2023

Variability in Causal Effects on A Binary Outcome and Noncompliance in A Multisite Randomized Trial

Xinxin Sun
Virginia Commonwealth University

VARIABILITY IN CAUSAL EFFECTS ON A BINARY OUTCOME AND NONCOMPLIANCE IN A MULTISITE RANDOMIZED TRIAL

A DISSERTATION SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

BY

XINXIN SUN

RICHMOND, VIRGINIA

2023

DOCTORAL COMMITTEE:

DR. YONGYUN SHIN (ADVISOR)

DR. LE KANG

DR. NITAI MUKHOPADHYAY

DR. STEPHEN W. RAUDENBUSH

DR. JENNIFER ELSTON LAFATA

DEPARTMENT OF BIOSTATISTICS

VIRGINIA COMMONWEALTH UNIVERSITY

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

First of all, I would like to thank Dr. Shin, my advisor, for his devoted mentorship and funding. Dr. Shin's passion for research leaves a strong impression on anyone who works with him. He is always calm and patient in explaining the existing methodologies, coding and new ideas. Whenever I encounter some difficulties during research, he can be counted on to show his advice and support. With his help, I can make progress along the research path continuously.

I would also like to thank my committee members, Dr. Le Kang, Dr. Nitai Mukhopadhyay, Dr. Stephen W. Raudenbush and Dr. Jennifer Elston Lafata. Their dedicated input allows me to develop a better version of the dissertation. Further, I would like to thank Dr. David Wheeler who partially supports my study at the Virginia Commonwealth University.

Lastly, I would like to thank my family for their continued support and encouragement that makes my successful completion of the program possible.

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305D210022 and by the National Cancer Institute through grant R01CA19720. The opinions expressed are those of the authors and do not represent the views of the Institute or the U.S. Department of Education or the National Cancer Institute.

ABSTRACT

Variability in Causal Effects on A Binary Outcome and Noncompliance in A Multisite Randomized Trial

By Xinxin Sun

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at Virginia Commonwealth University.

Virginia Commonwealth University, 2023.

Advisor: Yongyun Shin, Ph.D.,

Associate Professor, Department of Biostatistics

Noncompliance to treatment assignment is widespread in randomized trials and presents challenges in causal inference. In the presence of noncompliance, the most commonly estimated effect of treatment assignment, also known as intent-to-treat (ITT) effect, is biased. Of interest in this setting is the complier average causal effect (CACE), the ITT effect among compliers. Further complication arises when the outcome variable is partially observed.

My research focuses on estimating the distribution of a site-specific CACE in a multisite randomized controlled trial (MRCT) by maximum likelihood (ML). Assuming compliance missing at random (MAR). We express the likelihood as an integral with respect to random effects that cannot be evaluated analytically. We derive ML estimators by a combination of the EM algorithm and Newton Raphson method conditional on random effects and, then, integrate each estimator and the likelihood with respect to the random effects by adaptive Gauss Hermite quadrature (AGHQ). Next, we extend this approach to both outcome and imperfect compliance MAR. A distinctive feature of the approach is to estimate a site-specific CACE and its variance over sites efficiently.

We applied our method to data sets from two MRCTs with imperfect compliance: the e-assist intervention trial to assess the effectiveness of the e-assist intervention in promot-

ing colorectal cancer screening outcome fully observed; and the National Study of Learning Mindsets trial to evaluate the effectiveness of a growth mindset program in improving academic achievement outcomes MAR among ninth-graders.

CHAPTER 1

INTRODUCTION

1.1 Background

Unlike common statistical analysis that focuses on finding associations between outcome and covariates, causal analysis examines the changes in the outcome induced by external interventions or manipulations. For example, one may be interested in the effect of taking an aspirin to relieve someone's headache where taking aspirin is the external intervention and any reduction in headache induced by the intervention is of interest [Rubin, 1974]. A major framework in causal inference is the potential outcomes framework which considers both observed and counterfactual outcomes in the analysis. The formal notion of potential outcomes in a randomized experiment was introduced by Neyman [1923] in an agricultural experiment where he used an urn model to assign crop varieties to plots and estimated averages and the differences in potential yields between two crop varieties by difference in sample means. The concept of potential outcomes has been applied in randomized experiments for years till Rubin [1974] extended it to the observational studies by controlling extraneous variables.

Under the potential outcomes framework, the causal effect of an intervention or a treatment assigned to an individual within a specific time frame is defined as the difference in the potential outcomes of the individual [Rubin, 1974]. In the aspirin experiment, for example, if a subject takes an aspirin and after six hours her headache alleviates, and if she had not taken the pill and her headache would have alleviated after six hours, then the effect of aspirin is 0 (Table 1.1).

Table 1.1: Causal Effect of Taking an Aspirin on Headache

Treatment	Aspirin	No aspirin	Difference
Potential outcome	No headache	No headache	None

However, in a real randomized trial, we observe only one of the potential outcomes of a

subject, because we cannot assign the subject to both treatment and control. Consequently, multiple subjects are required to study causal inference of the average treatment effect [Holland, 1986, Imbens and Rubin, 2015]. Therefore, randomization is a critical procedure in a study because it generates two comparable groups of subjects that can be compared as random samples drawn from the target population.

Although randomization provides valid study results by minimizing allocation bias [Schulz and Grimes, 2002], in real experiments, the benefits of randomization are often undermined by noncompliance when study participants do not comply with their assignments. When compliance is perfect, the difference in the mean outcomes between the treatment and control arms is an unbiased estimate of the effect of the treatment. In the presence of noncompliance, this is no longer true. In reality, noncompliance to treatment assignment is widespread in randomized trials. Bloom [1984] illustrated a study of job counselling service for dislocated workers in Delaware, in which around 25% study eligibles never showed up for the service. Another example of noncompliance is the study of the impact of vitamin A on childhood mortality, where around 20% subjects in the treatment group failed to receive vitamin A as prescribed [Sommer et al., 1986, Sommer and Zeger, 1991]. In the Prostate Cancer Prevention Trial (PCPT) in 1994, Feigl et al. [1995] expected 5% noncompliance in the placebo arm and 14% nonadherance in the treatment arm when doing power calculation.

In the vitamin A study, the subjects in the treatment arm did not receive the treatment and, hence, failed to comply with their assignments. This kind of noncompliance is categorized as all-or-none compliance, as opposed to partial compliance by Baker [2004]. Specifically, Meier [1991] defined two sources of noncompliance, selection and treatment effects. The noncompliance due to selection effect usually occurs to subjects who are ill-disciplined and or expected to have poor outcomes, and can be considered as pre-treatment variable, whereas the noncompliance due to treatment effect is the result of the side effects of a treatment, and hence this type of noncompliance is considered as a post-treatment variable.

All-or-none compliance is a special type of noncompliance due to selection effect.

In a randomized trial where all subjects comply with their assignments, the intent-to-treat (ITT) effect is the treatment effect. However, when some subjects fail to comply with their assignments, the ITT effect only reveals the impact of assignment and produces a biased estimate of the treatment effect [Frangakis and Rubin, 1999, Imbens and Rubin, 2015]. In the presence of noncompliance, researchers often estimate the per-protocol (PP) effect of treatment assignment after dropping observed noncompliers and as-treated (AT) effect of treatment received in addition to the ITT effect. Both PP and AT analyses fail to keep the benefits of randomization and, thus, produce biased estimates of the treatment effect [Little et al., 2009].

With imperfect compliance, a better alternative to ITT, PP and AT effects is the mean treatment effect among compliers, also known as the complier average causal effect (CACE) [Angrist et al., 1996]. The CACE measures the average causal effect of the treatment in the subpopulation of compliers who participate in an intervention if assigned to it, and who would not otherwise. It is also known as the local average treatment effect (LATE) in the econometrics literature [Angrist and Imbens, 1995, Imbens and Rubin, 1997b]. With compliance as a pretreatment covariate, the CACE estimate will be causal and unbiased [Imbens and Rubin, 2015]. However, compliance is a latent variable [Imbens and Rubin, 1997b, Little and Yau, 1998] under the potential outcomes framework, which makes the estimation of the CACE challenging.

1.2 Existing Approaches

A major approach to estimate the average causal effect is the instrumental variable (IV) method from the econometrics literature. Angrist and Imbens [1995] discussed the identification and estimation of the average causal effect by instrumental variable, and Angrist et al. [1996] proposed a framework to identify and estimate CACE by the IV approach within

the Rubin Causal Model (RCM) [Holland, 1986] under specific assumptions, and applied the approach to study the impact of military service on civilian mortality [Hearst et al., 1986].

The IV approach is often used when the causal relationship between explanatory and outcome variables cannot be directly estimated. Thus, researchers use another variable as an instrument which does not have direct casual effect on the outcome, but indirectly influences the outcome through its effect on the explanatory variable. In a two-arm randomized trial, the instrument is the treatment assignment (T) which is absent of a direct effect on a primary outcome (Y), but passes its effect through treatment receipt ($D(T)$) on Y. The IV method consists of two steps: calculate the ITT (ITT_Y) effect of T on Y and the ITT (ITT_D) effect of T on D. The CACE is defined as their ratio $\frac{ITT_Y}{ITT_D}$ [Imbens and Rubin, 2015].

Although the IV approach is easy to implement, it does not estimate the marginal distribution of the outcome given treatment assignment among compliers, but only estimates the CACE. Based on the work of Angrist et al. [1996], Imbens and Rubin [1997a] presented Bayesian inference for CACE, described estimating the posterior distribution of the CACE by the EM and data augmentation algorithms, and applied the method to the causal effect of vitamin A supplement on child mortality [Sommer et al., 1986]. Imbens and Rubin [1997b] also estimated the distribution of the outcome given treatment assignment T for compliers under certain assumptions and presented the maximum likelihood (ML) estimators as alternatives to the IV estimator. Furthermore, they also compared the IV and ML estimates with the data from a "return to high school" study investigating the effect of years of education on earnings. Later, Little and Yau [1998] conceived of noncompliance as a missing data problem and estimated the CACE of a job training intervention on a depression score with compliance assumed missing at random (MAR). Cheng [2009] estimated the effect of an encouragement intervention to improve adherence to prescribed depression treatments among depressed elderly patients in PROSPECT study [Bruce et al., 2004, Ten Have et al., 2004] by ML approach. The outcomes of the study are the number of depression symptoms

and the depression levels (none, minor, major) measured at 4 months, and they considered both outcomes follow multinomial distributions.

These approaches have been extended to two-level studies. Small et al. [2006] extended the MAR framework to a hierarchical model for the average CACE of an encouragement intervention on a binary depression outcome. They estimated the effect of encouragement as an instrumental variable (IV) on the adherence of patients to prescribed depression treatments by a logistic regression at the first stage. Subsequently, they estimated a logistics mixed effect model for the longitudinal outcome that depended on the fitted adherence and compliance, estimating time-specific fixed CACEs. Raudenbush et al. [2012] estimated hierarchical models for the school-specific impacts of an algebra program intervention on participation and math achievement separately, combining the estimates to produce the mean and variance of the school-specific CACE on math achievement of ninth graders. This approach assumes independence between school-specific compliance and CACE. Walters [2015] estimated a bivariate random coefficients model for the causal effects of a randomized Head Start program attendance (vs denial to attendance) on a binary program participation and a continuous cognitive skills outcome by ML. The parameters of random coefficients included the mean and variance of CACE as well as the covariance between site-specific compliance and CACE. This approach, however, assumes that the within-site latent variable underlying participation and potential outcomes are multivariate normal, a strong and uncheckable assumption [Little et al., 2009]. Zhou et al. [2019] conducted a meta-analysis of studies comparing the rates of a cesarean section outcome among women in labor between epidural analgesia intervention and control (no or other analgesia treatment). With noncompliance to treatment assignment, the authors estimate the CACE of epidural analgesia on the cesarean delivery outcome by a logistic random intercept model where studies were conceived as clusters.

Outcome missingness is also a common problem in a randomized trial due to dropout or nonresponse at later assessments. The problem of missing outcomes complicates the

CACE analysis by adding more uncertainties. Researchers extended the CACE analysis framework [Angrist et al., 1996] to cover the missing outcomes under specific assumptions. Frangakis and Rubin [1999] showed that the common ITT estimator may be subject to bias if compliance is related to the missing pattern of the outcome, and suggested CACE estimator under assumption of latent ignorability. Latent ignorability implies that the outcome and outcome missing mechanism are independent within each level of compliance. Assuming outcome missing at random, Yau and Little [2001] extended the ML approach in Little and Yau [1998] to the partially observed outcome in the context of the JOBS II intervention trial [Vinokur et al., 1995]. They modeled the unobserved compliance and missing outcome data jointly which was estimated by the EM algorithm [Dempster et al., 1977]. Dunn et al. [2003] estimated the CACE of a psychological intervention for the treatment of depression measured by Beck Depression Inventory score in primary care, under the latent ignorabilty assumption.

In our case, the missing pattern of compliance depends on treatment assignment and, thus, compliance MAR is reasonable. However, the missing pattern of the outcome may depend on missing values of outcome or compliance. We assume that the pattern is independent of the outcome and compliance given all observed data to estimate our model under the MAR assumption [Frangakis and Rubin, 1999, Jo et al., 2010].

1.3 Motivation

In our study, we extend the previous approaches to a two-level multisite randomized controlled trial (MRCT) where there is a distribution of site-specific treatment effects. Our key contribution is to a multisite trial in this setting where each site generates its own causal effect.

One motivating example is the e-assist trial, a colon health study [Lafata et al., 2019]. The intervention is called "e-assist", an online program designed to improve the colorectal

cancer screenings (CRCS) among patients who are at risk of colorectal cancer. An MRCT was conducted to test the efficacy of the intervention. Eligible patients were randomly assigned to e-assist or control within physicians. A physician is conceived as a site. In the e-assist trial, a patient assigned to control was not allowed to access e-assist. Consequently, one assigned to e-assist is a complier if the patient participated in the intervention or a never taker otherwise. On the other hand, one assigned to control could be a complier if the patient would have participated in the intervention or a never taker otherwise under the hypothetical alternative assignment to e-assist. Therefore, compliance of a control is missing. We are interested in estimating the joint distribution of physician-specific complier average causal effect (CACE) and mean CRCS rates of never takers and compliers assigned to each arm.

The other motivating example comes from the National Study of Learning Mindsets (NSLM). The study assessed the effect of an online growth mindset intervention which teaches students in the secondary education in the US that intellectual abilities can be developed, leading to improved academic performance. In the study, students were randomized to either intervention or control within their schools, and their GPAs were collected before and after the intervention for evaluation. Similar to the e-assist trial, the students in the control arm could not access the intervention, and the compliance of the control is missing. Furthermore, the outcome, student GPA, was partially observed to make efficient estimation challenging.

The literature provides little guidance on how to study the joint distribution of site-specific CACEs and mean responses. We reasonably assume compliance MAR because it is missing for one assigned to control and, thus, depends on treatment assignment [Little and Yau, 1998]. We estimate a bivariate random coefficients model for the binary outcome and compliance by ML where site-specific mean outcomes of never takers, control compliers, and treatment compliers are expressed as coefficients randomly varying over sites. The CACE

is defined as the difference in mean outcomes between treatment and control compliers. We estimate the joint model efficiently by all observed data given partially observed outcome and compliance under the MAR assumption.

1.4 Thesis Objectives

Firstly, we build the bivariate distribution of primary outcome and compliance where site-specific CACEs and mean responses of compliers and noncompliers assigned to each arm are expressed as coefficients varying randomly over heterogeneous sites, and estimate the bivariate random coefficients model by ML. Because it is difficult to obtain close-form ML estimators from observed data, we derive the likelihood and estimators by a combination of the EM algorithm and Newton Raphson (NR) given random effects, and, then, integrate the likelihood and each estimator with respect to the random effects numerically by adaptive Gauss Hermite quadrature (AGHQ). Furthermore, we introduce a shared random effect modeling approach to alleviate the computation burden and multi-collinearity induced by multivariate random effects and to test nested models.

Secondly, we assess our ML estimators by simulations, manipulating compliance rates, cluster sample sizes and the number of clusters, and then apply the estimators to the e-assist study, compare the estimated CACE with ITT, PP and AT estimates, and illustrates estimation and interpretation of variabilities and covariances in site-specific CACE and mean responses of compliers and noncompliers assigned to treatment arms.

Thirdly, we extend the aforementioned approach to partially observed outcome and compliance under the MAR assumption. We evaluate the updated ML estimators with simulations likewise and apply it to the NSLM.

1.5 Thesis Outline

In the next chapter, we explain our approach of the causal inference for an MRCT. Chapter 3 describes application of the approach to the e-assist trial and simulation studies. Chapter 4 presents the application to the NSLM and related simulation studies. Chapter 5 summarizes the strengths and limitations of the approach.

CHAPTER 2

CAUSAL ANALYSIS IN A MULTISITE RANDOMIZED CONTROLLED TRIAL

We consider a two-arm multisite randomized trial like the e-assist study, with N patients nested in J sites (physicians). First we clarify our causal assumptions in section 2.1 and then present our model and likelihood in section 2.2. Section 2.3 shows how we estimate the parameters of the model.

2.1 Causal Assumptions

Let P_{ij} be the identification number of site j to which subject i is assigned, and $T_{ij} = 0$ (1) if the subject is assigned to control (treatment), for $i = 1, \ldots, n_j$, $j = 1, \ldots, J$ and $N = \sum_{j=1}^{J} n_j$, so that $\mathbf{P} = (P_{11}, P_{21}, \ldots, P_{n_J J}) = \mathbf{p}$ and $\mathbf{T} = (T_{11}, T_{21}, \ldots, T_{n_J J}) = \mathbf{t}$ for the entire population. A potential participation $D_{ij}(\mathbf{t}, \mathbf{p})$ is equal to 0 (1) if the subject receives control (treatment), and $\mathbf{D} = (D_{11}(\mathbf{t}, \mathbf{p}), D_{21}(\mathbf{t}, \mathbf{p}), \ldots, D_{n_J J}(\mathbf{t}, \mathbf{p})) = \mathbf{d}$ for the entire population. A potential outcome of a subject is $Y_{ij}(\mathbf{t}, \mathbf{d}, \mathbf{p})$ [Rubin, 1978].

In order to make causal inference using Rubin's causal inference framework [Imbens and Rubin, 1997a,b, 2015], we make the following assumptions [Angrist et al., 1996, Little and Yau, 1998, Imbens and Rubin, 2015], assuming that patients nested within physicians conceived as sites:

Assumption 1: *intact sites* [Hong and Raudenbush, 2006].

Because we cannot randomly assign patients to physicians, we consider the observed assignments of patients to physicians, but do not consider the vast majority of unlikely hypothetical assignments of patients to physicians. A patient is assumed to stay assigned to

the physician for the duration of the study. This assumption implies

$$D_{ij}(\mathbf{t}, \mathbf{p}) = D_{ij}(\mathbf{t})$$
$$Y_{ij}(\mathbf{t}, \mathbf{d}, \mathbf{p}) = Y_{ij}(\mathbf{t}, \mathbf{d}).$$

Assumption 2: *no interference between subjects* [Rubin, 1980].

This assumption is also known as the stable unit treatment value assumption (SUTVA) [Rubin, 1980]. It means that the treatment of one subject does not affect the potential outcomes of another subject.

$$D_{ij}(\mathbf{t}) = D_{ij}(t_{ij})$$
$$Y_{ij}(\mathbf{t}, \mathbf{d}) = Y_{ij}(t_{ij}, d_{ij}).$$

Assumption 3: *monotonicity.*

It is also called no-defier assumption. It means there is no defier who unreasonably receives the opposite of what is assigned no matter what, implying

$$D_{ij}(1) \geq D_{ij}(0).$$

In a randomized experiment, it is often plausible that there are no individuals who would take the opposite of what's assigned. It, however, is reasonable to think of the assignment to treatment as incentivising one to take the assigned treatment. These incentives need not be strong enough to induce everybody to take the treatment, but in many situations these incentives would rarely induce individuals to take the opposite of their assignments [Imbens and Rubin, 2015].

Assumption 4: *random assignment.*

Treatment assignment is random so that the potential outcomes and participations are

independent of the treatment assignment mechanism. The random assignment assumption reflects the unconfoundedness assumption for regular assignment mechanisms, implying the assignment of one subject does not influence potential outcomes of any other subject.

Assumption 5: *nonzero average causal effects of* T_{ij} *on* D_{ij}.

The expected difference in the proportion of subjects receiving treatment between those assigned to treatment and control is nonzero. That is,

$$E[D_{ij}(1) - D_{ij}(0)] \neq 0.$$

Along with the monotonicity assumption, it implies that the expected proportion of subjects participating in a treatment group is greater in the treatment arm than in the control arm.

Assumption 6: *exclusion restriction.* A potential outcome given treatment received does not depend on treatment assignment:

$$Y_{ij}(t, D_{ij}) = Y_{ij}(D_{ij}(t)), \quad t = 0, 1.$$

It states that given the treatment receipt, the treatment assignment does not influence the potential outcomes of a subject [Angrist et al., 1996]. This assumption is often justifiable by a study design. For a double-blinded trial as an example, neither physicians or patients are aware of the treatment arm to which a patient is assigned, and hence the assignment mechanism is not able to affect the potential outcomes of the patients.

In a two-arm randomized trial, the population of interest can be decomposed into compliers and three noncomplier groups based on the actual treatment assignment and receipt, and the hypothetical alternative assignment and receipt: i) compliers who would participate in the intervention if assigned to it and would not otherwise; ii) defiers who would take the opposite of what is assigned no matter what; iii) always takers who would find a way to participate in the intervention no matter what; and iv) never takers who would not

participate in the intervention no matter what [Angrist et al., 1996]. Let C_{ij} denote the compliance of patient i within site j. Table 2.1 (a) summarizes the compliance categories by T and $D(T)$ on a population level [Little and Yau, 1998]. In sample data we observe T_{ij} and $D_{ij} = T_{ij}D_{ij}(1) + (1 - T_{ij})D_{ij}(0)$. By monotonicity assumption, we overrule the existence of defiers who would take the opposite of what is assigned no matter what. Since compliance is missing when $T_{ij} = D_{ij}$, the missing pattern of compliance is reasonably MAR given T_{ij} and D_{ij} in Table 2.1 (b). The exclusion restriction assumption implies no causal effect of a never or always taker: $Y_{ij}(D_{ij}(1)) - Y_{ij}(D_{ij}(0)) = 0$.

Table 2.1: Compliance based on Treatment Assignments and Receipts

C_{ij}	$D_{ij}(1)$	$D_{ij}(0)$
complier (c)	1	0
never taker (n)	0	0
always taker (a)	1	1
defier (d)	0	1

(a)

		D_{ij}	
		1	0
T_{ij}	1	c or a	n
	0	a	c or n

(b)

2.2 Model and Likelihood

2.2.1 Model

The monotonicity assumption rules out the defiers, and hence the compliance becomes trinomial given fully observed T and D. Therefore, the compliance of the patients who are assigned to treatment and receive control and the patients who are assigned to control but receive treatment are observable in the off-diagonal cells of Table 2.1 (b). The patients who are assigned to and receive treatment or control are a mixture of compliers and noncompliers in the diagonal cells of the Table. To handle missing compliances efficiently, we jointly model the binary outcome Y_{ij} and compliance C_{ij}.

We model Y_{ij} given observed compliance in a hierarchical logistic random coefficients

model. Let C_n, C_a and C_c indicate a never taker, an always taker, and a complier, respectively. A vector of the compliances of n_j patients within site j (C_j) can be decomposed as observed (C_{oj}) and missing compliance (C_{mj}). Hence, the model for outcome Y_{ij} is

$$\begin{aligned} f(Y_{ij}|C_{oij}, u_j) &= \pi_{ij}^{y_{ij}}(1-\pi_{ij})^{1-y_{ij}} \\ &= exp[y_{ij} log\frac{\pi_{ij}}{1-\pi_{ij}} + log(1-\pi_{ij})] \\ &= exp[y_{ij}\eta_{yij} - log(1+e^{\eta_{yij}})] \end{aligned} \tag{2.1}$$

for

$$\begin{aligned} \eta_{yij} &= log\left(\frac{\pi_{yij}}{1-\pi_{yij}}\right) \\ &= A_{ij}^T\alpha + B_{ij}^T\mathbf{u}_j \end{aligned}$$

where $A_{ij}^T = \begin{pmatrix} C_{nij} & C_{aij} & C_{cij}(1-T_{ij}) & C_{cij}T_{ij} & W_{ij}^T \end{pmatrix}$ is a vector of covariates having main and interactive fixed effects $\begin{pmatrix} \alpha_n & \alpha_a & \alpha_{c0} & \alpha_{c1} & \alpha_w^T \end{pmatrix}$, and $B_{ij}^T = \begin{pmatrix} C_{nij} & C_{aij} & C_{cij}(1-T_{ij}) & C_{cij}T_{ij} \end{pmatrix}$ has main and interactive random effects $\mathbf{u}_j = \begin{pmatrix} u_{nj} \\ u_{aj} \\ u_{c0j} \\ u_{c1j} \end{pmatrix} \sim N\left(0, \tau = \begin{pmatrix} \tau_{nn} & \tau_{na} & \tau_{nc0} & \tau_{nc1} \\ \tau_{na} & \tau_{aa} & \tau_{ac0} & \tau_{ac1} \\ \tau_{nc0} & \tau_{ac0} & \tau_{c0c0} & \tau_{c0c1} \\ \tau_{nc1} & \tau_{ac1} & \tau_{c0c1} & \tau_{c1c1} \end{pmatrix}\right)$ for pre-treatment covariates W_{ij}. Here α_n, α_a, α_{c0} and α_{c1} are the mean outcome log odds of never takers, always takers and compliers assigned to control and treatment, respectively, and u_{nj}, u_{aj}, u_{c0j} and u_{c1j} account for random deviations of the site-specific log odds of the four compliance groups from the mean after controlling for the effects α_w of other pretreatment covariates W_{ij}. The site-

specific CACE is

$$(\alpha_{c1} - \alpha_{c0}) + (u_{c1j} - u_{c0j}) \sim N(\alpha_{c1} - \alpha_{c0}, \tau_{c0c0} + \tau_{c1c1} - 2\tau_{c0c1}) \tag{2.2}$$

We model compliance by a hierarchical logistic regression model

$$\begin{aligned} f(C_{ij}|\mathbf{b}_j) &= \pi_{aij}^{C_{aij}} \pi_{cij}^{C_{cij}} (1 - \pi_{aij} - \pi_{cij})^{1 - C_{aij} - C_{cij}} \\ &= exp(C_{aij}\eta_{aij} + C_{cij}\eta_{cij} - log(1 + e^{\eta_{aij}} + e^{\eta_{cij}})) \end{aligned} \tag{2.3}$$

for

$$\begin{aligned} \eta_{aij} &= log\left(\frac{\pi_{aij}}{\pi_{nij}}\right) = X_{ij}^T \gamma_a + Z_{ij}^T b_{aj}, \\ \eta_{cij} &= log\left(\frac{\pi_{cij}}{\pi_{nij}}\right) = X_{ij}^T \gamma_c + Z_{ij}^T b_{cj}, \end{aligned}$$

and the probabilities π_{nij}, π_{aij} and π_{cij} of a never taker, an always taker and a complier, respectively. Here X_{ij} and Z_{ij} are vectors of pre-treatment covariates having fixed effects γ and site-specific random effects $\mathbf{b}_j = \begin{pmatrix} b_{aj} \\ b_{cj} \end{pmatrix} \sim N\left(0, \Delta = \begin{pmatrix} \Delta_{aa} & \Delta_{ac} \\ \Delta_{ac} & \Delta_{cc}, \end{pmatrix}\right)$ respectively.

2.2.2 Likelihood

Let Y_j denote all outcomes within site j to express the likelihood $L(\theta) = \prod_j L_j$ of $\theta = (\alpha, \gamma, \tau, \Delta)$ for

$$L_j = \iint f(Y_j, C_{oj}|u_j, b_j)\phi(u_j|0, \tau)\phi(b_j|0, \Delta)du_j db_j \tag{2.4}$$

where $\phi(\cdot|\mu, \sigma^2)$ is the probability density function (pdf) of $N(\mu, \sigma^2)$ and

$$
\begin{aligned}
f(Y_j, C_{oj}|u_j, b_j) &= \prod_{i=1}^{n_j} f(Y_{ij}|C_{oij}, u_j) f(C_{oij}|b_j) \\
&= \prod_{\{i:T_{ij}=1, D_{ij}=0\}} f(Y_{ij}|C_n, u_j)\pi_{nij} \prod_{\{i:T_{ij}=0, D_{ij}=1\}} f(Y_{ij}|C_a, u_j)\pi_{aij} \\
&\quad \prod_{\{i:T_{ij}=0, D_{ij}=0\}} [f(Y_{ij}|C_n, u_j)\pi_{nij} + f(Y_{ij}|C_c, u_j)\pi_{cij}] \\
&\quad \prod_{\{i:T_{ij}=1, D_{ij}=1\}} [f(Y_{ij}|C_a, u_j)\pi_{aij} + f(Y_{ij}|C_c, u_j)\pi_{cij}].
\end{aligned} \tag{2.5}
$$

The joint distribution consists of four products: the first consists of never takers assigned to a treatment arm but do not participate; the second one comprises always takers assigned to a control group, but participate in an intervention group; the third includes never takers or compliers who are assigned to and receive a control; the last one includes always takers and compliers who are assigned to and participate in an intervention group. We are not able to observe the compliance of every subject in the third and forth products whose treatment assignment and receipt are identical.

We approximate the integral (2.4) by AGHQ. Let $\nu_j = [u_j^T \; b_j^T]^T$ be of length r, $f(\nu_j) = \phi(u_j|0, \tau)\phi(b_j|0, \Delta)$, and $h(\nu_j) = f(Y_j, C_{oj}|\nu_j)f(\nu_j)$, and we find $\tilde{\nu}_j = E(\nu_j|Y_j, C_{oj})$, $V_j = Var(\nu_j|Y_j, C_{oj})$ and lower triangular L_{vj} from Cholesky decomposition $2V_j = L_{\nu j}L_{\nu j}^T$ from the previous iteration. We use a change of variable $x_{\nu j} = L_{\nu j}^{-1}(\nu_j - \tilde{\nu}_j)$, Q weights $\mathbf{W} = (w_1, \ldots, w_Q)$, Q abscissas $\mathbf{A} = (a_1, \ldots, a_Q)$, and $\sum_{qr,\ldots,q1} = \sum_{qr=1}^{Q} \cdots \sum_{q1=1}^{Q}$ to

approximate Equation (2.4) by

$$\begin{aligned} L_j &= \int \frac{\phi(\nu_j|\tilde{v}_j, V_j)}{\phi(\nu_j|\tilde{v}_j, V_j)} h(\nu_j) d\nu_j \\ &= |L_{vj}| \int e^{-x_{vj}^T x_{vj}} e^{x_{vj}^T x_{vj}} h(L_{vj} x_{vj} + \tilde{v}_j) dx_{vj} \\ &\approx |L_{vj}| \sum_{qr \ldots q1} w_{q1} \ldots w_{qr} e^{A_{q1\ldots qr}^T A_{q1\ldots qr}} h(z_{q1\ldots qrj}) \end{aligned} \tag{2.6}$$

where $z_{q1\ldots qrj} = L_{vj} A_{q1\ldots qr} + \tilde{v}_j$ for $A_{q1\ldots qr} = [a_{q1} \ldots a_{qr}]$.

2.2.3 One-sided Noncompliance

The scenario we discussed above is called two-sided compliance, because the patients in both control and treatment arms may not comply with the assignment. Another possible scenario is that only patients assigned to the treatment can potentially circumvent their assigned treatment and receive the control while others assigned to the control are not allowed to access the treatment. This is called one-sided noncompliance. In this setting, always takers are forced to become compliers while defiers are forced to act like never takers [Little and Yau, 1998]. The one-sided noncompliance is displayed in table 2.2.

Table 2.2: One-sided Noncompliance

		D_{ij}	
		1	0
T_{ij}	1	c	n
	0	-	c or n

We adjust the model and likelihood accordingly. The model of interest is

$$f(Y_{ij}|C_{ij}, u_j) \sim Bin(1, \pi_{yij}), \quad \eta_{yij} = logit(\pi_{yij}) = A_{ij}^T \alpha + B_{ij}^T u_j \tag{2.7}$$

where $A_{ij}^T = \begin{pmatrix} C_{nij} & C_{cij}(1 - T_{ij}) & C_{cij} T_{ij} & W_{ij}^T \end{pmatrix}$ is a vector of covariates having main and

interactive fixed effects $\begin{pmatrix} \alpha_n & \alpha_{c0} & \alpha_{c1} & \alpha_w^T \end{pmatrix}$, and $B_{ij}^T = \begin{pmatrix} C_{nij} & C_{cij}(1-T_{ij}) & C_{cij}T_{ij} \end{pmatrix}$ has main and interactive random effects $u_j = \begin{pmatrix} u_{nj} \\ u_{c0j} \\ u_{c1j} \end{pmatrix} \sim N \left(0, \tau = \begin{pmatrix} \tau_{nn} & \tau_{nc0} & \tau_{nc1} \\ \tau_{nc0} & \tau_{c0c0} & \tau_{c0c1} \\ \tau_{nc1} & \tau_{c0c1} & \tau_{c1c1} \end{pmatrix} \right)$.
The compliance model by a hierarchical logistic regression model becomes

$$f(C_{ij}|b_j) \sim Bin(1, \pi_{cij}), \quad \eta_{cij} = logit(\pi_{cij}) = X_{ij}^T\gamma + Z_{ij}^T b_j, \tag{2.8}$$

where $b_j \sim N(0, \Delta)$ as before. The joint distribution of Y_j and C_{oj} is

$$\begin{aligned} f(Y_j, C_{oj}|u_j, b_j) = & \prod_{\{i:T_{ij}=1, D_{ij}=0\}} f(Y_{ij}|0, u_j)(1-\pi_{cij}) \prod_{\{i:T_{ij}=1, D_{ij}=1\}} f(Y_{ij}|1, u_j)\pi_{cij} \\ & \prod_{\{i:T_{ij}=0, D_{ij}=0\}} [f(Y_{ij}|0, u_j)(1-\pi_{cij}) + f(Y_{ij}|1, u_j)\pi_{cij}]. \end{aligned} \tag{2.9}$$

Because the population is now dichotomized into never takers and compliers, the joint distribution now consists of three products: the first consists of never takers who are assigned to an intervention group, but receives a control; the second consists of compliers who are assigned to and participate in an intervention group; and the final product includes never takers and compliers who are assigned to receive a control.

2.2.4 Shared Random Effects

The random effects may be near the boundary of the parameter space with a near-zero variance or high correlations to slow the convergence to ML. To test this hypothesis, we apply shared random effects by factor loadings to Equations (2.7) and (2.8), extending the method of Miyazaki and Frank [2006] to discrete outcomes. This approach will also make computation of numerical integration (2.6) by AGHQ efficient by the reduced dimension of

random effects.

Let $u_j' \sim N(0, \tau')$ be $r\,(<= 3)$ shared random effects such that $u_j = \Lambda u_j'$ for a $3 \times r$ factor loading matrix Λ. Integral (2.4) is now with respect to u_j' of a reduced dimension. With $r = 3$, Λ is a 3-by-3 identity matrix to imply $u_j = u_j'$. We consider two likely scenarios with u_j' of a dimension $r = 1$ or 2.

We consider the first likely scenario of $r = 1$ where never takers, control compliers and treatment compliers share a single random effect. Preliminary analysis of e-assist study implies that the random effect of observed never takers has a comparatively large variance to be away from zero and shared in our applications. Therefore, $u_j = \Lambda u_j' \sim N(0, \tau)$, for

$$\Lambda = \begin{pmatrix} 1 \\ \lambda_1 \\ \lambda_2 \end{pmatrix}, \quad u_j' = u_{nj}, \quad \tau = \Lambda \tau' \Lambda^T = \tau_{nn} \begin{pmatrix} 1 & \lambda_1 & \lambda_2 \\ \lambda_1 & \lambda_1^2 & \lambda_1\lambda_2 \\ \lambda_2 & \lambda_1\lambda_2 & \lambda_2^2 \end{pmatrix} \tag{2.10}$$

having three parameters, less than six parameters of the unconstrained τ. Now, the random effect of control compliers is $u_{c0j} = \lambda_1 u_{nj} \sim N(0, \lambda_1^2 \tau_{nn})$ and the random effect of treatment compliers is $u_{c0j} = \lambda_2 u_{nj} \sim N(0, \lambda_2^2 \tau_{nn})$.

The next likely scenario is $r = 2$ where there are two separate random effects in the Y model. In the context of e-assist study, the random effect of never takers is, this time, shared with e-assist compliers such that $u_j = \Lambda u_j' \sim N(0, \tau)$ for

$$\Lambda = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ \lambda & 0 \end{pmatrix}, \quad u_j' = \begin{pmatrix} u_{nj} \\ u_{c0j} \end{pmatrix}, \quad \tau = \Lambda \tau' \Lambda^T = \begin{pmatrix} \tau_{nn} & \tau_{nc0} & \lambda\tau_{nn} \\ \tau_{nc0} & \tau_{c0c0} & \lambda\tau_{nc0} \\ \lambda\tau_{nn} & \lambda\tau_{nc0} & \lambda^2\tau_{nn} \end{pmatrix} \tag{2.11}$$

having 4 parameters. The random effect of treatment compliers is then λu_{nj} and its variance is $\lambda^2 \tau_{nn}$. We can have other ways of sharing random effects by a different Λ. For example, if

we want treatment compliers sharing the random effect of control compliers, then, we form $\Lambda = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & \lambda \end{pmatrix}$. We can try difference sharings and then pick the best model by AIC.

2.3 Estimation

We view (Y_j, C_j, u_j', b_j) as complete data (CD) and (Y_j, C_{oj}) observed to estimate (τ', Δ) by the EM algorithm [Dempster et al., 1977, 1981] and $\Gamma = (\alpha^T, \gamma^T, \lambda^T)$ by the Newton Raphson (NR) method for a vector λ of factor loadings. The expected CD ML estimators $\hat{\tau} = \frac{\sum_j E(u_j u_j^T | Y_j, C_{oj})}{J}$ and $\hat{\Delta} = \frac{\sum_j E(\Delta_j \Delta_j^T | Y_j, C_{oj})}{J}$.

