Bayesian Design and Analysis for Studies with Intercurrent Events and Noncompliance

Anthony Sisti

A Dissertation submitted in partial fulfillment of the
requirements for the Degree of Doctor of Philosophy
in the Department of Biostatistics at Brown University

Providence, Rhode Island

October 2024

This dissertation by Anthony Sisti is accepted in its present form
by the Department of Biostatistics as satisfying the
dissertation requirement for the degree of Doctor of Philosophy.

Date _____ _____

Roee Gutman, Advisor

Recommended to the Graduate Council

Date _____ _____

Constantine Gatsonis, Reader

Date _____ _____

Vince Mor, Reader

Date _____ _____

Ellen McCreedy, Reader

Approved by the Graduate Council

Date _____ _____

Thomas A. Lewis, Dean of the Graduate School

Anthony Sisti

Contact Information

Email: anthony.j.sisti@gmail.com
Cell-Phone: (860) 502-5740
GitHub:// **AnthonySisti**
LinkedIn:// **anthonyjsisti**

Education

Brown University
Ph.D. in Biostatistics, September 2024
Advisor: Roee Gutman, Ph.D.
Thesis: *Bayesian Design and Analysis for Studies with Intercurrent Events and Noncompliance*

Brown University
M.A. in Biostatistics, May 2021

University of Connecticut
B.S. in Mathematics/Statistics, May 2019
Honors Advisors: Kun Chen, Ph.D. and Blair T. Johnson, Ph.D.

Programming

Advanced: R | *Proficient*: Python | *Intermediate*: SQL | *Familiar*: SAS, Stata, Git, Matlab

Professional Experience

Graduate Research Assistant **August 2019 - Present**
Brown University

- Graduate biostatistician for NIA-funded pragmatic cluster randomized controlled trial: Music & MEmory: A TRial for Nursing Home Residents With ALzheimer's Disease (METRICaL).
- Collaborates with Brown faculty to produce statistical analyses and methods, including missing data imputation and causal inference models, for the METRICaL pragmatic trial evaluating the effect of music therapy on 1,000 dementia patients in 54 nursing homes in the US.
- Develops and presents independent analyses of binary, ordinal and continuous clinical outcome data using both Bayesian and frequentist multi-level models, contributing to 5 papers submitted for publication.
- Led independent secondary analysis using Bayesian hierarchical multinomial model to evaluate music therapy's effect on patient mood and emotion using data from over 7,000 observations.
- First authored and submitted recently accepted manuscript for publication that showed reduction in verbal agitated behaviors and increase in pleasure during observations for those randomized to receive music therapy, a result not demonstrated in main trial analysis.

Biostatistics PhD Intern **June 2023 - August 2023**
Biogen

- Led research project to develop Bayesian Hierarchical Models (BHMs) that synthesize evidence of treatment efficacy for a class of therapeutics across several drugs, each the subject of multiple clinical trials with several correlated endpoints.
- Completed comprehensive literature review for Biogen's biostatistics group describing the current state of the literature on BHMs for treatment effect evaluation and how they compare to global statistical tests, another common multiple primary endpoint method.
- Derived and programmed 3 unique BHMs in R that allow statisticians to estimate the probability of treatment efficacy while accounting for heterogeneity across several drugs and studies with multiple continuous, binary or mixed outcomes.
- Developed simulation study framework for model testing in real-world scenarios. Implemented framework to cover more than 20 configurations that demonstrate the performance of the BHMs when compared to global statistical tests. Presented results to Biogen's biostatistics group.

Student Researcher **August 2018 - April 2021**
UConn Institute for Collaboration on Health, Intervention, and Policy (InCHIP)

- Implemented statistical analyses and created data visualizations for the Systematic Health Action Research Program (SHARP) to analyze more than 20 indicators of gun-violence in over 3000 U.S. counties between 2014 and 2017.
- Aggregated county-level characteristics and gun-violence records from government, private and non-profit sources to produce a diverse data set for analysis. Analyzed indicator interactions using regression and classification trees.

- Produced Bayesian spatial analyses in R for publication showing the association between income disparity and gun violence across 4 separate years. Provided figures illustrating associations and identifying outlier states and counties.

DISSERTATION WORK

A Bayesian Method for Adverse Effects Estimation in Observational Studies with Truncation by Death

- Developed Bayesian method for estimating the effect of an intervention on adverse event in the presence of mortality using a composite ordinal outcome.
- Led preparation of manuscript that describes the use of propensity scores and multiple imputation methods, a simulation study and a real data analysis on the toxicity profile of diabetes medications on over 1,000 nursing home residents.

A Latent Principal Stratification Method to Address Cluster and Individual Noncompliance in Cluster RCTs

- Developing principal stratification method for estimation in pragmatic cluster RCTs that accounts for latent similarities in compliance and characteristics among clusters. Allows researchers to estimate any estimand, such as ATE or CACE, in tandem using Bayesian sampling.
- Applying Bayesian mixture modeling and MCMC methods to estimate the effect of a personalized music intervention on the behaviors of over 1,000 nursing home residents with dementia while stratifying nursing home's by implementation levels and imputing unobserved compliance for individual's assigned to control.

A Two-Stage Adaptive Sequential Design for RCTs Subject to Noncompliance

- Proposing a new, two-stage, adaptive design for randomized trials subject to noncompliance that allows for interim analyses and adjustment of treatment implementation to improve adherence.
- Implementing Bayesian power prior to incorporate information from first stage to increase power of second stage while improving adherence with enhanced intervention.

PUBLICATIONS

Bitar R., Bozal S. B., **Sisti, A.**, ..., & Cornman-Homonoff J. (2024). Effect of Filtered Blood Return on Outcomes of Pulmonary Aspiration Thrombectomy. *Journal of Vascular and Interventional Radiology: JVIR*, S1051-0443.

Kaczynski M., Vassilopoulos A., Vassilopoulos S., **Sisti, A.**, ..., & Mylonakis E. (2023). Temporal Trends and Characteristics Associated with Racial, Ethnic, and Sex Representation in COVID-19 Clinical trials: A Systematic Review and Meta-Analysis, *Contemporary Clinical Trials*, 143, 107578.

Sisti, A., ..., & McCreedy, E. M. (2023). Using structured observations to evaluate the effects of a personalized music intervention on agitated behaviors and mood in nursing home residents with dementia: results from an embedded, pragmatic randomized controlled trial. *The American Journal of Geriatric Psychiatry.* 32(3), 300-311.

McCreedy, E. M., **Sisti, A.**, ... & Mor, V. (2022). Pragmatic Trial of Personalized Music for Agitation and Antipsychotic Use in Nursing Home Residents With Dementia. *Journal of the American Medical Directors Association*, 23(7), 1171-1177.

Johnson, B. T., **Sisti, A.**, ..., & Matos, M. (2021). Community-level factors and incidence of gun violence in the United States, 2014–2017. *Social Science & Medicine*, 280, 113969.

Majumdar, R., Mariano, P., Panzo, H., Peng, L., & **Sisti, A.** (2020). Lyapunov exponent and variance in the CLT for products of random matrices related to random Fibonacci sequences. *Discrete & Continuous Dynamical Systems-Series B*, 25(12).

SUBMITTED

Sisti A., Zullo, A., Gutman R. (2023). A Bayesian Method for Adverse Effects Estimation in Observational Studies with Truncation by Death. (Under review at *Statistical Methods in Medical Research*).

Conard R., Baier R., **Sisti A.**, Dionne, L., McCreedy E. (2023). Resident and Nursing Home Factors Associated with Adherence to a Personalized Music Intervention: Secondary Analyses from Music & MEmory: A Pragmatic TRial for Nursing Home Residents With Alzheimer's Disease (METRIcAL). (Under review at *Journal of Aging Research*).

McCreedy E., **Sisti, A.**, ..., & Thomas K. (2023). The effects of a personalized music intervention on depressive symptoms among nursing home residents with dementia: secondary analysis from Music & MEmory: A Pragmatic TRIal for Nursing Home Residents with ALzheimer's Disease (METRIcAL). (Under review at *American Journal of Alzheimers Disease & Other Dementias*).

WORKING

Sisti A., Gutman R. (2023). A Latent Principal Stratification Method to Address Cluster and Individual Noncompliance in Cluster RCTs.

Sisti A., Gutman R. (2023). A Two-Stage Sequential Design for RCTs Subject to Noncompliance.

INVITED RESEARCH TALKS

- *Multiplicative LLN and CLT and their Applications,* University of Connecticut Probability and Analysis Learning Seminar, Spring 2018
- *Option Pricing for the VIX Index Using a Historical Distribution,* Mathematics Institute for Secondary Teaching, WPI, Summer 2018

CONTRIBUTED RESEARCH TALKS

- *Using structured observations to evaluate the effects of a personalized music intervention on agitated behaviors and mood in nursing home residents with dementia: results from an embedded, pragmatic randomized controlled trial*, Academy Health Annual Research Meeting, "Caring for People Living with Dementia" Session, June 2023
- *A Bayesian Method for Causal Inference in Observational Studies Truncated by Death Using Composite Ordinal Outcomes*, Duke Industry Statistics Symposium (DISS 2023), Lighting Talk Session 1, March 2023.
- *A Bayesian Method for Causal Inference in Observational Studies Truncated by Death Using Composite Ordinal Outcomes*, 34th New England Statistical Symposium, Speed Presentation, August 2021.
- *A Bayesian Method for Causal Inference in Observational Studies Truncated by Death Using Composite Ordinal Outcomes*, Joint Statistical Meetings, Speed Session, August 2021.
- *The Black-Scholes Formula Using a Central Limit Theorem Argument,* AMS Special Session in Research in Applied Mathematics by Undergraduate and Post-Baccalaureate Students, JMM, January 2018.
- *Multiplicative LLN and CLT and their Applications,* Smith College Annual Students in Mathematics Conference, Fall 2017.
- *Multiplicative LLN and CLT and their Applications,* The 5th Northeast Mathematics Undergraduate Research Mini-Symposium, Summer 2017.

POSTERS

- *A Bayesian Method for Causal Inference in Observational Studies Truncated by Death Using Composite Ordinal Outcomes*, 34th New England Statistical Symposium, Speed Poster Presentation, August 2021.
- *Option Pricing for the VIX and TYVIX Indexes Using a Risk-Neutral Historical Distribution,* Frontiers in Undergraduate Research, University of Connecticut, April 2019.
- *Option Pricing for the VIX and TYVIX Indexes Using a Risk-Neutral Historical Distribution*, MAA Undergraduate Poster Session, JMM, January 2019
- *The Black-Scholes Formula Using a Central Limit Theorem Argument,* Frontiers in Undergraduate Research, University of Connecticut, April 2018.

- *The Black-Scholes Formula Using a Central Limit Theorem Argument,* MAA Undergraduate Poster Session, JMM, January 2018.

Relevant Coursework

- Big Data and Statistical Learning
- Causal Inference and Missing Data
- Bayesian Statistical Methods
- Applied Multivariate Analysis
- Survival Analysis
- Practical Data Analysis
- Introduction to Health Decision Analysis
- Linear & Generalized Linear Models
- Applied Time Series Analysis
- Statistical Inference I & II
- Financial Programming and Modeling
- Structural and Obj Oriented Programming
- Probability Theory and Stochastics I & II
- Mathematical Statistics I & II
- Advanced Financial Mathematics
- Real Analysis

Funding

- Summer 2018 research was supported in part by NSF Grant DMS-1757685 , and the Center for Industrial Mathematics and Statistics at WPI
- Summer 2017 research was supported in part by NSF Grant DMS-1262929.

Honors and Awards

- *Delta Omega Honor Society in Public Health* Epsilon Iota Chapter, Brown University, 2023
- *New England Scholar.* University of Connecticut, 2017.
- *Sophomore Honors Award.* University of Connecticut, 2017.
- *Babbidge Scholar.* University of Connecticut, 2016.
- *UConn Academic Excellence Scholarship.* University of Connecticut, 2015.

Related Experience

Teaching Assistant, Brown University **2019 - Present**

- Coordinator of four lab sections for PHP 1501, Essentials of Data Analysis
- Teaching assistant for PHP 2516, Applied Longitudinal Data Analysis and PHP 2517, Applied Multilevel Data Analysis

Workshop Coordinator, Brown University School of Public Health **2021 - Present**

- Lead and organized a week-long introductory statistics and programming workshop for incoming biostatistics graduate students
- Created virtual and in-person learning activities to facilitate the introduction to relevant statistical topics

Student Researcher, Worcester Polytechnic Institute Mathematics REU Program **Summer 2018**

- Worked alongside peers, faculty and industrial sponsors to conduct research in Financial Mathematics and Statistics.
- Collaborated with fellow REU members to construct an options pricing model for the VIX and TYVIX Indexes
- Prepared weekly and final presentations on completed research and future goals.

Student Researcher, UConn Mathematics REU Program **Summer 2017**

- Worked alongside graduate assistants and faculty to conduct research in the field of probability theory and stochastic processes.
- Collaborated with peers to write papers on the applications of Multiplicative LLN and CLT in the setting of Lyapunov Exponents and a derivation of the Black-Scholes option pricing formula using the Lindeberg-Feller Central Limit Theorem.
- Prepared weekly and final presentations on completed research and future goals.

Peer Tutor, Quantitative Learning Center **2016 - 2019**

- Facilitated individual and group study sessions in undergraduate mathematics and statistics.
- Collaborated with Graduate Assistants to improve tutoring methods and problem solving skills.
- Attended workshops to enhance knowledge of fundamental mathematical concepts and tutoring strategy.

Acknowledgements

As I reflect on my academic and professional pursuits that have culminated in this dissertation, I can't help but think of the many people who have made achieving my goals possible and, at the same time, a truly positive experience. There are countless individuals that deserve some share of the credit for this achievement, but I can only mention a few.

I am grateful for the tremendous amount of mentorship and guidance I have received from my dissertation advisor, **Roee Gutman**, who has supported me through the challenges I have faced and trained me to be a rigorous and thoughtful statistician. Our weekly meetings often consisted of long discussions about statistical theory and practice, and these conversations have equipped me with the tools to independently develop statistical methods in my professional career.

I would like to thank **Vince Mor**, **Ellen McCreedy**, and **Constantine Gatsonis** for being on my dissertation committee. Vince entrusted me with the statistical analyses for the METRIcAL trial, where I learned more about statistics than I could have in any class. Ellen was a second advisor to me, and I am grateful for her support and guidance which allowed me to publish my first, first-authored paper. Constantine has been a fantastic teacher and mentor, who helped make the School of Public Health a place I always enjoyed coming to.

Thanks to the **faculty, staff and students of the Biostatistics Department** who made being a doctoral student a fun and memorable experience. A special thanks to my cohort-mates **Patrick Gravelle** and **Nick Lewis** who will be my friends for life.

I would also like to thank **my friends and family** for always supporting and encouraging me. A special thanks to **my Mom**, **Ann** and **my Dad**, **Tony**, who have paved the way to this achievement. They have prioritized education at every stage of my life and are the reason that I have always believed I could accomplish whatever I put my mind to. **My sister**, **Julia**, and **my girlfriend**, **Gisella**, have been the best support system I could have asked for through this process.

Finally, I would like to thank **Dr. Dean Bajorin and his team**, as well as **Dr. Joel Sheinfeld and his team** for guiding me through a major life challenge while I worked on this dissertation. I wouldn't be here without the fantastic group of individuals who have helped me and continue to do so.

Contents

List of Tables

List of Figures

Introduction

Randomized controlled trials (RCTs) are often considered the gold standard for investigating the effects of an intervention. However, selective recruitment criteria often exclude vulnerable populations from RCTs, and tightly controlled designs make it difficult to determine how an intervention would perform in a more practical setting. To address these challenges, researchers typically rely on observational studies or pragmatic randomized controlled trials (PRCTs). Observational studies use data that has already been collected to analyze the effects of an intervention, and PRCTs are designed to assess an intervention's performance in routine clinical practice.

