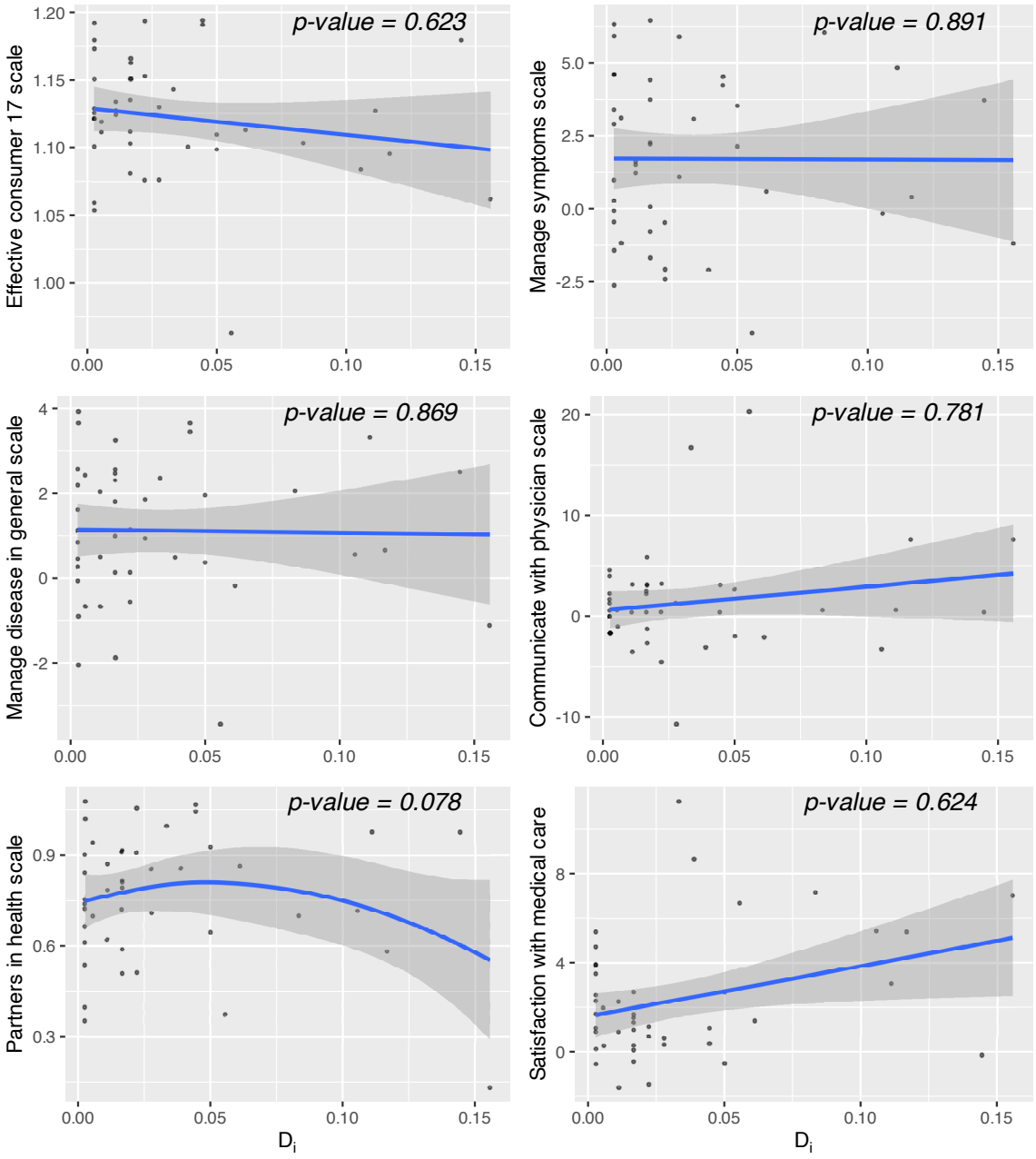

Figure 4.2: Plot of individual slopes versus the intervention compliance $D_i(T)$. Estimates for individual slopes are obtained based on the empirical Bayes approach. If the P-value is larger than 0.05, then there is not non-linearity.