Let φ_τ and φ_Δ be vectors of distinct elements of τ and Δ, respectively, such that $\theta = [\Gamma^T \ \varphi_\tau^T \ \varphi_\Delta^T]^T$. Denote the loglikelihood $l = \sum_j l_j$, score $S = \sum_j S_j$ and observed information $I_O = -H = -\sum H_j$ for $l_j = \ln L_j$, $S_j = L_j^{-1} \frac{\partial L_j}{\partial \theta}$ and $H_j = L_j^{-1} \frac{\partial^2 L_j}{\partial \theta^T \partial \theta} - L_j^{-2} S_j S_j^T$. Then, $S = [S_\Gamma^T \ S_{\varphi_\tau}^T \ S_{\varphi_\Delta}^T]^T$ and H consists of sub-matrices H_{ab} for $a, b = \Gamma, \varphi_\tau, \varphi_\Delta$. The NR estimator of Γ is $\hat{\Gamma}^{(m+1)} = \hat{\Gamma}^{(m)} + I_{O\Gamma\Gamma}^{-1} S_\Gamma$ for iteration-m estimates $\hat{\Gamma}^{(m)}$ and $I_{O\Gamma\Gamma} = -H_{\Gamma\Gamma}$.

Let $g(u_j, b_j) = u_j, u_j u_j^T, b_j, b_j b_j^T$. The final step of the EM step is to compute

$$E[g(u_j, b_j) | Y_j, C_{oj}] \ = \ \iint g(u_j, b_j) f(u_j, b_j | Y_j, C_{oj}) du_j db_j \tag{2.12}$$

by AGHQ; See Appendix C for details. We use Bayes theorem to find the nonstandard pdf

$$f(u_j, b_j | Y_j, C_{oj}) \ = \ L_j^{-1} f(Y_j, C_{oj} | u_j, b_j) \phi(u_j | 0, \tau) \phi(b_j | 0, \Delta). \tag{2.13}$$

Likewise, we compute the integral components of S_j and H_j in Appendix A by AGHQ. At convergence, we estimate standard errors by $var(\hat{\theta}) = I_O^{-1}$. See Appendix A for derivation of estimators given random effects and C for approximation of each estimator as an integral with respect to the random effects by AGHQ.

To find initial values, we fit the compliance model (2.8) given $T_{ij} = 1$, impute missing compliance by predictive mean matching [Rubin, 1986], and fit the outcome model (2.7) given the imputation. For the factor loadings, set $\lambda_1 = \frac{\tau_{nc0}}{\tau_{nn}}$ and $\lambda_2 = \frac{\tau_{nc1}}{\tau_{nn}}$ if $r = 1$, and $\lambda = \frac{\tau_{nc1}}{\tau_{nn}}$ if $r = 2$. Algorithm 1 summarizes the steps for generating initial values.

Algorithm 1 An algorithm for initial values

Fit two-level C model with C_o
Retrieve $\hat{\gamma}$ and $\hat{\Delta}$ from the fitted C model as initial values
Retrieve $\hat{\pi_c}$ from fitted C model for imputing C_m
Impute C_m
if Imputation method=Predictive mean matching **then**
    Impute a C_m with the C_o that has the closest $\hat{\pi_c}$ within a cluster
end if
if Imputation method=Predicted probability **then**
    Impute a C_m with $\hat{\pi_c}$
end if
Fit two-level Y model with imputed C
Retrieve $\hat{\alpha}$ and $\hat{\tau}$ from the fitted Y model as initial values
Retrieve $r \times r$ $\hat{\tau}$ from the fitted Y model
if $r' = 1$ **then**
    Initial values for $\lambda_1 = \frac{\hat{\tau_{nc0}}}{\hat{\tau_{nn}}}$ and $\lambda_2 = \frac{\hat{\tau_{nc1}}}{\hat{\tau_{nn}}}$
end if
if $r' = 2$ **then**
    Initial values for $\lambda = \frac{\hat{\tau_{nc1}}}{\hat{\tau_{nn}}}$
end if

CHAPTER 3

E-ASSIST STUDY

This chapter illustrates how the method explained above is applied to analyze the e-assist study data. Colorectal cancer screening (CRCS) is underused compared to other cancer screenings [Siegel et al., 2014]. While 93% of the insured due for CRCS receive a recommendation for screening when visiting a physician office, only 54% complete screening in the following year [Lafata et al., 2014]. To improve the screening rate, a team of investigators developed an online CRCS decision support program called "e-assist: Colon Health" or "e-assist" that reinforces benefits, facilitates informed decision making, and addresses common barriers to and questions regarding CRCS following a physician recommendation for screening [Lafata et al., 2019].

To test the efficacy of e-assist, an MRCT was conducted at primary care clinics located throughout the city of Detroit and the surrounding suburban tricounty area. Eligible patients were randomly assigned to e-assist or control within physicians that are conceived of as sites. As we discussed in section 1.3, a patient assigned to control was not allowed to access e-assist, and hence one assigned to e-assist is a complier if the patient participated in the e-assist intervention or a never taker otherwise. On the other hand, one assigned to control could be a complier or a never taker. We are interested in estimating the joint distribution of physician-specific CACE and mean CRCS rates of compliers and never takers assigned to each arm. We assume that the study is subject to all-or-none compliance [Baker, 2004]. We present data analysis results in the first section, followed by confirmatory simulations for estimator evaluation.

3.1 Data Analysis

We analyze 1825 participants nested within 170 physicians from the e-assist trial where 919 were assigned to e-assist and 906 to control electronically through patient portal accounts. The CRCS outcome is fully observed from the electronic health records and equal to 1 if a patient completed CRCS within a year after a physician recommendation for screening, and 0 otherwise. A patient assigned to e-assist is defined to have received the treatment if the patient submitted the reviewed program or saw at least an e-assist screen after the baseline survey when the intervention and control materials started to differ.

We summarize data for analysis according to the observed treatment assignment (T) and receipt (D), compliance (C) and outcome (Y) in Table 3.1. Out of the 1825 participants who all had patient portal accounts, 65% completed CRCS above the national average 54% from 2007 to 2009 [Lafata et al., 2014]. About 78.34% subjects in the treatment arm accessed e-assist. Published findings imply that portal account holders may differ in measured and unmeasured ways from those without a portal account [Tabriz et al., 2019]. Although 65.88% of compliers and 58.29% of never takers assigned to e-assist completed CRCS, the rates of those assigned to control are unknown due to missing compliance.

Table 3.1: Patients of the e-assist Trial Categorized by Treatment Assignment and Receipt, Compliance, and Outcome ($N = 1825$)

T	D	C	Y	# subjects
0	0	missing	0	322
0	0	missing	1	584
1	0	n	0	83
1	0	n	1	116
1	1	c	0	232
1	1	c	1	488

A popular instrumental variable (IV) approach is to estimate the ITT effect on the outcome (0.01, standard error (SE) = 0.02) and participation (0.78, SE = 0.01) and find the CACE as their ratio (0.02, SE = 0.03) in terms of risk differences [Imbens and Rubin,

2015]. The IV estimation is not an appropriate analysis of our hierarchical data from patients randomized within physicians.

We consider pretreatment patient covariates in Table 3.2. The marital status and CRCS order remained unbalanced across treatment arms unfortunately after randomization.

Table 3.2: Summary of Patient and Physician Characteristics

	e-assist ($n = 919$)	Control ($n = 906$)	P-value
Continuous, mean (sd)			
Patient's age	59.82 (7.29)	59.55 (7.22)	0.427
Physician's age (missing=123)	51.40 (11.09)	51.71 (10.89)	0.561
Charlson Comorbidity Score	1.13 (1.81)	1.08 (1.79)	0.385
Discrete, n (%)			
Female	560 (60.94%)	580 (64.02%)	0.190
Marital status (missing=8)			0.016
Married	626 (68.49%)	569 (63.01%)	
Single	288 (31.51%)	334 (36.99%)	
Colorectal cancer screening order			0.003
Colonoscopy	766 (83.35%)	705 (77.82%)	
Stool test only	153 (16.65%)	201 (22.19%)	

Kruskal-Wallis test for continuous variables, and Pearson's Chi-squared test for discrete variables

To test if marital status and CRCS order have significant effects on the outcome, we estimate the joint distribution in Equations (2.7) and (2.8) as the full model ($r = 3$) where W_{ij} comprises marital status and CRCS order in Equation (2.7), and X_{ij} consists of female indicator, Charlson comorbidity score [Deyo et al., 1992], CRCS order and patient's age, and $Z_{ij} = 1$ in Equation (2.8) based on our preliminary analysis. The reduced model has $\alpha_w = 0$ and others unchanged. We use four abscissas and the square root of summed squared differences in the estimates between two consecutive iterations $< 10^{-4}$ as the convergence criterion. To make sure that eight patients with missing martial status do not influence our estimates, we fitted the reduced model with $N = 1825$ and, then, with $N = 1817$ patients after dropping the eight patients missing marital status to find estimates barely changing and statistical inferences remaining the same between the two analyses (Table 3.3). Therefore,

the impact of the eight missing marital statuses on the estimates is trivial. Next, given $N = 1817$ patients, we find that the full model is a better fit than the reduced one with a likelihood ratio test statistic 17.85 and a p-value < 0.001.

Table 3.3: Parameter Estimation for e-assist Study ($r = 3$, $N = 1817\,\&\,1825$)

	Reduced Model ($N = 1825$)		Reduced Model ($N = 1817$)		Full Model ($N = 1817$)	
Parameter	Estimate (SE)	P-value	Estimate (SE)	P-value	Estimate (SE)	P-value
α_n	0.34 (0.15)	0.015	0.35 (0.16)	0.013	0.17 (0.20)	0.195
α_{c0}	0.66 (0.10)	< 0.001	0.66 (0.10)	< 0.001	0.51 (0.16)	< 0.001
α_{c1}	0.78 (0.10)	< 0.001	0.78 (0.10)	< 0.001	0.60 (0.16)	< 0.001
$\alpha_{c1} - \alpha_{c0}$	0.12 (0.13)	0.182	0.12 (0.13)	0.195	0.09 (0.13)	0.246
$\alpha_{marital\,status}$	-	-	-	-	0.44 (0.11)	< 0.001
$\alpha_{CRCS\,order}$	-	-	-	-	-0.14 (0.13)	0.141
γ_0	1.03 (0.23)	< 0.001	1.02 (0.23)	< 0.001	1.01 (0.23)	< 0.001
γ_{female}	0.24 (0.17)	0.081	0.22 (0.17)	0.10	0.22 (0.17)	0.095
γ_{CCS}	-0.05 (0.04)	0.121	-0.04 (0.04)	0.169	-0.04 (0.04)	0.170
$\gamma_{CRCS\,order}$	0.25 (0.21)	0.120	0.26 (0.21)	0.11	0.27 (0.21)	0.102
γ_{age}	-0.02 (0.01)	0.091	-0.02 (0.01)	0.080	-0.02 (0.01)	0.081
τ_{nn}	0.18		0.22		0.21	
τ_{nc0}	0.12		0.12		0.11	
τ_{nc1}	0.23		0.23		0.23	
τ_{c0c0}	0.08		0.06		0.06	
τ_{c0c1}	0.15		0.12		0.12	
τ_{c1c1}	0.29		0.25		0.24	
Δ	0.05		0.05		0.06	
logL	-1647.74		-1639.25		-1630.27	
AIC	3325.48		3308.50		3294.55	

CCS: Charlson comorbidity score

Based on $N = 1817$, we compare the full model with two reduced models having shared random effects ($r = 2$ and 1) and other elements unchanged as explained in Section 2.2.1. The AIC statistics are 3294.55, 3290.47, and 3288.28 for $r = 3$, 2 and 1, respectively [Akaike, 1974]. Therefore, we do not find evidence that larger models are better fits than the reduced one with $r = 1$ as shown in Tables 3.3-3.5.

Table 3.4: Parameter Estimation for e-assist Study ($r = 1$, $N = 1817\,\&\,1825$)

	Reduced Model ($N = 1825$)		Reduced Model ($N = 1817$)		Full Model ($N = 1817$)	
Parameter	Estimate (SE)	P-value	Estimate (SE)	P-value	Estimate (SE)	P-value
α_n	0.33 (0.16)	0.017	0.35 (0.16)	0.013	0.17 (0.20)	0.200
α_{c0}	0.65 (0.10)	< 0.001	0.66 (0.10)	< 0.001	0.51 (0.16)	< 0.001
α_{c1}	0.78 (0.10)	< 0.001	0.78 (0.10)	< 0.001	0.60 (0.16)	< 0.001
$\alpha_{c1} - \alpha_{c0}$	0.12 (0.14)	0.180	0.12 (0.14)	0.193	0.10 (0.14)	0.242
$\alpha_{marital\,status}$	-	-	-	-	0.44 (0.11)	< 0.001
$\alpha_{CRCS\,order}$	-	-	-	-	-0.14 (0.13)	0.142
γ_0	1.03 (0.23)	< 0.001	1.02 (0.23)	< 0.001	1.01 (0.23)	< 0.001
γ_{female}	0.24 (0.17)	0.083	0.22 (0.17)	0.100	0.22 (0.17)	0.100
γ_{CCS}	-0.05 (0.04)	0.128	-0.04 (0.04)	0.174	-0.04 (0.04)	0.180
$\gamma_{CRCS\,order}$	0.25 (0.21)	0.120	0.26 (0.21)	0.111	0.27 (0.21)	0.101
γ_{age}	-0.02 (0.01)	0.087	-0.02 (0.01)	0.078	-0.02 (0.01)	0.077
λ_1	0.45 (0.49)	0.180	0.40 (0.49)	0.206	0.30 (0.45)	0.249
λ_2	1.14 (0.60)	0.030	1.06 (0.58)	0.037	0.99 (0.57)	0.041
τ_{nn}	0.24		0.24		0.28	
τ_{nc0}	0.11		0.10		0.08	
τ_{nc1}	0.27		0.26		0.27	
τ_{c0c0}	0.05		0.04		0.03	
τ_{c0c1}	0.12		0.10		0.08	
τ_{c1c1}	0.31		0.28		0.27	
Δ	0.05		0.05		0.05	
logL	-1647.66		-1639.19		-1630.14	
AIC	3319.33		3302.38		3288.28	

CCS: Charlson comorbidity score

Table 3.5: Parameter Estimation for e-assist Study ($r = 2$, $N = 1817\,\&\,1825$, never taker and treatment complier sharing a random effect)

	Reduced Model ($N = 1825$)		Reduced Model ($N = 1817$)		Full Model ($N = 1817$)	
Parameter	Estimate (SE)	P-value	Estimate (SE)	P-value	Estimate (SE)	P-value
α_n	0.34 (0.15)	0.015	0.35 (0.16)	0.012	0.17 (0.20)	0.196
α_{c0}	0.66 (0.10)	< 0.001	0.66 (0.10)	< 0.001	0.51 (0.16)	< 0.001
α_{c1}	0.78 (0.10)	< 0.001	0.78 (0.10)	< 0.001	0.60 (0.16)	< 0.001
$\alpha_{c1} - \alpha_{c0}$	0.12 (0.13)	0.180	0.12 (0.14)	0.193	0.10 (0.14)	0.243
$\alpha_{marital\,status}$	-	-	-	-	0.44 (0.11)	< 0.001
$\alpha_{CRCS\,order}$	-	-	-	-	-0.14 (0.13)	0.143
γ_0	1.03 (0.23)	< 0.001	1.02 (0.23)	< 0.001	1.01 (0.23)	< 0.001
γ_{gender}	0.24 (0.17)	0.081	0.22 (0.17)	0.098	0.22 (0.17)	0.095
γ_{CCS}	-0.05 (0.04)	0.122	-0.04 (0.04)	0.169	-0.04 (0.04)	0.169
$\gamma_{CRCS\,order}$	0.25 (0.21)	0.120	0.26 (0.21)	0.111	0.27 (0.21)	0.102
γ_{age}	-0.02 (0.01)	0.092	-0.02 (0.01)	0.081	-0.02 (0.01)	0.081
λ	1.35 (1.00)	0.088	1.18 (0.80)	0.070	1.19 (0.41)	0.002
τ_{nn}	0.17 (0.24)	0.247	0.20 (0.26)	0.223	0.19 (0.09)	0.023
τ_{nc0}	0.11 (0.06)	0.036	0.10 (0.03)	< 0.001	0.10 (0.06)	0.054
τ_{nc1}	0.23 (0.18)	0.108	0.23 (0.17)	0.090	0.22 (0.09)	0.007
τ_{c0c0}	0.07 (-)	-	0.06 (0.05)	0.114	0.05 (-)	-
τ_{c0c1}	0.15 (0.11)	0.100	0.12 (0.06)	0.013	0.12 (-)	-
τ_{c1c1}	0.31 (0.16)	0.027	0.27 (0.15)	0.035	0.27 (0.15)	0.039
Δ	0.05 (0.11)	0.328	0.05 (0.12)	0.320	0.05 (0.12)	0.322
logL	-1647.73		-1639.22		-1630.24	
AIC	3321.46		3304.44		3290.47	

CCS: Charlson comorbidity score

We now interpret the model for $r = 1$ in Table 3.4. Marital status is positively, but CRCS order is not significantly associated with the outcome, ceteris paribus. Controlling for the covariates, the mean log odds of screening for never takers, and control and e-assist compliers are 0.17, 0.51 and 0.60, respectively, all statistically significant to imply a modest mean CACE 0.10 (SE = 0.14, odds ratio (OR) = 1.11) with p-value of 0.242. We identified physician H9028 who had the highest CACE 0.64 (SE = 0.37, OR = 1.90) and physician 129437 who had the lowest CACE -0.30 (SE = 0.27, OR = 0.74).

The estimates of τ were computed after convergence according to Equation (2.10). The positive $cov(u_{c1j} - u_{c0j}, u_{c0j}) = \tau_{c0c1} - \tau_{c0c0} = 0.05$ between physician-specific CACE and log odds of screening for control compliers implies that e-assist is effective among physicians whose complier patients would have completed CRCS well in the absence of the e-assist intervention, thereby increasing inequality in CRCS. However, physician-specific CACE co-varies also positively with the log odds for e-assist compliers, $cov(u_{c1j} - u_{c0j}, u_{c1j}) = 0.19$, leading to the weak impact of e-assist among the subpopulation of patients who are portal account users and, thus, who already produce higher CRCS rates than average overall, as explained before. The same inference is reached based on the model for $r = 3$ that produced $cov(u_{c1j} - u_{c0j}, u_{c0j}) = 0.06$ and $cov(u_{c1j} - u_{c0j}, u_{c1j}) = 0.12$, and respective correlations near 1 in support of the reduced model for r=1.

We compared the CACE 0.10 (SE = 0.14, OR = 1.11) of our final model with the ITT effect, as-treated (AT) effect of treatment received, and per-protocol (PP) effect as the ITT effect of 1619 patients except observed never takers. The ITT effect 0.04 (SE = 0.10, OR = 1.04) is well known to be the product of the CACE and complier proportion and, thus, lower than the CACE. The AT effect 0.20 (SE = 0.11, OR = 1.22) is higher than the CACE as e-assist compliers achieve a higher CRCS rate than do others. The PP effect 0.14 (SE = 0.11, OR = 1.15) is higher than the CACE because observed never takers have a lower CRCS rate than do observed e-assist compliers.

Based on the estimates in Table 3.4, we approximate $P(Y = 1|T = 1, C = 1) - P(Y = 1|T = 0, C = 1) \approx \frac{e^{0.6}}{1+e^{0.6}} - \frac{e^{0.51}}{1+e^{0.51}} = 0.645 - 0.624 = 0.021$, insignificant (SE = 0.03) and comparable to the IV estimate in terms of a risk difference. This is because our compliers seem highly health literate, thereby completing CRCS well above the national average at the time of Lafata et al. [2014] regardless of whether e-assist is present or not. Nevertheless, the IV estimate is not a good approach in our design involving randomization within physicians, and does not provide any information about variability in effects over physicians.

3.2 Confirmatory Simulation Study

We evaluate our estimators by simulating the e-assist sample data and analysis from the previous section closely. We simulate two cases of the final model with $r = 1$ in the previous section: the large cluster sizes with $n_j = 40$ patients or comparatively small cluster sizes of the e-assist sample ranging 1 to 44 patients (11 on average) nested within $J = 170$ physicians.

We simulate the large cluster sizes to validate our estimators and the correct execution of our program, and the e-assist cluster sizes to assess the accuracy and precision of our estimators in the previous section. For each case, we simulate data by the following steps:

1. $T_{ij} \sim Bernoulli(\pi_{Tj})$ by $logit(\pi_{Tj}) \sim N(0.2, 0.2)$ to target $\pi_{Tj} \approx 0.5$;

2. $C_{ij} \sim Bernoulli(\pi_{cij})$ where $logit(\pi_{cij}) = 1 - 0.5X_{1ij} + 0.5X_{2ij} + b_j$ and $b_j \sim N(0, \Delta = 0.3)$ for $X_{1ij} \sim N(1, 1)$ and $X_{2ij} \sim Bin(1, 0.65)$ to obtain $\pi_{cij} \approx 0.7$ on average;

3. $Y_{ij} \sim Bernoulli(\pi_{yij})$ in Equation (2.7) using $logit(\pi_{yij}) = 0.5(1 - C_{ij}) + 0.7C_{ij}(1 - T_{ij}) + 1.2C_{ij}T_{ij} - 0.5X_{1ij} + X_{2ij} + \Lambda u'_j$ to produce $\pi_{yij} \approx 0.69$ on average and $CACE = 0.5$. For $r = 1$, $u'_j \sim N(0, 0.5)$, $\lambda_1 = 0.7$ and $\lambda_2 = 1.1$. For $r = 2$, $u'_j = (u_{nj}, u_{c0j}) \sim N\left(0, \tau' = \begin{pmatrix} 0.5 & 0.3 \\ 0.3 & 0.2 \end{pmatrix}\right)$ and $\lambda = 1.3$.

4. Set $D_{ij} = 1$ if $T_{ij} = C_{ij} = 1$ and 0 otherwise such that $P(D_{ij} = 1) \approx 0.39$ on average with about 50% of compliance missing.

We repeat simulation of data and estimation of $\theta = (\alpha, \gamma, \Lambda, \tau', \Delta)$ 500 times, using four abscissas and convergence criterion as before. Tables 3.6 and 3.7 list the simulated parameters in the third column and resulting estimates in the subsequent columns as follows: average estimates (average standard error or ASE), % bias, the empirical estimate of the true standard error over simulated samples (ESE) and the coverage probabilities (CP).

Table 3.6 compares estimates given the unbalanced e-assist cluster sizes with those given a larger cluster size 40. The e-assist estimates exhibit small biases in general, except those of level-2 variance components λ_1, λ_2 and τ ranging about 3% to 10%. The accuracy improves substantially given the larger cluster size 40 with the largest bias of λ_1 less than 3%. With a large cluster size $n_j = 40$, ASEs appear small and near the ESEs, and the coverage probabilities are near the nominal 0.95 with some under-coverages for the factor loadings and τ'. When r is increased from 1 to 2 in Table 3.7, so the inaccuracy of the estimated factor loading Λ and variance τ_{c0c0} biased upward by 8% and 56%, respectively, given the e-assist cluster size. The %bias, however, improves to 3% and 26% respectively given the cluster size 40. Other estimates are comparable to the counterparts of Table 3.6. In particular, the expected CACE has a small bias of 1% or less for $r = 1$, and 2% or less for $r = 2$.

Figures 3.1 and 3.2 illustrate the results for $r = 1$ and 2, respectively. The horizontal axis labels each estimated parameter while the vertical axis draws the bias and precision of each estimate over the 500 simulations. The interval of each estimate draws one standard error above and below the ML estimate. As the intervals reveal, it improves the precision of all estimates and reduces bias of level-2 variance estimates substantially to increase the cluster sizes n_j from the e-assist average size 11 to 40 given J=170 physicians. According to the simulation results, an average of cluster size of 40 is recommended when $J = 170$, for an accurate and precise estimation.

Table 3.6: Simulated Data Analysis by CACE Model for $r = 1$ (e-assist)

Sample size	Parameter	Simulated	Estimate (ASE[a])	% Bias	ESE[b]	CP[c]
	α_n	0.50	0.51 (0.12)	1.45	0.10	0.95
	α_{c0}	0.70	0.71 (0.10)	0.94	0.10	0.93
	α_{c1}	1.20	1.20 (0.11)	0.27	0.09	0.96
	α_1	-0.50	-0.50 (0.03)	0.34	0.03	0.94
	α_2	1.00	1.00 (0.06)	0.33	0.06	0.95
	γ_0	1.00	0.99 (0.09)	-0.58	0.09	0.93
$n_j = 40$	γ_1	-0.50	-0.50 (0.04)	-0.83	0.04	0.94
	γ_2	0.50	0.50 (0.08)	0.67	0.08	0.94
	λ_1	0.70	0.72 (0.17)	2.74	0.18	0.87
	λ_2	1.10	1.11 (0.16)	1.12	0.17	0.86
	τ'	0.50	0.50 (0.13)	0.50	0.13	0.82
	Δ	0.30	0.30 (0.06)	0.54	0.07	0.93
	CACE	0.50	0.50 (0.10)	-0.68	0.09	0.95
	α_n	0.50	0.51 (0.18)	1.49	0.18	0.94
	α_{c0}	0.70	0.71 (0.18)	1.45	0.18	0.95
	α_{c1}	1.20	1.21 (0.16)	0.41	0.16	0.94
	α_1	-0.50	-0.50 (0.06)	0.87	0.06	0.96
	α_2	1.00	1.00 (0.12)	0.29	0.12	0.95
	γ_0	1.00	0.99 (0.16)	-1.11	0.16	0.95
$n_j = e - assist$	γ_1	-0.50	-0.50 (0.08)	-1.02	0.08	0.95
	γ_2	0.50	0.51 (0.15)	1.93	0.15	0.95
	λ_1	0.70	0.75 (0.38)	7.61	0.48	0.85
	λ_2	1.10	1.13 (0.35)	2.78	0.42	0.86
	τ'	0.50	0.55 (0.30)	10.23	0.27	0.87
	Δ	0.30	0.30 (0.14)	0.07	0.13	0.93
	CACE	0.50	0.50 (0.19)	-1.03	0.19	0.95

[a]average estimated standard error; [b]empirical estimate of the true standard error across simulations; [c]coverage probability

Table 3.7: Simulated Data Analysis by CACE Model for $r = 2$ (e-assist)

Sample size	Parameter	Simulated	Ave. Est (ASE[a])	% Bias	ESE[b]	CP[c]
	α_n	0.50	0.51 (0.11)	1.44	0.10	0.95
	α_{c0}	0.70	0.71 (0.10)	1.74	0.10	0.94
	α_{c1}	1.20	1.20 (0.11)	0.19	0.10	0.95
	α_1	-0.50	-0.50 (0.03)	0.56	0.03	0.95
	α_2	1.00	1.01 (0.06)	0.50	0.06	0.93
$n_j = 40$	γ_0	1.00	0.99 (0.09)	-0.80	0.09	0.93
	γ_1	-0.50	-0.50 (0.04)	-0.97	0.04	0.94
	γ_2	0.50	0.50 (0.08)	0.89	0.08	0.95
	λ	1.30	1.34 (0.14)	3.12	0.19	0.81
	τ'_{nn}	0.50	0.49 (0.09)	-2.36	0.13	0.79
	τ'_{nc0}	0.30	0.30 (0.06)	-1.10	0.06	0.93
	τ'_{c0c0}	0.20	0.25 (0.11)	22.51	0.09	0.94
	Δ	0.30	0.30 (0.06)	1.07	0.06	0.93
	CACE	0.50	0.49 (0.11)	-1.99	0.10	0.96
	α_n	0.50	0.50 (0.18)	-0.75	0.18	0.95
	α_{c0}	0.70	0.71 (0.18)	0.93	0.17	0.96
	α_{c1}	1.20	1.20 (0.17)	-0.14	0.17	0.95
	α_1	-0.50	-0.50 (0.06)	0.48	0.07	0.94
	α_2	1.00	1.01 (0.12)	1.32	0.12	0.95
	γ_0	1.00	0.99 (0.16)	-0.96	0.16	0.95
	γ_1	-0.50	-0.50 (0.08)	0.04	0.08	0.94
$n_j = e - assist$	γ_2	0.50	0.51 (0.15)	2.08	0.15	0.95
	λ	1.30	1.41 (0.38)	8.27	0.47	0.82
	τ'_{nn}	0.50	0.51 (0.23)	1.05	0.25	0.74
	τ'_{nc0}	0.30	0.29 (0.14)	-4.02	0.14	0.84
	τ'_{c0c0}	0.20	0.31 (0.33)	56.52	0.21	0.88
	Δ	0.30	0.30 (0.14)	0.39	0.12	0.95
	CACE	0.50	0.49 (0.21)	-1.65	0.18	0.97

[a]average estimated standard error; [b]empirical estimate of the true standard error across simulations; [c]coverage probability

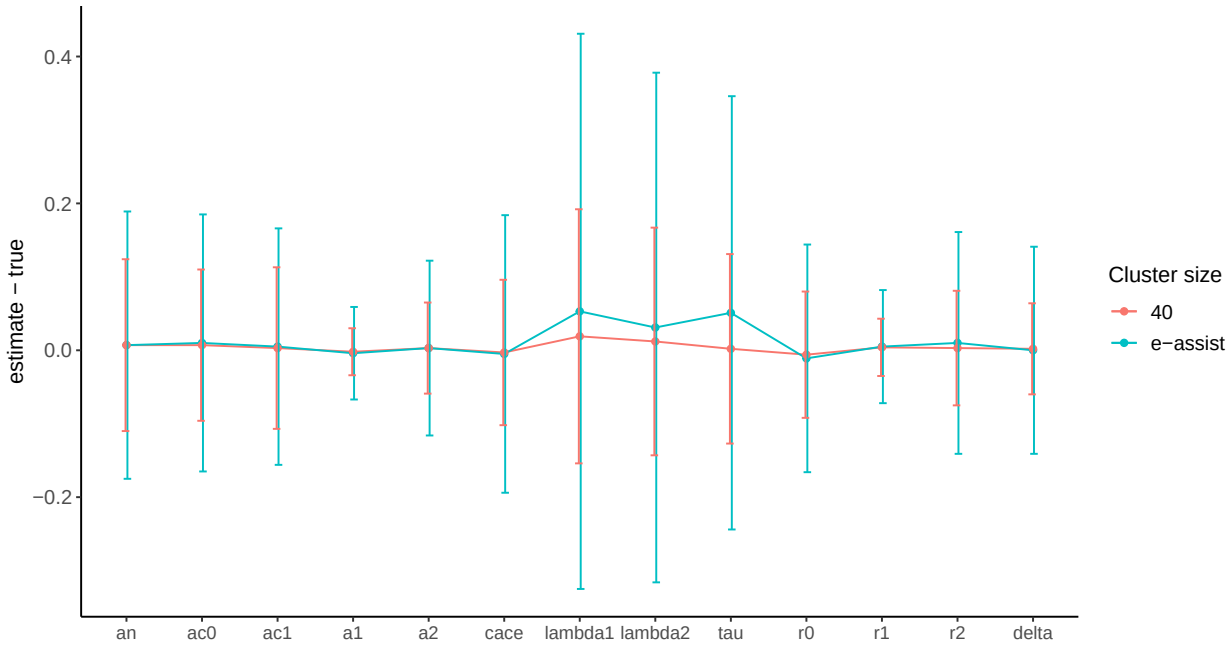

Figure 3.1: Simulation Results when r=1 (e-assist)

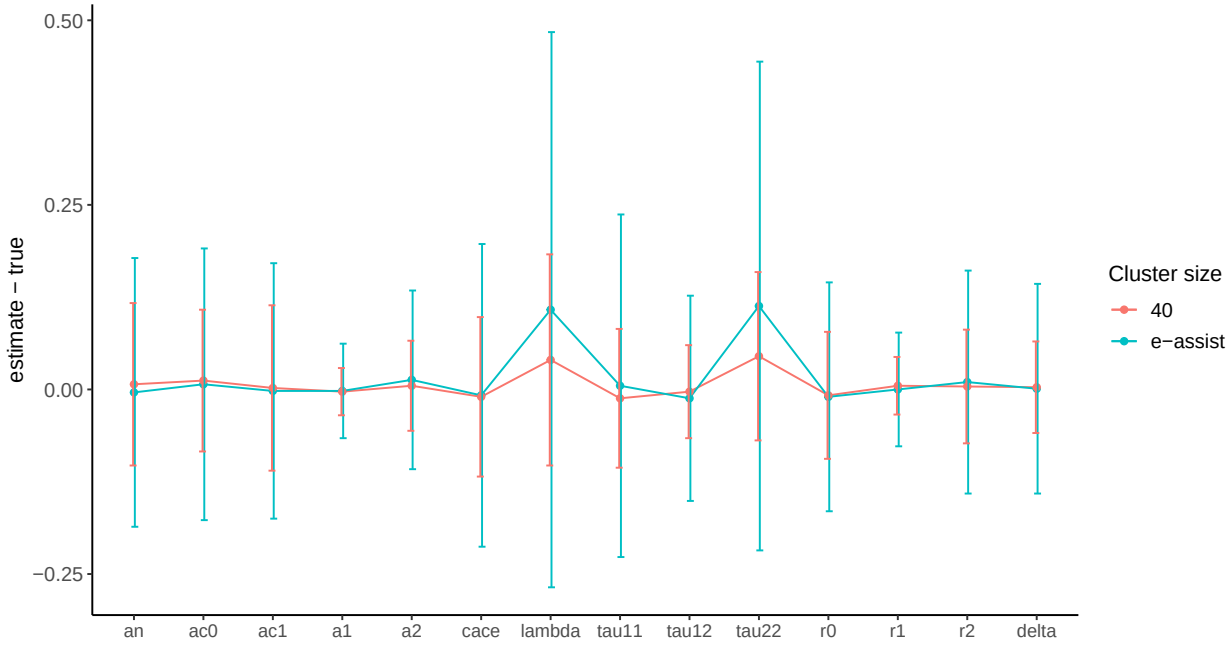

Figure 3.2: Simulation Results when r=2 (e-assist)

Table 3.8 compares the CACE of r=1 and e-assist cluster size in Table 3.6 with the ITT effect. As the table shows, unlike the CACE with a small bias less than 2% and the nominal

0.95 coverage probability, the ITT effect is 44.8% biased downwards with the resulting poor coverage of 0.48.

Table 3.8: ITT and CACE with Simulated Data

Model	Ave. Est (SE)	%bias	MSE	SD	CP
ITT	0.28 (0.11)	-44.80	0.06	0.12	0.48
CACE	0.50 (0.19)	-1.03	0.04	0.19	0.95

3.3 Conclusions

We estimated a distribution of the site-specific causal effects of an intervention on a binary outcome in a multisite trial. Patients were randomly assigned to the e-assist intervention that aimed to improve colorectal cancer screening (CRCS) or control within heterogeneous physicians. Compliance to the treatment assignment was imperfect to classify a patient as either never taker or complier. In this setting, of interest is the joint distribution of physician-specific complier average causal effects (CACEs) and mean responses of never takers, and compliers assigned to each arm. We estimated a bivariate random coefficients model for the outcome and compliance by maximum likelihood (ML) where the random coefficients have the joint distribution. We expressed the likelihood and derived ML estimators given random effects, and integrated the likelihood and each estimator numerically with respect to the random effects by AGHQ.

To test random effects near the boundary of a zero variance or correlations too high that will slow convergence to ML, we estimated reduced models with random effects shared by factor loadings and tested their goodness of fits. Shared random effects decreased the dimension of random effects, thereby reducing the computational expense by AGHQ.

We found no evidence that the average CACE of e-assist is significant. This may be because the trial targeted a population of eligible patients having an online portal account 65% of whom completed colorectal cancer screening (CRCS), higher than the national average

54% at the time of Lafata et al. [2014]. Therefore, the trial participants had comparatively little room for improved screening by the intervention which may have led to the modest treatment effect.

Our simulation reveals that our estimators are reasonably accurate and precise with coverage probabilities near nominal 0.95 overall with the sample size of the trial, and noticeably improve as cluster sizes increase.

Our causal assumptions seem reasonable for analysis of the e-assist trial data. We analyze the observed patient assignments to physicians realistically rather than considering unrealistic random assignments of patients to physicians under the intact physician assumption. The SUTVA assumption seems reasonable as patients do not appear to have interfered with each other, but may have been violated if some patients, communicating with other patients or physicians, were encouraged or discouraged to participate in the e-assist intervention. The monotonicity assumption holds by design. The nonzero average causal effect of treatment assignment T on receipt D seems very reasonable because 78.34% of those assigned to e-assist received the intervention while none assigned to control was allowed to access e-assist (Table 3.1). Since T, D and outcome Y are discrete, we verify the exclusion restriction assumption by a necessary condition on the joint probability distribution of $Y = y$, $T = t$ and $D = d$ by [Pearl, 1995]:

$$\max_{d} \sum_{y} [\max_{t} P(d, y|t)] \leq 1. \tag{3.1}$$

This inequality can be further decomposed into 4 inequalities when Y, T and D are all binary [Pearl, 2011]. Hence, we apply the four inequalities to our data in Table 3.1 and satisfy the

exclusion restriction assumption by

$$
\begin{aligned}
P(Y=0,D=0|T=0)+P(Y=1,D=0|T=1) &= \frac{322}{906}+\frac{116}{919}=0.48\leq 1,\\
P(Y=0,D=1|T=0)+P(Y=1,D=1|T=1) &= \frac{0}{906}+\frac{488}{919}=0.53\leq 1,\\
P(Y=1,D=0|T=0)+P(Y=0,D=0|T=1) &= \frac{584}{906}+\frac{83}{919}=0.73\leq 1,\\
P(Y=1,D=1|T=0)+P(Y=0,D=1|T=1) &= \frac{0}{906}+\frac{232}{919}=0.25\leq 1.
\end{aligned}
$$

CHAPTER 4

NATIONAL STUDY OF LEARNING MINDSETS (NSLM)

This chapter describes how the method explained above is extended to causal analysis of experimental data from the National Study of Learning Mindsets (NSLM) where the binary outcome variable as well as compliance is partially observed.

The compendium report by McFarland et al. [2018] shows that 4.8% of 15- to 24-year-olds left high school in grades 10 through 12 between October 2015 and 2016 without earning a high school diploma or equivalent credential in the United States. The overall dropout rate fluctuated over the past 40 years, and it decreased from 5.9% in 1976 to 3.8% in 2006 and rose to 4.8% in 2016 [McFarland et al., 2018]. The transition to secondary school is a critical period of flexibility in the educational trajectories for adolescents [Crosnoe, 2011]. Sutton et al. [2018] examined the seventh and tenth grades of the students in the National Longitudinal Study of Adolescent Health and discovered that the transition to high school presents risk of poorer academic performance for adolescents overall, and the white and black boys experienced the greatest drops in their GPAs. According to Dalton et al. [2009], students who had lower ninth-grade GPAs, and those who had lower tenth-grade scores on reading and math tests had higher dropout rates than those who had performed better academically.