Designing and analyzing observational studies and PRCTs may require additional statistical considerations related to intercurrent events and noncompliance. Intercurrent events are events that occur after the study has been initiated that may affect the existence or interpretation of an outcome. In studies that analyze the effects of an intervention in a vulnerable population, such as elderly or severly ill patients, intercurrent events events are more likely to occur. Noncompliance is when an individual is assigned to an intervention but does not take it. In PRCTs, noncompliance is more common because the administration of the intervention is typically done by medical providers or the patients themselves.

The objective of this dissertation is to develop Bayesian design and analysis methods that address statistical challenges related to intercurrent events and noncompliance in observational studies and PRCTs. For observational studies, we develop a composite ordinal outcome and Bayesian analysis method to account for death before follow up when analyzing adverse event side effects of medication. For PRCTs, we propose a Bayesian design and analysis method that enables researchers to augment an intervention at an interim stage to improve compliance while preserving statistical power. For pragmatic cluster randomized controlled trials (PCRCTs), we design and

implement a latent pricipal stratification method for analyzing studies with one-sided partial noncompliance at the cluster level and binary noncompliance at the individual level. We describe each of the specific aims in further detail below:

1. **A Bayesian Method for Adverse Effects Estimation in Observational Studies with Truncation by Death**

   Death among subjects is common in observational studies evaluating the causal effects of interventions among geriatric or severely ill patients. High mortality rates complicate the comparison of the prevalence of adverse events (AEs) between interventions. This problem is often referred to as outcome "truncation" by death. A possible solution is to estimate the survivor average causal effect (SACE), an estimand that evaluates the effects of interventions among those who would have survived under both treatment assignments. However, because the SACE does not include subjects who would have died under one or both arms, it does not consider the relationship between AEs and death. We propose a Bayesian method which imputes the unobserved mortality and AE outcomes for each participant under the intervention they did not receive. Using the imputed outcomes we define a composite ordinal outcome for each patient, combining the occurrence of death and the AE in an increasing scale of severity. This allows for the comparison of the effects of the interventions on death and the AE simultaneously among the entire sample. We implement the procedure to analyze the incidence of heart failure among geriatric patients being treated for Type II diabetes with sulfonylureas or dipeptidyl peptidase-4 inhibitors.

2. **A Latent Principal Stratification Method to Address Cluster and Individual Noncompliance in Pragmatic Cluster RCTs**

   In pragmatic cluster randomized controlled trials (PCRCTs), noncompliance to the intervention can occur among patients or clusters. As a consequence, investigators often collect data that describe the binary compliance status of individuals and the cluster-level implementation of the intervention. One approach to assess the effects of an intervention while accounting for noncompliance is to estimate the compiler average causal effect (CACE). In PCRCTs, the CACE is defined for the group of clusters and individuals that both comply with their treatment assignments. The CACE, however, has not been generalized to continuous non-

compliance at the cluster-level and therefore may ignore the relationship between cluster-level implementation and the effects of the intervention. We propose a new Bayesian method to analyze PCRCTs with one-sided binary noncompliance at the individual-level and one-sided partial compliance at the cluster-level. This method relies on a Bayesian mixture model to classify clusters into latent compliance strata based on cluster's pre-treatment characteristics, its partial compliance status and individual-level outcomes. Because compliance is only observed in the treatment arm, the mixture model imputes the unobserved compliance for clusters and individuals assigned to the control arm. The procedure can estimate finite and super-population estimands within strata defined by a combination of the latent cluster-level compliance and individual-level compliance. We apply this procedure to analyze the METRIcAL trial: a multi-part, pragmatic, nursing home (NH) based trial evaluating the effects of a personalized music intervention on the agitated behaviors of nursing home residents with dementia.

3. **A Bayesian Two-stage Sequential Design for Pragmatic RCTs with One-Sided Noncompliance**

   Clinical trials subject to noncompliance may have lower statistical power, higher costs and longer duration. To account for this, noncompliance should be considered in a study's design phase. Two-stage designs have been proposed that allow researchers to assess the efficacy of treatments at an intermediate point. However, these designs do not provide researchers with the opportunity to adjust treatment delivery to improve compliance when high levels of noncompliance have occurred. We propose a novel, Bayesian two-stage sequential design for PRCTs that addresses one-sided binary noncompliance. In the design, compliance and effectiveness data are accumulated during the first stage and used for interim analyses prior to initiating the second stage. Based on the interim analysis, the trial may be stopped for efficacy, or a new trial may be implemented with a multi-component intervention that improves compliance. The analyses of this design rely on Bayesian models that utilize a power prior distribution to incorporate information from Stage 1 in the analysis of the Stage 2 in order to improve the statistical power. We examine the operating characteristics of the proposed method compared to a single-stage design and evaluate its sensitivity to misspecifications

using simulations. In addition, we simulate a trial using the two-stage design with real data from METRIcAL: a multi-part, pragmatic, nursing home (NH) based trial evaluating the effects of a personalized music intervention on the agitated behaviors of nursing home residents with dementia.

Chapter 1

A Bayesian Method for Adverse Effects Estimation in Observational Studies with Truncation by Death

1.1 Introduction

In the United States in 2023, about 1 in 10 individuals had diabetes, with approximately 90-95% of these cases classified as type 2 diabetes mellitus (T2DM) (Centers for Disease Control and Prevention, 2023). The disease is even more prevalent among older adults (Kirkman et al., 2012). Sulfonylureas (SUs) are among the most common stand alone treatments for nursing home (NH) residents diagnosed with T2DM; however, usage rates of dipeptidyl peptidase-4 inhibitors (DPP4Is) have increased in recent years (Munshi et al., 2016; Zullo, Dore, Daiello, Baier, Gutman, Gifford and Smith, 2016). Studies among the general population have been inconclusive on the relative safety and efficacy of SUs and DPP4Is. Some studies have shown that SUs are associated with increased cardiovascular and hypoglycemia risk, while DPP4Is are associated with an increased risk of heart failure related hospitalizations (Scheen, 2018; Scirica et al., 2013; Kim et al., 2019). DPP4Is have also been found to be more effective for glycemic control compared to SU (Scheen, 2018; Fadini et al., 2018; Kim et al., 2019). Most of these studies have not been conducted among NH residents who are disproportionately effected by T2DM and are susceptible to negative side

effects of medications (Zullo et al., 2021, 2022; Zullo, Dore, Gutman, Mor and Smith, 2016; Munshi et al., 2016).

NH residents are often excluded from randomized studies because of efficacy concerns, ethical considerations and practical challenges of conducting clinical trials in nursing homes (Watts, 2012). In addition, because adverse events are generally rare, randomized studies with adverse events as the primary outcomes require large sample sizes to identify significant effects. Observational studies can address these limitation by relying on larger samples that include populations not commonly recruited to clinical trials. However, in observational studies, the decisions on which patients receive the interventions are often confounded with the outcomes.

One observational study that compared SUs and DPP4Is among older adults was a national retrospective cohort study of long-stay NH residents (Zullo et al., 2021). The study estimated the effects of DPP4Is versus SUs on severe adverse glycemic events, cardiovascular events, and death. The study concluded that DPP4I users had similar incidence of cardiovascular events and death, and a lower incidence of hypoglycemic events, over a 1-year time frame compared to SU users. However, many patients died before the 1-year follow up time was complete. This study considered death and the adverse event separately, which does not allow for a joint evaluation of the relationships between DPP4Is and SUs, adverse events and deaths. It also ignores the fact that a person is no longer at risk for an adverse event after death.

Estimating the effects of interventions on the occurrence of adverse events in the presence of mortality complicates the analysis because the occurrence of an adverse event at follow-up is not defined for individuals who die before the end of the follow up period (Colantuoni et al., 2018; McConnell, Stuart and Devaney, 2008; Rubin, 2006; Zhang and Rubin, 2003). One approach to address "truncation" by death is to estimate the intention to treat effect (ITT), which considers the difference in proportions of the adverse event that occurred between the two study arms. This approach utilizes the entire sample, but differences in mortality across interventions may result in misleading conclusions.

A different approach to address truncation by death relies on the principal stratification framework to estimate the survivor-average causal effect (SACE) (Frangakis and Rubin, 2002*a*; Rubin, 2006; Zhang and Rubin, 2003). This approach stratifies participants by mortality status under both interventions, and estimates the effect of the intervention on the outcome among participants who

would have survived under both interventions. Many applications that estimate SACE rely on the monotonicity and stochastic dominance assumptions (Rubin, 2006). The monotonicity assumption states that survival under the active intervention is the same or better than survival under the control intervention. However, in studies that compare two active interventions, this assumption may not be plausible. The stochastic dominance assumption requires that, on average, those who would survive under both interventions are less likely to experience the adverse event than those who would have died under either intervention. This assumption may also be violated in studies comparing two active interventions, because longer exposure to an intervention may lead to higher probability of experiencing an adverse event. Several approaches that modify these assumptions have been proposed to estimate the SACE in randomized and observational studies (Hayden, Pauler and Schoenfeld, 2005; Long and Hudgens, 2013; Wang, Zhou and Richardson, 2017), but procedures that use the SACE ignore individuals who would have died under either of the treatments. This includes individuals who experienced an adverse event that leads to death under one of the interventions. As such, the SACE may not provide a complete toxicity profile for the two interventions.

Another approach to address the truncation problem is to define a composite outcome that combines death and the adverse event (Cordoba et al., 2010). This is commonly implemented by defining a binary outcome that is equal to 1 if either death or the adverse event occur, and zero otherwise. This composite outcome is well defined for all participants in the study, and common methods for estimating the treatment effects with binary outcomes can be applied in randomized (Brown and Ezekowitz, 2017; Terheyden et al., 2021) and observational studies (Saager et al., 2021; Johnson and Tsiatis, 2004; Gutman and Rubin, 2015*a*). Interpretation of the composite approach is practically relevant when the adverse event and mortality have similar utility for all patients, and when the interventions influence both mortality and the adverse event in the same direction. However, when the rate of either the adverse event or mortality is high for one of the interventions, it is difficult to assess differences between the interventions on the less prevalent outcome (Colantuoni et al., 2018; Goldberg et al., 2014).

Another solution is to define a composite ordinal outcome that combines the occurrence of death and the adverse event in an increasing scale of severity. In randomized clinical trials, the desirability of outcome ranking (DOOR) was proposed as a possible ordinal outcome that combines the occurrence of multiple outcomes and ranks these combinations by their desirability (Evans

et al., 2015). The DOOR distributions are compared between the study arms by estimating the probability that an individual selected at random will have a better DOOR if assigned to the active intervention. Non-parametric tests and interval estimates can be used to evaluate whether this probability exceeds 50%. Analysis based on DOOR has been restricted to randomized studies and to compare marginal distributions without adjustments for covariates. Another composite statistic that has been proposed for randomized controlled trials is the win ratio (Pocock et al., 2011). Among matched pairs of units, one receiving the active intervention and one receiving the control intervention, the win ratio estimates how often a unit under the intervention fared better than a unit under the control. A "winner" is assigned in each matched pair based on which unit had the more desirable outcome first, or lasted longer without experiencing an adverse outcome. The number of "winners" under each intervention is computed and used to calculate the win ratio. The win ratio does not adjust for covariates and it relies on time to event data. In addition, when risk-matching among patients is not possible, obtaining confidence intervals and p-values for the win ratio is complex (Pocock et al., 2011).

We propose a new Bayesian method for observational studies that fits two conditional models for the adverse event and death in each study arm. Based on these models, we multiply impute the unobserved outcomes for each participant under the opposite intervention and construct a composite ordinal outcome. The proposed method generates statistically valid point and interval estimates for any causal estimand with ordinal outcomes. We apply this method to compare the 180-day risk of death and heart failure following the initiation of a SU or DPP4I among NH residents with T2DM. We show that a randomly selected patient is estimated to have a 5% higher risk of having a worse outcome under DPP4I than of having a worse outcome under SU, but this result was not significant at the 5% nominal level.

1.2 Methods

1.2.1 Notation and Definitions

In a sample from a population of N units indexed $i = 1, .., n \leq N$, let n_0 be the number of units receiving the control intervention and $n_1 = n - n_0$ units receiving the active intervention, where the control intervention may represent another active intervention. Let $W_i \in \{0, 1\}$ be an

indicator that is equal to 0 if unit i received the control intervention, and 1 if it received the active intervention. We define $\mathbf{Y}_i(0) = \{A_i(0), D_i(0)\}$ to be a joint outcome indicating whether an adverse event, $A_i(0)$, or death, $D_i(0)$, occurred between the initiation of the control intervention and the end of the study period for unit i. Similarly, let $\mathbf{Y}_i(1) = \{A_i(1), D_i(1)\}$ be the joint outcome for the active intervention. We write $\mathbf{Y}(0) = \{\mathbf{Y}_i(0)\}_{i=1}^N$ and $\mathbf{Y}(1) = \{\mathbf{Y}_i(1)\}_{i=1}^N$ to denote the collection of joint outcomes for the sample. Because individuals can receive only one of the interventions at a specific time point, only one of $\mathbf{Y}_i(0)$ and $\mathbf{Y}_i(1)$ can be realized and observed (Rubin, 1978*a*). Assuming the stable unit treatment value assumption (SUTVA) (Rubin, 1990*a*), the observed and missing outcomes are,

$$\mathbf{Y}_i^{obs} = \mathbf{Y}_i(1)W_i + \mathbf{Y}_i(0)(1 - W_i)$$
$$\mathbf{Y}_i^{mis} = \mathbf{Y}_i(1)(1 - W_i) + \mathbf{Y}_i(0)(W_i).$$

For each unit, we also observe a set of P covariates, $\mathbf{X}_i = \{X_i^1, ..., X_i^P\}$, that are recorded prior to initiation of the intervention.

To estimate the causal effects of the intervention, we need to identify the probability that units received the active intervention, $P(\mathbf{W}|\mathbf{Y}(1), \mathbf{Y}(0), \mathbf{X})$, commonly referred to as the assignment mechanism (Rubin, 1978*a*). Assuming that the assignment mechanism is strongly ignorable (Rosenbaum and Rubin, 1983*a*), and that the units are independent, we have

$$P(\mathbf{W}|\mathbf{Y}(0), \mathbf{Y}(1), \mathbf{X}) = \prod_i^N P(W_i = 1|\mathbf{Y}_i(1), \mathbf{Y}_i(0), \mathbf{X}_i, \phi) = \prod_i^N P(W_i = 1|\mathbf{X}_i, \phi) = \prod_i^N e(\mathbf{X}_i),$$

where ϕ comprises the parameters governing this distribution, and $e(\mathbf{X}_i)$ is the propensity score for unit i (Rosenbaum and Rubin, 1983*c*). In randomized controlled trials, the propensity score is known as part of the design phase. In observational studies, the probability of receiving each of the interventions is unknown, and only an estimate of it, $\hat{e}(\mathbf{X}_i)$, is available. Multiple procedures have been described for estimating the propensity score (Westreich, Lessler and Funk, 2010; Stuart, 2010; Guo, Fraser and Chen, 2020). We assume that the propensity scores have been estimated using any procedure selected by the investigators in the design phase of the study. Our method will utilize the estimated propensity score at the analysis stage.