the significant differences $(\beta_{T11k} - \beta_{C11k})$ for these two outcomes are positive, which implies the principal causal effects detected from the fourth and sixth outcomes are beneficial since higher values represent better results. Then we conducted a global hypothesis test to examine six endpoints simultaneously. Under the multivariate Wald test, the null hypothesis is there are no principal causal effects for all six outcomes simultaneously. The overall p-

value is 0.0001, which is highly significant compared with the significance level 0.05. The finding from Table 4.3 is consistent with the conclusion obtained in Chapter 2 (Guo et al. 2022), which employs the MCACE model to analyze the same AHJ dataset under all-or-none compliance. It's worth noting that the overall p-value detected under partial compliance is much lower than that (overall p-value = 0.008) in Chapter 2 (Guo et al. 2022). This is possibly caused by considering continuous partial compliance since dichotomization to all-or-none compliance or discretization to categorical compliance measures results in a loss of information.

Table 4.3: Parameters related to principal causal effects

$\beta_{T11k} - \beta_{C11k}$	Estimate	Standard error	P-value	Overall p-value
$\beta_{T111} - \beta_{C111}$	10.414	14.150	0.462	
$\beta_{T112} - \beta_{C112}$	17.395	22.415	0.438	
$\beta_{T113} - \beta_{C113}$	-14.675	17.089	0.390	
$\beta_{T114} - \beta_{C114}$	88.226	22.597	$< .0001$	0.0001
$\beta_{T115} - \beta_{C115}$	-2.037	17.900	0.909	
$\beta_{T116} - \beta_{C116}$	65.878	22.783	0.004	

Figure 4.3 displays principal causal effects versus intervention compliance $D_i(T)$ in the sixth month for each endpoint. The principal causal effects (solid lines in Figure 4.3) in the sixth month and corresponding confidence intervals (dashed lines in Figure 4.3) are calculated based on the estimates and standard errors in Table 4.3. The critical values of the corresponding confidence intervals are chosen under the significance level 0.008 (adjusted based on Bonferroni correction). Consistent with Table 4.3, the confidence intervals for the fourth and sixth endpoints in Figure 4.3 are above 0. Thus, the principal causal effects at the sixth month for the fourth and sixth endpoints are beneficial and significant. The confidence intervals for the first, second, third and fifth endpoints contained zero. Therefore, there are no statistically significant principal causal effects in the sixth month for these outcomes.

Table 4.4 shows the estimation results for $\boldsymbol{\alpha}$ in the Beta regression model (Eqn 4.3). The estimate for intercept term is -3.264, which indicates that the mean response (μ_{Di}) regarding intervention compliance for younger female participants with longer disease duration and low disease activity is 0.037 ($e^{-3.264}/(1+e^{-3.264})$). We noticed that the estimates for the coefficients of disease duration and gender are negative. This suggests that the mean response μ_{Di} is negatively correlated with disease duration and gender. Similarly, the mean response is positively related to disease activity and age since corresponding estimates are positive. As the covariates in Table 4.4 are binary variables, we conclude that older female patients with high disease activity and longer disease duration tend to use the AHJ tool more frequently (higher mean response μ_{Di}) if they are randomized to the intervention group.