One solution to the problem of early dropouts is to improve students' academic performances during the transition period to secondary education through social-psychological interventions [Walton and Wilson, 2018]. The purpose of NSLM is to study the effect of an online growth mindset (intellectual abilities) intervention in a nationally representative sample of high schools in the United States [Yeager et al., 2019]. The intervention in the study changes the belief of fixed mindset (fixed intelligence) the students might have and conveys a fact that the intellectual abilities are able to grow in response to devoted effort, and encourages students to try new methods and ask for help when they encounter difficul-

ties [Yeager et al., 2019]. The students who understand growth mindset are more likely to intentionally practice and improve their brains through schoolwork, resulting in consistent academic improvement.

The study was conducted from August 2015 to February 2016. There were 16000 ninth-grade students randomized within 76 public schools who agreed to attend the study. Randomization was conducted at the student level via the computer after students logged into the system in class. Students, teachers, facilitators and researchers were all blind to assignment. Due to undelivered data records, missing pre-treatment GPAs or random assignments, and the restriction of the data-sharing agreement of a school district, we finally included 10341 students from 63 schools. The intervention consisted of two 25-min sessions. The random assignment was made at the beginning of the first session. The first session conveyed the basic idea of a growth mindset that an individual's intellectual abilities can be improved in response to dedicated effort, and the second invited students to intensify the understanding of growth mindset and its application in their lives. The students in the control arm were presented with materials focusing on brain functions, and were not exposed to any material about intervention. The primary outcome variables are average ninth-grade GPA and post-treatment self-reported fixed mindset. The self-reported fixed mindset was measured by participants agreeing or disagreeing with the statement "You have a certain amount of intelligence, and you really can't do much to change it". If a student disagrees with the fixed mindset idea, in other words, agrees with the growth mindset, it is very likely for the student to make more efforts or seek for help when encountering difficulties in school.

The study is an MRCT with noncompliance and nonresponse. We account for the missing outcomes and apply our method to the study with dichotomized average GPA as the outcome variable. Different from the e-assist study, the outcome variable is not fully observed, which adds more complexities to efficient estimation of the treatment effect. Before we move to the analysis, we first describe the data in the following section.

4.1 Data Summary

This section gives descriptive analysis of the NSLM data, which includes the summary of baseline variables, compliance and outcome variables. We dichotomize the post-treatment average GPA into 1 if GPA $>$ median of all sample GPAs, and 0 if GPA $<=$ median of all sample GPAs. The student race is also dichotomized into 1 if a student is nonwhite and 0 otherwise. Parent education is summarized as 1 if a participant's mother has a degree higher than associate's degree and 0 otherwise. Challenge behavior, stated as peer norms in Yeager et al. [2019] paper, was measured by the number of challenging questions minus the number of easy questions on a worksheet task during the NSLM. It was hypothesized that the more supportive the environment for challenge seeking, the more motivated a student grows his(her) mindset [Yeager et al., 2019]. We regroup the variable challenge behavior as 1 if it is positive, and 0 otherwise. The school achievement variable was first constructed by GreatSchools.org ratings and related factors [Lavrakas et al., 2019], and then standardized in the population of more than 12000 US public schools [Yeager et al., 2019].

Compliance is a latent variable determined by the randomized treatment assignment and treatment receipt. The treatment receipt was measured by a student's note to future students who might be struggling in their ninth grade at the end of session 1 in the intervention arm. This exercise has been used in past social-psychological studies and has shown to be effective in helping students internalize the treatment message [Walton and Cohen, 2011, Yeager et al., 2016, 2019]. The notes from most students in the trial express encouragements and suggestions for help seeking when encountering difficulties, whereas some notes are just random words or letters. After reviewing more than 3000 notes, we derived a set of rules to select the notes that indicate the students took up the messages in the first session. The rules are: 1. the length of a note is more than 26 characters; 2. a note should contain any of the following keywords: "smart", "brain", "intelligent", "give up", "keep on going", "work hard", "work harder", "learning", "ask for help", "try your best", "asking question", "asking

Table 4.1: Students of NSLM Categorized by T and D, C, and Y ($N = 10341$)

T	D	C	Y	# students
0	0	missing	0	2588
0	0	missing	1	2384
1	0	n	0	334
1	0	n	1	170
1	1	c	0	2239
1	1	c	1	2211

questions", "mind". These students are considered to have received treatment. By doing so, we discover that about 89.63% students in the treatment arm received the treatment.

We summarize data for analysis according to the observed treatment assignment (T) and receipt (D), compliance (C) and outcome (Y) in Table 4.1. Because those assigned to control are not allowed to access the mindset treatment, we have one-sided noncompliance and, thus, those assigned to control have missing compliance.

Table 4.2 summarizes the outcome and baseline variables. The categorical variables are tabulated by frequencies (percentages) and continuous variables are summarized by median (interquartile range (IQR)). The post-treatment average GPA has 4.01% missing values overall, which do not differ significantly between treatment and control arms by fisher's exact test, with p-value of 0.395. School 73 has the highest missing rate of post-treatment average GPA (89.22%). Pre-treatment average GPA and challenge behavior are unbalanced between the intervention and control arms. Most baseline characteristics have missing values. The baseline characteristics such as English learner, receiving free or reduced lunch at school, special education and gifted class suffer from severe missingness. For example, 23 schools did not report English learning status, and 21 schools did not report if students receive free or reduced lunch at school. For the missing baseline characteristics in analysis, we impute the missing values by sample school means, because of the low measurement error due to the average school-specific sample size (165). If a school does not have a reported value, we impute the missing values of that school as 0.

Table 4.2: Summary of Outcomes and Baseline Characteristics

Variable	Intervention ($n = 5170$)	Control ($n = 5171$)	Total ($n = 10341$)	P-value
Outcomes				
Continuous, median (IQR)				
Post-trt GPA	NA=216 (4.18%) 2.75 (1.83, 3.50)	NA=199 (3.85%) 2.75 (1.80, 3.50)	NA=415 (4.01%) 2.75 (1.82, 3.50)	< 0.001
Discrete, n (%)				
Post-trt GPA	NA=216 (4.18%)	NA=199 (3.85%)	NA=415 (4.01%)	0.926
> sample median	2381 (48.06%)	2573 (47.95%)	4765 (48.00%)	
≤ sample median	2384 (51.94%)	2573 (52.05%)	4765 (52.00%)	
Post-trt fixedmindset	NA=560 (10.83%)	NA=512 (9.90%)	NA=1072 (10.37%)	< 0.001
Not fixed	2591 (56.20%)	1895 (40.67%)	4486 (43.38%)	
Fixed	2019 (43.80%)	2764 (56.20%)	4783 (46.25%)	
Baseline characteristics				
Continuous, median (IQR)				
Pre-trt GPA	2.97 (2.08, 3.58)	3.00 (2.18, 3.59)	3.00 (2.13, 3.58)	< 0.001
School achievement*	0.19 (-0.34, 0.73)	0.19 (-0.34, 0.73)	0.18 (-0.34, 0.73)	0.636
Discrete, n (%)				
Gender	NA=13 (0.25%)	NA=8 (0.15%)	NA=21 (0.20%)	0.938
Female	2545 (49.35%)	2543 (49.25%)	5088 (49.20%)	
Male	2612 (50.65%)	2620 (50.75%)	5232 (50.80%)	
Minority	NA=241 (4.66%)	NA=148 (2.86%)	NA=389 (3.76%)	0.877
Nonwhite	2558 (51.90%)	2598 (51.72%)	5156 (51.81%)	
White	2371 (48.10%)	2425 (48.28%)	4796 (48.19%)	
Pre-trt low GPA				0.226
< school median	2734 (52.88%)	2672 (51.67%)	5406 (52.28%)	
≥ school median	2436 (47.12%)	2499 (48.33%)	4935 (47.72%)	
Parent education	NA=1149 (22.22%)	NA=1057 (20.44%)	NA=2206 (21.33%)	0.806
Higher than Associate	1559 (38.77%)	1607 (39.06%)	3166 (38.92%)	
Equal to or lower than Associate	2462 (61.23%)	2507 (60.94%)	4969 (61.08%)	
Challenge behavior	NA=433 (8.38%)	NA=404 (7.81%)	NA=837 (8.09%)	< 0.001
More challenging	1739 (36.71%)	1346 (28.24%)	3085 (32.46%)	
Less challenging	2998 (63.29%)	3421 (71.76%)	6419 (67.54%)	
English learner	NA=2247 (48.46%)	NA=2254 (43.59%)	NA=4501 (43.53%)	0.486
Yes	192 (6.57%)	206 (7.06%)	398 (6.82%)	
No	2731 (93.43%)	2711 (92.94%)	5442 (93.18%)	
Free or reduced lunch	NA=2071 (40.06%)	NA=2089 (40.40%)	NA=4160 (40.23%)	0.384
Yes	1628 (52.53%)	1584 (51.40%)	3212 (51.97%)	
No	1741 (47.47%)	1498 (48.60%)	2969 (48.03%)	
Pre-trt fixedmindset	NA=9 (0.17%)	NA=8 (0.15%)	NA=17 (0.16%)	0.665
Not fixed	2331 (45.17%)	2309 (44.72%)	4640 (44.94%)	
Fixed	2830 (54.83%)	2854 (55.28%)	2969 (55.06%)	
Special education class	NA=2691 (52.05%)	NA=2699 (52.19%)	NA=5390 (52.12%)	0.217
Yes	256 (14.36%)	387 (15.66%)	743 (15.01%)	
No	2123 (85.64%)	2085 (84.34%)	4208 (84.99%)	
Gifted class	NA=3218 (62.24%)	NA=3208 (62.04%)	NA=6426 (62.14%)	0.555
Yes	336 (17.21%)	353 (17.98%)	689 (17.60%)	
No	1616 (82.79%)	1610 (82.02%)	3226 (82.40%)	

Note: trt stands for treatment; *School achievement is school-level characteristics; Wilcoxon Rank Sum tests were used for continuous variables, and Pearson's Chi-squared tests for discrete variables

The school-specific missing rate of post-treatment average GPA does not vary much from school to school, except for school 73 (Figure 4.1). Additionally, the missing outcome rates do not differ between compliers and noncompliers in the treatment arm by Fisher's exact test with p-value = 0.909. The school-specific missing outcome rates do not depend on school-specific compliance rates or school-specific sample means of baseline characteristics (Figures 4.2,4.3).

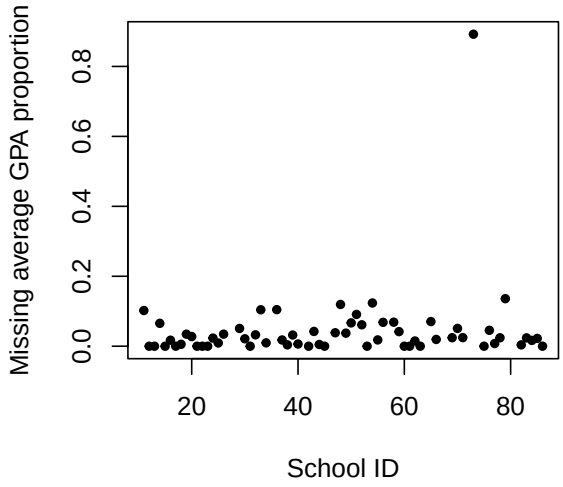

Figure 4.1: School Missing Rates of Post-treatment Average GPA

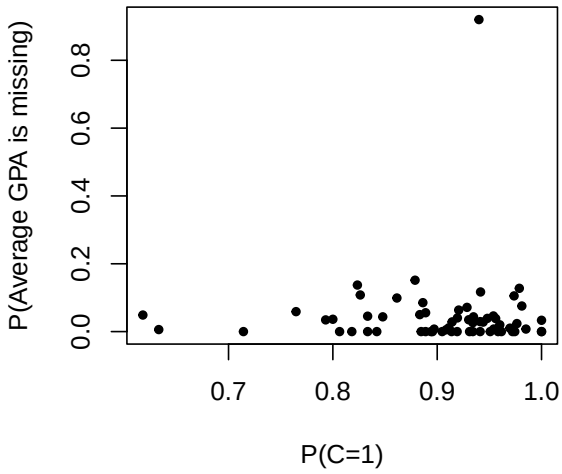

Figure 4.2: School Missing Rates of Post-treatment Average GPA vs. Compliance at School Level

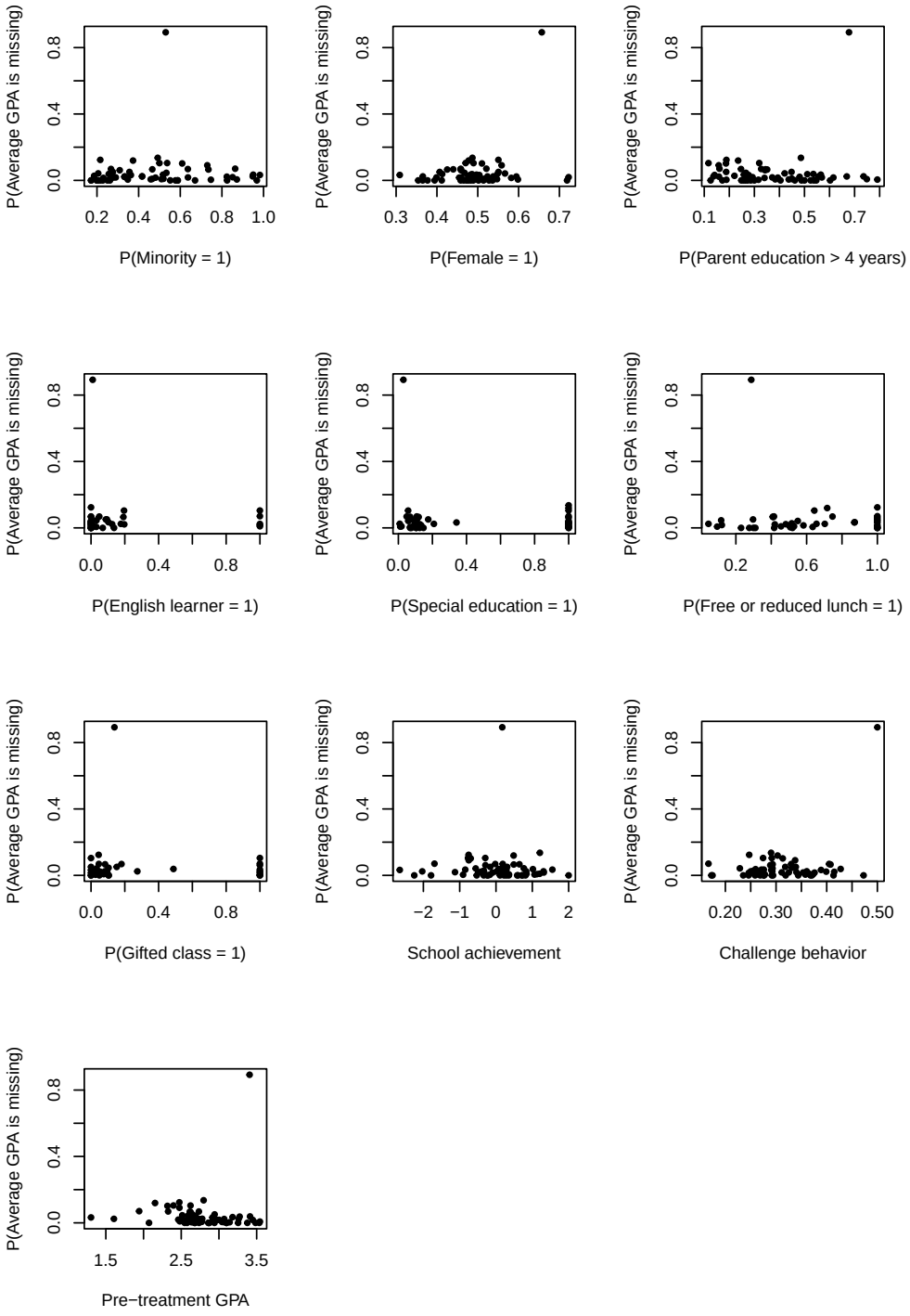

Figure 4.3: School Missing Rates of Post-treatment Average GPA vs. Baseline Characteristics at School Level

4.2 Analysis of the NSLM

Due to the co-occurrence of noncompliance and nonresponse, we have to make further assumptions regarding the missing data mechanism of the outcome. In the current dataset, we assume data missing at random [Rubin, 1976] as stated in section 1.2. We change $f(Y_j, C_{oj}|u_j, b_j)$, a major component of site-specific likelihood, to account for the missing values of the outcome. Let site-specific outcome Y_j be decomposed into observed Y_{oj} and unobserved Y_{mj}. The new $f(Y_{oj}, C_{oj}|u_j, b_j)$ is:

$$f(Y_{oj}, C_{oj}|u_j, b_j) = \prod_{Y_{ij} \in Y_{oj}} f(Y_{ij}|C_{oj}, u_j) f(C_{oj}|b_j) \prod_{Y_{ij} \in Y_{mj}} f(C_{oj}|b_j) \tag{4.1}$$

where

$$\begin{aligned}
\prod_{Y_{ij} \in Y_{oj}} f(Y_{ij}|C_{oj}, u_j) f(C_{oj}|b_j) = & \prod_{\{i:T_{ij}=1, D_{ij}=0\}} f(Y_{ij}|C_n, u_j)\pi_n \prod_{\{i:T_{ij}=1, D_{ij}=1\}} f(Y_{ij}|C_c, u_j)\pi_c \\
& \prod_{\{i:T_{ij}=0, D_{ij}=0\}} f(Y_{ij}|C_c, u_j)\pi_n + f(Y_{ij}|C_n, u_j)\pi_c, \\
\prod_{Y_{ij} \in Y_{mj}} f(C_{oj}|b_j) = & \prod_{\{i:T_{ij}=1, D_{ij}=0\}} \pi_{nij} \prod_{\{i:T_{ij}=1, D_{ij}=1\}} \pi_{cij}.
\end{aligned} \tag{4.2}$$

The updated score and hessian functions are presented in Appendix A.

We control for the unbalanced baseline characteristics, pre-treatment average GPA, and challenge behavior in the outcome Y model. Pre-treatment average GPA is standardized for modeling. School achievement score, challenge behavior, gender, parent education, and free or reduced lunch are selected for compliance C model according to the preliminary analysis of fully observed compliance of those assigned to the growth mindset intervention. We fit the shared random effect model for $r = 1$. We test the significance of average pre-treatment GPA and challenge behavior in the Y model by likelihood ratio test against the reduced model

without these covariates, and find that the full model with average pre-treatment GPA and challenge behavior is a better fit with a likelihood ratio test statistic 5402.37 and a p-value < 0.001. The results are summarized in Table 4.3. The mean CACE in the full model is 0.06 (OR = 1.06), with a p-value of 0.165, indicating an insignificant effect of the intervention on post-treatment GPA among 10341 students. Additionally, the variance of treatment effects among control compliers and treatment compliers are almost the same, higher than that of the never-taker group.

Table 4.3: Parameter Estimation for NSLM Study ($r = 1$, $N = 10341$)

	Reduced Model		Full Model	
Parameter	Estimate (SE)	P-value	Estimate (SE)	P-value
α_n	-0.73 (0.11)	< 0.001	-0.49 (0.14)	< 0.001
α_{c0}	-0.03 (0.03)	0.180	-0.58 (0.06)	< 0.001
α_{c1}	-0.06 (0.03)	0.031	-0.51 (0.06)	< 0.001
$\alpha_{c1} - \alpha_{c0}$	-0.03 (0.05)	0.279	0.06 (0.07)	0.165
$\alpha_{Challenge\,behavior}$	-	-	0.22 (0.07)	< 0.001
$\alpha_{Pre-treatment\,averageGPA}$	-	-	2.80 (0.06)	< 0.001
γ_0	2.31 (0.15)	< 0.001	2.41 (0.12)	< 0.001
$\gamma_{School\,achievement}$	0.34 (0.04)	< 0.001	0.37 (-)	-
$\gamma_{Challenge\,behavior}$	0.49 (0.11)	< 0.001	0.44 (0.11)	< 0.001
γ_{Gender}	0.49 (0.10)	< 0.001	0.37 (0.10)	< 0.001
$\gamma_{Parent\,education}$	-0.20 (0.13)	0.054	-0.34 (0.12)	0.003
$\gamma_{Free\,or\,reduced\,lunch}$	-0.50 (0.16)	< 0.001	-0.41 (0.13)	< 0.001
τ_{nn}	0.41		0.41	
τ_{nc0}	0.40		0.57	
τ_{nc1}	0.43		0.57	
τ_{c0c0}	0.39		0.78	
τ_{c0c1}	0.42		0.79	
τ_{c1c1}	0.45		0.79	
Δ	0.35		0.36	
logL	-8099.99		-5398.81	
AIC	16225.98		10827.61	

Because the intervention may be effective among lower-achieving students, we fit the full model to a smaller set of lower-achieving students whose pre-treatment GPA is less than

the school medians. Among lower-achieving students, about 19.39% students have post-treatment average GPA greater than sample median ($P(Y = 1)$). The estimated CACE is 0.18 (SE = 0.10, OR = 1.20, p-value = 0.033), significant and higher than the ITT effect 0.06 (SE = 0.06, OR = 1.06, p-value = 0.302). Per the results among lower-achieving students, school 82 that has the highest CACE 0.47 (SE = 0.22, OR = 1.60), while school 38 has the lowest CACE -0.11 (SE = 0.26, OR = 0.90). School 82 has 266 students, among whom 25.94% is minority, 50% female, 27.65% students chose more challenging questions, and 2.64 as pre-treatment GPA. In school 38, 74.71% students is minority, 51.53% female, 36.82% students chose more challenging questions, and the average pre-treatment GPA is 3.14.

Table 4.4: Parameter Estimation for Lower-Achieving Students in NSLM Study ($r = 1$, $N = 5406$)

Parameter	Estimate (SE)	P-value
α_n	-0.34 (0.19)	0.032
α_{c0}	-0.86 (0.09)	< 0.001
α_{c1}	-0.68 (0.08)	< 0.001
$\alpha_{c1} - \alpha_{c0}$	0.18 (0.10)	0.033
$\alpha_{Challenge\,behavior}$	0.19 (0.09)	0.021
$\alpha_{Pre-treatment\,averageGPA}$	2.36 (0.10)	< 0.001
γ_0	2.29 (0.15)	< 0.001
$\gamma_{School\,achievement}$	0.44 (-)	-
$\gamma_{Challenge\,behavior}$	0.50 (0.15)	< 0.001
γ_{Gender}	0.15 (0.13)	0.114
$\gamma_{Parent\,education}$	-0.50 (0.16)	0.001
$\gamma_{Free\,or\,reduced\,lunch}$	-0.29 (0.14)	0.019
τ_{nn}	0.27	
τ_{nc0}	0.46	
τ_{nc1}	0.44	
τ_{c0c0}	0.76	
τ_{c0c1}	0.74	
τ_{c1c1}	0.71	
Δ	0.43	
logL	-2862.79	
AIC	5755.58	

We assess the key causal assumptions for the CACE analysis. The SUTVA assump-

tion seems reasonable as students could not have interfered with each other in class. The monotonicity assumption holds by design. The nonzero average causal effect of treatment assignment T on receipt D seems very reasonable because 89.63% of those assigned to intervention received it while none assigned to control was allowed to access intervention. Similar to the e-assist study, we verify the exclusion restriction assumption by a necessary condition on the joint probability distribution of $Y = y$, $T = t$ and $D = d$ by Pearl [1995] and the assumption is satisfied.

$$
\begin{aligned}
P(Y=0,D=0|T=0)+P(Y=1,D=0|T=1) &= \frac{2588}{5171}+\frac{170}{5170}=0.53\leq 1,\\
P(Y=0,D=1|T=0)+P(Y=1,D=1|T=1) &= \frac{0}{5171}+\frac{2211}{5170}=0.43\leq 1,\\
P(Y=1,D=0|T=0)+P(Y=0,D=0|T=1) &= \frac{2384}{5171}+\frac{334}{5170}=0.53\leq 1,\\
P(Y=1,D=1|T=0)+P(Y=0,D=1|T=1) &= \frac{0}{5171}+\frac{2239}{5170}=0.43\leq 1.
\end{aligned}
$$

4.3 Confirmatory Simulations (NSLM)

Similar to the e-assist study, we evaluate our estimators by simulating the NSLM sample data and analysis. However, we simulate three cases of the model with two covariates for $r = 1$ as analyzed in the previous section: 5%, 10%, and 20% missing rates of the outcome, with 165 students nested within $J = 63$ schools. We also notice that the compliance rate of NLSM is higher than that of the e-assist study. Therefore, we simulate two compliance probabilities, $P(C = 1) = 0.70$ and $P(C = 1) = 0.85$, per outcome missing rate. For each of the six cases, we simulate data by the following steps:

1. $T_{ij} \sim Bernoulli(\pi_{Tj})$ by $logit(\pi_{Tj}) \sim N(0.2, 0.2)$ to target $\pi_{Tj} \approx 0.5$;

2. $C_{ij} \sim Bernoulli(\pi_{cij})$ where $logit(\pi_{cij}) = 1 - X_{1ij} + 0.1X_{2ij} + b_j$ or $logit(\pi_{cij}) = 1 + 0.5X_{1ij} + 2X_{2ij} + b_j$, and $b_j \sim N(0, \Delta = 0.3)$ for $X_{1ij} \sim N(0, 1)$ and $X_{2ij} \sim$

$Bin(1, 0.65)$ to obtain $\pi_{cij} \approx 0.7$ or 0.85 on average;

3. $Y_{ij} \sim Bernoulli(\pi_{yij})$ in Equation (2.7) using $logit(\pi_{yij}) = -(1 - C_{ij}) - 0.5C_{ij}(1 - T_{ij}) - 0.2C_{ij}T_{ij} - 0.5X_{1ij} + X_{2ij} + \Lambda u'_j$ to produce $\pi_{yij} \approx 0.51$ on average and $CACE = 0.3$. For $r = 1$, $u'_j \sim N(0, 0.5)$, $\lambda_1 = 0.9$ and $\lambda_2 = 1.1$. The missing outcome rate is the same between the treatment and control arms.

4. Set $D_{ij} = 1$ if $T_{ij} = C_{ij} = 1$ and 0 otherwise such that $P(D_{ij} = 1) \approx 0.39$ on average with about 50% of compliance missing.

We repeat the simulation and estimation of $\theta = (\alpha, \gamma, \Lambda, \tau', \Delta)$ and collect 100 converged results for each case, using four abscissas and the square root of summed squared differences in the estimates between two consecutive iterations $< 10^{-4}$ as the convergence criterion as before. Tables 4.5 and 4.6 summarize the results for $P(C = 1) = 0.85$ and 0.7, respectively.

Table 4.5: Simulated Data Analysis by CACE Model for $r = 1$, $P(C = 1) = 0.85$, $Q = 4$ (NSLM)

Missing Rate in Y	Parameter	Simulated	Estimate (ASE[a])	% Bias	ESE[b]	CP[c]
	α_n	-1.00	-1.00 (0.08)	-0.21	0.12	0.72
	α_{c0}	-0.50	-0.48 (0.07)	-3.51	0.10	0.74
	α_{c1}	-0.2	-0.19 (0.06)	-6.07	0.11	0.72
	α_1	-0.50	-0.50 (0.02)	0.65	0.02	0.97
	α_2	1.00	0.99 (0.05)	-0.66	0.05	0.97
	γ_0	1.00	1.01 (0.11)	0.56	0.08	0.95
$P(Y\ missing) = 0.05$	γ_1	0.50	0.50 (0.04)	-1.03	0.04	0.94
	γ_2	2.00	1.98 (0.09)	-0.83	0.08	0.96
	λ_1	0.90	0.92 (0.08)	2.03	0.14	0.67
	λ_2	1.10	1.11 (0.09)	0.99	0.14	0.67
	τ'	0.50	0.45 (0.08)	-10.48	0.13	0.71
	Δ	0.30	0.30 (0.08)	0.40	0.08	0.93
	CACE	0.30	0.30 (0.05)	-1.80	0.05	0.97
	α_n	-1.00	-1.00 (0.09)	0.30	0.12	0.78
	α_{c0}	-0.50	-0.49 (0.06)	-2.80	0.10	0.74
	α_{c1}	-0.20	-0.19 (0.07)	-3.87	0.12	0.65
	α_1	-0.50	-0.50 (0.03)	0.25	0.02	0.99
	α_2	1.00	0.99 (0.05)	-1.36	0.05	0.94
	γ_0	1.00	1.00 (0.10)	-0.45	0.08	0.95
$P(Y\ missing) = 0.10$	γ_1	0.50	0.50 (0.04)	-0.10	0.05	0.93
	γ_2	2.00	2.00 (0.09)	-0.15	0.10	0.94
	λ_1	0.90	0.91 (0.08)	1.59	0.16	0.55
	λ_2	1.10	1.13 (0.08)	2.31	0.16	0.61
	τ'	0.50	0.46 (0.09)	-8.11	0.14	0.67
	Δ	0.30	0.30 (0.08)	-1.76	0.07	0.93
	CACE	0.30	0.29 (0.06)	-2.08	0.05	0.96
	α_n	-1.00	-1.01 (0.10)	1.32	0.12	0.82
	α_{c0}	-0.50	-0.49 (0.07)	-1.72	0.11	0.75
	α_{c1}	-0.20	-0.20 (0.07)	1.08	0.11	0.73
	α_1	-0.50	-0.50 (0.03)	0.74	0.02	0.98
	α_2	1.00	0.99 (0.06)	-1.44	0.06	0.94
	γ_0	1.00	1.00 (0.10)	-0.18	0.09	0.94
$P(Y\ missing) = 0.20$	γ_1	0.50	0.50 (0.04)	0.46	0.05	0.92
	γ_2	2.00	1.99 (0.09)	-0.63	0.10	0.91
	λ_1	0.90	0.91 (0.08)	1.01	0.17	0.66
	λ_2	1.10	1.11 (0.08)	0.98	0.18	0.61
	τ'	0.50	0.48 (0.09)	-3.21	0.15	0.70
	Δ	0.30	0.31 (0.08)	4.20	0.08	0.94
	CACE	0.30	0.29 (0.06)	-3.58	0.06	0.93

[a]average estimated standard error; [b]empirical estimate of the true standard error across simulations; [c]coverage probability

Table 4.6: Simulated Data Analysis by CACE Model for $r = 1$, $P(C = 1) = 0.70$, $Q = 4$ (NSLM)

Missing Rate in Y	Parameter	Simulated	Estimate (ASE[a])	% Bias	ESE[b]	CP[c]
$P(Y\ missing) = 0.05$	α_n	-1.00	-1.00 (0.08)	0.34	0.12	0.72
	α_{c0}	-0.50	-0.50 (0.08)	0.11	0.10	0.78
	α_{c1}	-0.20	-0.21 (0.08)	3.12	0.11	0.68
	α_1	-0.50	-0.50 (0.03)	0.55	0.02	0.97
	α_2	1.00	1.00 (0.05)	-0.37	0.05	0.93
	γ_0	1.00	0.99 (0.10)	-0.55	0.09	0.95
	γ_1	-1.00	-1.00 (0.04)	0.16	0.03	0.97
	γ_2	0.10	0.09 (0.07)	-13.64	0.06	0.99
	λ_1	0.90	0.90 (0.11)	0.45	0.13	0.78
	λ_2	1.10	1.10 (0.10)	0.33	0.12	0.83
	τ'	0.50	0.48 (0.10)	-4.82	0.11	0.83
	Δ	0.30	0.31 (0.06)	1.92	0.07	0.92
	CACE	0.30	0.29 (0.07)	-1.91	0.06	0.97
$P(Y\ missing) = 0.10$	α_n	-1.00	-1.00 (0.09)	0.22	0.11	0.78
	α_{c0}	-0.50	-0.50 (0.08)	-0.62	0.10	0.81
	α_{c1}	-0.20	-0.20 (0.08)	1.42	0.11	0.70
	α_1	-0.50	-0.50 (0.03)	0.73	0.03	0.95
	α_2	1.00	1.00 (0.05)	0.25	0.05	0.94
	γ_0	1.00	1.00 (0.09)	-0.17	0.08	0.92
	γ_1	-1.00	-1.00 (0.04)	-0.03	0.04	0.96
	γ_2	0.10	0.09 (0.07)	-7.02	0.07	0.97
	λ_1	0.90	0.90 (0.10)	0.11	0.13	0.81
	λ_2	1.10	1.11 (0.09)	0.71	0.12	0.77
	τ'	0.50	0.49 (0.09)	-2.80	0.13	0.76
	Δ	0.30	0.30 (0.06)	0.94	0.07	0.89
	CACE	0.30	0.29 (0.07)	-1.98	0.07	0.96
$P(Y\ missing) = 0.20$	α_n	-1.00	-1.01 (0.09)	0.92	0.12	0.77
	α_{c0}	-0.50	-0.49 (0.08)	-1.47	0.11	0.78
	α_{c1}	-0.20	-0.20 (0.08)	2.23	0.12	0.66
	α_1	-0.50	-0.50 (0.03)	-0.75	0.03	0.98
	α_2	1.00	0.98 (0.05)	-0.31	0.05	0.96
	γ_0	1.00	1.00 (0.11)	-0.37	0.08	0.98
	γ_1	-1.00	-1.00 (0.04)	-0.09	0.03	0.97
	γ_2	0.10	0.10 (0.07)	-3.48	0.06	0.98
	λ_1	0.90	0.91 (0.11)	0.64	0.14	0.76
	λ_2	1.10	1.11 (0.09)	0.94	0.12	0.74
	τ'	0.50	0.49 (0.10)	-1.40	0.13	0.78
	Δ	0.30	0.30 (0.06)	-1.09	0.07	0.88
	CACE	0.30	0.29 (0.07)	-3.93	0.07	0.97

[a]average estimated standard error; [b]empirical estimate of the true standard error across simulations; [c]coverage probability

Table 4.5 compares estimates among three outcome missing rates when compliance rate is 0.85 using four abscissas. In the real data, the observed compliance rate is 0.89 and outcome missing rate is about 0.04. Oddly, the scenario with outcome missing rate 0.05 has the highest bias in τ. In particular, the biases in τ estimates increase as the missing rates in outcome increase. ASEs underestimate ESEs for the estimators of Y model for all missing outcome rates, resulting in under-coverages. Additionally, the expected CACE has for each outcome missing rate has a low bias of 4% or lower and a coverage close to 0.95.

Table 4.6 compares estimates among three outcome missing rates when compliance rate is 0.70 using four abscissas. There are slight improvements in terms of bias for $P(C = 1) = 0.7$ compared to $P(C = 1) = 0.85$. Decreased $P(C = 1)$ seems to have mild improvement on the estimates. We plan to formulate more simulations with different compliance rates to explore the impact of compliance on overall estimations. For both $P(C = 1) = 0.85$ and $P(C = 1) = 0.70$, estimates with highest 20% missing outcome rate appear to be strangely somewhat better that others in terms of bias and coverage. In order to evaluate if the odd behaviors are due to the inaccurate approximation of AGHQ with 4 abscissas, we did another set of simulations when $P(C = 1) = 0.85$ with three missing outcome rates as above and eight abscissas. We kept the same simulation setting except for $n_j = 160$. The updated simulation results are summarized in Table 4.7. When $Q = 8$, the simulation results improved in terms of bias and coverage overall. On the contrary to the cases when $Q = 4$, the bias increases as outcome missing rates increase. Therefore, we can confirm that the odd behaviors were due to the inaccurate approximation of AGHQ with less abscissas. We also conclude that for n_j over 100, the estimation requires around eight abscissas for a better AGHQ approximation.

Figures 4.4-4.6 draw estimates in the same way as did Figure 3.1. The accuracy of each estimate seems to improve in the following two cases: i) when the complier probability changes from 0.85 in Figure 4.4 to 0.7 in Figure 4.5, and ii) when the number of abscissas for the numerical integral by AGHQ increases from 4 in Figure 4.4 to 8 in Figure 4.6.

Table 4.7: Simulated Data Analysis by CACE Model for $r = 1$, $P(C = 1) = 0.85$, $Q = 8$ (NSLM)

Missing Rate in Y	Parameter	Simulated	Estimate (ASE[a])	% Bias	ESE[b]	CP[c]
	α_n	-1.00	-1.01 (0.11)	1.31	0.13	0.76
	α_{c0}	-0.50	-0.50 (0.10)	0.03	0.10	0.77
	α_{c1}	-0.2	-0.20 (0.11)	-1.22	0.12	0.75
	α_1	-0.50	-0.50 (0.03)	0.40	0.03	0.96
	α_2	1.00	1.00 (0.05)	0.43	0.05	0.96
	γ_0	1.00	1.00 (0.09)	0.40	0.09	0.96
$P(Y\ missing) = 0.05$	γ_1	0.50	0.50 (0.04)	0.70	0.04	0.97
	γ_2	2.00	1.99 (0.09)	-0.62	0.10	0.94
	λ_1	0.90	0.94 (0.11)	3.84	0.17	0.75
	λ_2	1.10	1.12 (0.11)	1.56	0.18	0.67
	τ'	0.50	0.51 (0.12)	2.27	0.18	0.78
	Δ	0.30	0.28 (0.08)	-5.46	0.07	0.92
	CACE	0.30	0.30 (0.06)	0.86	0.06	0.94
	α_n	-1.00	-1.01 (0.10)	0.85	0.12	0.84
	α_{c0}	-0.50	-0.49 (0.08)	-2.80	0.09	0.79
	α_{c1}	-0.20	-0.18 (0.08)	-8.96	0.11	0.84
	α_1	-0.50	-0.50 (0.03)	0.60	0.03	0.93
	α_2	1.00	0.99 (0.05)	-0.66	0.06	0.96
	γ_0	1.00	1.00 (0.09)	-0.54	0.09	0.97
$P(Y\ missing) = 0.10$	γ_1	0.50	0.50 (0.04)	0.88	0.05	0.96
	γ_2	2.00	2.00 (0.09)	-0.24	0.10	0.93
	λ_1	0.90	0.87 (0.10)	-3.15	0.28	0.70
	λ_2	1.10	1.06 (0.09)	-3.41	0.30	0.66
	τ'	0.50	0.51 (0.11)	2.67	0.19	0.78
	Δ	0.30	0.29 (0.08)	-4.44	0.07	0.91
	CACE	0.30	0.30 (0.06)	1.31	0.06	0.98
	α_n	-1.00	-1.00 (0.11)	-0.04	0.13	0.85
	α_{c0}	-0.50	-0.49 (0.08)	-2.01	0.11	0.73
	α_{c1}	-0.20	-0.18 (0.09)	-10.60	0.13	0.79
	α_1	-0.50	-0.50 (0.03)	0.75	0.03	0.95
	α_2	1.00	1.00 (0.06)	-0.45	0.06	0.98
	γ_0	1.00	1.01 (0.09)	0.99	0.10	0.94
$P(Y\ missing) = 0.20$	γ_1	0.50	0.50 (0.04)	0.42	0.04	0.96
	γ_2	2.00	1.99 (0.09)	-0.38	0.10	0.94
	λ_1	0.90	0.90 (0.11)	-1.26	0.22	0.67
	λ_2	1.10	1.08 (0.11)	-1.91	0.23	0.66
	τ'	0.50	0.55 (0.12)	9.49	0.20	0.75
	Δ	0.30	0.29 (0.08)	-4.16	0.08	0.89
	CACE	0.30	0.31 (0.06)	3.71	0.06	0.98

[a]average estimated standard error; [b]empirical estimate of the true standard error across simulations; [c]coverage probability

Consequently, bias seems to decrease as the complier probability tends to 0.5 or as the accuracy increases in numerical integration by AGHQ. However, we do not find the impact of the outcome missing rates 0.05, 0.1 and 0.2 on the estimates. The missing rates may not have been nontrivial to influence the estimates given comparatively a large sample size.