1.2.2 Composite Ordinal Outcome and Estimands

Estimands, which are functions of unit-level potential-outcomes, are used to summarize the effects of an intervention across a population of interest (Gutman and Rubin, 2015*a*; Rubin, 1978*a*). Possible estimands that are defined for the entire population and consider death and the adverse event outcome separately include the difference in proportions, $Pr(A_i(1) = 1) - Pr(A_i(0) = 1)$ and $Pr(D_i(1) = 1) - Pr(D_i(0) = 1)$, and their corresponding risk and odds ratios. These estimands do not consider the correlations between the outcomes, which may lead to misleading conclusions when mortality is differentiable between the two treatments. Other possible estimands to address this issue are based on a binary composite outcome. A binary composite outcome is equal to 1 if either of the events occur and zero otherwise. For example, the difference in proportions estimand for this outcome is $Pr(A_i(1) = 1 \cup D_i(1) = 1) - Pr(A_i(0) = 1 \cup D_i(0) = 1)$. This estimand may present an incomplete picture when the probability of mortality is much larger than the probability of adverse event and vice versa. A different estimand that attempts to address this limitation is the SACE. The SACE estimates the effects of the intervention on the adverse event among the subpopulation that survives under both interventions: $Pr(A_i(1) = 1 | D_i(0) = D_i(1) = 0) - Pr(A_i(0) = 1 | D_i(0) = D_i(1) = 0)$. This estimand may not present the entire toxicity profile of intervention, when mortality is higher for one of the interventions.

To address these limitations we define a composite ordinal potential outcome $G_i(w)$ that combines deaths and adverse events in an increasing scale of severity:

$$G_i(w) = \begin{cases} 1 & \text{if } \mathbf{Y}_i(w) = \{0, 0\} \\ 2 & \text{if } \mathbf{Y}_i(w) = \{1, 0\} \\ 3 & \text{if } \mathbf{Y}_i(w) = \{0, 1\} \\ 4 & \text{if } \mathbf{Y}_i(w) = \{1, 1\} \end{cases}.$$

This definition assumes that death is worse than undergoing an adverse event, but different rankings of severity and additional ordinal levels can be used in other situations. This outcome is well-defined on the whole sample, and enables researches to consider the associations between adverse events and death.

Let $p_k(w) = Pr(G_i(w) = k)$ for $k = 1, ..., 4$; possible estimands for the composite ordinal outcome are $p_k(1) - p_k(0)$, $\frac{p_k(1)}{p_k(0)}$ and $\frac{p_k(1)/(1-p_k(1))}{p_k(0)/(1-p_k(0))}$, which describe the difference in probabilities, the relative risk and the odds ratio of being in level k of the ordinal outcome under the active intervention and the control intervention, respectively.

A different estimand is the distributional estimand for ordinal outcomes (Ju and Geng, 2010),

$$\Delta_j = Pr\left(G_i(1) \leq j\right) - Pr\left(G_i(0) \leq j\right) = \sum_l \sum_{k \geq j} p_{kl} - \sum_k \sum_{l \geq j} p_{kl} \quad 1 \leq j \leq 4, \tag{1.1}$$

where $p_{kl} = Pr(G_i(1) = k, G_i(0) = l)$ for $k, l \in \{1, 2, 3, 4\}$. This estimand is comprised of four values corresponding to each level of the ordinal outcome, with positive Δ_j indicating that an individual is more likely to experience level j or lower under the active intervention than the control.

Two other estimands were proposed by Lu et. al.(Lu, Ding and Dasgupta, 2018),

$$\tau_{10} = Pr\left(G_i(1) \geq G_i(0)\right) = \sum_{k \geq l} \sum p_{kl}, \quad \text{and} \quad \kappa_{10} = Pr\left(G_i(1) > G_i(0)\right) = \sum_{k > l} \sum p_{kl}.$$

Estimand τ_{10} describes the probability that an individual has an outcome under the active intervention that is greater than or equal to their outcome under the control intervention, while κ_{10} describes the same quantity but for the strictly greater case. The estimands τ_{01} and κ_{01} correspond to τ_{10} and κ_{10}, but with the ordinal outcome being larger under the control intervention than the active intervention. The estimands $\kappa_{10} - \kappa_{01}$ and $\frac{\kappa_{10}}{\kappa_{01}}$ describe the difference in probabilities and relative risk of an individual doing strictly worse under the active intervention compared to doing strictly worse under the control intervention, respectively. The estimands τ_{10} and κ_{10} can be used to define the Mann-Whitney estimand (Agresti, 2010; Bross, 1958; Follmann et al., 2020),

$$U_{10} = Pr\left(G_i(1) > G_i(0)\right) + \frac{1}{2} Pr\left(G_i(1) = G_i(0)\right) = \frac{\kappa_{10} + \tau_{10}}{2}.$$

This estimand describes the probability that an individual selected at random has inferior outcomes under the active intervention compared to the control intervention, and $\frac{U_{10}}{1-U_{10}}$ describes the odds of this scenario.

Let, $\pi_{kl}(w) = Pr\left(G_i(1-w) = l | G_i(w) = k\right)$, the conditional probability matrix under interven-

tion w is another possible estimand,

$$\Pi(w) = \begin{pmatrix} \pi_{11}(w) & \pi_{12}(w) & \pi_{13}(w) & \pi_{14}(w) \\ \pi_{21}(w) & \pi_{22}(w) & \pi_{23}(w) & \pi_{24}(w) \\ \pi_{31}(w) & \pi_{32}(w) & \pi_{33}(w) & \pi_{34}(w) \\ \pi_{41}(w) & \pi_{42}(w) & \pi_{43}(w) & \pi_{44}(w) \end{pmatrix}$$

This matrix describes the probabilities that an individual would have outcome l under intervention $1 - w$ given that they have outcome k under intervention w. Individual probabilities within the matrices can be subtracted, $\Pi(1) - \Pi(0)$ or divided $\Pi(1)/\Pi(0)$ to obtain risk differences or risk ratios corresponding to specific events, respectively.

1.2.3 Bayesian Imputation of Counterfactual Outcomes

Deriving interval estimates for the estimands in Section 2.2 can be analytically complex, and Lu et al.(Lu, Ding and Dasgupta, 2018) describe sharp bounds to estimate τ_{10}, κ_{10} and U_{10}. We view causal inference from a missing data perspective (Rubin, 1976), and extend a method introduced by Gutman and Rubin (Gutman and Rubin, 2013a, 2015a) to multiply impute the unobserved potential outcomes. Let $\gamma = \nu(\mathbf{Y}(1), \mathbf{Y}(0))$, be a predefined estimand. A Bayesian approach to estimating γ will condition on the observed data $\mathbf{Y}^{obs}, \mathbf{W}, \text{and } \mathbf{X}$ to estimate

$$\begin{aligned} f(\gamma|\mathbf{Y}^{obs}, \mathbf{X}, \mathbf{W}) &= \int f(\gamma|\mathbf{Y}^{obs}, \mathbf{Y}^{mis}, \mathbf{X}, \mathbf{W}) f(\mathbf{Y}^{mis}|\mathbf{Y}^{obs}, \mathbf{X}, \mathbf{W}) d\mathbf{Y}^{mis} \\ &= \int f(\gamma|\mathbf{Y}^{obs}, \mathbf{Y}^{mis}, \mathbf{X}, \mathbf{W}) f(\mathbf{Y}^{mis}|\mathbf{Y}^{obs}, \mathbf{X}) d\mathbf{Y}^{mis}. \end{aligned}$$

Because the distribution of γ is a function of the potential outcomes, $\mathbf{X}$, and $\mathbf{W}$, only the conditional distribution of the missing outcomes, $f(\mathbf{Y}^{mis}|\mathbf{Y}^{obs}, \mathbf{X})$, is required to estimate the distribution of γ (Gutman and Rubin, 2015a). This distribution can be written as the product of the conditional distribution of the adverse event given the observed values, and the conditional distribution of death given the observed values and the adverse event,

$$f(\mathbf{Y}^{mis}|\mathbf{Y}^{obs},\mathbf{X}) = f(\mathbf{D}^{mis},\mathbf{A}^{mis}|\mathbf{D}^{obs},\mathbf{A}^{obs},\mathbf{X})$$
$$= f(\mathbf{D}^{mis}|\mathbf{A}^{mis},\mathbf{D}^{obs},\mathbf{A}^{obs},\mathbf{X})f(\mathbf{A}^{mis}|\mathbf{D}^{obs},\mathbf{A}^{obs},\mathbf{X}). \quad (1.2)$$

In observational studies with an unconfounded assignment mechanism, Gutman & Rubin (2015) proposed to estimate the response surface, $E_{\theta_w}(Y(w)|X,\theta_w) = h_w(X,\theta_w)$, using a spline along the propensity score and linear adjustments for the other covariates. This method relies on subclasses of the propensity scores to balance the covariates across treatment groups while increasing efficiency using splines. In addition, it relies on linear adjustments for the components of the covariates orthogonal to the propensity score to gain additional efficiency and control for minor residual imbalances. Formally, we approximate the conditional distributions in Equation (1.2) as

$$\tilde{Pr}(A_i(w) = 1|\mathbf{X}_i,\theta_w^a) = g^{-1}(f_w^a(g(\hat{e}(\mathbf{X}_i)),\mathbf{B}_w^a) + \mathbf{X}_i^*\beta_w^a) \quad (1.3)$$

and,

$$\tilde{Pr}(D_i(w) = 1|\mathbf{X}_i,\theta_w^d,A_i(w)) = g^{-1}(f_w^d(g(\hat{e}(\mathbf{X}_i)),\mathbf{B}_w^d) + \mathbf{X}_i^*\beta_w^d + \eta_w A_i(w)), \quad (1.4)$$

where g is the logit function, f_w^a and f_w^d are splines over the propensity scores, $\theta_w^a = \{\mathbf{B}_w^a,\beta_w^a\}$ and $\theta_w^d = \{\mathbf{B}_w^d,\beta_w^d,\eta_w\}$ are the sets of unknown parameters in Equations (1.3) and (1.4), respectively, and $\mathbf{X}_i^* = (X_i^1,...,X_i^{P-1})$, with X_{iP} being omitted.

Assuming that $\mathbf{Y}_i(1)$ and $\mathbf{Y}_i(0)$ are conditionally independent given $\mathbf{X}_i$, θ_w^a and θ_w^d, and that the prior distributions of θ_w^a and θ_w^d are independent, the posterior distributions of $\{\theta_1^a,\theta_1^d\}$ and $\{\theta_0^a,\theta_0^d\}$ are independent. This assumption is formally known as no contamination of imputation across treatments (Rubin, 2007), and is made by many causal inference methods (Gutman and Rubin, 2015*a*). In addition, we assume that $P(\theta_w^a,\theta_w^d) \propto P(\theta_w^a)P(\theta_w^d)$, then θ_1^a and θ_1^d, and θ_0^a and θ_0^d, have independent posterior distributions. Let $\psi_w^a(\theta_w^a|\mathbf{D}^{obs},\mathbf{A}^{obs},\mathbf{X})$ and $\psi_w^d(\theta_w^d|\mathbf{D}^{obs},\mathbf{A}^{mis},\mathbf{A}^{obs},\mathbf{X})$ be the posterior densities of the parameters in equations (1.3) and (1.4) under treatment w for adverse events and death, respectively. To obtain $m = 1,\ldots,M$ samples of $\mathbf{Y}(0)^{mis}$, we sample at iteration m, $\theta_0^{a(m)}$ and $\theta_0^{d(m)}$ from ψ_0^a and ψ_0^d, respectively. Given $\theta_0^{a(m)}$ and $\theta_0^{d(m)}$, we sample $\mathbf{Y}_i(0)^{mis(m)}$ from the posterior predictive distribution implied by Equations (1.3) and (1.4). Similarly, $\mathbf{Y}_i(1)^{mis}$ can be sampled from the corresponding posterior predictive distributions.

When estimating the posterior distribution of γ, it is important to distinguish between finite-sample and super-population estimands, denoted γ^{fp} and γ^{sp}, respectively. To estimate γ^{sp} we need to consider the distribution of the covariates in the super-population, $F(\mathbf{X}_i)$, which is not necessary when estimating γ^{fp}. Estimating $F(\mathbf{X}_i)$ is part of the design phase of observational studies because it does not involve any outcomes. When $\mathbf{X}$ is assumed to be a simple random sample from the population, the empirical distribution, $\hat{\mathbb{F}}_{\mathbf{X}_i}$, can be used to approximate the super-population distribution of the covariates. Under other sampling mechanisms, different approaches, such as weighting, may be required to estimate $F(\mathbf{X}_i)$. We now summarize the procedure for estimating γ:

1. Estimate the propensity score, $\hat{e}(X_i)$ for each of the n units in the sample, and partition these n units into K sub classes based on the quantiles of the estimated propensity scores.

2. Based on equations (1.3) and (1.4) estimate $\tilde{Pr}(A_i(w) = 1|\mathbf{X}_i, \theta_w^a)$, and $\tilde{Pr}(D_i(w) = 1|A_i(w), \mathbf{X}_i, \theta_w^d)$, respectively, within treatment groups.

3. Obtain M samples of θ_0^a, θ_0^d and θ_1^a, θ_1^d from their corresponding posterior distributions.

4. Using the m^{th} sample, $\tilde{\theta}_0^{a(m)}$, sample $A_i(0)$ from a Bernoulli distribution with probability $\tilde{Pr}(A_i(0) = 1|\mathbf{X}_i, \theta_0^{a(m)})$. Using $\tilde{\theta}_1^{a(m)}$ sample $A_i(1)$ for units with $W_i = 0$ from a Bernoulli distribution with probability $\tilde{Pr}(A_i(1) = 1|\mathbf{X}_i, \theta_1^{a(m)})$.

5. Using $\tilde{\theta}_0^{d(m)}$ and the sampled unobserved outcomes $\hat{\mathbf{A}}^{mis}$ from the previous step, impute the unobserved mortality for units with $W_i = 1$ by sampling from a Bernoulli distribution with probability $\tilde{Pr}(D_i(0) = 1|\mathbf{X}_i, \theta_0^{d(m)}, \hat{A}_i^{mis})$. Using $\tilde{\theta}_1^{d(m)}$, sample the unobserved death outcomes for units with $W_i = 0$ from a Bernoulli distribution with probability $\tilde{Pr}(D_i(1) = 1|\mathbf{X}_i, \theta_1^{d(m)}, \hat{A}_i^{mis})$.

6. Repeat steps 4 and 5 for $m = 1, \ldots, M$.

7. For $m = 1, \ldots, M$, let $\hat{\gamma}^{(m)} = \nu(\hat{\mathbf{Y}}(1)^{(m)}, \hat{\mathbf{Y}}(0)^{(m)})$ with $\hat{\mathbf{Y}}_i(w)^{(m)} = \{\hat{A}_i(w)^{(m)}, \hat{D}_i(w)^{(m)}\}$

where

$$\hat{A}_i(w)^{(m)} = \begin{cases} A_i^{obs} & \text{if } W_i = w \\ \hat{A}_i^{mis(m)} & \text{if } W_i \neq w \end{cases} \qquad \hat{D}_i(w)^{(m)} = \begin{cases} D_i^{obs} & \text{if } W_i = w \\ \hat{D}_i^{mis(m)} & \text{if } W_i \neq w \end{cases}$$

8. The point estimate of γ is $\hat{\gamma} = \frac{1}{M}\sum_{i=1}^{M}\hat{\gamma}^{(m)}$. Let the sampling variance of $\hat{\gamma}^{(m)}$ be $\hat{U}^{(m)}$ ($\hat{U}^{(m)}$ = 0 for finite-sample estimands). We let $\bar{U} = \frac{1}{M}\sum_{i=1}^{M}\hat{U}^{(m)}$ define the within imputation variance and $B = \frac{1}{M-1}\sum_{i=1}^{M}(\hat{\gamma}^{(m)} - \hat{\gamma})^2$ define the between imputation variance. The total estimate of the sampling variance for $\hat{\gamma}$ is $T = \bar{U} + (1 + \frac{1}{M})B$.

9. Posterior inferences are based on a t-distribution with ν_M degrees of freedom where $\nu_M = (M-1)\left(\frac{T}{(1+M^{-1})B}\right)^2$ (Rubin, 2004) . In small data sets, use the t-approximation derived by Barnard and Rubin (Barnard and Rubin, 1999) and also described by Yuan(Yuan, 2010) to estimate posterior intervals (Appendix Section A1.1).