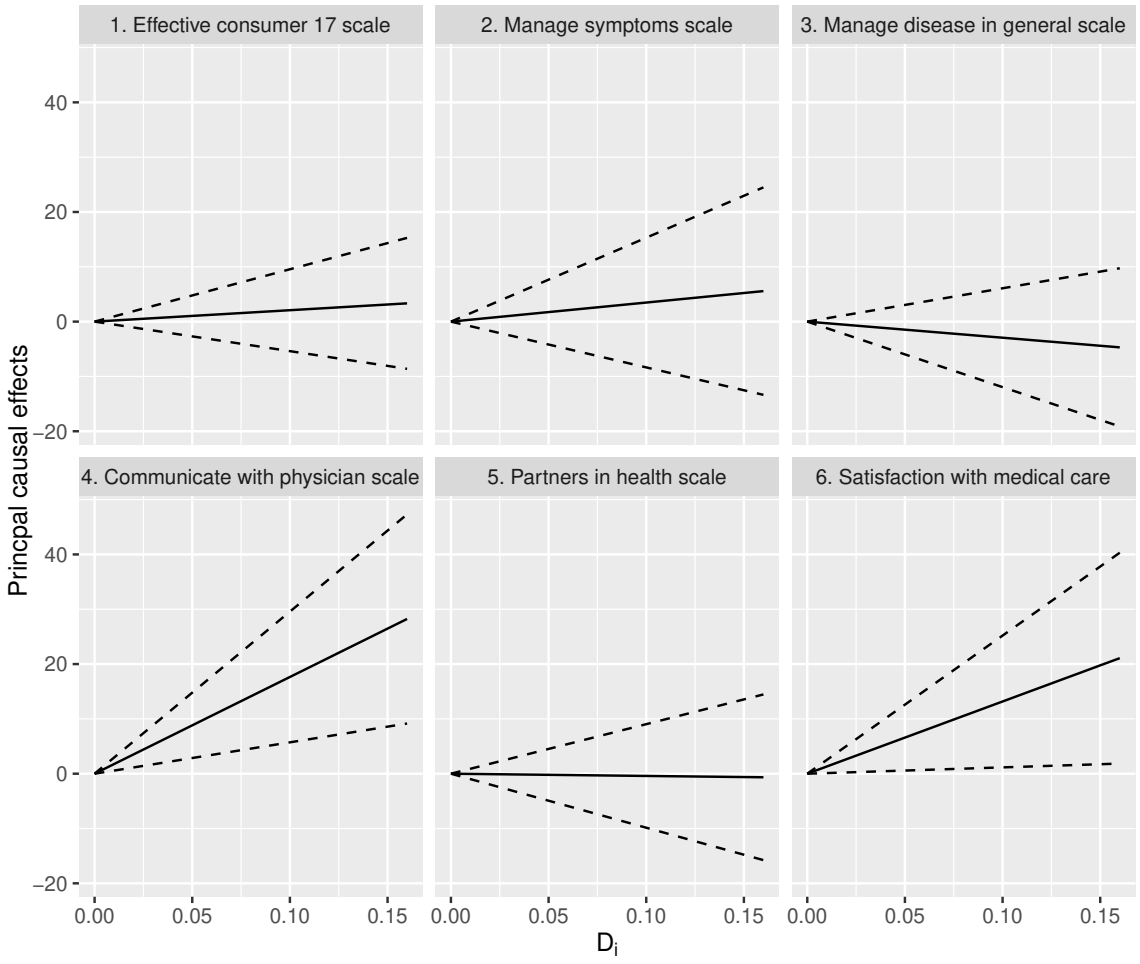

Figure 4.3: Principal causal effects at the sixth month. The solid lines are principal causal effects varying with D_i at the sixth month. The dashed lines are corresponding confidence intervals with adjusted critical values based on Bonferroni correction.

Table 4.4: Estimation results for the Beta regression

Covariates	Estimate	Standard error	P-value
Intercept	-3.264	0.234	$< .0001$
Disease duration	-0.244	0.339	0.473
Disease activity	0.315	0.233	0.176
Gender	-0.402	0.386	0.297
Age	0.066	0.168	0.694

Note: disease duration = 1 if early disease and = 0 if late disease, disease activity = 1 if there are high RAPID4 values and = 0 if there are moderate/low RAPID4 values, gender = 1 if male and 0 if female, age = 1 if above 54.50 years old and = 0 if below 54.50 years old