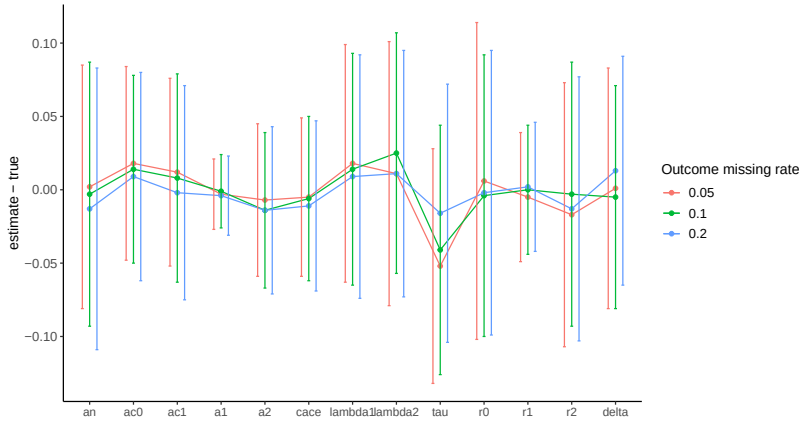

Figure 4.4: Simulation Results When r=1, P(C=1)=0.85, Q=4 (NSLM)

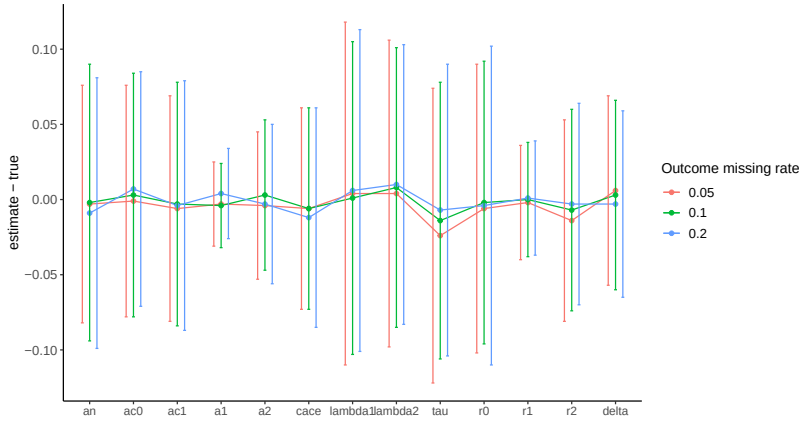

Figure 4.5: Simulation Results When r=1, P(C=1)=0.70, Q=4 (NSLM)

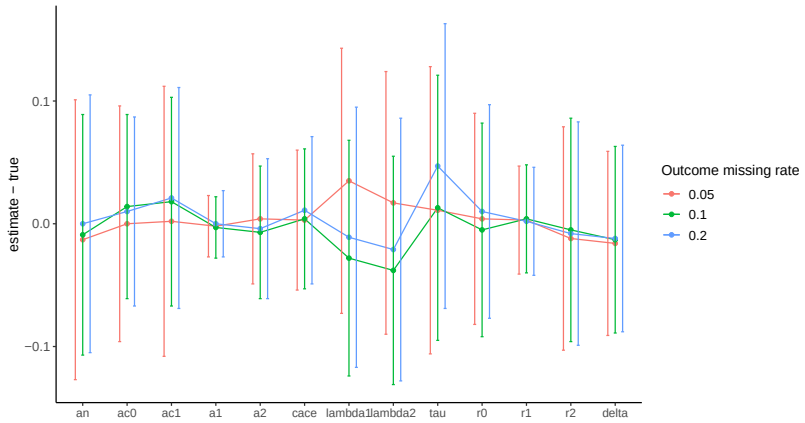

Figure 4.6: Simulation Results When r=1, P(C=1)=0.85, Q=8 (NSLM)

For all three cases when $P(C = 1) = 0.85$, CACE has a small bias up to 3.71% with

Table 4.8: Simulated Data Analysis by ITT Model for $r = 1$, $P(C = 1) = 0.85$ (NSLM)

Missing Rate in Y	Parameter	Simulated	Estimate (ASE[a])	% Bias	ESE[b]	CP[c]
$P(Y\ missing) = 0.05$	α_1	-0.50	-0.47 (0.02)	-6.80	0.04	0.72
	α_2	1.00	1.14 (0.05)	13.80	0.05	0.19
	γ_0	1.00	1.02 (0.09)	1.60	0.07	0.99
	γ_1	0.50	0.50 (0.04)	0.00	0.05	0.92
	γ_2	2.00	1.98 (0.09)	-0.95	0.09	0.93
	τ	0.50	0.50 (0.11)	-1.00	0.10	0.94
	Δ	0.30	0.30 (0.10)	0.67	0.08	1.00
	ITT	0.30	0.23 (0.05)	-22.67	0.05	0.61
$P(Y\ missing) = 0.10$	α_1	-0.50	-0.47 (0.02)	-6.80	0.03	0.73
	α_2	1.00	1.14 (0.05)	13.90	0.05	0.19
	γ_0	1.00	1.02 (0.09)	1.60	0.07	0.99
	γ_1	0.50	0.50 (0.04)	0.00	0.05	0.92
	γ_2	2.00	1.98 (0.09)	-0.95	0.09	0.93
	τ	0.50	0.50 (0.11)	-0.80	0.11	0.92
	Δ	0.30	0.30 (0.08)	-1.76	0.07	0.93
	ITT	0.30	0.23 (0.05)	-22.33	0.05	0.64
$P(Y\ missing) = 0.20$	α_1	-0.50	-0.47 (0.03)	-6.60	0.03	0.74
	α_2	1.00	1.14 (0.05)	13.60	0.05	0.28
	γ_0	1.00	1.02 (0.09)	1.60	0.07	0.99
	γ_1	0.50	0.50 (0.04)	0.00	0.05	0.92
	γ_2	2.00	1.98 (0.09)	-0.95	0.09	0.93
	τ	0.50	0.50 (0.12)	-1.00	0.11	0.93
	Δ	0.30	0.30 (0.10)	0.67	0.08	1.00
	ITT	0.30	0.23 (0.05)	-22	0.05	0.71

[a]average estimated standard error; [b]empirical estimate of the true standard error across simulations; [c]coverage probability

eight abscissas, while the ITT effect has bias up to -22.67%, seven times higher than that of the CACE (Table 4.8). Standard errors of the estimates of α_n, α_{c0}, α_{c1}, λ_1, λ_2 and τ' in the CACE model are smaller than the empirical estimates over samples, leading to under-coverages and inflated type one error rate (Table 4.7). In the presence of nonresponse, the estimates of the ITT models in Table 4.8 fares worse in terms of accuracy, precision and coverage.

4.4 Conclusions

We analyzed the causal effect of the growth mindset intervention and variability in the causal effect over schools on a binary GPA outcome. Students were randomly assigned to intervention or control within heterogeneous schools. In the presence of missing values of the outcome, we estimated the bivariate random coefficients model for the outcome and compliance by ML where the random coefficients have the joint distribution of interest.

We found no evidence that the average CACE of the intervention on post-treatment average GPA of all students was significant. However, the average CACE was significant among lower-achieving students. Additionally, the post-treatment GPA was positively related to pre-treatment GPA and challenge behavior (peer students tend to choose more challenging questions), indicating that among students with higher GPAs or those who preferred more challenges, there was comparatively less room for improving GPA.

The simulation study showed that the CACE estimator was reasonably accurate and precise with coverage probabilities near nominal 0.95 overall with different missing outcome rates. However, the estimation of parameters in Y model declined, compared to the scenario when outcome was fully observable. Changing the compliance rate influence the estimation of either the Y or C model mildly, indicating that probability of being compliers did not provide much information about the outcome, which can also confirmed by Figure 4.2. We need more simulations to form explore the relationship between compliance rate and estimation in presence of nonresponse. Further, as missing rates in Y increased, the bias in estimation increased as well, with eight abscissas for AGHQ. We also discovered that for a large cluster sample size ($>$100), eight abscissas were required for a better approximation of AGHQ.

Currently, we did the preliminary analysis ($r = 1$) for NSLM with our approach due to time limit. We plan to increase the dimension of random effects in the Y model and assess the correlation between CACE and the mean outcome rates of the three compliance groups.

Additionally, in the current dataset, the covariates also have missing values. We imputed

the missing values by school sample means and zeros due to time constraint. We could have used predictive mean matching, which is what we plan to do for the analysis.

Our method assumes outcome MAR. An important future research is to extend this approach to outcomes missing not at random.

CHAPTER 5

DISCUSSIONS

We developed a method to estimate the mean and site-specific CACEs from data in a two-level multisite randomized trial, by jointly modeling outcome and compliance, via a bivariate random coefficients model by ML where site-specific mean outcomes of never takers, control compliers, and treatment compliers are expressed as coefficients randomly varying over sites. Our key contribution is to a multisite trial where each site generates it own CACE and has correlations with the mean responses of the compliance groups. Hence, we may examine and learn from differences in sites with low and high CACEs.

We factorized the likelihood into the conditional distribution of the outcome given compliance and the marginal distribution of compliance, expressing it as an integral with respect to random effects. We also derived the ML estimators by a combination of the EM algorithm and Newton Raphson method as integrals with respect to random effects, and numerically approximated the likelihood and each estimator by AGHQ. We estimated the variance components stably by the EM algorithm. We fitted the shared random effects model to reduce the dimension of the random effects and, thus, to improve computation speed and convergence.

We applied our method to the e-assist study, a multisite randomized trial to assess the causal effect of the e-assist intervention on the CRCS outcome. The patients in the control arm could not access e-assist. The study has fully observed outcome and partially observed compliance. We estimated the mean CACE and the distribution of the site-specific CACE. We discovered no evidence that the CACE is significant. The insignificant CACE may have been due to samples that represent more health-literate patients who use portal accounts and enjoy comparatively higher CRCS rates. For the colon health study, we also noticed that the older the physician, the older their patients. The older patients might follow doctors' instructions and recommendation to be screened better. A future research for e-assist study is to control the site average patient age, and assess the CACE moderated by the site average

patient age.

We adjusted our model and applied our method to another multisite randomized study, National Study of Learning Mindsets (NSLM), a study to assess if the online growth mindset intervention can improve the GPAs of ninth-graders in American public schools. The study has 4.01% missing in the GPA outcome and compliance is also imperfect and partially observed. We found evidence of a significant CACE of the intervention among lower-achieving students, but the treatment effect is insignificant among all students including higher-achievers. We also discovered that pre-treatment GPAs and students' preference for challenges are significantly and positively associated to the post-treatment GPA outcome, which might be the reason why the intervention was not effective among higher-achieving students since there was no room for improvement in their GPAs. Therefore, a possible future research for NSLM is to apply the prognostic score [Hansen, 2008] which summarizes the associations of covariates with the potential outcomes and stratifies students according to the prognostic scores into different groups who might benefit from the treatment from the least to the most, and estimate the CACE within each level of prognostic scores.

One limitation of the method is the computationally intensive numerical integration by AGHQ. Taking the NSLM as an example, it took 14 hours for a model with a shared random effect using four abscissas to converge by parallel computing on a machine with 15 computation cores. The confirmatory simulation results also show that 8 abscissas are required to improve the convergence and estimation by AGHQ. One solution to the high computational cost by AGHQ is to adopt the multivariate Laplace approximation for numerical integration for estimation of a generalized linear model [Solomon and Cox, 1992, Liu and Pierce, 1993, Lin and Breslow, 1996]. According to Raudenbush et al. [2000], numerical integration by the multivariate Laplace approximation with the sixth-order differentiation resulted in a more efficient and better estimation than AGHQ with seven quadrature points. Another limitation is that we assume MAR for both missing outcome and compliance. Because the missing

outcome mechanism may violate the MAR assumption, an important future research topic is to model the missing data mechanism in the likelihood function.

REFERENCES

Clifford Adelman. The toolbox revisited: Paths to degree completion from high school through college. *US Department of Education*, 2006.

Hirotugu Akaike. A new look at the statistical model identification. *IEEE transactions on automatic control*, 19(6):716–723, 1974.

Joshua Angrist and Guido Imbens. Identification and estimation of local average treatment effects, 1995.

Joshua D Angrist, Guido W Imbens, and Donald B Rubin. Identification of causal effects using instrumental variables. *Journal of the American statistical Association*, 91(434): 444–455, 1996.

Stuart G Baker. Compliance, all-or-none. *Encyclopedia of Statistical Sciences*, 2, 2004.

Melanie Bannister-Tyrrell, Branko Miladinovic, Christine L Roberts, and Jane B Ford. Adjustment for compliance behavior in trials of epidural analgesia in labor using instrumental variable meta-analysis. *Journal of Clinical Epidemiology*, 68(5):525–533, 2015.

Howard S Bloom. Accounting for no-shows in experimental evaluation designs. *Evaluation review*, 8(2):225–246, 1984.

Martha L Bruce, Thomas R Ten Have, Charles F Reynolds III, Ira I Katz, Herbert C Schulberg, Benoit H Mulsant, Gregory K Brown, Gail J McAvay, Jane L Pearson, and George S Alexopoulos. Reducing suicidal ideation and depressive symptoms in depressed older primary care patients: a randomized controlled trial. *Jama*, 291(9):1081–1091, 2004.

George Casella and Roger L Berger. *Statistical inference*. Cengage Learning, 2021.

Jing Cheng. Estimation and inference for the causal effect of receiving treatment on a multinomial outcome. *Biometrics*, 65(1):96–103, 2009.

Linda M Collins, Joseph L Schafer, and Chi-Ming Kam. A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychological methods*, 6(4):330, 2001.

Robert Crosnoe. *Fitting in, standing out: Navigating the social challenges of high school to get an education*. Cambridge University Press, 2011.

Ben Dalton, Elizabeth Glennie, and Steven J Ingels. Late high school dropouts: Characteristics, experiences, and changes across cohorts. descriptive analysis report. nces 2009-307. *National center for education statistics*, 2009.

Arthur P Dempster, Nan M Laird, and Donald B Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1):1–22, 1977.

Arthur P Dempster, Donald B Rubin, and Robert K Tsutakawa. Estimation in covariance components models. *Journal of the American Statistical Association*, 76(374):341–353, 1981.

Richard A Deyo, Daniel C Cherkin, and Marcia A Ciol. Adapting a clinical comorbidity index for use with icd-9-cm administrative databases. *Journal of clinical epidemiology*, 45 (6):613–619, 1992.

Graham Dunn, Mohammad Maracy, Christopher Dowrick, Jose Luis Ayuso-Mateos, Odd Steffen Dalgard, Helen Page, Ville Lehtinen, Patricia Casey, Clare Wilkinson, José Luis Vázquez-Barquero, et al. Estimating psychological treatment effects from a randomised controlled trial with both non-compliance and loss to follow-up. *The British Journal of Psychiatry*, 183(4):323–331, 2003.

Carol S Dweck. *Self-theories: Their role in motivation, personality, and development*. Psychology press, 2000.

Jennifer Elston Lafata, Deirdre A Shires, Yongyun Shin, Susan Flocke, Kenneth Resnicow, Morgan Johnson, Ellen Nixon, Xinxin Sun, and Sarah Hawley. Opportunities and challenges when using the electronic health record for practice-integrated patient-facing interventions: The e-assist colon health randomized trial. *Medical Decision Making*, page 0272989X221104094, 2022.

Polly Feigl, Brent Blumenstein, Ian Thompson, John Crowley, Michael Wolf, Barnett S Kramer, Charles A Coltman Jr, Otis W Brawley, and Leslie G Ford. Design of the prostate cancer prevention trial (pcpt). *Controlled clinical trials*, 16(3):150–163, 1995.

Constantine E Frangakis and Donald B Rubin. Addressing complications of intention-to-treat analysis in the combined presence of all-or-none treatment-noncompliance and subsequent missing outcomes. *Biometrika*, 86(2):365–379, 1999.

Ben B Hansen. The prognostic analogue of the propensity score. *Biometrika*, 95(2):481–488, 2008.

Norman Hearst, Thomas B Newman, and Stephen B Hulley. Delayed effects of the military draft on mortality. *New England Journal of Medicine*, 314(10):620–624, 1986.

Donald Hedeker and Robert D Gibbons. A random-effects ordinal regression model for multilevel analysis. *Biometrics*, pages 933–944, 1994.

Nicholas C Henderson and Ravi Varadhan. Damped anderson acceleration with restarts and monotonicity control for accelerating em and em-like algorithms. *Journal of Computational and Graphical Statistics*, 28(4):834–846, 2019.

Paul W Holland. Statistics and causal inference. *Journal of the American statistical Association*, 81(396):945–960, 1986.

Guanglei Hong and Stephen W Raudenbush. Evaluating kindergarten retention policy: A case study of causal inference for multilevel observational data. *Journal of the American Statistical Association*, 101(475):901–910, 2006.

Guido W Imbens and Donald B Rubin. Bayesian inference for causal effects in randomized experiments with noncompliance. *The annals of statistics*, pages 305–327, 1997a.

Guido W Imbens and Donald B Rubin. Estimating outcome distributions for compliers in instrumental variables models. *The Review of Economic Studies*, 64(4):555–574, 1997b.

Guido W Imbens and Donald B Rubin. *Causal inference in statistics, social, and biomedical sciences*. Cambridge University Press, 2015.

Booil Jo, Elizabeth M Ginexi, and Nicholas S Ialongo. Handling missing data in randomized experiments with noncompliance. *Prevention Science*, 11(4):384–396, 2010.

Jennifer Elston Lafata, Greg Cooper, George Divine, Nancy Oja-Tebbe, and Susan A Flocke. Patient–physician colorectal cancer screening discussion content and patients' use of colorectal cancer screening. *Patient education and counseling*, 94(1):76–82, 2014.

Jennifer Elston Lafata, Yongyun Shin, Susan A Flocke, Sarah T Hawley, Resa M Jones, Ken Resnicow, Michelle Schreiber, Deirdre A Shires, and Shin-Ping Tu. Randomised trial to evaluate the effectiveness and impact of offering postvisit decision support and assistance in obtaining physician-recommended colorectal cancer screening: the e-assist: Colon health study—a protocol study. *BMJ open*, 9(1):e023986, 2019.

Paul J Lavrakas, Michael W Traugott, Courtney Kennedy, Allyson L Holbrook, Edith D de Leeuw, and Brady T West. *Experimental methods in survey research: Techniques that combine random sampling with random assignment*. John Wiley & Sons, 2019.

Xihong Lin and Norman E Breslow. Bias correction in generalized linear mixed models with multiple components of dispersion. *Journal of the American Statistical Association*, 91 (435):1007–1016, 1996.

Roderick J Little and Linda HY Yau. Statistical techniques for analyzing data from prevention trials: Treatment of no-shows using rubin's causal model. *Psychological Methods*, 3 (2):147, 1998.

Roderick J Little, Qi Long, and Xihong Lin. A comparison of methods for estimating the causal effect of a treatment in randomized clinical trials subject to noncompliance. *Biometrics*, 65(2):640–649, 2009.

Qing Liu and Donald A Pierce. Heterogeneity in mantel-haenszel-type models. *Biometrika*, 80(3):543–556, 1993.

Qing Liu and Donald A Pierce. A note on gauss—hermite quadrature. *Biometrika*, 81(3): 624–629, 1994.

Joel McFarland, Jiashan Cui, Amy Rathbun, and Juliet Holmes. Trends in high school dropout and completion rates in the united states: 2018. compendium report. nces 2019-117. *National Center for Education Statistics*, 2018.

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

Yasuo Miyazaki and Kenneth A Frank. A hierarchical linear model with factor analysis structure at level 2. *Journal of Educational and Behavioral Statistics*, 31(2):125–156, 2006.

John C Naylor and Adrian FM Smith. Applications of a method for the efficient computation of posterior distributions. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 31(3):214–225, 1982.

Jerzy S Neyman. On the application of probability theory to agricultural experiments. essay on principles. section 9.(tlanslated and edited by dm dabrowska and tp speed, statistical science (1990), 5, 465-480). *Annals of Agricultural Sciences*, 10:1–51, 1923.

Judea Pearl. On the testability of causal models with latent and instrumental variables. *Uncertainty in Artificial Intelligence*, 11:435–443, 02 1995.

Judea Pearl. Aspects of graphical models connected with causality. *UCLA: Department of Statistics, UCLA. Retrieved from https://escholarship.org/uc/item/9zx0h8k6*, 2011.

José C Pinheiro and Douglas M Bates. Approximations to the log-likelihood function in the nonlinear mixed-effects model. *Journal of computational and Graphical Statistics*, 4(1): 12–35, 1995.

Sophia Rabe-Hesketh, Anders Skrondal, and Andrew Pickles. Reliable estimation of generalized linear mixed models using adaptive quadrature. *The Stata Journal*, 2(1):1–21, 2002.

Stephen W Raudenbush, Meng-Li Yang, and Matheos Yosef. Maximum likelihood for generalized linear models with nested random effects via high-order, multivariate laplace approximation. *Journal of computational and Graphical Statistics*, 9(1):141–157, 2000.

Stephen W Raudenbush, Sean F Reardon, and Takako Nomi. Statistical analysis for multisite trials using instrumental variables with random coefficients. *Journal of research on Educational Effectiveness*, 5(3):303–332, 2012.

Donald Rubin. Discussion of" randomization analysis of experimental data in the fisher randomization test" by d. basu. *Journal of the American statistical association*, 75:591–593, 1980.

Donald B Rubin. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of educational Psychology*, 66(5):688, 1974.

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

Donald B Rubin. Bayesian inference for causal effects: The role of randomization. *The Annals of statistics*, pages 34–58, 1978.

Donald B Rubin. Statistical matching using file concatenation with adjusted weights and multiple imputations. *Journal of Business & Economic Statistics*, 4(1):87–94, 1986.

Kenneth F Schulz and David A Grimes. Generation of allocation sequences in randomised trials: chance, not choice. *The Lancet*, 359(9305):515–519, 2002.

Claude E. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27(3):379–423, 1948. doi:10.1002/j.1538-7305.1948.tb01338.x.

Yongyun Shin and Stephen W Raudenbush. The causal effect of class size on academic achievement: Multivariate instrumental variable estimators with data missing at random. *Journal of Educational and Behavioral Statistics*, 36(2):154–185, 2011.

Yongyun Shin and Stephen W Raudenbush. Maximum likelihood estimation of hierarchical linear models from incomplete data: Random coefficients, statistical interactions, and measurement error. *Submitted*, 2023.

Rebecca Siegel, Carol DeSantis, and Ahmedin Jemal. Colorectal cancer statistics, 2014. *CA: a cancer journal for clinicians*, 64(2):104–117, 2014.

Elena Silva and Taylor White. Pathways to improvement: Using psychological strategies to help college students master developmental math. *Carnegie Foundation for the Advancement of Teaching*, 2013.

Dylan S Small, Thomas R Ten Have, Marshall M Joffe, and Jing Cheng. Random effects logistic models for analysing efficacy of a longitudinal randomized treatment with non-adherence. *Statistics in Medicine*, 25(12):1981–2007, 2006.

PJ Solomon and David Roxbee Cox. Nonlinear component of variance models. *Biometrika*, 79(1):1–11, 1992.

Alfred Sommer and Scott L Zeger. On estimating efficacy from clinical trials. *Statistics in medicine*, 10(1):45–52, 1991.

Alfred Sommer, Edi Djunaedi, AA Loeden, Ignatius Tarwotjo, KeithP West JR, Robert Tilden, Lisa Mele, Aceh Study Group, et al. Impact of vitamin a supplementation on childhood mortality: a randomised controlled community trial. *The Lancet*, 327(8491): 1169–1173, 1986.

April Sutton, Amy G Langenkamp, Chandra Muller, and Kathryn S Schiller. Who gets ahead and who falls behind during the transition to high school? academic performance at the intersection of race/ethnicity and gender. *Social problems*, 65(2):154–173, 2018.

Amir Alishahi Tabriz, Patrice Jordan Fleming, Yongyun Shin, Ken Resnicow, Resa M Jones, Susan A Flocke, Deirdre A Shires, Sarah T Hawley, David Willens, and Jennifer Elston Lafata. Challenges and opportunities using online portals to recruit diverse patients to behavioral trials. *Journal of the American Medical Informatics Association*, 26(12):1637–1644, 2019.

Thomas R Ten Have, Michael R Elliott, Marshall Joffe, Elaine Zanutto, and Catherine Datto. Causal models for randomized physician encouragement trials in treating primary care depression. *Journal of the American Statistical Association*, 99(465):16–25, 2004.

Amiram D Vinokur, Richard H Price, and Yaacov Schul. Impact of the jobs intervention on unemployed workers varying in risk for depression. *American journal of community psychology*, 23(1):39–74, 1995.

Christopher R Walters. Inputs in the production of early childhood human capital: Evidence from head start. *American Economic Journal: Applied Economics*, 7(4):76–102, 2015.

Gregory M Walton and Geoffrey L Cohen. A brief social-belonging intervention improves academic and health outcomes of minority students. *Science*, 331(6023):1447–1451, 2011.

Gregory M Walton and Timothy D Wilson. Wise interventions: Psychological remedies for social and personal problems. *Psychological review*, 125(5):617, 2018.

Linda HY Yau and Roderick J Little. 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, 2001.

David S Yeager, Carissa Romero, Dave Paunesku, Christopher S Hulleman, Barbara Schneider, Cintia Hinojosa, Hae Yeon Lee, Joseph O'Brien, Kate Flint, Alice Roberts, et al. Using design thinking to improve psychological interventions: The case of the growth mindset during the transition to high school. *Journal of educational psychology*, 108(3): 374, 2016.

David S Yeager, Paul Hanselman, Gregory M Walton, Jared S Murray, Robert Crosnoe, Chandra Muller, Elizabeth Tipton, Barbara Schneider, Chris S Hulleman, Cintia P Hinojosa, et al. A national experiment reveals where a growth mindset improves achievement. *Nature*, 573(7774):364–369, 2019.

Jincheng Zhou, James S Hodges, M Fareed K Suri, and Haitao Chu. A bayesian hierarchical model estimating cace in meta-analysis of randomized clinical trials with noncompliance. *Biometrics*, 75(3):978–987, 2019.

APPENDIX A

NR ESTIMATORS

A.1 One-sided Noncompliance

The score function of $\theta = (\alpha, \gamma, \lambda, \varphi_\tau, \varphi_\Delta,)$ are

$$\begin{aligned}\frac{\partial l}{\partial \theta^T} &= \sum_{j=1}^{J} \frac{\partial l_j}{\partial \theta^T} = \sum_{j=1}^{J} \frac{1}{L_j} \frac{\partial L_j}{\partial \theta^T} \\ &= \sum_{j=1}^{J} \frac{1}{L_j} \iint \frac{\partial f(Y_j, C_{oj}|u_j, b_j)}{\partial \theta^T} \phi(u_j)\phi(b_j) du_j db_j \\ &= \sum_{j=1}^{J} \frac{1}{L_j} \iint \frac{\partial log(f(Y_j, C_{oj}|u_j, b_j))}{\partial \theta^T} h(u_j, b_j) du_j db_j\end{aligned} \tag{A.1}$$

where $\phi(u_j) = \phi(u_j|0, \tau), \phi(b_j) = \phi(b_j|0, \Delta)$, and $h(u_j, b_j) = f(Y_j, C_{oj}|u_j, b_j)\phi(u_j)\phi(b_j)$,

$$\begin{aligned}\frac{\partial l}{\partial \alpha^T} &= \sum_{j=1}^{J} \frac{1}{L_j} \iint \frac{\partial log(f(Y_j, C_{oj}|u_j, b_j))}{\partial \alpha^T} h(u_j, b_j) du_j db_j, \\ \frac{\partial l}{\partial \gamma_c^T} &= \sum_{j=1}^{J} \frac{1}{L_j} \iint \frac{\partial log(f(Y_j, C_{oj}|u_j, b_j))}{\partial \gamma_c^T} h(u_j, b_j) du_j db_j, \\ \frac{\partial l}{\partial \varphi_\tau^T} &= \sum_{j=1}^{J} \frac{1}{L_j} \iint \frac{\partial log(\phi(u_j))}{\partial \varphi_\tau^T} h(u_j, b_j) du_j db_j, \\ \frac{\partial l}{\partial \varphi_\Delta^T} &= \sum_{j=1}^{J} \frac{1}{L_j} \iint \frac{\partial log(\phi(b_j))}{\partial \varphi_\Delta^T} h(u_j, b_j) du_j db_j, \\ \frac{\partial l}{\partial \lambda_k^T} &= \sum_{j=1}^{J} \frac{1}{L_j} \iint \frac{\partial log(f(Y_j, C_{oj}|u_j, b_j))}{\partial \lambda_k^T} h(u_j, b_j) du_j db_j.\end{aligned} \tag{A.2}$$

We define $f_j = f(Y_j, C_{oj}|u_j, b_j)$ and $log(f_j) = lnf_j$ such that

$$\begin{aligned}
\frac{\partial lnf_j}{\partial \alpha^T} =& \sum_{\{i:T_{ij}=1,D_{ij}=0\}} (Y_{ij} - \pi_{yn_{ij}})A_n^T + \sum_{\{i:T_{ij}=0,D_{ij}=0\}} \frac{D'_{00}}{D_{00}} + \sum_{\{i:T_{ij}=1,D_{ij}=1\}} (Y_{ij} - \pi_{ycij})A_c^T, \\
\frac{\partial lnf_j}{\partial \gamma_c^T} =& \sum_{\{i:T_{ij}=0,D_{ij}=0\}} \frac{exp[Y_{ij}\eta_{ycij} - log(1+e^{\eta_{ycij}}) + \eta_{cij}]X_{ij}^T}{D_{00}} \\
&+ \sum_{\{i:T_{ij}=1,D_{ij}=1\}} X_{ij}^T - \sum_{i=1}^{n_j} \frac{e^{\eta_{cij}}X_{ij}^T}{1+e^{\eta_{cij}}}, \\
\frac{\partial lnf_j}{\partial \lambda_k} =& \sum_{\{i:T_{ij}=1,D_{ij}=0\}} \left(Y_{ij} - \frac{e^{\eta_{ynij}}}{1+e^{\eta_{ynij}}}\right)\frac{\partial \eta_{ynij}}{\partial \lambda_k} + \sum_{\{i:T_{ij}=0,D_{ij}=0\}} \frac{\frac{\partial D_{00n}}{\partial \lambda_k} + \frac{\partial D_{00c}}{\partial \lambda_k}}{D_{00}} \\
&+ \sum_{\{i:T_{ij}=1,D_{ij}=1\}} \left(Y_{ij} - \frac{e^{\eta_{ycij}}}{1+e^{\eta_{ycij}}}\right)\frac{\partial \eta_{ycij}}{\partial \lambda_k}, \\
\frac{\partial ln(\phi(u_j))}{\partial \varphi_\tau^T} =& \frac{1}{2}\left[\frac{\partial vec(\tau)}{\partial \varphi_\tau^T}\right]^T (\tau^{-1} \otimes \tau^{-1})vec(u_j u_j^T - \tau), \\
\frac{\partial ln(\phi(b_j))}{\partial \varphi_\Delta^T} =& \frac{1}{2}\left[\frac{\partial vec(\Delta)}{\partial \varphi_\Delta^T}\right]^T (\Delta^{-1} \otimes \Delta^{-1})vec(b_j b_j^T - \Delta).
\end{aligned}$$

where

$$\begin{aligned}
D_{00} =& exp[Y_{ij}\eta_{yn_{ij}} - log(1+e^{\eta_{yn_{ij}}})] + exp[Y_{ij}\eta_{yc_{ij}} - log(1+e^{\eta_{yc_{ij}}}) + \eta_{cij}], \\
D'_{00} =& \frac{\partial D_{00}}{\partial \alpha^T} = exp[Y_{ij}\eta_{yn_{ij}} - log(1+e^{\eta_{yn_{ij}}})](Y_{ij} - \pi_{yn_{ij}})A_n^T \\
&+ exp[Y_{ij}\eta_{yc_{ij}} - log(1+e^{\eta_{yc_{ij}}}) + \eta_{cij}](Y_{ij} - \pi_{yc_{ij}})A_c^T, \\
\frac{\partial D_{00n}}{\partial \lambda_k} =& exp[Y_{ij}\eta_{ynij} - log(1+e^{\eta_{ynij}})](Y_{ij} - \pi_{ynij})\frac{\partial \eta_{ynij}}{\partial \lambda_k}, \\
\frac{\partial D_{00c}}{\partial \lambda_k} =& exp[Y_{ij}\eta_{ycij} - log(1+e^{\eta_{ycij}}) + \eta_c](Y_{ij} - \pi_{ycij})\frac{\partial \eta_{ycij}}{\partial \lambda_k},
\end{aligned}$$

and $\frac{\partial \eta_y}{\partial \lambda_k} = B_{ij}^T M u'_j$. Let $M_1 = \begin{pmatrix} 0 & 1 & 0 \end{pmatrix}^T$, $M_2 = \begin{pmatrix} 1 & 0 & 0 \end{pmatrix}^T$ and $M_3 = \begin{pmatrix} 0 & 0 & 1 \end{pmatrix}^T$. When $r = 1$, then $M = M_1$, M_2 or M_3; when $r = 2$, $M = [M_1\, 0]$ or $[M_2\, 0]$ or $[0\, M_3]$.