When M is large enough, posterior interval estimation for finite-sample estimands can also be derived using the percentiles of the distribution of $\gamma^{(m)}$ where $m = 1, ..., M$. Some super-population estimands can be estimated using $\theta_0^{a(m)}, \theta_0^{d(m)}, \theta_1^{a(m)}, \theta_1^{d(m)}$, and omitting steps 4-9 in the estimation procedure. For example, the super-population average treatment effect on adverse events, $\gamma^{sp} = \gamma(\theta_1^a, \theta_0^a) = E_{\mathbf{X}_i}[\gamma(\theta_1^a, \theta_0^a, \mathbf{X}_i)|\theta_1^a, \theta_0^a]$, where

$$\begin{aligned}\gamma(\theta_1^a, \theta_0^a, \mathbf{X}_i) &= E_{A_i(1),A_i(0)}[A_i(1) - A_i(0)|\theta_1^a, \theta_0^a, \mathbf{X}_i] \\ &= \tilde{Pr}(A_i(1) = 1|\mathbf{X}_i, \theta_1^a) - \tilde{Pr}(A_i(0) = 1|\mathbf{X}_i, \theta_0^a).\end{aligned}$$

Assuming that the data originates from a simple random sample, and using the empirical distribution to approximate the covariate distribution in the super-population, we have

$$\gamma^{sp(m)} = \frac{1}{n}\sum_{i=1}^{n}\left(\tilde{Pr}(A_i(1) = 1|\mathbf{X}_i, \theta_1^{a(m)}) - \tilde{Pr}(A_i(0) = 1|\mathbf{X}_i, \theta_0^{a(m)})\right) = \gamma(\theta_1^{a(m)}, \theta_0^{a(m)}).$$

The posterior distribution of γ^{sp} can thus be obtained using the M samples of $\theta_0^{a(m)}, \theta_0^{d(m)}, \theta_1^{a(m)}$ and $\theta_1^{d(m)}$. Point and posterior interval estimates can be derived from this posterior distribution using the mean and quantiles of the distribution, respectively.

1.2.4 Sensitivity Analysis

The proposed Bayesian estimation procedure relies on the strong ignorability assumption, which cannot be tested with observed data. To assess the validity this assumption we describe an interpretable sensitivity analysis.

Let $\mathbf{Z} = \{Z_1, .., Z_n\}$ denote an unobserved covariate, independent from observed covariates, with $E[Z_i] = \mu_1^z * W_i + \mu_0^z * (1 - W_i)$. The parameters μ_1^z and μ_0^z denote the expected value of the unobserved covariate among those assigned to the active and control treatments, respectively. Without loss of generality, we set $\mu_1^z = 0$ so that μ_0^z describes the bias in $\mathbf{Z}$ between subjects assigned to receive the control intervention and subjects assigned to receive the active intervention. We define the following relationship between the estimated log odds of experiencing an adverse event and death, and the unobserved covariate:

$$\begin{aligned}
\text{logit}\left(\tilde{Pr}(A_i(w) = 1|\mathbf{X}_i, \theta_w^a)\right) =& f_w^a(g(\hat{e}(\mathbf{X}_i)), \hat{\mathbf{B}}_w^a) + \mathbf{X}_i^* \hat{\beta}_w^a + \delta_a \mathbf{Z}_i \\
\text{logit}\left(\tilde{Pr}(D_i(w) = 1|\mathbf{X}_i, \theta_w^d, A_i(w))\right) =& f_w^d(g(\hat{e}(\mathbf{X}_i)), \hat{\mathbf{B}}_w^d) + \mathbf{X}_i^* \hat{\beta}_w^d + \hat{\eta}_w A_i(w) + \delta_d \mathbf{Z}_i.
\end{aligned}$$

The parameter δ_a describes the linear change in the conditional log odds of experiencing the adverse event under both treatments due to a one unit change in the unobserved covariate $\mathbf{Z}$. δ_d defines the same relationship, but for the occurrence of death. To assess the sensitivity of estimates to violations of the strong ignorability assumptions, we incorporate $\mathbf{Z}$ into the estimation procedure for different μ_0^z, δ_a and δ_d.

1.3 Description of Data for Comparing Interventions for Type II Diabetes

To illustrate the method proposed in Section 1.2.3, we analyze an observational study that compares the effects of sulfonylureas (SUs) and dipeptidyl peptidase-4 inhibitors (DPP4Is) among NH residents with type 2 diabetes mellitus (T2DM) (Zullo, 2017). The data consists of U.S. NH residents aged 65 and older who initiated a DPP4I or SU between January 1, 2008 and September 30, 2010, and reside in a NH for at least three months preceding the initiation. After exclusions, the initial cohort included 1,064 new DPP4I users and 6,821 new SU users. The propensity score was

estimated using logistic regression with a set of 198 covariates extracted from the Minimum Data Set (MDS) (Centers for Medicare & Medicaid Services, 2021) and Medicare parts A (inpatient), B (outpatient) and D (prescription drug) claims data. A one-to-one matching procedure using the propensity score yielded 1,008 patients initiating DPP4I and 1,008 initiating SU. Although propensity score matching is not necessary to apply the proposed method, we rely on the final data set from this study and estimate average effects among the treated when conducting the data analysis. This ensured that only minor imbalances among observed covariates exists between the two treatment arms. The outcomes of this study were the occurrence of adverse cardiovascular events and mortality within 180 days of initiation.

1.4 Simulated Case Studies

We design simulated case studies in which we control the assignment mechanism and the treatment effects for two interventions to display the differences in interpretation between estimands and to show the operating characteristics of the proposed method. The simulated data is based on the matched subjects described in Section 1.3. Out of the 198 covariates, we selected three continuous covariates: number of comorbidities, Morris activities of daily living scale, and number of days per week using a diuretic, and two binary covariates: an indicator of treatment for skin conditions and an indicator of hypertension. We define the probability that unit i receives the active intervention as

$$Pr(W_i = 1|\mathbf{X}_i, \alpha) = u^{-1}(\mathbf{X}_i\alpha), \tag{1.5}$$

where u is a link function, $\mathbf{X}_i$ consists of the 5 covariates described above, and α is a coefficient vector. The expected probability that unit i experiences the adverse event under treatment w is

$$Pr(A_i(w) = 1|\mathbf{X}_i, \varphi_w^a, \xi_w^a) = u^{-1}(\varphi_w^a + \mathbf{X}_i\xi_w^a) \tag{1.6}$$

where φ_w^a is a scalar that controls the overall probability of experiencing the adverse event outcome in treatment group w, and ξ_w^a is a vector parameter that defines the relationship between the covariates and the adverse event outcome. The expected probability that unit i experiences death under treatment w is defined as

$$Pr(D_i(w) = 1|\mathbf{X}_i, \varphi_w^d, \xi_w^d, A_i(w)) = u^{-1}(\varphi_w^d + \mathbf{X}_i\xi_w^d + \zeta_w * A_i(w)) \tag{1.7}$$

where the parameter ζ_w defines the dependence between the adverse event and death outcomes. We examined multiple estimands to summarize the results of the data: the ITT for the adverse event and death separately, the ITT for a composite binary outcome, the SACE, $\kappa_{10} - \kappa_{01}$, $\frac{\kappa_{10}}{\kappa_{01}}$ and the probability that $G_i(w) = k$, for all $k \in \{1, \ldots, 4\}$ and $w \in \{0, 1\}$, where $G_i(w)$ is defined:

$$G_i(w) = \begin{cases} 1 & \text{if patient } i \text{ neither dies nor experiences heart failure under treatment } w \\ 2 & \text{if patient } i \text{ experiences heart failure but does not die} \\ 3 & \text{if patient } i \text{ does not experience the heart failure but does die} \\ 4 & \text{if patient } i \text{ experiences both heart failure and death.} \end{cases} \tag{1.8}$$

1.4.1 Case Studies

For each case study, we assume that u is the logistic function and that the values of ξ_w^a are fixed at the posterior mean of the logistic regression parameters estimated using the data in Section 1.3. We set ζ_1, ζ_0, φ_1^a, φ_0^a ,φ_1^d and φ_0^d to the values depicted in Appendix Tables A1.1 and A1.2. Each setting provides different data generating mechanisms that are used to examine the operating characteristics of proposed Bayesian method. In the first case study, the adverse events and death occur together more frequently when $W_i = 1$ compared to $W_i = 0$. In the second case study, fewer adverse events occur when $W_i = 1$ compared to $W_i = 0$, but the adverse events that occur under $W_i = 1$ result in higher mortality than the adverse events that occur under $W_i = 0$.

In Case Study 1, the ITT effects for the occurrence of the adverse event and mortality are 0.035 and 0.030, respectively, which implies that individuals under the active intervention suffer from more deaths and adverse effects on average. The SACE for adverse events is -0.065 which indicates that among those who survive under both treatments, less adverse events occur for those receiving the active intervention on average. The ITT effect of the composite binary outcome is -0.004. This indicates that, on average, the probability of suffering from either an adverse event or death is practically the same under both interventions. These estimand values illustrate the

Table 1.1: Values of different estimands under the data generating mechanism defined in Section 1.4

Estimand	Case Study 1			Case Study 2		
	$w=1$	$w=0$	Value	$w=1$	$w=0$	Value
ITT Adverse	0.380	0.345	**0.035**	0.193	0.226	**-0.033**
ITT Death	0.568	0.538	**0.030**	0.378	0.252	**0.127**
ITT Composite	0.634	0.638	**-0.004**	0.415	0.363	**0.052**
SACE	0.152	0.217	**-0.065**	0.059	0.148	**-0.089**
$Pr(G_i(w)=1)$	0.366	0.362	**0.005**	0.585	0.637	**-0.052**
$Pr(G_i(w)=2)$	0.066	0.100	**-0.034**	0.037	0.111	**-0.074**
$Pr(G_i(w)=3)$	0.253	0.292	**-0.039**	0.222	0.136	**0.086**
$Pr(G_i(w)=4)$	0.315	0.245	**0.070**	0.156	0.115	**0.041**
$\kappa_{10}-\kappa_{01}$	-	-	**0.092**	-	-	**0.120**
$\frac{\kappa_{10}}{\kappa_{01}}$	-	-	**1.445**	-	-	**1.608**

difficulties that can arise from using traditional estimands, in which ITT estimands show that the active intervention increases the probability of adverse events and death, but the SACE shows an effect in the opposite direction and the composite binary outcome shows no effect.

The composite ordinal outcome provides a clearer interpretation. The probability of composite ordinal outcome levels 2 and 3 are higher under the control intervention, indicating more patients are experiencing adverse events and death separately. The probability of composite ordinal outcome level 4 is higher under the active intervention, implying that death and the adverse event are occurring more often together under the active intervention. The composite ordinal outcome estimands $\kappa_{10}-\kappa_{01}$ and $\frac{\kappa_{10}}{\kappa_{01}}$ indicate that patients are expected to do worse under the active intervention more often than they would under the control intervention.

In Case Study 2, the ITT effects for the occurrence of the adverse event and mortality are -0.033 and 0.127, respectively. This indicates the probability of suffering from an adverse event is higher under the control intervention, however, the probability of mortality is higher under the active intervention. The SACE for the adverse events is -0.089, which indicates that among those who survive under both treatments, the probability of experiencing an adverse event is higher for those receiving the control intervention. The ITT effect for the composite binary outcome is 0.052. This indicates that the probability of suffering from either an adverse event or death is larger under the active intervention. Each of these estimands may lead to different conclusions regarding the toxicity profiles of the interventions. The composite ordinal outcome provides a more

complete description for the effects of the intervention by considering the effects on death and the adverse event simultaneously. Overall, patients are expected to do worse more often under the active intervention than they would under the control intervention. This can also be seen using $G_i(w)$, where the individuals with $W_i = 1$ have higher probability of being at levels 3 and 4 of the composite ordinal outcome.

1.4.2 Simulation Results

The case studies in Section 1.4.1 were replicated with 1000 datasets. The vector parameter α in Equation (1.5) is sampled from a multivariate normal distribution with mean 0 and identity covariance matrix at each replication. This resulted in different subpopulations assigned to the active intervention. We estimated all of the estimands in Table 1 using the proposed Bayesian Method with a natural cubic spline on the propensity score. We compared the method to a Doubly Robust (DR) estimator (Robins, Rotnitzky and Zhao, 1994) that estimates both the propensity score and outcomes using logistic regression models with all available covariates. The SACE, $\kappa_{10} - \kappa_{01}$, and $\frac{\kappa_{10}}{\kappa_{01}}$ were not estimated using a DR method because we did not identify a valid DR method to estimate these estimands in observational studies. When defining the natural cubic spline for the proposed Bayesian method, we start by defining six subclasses using quantiles of the propensity score distribution. In order to be able to estimate a separate cubic polynomials within a subclass, we need three observations from each treatment group. When the number of observations in each subclass for each treatment group was smaller than 3, we decrease the number of subclasses until we have at least three observations from each treatment group in each subclass. Six subclasses were selected intitially because it was shown to have good performance in many scenarios when estimating treatment effects for dichotomous outcomes (Gutman and Rubin, 2013*a*). Additionally, the Cauchy prior distribution with mode 0 and scale parameter of 2.5 has been shown to have good operating characteristics when used for Bayesian logistic regression in many scenarios, including complete separation (Gelman et al., 2008). Thus, we assume independent Cauchy distributions with mode 0 and scale 2.5 as the prior distributions for all model parameters in all logistic regression outcome models.

When there is no analytical form to the posterior distribution, sampling from it can be computationally intensive when repeating for 1000 replications of each case study. To reduce the

computation at each replication of each case study, we estimated the Bayesian logistic regressions for the outcomes with the *baysglm* function in the `arm` library (Gelman and Su, 2022). This method utilizes an approximate EM algorithm to obtain the maximum a posteriori probability (MAP) estimates for the model's parameters (Gelman et al., 2008). In large samples, the posterior distribution of these estimates can be approximated with a multivariate normal distribution centered around the MAP estimate with the expected Fisher Information as the variance-covariance matrix. In our simulations, we relied on this approximation to draw samples from the posterior distribution of θ_0^a, θ_1^a, θ_0^d and θ_1^d.

Table 1.2: Simulation results for the first case study

Estimand	Bayesian Method				DR			
	Coverage*	Bias	I.W.†	RMSE	Coverage⋆	Bias	I.W.	RMSE
ITT Adverse	94	-0.001	0.080	0.021	95	0.000	0.083	0.022
ITT Death	95	0.000	0.092	0.023	96	0.001	0.097	0.024
ITT Composite	94	0.001	0.085	0.022	96	0.001	0.092	0.023
SACE	94	0.001	0.104	0.028	-	-	-	-
$Pr(G_i(1)=1) - Pr(G_i(0)=1)$	94	-0.001	0.085	0.023	96	-0.001	0.092	0.023
$Pr(G_i(1)=2) - Pr(G_i(0)=2)$	94	0.000	0.052	0.013	94	0.000	0.056	0.015
$Pr(G_i(1)=3) - Pr(G_i(0)=3)$	95	0.002	0.085	0.022	96	0.001	0.092	0.023
$Pr(G_i(1)=4) - Pr(G_i(0)=4)$	95	-0.001	0.074	0.018	95	-0.001	0.079	0.020
$\kappa_{10} - \kappa_{01}$	96	0.001	0.102	0.027	-	-	-	-
$\frac{\kappa_{10}}{\kappa_{01}}$	96	0.005	0.399	0.101	-	-	-	-

* Coverage of 95% credible interval
⋆ Coverage of 95% confidence interval
† Interval Width

Table 1.2 depicts the performance of the estimation methods for each of the estimands in the first case study when g in Equations (1.5), (1.6), and (1.7) is the logistic link function. The Bayesian and DR methods generally result in well calibrated interval estimates. Interval lengths and root mean squared errors are slightly larger for the DR method compared to the Bayesian methods. The proposed Bayesian method also produced well calibrated credible intervals for the SACE and for ordinal outcomes estimands. For both Bayesian and DR methods, the bias of estimates is relatively small.