4.6 Discussion

Under the framework of principal stratification proposed by Frangakis & Rubin (2002), principal strata can be formed based on the potential values of an intermediate outcome variable $D_i(A)$. In the AHJ study, $D_i(A)$ is defined as the frequency rate of subject i's AHJ tool uses within the follow-up period when randomized to group A. When patients in the control group do not have access to the behaviour intervention which implies $D_i(C) \equiv 0$, the principal stratum is uniquely determined by the partially observed variable $D_i(T)$. Instead of considering $D_i(T) = 0$ or 1 under all-or-none compliance, we consider partial compliance where $D_i(T)$ is a continuous variable ranging from 0 to 1. A Beta regression model is employed to model the intervention compliance $D_i(T)$ so that covariates could be considered to help predict the intervention compliance. When modeling primary potential outcomes $\boldsymbol{Y}_i(A_i)$ given the principal stratum defined by $D_i(T)$, we proposed a multivariate hierarchical random-effects approach to model multidimensional potential outcomes in the treatment group $\boldsymbol{Y}_i(T)$ and potential outcomes in the control group $\boldsymbol{Y}_i(C)$ separately. The proposed model can address partial compliance properly and account for the correlations across multiple endpoints over all time points by introducing random effects in stage 2 in the hierarchical random-effects approach. Furthermore, the principal causal effects within stratum S defined by $D_i(T)$ can be easily obtained under our approach.

After verifying the performance of the proposed model in the simulation study, we apply the model to the AHJ data. We detected significant beneficial principal causal effects for the fourth and sixth outcomes after adjusting the multiple testing issues. Moreover, when examining whether there are principal causal effects or not for all six endpoints simultaneously, we noticed that the overall p-value for all six outcomes is highly significant, which implies AHJ tool did help participants manage their disease very well. This finding is consistent with the conclusion obtained in Chapter 2 (Guo et al. 2022) which analyzed the same dataset based on a similar hierarchical random-effects approach under all-or-none compliance. It's notable that the overall p-value is much smaller in the case of partial compliance compared with all-or-none compliance since the dichotomization in all-or-none compliance can lead to loss of information. Besides, the estimation results (Table 4.4) from the Beta regression model establishes the bridge between compliance behavior and baseline covariates. Although none of the covariates is significant at the 5% level, this analysis is still useful if we want to study the compliance behavior of patients based on baseline characteristics in other similar studies.

The proposed model can be extended to other situations. Firstly, intervention compliance is considered as a continuously-measured variable and is modelled through the Beta regression model. If the intervention compliance is treated as count data, Poisson distribution can be applied instead. Secondly, instead of considering the receipt of the intervention only, real-world RCTs may involve a placebo in the control group. For example, the Efron-

Feldman data analyzed by Jin & Rubin (2008) is from a placebo-controlled double-blind randomized controlled trial, where patients in the control group have access to the placebo. If we define $d_i(C)$ as the potential amount of the placebo received by subject i if assigned to the control group, our model can be easily extended to incorporate $d_i(C)$ into Eqn 4.5 as a predictor. Lastly, it maybe restrictive to assume the error terms in Eqn 4.4 and random effects in Eqn 4.5 follow multivariate normal distributions. Other distributions can be considered to increase the applicability of the method.

Chapter 5

Final Discussion

In real-world longitudinal randomized controlled trials (RCTs), noncompliance issue often occurs and multiple outcomes are commonly used to measure complex traits when evaluating multifaceted behaviour interventions. To assess the treatment efficacy of such behaviour intervention efficiently, it's important to tackle the noncompliance issue properly and consider the correlations among multivariate outcomes. This dissertation focuses on developing novel statistical methodologies with improved efficiency to estimate treatment efficacy while addressing these two challenges simultaneously.

Throughout the dissertation, we deal with the noncompliance issue within the principal stratification framework proposed by Frangakis & Rubin (2002). Both all-or-none compliance (binary compliance) and partial compliance (continuously-measured compliance) are discussed here. Under all-or-none compliance, the whole population is divided into four groups: compliers, never-takers, always-takers and defiers. Chapters 2 and 3 considered all-or-none compliance and aimed at estimating CACEs, the causal treatment effects among compliers in the population. Chapter 4 considered partial compliance and focused on estimating causal effects within principal strata defined by combinations of continuous potential intermediate outcomes.