Next we derive $\frac{\partial^2 l}{\partial\theta\partial\theta^T} = \sum_{j=1}^{J} \frac{\partial^2 l_j}{\partial\theta\partial\theta^T} = \sum_{j=1}^{J} (\frac{1}{L_j}\frac{\partial^2 L_j}{\partial\theta\partial\theta^T} - \frac{1}{L_j^2}\frac{\partial L_j}{\partial\theta}\frac{\partial L_j}{\partial\theta^T})$ where

$$
\begin{aligned}
\frac{\partial^2 L_j}{\partial\alpha\partial\alpha^T} &= \iint \left[\frac{\partial^2 lnf_j}{\partial\alpha\partial\alpha^T} + \frac{\partial lnf_j}{\partial\alpha}\frac{\partial lnf_j}{\partial\alpha^T}\right] h(u_j,b_j)du_j db_j,\\
\frac{\partial^2 L_j}{\partial\gamma_c\partial\gamma_c^T} &= \iint \left[\frac{\partial^2 lnf_j}{\partial\gamma_c\partial\gamma_c^T} + \frac{\partial lnf_j}{\partial\gamma_c}\frac{\partial lnf_j}{\partial\gamma_c^T}\right] h(u_j,b_j)du_j db_j,\\
\frac{\partial^2 L_j}{\partial\lambda_k\partial\lambda_k^T} &= \iint \left[\frac{\partial^2 lnf_j}{\partial\lambda_k\partial\lambda_k^T} + \frac{\partial lnf_j}{\partial\lambda_k}\frac{\partial lnf_j}{\partial\lambda_k^T}\right] h(u_j,b_j)du_j db_j,\\
\frac{\partial^2 L_j}{\partial\varphi_\tau\partial\varphi_\tau^T} &= \iint \left[\frac{\partial^2 ln(\phi(u_j))}{\partial\varphi_\tau\partial\varphi_\tau^T} + \frac{\partial ln(\phi(u_j))}{\partial\varphi_\tau}\frac{\partial ln(\phi(u_j))}{\partial\varphi_\tau^T}\right] h(u_j,b_j)du_j db_j,\\
\frac{\partial^2 L_j}{\partial\varphi_\Delta\partial\varphi_\Delta^T} &= \iint \left[\frac{\partial^2 ln(\phi(b_j))}{\partial\varphi_\Delta\partial\varphi_\Delta^T} + \frac{\partial ln(\phi(b_j))}{\partial\varphi_\Delta}\frac{\partial ln(\phi(b_j))}{\partial\varphi_\Delta^T}\right] h(u_j,b_j)du_j db_j,\\
\frac{\partial^2 L_j}{\partial\gamma_c\partial\alpha^T} &= \iint \left[\frac{\partial^2 lnf_j}{\partial\gamma_c\partial\alpha^T} + \frac{\partial lnf_j}{\partial\gamma_c}\frac{\partial lnf_j}{\partial\alpha^T}\right] h(u_j,b_j)du_j db_j,\\
\frac{\partial^2 L_j}{\partial\lambda_l\partial\lambda_k^T} &= \iint \left[\frac{\partial^2 lnf_j}{\partial\lambda_l\partial\lambda_k^T} + \frac{\partial lnf_j}{\partial\lambda_l}\frac{\partial lnf_j}{\partial\lambda_k^T}\right] h(u_j,b_j)du_j db_j,\\
\frac{\partial^2 L_j}{\partial\alpha\partial\lambda_k^T} &= \iint \left[\frac{\partial^2 lnf_j}{\partial\alpha\partial\lambda_k^T} + \frac{\partial lnf_j}{\partial\alpha}\frac{\partial lnf_j}{\partial\lambda_k^T}\right] h(u_j,b_j)du_j db_j,\\
\frac{\partial^2 L_j}{\partial\gamma_c\partial\lambda_k^T} &= \iint \left[\frac{\partial^2 lnf_j}{\partial\gamma_c\partial\lambda_k^T} + \frac{\partial lnf_j}{\partial\gamma_c}\frac{\partial lnf_j}{\partial\lambda_k^T}\right] h(u_j,b_j)du_j db_j,\\
\frac{\partial^2 L_j}{\partial\varphi_\tau\partial\alpha^T} &= \iint \frac{\partial ln(\phi(u_j))}{\varphi_\tau}\frac{\partial lnf_j}{\partial\alpha^T} h(u_j,b_j)du_j db_j,\\
\frac{\partial^2 L_j}{\partial\varphi_\Delta\partial\alpha^T} &= \iint \frac{\partial ln(\phi(b_j))}{\varphi_\Delta}\frac{\partial lnf_j}{\partial\alpha^T} h(u_j,b_j)du_j db_j,
\end{aligned}
\tag{A.3}
$$

for

$$\frac{\partial^2 lnf_j}{\partial\alpha\partial\alpha^T} = -\sum_{\{i:T_{ij}=1,D_{ij}=0\}}\pi_{ynij}(1-\pi_{ynij})A_nA_n^T - \sum_{\{i:T_{ij}=1,D_{ij}=1\}}\pi_{ycij}(1-\pi_{ycij})A_cA_c^T + \sum_{\{i:T_{ij}=0,D_{ij}=0\}}\frac{D_{00}''D_{00}-(D_{00}')^2}{D_{00}^2},$$

$$\frac{\partial^2 lnf_j}{\partial\gamma_c\partial\gamma_c^T} = \sum_{\{i:T_{ij}=0,D_{ij}=0\}}\frac{\frac{\partial^2 D_{00}}{\partial\gamma_c\partial\gamma_c^T}D_{00}-\frac{\partial D_{00}}{\partial\gamma_c}\frac{\partial D_{00}}{\partial\gamma_c^T}}{D_{00}^2} - \sum_{i=1}^{n_j}\pi_{cij}(1-\pi_{cij})X_{ij}X_{ij}^T,$$

$$\frac{\partial^2 lnf_j}{\partial\lambda_k^2} = \sum_{\{i:T_{ij}=1,D_{ij}=0\}}\left[-\frac{e^{\eta_{ynij}}}{(1+e^{\eta_{ynij}})^2}\left(\frac{\partial\eta_{ynij}}{\partial\lambda_k}\right)^2\right] + \sum_{\{i:T_{ij}=1,D_{ij}=1\}}\left[-\frac{e^{\eta_{ycij}}}{(1+e^{\eta_{ycij}})^2}\left(\frac{\partial\eta_{ycij}}{\partial\lambda_k}\right)^2\right] + \sum_{\{i:T_{ij}=0,D_{ij}=0\}}\frac{(\frac{\partial^2 D_{00n}}{\partial\lambda_k^2}+\frac{\partial^2 D_{00c}}{\partial\lambda_k^2})D_{00}-\left(\frac{\partial D_{00}}{\partial\lambda_k}\right)^2}{D_{00}^2},$$

$$\frac{\partial^2 ln(\phi(u_j))}{\partial\varphi_\tau\partial\varphi_\tau^T} = \frac{1}{2}\left(\frac{dvec(\tau)}{d\varphi_\tau^T}\right)^T(\tau^{-1}-2\tau^{-1}u_ju_j^T\tau^{-1})\otimes\tau^{-1}\frac{dvec(\tau)}{d\varphi_\tau^T},$$

$$\frac{\partial^2 ln(\phi(b_j))}{\partial\varphi_\Delta\partial\varphi_\Delta^T} = \frac{1}{2}\left(\frac{dvec(\Delta)}{d\varphi_\Delta^T}\right)^T(\Delta^{-1}-2\Delta^{-1}b_jb_j^T\Delta^{-1})\otimes\Delta^{-1}\frac{dvec(\Delta)}{d\varphi_\Delta^T},$$

$$\frac{\partial^2 lnf_j}{\partial\gamma_c\partial\alpha^T} = \sum_{\{i:T_{ij}=0,D_{ij}=0\}}\frac{\frac{\partial D_{00}'}{\partial\gamma_c}D_{00}-\frac{\partial D_{00}}{\partial\gamma_c}D_{00}'}{D_{00}^2},$$

$$\frac{\partial^2 lnf_j}{\partial\lambda_k\partial\lambda_l} = \sum_{\{i:T_{ij}=1,D_{ij}=0\}}\left[-\frac{e^{\eta_{ynij}}}{(1+e^{\eta_{ynij}})^2}\frac{\partial\eta_{ynij}}{\partial\lambda_l}\frac{\partial\eta_{ynij}}{\partial\lambda_k}\right] + \sum_{\{i:T_{ij}=1,D_{ij}=1\}}\left[-\frac{e^{\eta_{ycij}}}{(1+e^{\eta_{ycij}})^2}\frac{\partial\eta_{ycij}}{\partial\lambda_l}\frac{\partial\eta_{ycij}}{\partial\lambda_k}\right] + \sum_{\{i:T_{ij}=0,D_{ij}=0\}}\frac{(\frac{\partial^2 D_{00n}}{\partial\lambda_k\partial\lambda_l}+\frac{\partial^2 D_{00c}}{\partial\lambda_k\partial\lambda_l})D_{00}-\frac{\partial D_{00}}{\partial\lambda_k}\frac{\partial D_{00}}{\partial\lambda_l}}{D_{00}^2},$$

$$\frac{\partial^2 lnf_j}{\partial\lambda_k\partial\alpha} = \sum_{\{i:T_{ij}=1,D_{ij}=0\}}\left[-\frac{e^{\eta_{ynij}}}{(1+e^{\eta_{ynij}})^2}A_n^T\right]\frac{\partial\eta_{ynij}}{\partial\lambda_k} + \sum_{\{i:T_{ij}=1,D_{ij}=1\}}\left[-\frac{e^{\eta_{ycij}}}{(1+e^{\eta_{ycij}})^2}A_c^T\right]\frac{\partial\eta_{ycij}}{\partial\lambda_k} + \sum_{\{i:T_{ij}=0,D_{ij}=0\}}\frac{(\frac{\partial^2 D_{00n}}{\partial\lambda_k\partial\alpha}+\frac{\partial^2 D_{00c}}{\partial\lambda_k\partial\alpha})D_{00}-\frac{\partial D_{00}}{\partial\lambda_k}\frac{\partial D_{00}}{\partial\alpha}}{D_{00}^2},$$

$$\frac{\partial^2 lnf_j}{\partial\lambda_k\partial\gamma_c} = \sum_{\{i:T_{ij}=0,D_{ij}=0\}}\frac{(\frac{\partial^2 D_{00c}}{\partial\lambda_k\partial\gamma_c})D_{00}-\frac{\partial D_{00c}}{\partial\lambda_k}\frac{\partial D_{00}}{\partial\gamma_c}}{D_{00}^2}.$$

Here,

$$
\begin{aligned}
D''_{00} =& \frac{\partial^2 D_{00}}{\partial\alpha\partial\alpha^T} \\
=& exp[Y_{ij}\eta_{ynij} - log(1+e^{\eta_{ynij}})][(Y_{ij}-\pi_{ynij})^2 - \pi_{ynij}(1-\pi_{ynij})]A_nA_n^T \\
&+ exp[Y_{ij}\eta_{ycij} - log(1+e^{\eta_{ycij}}) + \eta_{cij}][(Y_{ij}-\pi_{ycij})^2 - \pi_{ycij}(1-\pi_{ycij})]A_cA_c^T, \\
\frac{\partial D_{00}}{\partial\gamma_c} =& exp[Y_{ij}\eta_{ycij} - log(1+e^{\eta_{ycij}}) + \eta_{cij}]X_{ij}, \\
\frac{\partial D'_{00}}{\partial\gamma_c} =& exp[Y_{ij}\eta_{ycij} - log(1+e^{\eta_{ycij}}) + \eta_{cij}](Y_{ij}-\pi_{cij})X_{ij}A_c^T, \\
\frac{\partial^2 D_{00}}{\partial\gamma_c\partial\gamma_c^T} =& exp[Y_{ij}\eta_{ycij} - log(1+e^{\eta_{ycij}}) + \eta_{cij}]X_{ij}X_{ij}^T, \\
\frac{\partial^2 D_{00n}}{\partial\lambda_k^2} =& exp[Y_{ij}\eta_{ynij} - log(1+e^{\eta_{ynij}})]\left(\frac{\partial\eta_{ynij}}{\partial\lambda_k}\right)^2\left[\left(Y_{ij}-\frac{e^{\eta_{ynij}}}{1+e^{\eta_{ynij}}}\right)^2 - \frac{e^{\eta_{ynij}}}{(1+e^{\eta_{ynij}})^2}\right], \\
\frac{\partial^2 D_{00c}}{\partial\lambda_k^2} =& exp[Y_{ij}\eta_{ycij} - log(1+e^{\eta_{ycij}}) + \eta_c]\left(\frac{\partial\eta_{ycij}}{\partial\lambda_k}\right)^2\left[\left(Y_{ij}-\frac{e^{\eta_{ycij}}}{1+e^{\eta_{ycij}}}\right)^2 - \frac{e^{\eta_{ycij}}}{(1+e^{\eta_{ycij}})^2}\right], \\
\frac{\partial^2 D_{00n}}{\partial\lambda_k\partial\lambda_l} =& exp[Y_{ij}\eta_{ynij} - log(1+e^{\eta_{ynij}})]\left(\frac{\partial\eta_{ynij}}{\partial\lambda_l}\right)\left[\left(Y_{ij}-\frac{e^{\eta_{ynij}}}{1+e^{\eta_{ynij}}}\right)^2 - \frac{e^{\eta_{ynij}}}{(1+e^{\eta_{ynij}})^2}\right]\left(\frac{\partial\eta_{ynij}}{\partial\lambda_k}\right), \\
\frac{\partial^2 D_{00c}}{\partial\lambda_k\partial\lambda_l} =& exp[Y_{ij}\eta_{ycij} - log(1+e^{\eta_{ycij}}) + \eta_c]\left(\frac{\partial\eta_{ycij}}{\partial\lambda_l}\right)\left[\left(Y_{ij}-\frac{e^{\eta_{ycij}}}{1+e^{\eta_{ycij}}}\right)^2 - \frac{e^{\eta_{ycij}}}{(1+e^{\eta_{ycij}})^2}\right]\left(\frac{\partial\eta_{ycij}}{\partial\lambda_k}\right), \\
\frac{\partial^2 D_{00n}}{\partial\lambda_k\partial\alpha} =& exp[Y_{ij}\eta_{ynij} - log(1+e^{\eta_{ynij}})]A_n^T\left(\frac{\partial\eta_{ynij}}{\partial\lambda_k}\right)\left[\left(Y_{ij}-\frac{e^{\eta_{ynij}}}{1+e^{\eta_{ynij}}}\right)^2 - \frac{e^{\eta_{ynij}}}{(1+e^{\eta_{ynij}})^2}\right], \\
\frac{\partial^2 D_{00c}}{\partial\lambda_k\partial\alpha} =& exp[Y_{ij}\eta_{ycij} - log(1+e^{\eta_{ycij}}) + \eta_c]A_c^T\left(\frac{\partial\eta_{ycij}}{\partial\lambda_k}\right)\left[\left(Y_{ij}-\frac{e^{\eta_{ycij}}}{1+e^{\eta_{ycij}}}\right)^2 - \frac{e^{\eta_{ycij}}}{(1+e^{\eta_{ycij}})^2}\right], \\
\frac{\partial^2 D_{00c}}{\partial\lambda_k\partial\gamma_c} =& exp[Y_{ij}\eta_{ycij} - log(1+e^{\eta_{ycij}}) + \eta_c]X_{ij}^T\left(\frac{\partial\eta_{ycij}}{\partial\lambda_k}\right)(Y_{ij}-\pi_{ycij}), \\
\frac{\partial D_{00c}}{\partial\lambda_k} =& exp[Y_{ij}\eta_{ycij} - log(1+e^{\eta_{ycij}}) + \eta_c](Y_{ij}-\pi_{ycij})\frac{\partial\eta_{ycij}}{\partial\lambda_k}.
\end{aligned}
$$

A.2 Two-sided Noncompliance

The difference in the likelihood between one-sided and two-sided noncompliance is the inclusion of always takers in the model. However, the general presentations of the derivatives of the log-likelihood with respect to α or γ_c do not change. The first and second derivatives

of the logarithmic joint probability mass function with respect to α, γ_a and γ_c are:

$$\frac{\partial log(f(y_j, c_{oj}|u_j, b_j))}{\partial \alpha^T} = \sum_{\{i:T_{ij}=1, D_{ij}=0\}} (y_{ij} - \pi_{yn_{ij}}) A_n^T + \sum_{\{i:T_{ij}=0, D_{ij}=1\}} (y_{ij} - \pi_{ya_{ij}}) A_a^T + \sum_{\{i:T_{ij}=0, D_{ij}=0\}} \frac{D'_{00}}{D_{00}} + \sum_{\{i:T_{ij}=1, D_{ij}=1\}} \frac{D'_{11}}{D_{11}}$$

for

$$\begin{aligned}
D_{00} =& exp[y_{ij}\eta_{yn_{ij}} - log(1 + e^{\eta_{yn_{ij}}})] + exp[y_{ij}\eta_{yc_{ij}} - log(1 + e^{\eta_{yc_{ij}}}) + \eta_{cij}] \\
D'_{00} =& \frac{\partial D_{00}}{\partial \alpha^T} \\
=& exp[y\eta_{yn_{ij}} - log(1 + e^{\eta_{yn_{ij}}})](y_{ij} - \pi_{yn_{ij}}) A_n^T \\
& + exp[y_{ij}\eta_{yc_{ij}} - log(1 + e^{\eta_{yc_{ij}}}) + \eta_{cij}](y_{ij} - \pi_{yc_{ij}}) A_c^T \\
D_{11} =& exp[y_{ij}\eta_{ya_{ij}} - log(1 + e^{\eta_{ya_{ij}}}) + \eta_{aij}] + exp[y_{ij}\eta_{yc_{ij}} - log(1 + e^{\eta_{yc_{ij}}}) + \eta_{cij}] \\
D'_{11} =& \frac{\partial D_{11}}{\partial \alpha^T} \\
=& exp[y\eta_{ya_{ij}} - log(1 + e^{\eta_{ya_{ij}}})](y_{ij} - \pi_{ya_{ij}}) A_a^T \\
& + exp[y_{ij}\eta_{yc_{ij}} - log(1 + e^{\eta_{yc_{ij}}}) + \eta_{cij}](y_{ij} - \pi_{yc_{ij}}) A_c^T;
\end{aligned}$$

$$\frac{\partial log(f(y_j, c_{oj}|u_j, b_j))}{\partial \gamma_a^T} = \sum_{\{i:T_{ij}=0, D_{ij}=1\}} X_{ij}^T + \sum_{\{i:T_{ij}=1, D_{ij}=1\}} \frac{exp[y_{ij}\eta_{ya_{ij}} - log(1 + e^{\eta_{yaij}}) + \eta_{aij}]}{D_{11}} - \sum_{i=1}^{n_j} \frac{e^{\eta_{aij}} X_{ij}^T}{1 + e^{\eta_{aij}} + e^{\eta_{cij}}};$$

$$\frac{\partial log(f(y_j, c_{oj}|u_j, b_j))}{\partial \gamma_c^T} = \sum_{\{i:T_{ij}=0, D_{ij}=0\}} \frac{exp[y_{ij}\eta_{ycij} - log(1+e^{\eta_{ycij}}) + \eta_{cij}]X_{ij}^T}{D_{00}} + \sum_{\{i:T_{ij}=1, D_{ij}=1\}} \frac{exp[y_{ij}\eta_{ycij} - log(1+e^{\eta_{ycij}}) + \eta_{cij}]X_{ij}^T}{D_{11}} - \sum_{i=1}^{n_j} \frac{e^{\eta_{cij}} X_{ij}^T}{1 + e^{\eta_{aij}} + e^{\eta_{cij}}}.$$

A.3 One-sided Noncompliance with Missing Outcomes

General Form of Score Function

$$\frac{\partial l}{\partial \theta} = \sum_j \frac{\partial l_j}{\partial \theta} = \sum_j \frac{1}{L_j} \frac{\partial L_j}{\partial \theta} = \sum_j \frac{1}{L_j} \int \frac{\partial log(f(Y_j, C_{oj}|u_j, b_j))}{\partial \theta} h(u_j, b_j) du_j db_j$$

In the presence of missing outcomes, the log-likelihood becomes

$$log(f(Y_j, C_{oj}|u_j, b_j)) = \sum_{\{i:Y_{ij} \in Yoj\}} log[f(Y_{ij}, C_{oj}|u_j, b_j)] + \sum_{\{i:Y_{ij} \in Ymj\}} log[f(Y_{ij}, C_{oj}|u_j, b_j)].$$

Only the γ_c estimators change in this case.

Score Function of γ_c

$$\frac{\partial log(f(Y_j, C_{oj}|u_j, b_j))}{\partial \gamma_c^T} = \sum_{\{i:Y_{ij} \in Y_{oj}, T_{ij}=0, D_{ij}=0\}} \frac{exp[Y_{ij}\eta_{ycij} - log(1+e^{\eta_{ycij}}) + \eta_{cij}]X_{ij}^T}{D_{00}} + \sum_{\{i:Y_{ij} \in Y_{oj}, T_{ij}=1, D_{ij}=1\}} X_{ij}^T - \sum_{\{i:y_{ij} \in Y_{oj}\}} P_{cij} e^{\eta_{cij}} X_{ij}^T - \sum_{\{i:y_{ij} \in Y_{mj}, T_{ij}=1, D_{ij}=0\}} P_{cij} X_{ij}^T + \sum_{\{i:y_{ij} \in Y_{mj}, T=1, D=1\}} P_{nij} X_{ij}^T$$

where $P_c = \frac{e^{\eta c}}{1+e^{\eta c}}$ and $P_n = \frac{1}{1+e^{\eta c}}$.

Hessian of γ_c

$$\begin{aligned}\frac{\partial^2 log(f(y_j, c_{oj}|u_j, b_j))}{\partial\gamma_c \partial\gamma_c^T} = & \sum_{\{i:y_{ij}\in Y_{oj}, T_{ij}=0, D_{ij}=0\}} \frac{\frac{\partial^2 D_{00}}{\partial\gamma_c \partial\gamma_c^T} D_{00} - \frac{\partial D_{00}}{\partial\gamma_c}\frac{\partial D_{00}}{\partial\gamma_a^T}}{D_{00}^2} \\ & - \sum_{\{i:y_{ij}\in Y_{oj}\}} \pi_{cij}(1-\pi_{cij}) X_{ij} X_{ij}^T \\ & - \sum_{\{i:y_{ij}\in Y_{mj}, T_{ij}=1, D_{ij}=0\}} P_{nij} P_{cij} X_{ij} X_{ij}^T \\ & - \sum_{\{i:y_{ij}\in Y_{mj}, T_{ij}=1, D_{ij}=1\}} P_{nij} P_{cij} X_{ij} X_{ij}^T\end{aligned}$$

APPENDIX B

THE EM ALGORITHM

We estimate τ and Δ by the EM algorithm. The complete data (CD) likelihood is

$$l_{CD} = \sum_{j=1}^{J}\sum_{i=1}^{n_j}[Y_{ij}\eta_{yij} - log(1+e^{\eta_{yij}}) + C_{ij}\eta_{cij} - log(1+\eta_{cij})]$$
$$- 3Jlog(2\pi) - \frac{J}{2}log|\tau| - \frac{1}{2}\sum_{j=1}^{J} u_j^T\tau^{-1}u_j - \frac{J}{2}log|\Delta| - \frac{1}{2}\sum_{j=1}^{J} b_j^T\Delta^{-1}b_j.$$

We maximize l_{CD} with respect to τ and Δ and find the expected CD ML estimators. We let

$$Q(\theta_m;\theta_{m-1}) := E\left(-\frac{J}{2}log|\tau| - \frac{1}{2}\sum_{j=1}^{J} u_j^T\tau^{-1}u_j - \frac{J}{2}log|\Delta| - \frac{1}{2}\sum_{j=1}^{J} b_j^T\Delta^{-1}b_j|\theta_{m-1}, Y_j, C_{oj}\right)$$
$$= -\frac{J}{2}log|\tau| - \frac{1}{2}\sum_{j=1}^{J} E(u_j^T\tau^{-1}u_j|Y_j, C_{oj}, \tau_{m-1})$$
$$- \frac{J}{2}log|\Delta| - \frac{1}{2}\sum_{j=1}^{J} E(b_j^T\Delta^{-1}b_j|Y_j, C_{oj}, \Delta_{m-1})$$
$$= -\frac{J}{2}log|\tau| - \frac{1}{2}\sum_{j=1}^{J} tr[\tau^{-1}E(u_ju_j^T|Y_j, C_{oj})]$$
$$- \frac{J}{2}log|\Delta| - \frac{1}{2}\sum_{j=1}^{J} tr[\Delta^{-1}E(b_jb_j^T|Y_j, C_{oj})]$$
$$= -\frac{J}{2}log|\tau| - \frac{1}{2}tr\left[\tau^{-1}E\left(\sum_{j=1}^{J} u_ju_j^T|Y_j, C_{oj}\right)\right]$$
$$- \frac{J}{2}log|\Delta| - \frac{1}{2}tr\left[\Delta^{-1}E\left(\sum_{j=1}^{J} b_jb_j^T|Y_j, C_{oj}\right)\right],$$

and set $\theta_m := \max_{\theta} Q(\theta_m; \theta_{m-1})$, and have

$$\begin{aligned}
\frac{\partial Q}{\partial \tau} &= -\frac{J}{2} tr(\tau^{-1}) + \frac{1}{2} tr\left[\tau^{-2} E\left(\sum_{j=1}^{J} u_j u_j^T | Y_j, C_{oj}\right)\right] \\
&= -\frac{1}{2} tr\left[J\tau^{-1} - \tau^{-2} E\left(\sum_{j=1}^{J} u_j u_j^T | Y_j, C_{oj}\right)\right] \\
&= -\frac{1}{2} tr\left[\tau^{-1}\left(J - \tau^{-1} E\left(\sum_{j=1}^{J} u_j u_j^T | Y_j, C_{oj}\right)\right)\right] := 0 \\
&\implies J - \tau^{-1} E\left(\sum_{j=1}^{J} u_j u_j^T | Y_j, C_{oj}\right) = 0 \\
&\implies \hat{\tau} = \frac{E(\sum_{j=1}^{J} u_j u_j^T | Y_j, C_{oj})}{J}.
\end{aligned}$$

Similarly, we have

$$\hat{\Delta} = \frac{E(\sum_{j=1}^{J} b_j b_j^T | Y_j, C_{oj})}{J}.$$

APPENDIX C

APPROXIMATING ESTIMATORS BY AGHQ

For each estimator expressed as an integral of the random effects, we approximate numerically by AGHQ. The formula to approximate it is as below. We estimate an EM or NR component $g(\nu_j)$, a function of ν_j in `Equation` (2.12) by

$$\begin{aligned}\int g(\nu_j) f(\nu_j|Y_j, C_{oj}) d\nu_j &= \int \frac{\phi(\nu_j|\tilde{v}_j, V_j)}{\phi(\nu_j|\tilde{v}_j, V_j)} \frac{g(\nu_j) f(Y_j, C_{oj}, \nu_j)}{f(Y_j, C_{oj})} d\nu_j \\ &= \frac{|L_{vj}|}{L_j} \int e^{-x_{vj}^T x_{vj}} e^{x_{vj}^T x_{vj}} g(L_{vj} x_{vj} + \tilde{v}_j) h(L_{vj} x_{vj} + \tilde{v}_j) dx_{vj} \\ &\approx \frac{|L_{vj}|}{L_j} \sum_{qr \ldots q1} w_{q1} \ldots w_{qr} e^{A_{q1\ldots qr}^T A_{q1\ldots qr}} g(z_{q1\ldots qrj}) h(z_{q1\ldots qrj}).\end{aligned}$$