In order to assess the methods' sensitivity to specification of the link function, we modify u in Equations (1.5), (1.6), and (1.7) to be from the Burr family which takes the form,

$$F_c(x) = 1 - (1 + e^x)^{-c}.$$

Table 1.3: Simulation results for the first case study under Burr link function with $c = 0.5$.

Estimand	Bayesian Method				DR			
	Coverage	Bias	I.W.	RMSE	Coverage	Bias	I.W.	RMSE
ITT Adverse	95	0.004	0.081	0.021	95	0.001	0.082	0.021
ITT Death	94	0.004	0.095	0.024	96	0.000	0.098	0.025
ITT Composite	95	0.005	0.095	0.025	95	0.000	0.099	0.026
SACE	93	0.003	0.085	0.023	-	-	-	-
$Pr(G_i(1) = 1) - Pr(G_i(0) = 1)$	95	-0.005	0.095	0.025	95	0.000	0.099	0.026
$Pr(G_i(1) = 2) - Pr(G_i(0) = 2)$	94	0.002	0.060	0.016	94	0.000	0.062	0.017
$Pr(G_i(1) = 3) - Pr(G_i(0) = 3)$	96	0.002	0.082	0.021	95	-0.001	0.086	0.022
$Pr(G_i(1) = 4) - Pr(G_i(0) = 4)$	94	0.002	0.063	0.017	95	0.001	0.067	0.018
$\kappa_{10} - \kappa_{01}$	95	0.006	0.106	0.028	-	-	-	-
$\frac{\kappa_{10}}{\kappa_{01}}$	95	0.031	0.414	0.112	-	-	-	-

When $c = 1$, $F_c(x)$ corresponds to the inverse of the logistic link function. For our simulations, we set $c = 0.5$ as in Gutman and Rubin(Gutman and Rubin, 2013*a*). Table 1.3 describes the performance of the Bayesian and DR estimating methods for each of the estimands in the first case study with the Burr link function. The Bayesian estimation method results in approximately nominal coverage for all traditional and ordinal ITT estimands under the misspecified link function (Table 1.3). The 95% confidence intervals for the DR methods have similar coverage rates, interval length and RMSE. Similar results are observed for the second case study (Tables A1.3 and A1.4 in the Appendix).

1.5 Comparing Interventions for Type II Diabetes Mellitus among Nursing Home Residents

We analyze the effect of treatment of T2DM with DPP4Is and SUs in NH residents using the proposed Bayesian method and the data described in Section 1.3. Because we use the propensity score matched sample, we estimate the treatment effects among those treated with DPP4Is. Let $W_i = 1$ indicate treatment initiation with DPP4I and $W_i = 0$ indicate treatment initiation with SU. $D_i(w)$ and $A_i(w)$ indicate the occurrence of death and heart failure, respectively, under treatment

w for patient i. Based on the functional forms described in Equations (1.3) and (1.4), we assume:

$$
\begin{aligned}
A_i(w) &\sim \text{Bern}(\text{logit}^{-1}(f_w^a(g(\hat{e}(\mathbf{X}_i)), \mathbf{B}_w^a) + \mathbf{X}_i^* \cdot \beta_w^a)) \\
D_i(w) &\sim \text{Bern}(\text{logit}^{-1}(f_w^d(g(\hat{e}(\mathbf{X}_i)), \mathbf{B}_w^d) + \mathbf{X}_i^* \cdot \beta_w^d + A_i(w) \cdot \eta_w)) \\
\beta_w^\ell &\overset{i.i.d}{\sim} N\left(0, \frac{3^2}{\lambda_{\beta_w^\ell}}\right) \text{ for } \ell \in \{a, d\}, w \in \{0, 1\} \\
\mathbf{B}_w^\ell &\overset{i.i.d}{\sim} N\left(0, 8^2\right) \text{ for } \ell \in \{a, d\}, w \in \{0, 1\} \\
\eta_w &\sim N(0, 8^2) \text{ for } w \in \{0, 1\} \\
\lambda_{\beta_w^\ell} &\sim \text{half-Cauchy}(0, 1) \text{ for } \ell \in \{a, d\}, w \in \{0, 1\},
\end{aligned}
$$

where $f(\hat{e}(\mathbf{X}_i), \mathbf{B}_w^a)$ is a natural cubic spline on the logit of the estimated propensity score with 5 internal knots, $\mathbf{B}_w^a$ and $\mathbf{B}_w^d$ represent vectors of unknown spline coefficients, β_w^a and β_w^d are unknown linear adjustment coefficients, and η_w is an unknown coefficient defining the relationship between the adverse event and mortality. Because we already include a spline on the logit of the estimated propensity score and there are a large number of covariates in each regression model, we assume $\beta_w^\ell \overset{i.i.d}{\sim} N\left(0, \frac{3^2}{\lambda_{\mathbf{B}_w^\ell}}\right)$ and $\lambda_{\beta_w^\ell} \sim$ half-Cauchy$(0, 1)$ for $\ell \in \{a, d\}, w \in \{0, 1\}$. These prior distributions shrink the estimates of the parameters towards zero and are commonly referred to as the Ridge shrinkage prior distribution (van Erp, Oberski and Mulder, 2019). To test the sensitivity of the results to the prior distribution, we also conducted the analysis assuming the components of β_w^ℓ independently follow the Laplace distribution with location parameter 0 and scale parameter $\frac{3^2}{\lambda_{\beta_w^\ell}}$ for $\ell \in \{a, d\}, w \in \{0, 1\}$. This prior distribution is commonly referred to as the Lasso shrinkage prior distribution (van Erp, Oberski and Mulder, 2019). The results are quantitatively similar for both prior distributions and we report only the ones using Ridge prior distributions.

Treatment effects were summarized with both traditional estimands and composite ordinal outcome estimands using $G_i(w)$, where $G_i(w)$ is defined as it is in Equation (1.8). Because we are analyzing treatment effects in the super-population, we estimated the estimands by sampling from the posterior distributions of functions of θ_0^a, θ_1^a, θ_0^d and θ_1^d using the sampling procedure in Section 1.2.3. To sample from the posterior distributions we used the JAGS software (Depaoli, Clifton and Cobb, 2016) and the rjags package (Denwood, 2016). The Gelman-Rubin convergence statistics

for all model parameters were less than 1.1 and posterior predictive checks demonstrated suitable model fit (see Figures A1.2 and A1.3 in Appendix).

1.5.1 Results

Table 1.4 summarizes the results for the ITT for heart failure and death, the composite binary outcome and the SACE. Except for the death ITT, the other three estimands are greater than zero, but do not reach the 5% significance level. Nursing home residents with T2DM have 2% (95% CrI [-0.01, 0.04]) higher probability of heart failures within 180 days when treated with DPP4I (0.08, 95% CrI [0.07, 0.09]) compared to SU (0.06, 95% [0.05, 0.08]). Mortality within 180 days is similar between DPP4I (0.27; 95% CrI [0.24, 0.29]) and SU (0.27; 95% CrI [0.24, 0.29]). Lastly, among NH residents who survive for 180 days under both drugs, those who used DPP4I have 1% (95% CrI [0.00, 0.04]) higher probability of experiencing heart failure compared to those that use SUs.

Table 1.4: Posterior mean and 95 % super-population credible interval for traditional estimands estimated using the proposed Bayesian procedure

Treatment	Heart Failure	Death	Composite Binary	SACE
DPP4I	0.08 [0.07, 0.09	0.27 [0.24, 0.29]	0.30 [0.28, 0.33]	0.04 [0.03, 0.05]
SU	0.06 [0.05, 0.08]	0.27 [0.24, 0.29]	0.29 [0.26, 0.31]	0.027 [0.02, 0.04]
Difference	0.02 [−0.01, 0.04]	0.00 [−0.04, 0.04]	0.02 [−0.01, 0.04]	0.01 [0.00, 0.03]

Table 1.5 presents the results for treatment effects with ordinal outcomes. After 180 days from the initiation of either DPP4I or SU for T2DM, DPP4I and SU users had similar risk of experiencing heart failure but not death (Risk Difference (RD): 0.01, 95% CrI: [0.00, 0.02]), death but not heart failure (RD: -0.01, 95% CrI: [-0.04, 0.03]) and both heart failure and death (RD: 0.01, 95% CrI: [-0.01, 0.02]). The mortality under both treatments is similar, with approximately 27% of patients having outcomes in level 3 and 4 for both DPP4I and SU. Nursing home residents with T2DM were estimated to have a 5% higher risk of a worse composite ordinal outcome under DPP4I compared to SU, but this estimate was not statistically significant at the 5% nominal level (κ_{10}/κ_{01}: 1.05, 95% CrI: [0.87, 1.24]). The estimated posterior probability that a resident selected at random would experience a worse composite ordinal outcome under DPP4I compared to SU is 0.72.

Table 1.5: Posterior mean and 95 % super-population credible interval for proportion of patients who would experience each level of the outcome under each treatment

Treatment	$Pr(G_i(w)=1)$	$Pr(G_i(w)=2)$	$Pr(G_i(w)=3)$	$Pr(G_i(w)=4)$
DPP4I	0.70 [0.68, 0.73]	0.03 [0.02, 0.04]	0.22 [0.19, 0.24]	0.050 [0.04, 0.06]
SU	0.71 [0.69, 0.74]	0.02 [0.01, 0.03]	0.22 [0.20, 0.25]	0.04 [0.03, 0.06]
Difference	−0.01 [−0.05, 0.03]	0.01 [0.00, 0.02]	0.01 [0.04, 0.03]	0.006 [0.01, 0.02]

The third row in Table 1.6 describes the estimates of $\Delta_j, j \in \{1,2,3\}$ in Equation (1.1). At each level of the composite scale, the posterior mean of Δ_j is either 0 or negative. This indicates that on-average residents who received SU have lower ordinal outcome than those who receive DPP4I. However, none of these estimates are significant at the 5% nominal level.

Table 1.6: Estimates and 95% CI for Distributional Causal Estimands

Treatment	$Pr(G_i(w)=1)$	$Pr(G_i(w)\leq 2)$	$Pr(G_i(w)\leq 3)$
DPP4I	0.70 [0.68, 0.73]	0.73 [0.71 0.76]	0.95 [0.94, 0.96]
SU	0.71 [0.69, 0.74]	0.73 [0.71, 0.76]	0.96 [0.94, 0.97]
Difference	−0.01 [−0.05, 0.03]	0.00 [−0.04, 0.04]	−0.01 [−0.02, 0.01]

1.5.2 Sensitivity Analysis for Data Application

We use the procedure described in Section 1.2.4 to examine the sensitivity of estimates of the relative treatment effect (RTE), $\hat{\kappa}_{10} - \hat{\kappa}_{01}$, to the strong ignorability assumption. For selected values of μ^z_{SU}, we calculate $\hat{\kappa}_{10} - \hat{\kappa}_{01}$ and its standard error (SE) at different levels of δ_a and δ_d. We then compute a standardized effect, $\frac{\hat{\kappa}_{10}-\hat{\kappa}_{01}}{\text{SE}(\hat{\kappa}_{10}-\hat{\kappa}_{01})}$, and plot its value at each combination of δ_a and δ_d. We let the unobserved covariate $\mathbf{Z}$ be normally distributed with unit variance because covariates in our model have been standardized. Thus, μ^z_{SU} describes the standardized bias between the distribution of Z_i in patients who initiated a SU and patients who initiated a DPP4I. Figure 3 displays the results for $\delta_a, \delta_d \in [-1,1]$ and $\mu^z_{SU} = 1$. Under the strong ignorability assumption, $\hat{\kappa}_{10} - \hat{\kappa}_{01}$ was 0.01 (SE 0.02), which implies a standardized effect of approximately 0.5.

All standardized values of $\hat{\kappa}_{10} - \hat{\kappa}_{01}$ are between -0.11 and 0.77, with the estimated value dropping below 0 only when δ_a approaches -1 and δ_d approaches 1. This shows that even with a relatively large standardized bias of 1 (Rubin, 2001) the standardized estimate is between the 2.5% to the

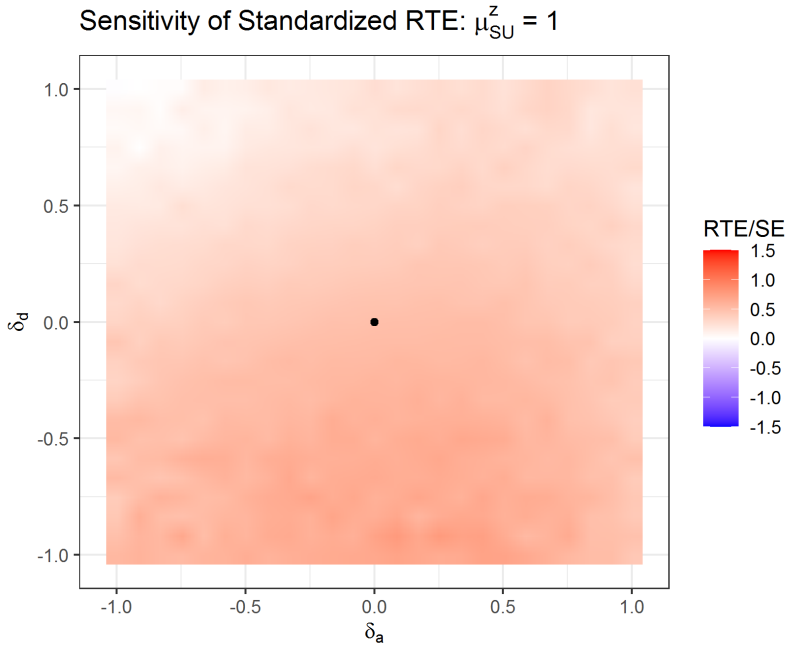

Figure 1.1: Sensitivity of the standardized relative treatment effect at different levels of δ_a and δ_d when $\mu^z_{\text{SU}} = 1$. The black dot denotes where δ_a and δ_d equal 0.

97.5% percentile of the Normal distribution. These results hold even when the unobserved covariate has a conditional log odds below -5 on the adverse events and larger than 5 on death. These are large effect sizes, and none of the observed variables had a conditional log odds of this magnitude for either the adverse event or death.

Figure 1.2 depicts the sensitivity results for $\mu^z_{SU} = -1$, which displays large initial bias in the opposite direction. Under this configuration, the standardized effect estimates remain between -0.02 and 0.54, with estimates below 0 occurring only when δ_a approaches 1 while δ_d approaches -1. This shows that with a standardized of -1 the standardized estimates of the RTE are between the 2.5% and the 97.5% quantiles of the Normal distribution. Similar observations can be made with a conditional log odds of more than 5 on the adverse effects and below -5 on death. This demonstrates that the estimates of the relative treatment effect of DPP4I compared to SU are relatively robust to violations of the strong ignorability assumption.