When dealing with multiple outcomes, we employed a hierarchical random-effects approach and a latent-factor approach. Under the hierarchical random-effects approach, chapter 2 developed a multivariate longitudinal potential outcome model stratified on latent compliance types under all-or-none compliance. The random effects in the model combined all information for multiple outcomes across all visits. Chapter 4 extended the hierarchical random-effects approach to estimating causal effects under a partial compliance setting. By introducing latent factors, the latent-factor MCACE model proposed in Chapter 3 leads to dimensionality reduction. Under the model in Chapter 3, high-dimensional outcomes are reduced to low-dimensional latent factors. These latent factors represent underlying constructs measured by high-dimensional outcomes. This allows us to make statistical inference on latent factors (underlying constructs) directly.

In each chapter, simulation studies are conducted to demonstrate the validity and evaluate the performance of these developed novel methodologies. In Chapter 2, compared with univariate complier-average causal effect analysis, the multivariate longitudinal potential outcome model shows a significant gain in the estimation efficiency, including up to 50% reduction in standard errors of CACE estimates and a one-fold increase in the power to detect the CACE. Furthermore, the simulation study in Chapter 3 shows there is a four-fold increase in the power to detect CACE under the latent-factor MCACE model compared with existing univariate CACE analysis. These simulation studies demonstrate that these proposed models can lead to a significant improvement in the estimation efficiency for treatment efficacy in longitudinal RCTs with noncompliance and multiple outcomes.

After applying the proposed methodologies to the AHJ data, some novel findings are presented. In Chapter 2, after Bonferroni correction, the multivariate longitudinal potential outcome model detects beneficial CACEs on the fourth outcome (communicate with a physician) and the sixth outcome (satisfaction with medical care), while univariate CACE analysis fails to detect CACEs on any outcomes. It's worth noting that the multivariate longitudinal potential outcome model detects an overall CACE by giving a significant overall p-value, 0.008. However, the direction of the overall CACE is undetermined. We still need to examine each outcome individually to interpret treatment effects, which leads to multiple testing issues. To overcome these limitations, chapter 3 introduces latent factors which represent underlying constructs measured by multiple outcomes. By making statistical inference on these latent factors directly, the directions of treatment effects on underlying constructs can be determined and multiple testing issues can be mitigated. Specifically, the application in Chapter 3 considers two underlying constructs (self-efficacy and interactions with caregivers). And we found significant and beneficial CACEs on both underlying constructs. In Chapter 4, we analyzed the AHJ data under a hierarchical random-effects approach in the partial compliance setting. Compared with Chapter 2 which analyzes the same dataset under a similar approach in the binary compliance setting, a smaller overall p-value is detected in Chapter 4. This verifies the rationale that considering partial compliance makes use of more information than considering all-or-none compliance since dichotomization in binary compliance can cause a loss of information.

In this dissertation, we assume that potential outcomes within certain principal strata follow multivariate normal distributions. To avoid possible model misspecification brought by prespecifying the multivariate normal distribution, future work may relax this assumption by considering other distributions, such as the multivariate t-distribution. Furthermore, the proposed methodologies mentioned above focus on studying continuous potential outcomes. One possible future extension of current work is to consider discrete and mixed-type responses. In addition, although we only illustrate these proposed methodologies in AHJ data, these models developed in this dissertation can be applied broadly to other similar longitudinal RCTs with noncompliance and multiple outcomes.

Bibliography

An, X., Yang, Q. & Bentler, P. M. (2013), 'A latent factor linear mixed model for high-dimensional longitudinal data analysis', *Statistics in Medicine* **32**(24), 4229–4239.

Angrist, J. D., Imbens, G. W. & Rubin, D. B. (1996), 'Identification of causal effects using instrumental variables', *Journal of the American Statistical Association* **91**(434), 444–455.

Baker, S. G., Kramer, B. S. & Lindeman, K. S. (2016), 'Latent class instrumental variables: a clinical and biostatistical perspective', *Statistics in Medicine* **35**(1), 147–160. [correction 2019; 38(5): 901].