APPENDIX D

R CODES

D.1 Functions

```
#cace_1side_parallel_lambda() is the main function for estimating
   CACE with parallel computing
#inputs
#r: full dimension of random effects of Y model
#k: full dimension of random effects of C model. k=1 for one-sided
   noncompliance
#Q: number of abscissas for AGHQ
#J: number of clusters/sites
#x0: covariates controlled in Y model
#x1: covariates controlled in C model
#init: initial values for parameters
#data: dataframe containing outcome, treatment assignment and receipt
   , cluster/site level variable
#col.clinic: column number of cluster/site level variable in the
   dataframe
#col.trt: column number of treatment assignment in the dataframe
#col.D: column number of treatment receipt in the dataframe
#col.Y: column number of outcome variable in the dataframe
#niter: maximum number of iteration. suggest 500
#tol: convergence criteria. suggest 10^-4
#Share: sharing of random effect.
  #When r=1, Share=1 means u'=un, Share=2 means u'=uc0 and Share=3
     means u'=uc1
```

```
  #When r=2, Share=1 means never takers and treatment compliers
     sharing one random effect, Share=2 means never takers and
     control compliers sharing one random effect, Share=3 means
     control compliers and trt compliers sharing one random effect
#r_prime: reduced number of random effects

cace_1side_parallel_lambda <- function(r,k=1,Q,J,data,x0,x1,init,col.
   clinic,col.trt,col.D,col.Y,niter,tol,Share,r_prime){
  start_time <- Sys.time()

  #first derivative lnf/a * f(y)
  lnf_a_fy_1 <- function(u,b,Alpha,gamma,data_j,ind.x0,ind.x1,lambda)
     {
    fun_der_lnf_alpha(u,b,Alpha,gamma,data_j,ind.x0,ind.x1,lambda) *
       f_y_incomplete(u,b,Alpha,gamma,data_j,ind.x0,ind.x1,lambda)
  }

  #second derivative lnf/a * f(y)
  lnf_a_fy_2 <- function(u,b,Alpha,gamma,data_j,ind.x0,ind.x1,lambda)
     {
    fun_2nd_deriv_L_a(u,b,Alpha,gamma,data_j,ind.x0,ind.x1,lambda) *
       f_y_incomplete(u,b,Alpha,gamma,data_j,ind.x0,ind.x1,lambda)
  }

  #first derivative lnf/tau * f(y)
  lnf_t_fy_1 <- function(u,b,Alpha,gamma,data_j,ind.x0,ind.x1,lambda,
     tau){
```

```
  fun_der_1st_ln_phitau(u,tau) * f_y_incomplete(u,b,Alpha,gamma,
     data_j,ind.x0,ind.x1,lambda)
}

#second derivative lnf/tau * f(y)
lnf_t_fy_2 <- function(u,b,Alpha,gamma,data_j,ind.x0,ind.x1,lambda,
   tau){
  fun_2nd_L_tau(u,tau) * f_y_incomplete(u,b,Alpha,gamma,data_j,ind.
     x0,ind.x1,lambda)
}

#first derivative lnf/rc * f(y)
lnf_rc_fy_1 <- function(u,b,Alpha,gamma,data_j,ind.x0,ind.x1,lambda
   ){
  deriv_lnf_rc_1st(u,b,Alpha,gamma,data_j,ind.x0,ind.x1,lambda) * f
     _y_incomplete(u,b,Alpha,gamma,data_j,ind.x0,ind.x1,lambda)
}

#seconc derivative lnf/rc * f(y)
lnf_rc_fy_2 <- function(u,b,Alpha,gamma,data_j,ind.x0,ind.x1,lambda
   ){
  fun_2nd_deriv_L_rc(u,b,Alpha,gamma,data_j,ind.x0,ind.x1,lambda) *
      f_y_incomplete(u,b,Alpha,gamma,data_j,ind.x0,ind.x1,lambda)
}

#second derivative lnf/tau*a * f(y)
lnf_t_a_fy_2 <- function(u,b,Alpha,gamma,data_j,ind.x0,ind.x1,
   lambda,tau){
```

```
  f_y_incomplete(u,b,Alpha,gamma,data_j,ind.x0,ind.x1,lambda) * fun
     _der_1st_ln_phitau(u,tau) %*% fun_der_lnf_alpha(u,b,Alpha,
     gamma,data_j,ind.x0,ind.x1,lambda)
}

#first derivative lnf/delta * f(y)
lnf_delta_fy_1 <- function(u,b,Alpha,gamma,data_j,ind.x0,ind.x1,
   lambda,delta){
  deriv_lnf_delta_1st(b,delta) * f_y_incomplete(u,b,Alpha,gamma,
     data_j,ind.x0,ind.x1,lambda)
}

#second derivative lnf/delta * f(y)
lnf_delta_fy_2 <- function(u,b,Alpha,gamma,data_j,ind.x0,ind.x1,
   lambda,delta){
  fun_2nd_L_delta(b,delta) * f_y_incomplete(u,b,Alpha,gamma,data_j,
     ind.x0,ind.x1,lambda)
}

#second derivative lnf/delta*a * f(y)
lnf_delta_a_fy_2 <- function(u,b,Alpha,gamma,data_j,ind.x0,ind.x1,
   lambda,delta){
  f_y_incomplete(u,b,Alpha,gamma,data_j,ind.x0,ind.x1,lambda) *
     deriv_lnf_delta_1st(b,delta) %*% fun_der_lnf_alpha(u,b,Alpha,
     gamma,data_j,ind.x0,ind.x1,lambda)
}

#second derivative lnf/tau*rc * f(y)
```

```
lnf_t_rc_fy_2 <- function(u,b,Alpha,gamma,data_j,ind.x0,ind.x1,
   lambda,tau){
  f_y_incomplete(u,b,Alpha,gamma,data_j,ind.x0,ind.x1,lambda) * fun
     _der_1st_ln_phitau(u,tau) %*% t(deriv_lnf_rc_1st(u,b,Alpha,
     gamma,data_j,ind.x0,ind.x1,lambda))
}

#second derivative lnf/delta*rc * f(y)
lnf_delta_rc_fy_2 <- function(u,b,Alpha,gamma,data_j,ind.x0,ind.x1,
   lambda,delta){
  f_y_incomplete(u,b,Alpha,gamma,data_j,ind.x0,ind.x1,lambda) *
     deriv_lnf_delta_1st(b,delta) %*% t(deriv_lnf_rc_1st(u,b,Alpha,
     gamma,data_j,ind.x0,ind.x1,lambda))
}

#second derivative lnf/tau*delta * f(y)
lnf_delta_t_fy_2 <- function(u,b,Alpha,gamma,data_j,ind.x0,ind.x1,
   lambda,tau,delta){
  deriv_lnf_delta_1st(b,delta) %*% t(fun_der_1st_ln_phitau(u,tau))
     * f_y_incomplete(u,b,Alpha,gamma,data_j,ind.x0,ind.x1,lambda)
}

#second derivative lnf/rc*a * f(y)
lnf_rc_a_fy_2 <- function(u,b,Alpha,gamma,data_j,ind.x0,ind.x1,
   lambda){
  f_y_incomplete(u,b,Alpha,gamma,data_j,ind.x0,ind.x1,lambda)*fun_2
     nd_deriv_L_alpha_rc(u,b,Alpha,gamma,data_j,ind.x0,ind.x1,
     lambda)
```

```
}

#f(y)*z E(u,b)
f_y_z <- function(u,b,Alpha,gamma,data_j,ind.x0,ind.x1,lambda){
  f_y_incomplete(u,b,Alpha,gamma,data_j,ind.x0,ind.x1,lambda) * c(u
     ,b)
}

#f(y)*z^2 E((u,b)^2)
f_y_z2 <- function(u,b,Alpha,gamma,data_j,ind.x0,ind.x1,lambda){
  f_y_incomplete(u,b,Alpha,gamma,data_j,ind.x0,ind.x1,lambda) * c(u
     ,b) %*% t(c(u,b))
}

#first derivative lnf/lambda * f(y)
lnf_lambda_fy_1 <- function(u,b,Alpha,gamma,data_j,ind.x0,ind.x1,
   lambda,M){
  fun_der_lnf_lambda(u,b,Alpha,gamma,data_j,ind.x0,ind.x1,lambda,M)
      * f_y_incomplete(u,b,Alpha,gamma,data_j,ind.x0,ind.x1,lambda)
}

#second derivative lnf/lambda * f(y)
lnf_lambda_fy_2 <- function(u,b,Alpha,gamma,data_j,ind.x0,ind.x1,
   lambda,M){
  fun_2nd_deriv_L_lambda(u,b,Alpha,gamma,data_j,ind.x0,ind.x1,
     lambda,M) * f_y_incomplete(u,b,Alpha,gamma,data_j,ind.x0,ind.
     x1,lambda)
}
```

```
#second derivative lnf/a*lambda * f(y)
lnf_a_lambda_fy_2 <- function(u,b,Alpha,gamma,data_j,ind.x0,ind.x1,
   lambda,M){
  fun_2nd_deriv_L_lambda_a(u,b,Alpha,gamma,data_j,ind.x0,ind.x1,
     lambda,M) * f_y_incomplete(u,b,Alpha,gamma,data_j,ind.x0,ind.
     x1,lambda)
}

#second derivative lnf/rc*lambda * f(y)
lnf_rc_lambda_fy_2 <- function(u,b,Alpha,gamma,data_j,ind.x0,ind.x1
   ,lambda,M){
  fun_2nd_deriv_L_lambda_rc(u,b,Alpha,gamma,data_j,ind.x0,ind.x1,
     lambda,M) * f_y_incomplete(u,b,Alpha,gamma,data_j,ind.x0,ind.
     x1,lambda)
}

#second derivative lnf/lambda1*lambda2 * f(y)
lnf_lambda1_lambda2_fy_2 <- function(u,b,Alpha,gamma,data_j,ind.x0,
   ind.x1,lambda,M){
  fun_2nd_deriv_L_lambda1_lambda2(u,b,Alpha,gamma,data_j,ind.x0,ind
     .x1,lambda,M) * f_y_incomplete(u,b,Alpha,gamma,data_j,ind.x0,
     ind.x1,lambda)
}

lnf_t_lambda_fy_2 <- function(u,b,Alpha,gamma,data_j,ind.x0,ind.x1,
   lambda,tau,M){
```

```
  f_y_incomplete(u,b,Alpha,gamma,data_j,ind.x0,ind.x1,lambda) * fun
     _der_1st_ln_phitau(u,tau) %*% fun_der_lnf_lambda(u,b,Alpha,
     gamma,data_j,ind.x0,ind.x1,lambda,M)
}

lnf_delta_lambda_fy_2 <- function(u,b,Alpha,gamma,data_j,ind.x0,ind
   .x1,lambda,delta,M){
  f_y_incomplete(u,b,Alpha,gamma,data_j,ind.x0,ind.x1,lambda) *
     deriv_lnf_delta_1st(b,delta) %*% fun_der_lnf_lambda(u,b,Alpha,
     gamma,data_j,ind.x0,ind.x1,lambda,M)
}

colnames(data)[c(col.clinic,col.trt, col.D, col.Y)] <- c("clinic",
   "trt", "D", "Y")

if(anyNA(x0)){
  ind.x0 <- NA
  l.alpha <- 3
} else {
  ind.x0 <- which(colnames(data) %in% c(x0))
  l.alpha <- 3 + length(x0)
}

if(anyNA(x1)){
  ind.x1 <- NA
  l.gamma <- 1
} else {
```

```
  ind.x1 <- which(colnames(data) %in% c(x1)) #x1 should be input as
      a vector of character with "". covariate order need to comply
      with the order of collumns in dataset
  l.gamma <- 1+length(x1)
}

for(i in 1:nrow(data)){
  if(data$trt[i]==1 & data$D[i]==1){
    data$C[i] <- 1
  }
  if(data$trt[i]==1 & data$D[i]==0){
    data$C[i] <- 0
  }
  if(data$trt[i]==0 & data$D[i]==0){
    data$C[i] <- NA
  }
}

nlambda <- r-r_prime

if(nlambda==1){
  lambda1 <- init[l.alpha+l.gamma+1]
}
if(nlambda==2){
  lambda1 <- init[l.alpha+l.gamma+1]
  lambda2 <- init[l.alpha+l.gamma+2]
}
```

```
if(!missing(r_prime)){
  if(r_prime==1){
    if(Share==1) lambda <- as.matrix(c(1,lambda1,lambda2)) #share
       never taker's random effect. u'=un
    if(Share==2) lambda <- as.matrix(c(lambda1,1,lambda2)) #share
       control complier's random effect. u'=uc0
    if(Share==3) lambda <- as.matrix(c(lambda1,lambda2,1)) ##share
       trt complier's random effect. u'=uc1
  }
  if(r_prime==2){
    if(Share==1) lambda <- matrix(c(1,0,lambda1,0,1,0),nrow = 3,
       ncol = r_prime,byrow = F) #never takers and treatment
       compliers sharing one random effect
    if(Share==2) lambda <- matrix(c(1,lambda1,0,0,0,1),nrow = 3,
       ncol = r_prime,byrow = F) #never takers and control
       compliers sharing one random effect
    if(Share==3) lambda <- matrix(c(1,0,0,0,1,lambda1),nrow = 3,
       ncol = r_prime,byrow = F) #control compliers and trt
       compliers sharing one random effect
  }
}

theta <- matrix(NA,niter,(r_prime*(r_prime+1)/2+l.alpha+l.gamma+k*(
   k+1)/2+nlambda))
#  'alpha_n','alpha_c0','alpha_c1',gamma,lambda1,lambda2,'tou11','
   tou12','tou13','tou22','tou23','tou33',delta)
theta[1,] <- init
```

```
if(r_prime==1){
  Vu <- theta[1,(l.alpha+l.gamma+nlambda+1):(r_prime*(r_prime+1)/2+
     l.alpha+l.gamma+nlambda)]
} else {
  Vu <- VechToCovM(theta[1,(l.alpha+l.gamma+nlambda+1):(r_prime*(r_
     prime+1)/2+l.alpha+l.gamma+nlambda)], r_prime)
}

if(k==1){
  Vb <- init[((length(init)-k*(k+1)/2+1):length(init))]
} else {
  Vb <- VechToCovM(init[((length(init)-k*(k+1)/2+1):length(init))],
      k)
}

Vub <- bdiag(Vu, Vb) %>% as.matrix
Vu.b <- do.call("rbind",replicate(J, MtoVech(Vub,(r_prime+k)),
   simplify = FALSE))

E_u <- do.call("rbind",replicate(J, rep(0,(r_prime+k)), simplify =
   FALSE))

E_u2 <- NA

X <- gaussHermiteData(Q)

#aghq matrix for u
A <- matrix(NA, Q^(r_prime+k),(r_prime+k))
```

```
for(i in 1:(r_prime+k)){
  A[,i] <- rep(X$x,each=Q^((r_prime+k)-i))
}

W <- matrix(NA, Q^(r_prime+k),(r_prime+k))
for(i in 1:(r_prime+k)){
  W[,i] <- rep(X$w,each=Q^((r_prime+k)-i))
}

Ws <- W*exp(A^2)
Ws <- apply(Ws,1,prod)

L2.norm <- c() #l2 norm between theta^k and theta^(k+1)
ll <- c() #likelihood

if(r_prime==1){
  M1 <- c(0,1,0)
  M2 <- c(0,0,1)
  M12 <- cbind(M1,M2)
}
if(r_prime==2){
  if(Share==1) M1 <- matrix(c(0,0,1,rep(0,3)),nrow=3,ncol=2,byrow =
      F)
  if(Share==2) M1 <- matrix(c(0,1,rep(0,4)),nrow=3,ncol=2,byrow = F
     )
  if(Share==3) M1 <- matrix(c(rep(0,5),1),nrow=3,ncol=2,byrow = F)
}
```

```
for(iter in 1:niter){
  Alpha <- theta[iter,1:l.alpha]
  gamma <- theta[iter,(l.alpha+1):(l.alpha+l.gamma)]

  if(r_prime==1){
    tau <- theta[iter,(length(c(Alpha,gamma))+nlambda+1):(r_prime*(
       r_prime+1)/2+l.alpha+l.gamma+nlambda)]
  } else {
    tau <- VechToCovM(theta[iter,(length(c(Alpha,gamma))+nlambda+1)
       :(r_prime*(r_prime+1)/2+l.alpha+l.gamma+nlambda)],r_prime)
  }

  if(k==1){
    delta <- theta[iter,((length(init)-k*(k+1)/2+1):length(init))]
  } else {
    delta <- VechToCovM(theta[iter,((length(init)-k*(k+1)/2+1):
       length(init))],k)
  }

  tau.delta <- bdiag(tau, delta) %>% as.matrix

  print(paste0("iteration␣=␣", iter))

  # Integration by AGHQ starts here
  list1 <- foreach(j = 1:J,.verbose=T,.errorhandling='stop',.export
     =ls(.GlobalEnv),.packages=c('mvtnorm')) %dopar% {
    #filter jth clinic data
```

```
data_j <- data[data$clinic==j,]

#change of variable
V.z <- VechToCovM(Vu.b[j,],(k+r_prime))

Lz <- t(chol(2*V.z))
z <- matrix(NA,Q^(r_prime+k),(r_prime+k))

Wphi <- numeric(length=nrow(z))
for(i in 1:nrow(z)){
  z[i,] <- Lz %*% A[i,]+E_u[j,]
  Wphi[i] <- Ws[i]*dmvnorm(z[i,], mean = rep(0, (r_prime+k)),
     sigma = tau.delta, log = FALSE)
}

#AGHQ
Lj <- det(Lz)*AGHQ(Q,Wphi,f_y_incomplete,z,r_prime,k,Alpha,
   gamma,data_j,ind.x0,ind.x1,lambda)
if(is.nan(Lj) | Lj==0) stop("lj is not positive")

E_u_j <- (det(Lz)/Lj) * AGHQ(Q,Wphi,f_y_z,z,r_prime,k,Alpha,
   gamma,data_j,ind.x0,ind.x1,lambda)
E_u2 <- (det(Lz)/Lj) * AGHQ(Q,Wphi,f_y_z2,z,r_prime,k,Alpha,
   gamma,data_j,ind.x0,ind.x1,lambda)

V <- E_u2 - E_u_j %*% t(E_u_j)

Vub_j <- MtoVech(V,(r_prime+k))
```

```
S_L_a <- det(Lz) * AGHQ(Q,Wphi,lnf_a_fy_1,z,r_prime,k,Alpha,
   gamma,data_j,ind.x0,ind.x1,lambda)
H_L_a <- det(Lz) * AGHQ(Q,Wphi,lnf_a_fy_2,z,r_prime,k,Alpha,
   gamma,data_j,ind.x0,ind.x1,lambda)
S_L_rc <- det(Lz) * AGHQ(Q,Wphi,lnf_rc_fy_1,z,r_prime,k,Alpha,
   gamma,data_j,ind.x0,ind.x1,lambda)
H_L_rc <- det(Lz) * AGHQ(Q,Wphi,lnf_rc_fy_2,z,r_prime,k,Alpha,
   gamma,data_j,ind.x0,ind.x1,lambda)
H_L_a_rc <- det(Lz) * AGHQ(Q,Wphi,lnf_rc_a_fy_2,z,r_prime,k,
   Alpha,gamma,data_j,ind.x0,ind.x1,lambda)

if(nlambda==1){
  S_L_lambda1 <- det(Lz) * AGHQ(Q,Wphi,lnf_lambda_fy_1,z,r_
     prime,k,Alpha,gamma,data_j,ind.x0,ind.x1,lambda,M1)
  H_L_lambda1 <- det(Lz) * AGHQ(Q,Wphi,lnf_lambda_fy_2,z,r_
     prime,k,Alpha,gamma,data_j,ind.x0,ind.x1,lambda,M1)
  H_L_a_lambda1 <- det(Lz) * AGHQ(Q,Wphi,lnf_a_lambda_fy_2,z,r_
     prime,k,Alpha,gamma,data_j,ind.x0,ind.x1,lambda,M1)
  H_L_rc_lambda1 <- det(Lz) * AGHQ(Q,Wphi,lnf_rc_lambda_fy_2,z,
     r_prime,k,Alpha,gamma,data_j,ind.x0,ind.x1,lambda,M1)
}

if(nlambda==2){
  S_L_lambda1 <- det(Lz) * AGHQ(Q,Wphi,lnf_lambda_fy_1,z,r_
     prime,k,Alpha,gamma,data_j,ind.x0,ind.x1,lambda,M1)
  S_L_lambda2 <- det(Lz) * AGHQ(Q,Wphi,lnf_lambda_fy_1,z,r_
     prime,k,Alpha,gamma,data_j,ind.x0,ind.x1,lambda,M2)
```

```
  H_L_lambda1 <- det(Lz) * AGHQ(Q,Wphi,lnf_lambda_fy_2,z,r_
     prime,k,Alpha,gamma,data_j,ind.x0,ind.x1,lambda,M1)
  H_L_lambda2 <- det(Lz) * AGHQ(Q,Wphi,lnf_lambda_fy_2,z,r_
     prime,k,Alpha,gamma,data_j,ind.x0,ind.x1,lambda,M2)
  H_L_a_lambda1 <- det(Lz) * AGHQ(Q,Wphi,lnf_a_lambda_fy_2,z,r_
     prime,k,Alpha,gamma,data_j,ind.x0,ind.x1,lambda,M1)
  H_L_a_lambda2 <- det(Lz) * AGHQ(Q,Wphi,lnf_a_lambda_fy_2,z,r_
     prime,k,Alpha,gamma,data_j,ind.x0,ind.x1,lambda,M2)
  H_L_rc_lambda1 <- det(Lz) * AGHQ(Q,Wphi,lnf_rc_lambda_fy_2,z,
     r_prime,k,Alpha,gamma,data_j,ind.x0,ind.x1,lambda,M1)
  H_L_rc_lambda2 <- det(Lz) * AGHQ(Q,Wphi,lnf_rc_lambda_fy_2,z,
     r_prime,k,Alpha,gamma,data_j,ind.x0,ind.x1,lambda,M2)
  H_L_lambda1_lambda2 <- det(Lz) * AGHQ(Q,Wphi,lnf_lambda1_
     lambda2_fy_2,z,r_prime,k,Alpha,gamma,data_j,ind.x0,ind.x1,
     lambda,M12)
}

logLj <- log(Lj)

Sa_j <- S_L_a/Lj
Src_j <- S_L_rc/Lj

Ha_j <- (H_L_a/Lj - S_L_a %*% t(S_L_a)/(Lj^2))
Hrc_j <-  (H_L_rc/Lj - S_L_rc %*% t(S_L_rc)/(Lj^2))
Harc_j <- (H_L_a_rc/Lj - S_L_rc %*% t(S_L_a)/(Lj^2))

Slambda1_j <- S_L_lambda1/Lj
```

```
    Hlambda1_j <- (H_L_lambda1/Lj - S_L_lambda1 %*% t(S_L_lambda1)/
       (Lj^2))
    Halambda1_j <- (H_L_a_lambda1/Lj - S_L_lambda1 %*% t(S_L_a)/(Lj
       ^2))
    Hrclambda1_j <- (H_L_rc_lambda1/Lj - S_L_lambda1 %*% t(S_L_rc)/
       (Lj^2))

    Slambda2_j <-0
    Hlambda2_j <-0
    Halambda2_j <-0
    Hrclambda2_j <-0
    Hlambda1_2_j <-0

    if(nlambda==2){
      Slambda2_j <- S_L_lambda2/Lj
      Hlambda2_j <- (H_L_lambda2/Lj - S_L_lambda2 %*% t(S_L_lambda2
         )/(Lj^2))
      Halambda2_j <- (H_L_a_lambda2/Lj - S_L_lambda2 %*% t(S_L_a)/(
         Lj^2))
      Hrclambda2_j <- (H_L_rc_lambda2/Lj - S_L_lambda2 %*% t(S_L_rc
         )/(Lj^2))
      Hlambda1_2_j <- (H_L_lambda1_lambda2/Lj - S_L_lambda2 %*% t(S
         _L_lambda1)/(Lj^2))
}

    list2 <- list(E_u_j=E_u_j,E_u2=E_u2,Vub_j=Vub_j, logLj=logLj,
       Sa_j=Sa_j, Src_j=Src_j, Ha_j=Ha_j, Hrc_j=Hrc_j, Harc_j=Harc_
       j,
```

```
                Slambda1_j=Slambda1_j, Hlambda1_j=Hlambda1_j,
                   Halambda1_j=Halambda1_j, Hrclambda1_j=
                   Hrclambda1_j,
                Slambda2_j=Slambda2_j, Hlambda2_j=Hlambda2_j,
                   Halambda2_j=Halambda2_j, Hrclambda2_j=
                   Hrclambda2_j, Hlambda1_2_j=Hlambda1_2_j)
}#end of loop for clinic

Sa <- 0
Src <- 0

Ha <- 0
Harc <- 0
Hrc <- 0

Slambda1 <- 0
Slambda2 <- 0
Hlambda1 <- 0
Hlambda2 <- 0
Halambda1 <- 0
Halambda2 <- 0
Hrclambda1 <- Hrclambda2 <- 0
Hlambda1_2 <- 0

logL <- 0
tau.del.hat <-0

for(j in 1:J) {
```

```
  tau.del.hat <- tau.del.hat+list1[[j]]$E_u2
  logL <- logL+list1[[j]]$logLj
  Sa <- Sa+list1[[j]]$Sa_j
  Src <- Src+list1[[j]]$Src_j
  Ha <- Ha+list1[[j]]$Ha_j
  Hrc <- Hrc+list1[[j]]$Hrc_j
  Harc <- Harc+list1[[j]]$Harc_j
  Slambda1 <- Slambda1+list1[[j]]$Slambda1_j
  Slambda2 <- Slambda2+list1[[j]]$Slambda2_j
  Hlambda1 <- Hlambda1+list1[[j]]$Hlambda1_j
  Hlambda2 <- Hlambda2+list1[[j]]$Hlambda2_j
  Halambda1 <- Halambda1+list1[[j]]$Halambda1_j
  Halambda2 <- Halambda2+list1[[j]]$Halambda2_j
  Hrclambda1 <- Hrclambda1+list1[[j]]$Hrclambda1_j
  Hrclambda2 <- Hrclambda2+list1[[j]]$Hrclambda2_j
  Hlambda1_2 <- Hlambda1_2+list1[[j]]$Hlambda1_2_j
}

E_u <- list1[[1]]$E_u_j
Vu.b <- list1[[1]]$Vub_j

for(j in 2:J){
  E_u <- rbind(E_u, list1[[j]]$E_u_j)
  Vu.b <- rbind(Vu.b, list1[[j]]$Vub_j)
}

tau.d.h <- tau.del.hat/J
tau.h <- tau.d.h[1:r_prime,1:r_prime]
```

```
d.h <- tau.d.h[(nrow(tau.d.h)-k+1):nrow(tau.d.h),(nrow(tau.d.h)-k
   +1):nrow(tau.d.h)]
#log likelihood of J clinics
print(paste0("loglike = ", logL))

if(nlambda==1){
  S <- c(Sa,Src,Slambda1)

  H1 <- cbind(Ha,t(Harc),t(Halambda1))
  H2 <- cbind(Harc,Hrc,t(Hrclambda1))
  H3 <- cbind(Halambda1,Hrclambda1,Hlambda1)
  H <- rbind(H1,H2,H3)
}

if(nlambda==2){
  S <- c(Sa,Src,Slambda1,Slambda2)

  H1 <- cbind(Ha,t(Harc),t(Halambda1),t(Halambda2))
  H2 <- cbind(Harc,Hrc,t(Hrclambda1),t(Hrclambda2))
  H3 <- cbind(Halambda1,Hrclambda1,Hlambda1,Hlambda1_2)
  H4 <- cbind(Halambda2,Hrclambda2,Hlambda1_2,Hlambda2)
  H <- rbind(H1,H2,H3,H4)
}

if(!is.symmetric.matrix(H)) {
  print(paste0("iter= ",iter))
  print(H)
  print("H asymmetric")
```

```
}

Hinv <- solve(H)

theta[(iter+1),1:(l.alpha+l.gamma+nlambda)] <- theta[iter,1:(l.
   alpha+l.gamma+nlambda)] - as.vector(Hinv %*% S)
lambda1 <- theta[(iter+1),(l.alpha+l.gamma+1)]
if(nlambda==2){
  lambda2 <- theta[(iter+1),(l.alpha+l.gamma+2)]
  if(Share==1) lambda <- c(1,lambda1,lambda2)
}
if(nlambda==1){
  if(Share==1)  lambda <- matrix(c(1,0,lambda1,0,1,0),nrow = 3,
     ncol = r_prime,byrow = F)
  if(Share==2)  lambda <- matrix(c(1,lambda1,rep(0,3),1),nrow =
     3,ncol = r_prime, byrow = F)
  if(Share==3)  lambda <- matrix(c(1,rep(0,3),1,lambda1),nrow =
     3, ncol = r_prime, byrow = F)
}

if(r_prime==1){
  theta[(iter+1),(l.alpha+l.gamma+1+nlambda):(r_prime*(r_prime+1)
     /2+l.alpha+l.gamma+nlambda)] <- tau.h
} else {
  theta[(iter+1),(l.alpha+l.gamma+1+nlambda):(r_prime*(r_prime+1)
     /2+l.alpha+l.gamma+nlambda)] <- MtoVech(tau.h,r_prime)
}
if(k==1){
```

```
    theta[(iter+1),(l.alpha+l.gamma+r_prime*(r_prime+1)/2+nlambda
       +1):(r_prime*(r_prime+1)/2+l.alpha+l.gamma+k*(k+1)/2+nlambda
       )] <- d.h
  } else {
    theta[(iter+1),(l.alpha+l.gamma+r_prime*(r_prime+1)/2+nlambda
       +1):(r_prime*(r_prime+1)/2+l.alpha+l.gamma+k*(k+1)/2+nlambda
       )] <- MtoVech(d.h,k)
  }

  print(paste0("theta= ",theta[(iter+1),]))

  if(iter>1){
    L2.norm[iter] <- sqrt(sum((theta[(iter),]-theta[(iter-1),])^2))
    print(paste0("L2.norm= ",L2.norm[iter]))
    if(L2.norm[iter] < tol) {break}
  }

  ll[iter] <- logL

} #end of iter loop

para0 <- theta[(iter+1),]

if(r_prime==1){
  Vu <- para0[(l.alpha+l.gamma+nlambda+1):(r_prime*(r_prime+1)/2+l.
     alpha+l.gamma+nlambda)]
} else {
```

```
  Vu <- VechToCovM(para0[(l.alpha+l.gamma+nlambda+1):(r_prime*(r_
     prime+1)/2+l.alpha+l.gamma+nlambda)], r_prime)
}

if(k==1){
  Vb <- para0[((length(para0)-k*(k+1)/2+1):length(para0))]
} else {
  Vb <- VechToCovM(para0[((length(para0)-k*(k+1)/2+1):length(para0)
     )], k)
}

Vub <- bdiag(Vu, Vb) %>% as.matrix
Vu.b <- do.call("rbind",replicate(J, MtoVech(Vub,(r_prime+k)),
   simplify = FALSE))

E_u <- do.call("rbind",replicate(J, rep(0,(r_prime+k)), simplify =
   FALSE))

E_u2 <- NA

Alpha <- para0[1:l.alpha]
gamma <- para0[(l.alpha+1):(l.alpha+l.gamma)]

if(r_prime==1){
  tau <- para0[(l.alpha+l.gamma+nlambda+1):(r_prime*(r_prime+1)/2+l
     .alpha+l.gamma+nlambda)]
} else {
```

```
  tau <- VechToCovM(para0[(l.alpha+l.gamma+nlambda+1):(r_prime*(r_
     prime+1)/2+l.alpha+l.gamma+nlambda)], r_prime)
}

if(k==1){
  delta <- para0[((length(para0)-k*(k+1)/2+1):length(para0))]
} else {
  delta <- VechToCovM(para0[((length(para0)-k*(k+1)/2+1):length(
     para0))], k)
}

tau.delta <- bdiag(tau, delta) %>% as.matrix

#for computing H matrix only
list1 <- foreach(j = 1:J,.verbose=T,.errorhandling='stop',.export=
   ls(.GlobalEnv),.packages=c('mvtnorm')) %dopar% {
  #filter jth clinic data
  data_j <- data[data$clinic==j,]

  #change of variable
  V.z <- VechToCovM(Vu.b[j,],(k+r_prime))

  Lz <- t(chol(2*V.z))
  z <- matrix(NA,Q^(r_prime+k),(r_prime+k))

  Wphi <- numeric(length=nrow(z))
  for(i in 1:nrow(z)){
    z[i,] <- Lz %*% A[i,]+E_u[j,]
```

```
  Wphi[i] <- Ws[i]*dmvnorm(z[i,], mean = rep(0, (r_prime+k)),
     sigma = tau.delta, log = FALSE)
}

#AGHQ
Lj <- det(Lz)*AGHQ(Q,Wphi,f_y_incomplete,z,r_prime,k,Alpha,gamma,
   data_j,ind.x0,ind.x1,lambda)
if(is.nan(Lj) | Lj==0) stop("lj is not positive")

E_u_j <- (det(Lz)/Lj) * AGHQ(Q,Wphi,f_y_z,z,r_prime,k,Alpha,gamma
   ,data_j,ind.x0,ind.x1,lambda)
E_u2 <- (det(Lz)/Lj) * AGHQ(Q,Wphi,f_y_z2,z,r_prime,k,Alpha,gamma
   ,data_j,ind.x0,ind.x1,lambda)

S_L_a <- det(Lz) * AGHQ(Q,Wphi,lnf_a_fy_1,z,r_prime,k,Alpha,gamma
   ,data_j,ind.x0,ind.x1,lambda)
H_L_a <- det(Lz) * AGHQ(Q,Wphi,lnf_a_fy_2,z,r_prime,k,Alpha,gamma
   ,data_j,ind.x0,ind.x1,lambda)
S_L_rc <- det(Lz) * AGHQ(Q,Wphi,lnf_rc_fy_1,z,r_prime,k,Alpha,
   gamma,data_j,ind.x0,ind.x1,lambda)
H_L_rc <- det(Lz) * AGHQ(Q,Wphi,lnf_rc_fy_2,z,r_prime,k,Alpha,
   gamma,data_j,ind.x0,ind.x1,lambda)
H_L_a_rc <- det(Lz) * AGHQ(Q,Wphi,lnf_rc_a_fy_2,z,r_prime,k,Alpha
   ,gamma,data_j,ind.x0,ind.x1,lambda)

S_L_t <- det(Lz) * AGHQ(Q,Wphi,lnf_t_fy_1,z,r_prime,k,Alpha,gamma
   ,data_j,ind.x0,ind.x1,lambda,tau=tau) #r(r+1)/2 x 1
```

```
H_L_t <- det(Lz) * AGHQ(Q,Wphi,lnf_t_fy_2,z,r_prime,k,Alpha,gamma
   ,data_j,ind.x0,ind.x1,lambda,tau=tau)
H_L_at <- det(Lz) * AGHQ(Q,Wphi,lnf_t_a_fy_2,z,r_prime,k,Alpha,
   gamma,data_j,ind.x0,ind.x1,lambda,tau=tau)
S_L_d <- det(Lz) * AGHQ(Q,Wphi,lnf_delta_fy_1,z,r_prime,k,Alpha,
   gamma,data_j,ind.x0,ind.x1,lambda,delta = delta)
H_L_d <- det(Lz) * AGHQ(Q,Wphi,lnf_delta_fy_2,z,r_prime,k,Alpha,
   gamma,data_j,ind.x0,ind.x1,lambda,delta = delta)
H_L_ad <- det(Lz) * AGHQ(Q,Wphi,lnf_delta_a_fy_2,z,r_prime,k,
   Alpha,gamma,data_j,ind.x0,ind.x1,lambda, delta=delta)
H_L_rc_t <- det(Lz) * AGHQ(Q,Wphi,lnf_t_rc_fy_2,z,r_prime,k,Alpha
   ,gamma,data_j,ind.x0,ind.x1,lambda,tau = tau)
H_L_rc_d <- det(Lz) * AGHQ(Q,Wphi,lnf_delta_rc_fy_2,z,r_prime,k,
   Alpha,gamma,data_j,ind.x0,ind.x1,lambda,delta = delta)
H_L_t_d <- det(Lz) * AGHQ(Q,Wphi,lnf_delta_t_fy_2,z,r_prime,k,
   Alpha,gamma,data_j,ind.x0,ind.x1,lambda,tau = tau,delta =
   delta)

S_L_lambda1 <- det(Lz) * AGHQ(Q,Wphi,lnf_lambda_fy_1,z,r_prime,k,
   Alpha,gamma,data_j,ind.x0,ind.x1,lambda,M1)
H_L_lambda1 <- det(Lz) * AGHQ(Q,Wphi,lnf_lambda_fy_2,z,r_prime,k,
   Alpha,gamma,data_j,ind.x0,ind.x1,lambda,M1)
H_L_a_lambda1 <- det(Lz) * AGHQ(Q,Wphi,lnf_a_lambda_fy_2,z,r_
   prime,k,Alpha,gamma,data_j,ind.x0,ind.x1,lambda,M1)
H_L_rc_lambda1 <- det(Lz) * AGHQ(Q,Wphi,lnf_rc_lambda_fy_2,z,r_
   prime,k,Alpha,gamma,data_j,ind.x0,ind.x1,lambda,M1)
H_L_t_lambda1 <- det(Lz) * AGHQ(Q,Wphi,lnf_t_lambda_fy_2,z,r_
   prime,k,Alpha,gamma,data_j,ind.x0,ind.x1,lambda, tau = tau, M1
```

```
    )
 H_L_d_lambda1 <- det(Lz) * AGHQ(Q,Wphi,lnf_delta_lambda_fy_2,z,r_
    prime,k,Alpha,gamma,data_j,ind.x0,ind.x1,lambda, delta, M1)

 if(nlambda==2){
   S_L_lambda2 <- det(Lz) * AGHQ(Q,Wphi,lnf_lambda_fy_1,z,r_prime,
      k,Alpha,gamma,data_j,ind.x0,ind.x1,lambda,M2)
   H_L_lambda2 <- det(Lz) * AGHQ(Q,Wphi,lnf_lambda_fy_2,z,r_prime,
      k,Alpha,gamma,data_j,ind.x0,ind.x1,lambda,M2)
   H_L_a_lambda2 <- det(Lz) * AGHQ(Q,Wphi,lnf_a_lambda_fy_2,z,r_
      prime,k,Alpha,gamma,data_j,ind.x0,ind.x1,lambda,M2)
   H_L_rc_lambda2 <- det(Lz) * AGHQ(Q,Wphi,lnf_rc_lambda_fy_2,z,r_
      prime,k,Alpha,gamma,data_j,ind.x0,ind.x1,lambda,M2)
   H_L_lambda1_lambda2 <- det(Lz) * AGHQ(Q,Wphi,lnf_lambda1_
      lambda2_fy_2,z,r_prime,k,Alpha,gamma,data_j,ind.x0,ind.x1,
      lambda,M12)
   H_L_t_lambda2 <- det(Lz) * AGHQ(Q,Wphi,lnf_t_lambda_fy_2,z,r_
      prime,k,Alpha,gamma,data_j,ind.x0,ind.x1,lambda, tau = tau,
      M2)
   H_L_d_lambda2 <- det(Lz) * AGHQ(Q,Wphi,lnf_delta_lambda_fy_2,z,
      r_prime,k,Alpha,gamma,data_j,ind.x0,ind.x1,lambda, delta, M2
      )
}

 logLj <- log(Lj)

 Ha_j <- (H_L_a/Lj - S_L_a %*% t(S_L_a)/(Lj^2))
 Hrc_j <-  (H_L_rc/Lj - S_L_rc %*% t(S_L_rc)/(Lj^2))
```

```
Harc_j <- (H_L_a_rc/Lj - S_L_rc %*% t(S_L_a)/(Lj^2))

Ht_j <- (H_L_t/Lj - S_L_t %*% t(S_L_t)/(Lj^2))
Hat_j <- (H_L_at/Lj - S_L_t %*% S_L_a/(Lj^2))
Hrct_j <- (H_L_rc_t/Lj - t(S_L_rc %*% t(S_L_t))/(Lj^2))

Hd_j <- (H_L_d/Lj - S_L_d %*% t(S_L_d)/(Lj^2))
Had_j <- (H_L_ad/Lj - S_L_d %*% S_L_a/(Lj^2))
Htd_j <- (H_L_t_d/Lj - t(S_L_t %*% t(S_L_d))/(Lj^2))
Hrcd_j <- (H_L_rc_d/Lj - t(S_L_rc %*% t(S_L_d))/(Lj^2))

Hlambda1_j <- (H_L_lambda1/Lj - S_L_lambda1 %*% t(S_L_lambda1)/(
   Lj^2))
Halambda1_j <- (H_L_a_lambda1/Lj - S_L_lambda1 %*% t(S_L_a)/(Lj
   ^2))
Hrclambda1_j <- (H_L_rc_lambda1/Lj - S_L_lambda1 %*% t(S_L_rc)/(
   Lj^2))
Htlambda1_j <- (H_L_t_lambda1/Lj - S_L_t %*% S_L_lambda1/(Lj^2))
Hdlambda1_j <- (H_L_d_lambda1/Lj - S_L_d %*% S_L_lambda1/(Lj^2))

Hlambda2_j <-0
Halambda2_j <-0
Hrclambda2_j <-0
Hlambda1_2_j <-0
Htlambda2_j <-0
Hdlambda2_j <-0

if(nlambda==2){
```

```
  #Slambda2_j <- S_L_lambda2/Lj
  Hlambda2_j <- (H_L_lambda2/Lj - S_L_lambda2 %*% t(S_L_lambda2)/
     (Lj^2))
  Halambda2_j <- (H_L_a_lambda2/Lj - S_L_lambda2 %*% t(S_L_a)/(Lj
     ^2))
  Hrclambda2_j <- (H_L_rc_lambda2/Lj - S_L_lambda2 %*% t(S_L_rc)/
     (Lj^2))
  Hlambda1_2_j <- (H_L_lambda1_lambda2/Lj - S_L_lambda2 %*% t(S_L
     _lambda1)/(Lj^2))
  Htlambda2_j <- (H_L_t_lambda2/Lj - S_L_t %*% S_L_lambda2/(Lj^2)
     )
  Hdlambda2_j <- (H_L_d_lambda2/Lj - S_L_d %*% S_L_lambda2/(Lj^2)
     )

}

  list2 <- list(E_u_j=E_u_j,E_u2=E_u2, logLj=logLj,
                Ha_j=Ha_j, Hrc_j=Hrc_j, Harc_j=Harc_j,
                Ht_j=Ht_j, Hat_j=Hat_j, Hrct_j=Hrct_j,
                Hd_j=Hd_j, Had_j=Had_j, Hrcd_j=Hrcd_j, Htd_j=Htd_
                   j,
                Hlambda1_j=Hlambda1_j, Halambda1_j=Halambda1_j,
                   Hrclambda1_j=Hrclambda1_j,
                Htlambda1_j=Htlambda1_j,Hdlambda1_j=Hdlambda1_j,
                Hlambda2_j=Hlambda2_j, Halambda2_j=Halambda2_j,
                   Hrclambda2_j=Hrclambda2_j, Hlambda1_2_j=
                   Hlambda1_2_j,
                Htlambda2_j=Htlambda2_j,Hdlambda2_j=Hdlambda2_j)
```

```
}#end of loop for clinic
logL <- 0
tau.del.hat <-0

Ha <- 0
Harc <- 0
Hrc <- 0
Hat <- 0
Ht <- 0
Had <- 0
Hrct <- 0
Hrcd <- 0
Hd <- 0
Htd <- 0
Hlambda1 <- 0
Hlambda2 <- 0
Halambda1 <- 0
Halambda2 <- 0
Hrclambda1 <- Hrclambda2 <- 0
Hlambda1_2 <- 0
Htlambda1 <- Htlambda2 <- Hdlambda1 <- Hdlambda2 <- 0

for(j in 1:J) {
  tau.del.hat <- tau.del.hat+list1[[j]]$E_u2
  logL <- logL+list1[[j]]$logLj

  Ha <- Ha+list1[[j]]$Ha_j
```

```
    Hrc <- Hrc+list1[[j]]$Hrc_j
    Harc <- Harc+list1[[j]]$Harc_j

    Ht <- Ht + list1[[j]]$Ht_j
    Hat <- Hat + list1[[j]]$Hat_j
    Had <- Had + list1[[j]]$Had_j
    Hrct <- Hrct + list1[[j]]$Hrct_j
    Hrcd <- Hrcd + list1[[j]]$Hrcd_j
    Hd <- Hd + list1[[j]]$Hd_j
    Htd <- Htd + list1[[j]]$Htd_j

    Hlambda1 <- Hlambda1+list1[[j]]$Hlambda1_j
    Hlambda2 <- Hlambda2+list1[[j]]$Hlambda2_j
    Halambda1 <- Halambda1+list1[[j]]$Halambda1_j
    Halambda2 <- Halambda2+list1[[j]]$Halambda2_j
    Hrclambda1 <- Hrclambda1+list1[[j]]$Hrclambda1_j
    Hrclambda2 <- Hrclambda2+list1[[j]]$Hrclambda2_j
    Hlambda1_2 <- Hlambda1_2+list1[[j]]$Hlambda1_2_j

    Htlambda1 <- Htlambda1+list1[[j]]$Htlambda1_j
    Htlambda2 <- Htlambda2+list1[[j]]$Htlambda2_j
    Hdlambda1 <- Hdlambda1+list1[[j]]$Hdlambda1_j
    Hdlambda2 <- Hdlambda2+list1[[j]]$Hdlambda2_j
}

 E_u <- list1[[1]]$E_u_j

 for(j in 2:J){
```

```
  E_u <- rbind(E_u, list1[[j]]$E_u_j)
}

print(paste0("loglike␣=␣", logL))
if(nlambda==1){
  H1 <- cbind(Ha,t(Harc),t(Halambda1),t(Hat),t(Had))
  H2 <- cbind(Harc,Hrc,t(Hrclambda1),t(Hrct),t(Hrcd))
  H3 <- cbind(Halambda1,Hrclambda1,Hlambda1,t(Htlambda1),t(
     Hdlambda1))
  H4 <- cbind(Hat,Hrct,Htlambda1,Ht,t(Htd))
  H5 <- cbind(Had,Hrcd,Hdlambda1,Htd,Hd)
  H <- rbind(H1,H2,H3,H4,H5)
}

if(nlambda==2){
  H1 <- cbind(Ha,t(Harc),t(Halambda1),t(Halambda2),t(Hat),t(Had))
  H2 <- cbind(Harc,Hrc,t(Hrclambda1),t(Hrclambda2),t(Hrct),t(Hrcd))
  H3 <- cbind(Halambda1,Hrclambda1,Hlambda1,t(Hlambda1_2),t(
     Htlambda1),t(Hdlambda1))
  H4 <- cbind(Halambda2,Hrclambda2,Hlambda1_2,Hlambda2, t(Htlambda2
     ), t(Hdlambda2))
  H5 <- cbind(Hat,Hrct,Htlambda1,Htlambda2, Ht,t(Htd))
  H6 <- cbind(Had,Hrcd,Hdlambda1,Hdlambda2,Htd,Hd)
  H <- rbind(H1,H2,H3,H4,H5,H6)
}

print(H)
print(paste0("symmetric␣H?␣",isSymmetric(H, tol=1e-14)))
```

```
se<-sqrt(diag(solve(-H)))

pval <- c()
for(i in 1:length(para0)){
  if(para0[i]>=0 & !is.nan(se[i])){
    pval[i] <- pnorm((para0[i]-0)/se[i],lower.tail = F) #para0 here
        is the converged estimates
  }
  if(para0[i]<0 & !is.nan(se[i])){
    pval[i] <- pnorm((para0[i]-0)/se[i],lower.tail = T)
  }
}

#ac1-ac0
cace <- para0[3]-para0[2]
covc0c1 <- -solve(H)[2,3]
cace.se <- sqrt(se[3]^2+se[2]^2-2*covc0c1)

if(para0[3]-para0[2]>=0){
  cace.pval <- pnorm((para0[3]-para0[2])/cace.se,lower.tail = F)
} else {
  cace.pval <- pnorm((para0[3]-para0[2])/cace.se,lower.tail = T)
}

end_time <- Sys.time()

time <- end_time-start_time
print(time)
```

```
  return(list(par=para0,se=se,pval=pval,logL=logL,L2.norm=L2.norm[
     iter],time=time,cace=cace,cace.se=cace.se,cace.pval=cace.pval,
     niter=iter,E_u=E_u,H=H))
} #end of function

#The function returns par=estimates of parameters, se=standard errors
    of estimates, pval=p-values of wald test for estimates, logL=
   loglikelihood
#L2.norm=L2 norm, time=computation time, cace=mean cace, cace.se=
   standard error of CACE, cace.pval=p-value of wald test for cace,
   niter=# iterations,
#E_u= site-level random effects of reduced dimension, H=Hessian
   matrix


#AGHQ() approximates an integral with a sum of weighted Q abscissas
#inputs:
#Q: number of abscissas
#Wphi: weights * probabilities of a vector of z (abscissas after
   change of variables) that follow multivariate standard normals
   distribution
#f: g()f(Y,C), g() is the function to be approximated, f(Y,C) is the
   joint distribution of outcome Y and compliance
#z: a vector or a matrix of abscissas after change of variable
#r: reduced dimension of random effects of Y model
#k: reduced dimension of random effects of C model. it is 1 for one-
   sided noncompliance
#Alpha: alpha vector in Y model
```

```
#gamma: gamma vector in C model
#data_j: data frame of jth site
#ind.x0: column number of covariates controlled in Y model
#ind.x1: column number of covariates controlled in C model
#lambda: factor loading matrix
#tau: variance (matrix) of random effects of Y model
#delta: varianceof random effect of C model
#M: derivative of factor loading matrix

#Output: approximation of an integral

AGHQ <- function(Q,Wphi,f,z,r,k,Alpha,gamma,data_j,ind.x0,ind.x1,
   lambda,tau,delta,M){ #f:function. add z as an additional arguement
  p <- Q^(r+k)
  L=0
  if(missing(M)){
    if(missing(tau) & missing(delta)){
      for(q in 1:p){
        L <- L + f(z[q,1:r],z[q,(r+1):(r+k)],Alpha,gamma,data_j,ind.
            x0,ind.x1,lambda)*Wphi[q]
      }
    }
    if (missing(tau) & !missing(delta)) {
      for(q in 1:p){
        L <- L + f(z[q,1:r],z[q,(r+1):(r+k)],Alpha,gamma,data_j,ind.
            x0,ind.x1,lambda,delta)*Wphi[q]
      }
    }
```

```
  if (!missing(tau) & missing(delta)) {
    for(q in 1:p){
      L <- L + f(z[q,1:r],z[q,(r+1):(r+k)],Alpha,gamma,data_j,ind.
          x0,ind.x1,lambda,tau)*Wphi[q]
    }
  }
  if (!missing(tau) & !missing(delta)) {
    for(q in 1:p){
      L <- L + f(z[q,1:r],z[q,(r+1):(r+k)],Alpha,gamma,data_j,ind.
          x0,ind.x1,lambda,tau,delta)*Wphi[q]
    }
  }
} else {#end of missing(M)
  if (missing(tau) & missing(delta)) {
    for(q in 1:p){
      L <- L + f(z[q,1:r],z[q,(r+1):(r+k)],Alpha,gamma,data_j,ind.
          x0,ind.x1,lambda,M)*Wphi[q]
    }
  }
  if (!missing(tau) & missing(delta)) {
    for(q in 1:p){
      L <- L + f(z[q,1:r],z[q,(r+1):(r+k)],Alpha,gamma,data_j,ind.
          x0,ind.x1,lambda,tau,M)*Wphi[q]
    }
  }
  if (missing(tau) & !missing(delta)) {
    for(q in 1:p){
```

```
        L <- L + f(z[q,1:r],z[q,(r+1):(r+k)],Alpha,gamma,data_j,ind.
            x0,ind.x1,lambda,delta,M)*Wphi[q]
      }
    }
  } #end of !missing(M)
  return(L)
}

#f_y_incomplete() calculates the joint pmf of Y and C.
#inputs:
#u: random effects of y model. usually the vector z of length r'
   after change of variable
#b: random effect of c model. usually the last element of vector z
#Alpha: alpha vector in Y model
#gamma: gamma vector in C model
#data_j: data frame of jth site
#ind.x0: column number of covariates controlled in Y model
#ind.x1: column number of covariates controlled in C model
#lambda: factor loading matrix

f_y_incomplete <- function(u,b,Alpha,gamma,data_j,ind.x0,ind.x1,
   lambda){
   data_j <- data_j[!is.na(data_j$Y) | !is.na(data_j$C),] #remove obs
       with both missing C and Y

   n.j <- nrow(data_j)

   if(anyNA(ind.x1)){
```

```
  Xij <- as.matrix(rep(1,n.j)) #intercept only for compliance
     model
} else {
  Xij <- as.matrix(cbind(rep(1,n.j), data_j[,ind.x1])) #intercept
     and covariates
}

Bij <- diag(rep(1,3))
Bij <- Bij %*% lambda

data_j$f_ij <- 0

for(i in 1:n.j){
  if(is.na(data_j$Y[i])){
    eta_c <- as.numeric(t(as.matrix(gamma)) %*% as.matrix(Xij[i,])
        + b) #compliance model. eta_c
    Pc <- 1/(1+exp(-eta_c))
    Pn <- 1-Pc
    if(data_j$trt[i]==1 & data_j$D[i]==0) data_j$f_ij[i] <- Pn
    if(data_j$trt[i]==1 & data_j$D[i]==1) data_j$f_ij[i] <- Pc
  } else {
    if(anyNA(ind.x0)){
      Aij <- diag(rep(1,3))
    } else {
      Aij <- cbind(diag(rep(1,3)),matrix(unlist(replicate(3,as.
         matrix(data_j[,ind.x0])[i,])),nrow = 3,byrow = T))
    }
```

```
    eta_y <- Aij %*% as.matrix(Alpha) + Bij %*% u
    eta_c <- as.numeric(t(as.matrix(gamma)) %*% as.matrix(Xij[i,])
        + b)

    Pc <- 1/(1+exp(-eta_c))
    Pn <- 1-Pc

    if(data_j$trt[i]==1 & data_j$D[i]==0){ #never taker
      data_j$f_ij[i] <- exp(data_j$Y[i]*eta_y[1]-log(1+exp(eta_y
          [1])))*Pn
    }
    if(data_j$trt[i]==0 & data_j$D[i]==0){ #control complier+never
        taker
      data_j$f_ij[i] <- exp(data_j$Y[i]*eta_y[1]-log(1+exp(eta_y
          [1])))*Pn + exp(data_j$Y[i]*eta_y[2]-log(1+exp(eta_y[2]))
          )*Pc
    }
    if(data_j$trt[i]==1 & data_j$D[i]==1){ #trt complier
      data_j$f_ij[i] <- exp(data_j$Y[i]*eta_y[3]-log(1+exp(eta_y
          [3])))*Pc
    }
  }#end of ifelse
}#end of loop
if(0 %in% data_j$f_ij){
  f_ij <- data_j$f_ij[-which(data_j$f_ij %in% 0)]
} else {
  f_ij <- data_j$f_ij
}
```

```
    f_y_j <- prod(f_ij)

   return(f_y_j)
 }

#fun_der_lnf_alpha() gives the first derivative of alpha
#inputs:
#u: random effects of y model. usually the vector z of length r'
   after change of variable
#b: random effect of c model. usually the last element of vector z
#Alpha: alpha vector in Y model
#gamma: gamma vector in C model
#data_j: data frame of jth site
#ind.x0: column number of covariates controlled in Y model
#ind.x1: column number of covariates controlled in C model
#lambda: factor loading matrix

fun_der_lnf_alpha <- function(u,b,Alpha,gamma,data_j,ind.x0,ind.x1,
   lambda){
  data_j <- data_j[!is.na(data_j$Y),]
  n.j <- nrow(data_j)

  if(anyNA(ind.x1)){
    Xij <- as.matrix(rep(1,n.j)) #intercept only for compliance model
  } else {
    Xij <- as.matrix(cbind(rep(1,n.j),data_j[,ind.x1])) #intercept
       and covariates
```

```
}

Bij <- diag(rep(1,3))
Bij <- Bij %*% lambda

first.deriv <- list()

for(i in 1:n.j){
  if(anyNA(ind.x0)){
    Aij <- diag(rep(1,3))
  } else {
    Aij <- cbind(diag(rep(1,3)),matrix(unlist(replicate(3,as.matrix
       (data_j[,ind.x0])[i,])),nrow = 3,byrow = T))
  }

  eta_y <- Aij %*% Alpha + Bij %*% u
  Py <- 1/(1+exp(-eta_y))
  eta_c <- as.numeric(t(as.matrix(gamma)) %*% as.matrix(Xij[i,]) +
     b)

  if(data_j$D[i]==0 & data_j$trt[i]==1){ #never taker
    first.deriv[[i]] <- (data_j$Y[i]-Py[1]) * Aij[1,]
  }

  if(data_j$D[i]==0 & data_j$trt[i]==0){ #control complier
    first.deriv[[i]] <- (exp(data_j$Y[i]*eta_y[1]-log(1+exp(eta_y
       [1])))*(data_j$Y[i]-Py[1])*Aij[1,] +
```

```
                        exp(data_j$Y[i]*eta_y[2]-log(1+exp(eta_y
                           [2]))+eta_c)*(data_j$Y[i]-Py[2])*Aij
                           [2,]) /