1.6 Conclusion

We propose a method for estimating the effects of interventions on adverse events in the presence of mortality. Our method views causal inference as a missing data problem and explicitly imputes

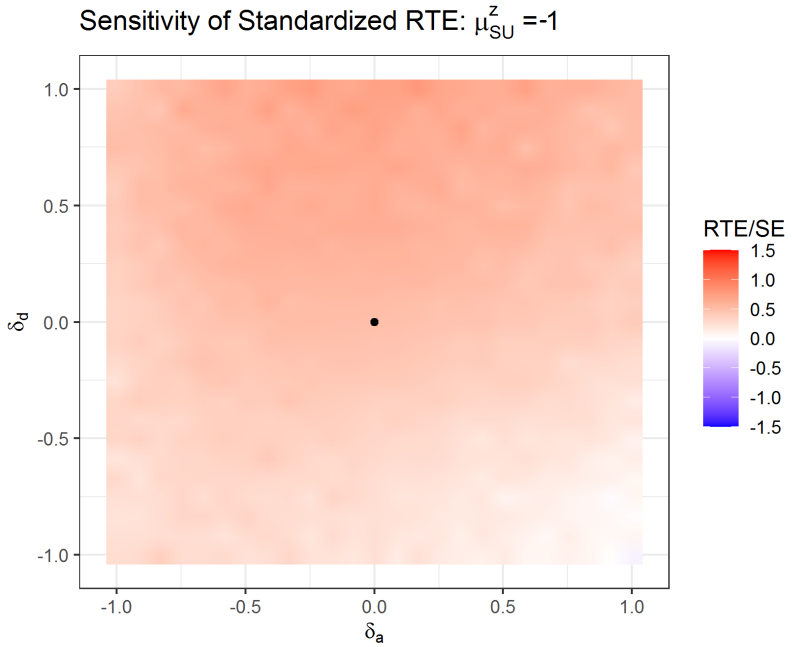

Figure 1.2: Sensitivity of the standardized relative treatment effect at different levels of δ_a and δ_d when $\mu^z_{\text{SU}} = -1$. The black dot denotes where δ_a and δ_d equal 0.

the missing vectors of potential outcomes. This enables researchers to make inference on any finite-sample or super-population estimand. The method can be applied to observational studies with the strong ignorability assumption and to randomized trials. For observational studies, we also provide a sensitivity analysis for the strong ignorability assumption. Combining the estimation procedure with the sensitivity analysis can generate real world evidence on possible adverse effects of interventions in the presence of differential mortality.

To address possible limitations of current estimands such as the ITT and SACE, we proposed a composite ordinal outcome to analyze the effects of interventions on adverse events. By combining the occurrence of death and an adverse event on an increasing scale of severity, the outcomes can be analyzed simultaneously so that the impact of the interventions on both is considered. The proposed method can be applied to the entire population that participated in the study, in contrast to the SACE estimand, which describes a subpopulation that may not be of interest when considering the toxicity of interventions.

We provide case studies to display the differences in interpretation that arise from using the traditional and the composite ordinal outcome estimands. The proposed ordinal estimand can provide a more complete picture on the adverse effects of the different interventions. Simulating

the case studies displays similar or superior performance of the proposed Bayesian method when compared to a Doubly Robust method for estimating causal estimands under correctly specified and misspecified link functions.

We apply the proposed Bayesian method to an observational study that compares the effects of SUs and DPP4Is among NH residents with T2DM. Estimates of traditional estimands reveal that in this population, more heart failure is expected to occur under DPP4I than SU. Specifically, the composite ordinal outcome estimands show that a randomly selected patient is estimated to have a 5% higher risk of having a worse composite ordinal outcome under DPP4I than having a worse outcome under SU. However, this result was not significant at the 5% nominal level. Prior analyses of these data utilize Cox proportional hazard models that include the occurrence of death as independent censoring when estimating the treatment effects on the rate of heart failure (Zullo et al., 2021; Zullo, 2017). These models ignore the possibility that one treatment may lead to higher mortality because of more lethal heart failures or is less effective at treating T2DM. The proposed Bayesian method analyzes death and heart failure simultaneously while ranking death as the more severe adverse event to provide a more complete assessment of the risk profiles of the two interventions. This enables investigators to assess relationships between heart failure and death by producing estimates for the risk of each combination of these events under DPP4Is compared to SUs. Additionally, the proposed method provides posterior probabilties that quantify the liklihood that a patient experiences a worse composite ordinal outcome under either treatment. When using the proposed Bayesian method for this analysis, no differential risks of heart failure or death were found 180 days after initiating DPP4Is or SUs. However, at longer follow-up times, there may be benefits to using the composite ordinal outcome scale because differential risks in mortality may be more likely.

A possible limitation of the proposed composite ordinal outcome is that individuals who die during the study period may survive longer under one intervention and therefore have higher probability of experiencing an adverse event. One possible solution is to increase the number of ordinal levels in the composite outcome to distinguish between the occurrence of death during different time periods. This would require replacing the current binary model for mortality with a time-to-death model or a series of conditional binary models that represent mortality at different time points. The combination of experiencing the adverse event and death during an earlier time

period would be ranked as the most severe outcome, and the combinations of later mortality and adverse events would be considered less severe outcomes. Another approach relies on a time-to-death model to make inference on the win-ratio using the imputed missing time-to-death and adverse events potential outcomes.

In conclusion, the proposed Bayesian procedure provides a new methodology to analyze observational studies with binary outcomes in the presence of mortality. The procedure imputes the adverse events and death from their joint posterior distribution which preserves the relationship between these outcomes. This relationship is used in defining a composite ordinal outcome that summarizes the effects of the interventions on the entire study population and provide a more complete picture on the toxicity profiles of these interventions. The proposed method can be used with randomized and observational studies, and for finite-sample and super-population estimands. We have applied the method for one type of adverse events, but this method can be extended to multiple types of adverse events. This extension includes incorporating additional conditional models to Equations (1.3) and (1.4) and adjusting the estimands accordingly. Other possible extensions may comprise models that include continuous and time-to-event outcomes.

Chapter 2

A Latent Principal Stratification Method to Address Cluster and Individual Noncompliance in Pragmatic Cluster RCTs

2.1 Introduction

Agitation and aggressive behaviors are common among cognitively impaired individuals that reside in nursing homes (NHs) (Zuidema, Koopmans and Verhey, 2007). These behaviors can decrease quality of life and enhance cognitive decline (García-Alberca et al., 2019; Shin et al., 2005) while also being major source of stress for NH staff (Karlsen et al., 2023) and other residents (Rosen et al., 2008). Commonly used antipsychotic and mood stabilizer medications are associated with significant adverse events such as falls and deaths [34]. As a consequence, it is important to examine the effects of non-pharmaceutical alternatives to mitigate aggressive behaviors and reduce antipsychotic use.

METRIcAL is a multi-stage, pragmatic, NH based, cluster randomized control trial (PCRCT) that was implemented NHs across 4 corporate chains in the US (Mor, 2021). The trial evaluated the effects of personalized music on the agitated behaviors and medication usage of nursing home

residents with moderate to severe dementia. However, like many other pragmatic trials in patients with dementia (Allore et al., 2020), METRIcAL suffered from a large portion of individuals in the active intervention arm that did not comply with the intervention.

Pragmatic randomized controlled trials (PRCTs) aim to estimate the effects of interventions in real-world settings where participants are subject to different degrees of compliance to the intervention (Patsopoulos, 2011). PRCT analyses commonly follow the intention to treat (ITT) principle where units are analyzed according to the trial arm to which they were originally assigned, regardless of their compliance status (Sheiner and Rubin, 1995; Gupta, 2011; Jo, Asparouhov and Muthén, 2008). These estimands may not summarize the effects of receipt of the intervention, and may not provide insights into the effects of the intervention in settings with different patterns of compliance (Hernán and Robins, 2017). Possible methods to address non-compliance are the as-treated and per-protocol estimands. The as-treated estimands allocate units to the intervention that they received (Ellenberg, 1996), while per-protocol estimands consider only units who received their assigned intervention (Matilde Sanchez and Chen, 2006). Because these estimands consider samples that do not correspond to the original randomization design, they may result in biased estimates of the causal effects (Goldberg and Belitskaya-Levy, 2014).

When compliance status is binary, an estimand that preserves the randomization design is the complier average causal effect (CACE). This estimand estimates the effects of an intervention among units that would comply to the intervention they are assigned (Imbens and Rubin, 1997; Frangakis, 2002). However, because compliance is only observed under one intervention for each unit, additional assumptions are required to estimate the CACE. Under the monotonicity assumption, the exclusion restriction assumption, and individual-level randomization, instrumental variables (IV) methods can be used to estimate the CACE (Angrist, Imbens and Rubin, 1996). Bayesian procedures have shown improved operating characteristics compared to econometric IV techniques (Imbens and Rubin, 1997) and can evaluate the impact of violations of the monotonicity and exclusion restrictions assumptions (Imbens and Rubin, 1997; Hirano et al., 2000). Another estimation procedure for the CACE relies on multiple imputation (MI) to explicitly impute the missing compliance status. In smaller sample sizes, this procedure was shown to have better operating characteristics than maximum likelihood and method of moments estimators (Taylor and Zhou, 2009). In cluster randomized designs, Forastiere, Mealli and VanderWeele (2016) described

a framework to estimate the CACE when there is individual-level noncompliance. Imai, Jiang and Malani (2021) proposed a different method for two-stage randomized experiments that have both interference and noncompliance at the individual-level. Bayesian methods have also been proposed to address individual-level noncompliance in cluster randomized designs (Frangakis, Rubin and Zhou, 2002; Taylor and Zhou, 2011; Ohnishi and Sabbaghi, 2022). However, in pragmatic cluster randomized trials, noncompliance to the intervention can occur at both the individual and cluster-levels.

The principal stratification framework enables investigators to adjust for post-treatment variables to obtain valid causal effects estimands (Frangakis and Rubin, 2002*b*). This framework is more flexible than the instrumental variable approach and can be used to define estimands with two-level compliance status. With binary compliance at the cluster-level and the individual-level, Schochet and Chiang (2011) defined 16 principal strata that summarize all combinations of compliance at both levels. The CACE is defined for the group of clusters and individuals that comply with their treatment assignments. To estimate this version of the CACE, a method of moments procedure under specific identifiability conditions is derived. An extension of this method considers a variety of identifying assumptions and proposes a different maximum liklihood estimation procedure (Sheng, Li and Zhou, 2019).

In some studies, compliance to an intervention may be partial. For example, when individuals consume different proportions of the assigned intervention dose (Jin and Rubin, 2008). In PCRCTs, partial compliance can also occur at the cluster-level. Zhai and Gutman (2020) proposed a Bayesian MI procedure that imputes the unobserved partial compliance status of healthcare providers and the binary compliance status of patients to estimate the effects of video for advanced directives on healthcare utilization. This method relied on a single, discretized, compliance variable to define partial compliance levels at the cluster-level, and it did not specify a joint distribution for the outcomes and the compliance status.

We propose a new Bayesian method to analyze PCRCTs with one-sided binary noncompliance (Imbens and Rubin, 2016) at the individual-level and one-sided partial compliance at the cluster-level. This method relies on a Bayesian mixture model to classify clusters into latent compliance strata based on cluster's pre-treatment characteristics, its partial compliance status and individual-level outcomes. Because compliance is only observed in the treatment arm, the mix-

ture model imputes the unobserved compliance for clusters and individuals assigned to the control arm. The procedure can estimate finite and super-population estimands within strata defined by a combination of the latent cluster-level compliance and individual-level compliance. We apply the procedure to analyze METRIcAL and find that residents who listened to the personalized music playlists and lived in NHs that were more likely to target agitated behaviors reduced their antipsychotic use at 4-month follow-up compared to those assigned to standard of care. However, this reduction was not statistically significant at the 5% nominal level.

2.2 Motivating Example: METRIcAL

METRIcAL is a PCRCT that examined the effects of a personalized music intervention on agitated behaviors and antipsychotic use among NH residents with dementia. The first stage of METRIcAL included 976 residents in 54 NH facilities. Half of the facilities were assigned to the personalized music intervention and the other half were assigned to the control condition (Mor, 2021). Participants assigned to the personalized music intervention received iPods that included music that they listened to as young adults. The control condition was usual care, which could include the use of ambient or group music. In the first stage of METRIcAL, no effect of personalized music intervention was observed (McCreedy et al., 2022).

2.2.1 Description of Data

Facility-level baseline characteristics were obtained from LTCfocus, a long term care database sponsored by the National Institute on Aging (1P01AG027296) through collaboration with Brown University. These characteristics included the previous year's average daily census, racial and gender demographics, latitude and longitude, the average age of residents, and the percent of walking residents. Participant level baseline characteristics were obtained from the minimum data set (MDS) (Centers for Medicare & Medicaid Services, 2021). The MDS contains comprehensive data on each participant's health status and functional capabilities. These characteristics include demographics such as sex and age, co-morbid conditions such as depression, and measures of functional status such as the Activities of Daily Living Scale (ADL) (Graf, 2008), Cognitive Function Scale (CFS) (Thomas et al., 2017) and Agitated and Reactive Behavior Scale (ARBS) (McCreedy et al., 2019)

(Appendix Table A2.2). Outcome measures including the Cohen-Mansfield Agitation Inventory (CMAI) (Cohen-Mansfield, Marx and Rosenthal, 1989) and antipsychotic use were collected at baseline and at up to two follow-up time points for each study participant. The trial also collected information about the fidelity of the intervention. At the individual-level, the amount of personalized music played by each resident was recorded in iPod metadata. At the NH level, investigators documented the proportion of nurse involvement, the proportion of unique songs loaded on the iPods for each resident, and the proportion of enrolled residents who were identified by NH-staff on a pre-treatment survey as individuals whose behaviors they intended to directly target with the personalized music intervention. Detailed description of the trial design is reported in McCreedy et al. (2021).

2.3 Methods

2.3.1 Notation and Definitions

We consider a super-population of I^{sp} clusters, such that each cluster consists of N_i units, and the total population size is $N = \sum_{i=1}^{I^{sp}} N_i$, where N could be infinite. To study the effect of a binary intervention, a cluster randomized trial is conducted in a finite-sample with I^{fp} clusters indexed $i = 1, .., I^{fp}$, where I_0^{fp} is the number of clusters assigned to the control arm, and $I_1^{fp} = I^{fp} - I_0^{fp}$ is the number of clusters assigned to the treatment arm. Cluster i comprises $n_i \leq N_i$ units such that $n = \sum_{i=1}^{I^{fp}} n_i$ denotes the total number of study participants. Let $W_i \in \{0, 1\}$ be a treatment assignment indicator that equals 0 if cluster i is assigned to the control condition and equals 1 if it is assigned to the active intervention. We define $\mathbf{W} = \{W_1, \ldots, W_{I^{fp}}\}$ as the treatment assignment vector for all I^{fp} clusters. In addition, let $\mathbf{C}_i(W_i) = \{C_i^1(W_i), ..., C_i^\ell(W_i)\}$ denote a vector of ℓ potential continuous compliance metrics that describe the implementation of the active intervention in cluster i when it is assigned to W_i. We assume that clusters assigned to the control arm have no access to the active intervention so that $\mathbf{C}_i(0) = 0$. For simplicity we define $\mathbf{C}_i := \mathbf{C}_i(1)$. For each cluster, we observe a set of P covariates, $\mathbf{Z}_i = \{Z_i^1, ..., Z_i^P\}$, that consists of variables recorded prior to the intervention being implemented. We assume that the values of $\mathbf{Z}_i$ and $\mathbf{C}_i$ are determined by a clusters' membership to one of K latent compliance strata. Let $S_i \in \{1, \ldots, K\}$ be an allocation variable that indicates which of the strata cluster i belongs to. We define $\mathbf{C} = \{\mathbf{C}_i\}$ and $\mathbf{S} = \{S_i\}$

to be the cluster-level compliance vector and latent compliance strata vector for all I^{fp} clusters.

We define $D_{ij}(W_i) \in \{0, 1\}$ to be an indicator that equals 1 when individual j in cluster i receives the active intervention when assigned W_i, and 0 otherwise. Because individuals in control clusters have no access to the active intervention, $D_{ij}(0) = 0$ for all i, j. For simplicity, we define $D_{ij} := D_{ij}(1)$ and identify an individual as a complier when $D_{ij} = 1$. Let $\mathbf{Y}(0, \mathbf{D}) = \{Y_{ij}(0, D_{ij})\}$ denote the set of potential outcomes for all n individuals under the control condition, and $\mathbf{Y}(1, \mathbf{D}) = \{Y_{ij}(1, D_{ij})\}$ denote the set of potential outcomes for all n individuals under the active intervention. Similarly, we let $\mathbf{D} = \{D_{ij}\}$ denote the set of individual compliance indicators. Lastly, for individual j in cluster i we have M pre-treatment covariates $\mathbf{X}_{ij} = \{X_{ij}^1, \ldots, X_{ij}^M\}$.