Baker, S. G. & Lindeman, K. S. (1994), 'The paired availability design: a proposal for evaluating epidural analgesia during labor', *Statistics in Medicine* **13**(21), 2269–2278.

Barlow, J., Wright, C., Sheasby, J., Turner, A. & Hainsworth, J. (2002), 'Self-management approaches for people with chronic conditions: a review', *Patient Education and Counseling* **48**(2), 177–187.

Bartolucci, F. & Grilli, L. (2011), 'Modeling partial compliance through copulas in a principal stratification framework', *Journal of the American Statistical Association* **106**(494), 469–479.

Berkhof, J. & Snijders, T. A. (2001), 'Variance component testing in multilevel models', *Journal of Educational and Behavioral Statistics* **26**(2), 133–152.

Choi, H. K., Hernán, M. A., Seeger, J. D., Robins, J. M. & Wolfe, F. (2002), 'Methotrexate and mortality in patients with rheumatoid arthritis: a prospective study', *The Lancet* **359**(9313), 1173–1177.

Efron, B. & Feldman, D. (1991), 'Compliance as an explanatory variable in clinical trials', *Journal of the American Statistical Association* **86**(413), 9–17.

Everitt, B. & Hothorn, T. (2011), *An introduction to applied multivariate analysis with R*, Springer Science & Business Media.

Frangakis, C. E. & Rubin, D. B. (2002), 'Principal stratification in causal inference', *Biometrics* **58**(1), 21–29.

Gao, X., Brown, G. K. & Elliott, M. R. (2014), 'Joint modeling compliance and outcome for causal analysis in longitudinal studies', *Statistics in Medicine* **33**(20), 3453–3465.

Guo, L., Qian, Y. & Xie, H. (2022), 'Assessing complier average causal effects from longitudinal trials with multiple endpoints and treatment noncompliance: An application to a study of arthritis health journal', *Statistics in Medicine* **41**(13), 2448–2465.

Hernán, M. A. & Hernández-Díaz, S. (2012), 'Beyond the intention-to-treat in comparative effectiveness research', *Clinical trials* **9**(1), 48–55.

Hirano, K., Imbens, G. W., Rubin, D. B. & Zhou, X.-H. (2000), 'Assessing the effect of an influenza vaccine in an encouragement design', *Biostatistics* **1**(1), 69–88.

Holland, P. W. (1986), 'Statistics and causal inference', *Journal of the American Statistical Association* **81**(396), 945–960.

Imbens, G. W. & Angrist, J. D. (1994), 'Identification and estimation of local average treatment effects', *Econometrica.* **62**(2), 467–475.

Imbens, G. W. & Rubin, D. B. (1997), 'Bayesian inference for causal effects in randomized experiments with noncompliance', *The Annals of Statistics* pp. 305–327.

Imbens, G. W. & Rubin, D. B. (2015), *Causal inference in statistics, social, and biomedical sciences*, Cambridge University Press, New York.

Jin, H. & Rubin, D. B. (2008), 'Principal stratification for causal inference with extended partial compliance', *Journal of the American Statistical Association* **103**(481), 101–111.

Jo, B. & Muthén, B. O. (2001), Modeling of intervention effects with noncompliance: a latent variable approach for randomized trials, *in* G. A. Marcoulides & R. E. Schumacker, eds, 'New Developments and Techniques in Structural Equation Modeling', Lawrence Erlbaum Associates, Mahwah, NJ, pp. 57–87.

Lacaille, D., Carruthers, E., van As, B., Goldsmith, C. H., Horlock, H., Li, L., Townsend, A. F., Mitchell, B. & Adam, P. (2015), 'Proof of concept study of the arthritis health journal: An online tool to promote self-monitoring in people with rheumatoid arthritis', *American College of Rheumatology Annual Meeting Abstract 2334* .

Laird, N. M. & Ware, J. H. (1982), 'Random-effects models for longitudinal data', *Biometrics* **38**(4), 963–974.