                        (exp(data_j$Y[i]*eta_y[1]-log(1+exp(eta_y
                           [1])))+exp(data_j$Y[i]*eta_y[2]-log(1+
                           exp(eta_y[2]))+eta_c))
    }
    if(data_j$D[i]==1 & data_j$trt[i]==1){ #trt complier
      first.deriv[[i]] <- (data_j$Y[i]-Py[3]) * Aij[3,]
    }
  }#end of loop

  first_der_lj_alpha <- Reduce('+', first.deriv)

  return(first_der_lj_alpha)
}

#fun_2nd_deriv_lnf_a() gives the second derivative of alpha
#inputs:
#u: random effects of y model. usually the vector z of length r'
   after change of variable
#b: random effect of c model. usually the last element of vector z
#Alpha: alpha vector in Y model
#gamma: gamma vector in C model
#data_j: data frame of jth site
#ind.x0: column number of covariates controlled in Y model
#ind.x1: column number of covariates controlled in C model
#lambda: factor loading matrix
```

```
fun_2nd_deriv_lnf_a <- function(u,b,Alpha,gamma,data_j,ind.x0,ind.x1,
   lambda){
  data_j <- data_j[!is.na(data_j$Y),]
  n.j <- nrow(data_j)

  if(anyNA(ind.x1)){
    Xij <- as.matrix(rep(1,n.j)) #intercept only for compliance model
  } else {
    Xij <- as.matrix(cbind(rep(1,n.j),data_j[,ind.x1])) #intercept
       and covariates
  }

  Bij <- diag(rep(1,3))
  Bij <- Bij %*% lambda

  deriv_2nd_lnf_a <- list()

  for(i in 1:n.j){
    if(anyNA(ind.x0)){
      Aij <- diag(rep(1,3))
    } else {
      Aij <- cbind(diag(rep(1,3)),matrix(unlist(replicate(3,as.matrix
         (data_j[,ind.x0])[i,])),nrow = 3,byrow = T))
    }

    eta_y <- Aij %*% Alpha + Bij %*% u
    Py <- 1/(1+exp(-eta_y)) #probability of y being 1. a vector of 3.
```

```
  eta_c <- as.numeric(t(as.matrix(gamma)) %*% as.matrix(Xij[i,]) +
     b)

  if(data_j$D[i]==0 & data_j$trt[i]==1){ #never taker
    deriv_2nd_lnf_a[[i]] <- -Py[1]*(1-Py[1]) * Aij[1,]%*%t(Aij[1,])
  }

  if(data_j$D[i]==0 & data_j$trt[i]==0){ #control complier+never
     taker
    D00 <- exp(data_j$Y[i]*eta_y[1]-log(1+exp(eta_y[1]))) + exp(
       data_j$Y[i]*eta_y[2]-log(1+exp(eta_y[2]))+eta_c)
    D00_p <- exp(data_j$Y[i]*eta_y[1]-log(1+exp(eta_y[1])))*(data_j
       $Y[i]-Py[1]) * Aij[1,] +
              exp(data_j$Y[i]*eta_y[2]-log(1+exp(eta_y[2]))+eta_c)*
                 (data_j$Y[i]-Py[2]) * Aij[2,]
    D00_pp <- exp(data_j$Y[i]*eta_y[1]-log(1+exp(eta_y[1])))*((data
       _j$Y[i]-Py[1])^2-Py[1]*(1-Py[1])) * Aij[1,] %*% t(Aij[1,]) +
              exp(data_j$Y[i]*eta_y[2]-log(1+exp(eta_y[2]))+eta_c)*
                 ((data_j$Y[i]-Py[2])^2-Py[2]*(1-Py[2]))* Aij[2,] %
                 *% t(Aij[2,])
    deriv_2nd_lnf_a[[i]] <- ( D00_pp*D00 - D00_p %*% t(D00_p) )/D00
       ^2
  }

  if(data_j$D[i]==1 & data_j$trt[i]==1){ #trt complier
    deriv_2nd_lnf_a[[i]] <- -Py[3]*(1-Py[3]) * Aij[3,]%*%t(Aij[3,])
  }
}#end of loop
```

```
  deriv_2nd_lnf_alpha <- Reduce('+', deriv_2nd_lnf_a)

  return(deriv_2nd_lnf_alpha)
}

## fun_2nd_deriv_L_a() gives second derivative Lj/alpha
fun_2nd_deriv_L_a <- function(u,b,Alpha,gamma,data_j,ind.x0,ind.x1,
   lambda){

  deriv_2nd <- fun_2nd_deriv_lnf_a(u,b,Alpha,gamma,data_j,ind.x0,ind.
     x1,lambda) + fun_der_lnf_alpha(u,b,Alpha,gamma,data_j,ind.x0,ind
     .x1,lambda) %*% t(fun_der_lnf_alpha(u,b,Alpha,gamma,data_j,ind.
     x0,ind.x1,lambda))

  return(deriv_2nd)
}

#fun_der_lnf_lambda() gives the first derivative of lambda \pratial
   lnf(Yj,c|u,b) / \partial lambda for subject i in clinic j
#inputs:
#u: random effects of y model. usually the vector z of length r'
   after change of variable
#b: random effect of c model. usually the last element of vector z
#Alpha: alpha vector in Y model
#gamma: gamma vector in C model
#data_j: data frame of jth site
#ind.x0: column number of covariates controlled in Y model
```

```
#ind.x1: column number of covariates controlled in C model
#lambda: factor loading matrix
#M is the direvative of factor loading matrix

fun_der_lnf_lambda <- function(u,b,Alpha,gamma,data_j,ind.x0,ind.x1,
   lambda,M){
  data_j <- data_j[!is.na(data_j$Y),]
  n.j <- nrow(data_j)

  if(anyNA(ind.x1)){
    Xij <- as.matrix(rep(1,n.j)) #intercept only for compliance model
  } else {
    Xij <- as.matrix(cbind(rep(1,n.j),data_j[,ind.x1])) #intercept
       and covariates
  }

  Bij <- diag(rep(1,3))

  first.deriv <- list()

  for(i in 1:nrow(data_j)){
    if(anyNA(ind.x0)){
      Aij <- diag(rep(1,3))
    } else {
      Aij <- cbind(diag(rep(1,3)),matrix(unlist(replicate(3,as.matrix
         (data_j[,ind.x0])[i,])),nrow = 3,byrow = T))
    }
```

```
  eta_y <- Aij %*% Alpha + Bij %*% lambda %*% u
  Py <- 1/(1+exp(-eta_y))
  eta_c <- as.numeric(t(as.matrix(gamma)) %*% as.matrix(Xij[i,]) +
     b)

  deriv.etay.lambda <- t(Bij) %*% M %*% u


  if(data_j$D[i]==0 & data_j$trt[i]==1){ #never taker
   first.deriv[[i]] <- (data_j$Y[i]-Py[1]) * deriv.etay.lambda[1]
  }

  if(data_j$D[i]==0 & data_j$trt[i]==0){ #control complier
   D00 <- exp(data_j$Y[i]*eta_y[1]-log(1+exp(eta_y[1]))) + exp(data
      _j$Y[i]*eta_y[2]-log(1+exp(eta_y[2]))+eta_c)
   first.deriv[[i]] <- (exp(data_j$Y[i]*eta_y[1]-log(1+exp(eta_y
      [1])))*(data_j$Y[i]-Py[1])* deriv.etay.lambda[1] +
                          exp(data_j$Y[i]*eta_y[2]-log(1+exp(eta_y
                             [2]))+eta_c)*(data_j$Y[i]-Py[2])*deriv
                             .etay.lambda[2]) / D00
  }

  if(data_j$D[i]==1 & data_j$trt[i]==1){ #trt complier
   first.deriv[[i]] <- (data_j$Y[i]-Py[3]) * deriv.etay.lambda[3]
  }

}#end of loop
```

```
  first_der_lj_lambda <- Reduce('+', first.deriv)

  return(first_der_lj_lambda)
}

#fun_2nd_deriv_lnf_lambda() gives second derivative of lambda, lnf /
   lambda_k
#contains n,c0,c1
#inputs:
#u: random effects of y model. usually the vector z of length r'
   after change of variable
#b: random effect of c model. usually the last element of vector z
#Alpha: alpha vector in Y model
#gamma: gamma vector in C model
#data_j: data frame of jth site
#ind.x0: column number of covariates controlled in Y model
#ind.x1: column number of covariates controlled in C model
#lambda: factor loading matrix
#M is the direvative of factor loading matrix

fun_2nd_deriv_lnf_lambda <- function(u,b,Alpha,gamma,data_j,ind.x0,
   ind.x1,lambda,M){
  data_j <- data_j[!is.na(data_j$Y),]
  n.j <- nrow(data_j)

  if(anyNA(ind.x1)){
    Xij <- as.matrix(rep(1,n.j)) #intercept only for compliance model
  } else {
```

```
  Xij <- as.matrix(cbind(rep(1,n.j),data_j[,ind.x1])) #intercept
     and covariates
}

Bij <- diag(rep(1,3))

deriv_2nd_lnf_lambda <- list()

for(i in 1:nrow(data_j)){
  if(anyNA(ind.x0)){
    Aij <- diag(rep(1,3))
  } else {
    Aij <- cbind(diag(rep(1,3)),matrix(unlist(replicate(3,as.matrix
       (data_j[,ind.x0])[i,])),nrow = 3,byrow = T))
  }

  eta_y <- Aij %*% Alpha + Bij %*% lambda %*% u
  Py <- 1/(1+exp(-eta_y)) #probability of y being 1. a vector of 3.
  eta_c <- as.numeric(t(as.matrix(gamma)) %*% as.matrix(Xij[i,]) +
     b)

  deriv.etay.lambda <- t(Bij) %*% M %*% u


    if(data_j$D[i]==0 & data_j$trt[i]==1){ #never taker
      deriv_2nd_lnf_lambda[[i]] <- -Py[1]*(1-Py[1]) * deriv.etay.
         lambda[1]^2
    }
```

```
    if(data_j$D[i]==0 & data_j$trt[i]==0){ #control complier+never
       taker
      D00 <- exp(data_j$Y[i]*eta_y[1]-log(1+exp(eta_y[1]))) + exp(
         data_j$Y[i]*eta_y[2]-log(1+exp(eta_y[2]))+eta_c)

      D00_p_lambda <- exp(data_j$Y[i]*eta_y[1]-log(1+exp(eta_y[1]))
         )*(data_j$Y[i]-Py[1]) * deriv.etay.lambda[1] +
        exp(data_j$Y[i]*eta_y[2]-log(1+exp(eta_y[2]))+eta_c)*(data_
           j$Y[i]-Py[2]) * deriv.etay.lambda[2]

      D00_pp_lambda <- exp(data_j$Y[i]*eta_y[1]-log(1+exp(eta_y[1])
         ))* deriv.etay.lambda[1]^2 * ((data_j$Y[i]-Py[1])^2- Py[1]
         *(1-Py[1]))+
        exp(data_j$Y[i]*eta_y[2]-log(1+exp(eta_y[2]))+eta_c)* deriv
           .etay.lambda[2]^2 * ((data_j$Y[i]-Py[2])^2- Py[2]*(1-Py
           [2]))

      deriv_2nd_lnf_lambda[[i]] <- ( D00_pp_lambda*D00 - D00_p_
         lambda %*% t(D00_p_lambda) )/D00^2
    }

    if(data_j$D[i]==1 & data_j$trt[i]==1){ #trt complier
      deriv_2nd_lnf_lambda[[i]] <- -Py[3]*(1-Py[3]) * deriv.etay.
         lambda[3]^2
    }

}#end of loop
```

```
  deriv_2nd_lnf_lambda <- Reduce('+', deriv_2nd_lnf_lambda)

  return(deriv_2nd_lnf_lambda)
}

##fun_2nd_deriv_L_lambda() gives second derivative Lj/lambda
fun_2nd_deriv_L_lambda <- function(u,b,Alpha,gamma,data_j,ind.x0,ind.
   x1,lambda,M){

  deriv_2nd <- fun_2nd_deriv_lnf_lambda(u,b,Alpha,gamma,data_j,ind.x0
     ,ind.x1,lambda,M) + fun_der_lnf_lambda(u,b,Alpha,gamma,data_j,
     ind.x0,ind.x1,lambda,M) %*% t(fun_der_lnf_lambda(u,b,Alpha,gamma
     ,data_j,ind.x0,ind.x1,lambda,M))

  return(deriv_2nd)
}

#fun_2nd_deriv_lnf_lambda_a() gives the second derivative of lnf /
   lambda_k*alpha
#inputs:
#u: random effects of y model. usually the vector z of length r'
   after change of variable
#b: random effect of c model. usually the last element of vector z
#Alpha: alpha vector in Y model
#gamma: gamma vector in C model
#data_j: data frame of jth site
#ind.x0: column number of covariates controlled in Y model
```

```
#ind.x1: column number of covariates controlled in C model
#lambda: factor loading matrix
#M is the direvative of factor loading matrix

fun_2nd_deriv_lnf_lambda_a <- function(u,b,Alpha,gamma,data_j,ind.x0,
   ind.x1,lambda,M){
  data_j <- data_j[!is.na(data_j$Y),]
  n.j <- nrow(data_j)

  if(anyNA(ind.x1)){
    Xij <- as.matrix(rep(1,n.j)) #intercept only for compliance model
  } else {
    Xij <- as.matrix(cbind(rep(1,n.j),data_j[,ind.x1])) #intercept
       and covariates
  }

  Bij <- diag(rep(1,3))

  deriv_2nd_lnf_lambda_a <- list()

  for(i in 1:nrow(data_j)){
    if(anyNA(ind.x0)){
      Aij <- diag(rep(1,3))
    } else {
      Aij <- cbind(diag(rep(1,3)),matrix(unlist(replicate(3,as.matrix
         (data_j[,ind.x0])[i,])),nrow = 3,byrow = T))
    }
```

```
eta_y <- Aij %*% Alpha + Bij %*% lambda %*% u
Py <- 1/(1+exp(-eta_y)) #probability of y being 1. a vector of 3.
eta_c <- as.numeric(t(as.matrix(gamma)) %*% as.matrix(Xij[i,]) +
   b)

deriv.etay.lambda <- t(Bij) %*% M %*% u

if(data_j$D[i]==0 & data_j$trt[i]==1){ #never taker
  deriv_2nd_lnf_lambda_a[[i]] <- -Py[1]*(1-Py[1]) * deriv.etay.
     lambda[1] * Aij[1,]
}

if(data_j$D[i]==0 & data_j$trt[i]==0){ #control complier+never
   taker
  D00 <- exp(data_j$Y[i]*eta_y[1]-log(1+exp(eta_y[1]))) + exp(
     data_j$Y[i]*eta_y[2]-log(1+exp(eta_y[2]))+eta_c)

  D00_p <- exp(data_j$Y[i]*eta_y[1]-log(1+exp(eta_y[1])))*(data_j
     $Y[i]-Py[1]) * Aij[1,] +
    exp(data_j$Y[i]*eta_y[2]-log(1+exp(eta_y[2]))+eta_c)*(data_j$
       Y[i]-Py[2]) * Aij[2,]

  D00_p_lambda <- exp(data_j$Y[i]*eta_y[1]-log(1+exp(eta_y[1])))*
     (data_j$Y[i]-Py[1]) * deriv.etay.lambda[1] +
    exp(data_j$Y[i]*eta_y[2]-log(1+exp(eta_y[2]))+eta_c)*(data_j$
       Y[i]-Py[2]) * deriv.etay.lambda[2]
```

```
      D00_pp_lambda_a <- exp(data_j$Y[i]*eta_y[1]-log(1+exp(eta_y[1])
         ))* deriv.etay.lambda[1] * ((data_j$Y[i]-Py[1])^2- Py[1]*(1-
         Py[1])) * Aij[1,]+
        exp(data_j$Y[i]*eta_y[2]-log(1+exp(eta_y[2]))+eta_c)* deriv.
           etay.lambda[2] * ((data_j$Y[i]-Py[2])^2- Py[2]*(1-Py[2]))
           * Aij[2,]

      deriv_2nd_lnf_lambda_a[[i]] <- ( D00_pp_lambda_a*D00 - D00_p_
         lambda %*% t(D00_p) )/D00^2
    }

    if(data_j$D[i]==1 & data_j$trt[i]==1){ #trt complier
      deriv_2nd_lnf_lambda_a[[i]] <- -Py[3]*(1-Py[3]) * deriv.etay.
         lambda[3] * Aij[3,]
    }

  }#end of loop

  deriv_2nd_lnf_lambda_a <- Reduce('+', deriv_2nd_lnf_lambda_a)

  return(deriv_2nd_lnf_lambda_a)
}

##fun_2nd_deriv_L_lambda_a() gives the second derivative Lj/lambda*
   alpha
fun_2nd_deriv_L_lambda_a <- function(u,b,Alpha,gamma,data_j,ind.x0,
   ind.x1,lambda,M){
```

```
  deriv_2nd <- fun_2nd_deriv_lnf_lambda_a(u,b,Alpha,gamma,data_j,ind.
     x0,ind.x1,lambda,M) + fun_der_lnf_lambda(u,b,Alpha,gamma,data_j,
     ind.x0,ind.x1,lambda,M) %*% t(fun_der_lnf_alpha(u,b,Alpha,gamma,
     data_j,ind.x0,ind.x1,lambda))

  return(deriv_2nd)
}

#fun_2nd_deriv_lnf_lambda1_lambda2() gives the second derivative of
   lnf / lambda_k*lambda_l
#inputs:
#u: random effects of y model. usually the vector z of length r'
   after change of variable
#b: random effect of c model. usually the last element of vector z
#Alpha: alpha vector in Y model
#gamma: gamma vector in C model
#data_j: data frame of jth site
#ind.x0: column number of covariates controlled in Y model
#ind.x1: column number of covariates controlled in C model
#lambda: factor loading matrix
#M is the direvative of factor loading matrix wrt lambda_k and lambda
   _l

fun_2nd_deriv_lnf_lambda1_lambda2 <- function(u,b,Alpha,gamma,data_j,
   ind.x0,ind.x1,lambda,M){
  data_j <- data_j[!is.na(data_j$Y),]
  n.j <- nrow(data_j)
```

```
if(anyNA(ind.x1)){
  Xij <- as.matrix(rep(1,n.j)) #intercept only for compliance model
} else {
  Xij <- as.matrix(cbind(rep(1,n.j),data_j[,ind.x1])) #intercept
     and covariates
}

Bij <- diag(rep(1,3))

deriv_2nd_lnf_lambda1_lambda2 <- list()

for(i in 1:nrow(data_j)){
  if(anyNA(ind.x0)){
    Aij <- diag(rep(1,3))
  } else {
    Aij <- cbind(diag(rep(1,3)),matrix(unlist(replicate(3,as.matrix
       (data_j[,ind.x0])[i,])),nrow = 3,byrow = T))
  }

  eta_y <- Aij %*% Alpha + Bij %*% lambda %*% u
  Py <- 1/(1+exp(-eta_y)) #probability of y being 1. a vector of 3.
  eta_c <- as.numeric(t(as.matrix(gamma)) %*% as.matrix(Xij[i,]) +
     b)

  if(r_prime==1){
    deriv.etay.lambda1 <- t(Bij) %*% M[,1] %*% u
    deriv.etay.lambda2 <- t(Bij) %*% M[,2] %*% u
```

```
if(data_j$D[i]==0 & data_j$trt[i]==1){ #never taker
  deriv_2nd_lnf_lambda1_lambda2[[i]] <- -Py[1]*(1-Py[1]) *
     deriv.etay.lambda1[1]*deriv.etay.lambda2[1]
}

if(data_j$D[i]==0 & data_j$trt[i]==0){ #control complier+never
   taker
  D00 <- exp(data_j$Y[i]*eta_y[1]-log(1+exp(eta_y[1]))) + exp(
     data_j$Y[i]*eta_y[2]-log(1+exp(eta_y[2]))+eta_c)

  D00_p_lambda_l <- exp(data_j$Y[i]*eta_y[1]-log(1+exp(eta_y
     [1])))*(data_j$Y[i]-Py[1]) * deriv.etay.lambda1[1] +
    exp(data_j$Y[i]*eta_y[2]-log(1+exp(eta_y[2]))+eta_c)*(data_
       j$Y[i]-Py[2]) * deriv.etay.lambda1[2]

  D00_p_lambda_k <- exp(data_j$Y[i]*eta_y[1]-log(1+exp(eta_y
     [1])))*(data_j$Y[i]-Py[1]) * deriv.etay.lambda2[1] +
    exp(data_j$Y[i]*eta_y[2]-log(1+exp(eta_y[2]))+eta_c)*(data_
       j$Y[i]-Py[2]) * deriv.etay.lambda2[2]

  D00_pp_lambda <- exp(data_j$Y[i]*eta_y[1]-log(1+exp(eta_y[1])
     ))* deriv.etay.lambda1[1] * deriv.etay.lambda2[1] * ((data
     _j$Y[i]-Py[1])^2- Py[1]*(1-Py[1]))+
    exp(data_j$Y[i]*eta_y[2]-log(1+exp(eta_y[2]))+eta_c)* deriv
       .etay.lambda1[2] * deriv.etay.lambda2[2] * ((data_j$Y[i
       ]-Py[2])^2- Py[2]*(1-Py[2]))
```

```
    deriv_2nd_lnf_lambda1_lambda2[[i]] <- ( D00_pp_lambda*D00 -
       D00_p_lambda_l %*% D00_p_lambda_k )/D00^2
  }

  if(data_j$D[i]==1 & data_j$trt[i]==1){ #trt complier
    deriv_2nd_lnf_lambda1_lambda2[[i]] <- -Py[3]*(1-Py[3]) *
       deriv.etay.lambda1[3] * deriv.etay.lambda2[3]
  }
} else {
  deriv.etay.lambda1 <- t(Bij) %*% M[,1:2] %*% u
  deriv.etay.lambda2 <- t(Bij) %*% M[,3:4] %*% u

  if(data_j$D[i]==0 & data_j$trt[i]==1){ #never taker
    deriv_2nd_lnf_lambda1_lambda2[[i]] <- -Py[1]*(1-Py[1]) *
       deriv.etay.lambda1[1,] %*% deriv.etay.lambda2[1,]
  }

  if(data_j$D[i]==0 & data_j$trt[i]==0){ #control complier+never
     taker
    D00 <- exp(data_j$Y[i]*eta_y[1]-log(1+exp(eta_y[1]))) + exp(
       data_j$Y[i]*eta_y[2]-log(1+exp(eta_y[2]))+eta_c)

    D00_p_lambda_l <- exp(data_j$Y[i]*eta_y[1]-log(1+exp(eta_y
       [1])))*(data_j$Y[i]-Py[1]) * deriv.etay.lambda1[1,] +
      exp(data_j$Y[i]*eta_y[2]-log(1+exp(eta_y[2]))+eta_c)*(data_
         j$Y[i]-Py[2]) * deriv.etay.lambda1[2,]
```

```
      D00_p_lambda_k <- exp(data_j$Y[i]*eta_y[1]-log(1+exp(eta_y
         [1])))*(data_j$Y[i]-Py[1]) * deriv.etay.lambda2[1,] +
        exp(data_j$Y[i]*eta_y[2]-log(1+exp(eta_y[2]))+eta_c)*(data_
           j$Y[i]-Py[2]) * deriv.etay.lambda2[2,]

      D00_pp_lambda <- exp(data_j$Y[i]*eta_y[1]-log(1+exp(eta_y[1])
         ))* deriv.etay.lambda1[1,] %*% deriv.etay.lambda2[1,] * ((
         data_j$Y[i]-Py[1])^2- Py[1]*(1-Py[1]))+
        exp(data_j$Y[i]*eta_y[2]-log(1+exp(eta_y[2]))+eta_c)* deriv
           .etay.lambda1[2,] %*% deriv.etay.lambda2[2,] * ((data_j$
           Y[i]-Py[2])^2- Py[2]*(1-Py[2]))

      deriv_2nd_lnf_lambda1_lambda2[[i]] <- ( D00_pp_lambda*D00 -
         D00_p_lambda_l %*% D00_p_lambda_k )/D00^2
    }

    if(data_j$D[i]==1 & data_j$trt[i]==1){ #trt complier
      deriv_2nd_lnf_lambda1_lambda2[[i]] <- -Py[3]*(1-Py[3]) *
         deriv.etay.lambda1[3,] %*% deriv.etay.lambda2[3,]
    }
  }

}#end of loop

deriv_2nd_lnf_lambda1_lambda2 <- Reduce('+', deriv_2nd_lnf_lambda1_
   lambda2)

return(deriv_2nd_lnf_lambda1_lambda2)
```

```
}

##fun_2nd_deriv_L_lambda1_lambda2() gives the second derivative Lj/
   lambda_k*lambda_l
fun_2nd_deriv_L_lambda1_lambda2 <- function(u,b,Alpha,gamma,data_j,
   ind.x0,ind.x1,lambda,M){

  deriv_2nd <- fun_2nd_deriv_lnf_lambda1_lambda2(u,b,Alpha,gamma,data
     _j,ind.x0,ind.x1,lambda,M) + fun_der_lnf_lambda(u,b,Alpha,gamma,
     data_j,ind.x0,ind.x1,lambda,M[,1]) %*% t(fun_der_lnf_lambda(u,b,
     Alpha,gamma,data_j,ind.x0,ind.x1,lambda,M[,2]))

  return(deriv_2nd)
}

#fun_2nd_deriv_lnf_lambda_rc() gives the second derivative of lnf /
   lambda_k*gamma

#inputs:
#u: random effects of y model. usually the vector z of length r'
   after change of variable
#b: random effect of c model. usually the last element of vector z
#Alpha: alpha vector in Y model
#gamma: gamma vector in C model
#data_j: data frame of jth site
#ind.x0: column number of covariates controlled in Y model
#ind.x1: column number of covariates controlled in C model
#lambda: factor loading matrix
```

```
#M is the direvative of factor loading matrix

fun_2nd_deriv_lnf_lambda_rc <- function(u,b,Alpha,gamma,data_j,ind.x0
   ,ind.x1,lambda,M){
  data_j <- data_j[!is.na(data_j$Y),]
  n.j <- nrow(data_j)

  if(anyNA(ind.x1)){
    Xij <- as.matrix(rep(1,n.j)) #intercept only for compliance model
  } else {
    Xij <- as.matrix(cbind(rep(1,n.j),data_j[,ind.x1])) #intercept
       and covariates
  }

  Bij <- diag(rep(1,3))

  deriv_2nd_lnf_lambda_rc <- list()

  for(i in 1:nrow(data_j)){
    if(anyNA(ind.x0)){
      Aij <- diag(rep(1,3))
    } else {
      Aij <- cbind(diag(rep(1,3)),matrix(unlist(replicate(3,as.matrix
         (data_j[,ind.x0])[i,])),nrow = 3,byrow = T))
    }

    eta_y <- Aij %*% Alpha + Bij %*% lambda %*% u
    Py <- 1/(1+exp(-eta_y)) #probability of y being 1. a vector of 3.
```

```
eta_c <- as.numeric(t(as.matrix(gamma)) %*% as.matrix(Xij[i,]) +
   b)

deriv.etay.lambda <- t(Bij) %*% M %*% u

if(data_j$D[i]==0 & data_j$trt[i]==1){ #never taker
 deriv_2nd_lnf_lambda_rc[[i]] <- 0
}

if(data_j$D[i]==0 & data_j$trt[i]==0){ #control complier+never
   taker
 D00 <- exp(data_j$Y[i]*eta_y[1]-log(1+exp(eta_y[1]))) + exp(data
    _j$Y[i]*eta_y[2]-log(1+exp(eta_y[2]))+eta_c)

 D00_p_lambda <- exp(data_j$Y[i]*eta_y[2]-log(1+exp(eta_y[2]))+
    eta_c)*(data_j$Y[i]-Py[2]) * deriv.etay.lambda[2]

 D00_pp_lambda_rc <- exp(data_j$Y[i]*eta_y[2]-log(1+exp(eta_y[2])
    )+eta_c)* deriv.etay.lambda[2] %*% t(Xij[i,]) * (data_j$Y[i]-
    Py[2])

 D00_p_rc <- exp(data_j$Y[i]*eta_y[2]-log(1+exp(eta_y[2]))+eta_c)
     %*% t(Xij[i,])

 deriv_2nd_lnf_lambda_rc[[i]] <- (D00_pp_lambda_rc*D00 - D00_p_
    lambda %*% D00_p_rc)/D00^2
}
```

```
    if(data_j$D[i]==1 & data_j$trt[i]==1){ #trt complier
     deriv_2nd_lnf_lambda_rc[[i]] <- 0
    }

  }#end of loop

  deriv_2nd_lnf_lambda_rc <- Reduce('+', deriv_2nd_lnf_lambda_rc)

  return(deriv_2nd_lnf_lambda_rc)
}

##fun_2nd_deriv_L_lambda_rc() gives the second derivative Lj/lambda*r
fun_2nd_deriv_L_lambda_rc <- function(u,b,Alpha,gamma,data_j,ind.x0,
   ind.x1,lambda,M){

  deriv_2nd <- fun_2nd_deriv_lnf_lambda_rc(u,b,Alpha,gamma,data_j,ind
     .x0,ind.x1,lambda,M) + fun_der_lnf_lambda(u,b,Alpha,gamma,data_j
     ,ind.x0,ind.x1,lambda,M) %*% t(deriv_lnf_rc_1st(u,b,Alpha,gamma,
     data_j,ind.x0,ind.x1,lambda))

  return(deriv_2nd)
}

#deriv_lnf_rc_1st() gives the first derivative wrt gamma
#inputs:
#u: random effects of y model. usually the vector z of length r'
   after change of variable
#b: random effect of c model. usually the last element of vector z
```

```
#Alpha: alpha vector in Y model
#gamma: gamma vector in C model
#data_j: data frame of jth site
#ind.x0: column number of covariates controlled in Y model
#ind.x1: column number of covariates controlled in C model
#lambda: factor loading matrix
deriv_lnf_rc_1st <- function(u,b,Alpha,gamma,data_j,ind.x0,ind.x1,
   lambda){
  data_j <- data_j[!is.na(data_j$Y) | !is.na(data_j$C),] #remove obs
     with both missing C and Y

  n.j <- nrow(data_j)

  if(anyNA(ind.x1)){
    Xij <- as.matrix(rep(1,n.j)) #intercept only for compliance model
  } else {
    Xij <- as.matrix(cbind(rep(1,n.j),data_j[,ind.x1])) #intercept
       and covariates
  }

  Bij <- diag(rep(1,3))
  Bij <- Bij %*% lambda

  first <- 0
  second <- 0
  third <- 0
  forth <- 0
  fifth <- 0
```

```
for(i in 1:n.j){
  if(is.na(data_j$Y[i])){
    eta_c <- as.numeric(t(as.matrix(gamma)) %*% as.matrix(Xij[i,])
       + b) #compliance model
    Pc <- 1/(1+exp(-eta_c))
    Pn <- 1-Pc
    if(data_j$trt[i]==1 & data_j$D[i]==0) forth <- forth - Pc*Xij[i
       ,]
    if(data_j$trt[i]==1 & data_j$D[i]==1) fifth <- fifth + Pn*Xij[i
       ,]

  } else {

  if(anyNA(ind.x0)){
    Aij <- diag(rep(1,3))
  } else {
    Aij <- cbind(diag(rep(1,3)),matrix(unlist(replicate(3,as.matrix
       (data_j[,ind.x0])[i,])),nrow = 3,byrow = T))
  }

  eta_y <- Aij %*% Alpha + Bij %*% u #y model
  eta_c <- as.numeric(t(as.matrix(gamma)) %*% as.matrix(Xij[i,]) +
     b)

  if(data_j$trt[i]==0 & data_j$D[i]==0){
    first <- first + exp(data_j$Y[i]*eta_y[2]-log(1+exp(eta_y[2]))+
       eta_c) * Xij[i,] /
```

```
        (exp(data_j$Y[i]*eta_y[1]-log(1+exp(eta_y[1])))+exp(data_j$Y[
           i]*eta_y[2]-log(1+exp(eta_y[2]))+eta_c))
    }
    if(data_j$trt[i]==1 & data_j$D[i]==1){
      second <- second + as.matrix(Xij[i,])
    } #end of if
    third <- third + as.matrix(exp(eta_c) * Xij[i,] / (1+exp(eta_c)))
    }#end of else
  } #end of loop

  return(first + second - third + forth + fifth)
}#end of function

#fun_2nd_deriv_lnf_rc() gives the second derivative wrt to gamma
#inputs:
#u: random effects of y model. usually the vector z of length r'
   after change of variable
#b: random effect of c model. usually the last element of vector z
#Alpha: alpha vector in Y model
#gamma: gamma vector in C model
#data_j: data frame of jth site
#ind.x0: column number of covariates controlled in Y model
#ind.x1: column number of covariates controlled in C model
#lambda: factor loading matrix

fun_2nd_deriv_lnf_rc <- function(u,b,Alpha,gamma,data_j,ind.x0,ind.x1
   ,lambda){
```

```
data_j <- data_j[!is.na(data_j$Y) | !is.na(data_j$C),] #remove obs
   with both missing C and Y

n.j <- nrow(data_j)

if(anyNA(ind.x1)){
  Xij <- as.matrix(rep(1,n.j)) #intercept only for compliance model
} else {
  Xij <- as.matrix(cbind(rep(1,n.j),data_j[,ind.x1])) #intercept
     and covariates
}

Bij <- diag(rep(1,3))
Bij <- Bij %*% lambda

first <- 0
second <- 0
third <- 0
forth <- 0

for(i in 1:nrow(data_j)){
  if(is.na(data_j$Y[i])){
    eta_c <- as.numeric(t(as.matrix(gamma)) %*% as.matrix(Xij[i,])
       + b) #compliance model. eta_c
    if(data_j$trt[i]==1 & data_j$D[i]==0) third <- third - exp(eta_
       c)/(1+exp(eta_c))^2 * Xij[i,] %*% t(Xij[i,])
    if(data_j$trt[i]==1 & data_j$D[i]==1) forth <- forth - exp(eta_
       c)/(1+exp(eta_c))^2 * Xij[i,] %*% t(Xij[i,])
```

```
} else {

if(anyNA(ind.x0)){
  Aij <- diag(rep(1,3))
} else {
  Aij <- cbind(diag(rep(1,3)),matrix(unlist(replicate(3,as.matrix
     (data_j[,ind.x0])[i,])),nrow = 3,byrow = T))
}

eta_y <- Aij %*% Alpha + Bij %*% u
Py <- 1/(1+exp(-eta_y)) #probability of y being 1.
eta_c <- as.numeric(t(as.matrix(gamma)) %*% as.matrix(Xij[i,]) +
   b)
Pc <- 1/(1+exp(-eta_c))

if(data_j$D[i]==0 & data_j$trt[i]==0){
  D00 <- exp(data_j$Y[i]*eta_y[1]-log(1+exp(eta_y[1]))) + exp(
     data_j$Y[i]*eta_y[2]-log(1+exp(eta_y[2]))+eta_c)
  first <- first + (exp(data_j$Y[i]*eta_y[2]-log(1+exp(eta_y[2]))
     +eta_c) * (Xij[i,]) %*% t(Xij[i,]) * D00 -
            exp(data_j$Y[i]*eta_y[2]-log(1+exp(eta_y[2]))+eta_c)
               * Xij[i,] %*% t(exp(data_j$Y[i]*eta_y[2]-log(1+exp
               (eta_y[2]))+eta_c) * Xij[i,]))/D00^2
} #end of if

second <- second+Pc*(1-Pc) * Xij[i,] %*% t(Xij[i,])
}
```

```
  }#end of loop
  output <- first-second+third+forth
  output <- (output+t(output))/2 #fix symmetry issue
  return(output)
}#end of function

#fun_2nd_deriv_L_rc() gives the second derivative Lj/gamma
fun_2nd_deriv_L_rc <- function(u,b,Alpha,gamma,data_j,ind.x0,ind.x1,
   lambda){

  deriv_2nd <- fun_2nd_deriv_lnf_rc(u,b,Alpha,gamma,data_j,ind.x0,ind
     .x1,lambda) + deriv_lnf_rc_1st(u,b,Alpha,gamma,data_j,ind.x0,ind
     .x1,lambda) %*% t(deriv_lnf_rc_1st(u,b,Alpha,gamma,data_j,ind.x0
     ,ind.x1,lambda))

  return(deriv_2nd)
}

#fun_2nd_deriv_lnf_alpha_rc() gives a second derivative wrt gamma and
    alpha
#inputs:
#u: random effects of y model. usually the vector z of length r'
   after change of variable
#b: random effect of c model. usually the last element of vector z
#Alpha: alpha vector in Y model
#gamma: gamma vector in C model
#data_j: data frame of jth site
#ind.x0: column number of covariates controlled in Y model
```

```
#ind.x1: column number of covariates controlled in C model
#lambda: factor loading matrix

fun_2nd_deriv_lnf_alpha_rc <- function(u,b,Alpha,gamma,data_j,ind.x0,
   ind.x1,lambda){
  data_j <- data_j[!is.na(data_j$Y),]
  n.j <- nrow(data_j)

  if(anyNA(ind.x1)){
    Xij <- as.matrix(rep(1,n.j)) #intercept only for compliance model
  } else {
    Xij <- as.matrix(cbind(rep(1,n.j),data_j[,ind.x1])) #intercept
       and covariates
  }

  Bij <- diag(rep(1,3))
  Bij <- Bij %*% lambda

  first <- 0

  for(i in 1:nrow(data_j)){
    if(anyNA(ind.x0)){
      Aij <- diag(rep(1,3))
    } else {
      Aij <- cbind(diag(rep(1,3)),matrix(unlist(replicate(3,as.matrix
         (data_j[,ind.x0])[i,])),nrow = 3,byrow = T))
    }
```

```
    eta_y <- Aij %*% Alpha + Bij %*% u
    Py <- 1/(1+exp(-eta_y))
    eta_c <- as.numeric(t(as.matrix(gamma)) %*% as.matrix(Xij[i,]) +
       b)

    if(data_j$D[i]==0 & data_j$trt[i]==0){
      D00 <- exp(data_j$Y[i]*eta_y[1]-log(1+exp(eta_y[1]))) + exp(
         data_j$Y[i]*eta_y[2]-log(1+exp(eta_y[2]))+eta_c)
      D00_p <- exp(data_j$Y[i]*eta_y[1]-log(1+exp(eta_y[1])))*(data_j
         $Y[i]-Py[1]) * Aij[1,] +
              exp(data_j$Y[i]*eta_y[2]-log(1+exp(eta_y[2]))+eta_c)*(
                 data_j$Y[i]-Py[2]) * Aij[2,]
      first <- first + (exp(data_j$Y[i]*eta_y[2]-log(1+exp(eta_y[2]))
         +eta_c)*(data_j$Y[i]-Py[2]) * (Xij[i,]) %*% t(Aij[2,]) * D00
          -
                        exp(data_j$Y[i]*eta_y[2]-log(1+exp(eta_y[2]))
                           +eta_c) * (Xij[i,]) %*% t(D00_p)) / D00^2
    }#end of if

  }#end of loop
  return(first)
}

##fun_2nd_deriv_L_alpha_rc() gives the second derivative Lj/gamma*
   alpha
fun_2nd_deriv_L_alpha_rc <- function(u,b,Alpha,gamma,data_j,ind.x0,
   ind.x1,lambda){
```

```
  deriv_2nd <- fun_2nd_deriv_lnf_alpha_rc(u,b,Alpha,gamma,data_j,ind.
     x0,ind.x1,lambda) + deriv_lnf_rc_1st(u,b,Alpha,gamma,data_j,ind.
     x0,ind.x1,lambda) %*% t(fun_der_lnf_alpha(u,b,Alpha,gamma,data_j
     ,ind.x0,ind.x1,lambda))

  return(deriv_2nd)
}

#fun_der_1st_ln_phitau() gives the first derivative wrt tau, \partial
    ln(phi) / \partial phi(tau)^T for subject i in clinic j
#inputs:
#u: random effects of y model. usually the vector z of length r'
   after change of variable
#tau: initial or estimated values for tau

fun_der_1st_ln_phitau <- function(u,tau){
  lu <- length(u)
  if(lu==1){
    deriv_1st_lnph_tau <- 0.5 * tau^(-2) * (u^2-tau)
  }
  if(lu>1){
    #locate index of 1 and create derivative matrix of vector tau /
       phi tau
    phi_tau <- c(1:(lu*(lu+1)/2))
    matrix.ind <-  matrix(NA,lu,lu)
    matrix.ind[lower.tri(matrix.ind, diag=T)] <- phi_tau
    matrix.ind <- t(matrix.ind)
    matrix.ind[lower.tri(matrix.ind, diag=T)] <-  phi_tau
```

```
    vec_tau <- stack(as.data.frame(matrix.ind))[,1]
    deriv_vectau_phitau <- matrix(NA,lu^2,lu*(lu+1)/2,byrow = T)
    for(i in 1:length(phi_tau)){
      deriv_vectau_phitau[,i] <- vec_tau %in% phi_tau[i]
    }
    deriv_1st_lnph_tau <- 0.5 * t(deriv_vectau_phitau) %*% kronecker(
       invMatrix(tau),invMatrix(tau)) %*% vec(u %*% t(u)-tau)
  }

  return(deriv_1st_lnph_tau)
}

#fun_2nd_L_tau() gives the second derivative of L wrt tau
#inputs:
#u: random effects of y model. usually the vector z of length r'
   after change of variable
#tau: initial or estimated values for tau
fun_2nd_L_tau <- function(u,tau){
    lu <- length(u)
    if(lu==1){
     integrand <- 1/2 * tau^(-2) - tau^(-3) * u^2  + fun_der_1st_ln_
        phitau(u,tau)^2
    }
    if(lu>1){
      phi_tau <- c(1:(lu*(lu+1)/2))
      matrix.ind <-  matrix(NA,lu,lu)
      matrix.ind[lower.tri(matrix.ind, diag=T)] <- phi_tau
      matrix.ind <- t(matrix.ind)
```

```
      matrix.ind[lower.tri(matrix.ind, diag=T)] <-  phi_tau
      vec_tau <- stack(as.data.frame(matrix.ind))[,1]
      deriv_vectau_phitau <- matrix(NA,lu^2,lu*(lu+1)/2,byrow = T)
      for(i in 1:length(phi_tau)){
        deriv_vectau_phitau[,i] <- vec_tau %in% phi_tau[i]
      }

      deriv_vectau_phitau <- ifelse(deriv_vectau_phitau==T, 1, 0) #
         newly added

      product_u_tau <- invMatrix(tau)%*%u%*%t(u)%*%invMatrix(tau) #
         newly added for solving asymmertric Htau

      second.der <- 1/2 * t(deriv_vectau_phitau) %*% kronecker((
         invMatrix(tau)-2*product_u_tau),invMatrix(tau)) %*% deriv_
         vectau_phitau

      integrand <- second.der + fun_der_1st_ln_phitau(u,tau) %*% t(
         fun_der_1st_ln_phitau(u,tau))
     # integrand <- (integrand + t(integrand))/2 #fix asymmetric Ht
        issue
    }
   return(integrand)
}

deriv_lnf_delta_1st <- function(b,delta){
  lb <- length(b)
  if(lb==1){
```

```
    deriv_1st_lnph_delta <- 0.5 * delta^(-2) * (b^2-delta)
  }
  if(lb>1){
    #locate index of 1 and create derivative matrix of vector delta /
        phi delta
    phi_delta <- c(1:(k*(k+1)/2))
    matrix.ind <-  matrix(NA,k,k)
    matrix.ind[lower.tri(matrix.ind, diag=T)] <- phi_delta
    matrix.ind <- t(matrix.ind)
    matrix.ind[lower.tri(matrix.ind, diag=T)] <-  phi_delta
    vec_delta <- stack(as.data.frame(matrix.ind))[,1]
    deriv_vecdelta_phidelta <- matrix(NA,k^2,k*(k+1)/2,byrow = T)
    for(i in 1:length(phi_delta)){
      deriv_vecdelta_phidelta[,i] <- vec_delta %in% phi_delta[i]
    }
    deriv_1st_lnph_delta <- 0.5 * t(deriv_vecdelta_phidelta) %*%
       kronecker(invMatrix(delta),invMatrix(delta)) %*% vec(b %*% t(b
       )-delta)
  }

  return(deriv_1st_lnph_delta)
}

#fun_2nd_L_delta() gives the second derivative of L wrt delta
#inputs:
#b: random effects of C model. usually the last element of vector z
   after change of variable
#delta: initial or estimated values for delta
```

```
fun_2nd_L_delta <- function(b,delta){
  lb <- length(b)
  if(lb==1){
    integrand <- 1/2 * delta^(-2) - delta^(-3) * b^2  + deriv_lnf_
       delta_1st(b,delta)^2
  }
  if(lb>1){
    phi_delta <- c(1:(lb*(lb+1)/2))
    matrix.ind <-  matrix(NA,lb,lb)
    matrix.ind[lower.tri(matrix.ind, diag=T)] <- phi_delta
    matrix.ind <- t(matrix.ind)
    matrix.ind[lower.tri(matrix.ind, diag=T)] <-  phi_delta
    vec_delta <- stack(as.data.frame(matrix.ind))[,1]
    deriv_vecdelta_phidelta <- matrix(NA,lb^2,lb*(lb+1)/2,byrow = T)
    for(i in 1:length(phi_delta)){
      deriv_vecdelta_phidelta[,i] <- vec_delta %in% phi_delta[i]
    }

    second.der <- 1/2 * t(deriv_vecdelta_phidelta) %*% kronecker((
       invMatrix(delta)-2*invMatrix(delta)%*%b%*%t(b)%*%invMatrix(
       delta)),invMatrix(delta)) %*% deriv_vecdelta_phidelta

    integrand <- second.der + deriv_lnf_delta_1st(b,delta) %*% t(
       deriv_lnf_delta_1st(b,delta))

  }
  return(integrand)
}
```

```
#set_init() gives the initial values by fitting logit model by glmer
   ()
#inputs:
#data: dataframe containing outcome, treatment assignment and receipt
   , cluster/site level variable
#r: full dimension of random effects of Y model
#side: one sided or two sided noncompliance. default is one
#formula.c: formula for C model
#formula.y: formula for Y model
#subname: name of the cluster/site variable
#method: either "pmm" predictive mean matching or "prob" probability
#r_prime: reduced dimension of random effects of Y model

set_init <- function(data,r,side=1,formula.c,formula.y,subname,method
   ,r_prime){
  cluster.ind <- which(colnames(data)==subname)
  if(side==1){
    for(i in 1:nrow(data)){
      if(data$trt[i]==1 & data$D[i]==1){
        data$C[i] <- 1
      }
      if(data$trt[i]==1 & data$D[i]==0){
        data$C[i] <- 0
      }
      if(data$trt[i]==0 & data$D[i]==0){
        data$C[i] <- NA
      }
```

```
}#end of loop
P.c <- sum(data$C, na.rm = TRUE)/nrow(data[!is.na(data$C),])
formC <- as.formula(formula.c)
fit.c <- glmer(formC, data=data, family = binomial, nAGQ=20)
C.init <- fit.c@beta
delta.init <- fit.c@theta^2
predicted <- predict(fit.c,type="response",newdata=data,allow.new
   .levels=T) #no random effect
data$predicted <- predicted
data$sub_id <- data[,cluster.ind]

J <- dim(table(data$clinic))

if(method=="pmm"){
  for(j in 1:J){
    subset <- data[data$clinic==j,]
    if(all(is.na(subset$C))) next
    for(i in 1:nrow(subset)){
      if(!is.na(subset$C[i])) next
      else {
        matchid <- subset[!is.na(subset$C),]$sub_id[which.min(abs
           (subset$predicted[i]-subset$predicted[!is.na(subset$C)
           ]))]
        subset$C[i] <- subset$C[subset$sub_id==matchid]
        data$C[data$sub_id==subset$sub_id[i] & data$clinic==j] <-
            subset$C[i]
      }
    }#end of for loop i
```

```
  }#end of for loop j
}#end of "pmm"

if(method=="prob"){
  for(i in 1:nrow(data)){
    if(is.na(data$C[i])){
      data$C[i] <- data$predicted[i]
    }
  }
}#end of "prob"
if(r==1){
  fit.y <- glmer(Y~ -1 + I(1-C) + I((1-trt)*C) + I(trt*C)+(1|
     clinic), data=data, family=binomial)
}
if(r==2){
  fit.y <- glmer(Y~ -1 + I(1-C) + I((1-trt)*C) + I(trt*C)+(-1+I
     (1-C)+C|clinic), data=data, family=binomial)
}
if(r==3){
  fit.y <- glmer(Y~ -1 + I(1-C) + I((1-trt)*C) + I(trt*C)+(-1 + I
     (1-C) + I((1-trt)*C) + I(trt*C)|clinic), data=data, family=
     binomial)
}

alpha.init <- fit.y@beta

formY <- as.formula(formula.y)
fit.itt <- glmer(formY, data=data, family = binomial, nAGQ=20)
```

```
if(r==1) {
  tau.init <- fit.y@theta^2
  lambda.init <- 0
}
if(r==2) {
  tau.vec <- as.data.frame(summary(fit.y)$varcor)[,4] #tau11,
     tau22, tau12
  tau.init <- matrix(c(tau.vec[1],tau.vec[3],tau.vec[3],tau.vec
     [2]),2,2)
  lambda.init <- 0
}

if(r==3) {
  if(r_prime==1){
    tau.vec <- as.data.frame(summary(fit.y)$varcor)[,4] #tau11,
       tau22, tau33, tau12, tau13, tau23
    tau.vec <- tau.vec[c(1,4,5,2,6,3)]
    tau.m <- VechToCovM(tau.vec, r)
    tau.init <- tau.m[1,1]
    lambda1.init <- tau.m[1,2]/tau.m[1,1]
    lambda2.init <- tau.m[1,3]/tau.m[1,1]
    lambda.init <- c(lambda1.init,lambda2.init)
  }
  if(r_prime==2){
    tau.vec <- as.data.frame(summary(fit.y)$varcor)[,4] #tau11,
       tau22, tau33, tau12, tau13, tau23
    tau.vec <- tau.vec[c(1,4,5,2,6,3)]
```

```
        tau.m <- VechToCovM(tau.vec, r)
        tau.init <- tau.m[1:2,1:2]
        lambda.init <- tau.m[1,3]/tau.m[1,1]
      }

      if(r_prime==3){
        tau.vec <- as.data.frame(summary(fit.y)$varcor)[,4] #tau11,
           tau22, tau33, tau12, tau13, tau23
        tau.vec <- tau.vec[c(1,4,5,2,6,3)]
        tau.init <- VechToCovM(tau.vec, r)
      }
    } #end of if(r=3)

    y.itt.est <- fit.itt@beta
  }#end of if(side==1)

  return(list(alpha.intercept.init=alpha.init,tau.init=tau.init,C.
     init=C.init,delta.init=delta.init,lambda.init=lambda.init,
              estimatePC=P.c, y.itt.est=y.itt.est))
}#end of function

#invMatrix() finds the inverse of a matrix
invMatrix <- function(matrix){
  L <- chol(matrix)
  invM <- solve(L)%*%t(solve(L))
  return(invM)
}
```

```
#VechToCovM() creates a symmetric matrix from a vector of unique
   elements by row. i.e. from \phi_tau to tau

VechToCovM <- function(phi_tau,d){
  CovM <- matrix(NA,d,d)
  irow=0
  for(i in 1:d){
    for(j in i:d){
      irow=irow+1
      CovM[i,j] <- CovM[j,i] <- phi_tau[irow]
    }#end of loop j
  } #end of loop i
  return(CovM)
}#end of function

#MtoVech() converts covariance matrix to a vector of unique elements
MtoVech <- function(CovM,d){
  phi_tau <- numeric(length = d*(d+1)/2)
  irow=0
  for(i in 1:d){
    for(j in i:d){
      irow=irow+1
      phi_tau[irow] <- CovM[i,j]
    }
  }
  return(phi_tau)
}
```