We assume the stable unit treatment value assumption (SUTVA) (Rubin, 1990*b*) for compliance status and the potential outcomes in the cluster randomized trial so that the observed and missing potential outcomes are

$$Y_{ij}^{obs} = Y_{ij}(1, D_{ij}) * W_i + Y_{ij}(0, D_{ij}) * (1 - W_i)$$
$$Y_{ij}^{mis} = Y_{ij}(1, D_{ij}) * (1 - W_i) + Y_{ij}(0, D_{ij}) * (W_i).$$

The sets of observed and missing compliance and variables are

$$\mathbf{D}^{obs} = \{D_{ij} | W_i = 1\} \text{ and } \mathbf{D}^{mis} = \{D_{ij} | W_i = 0\}$$
$$\mathbf{C}^{obs} = \{\mathbf{C}_i | W_i = 1\} \text{ and } \mathbf{C}^{mis} = \{\mathbf{C}_i | W_i = 0\}.$$

Table A2.1 in the Appendix describes the observed and unobserved quantities in a PCRCT with one-sided noncompliance.

2.3.2 Principal Strata and Estimands

Causal estimands are statistics that can be used to summarize the effects of an intervention across a population of interest (Gutman and Rubin, 2015*b*; Rubin, 1978*b*). The population of interest can comprise only units in the sample, resulting in finite-sample estimands. Alternatively, the population can comprise all units in a target super-population, resulting in super-population estimands. The intention to treat (ITT) estimand summarizes the effects of randomization across the

population of interest (Imbens and Rubin, 2016). When individuals adhere with their assigned intervention the ITT estimand can be used to infer the effect of the intervention. However, when trial participants do not comply with their assigned intervention, the ITT effect may not represent the effects of recieving the intervention (Imbens and Rubin, 2016). Possible finite and super-population estimands under the ITT principle are,

$$\text{ITT}^{fp} = \frac{1}{n}\sum_{ij} Y(1, D_{ij}) - Y(0, D_{ij})$$

$$\text{ITT}^{sp} = E_{sp}[Y_{ij}(1, D_{ij}) - Y_{ij}(0, D_{ij})],$$

where the expectation for ITT^{sp} is taken over the super-population.

Principal stratification estimands can be used to estimate the effects of receiving the intervention among certain sub-populations of participants. Zhai and Gutman [50] defined a set of principal strata estimands for a pragmatic CRCT using a continuous compliance measure at the cluster-level and a binary compliance indicator at the unit level. These principal strata discretized the cluster-level continuous compliance metric into ordinal subclasses, and within cluster-level compliance strata it defined defined individual-level compliance strata. These estimands are based on a single compliance measure, and require cutoff points to discretize the continuous cluster-level compliance level.

With multiple possible continuous compliance metrics, we define principal strata estimands that stratify individuals by their cluster's latent compliance class, S_i, and their individual compliance status, D_{ij}. One possible set of estimands describe the ITT effect among individuals in clusters that belong to latent compliance stratum k,

$$\text{ITT}_k^{fp} = \frac{\sum_{ij:S_i=k} Y(1, D_{ij}) - Y(0, D_{ij})}{\sum_i n_i * I(S_i = 1)} \tag{2.1}$$

$$\text{ITT}_k^{sp} = E_{sp}[Y_{ij}(1, D_{ij}) - Y_{ij}(0, D_{ij})|S_i = k], \tag{2.2}$$

where $I(S_i = k)$ is an indicator function that equals 1 if cluster i belongs to latent compliance stratum k, and 0 otherwise. A different estimand is the CACE, which describes the effects of the

intervention among compliers regardless of their cluster's compliance stratum,

$$\text{CACE}^{fp} = \frac{\sum_{ij:D_{ij}=1} Y(1, D_{ij}) - Y(0, D_{ij})}{\sum_{ij} I(D_{ij} = 1)} \tag{2.3}$$

$$\text{CACE}^{sp} = E_{sp}[Y_{ij}(1, D_{ij}) - Y_{ij}(0, D_{ij})|D_{ij} = 1]. \tag{2.4}$$

A principal strata estimand that combines individual compliance status and cluster latent compliance strata is the CACE within each latent compliance stratum,

$$\text{CACE}_k^{fp} = \frac{\sum_{ij:D_{ij}=1,\ S_i=k} Y(1, D_{ij}) - Y(0, D_{ij})}{\sum_{ij} I(D_{ij} = 1) * I(S_i = 1)} \tag{2.5}$$

$$\text{CACE}_k^{sp} = E_{sp}[Y_{ij}(1, D_{ij}) - Y_{ij}(0, D_{ij})|D_{ij} = 1, S_i = k]. \tag{2.6}$$

To compare the effects of an intervention between cluster latent compliance strata, we can estimate contrasts between the ITT_k estimands. For example, $\text{ITT}_1^{sp} - \text{ITT}_2^{sp}$ estimates the difference in being randomized to the intervention in compliance stratum 1 compared to the same effect in compliance stratum 2. Another example is the $\text{CACE}_1^{sp} - \text{CACE}_2^{sp}$ that estimates the same effect, but only among compliers in each cluster-level compliance stratum.

2.3.3 Bayesian Latent Class Model

Distributional Assumptions

We rely on a Bayesian model-based approach to estimate the effects of the intervention within principal strata. Model-based estimation is a principled approach to incorporate restrictions on the joint distribution of observed variables and include covariates in a flexible manner (Imbens and Rubin, 2016). In addition, using a Bayesian approach enables investigators to examine aspects of model fit using posterior predictive checks to ensure sensible model specifications (Mattei, Li and Mealli, 2013). The joint distribution of the variables described in Section 2.3.1 is

$$\begin{aligned} &P(\mathbf{W}, \mathbf{Y}(1, \mathbf{D}), \mathbf{Y}(0, \mathbf{D}), \mathbf{D}, \mathbf{C}, \mathbf{Z}, \mathbf{X}, \mathbf{S}) \propto \\ &\quad P(\mathbf{Y}(1, \mathbf{D}), \mathbf{Y}(0, \mathbf{D})|\mathbf{D}, \mathbf{C}, \mathbf{Z}, \mathbf{X}, \mathbf{S}) P(\mathbf{D}|\mathbf{C}, \mathbf{Z}, \mathbf{X}, \mathbf{S}) P(\mathbf{C}, \mathbf{Z}, \mathbf{X}, \mathbf{S}) P(\mathbf{W}|\mathbf{Z}). \end{aligned} \tag{2.7}$$

In a pragmatic CRCT, the cluster's treatment assignment is controlled by the investigators and is decided prior to study initiation. Thus, it is independent from both the compliance status and the potential outcomes,

$$P(\mathbf{W}|\mathbf{Y}(1,\mathbf{D}),\mathbf{Y}(0,\mathbf{D}),\mathbf{D},\mathbf{C},\mathbf{Z},\mathbf{X}) = P(\mathbf{W}|\mathbf{Z}). \tag{2.8}$$

Units are considered exchangeable when the joint distribution of their potential outcomes are the same regardless of their ordering (Bernardo, 1996). Cluster-level assignment to treatment informed whether a clusters' potential compliance metrics were observed, but it did not inform their values. Similarly, within clusters, individual-level covariates, compliance status and the potential outcomes are not informed by their assignment to treatment. Thus, we make the following partial exchangeablitly assumptions.

Assumption 1: Exchangability

(a) The individual-level covariates, $\mathbf{X}_{ij}$, the facility compliance metrics, $\mathbf{C}_i$, and characteristics, $\mathbf{Z}_i$, and the compliance strata allocation variable, S_i, are exchangeable. Assuming the prior $P(\theta_S)$ and appealing to De-Finetti's theorem (de Finetti, 1992), we can write $P(\mathbf{X},\mathbf{C},\mathbf{Z})$ as

$$P(\mathbf{X},\mathbf{C},\mathbf{Z},\mathbf{S}) \propto \int \prod_{i=1}^{I^{fp}} \prod_{j=1}^{n_i} P(\mathbf{X}_{ij},\mathbf{C}_i,\mathbf{Z}_i,S_i|\theta_S)P(\theta_S)d\theta_S.$$

Furthermore, we assume that $\mathbf{X}_{ij}$ is independent of $\mathbf{C}_i$ and $\mathbf{Z}_i$ given the allocation variable S_i. This implies that

$$P(\mathbf{X},\mathbf{C},\mathbf{Z},\mathbf{S}) \propto \int \prod_{i=1}^{I^{fp}} \prod_{j=1}^{n_i} P(\mathbf{X}_{ij}|S_i,\theta_S)P(\mathbf{C}_i,\mathbf{Z}_i|S_i,\theta_S)P(S_i|\theta_S)P(\theta_S)d\theta_S.$$

(b) The individual compliance status, $\mathbf{D}_{ij}$, is exchangeable conditional on $\mathbf{X}_{ij}$, $\mathbf{C}_i$, $\mathbf{Z}_i$ Assuming prior distribution $P(\theta_D)$ and appealing to De-Finetti's theorem we have

$$P(\mathbf{D}|\mathbf{C},\mathbf{Z},\mathbf{X},\mathbf{S}) \propto \int \prod_{i=1}^{I^{fp}} \prod_{j=1}^{n_i} P(D_{ij}|\theta_D,\mathbf{C}_i,\mathbf{Z}_i,\mathbf{X}_{ij},S_i)P(\theta_D)d\theta_D.$$

(c) The potential outcomes are exchangeable conditional on D_{ij}, $\mathbf{X}_{ij}$, $\mathbf{C}_i$ and $\mathbf{Z}_i$. Assuming prior

distribution $P(\theta_Y)$ and appealing to De-Finetti's theorem we have

$$P(\mathbf{Y}(1,\mathbf{D}),\mathbf{Y}(0,\mathbf{D})|\mathbf{D},\mathbf{C},\mathbf{Z},\mathbf{X},\mathbf{S}) \propto$$
$$\int \prod_{i=1}^{I^{fp}} \prod_{j=1}^{n_i} P(Y_{ij}(1,D_{ij}),Y_{ij}(0,D_{ij})|\theta_Y,D_{ij},\mathbf{C}_i,\mathbf{Z}_i,\mathbf{X}_{ij},S_i)P(\theta_Y)d\theta_Y.$$

A commonly made assumption in randomized trials with noncompliance is the exclusion restriction assumption (Angrist, Imbens and Rubin, 1996), which rules-out the effects of the assignment on the outcome for non-compliers. We assume a slightly weaker version of this assumption that requires it to hold in distribution.

Assumption 2: Stochastic exclusion restriction for individuals that do not comply with the active intervention (Imbens and Rubin, 2016). $P(Y_{ij}(1,0)) = P(Y_{ij}(0,0))$.

This assumption presumes that the assignment of a cluster to the intervention does not influence the distribution of the potential outcomes except through an individual's compliance to the intervention. This assumption is reasonable in PCRCTs that examine interventions that are embedded in regular patient care because the assignment of cluster to the active intervention generally does not influence individuals' outcomes unless they receive the intervention.

Lastly, we rely on two assumptions that simplify the parametric modeling of the different distributions. The first assumption presumes that clusters' characteristics and compliance metrics only inform individual compliance status and the potential outcomes through the latent compliance strata.

Assumption 3: Conditional independence of individual compliance and potential outcomes from cluster's characteristics. Formally, $P(D_{ij}|\mathbf{C}_i,\mathbf{Z}_i,S_i,\mathbf{X}_{ij},\theta_D) = P(D_{ij}|S_i,\mathbf{X}_{ij},\theta_D)$ and $P(Y_{ij}(1,D_{ij}),Y_{ij}(0,D_{ij})|\mathbf{C}_i,\mathbf{Z}_i,S_i,D_{ij},\mathbf{X}_{ij},\theta_Y) = P(Y_{ij}(1,D_{ij}),Y_{ij}(0,D_{ij})|S_i,D_{ij},\mathbf{X}_{ij},\theta_Y)$.

This assumption can be removed, but it is expected that the latent cluster-level compliance encapsulates the cluster-level characteristics that influence individual-level compliance. The second

simplifying assumption is that the distributions of the potential outcomes are independent given cluster-level and individual compliance as well as individual-level characteristics.

Assumption 4: No contamination of imputation across treatments (Rubin, 2008). Formally, $P(Y_{ij}(1, D_{ij}), Y_{ij}(0, D_{ij})|S_i, D_{ij}, \mathbf{X}_{ij}, \theta_Y) = P(Y_{ij}(1, D_{ij})|S_i, D_{ij}, \mathbf{X}_{ij}, \theta_Y)P(Y_{ij}(0, D_{ij})|S_i, D_{ij}, \mathbf{X}_{ij}, \theta_Y)$.

This assumption is commonly made by causal inference methods because the observed data does not include information on the relationship between the potential outcomes for the same individual (Hirano et al., 2000; Imbens and Rubin, 2016). When the units in a study are a random sample from a super-population, and the estimand is the average difference in this super-population, this assumption has little inferential effect for average treatment effects (Hirano et al., 2000; Imbens and Rubin, 2016; Rubin, 1978*b*; Imbens and Rubin, 1997).

Parametric Model Specifications

To estimate finite-sample and super-population estimands we develop a model based approach. This approach provides a principled way to incorporate the distributional assumption in Section 2.3.3 as well as cluster and individual-level covariates. To identify the latent compliance strata we assume that the joint distribution of $\mathbf{X}_{ij}, \mathbf{C}_i$ and $\mathbf{Z}_i$ follow a K component mixture distribution with allocation variable S_i,

$$P(\mathbf{X}_i, \mathbf{C}_i, \mathbf{Z}_i, S_i|\theta_S) = \prod_{k=1}^{K} \left[\prod_{j=1}^{n_i} P_k(\mathbf{X}_{ij})^{I(S_i=k)} \right] \text{MVN}(\mathbf{C}_i, \mathbf{Z}_i|\boldsymbol{\mu}_k^S, \Sigma)^{I(S_i=k)} \prod_{k=1}^{K} \pi_k^{I(S_i=k)}, \qquad (2.9)$$

where MVN is the Multivariate Normal density function with mean $\boldsymbol{\mu}_k^S = \left\{ \mu_k^{C_1}, \ldots, \mu_k^{C_\ell}, \mu_k^{Z_1}, \ldots, \mu_k^{Z_P} \right\}$ and covariance matrix Σ. P_k is the distribution of individual-level covariates in mixture component k and π_k is the prior probability that a cluster is in latent compliance stratum k. We define the set of parameters across all clusters as $\theta_S = \{\boldsymbol{\mu}_1^S, \ldots, \boldsymbol{\mu}_K^S, \pi_1, \ldots, \pi_K, \Sigma\}$. P_k can be modeled as either the empirical distribution across all observed individual-level covariate vectors, $\hat{F}(\mathbf{X}_{ij})$, or the empirical distribution across observed covariate vectors for individuals in clusters that belong to latent compliance stratum k, $\hat{F}_k(\mathbf{X}_{ij})$. If the baseline covariates of trial participants are expected

to be systematically different based on their cluster's latent compliance stratum, then $\hat{F}_k(\mathbf{X}_{ij})$ is preferred.