Lee, Y. J., Ellenberg, J. H., Hirtz, D. G. & Nelson, K. B. (1991), 'Analysis of clinical trials by treatment actually received: is it really an option?', *Statistics in Medicine* **10**(10), 1595–1605.

Lin, J. Y., Ten Have, T. R. & Elliott, M. R. (2008), 'Longitudinal nested compliance class model in the presence of time-varying noncompliance', *Journal of the American Statistical Association* **103**(482), 462–473.

Mattei, A., Li, F., Mealli, F. et al. (2013), 'Exploiting multiple outcomes in bayesian principal stratification analysis with application to the evaluation of a job training program', *Annals of Applied Statistics* **7**(4), 2336–2360.

Mealli, F. & Pacini, B. (2013), 'Using secondary outcomes to sharpen inference in randomized experiments with noncompliance', *Journal of the American Statistical Association* **108**(503), 1120–1131.

Meier, P. (1991), 'Compliance as an explanatory variable in clinical trials: comment', *Journal of the American Statistical Association* **86**(413), 19–22.

Roy, J. & Lin, X. (2000), 'Latent variable models for longitudinal data with multiple continuous outcomes', *Biometrics* **56**(4), 1047–1054.

Rubin, D. B. (1978), 'Bayesian inference for causal effects: the role of randomization', *The Annals of Statistics* **6**(1), 34–58.

Rubin, D. B. (1980), 'Randomization analysis of experimental data: the fisher randomization test comment', *Journal of the American Statistical Association* **75**(371), 591–593.

Rubin, D. B. (1990), 'Comment: Neyman (1923) and causal inference in experiments and observational studies', *Statistical Science* **5**(4), 472–480.

Sanders, E., Gustafson, P. & Karim, M. E. (2021), 'Incorporating partial adherence into the principal stratification analysis framework', *Statistics in Medicine* **40**(15), 3625–3644.

Schwartz, S. L., Li, F. & Mealli, F. (2011), 'A bayesian semiparametric approach to intermediate variables in causal inference', *Journal of the American Statistical Association* **106**(496), 1331–1344.

Sheiner, L. B. & Rubin, D. B. (1995), 'Intention-to-treat analysis and the goals of clinical trials', *Clinical Pharmacology & Therapeutics* **57**(1), 6–15.

Shrier, I., Platt, R. W., Steele, R. J. & Schnitzer, M. (2018), 'Estimating causal effects of treatment in a randomized trial when some participants only partially adhere', *Epidemiology* **29**(1), 78–86.

Sommer, A. & Zeger, S. L. (1991), 'On estimating efficacy from clinical trials', *Statistics in Medicine* **10**(1), 45–52.

Steele, R. J., Shrier, I., Kaufman, J. S. & Platt, R. W. (2015), 'Simple estimation of patient-oriented effects from randomized trials: an open and shut cace', *American Journal of Epidemiology* **182**(6), 557–566.

Tam, J., Lacaille, D., Liu-Ambrose, T., Shaw, C., Xie, H., Backman, C. L., Esdaile, J. M., Miller, K., Petrella, R. & Li, L. C. (2019), 'Effectiveness of an online self-management tool, operas (an on-demand program to empower active self-management), for people with rheumatoid arthritis: a research protocol', *Trials* **20**, 1–8.

Van Den Bemt, B. J., Zwikker, H. E. & Van Den Ende, C. H. (2012), 'Medication adherence in patients with rheumatoid arthritis: a critical appraisal of the existing literature', *Expert Review of Clinical Immunology* **8**(4), 337–351.

Xie, H. & Heitjan, D. F. (2004), 'Sensitivity analysis of causal inference in a clinical trial subject to crossover', *Clinical Trials* **1**(1), 21–30.

Yau, L. H. & Little, R. J. (2001), 'Inference for the complier-average causal effect from longitudinal data subject to noncompliance and missing data, with application to a job training assessment for the unemployed', *Journal of the American Statistical Association* **96**(456), 1232–1244.