D.2 Simulation

The following codes are for simulations when one random effect is shared among three groups in Y model, when outcome is fully observed.

```
rm(list=ls())
library(dplyr)
library(matrixcalc) #vec(), vech(), is.sysmetric.matrix
library(fastGHQuad) #integration
library(mvtnorm)
library(doParallel)
library(foreach)
library(lme4)


source("AGHQ.R")
source("f_y.R")
source("deriv_lnf_alpha_1st.R")
source("deriv_lnf_alpha_2nd.R")
source("deriv_lnf_tau_1st.R")
source("deriv_lnf_delta_1st.R")
source("deriv_lnf_rc_1st.R")
source("fun_2nd_deriv_lnf_alpha_rc.R")
source("fun_2nd_deriv_lnf_rc.R")
source("deriv_lnf_lambda_1st.R")
source("deriv_lnf_lambda_2nd.R")
source("deriv_lnf_lambda_a_2nd.R")
source("deriv_lnf_lambda_rc_2nd.R")
source("deriv_lnf_lambda1_lambda2_2nd.R")
source("deriv_L_tau_2nd.R")
```

```
source("fun_2nd_L_delta.R")
source("VechToCovM.R")
source("MtoVech.R")
source("invMatrix.R")
source("set_init.R")
source("cace_1side_parallel_lambda.R")

registerDoParallel(cores=40)

miu.etaT=0.2
r=3
Q=4
S=1
r_prime=1 #reduced dimension of random effects of Y model
k=1
n <- 40
J=170
niter=500
tol <- 1e-04

alpha.true <- c(0.5,0.7,1.2,-0.5,1) #alpha_x1=1, alpha_x2=2
gamma.true <- c(1,-0.5, 0.5) #gamma_x1=0.5, gamma_x2=1
lambda.true <- c(1,0.7,1.1)
delta.true <- 0.3
tau.true <- 0.5

true <- c(alpha.true,gamma.true,lambda.true[2:3],tau.true, delta.true
   )
```

```
sim_x1x2_2level <- function(seed, r, miu.etaT, n, J, alpha.true,
   gamma.true, lambda.true, tau.true, delta.true, S){
  set.seed(seed)
  alpha <- alpha.true
  gamma <- gamma.true
  N <- n*J

  #simulate T
  etaT=rnorm(J,miu.etaT,0.2) #
  piT=1/(1+exp(-etaT)) # P(T=1), constant within practice j
  summary(piT)
  #    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  # 0.3540  0.4652  0.4982  0.5011  0.5339  0.6818
  id=T=rep(0,N)       # id, T, etaY, Y

  for (j in 1:J) {
    id[((j-1)*n+1):(j*n)] = j      # school ID
    T[((j-1)*n+1):(j*n)] = rbinom(n,1,piT[j]) # T~Bernoulli(piT[j])
  }

  x1 <- round(rnorm(N,1,1),4)
  x2 <- rbinom(N,1,0.65)

  #simulate C
  tau=delta.true  # var(delta)=tau
  L=sqrt(tau)
  etaC=rep(0,N)  # etaC=r+b  compliance model
```

```
Xc <- cbind(rep(1,N), x1, x2)

b1 <- L * rnorm(J,0,1)
b2 <- list()
for(j in 1:J){
  b2[[j]] <- rep(b1[j],n)
}

b <- unlist(b2)

for(i in 1:N){
  etaC[i]=Xc[i,] %*% gamma +b[i]
}
summary(etaC)
#mean=0.79

pin=1/(1+exp(etaC))    # pin=P(C=n)
pic=exp(etaC)/(1+exp(etaC))

summary(pin)
#mean=0.33
summary(pic)
#mean=0.67

C=matrix(0,N,2)        # C=[Cn Cc]=[1{C=n} 1{C=c}]
for(i in 1:N){
  C[i,]=t(rmultinom(1,1,c(pin[i],pic[i])))
```

```
} # C~binom(pic)
summary(C)

#simulate Y
lambda <- lambda.true

if(r==1){
  taubeta <- lambda %*% t(lambda) * tau.true
} else {
  taubeta <- lambda %*% tau.true %*% t(lambda)
}

Lbeta=t(chol(taubeta))

u=matrix(0,J,3)       #random effects of 3 compliance groups

for (j in 1:J){
  u[j,] = t(Lbeta %*% rnorm(3,0,1))
}

u1 = u[rep(seq_len(nrow(u)), each = n),]

Xy = cbind(x1,x2)

beta=matrix(0,N,3)

for (i in 1:N){
  Aij <- cbind(diag(rep(1,3)), Xy[rep(i, 3),])
```

```
  beta[i,] <- Aij %*% alpha.true + u1[i,]
}

etaY=C[,1]*beta[,1]+C[,2]*((1-T)*beta[,2]+T*beta[,3])

summary(etaY)
#mean=0.97

piY = 1/(1+exp(-etaY))

Y=rep(0,N)

for(i in 1:N){
  Y[i] = rbinom(1,1,piY[i])
} #generate outcome Y
summary(Y)
#mean=0.69. cor(Y,x2)=0.23, cor(Y,x1)=-0.21

D=rep(0,N)  # set all D=0

D[T*C[,2]==1]=1 # set D=1 of compliers assigned to T=1
summary(D) #mean=0.39
summary(id)

L1=as.data.frame(cbind(id, Y, T, D, C, x1, x2))   # L1 complete
   data=[id Y T D Cn Cc x1]
names(L1)[5:6]=c("Cn", "Cc")    # Cn=1(C=n),Cc=1(C=c); for binary
   cases, C = Cc
```

```
  colnames(L1)[c(1,3)] = c("clinic","trt")
  L1$id <- rep(1:n,J)
  L1o=L1
  # L1o observed data=[id Y T D Cn Cc]

  L1o[T==0&D==0,5:6]=NA
  return(list(L1=L1,L1o=L1o))
}

#simulation
  print(paste0("ISIM=␣",isim))
  L1o <- sim_x1x2_2level(seed = (isim), r=r_prime, miu.etaT, n, J,
     alpha.true, gamma.true, lambda.true, tau.true, delta.true, S=1)$
     L1o
  init.list <- set_init(L1o,r=3,side = 1,C~x1+x2+(1|clinic), Y~x1+x2
     +(1|clinic),"id","pmm",r_prime)
  if(init.list$tau.init==0) init.list$tau.init <- init.list$tau.init
     +0.1
  if(init.list$delta.init==0) init.list$delta.init <- init.list$delta
     .init+0.1
  if(0 %in% init.list$lambda.init) init.list$lambda.init[which(init.
     list$lambda.init==0)] <- 0.1

  init <- c(init.list$alpha.intercept.init, init.list$y.itt.est[2:
     length(init.list$y.itt.est)],
            init.list$C.init,
            init.list$lambda.init,
            init.list$tau.init,
```

```
            init.list$delta.init)
  #initial values should be in this order: alpha, gamma, lambda, tau,
      delta

  print(init)

    res <- cace_1side_parallel_lambda(r=r,k=k,Q=4,J=J,L1o,x0=c("x1","
       x2"),x1=c("x1","x2"),init = init, 1,3,niter=niter,tol=tol,S=1,
       r_prime=1)

saveRDS(res,file = "r1_n40_J170_x1x2_isim.rds")
```

The following codes are for simulations when one random effect is shared among three groups in Y model, when outcome is partially observed.

```
rm(list=ls())
library(dplyr)
library(matrixcalc) #vec(), vech(), is.sysmetric.matrix
library(fastGHQuad) #integration
library(mvtnorm)
library(doParallel)
library(foreach)
library(lme4)

source("AGHQ.R")
source("f_y.R")
source("deriv_lnf_alpha_1st.R")
source("deriv_lnf_alpha_2nd.R")
source("deriv_lnf_tau_1st.R")
```

```
source("deriv_lnf_delta_1st.R")
source("deriv_lnf_rc_1st.R")
source("fun_2nd_deriv_lnf_alpha_rc.R")
source("fun_2nd_deriv_lnf_rc.R")
source("deriv_lnf_lambda_1st.R")
source("deriv_lnf_lambda_2nd.R")
source("deriv_lnf_lambda_a_2nd.R")
source("deriv_lnf_lambda_rc_2nd.R")
source("deriv_lnf_lambda1_lambda2_2nd.R")
source("deriv_L_tau_2nd.R")
source("fun_2nd_L_delta.R")
source("VechToCovM.R")
source("MtoVech.R")
source("invMatrix.R")
source("set_init.R")
source("cace_1side_parallel_lambda.R")
source("cace_1side_nodaarem.R")

registerDoParallel(cores=46)

#simulation
##fit intercept model
miu.etaT=0.2
r=3
Q=8
S=1
r_prime=1
k=1
```

```
n <- 160
J=63
niter=500
tol <- 1e-04

alpha.true <- c(-1,-0.5, -0.2, -0.5,1) #alpha_x1=1, alpha_x2=2
gamma.true <- c(1,0.5, 2) #gamma_x1=0.5, gamma_x2=1
lambda.true <- c(0.9,1.1)
delta.true <- 0.3 #
tau.true <- 0.5 #
missrate <- 0.05
true <- c(alpha.true,gamma.true,lambda.true,tau.true, delta.true)

sim_x1x2_2level <- function(seed, r, miu.etaT, n, J, alpha.true,
   gamma.true, lambda.true, tau.true, delta.true, S, missrate){
  set.seed(seed)
  alpha <- alpha.true
  gamma <- gamma.true
  N <- n*J

  #simulate T
  etaT=rnorm(J,miu.etaT,0.2) #
  piT=1/(1+exp(-etaT)) # P(T=1), constant within practice j
  summary(piT)
  #    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  # 0.3540  0.4652  0.4982  0.5011  0.5339  0.6818
  id=T=rep(0,N)       # id, T, etaY, Y
```

```
for (j in 1:J) {
  id[((j-1)*n+1):(j*n)] = j      # school ID
  T[((j-1)*n+1):(j*n)] = rbinom(n,1,piT[j]) # T~Bernoulli(piT[j])
}

x1 <- round(rnorm(N,0,1),4)
x2 <- rbinom(N,1,0.65)

#simulate C
tau=delta.true  # var(delta)=tau
L=sqrt(tau)
etaC=rep(0,N)  # etaC=r+b  compliance model

Xc <- cbind(rep(1,N), x1, x2)

b1 <- L * rnorm(J,0,1)
b2 <- list()
for(j in 1:J){
  b2[[j]] <- rep(b1[j],n)
}

b <- unlist(b2)

for(i in 1:N){
  etaC[i]=Xc[i,] %*% gamma +b[i]
}
summary(etaC)
#mean=0.79
```

```
pin=1/(1+exp(etaC))     # pin=P(C=n)
pic=exp(etaC)/(1+exp(etaC))

summary(pin)
#mean=0.15
summary(pic)
#mean=0.85

C=matrix(0,N,2)         # C=[Cn Cc]=[1{C=n} 1{C=c}]
for(i in 1:N){
  C[i,]=t(rmultinom(1,1,c(pin[i],pic[i])))
} # C~binom(pic)
summary(C)

#simulate Y

taubeta <- tau.true
Lbeta <- as.numeric(chol(taubeta))
u=matrix(0,J,3)        #random effects of 3 compliance groups

if(r==1){#only simulate s=1
  u[,1] = Lbeta %*% rnorm(J,0,1)
  u[,2] = lambda.true[1] * u[,1]
  u[,3] = lambda.true[2] * u[,1]
}
if(r==2){
  selectrandom <- matrix(NA,nrow=J,ncol=r)
```

```
  for(j in 1:j){
    selectrandom <- Lbeta %*% rnorm(2,0,1)
  }
  if(S==1){
    u[,1] = selectrandom[1]
    u[,2] = selectrandom[2]
    u[,3] = lambda.true * u[,1]
  }
  if(S==2){
    u[,1] = selectrandom[1]
    u[,2] = lambda.true * u[,1]
    u[,3] = selectrandom[2]
  }
  if(S==3){
    u[,1] = selectrandom[1]
    u[,2] = selectrandom[2]
    u[,3] = lambda.true * u[,2]
  }
} #end of r=2

if(r==3){
  for(j in 1:J){
    u[j,] = t(Lbeta %*% rnorm(3,0,1))
  }
 }
u1 = u[rep(seq_len(nrow(u)), each = n),]

Xy = cbind(x1,x2)
```

```
beta=matrix(0,N,3)

for (i in 1:N){
  Aij <- cbind(diag(rep(1,3)), Xy[rep(i, 3),])
  beta[i,] <- Aij %*% alpha.true + u1[i,]
}

etaY=C[,1]*beta[,1]+C[,2]*((1-T)*beta[,2]+T*beta[,3])

summary(etaY)
#mean=0.18

piY = 1/(1+exp(-etaY))

Y=rep(0,N)

for(i in 1:N){
  Y[i] = rbinom(1,1,piY[i])
} #generate outcome Y
summary(Y)
#mean=0.542

D=rep(0,N)  # set all D=0

D[T*C[,2]==1]=1 # set D=1 of compliers assigned to T=1
summary(D) #mean=0.456
summary(id)
```

```
  L1=as.data.frame(cbind(id, Y, T, D, C, x1, x2))   # L1 complete
     data=[id Y T D Cn Cc x1]
  names(L1)[5:6]=c("Cn", "Cc")    # Cn=1(C=n),Cc=1(C=c); for binary
     cases, C = Cc
  colnames(L1)[c(1,3)] = c("clinic","trt")
  L1$id <- rep(1:n,J)
  L1o=L1

  L1o[T==0&D==0,5:6]=NA

  #sample missing y
  miss.n <- missrate*N
  miss.sample <- sample(1:N,miss.n,replace = F)
  L1o$M <- 1:N
  L1o$Y[which(L1o$M %in% miss.sample)] <- NA
  L1o <- L1o[,-length(L1o)]
  return(list(L1=L1,L1o=L1o))
}

#simulation
print(paste0("ISIM=␣",isim))
L1o <- sim_x1x2_2level(seed = (isim), r=r_prime, miu.etaT, n, J,
   alpha.true, gamma.true, lambda.true, tau.true, delta.true, S,
   missrate)$L1o
init.list <- set_init(L1o,r=3,side = 1,C~x1+x2+(1|clinic), Y~x1+x2
   +(1|clinic),"id","pmm",r_prime)
if(init.list$tau.init==0) init.list$tau.init <- init.list$tau.init
```

```
   +0.1
if(init.list$delta.init==0) init.list$delta.init <- init.list$delta.
   init+0.1
if(0 %in% init.list$lambda.init) init.list$lambda.init[which(init.
   list$lambda.init==0)] <- 0.1

init <- c(init.list$alpha.intercept.init, init.list$y.itt.est[2:
   length(init.list$y.itt.est)],
           init.list$C.init,
           init.list$lambda.init,
           init.list$tau.init,
           init.list$delta.init)
#initial values should be in this order: alpha, gamma, lambda, tau,
   delta

print(init)

res <- cace_1side_parallel_lambda(r=r,k=k,Q=8,J=J,L1o,x0=c("x1","x2")
   ,x1=c("x1","x2"),init = init, col.clinic=1,col.trt=3,col.D = 4,col
   .Y=2,niter=niter,tol=tol,S=1,r_prime=1)

saveRDS(res,file = "r1my5_q8_n160_J63_x1x2_isim.rds")
```

The following shell codes are for sending 100 simulation jobs to computation clusters.

```
    for ((i=1; i<=100; i++)) #
        do
        cp r1ym5_q8_n160_J63_x1x2_temp.R r1ym5_q8_n160_J63_x1x2_$i.R
        sed -i s/isim/$i/g r1ym5_q8_n160_J63_x1x2_$i.R
```

```
    qR r1ym5_q8_n160_J63_x1x2_$i.R
done
```

D.3 Real Data Analysis

D.3.1 e-assist

The following codes fit a CACE model controlling marital status and CRCS order in Y model and gender, charlson comordity score, CRCS order, centered patient age in C model, with a random effect shared among three groups for Y model.

```
rm(list=ls())
library(dplyr)
library(matrixcalc) #vec(), vech(), is.sysmetric.matrix
library(lme4)
library(fastGHQuad) #integration
library(mvtnorm)
library(doParallel)
library(foreach)

source("AGHQ.R")
source("f_y.R")
source("deriv_lnf_alpha_1st.R")
source("deriv_lnf_alpha_2nd.R")
source("deriv_lnf_tau_1st.R")
source("deriv_lnf_delta_1st.R")
source("deriv_lnf_rc_1st.R")
source("fun_2nd_deriv_lnf_alpha_rc.R")
source("fun_2nd_deriv_lnf_rc.R")
```

```
source("deriv_lnf_lambda_1st.R")
source("deriv_lnf_lambda_2nd.R")
source("deriv_lnf_lambda_a_2nd.R")
source("deriv_lnf_lambda_rc_2nd.R")
source("deriv_lnf_lambda1_lambda2_2nd.R")
source("deriv_L_tau_2nd.R")
source("fun_2nd_L_delta.R")
source("VechToCovM.R")
source("MtoVech.R")
source("invMatrix.R")
source("set_init.R")
source("cace_1side_parallel_lambda.R")

registerDoParallel(cores=48)

df <- read.csv("df_1817.csv", header = T)

J <- dim(table(df$clinic))

r=3
r_prime=1
k=1
niter=1000
Q=7
tol <- 1e-04

init.list <- set_init(df,r=r,side = 1,C~gender+charlson+first.colon.
   order+age.c+(1|clinic),
```

```
                        Y~first.colon.order+marry+(1|clinic),
                        "study_id","pmm",r_prime = r_prime)
init <- c(init.list$alpha.intercept.init, init.list$y.itt.est[2:
   length(init.list$y.itt.est)],
           init.list$C.init,
           init.list$lambda.init,
           init.list$tau.init,
           init.list$delta.init)

print(init)

res <- cace_1side_parallel_lambda(r=r,k=k,Q=Q,J=J,df,x0=c("first.
   colon.order","marry"), x1=c("gender","charlson","first.colon.order
   ","age.c"), init = init, col.clinic = 11, col.trt = 4,niter=niter,
   tol=tol,Share=1,r_prime=r_prime)

saveRDS(res,file = "r1q7_YMF_1817_CGenderFAgecCharlson.rds")
```

D.3.2 NSLM

The following codes fit a CACE model controlling challenge behavior and standardized pre-treatment GPA in Y model and school achievement, challenge behavior, gender, parent education, free/reduced lunch or not in C model, with a random effect shared among three groups for Y model, to lower-achieving students in NSLM.

```
df <- read.csv("df10341.csv", header = T)

library(dplyr)
library(matrixcalc) #vec(), vech(), is.sysmetric.matrix
```

```
library(fastGHQuad) #integration
library(mvtnorm)
library(doParallel)
library(foreach)
library(lme4)

source("AGHQ.R")
source("f_y.R")
source("deriv_lnf_alpha_1st.R")
source("deriv_lnf_alpha_2nd.R")
source("deriv_lnf_tau_1st.R")
source("deriv_lnf_delta_1st.R")
source("deriv_lnf_rc_1st.R")
source("fun_2nd_deriv_lnf_alpha_rc.R")
source("fun_2nd_deriv_lnf_rc.R")
source("deriv_lnf_lambda_1st.R")
source("deriv_lnf_lambda_2nd.R")
source("deriv_lnf_lambda_a_2nd.R")
source("deriv_lnf_lambda_rc_2nd.R")
source("deriv_lnf_lambda1_lambda2_2nd.R")
source("deriv_L_tau_2nd.R")
source("fun_2nd_L_delta.R")
source("VechToCovM.R")
source("MtoVech.R")
source("invMatrix.R")
source("set_init.R")
source("cace_1side_parallel_lambda.R")
```

```
registerDoParallel(cores=15)

df_lowgpa <- df[df$lowgpa==1,]
df_highgpa <- df[df$lowgpa==0,]

J <- dim(table(df_lowgpa$clinic))
r=3
Q=4
S=1
r_prime=1
k=1
niter=500
tol <- 1e-04

###################################
#test post-trt gpa in low achieveing students

init.list <- set_init(df_lowgpa,r=3,side = 1,C~schoolquality +
   challenge_behavior_binary + gender_imputed_wmean + pared_imputed_
   wmean + free_reduced_lunch_imputed_wmean + (1|clinic),
                      Y~ challenge_behavior_binary + pregpa_standard
                         + (1|clinic),"stu_id","pmm",r_prime)

init <- c(init.list$alpha.intercept.init, init.list$y.itt.est[2:
   length(init.list$y.itt.est)],
          init.list$C.init,
          init.list$lambda.init,
          init.list$tau.init,
```

```
          init.list$delta.init)
#initial values should be in this order: alpha, gamma, lambda, tau,
   delta

print(init)
length(init)
res <- cace_1side_parallel_lambda(r=r,k=k,Q=4,J=J,df_lowgpa,
                                  x0=c("challenge_behavior_binary","
                                     pregpa_standard"),
                                  x1=c("schoolquality", "challenge_
                                     behavior_binary", "gender_
                                     imputed_wmean", "pared_imputed_
                                     wmean", "free_reduced_lunch_
                                     imputed_wmean"),
                                  init = init, col.clinic = 38, col.
                                     trt = 3, niter=niter,tol=tol,S
                                     =1,r_prime=1)

#x0 is for y model. x1 for c model.

saveRDS(res,file = "r1my_Y_gpa_lowgpa_2.rds")
```