We model the conditional distribution of individual-level binary compliance indicators as

$$P(D_{ij}|S_i, \mathbf{X}_{ij}, \theta_D) = \prod_{k=1}^{K}[\text{probit}^{-1}(\mu_k^D + \mathbf{X}_{ij}\alpha_k + \phi_i^D)^{D_{ij}}(1 - \text{probit}^{-1}(\mu_k^D + \mathbf{X}_{ij}\alpha_k + \phi_i^D))^{1-D_{ij}}]^{I(S_i=k)}, \quad (2.10)$$

where $\theta_D = \{\theta_{D_1}, ..., \theta_{D_K}, \phi^D\}$, $\theta_{D_k} = \{\mu_k^D, \alpha_k\}$ and $\{\phi_i^D\}$ are cluster-level effects. The conditional distribution for the potential outcomes is modeled as

$$\begin{aligned}
&P(Y_{ij}(1, D_{ij}), Y_{ij}(0, D_{ij})|S_i, D_{ij}, \mathbf{X}_{ij}, \theta_Y) = \\
&P(Y_{ij}(1, D_{ij})|S_i, D_{ij}, \mathbf{X}_{ij}, \theta_Y)P(Y_{ij}(0, D_{ij})|S_i, D_{ij}, \mathbf{X}_{ij}, \theta_Y) = \\
&\prod_{k=1}^{K}[N(Y_{ij}(1, D_{ij})|\mu_k^Y + \mathbf{X}_{ij}\beta_{0,k} + D_{ij} * (\delta_{1,k} + \mathbf{X}_{ij}\beta_{1,k}) + \phi_i^Y, \sigma_k^2)\times \\
&\quad N(Y_{ij}(0, D_{ij})|\mu_k^Y + \mathbf{X}_{ij}\beta_{0,k} + D_{ij} * \delta_{0,k} + \phi_i^Y, \sigma_k^2)]^{I(S_i=k)}, \qquad (2.11)
\end{aligned}$$

where $\theta_Y = \{\theta_{Y_1}, \ldots, \theta_{Y_K}, \phi^Y\}$, $\theta_{Y_k} = \{\mu_k^Y, \beta_{1,k}, \beta_{0,k}, \delta_{1,k}, \delta_{0,k}, \sigma_k^2\}$ and $\phi^Y = \{\phi_i^Y\}$. For individuals in clusters that belong to latent compliance stratum k, $\delta_{0,k}$ represents the conditional difference in expected outcomes between compliers and noncompliers if assigned to control. Similarly, $\delta_{1,k}$ represents the conditional difference in expected outcomes for compliers and noncompliers when assigned to the intervention. Lastly and the parameter vector $\beta_{1,k}$ represents the interaction effects on the expected outcomes between individual-level compliance status and the individual's observed covariates. Cluster-level effects are distributed as

$$\begin{aligned}
P(\phi_i^D|\tau_D^2) &= N(\phi_i^D|0, \tau_D^2) \text{ and} \qquad (2.12)\\
P(\phi_i^Y|\tau_Y^2) &= N(\phi_i^Y|0, \tau_Y^2).
\end{aligned}$$

These distributions imply a common variance parameter across the latent compliance strata. The parameter vector $\boldsymbol{\theta} = \{\theta_Y, \theta_D, \theta_S, \tau_Y^2, \tau_D^2\}$ comprises the set of parameters that define the distributions of all observable variables.

Estimation Procedures

A Bayesian approach for estimating finite-sample estimands, denoted by γ^{fp}, and super-populaiton estimands, denoted by γ^{sp}, conditions on the observed data to calculate the posterior distribution, $P(\gamma|\mathbf{Y}^{obs}, \mathbf{D}^{obs}, \mathbf{C}^{obs}, \mathbf{Z}, \mathbf{X}, \mathbf{W})$. We describe a procedure for estimating the posterior distribution of γ^{fp} (Procedure 1), and two procedures for estimating the posterior distribution of γ^{sp} (Procedures 2 and 3) under the assumptions in Sections 2.3.3 and 2.3.3.

Procedure 1: Finite-sample estimands can be written as functions of the observed data, missing data and treatment assignment indicator. For example,

$$\text{ITT}^{fp} = \frac{1}{n}\sum_{ij} Y_{ij}(1, D_{ij}) - Y_{ij}(0, D_{ij}) = \frac{1}{n}\left(\sum_{ij:W_i=1}\left[Y_{ij}^{obs} - Y_{ij}^{mis}\right] + \sum_{ij:W_i=0}\left[Y_{ij}^{mis} - Y_{ij}^{obs}\right]\right).$$

The posterior distribution of γ^{fp} is

$$\begin{aligned}
&P(\gamma^{fp}|\mathbf{Y}^{obs}, \mathbf{D}^{obs}, \mathbf{C}^{obs}, \mathbf{Z}, \mathbf{X}, \mathbf{W}) \\
&= \int\int\int P(\gamma^{fp}|\mathbf{Y}^{obs}, \mathbf{D}^{obs}, \mathbf{C}^{obs}, \mathbf{Y}^{mis}, \mathbf{D}^{mis}, \mathbf{C}^{mis}, \mathbf{S}, \mathbf{Z}, \mathbf{X}, \mathbf{W}) \times \\
&\quad P(\mathbf{Y}^{mis}, \mathbf{D}^{mis}, \mathbf{C}^{mis}, \mathbf{S}|\mathbf{Y}^{obs}, \mathbf{D}^{obs}, \mathbf{C}^{obs}, \mathbf{Z}, \mathbf{X}, \mathbf{W}) d\mathbf{Y}^{mis} d\mathbf{D}^{mis} d\mathbf{C}^{mis} d\mathbf{S} \\
&= \int\int\int P(\gamma^{fp}|\mathbf{Y}^{obs}, \mathbf{D}^{obs}, \mathbf{C}^{obs}, \mathbf{Y}^{mis}, \mathbf{D}^{mis}, \mathbf{C}^{mis}, \mathbf{S}, \mathbf{Z}, \mathbf{X}, \mathbf{W}) \times \\
&\quad P(\mathbf{Y}^{mis}, \mathbf{D}^{mis}, \mathbf{C}^{mis}, \mathbf{S}|\mathbf{Y}^{obs}, \mathbf{D}^{obs}, \mathbf{C}^{obs}, \mathbf{Z}, \mathbf{X}) d\mathbf{Y}^{mis} d\mathbf{D}^{mis} d\mathbf{C}^{mis} d\mathbf{S}. \qquad (2.13)
\end{aligned}$$

Equation (2.13) shows that to estimate γ^{fp} we need to estimate $P(\mathbf{Y}^{mis}, \mathbf{D}^{mis}, \mathbf{C}^{mis}, \mathbf{S}|\mathbf{Y}^{obs}, \mathbf{D}^{obs}, \mathbf{C}^{obs}, \mathbf{Z}, \mathbf{X})$ because $P(\gamma^{fp}|\mathbf{Y}^{obs}, \mathbf{D}^{obs}, \mathbf{C}^{obs}, \mathbf{Y}^{mis}, \mathbf{D}^{mis}, \mathbf{C}^{mis}, \mathbf{S}, \mathbf{Z}, \mathbf{X}, \mathbf{W})$ is a degenerate distribution. Conditioning on $\boldsymbol{\theta}$,

$$\begin{aligned}
&P(\mathbf{Y}^{mis}, \mathbf{D}^{mis}, \mathbf{C}^{mis}, \mathbf{S}|\mathbf{Y}^{obs}, \mathbf{D}^{obs}, \mathbf{C}^{obs}, \mathbf{Z}, \mathbf{X}) = \\
&\int_{\boldsymbol{\theta}} P(\mathbf{Y}^{mis}, \mathbf{D}^{mis}, \mathbf{C}^{mis}, \mathbf{S}|\mathbf{Y}^{obs}, \mathbf{D}^{obs}, \mathbf{C}^{obs}, \mathbf{Z}, \mathbf{X}, \boldsymbol{\theta}) P(\boldsymbol{\theta}|\mathbf{Y}^{obs}, \mathbf{D}^{obs}, \mathbf{C}^{obs}, \mathbf{Z}, \mathbf{X}) d\boldsymbol{\theta} \qquad (2.14)
\end{aligned}$$

The posterior distribution, $P(\boldsymbol{\theta}|\mathbf{Y}^{obs}, \mathbf{D}^{obs}, \mathbf{C}^{obs}, \mathbf{Z}, \mathbf{X})$, does not have an analytical form, so we developed a Gibbs sampling algorithm (Geman and Geman, 1984; Gelfand and Adrian, 1990; Imbens and Rubin, 1997) to obtain posterior draws. Based on Equation (2.14), the algorithm for

estimating γ^{fp} is

1. Select initial values: $\boldsymbol{\theta}^{(0)} = \{\theta_D^{(0)}, \theta_Y^{(0)}, \theta_S^{(0)}, \tau_Y^{2(0)}, \tau_D^{2(0)}\}$.

2. For m in $1, \ldots, M$:

   (a) Sample $\boldsymbol{\theta}^{(m)} \sim (\boldsymbol{\theta}|\mathbf{Y}^{obs}, \mathbf{D}^{obs}, \mathbf{C}^{obs}, \mathbf{Z}, \mathbf{X})$ using the sampling procedure outlined in Appendix Section A2.1.

   (b) Given $\boldsymbol{\theta}^{(m)}$, draw $\{\mathbf{Y}^{mis(m)}, \mathbf{D}^{mis(m)}, \mathbf{C}^{mis(m)}, \mathbf{S}\}$ from $P(\mathbf{Y}^{mis}, \mathbf{D}^{mis}, \mathbf{C}^{mis}, \mathbf{S}|\mathbf{Y}^{obs}, \mathbf{D}^{obs}, \mathbf{C}^{obs}, \mathbf{Z}, \mathbf{X}, \boldsymbol{\theta}^{(m)})$ using the procudure outlined in Appendix Section A2.2.

   (c) Calculate $\hat{\gamma}^{fp(m)}$ as a function of the observed values and $\{\mathbf{Y}^{mis(m)}, \mathbf{D}^{mis(m)}, \mathbf{C}^{mis(m)}, \mathbf{S}\}$.

3. The posterior intervals are obtained using percentiles of the posterior samples $\{\hat{\gamma}^{fp(1)}, \ldots, \hat{\gamma}^{fp(M)}\}$.

Procedure 2: Super-population estimands can often be written as a functions of the model parameters. The formulas for the four super-population estimands described in Section 2.3.2 are displayed in Table 2.1 (analytical derivations are in Appendix Section A2.4).

Estimand	$P_k(\mathbf{X}_{ij}) = \hat{F}(\mathbf{X}_{ij})$	$P_k(\mathbf{X}_{ij}) = \hat{F}_k(\mathbf{X}_{ij})$
ITT^{sp}	$\frac{1}{n}\sum_{ij}\sum_k \pi_k p_{ijk}(\mathbf{X}_{ij}\beta_{1,k} + \delta_k)$	$\sum_k \pi_k \left(\frac{1}{\sum_{i\in\mathcal{I}_k} n_i}\sum_{i\in\mathcal{I}_k}\sum_{j=1}^{n_i} p_{ijk} * (\mathbf{X}_{ij}\beta_{1,k} + \delta_k)\right)$
ITT_k^{sp}	$\frac{1}{n}\sum_{ij} p_{ijk} * (\mathbf{X}_{ij}\beta_{1,k} + \delta_k)$	$\frac{1}{\sum_{i\in\mathcal{I}_k} n_i}\sum_{i\in\mathcal{I}_k}\sum_{j=1}^{n_i} p_{ijk} * (\mathbf{X}_{ij}\beta_{1,k} + \delta_k)$
CACE^{sp}	$\mu_k^Y + \frac{\frac{1}{n}\sum_{ij}\mathbf{X}_{ij}\beta_{1,k}p_{ijk}}{\frac{1}{n}\sum_{ij} p_{ijk}} + \delta_k$	$\frac{\sum_k \pi_k \frac{1}{\sum_{i\in\mathcal{I}_k} n_i}\sum_{i\in\mathcal{I}_k}\sum_{j=1}^{n_i} p_{ijk}\left(\frac{\frac{1}{\sum_{i\in\mathcal{I}_k} n_i}\sum_{i\in\mathcal{I}_k}\sum_{j=1}^{n_i} p_{ijk}\mathbf{X}_{ij}\beta_{0,k}}{\frac{1}{\sum_{i\in\mathcal{I}_k} n_i}\sum_{i\in\mathcal{I}_k}\sum_{j=1}^{n_i} p_{ijk}} + \delta_k\right)}{\sum_k \pi_k \frac{1}{\sum_{i\in\mathcal{I}_k} n_i}\sum_{i\in\mathcal{I}_k}\sum_{j=1}^{n_i} p_{ijk}}$
CACE_k^{sp}	$\frac{\frac{1}{n}\sum_{ij} p_{ijk}*\left(\mathbf{X}_{ij}\beta_{1,k}+\delta_{1,k}\right)}{\frac{1}{n}\sum_{ij} p_{ijk}}$	$\frac{\frac{1}{\sum_{i\in\mathcal{I}_k} n_i}\sum_{i\in\mathcal{I}_k}\sum_{j=1}^{n_i} p_{ijk}\mathbf{X}_{ij}\beta_{1,k}}{\frac{1}{\sum_{i\in\mathcal{I}_k} n_i}\sum_{i\in\mathcal{I}_k}\sum_{j=1}^{n_i} p_{ijk}} + \delta_k$

$p_{ijk} = \text{probit}^{-1}\left(\left(\mathbf{X}_{ij}\alpha_k + \mu_k^D\right)/\sqrt{\tau_D^2 + 1}\right)$; $\delta_k = \delta_{1,k} - \delta_{0,k}$
$\mathcal{I}_k : \{i|S_i = k\}$; The set of indexes of clusters that belong to latent compliance stratum k.
$\text{ITT}^{sp} = E_{sp}[Y_{ij}(1, D_{ij}) - Y_{ij}(0, D_{ij})]$; $\text{ITT}_k^{sp} = E_{sp}[Y_{ij}(1, D_{ij}) - Y_{ij}(0, D_{ij})|S_i = k]$
$\text{CACE}^{sp} = E_{sp}[Y_{ij}(1, D_{ij}) - Y_{ij}(0, D_{ij})|D_{ij} = 1]$; $\text{ITT}_k^{sp} = E_{sp}[Y_{ij}(1, D_{ij}) - Y_{ij}(0, D_{ij})|S_i = k, D_{ij} = 1]]$

Table 2.1: Estimands as a function of the model parameters assuming the full empirical distirbution of coviariates ($\hat{F}(\mathbf{X}_{ij})$), or the strata specific empirical distribution of covariates ($\hat{F}_k(\mathbf{X}_{ij})$).

The posterior distribution of γ^{sp} can be written as

$$P(\gamma^{sp}|\mathbf{Y}^{obs},\mathbf{D}^{obs},\mathbf{C}^{obs},\mathbf{Z},\mathbf{X}) = \int P(\gamma^{sp}|\mathbf{Y}^{obs},\mathbf{D}^{obs},\mathbf{C}^{obs},\mathbf{Z},\mathbf{X},\boldsymbol{\theta})P(\boldsymbol{\theta}|\mathbf{Y}^{obs},\mathbf{D}^{obs},\mathbf{C}^{obs},\mathbf{Z},\mathbf{X})d\boldsymbol{\theta}. \tag{2.15}$$

Because γ^{sp} is a function of the $\boldsymbol{\theta}$, posterior samples from $P(\gamma^{sp}|\mathbf{Y}^{obs},\mathbf{D}^{obs},\mathbf{C}^{obs},\mathbf{Z},\mathbf{X})$ can be obtaind by sampling from $P(\boldsymbol{\theta}|\mathbf{Y}^{obs},\mathbf{D}^{obs},\mathbf{C}^{obs},\mathbf{Z},\mathbf{X})$. To obtain interval estimates of γ^{sp}, we skip step 2(b) in Procedure 1 and calculate $\hat{\gamma}^{sp(m)}$ as a function of $\boldsymbol{\theta}^{(m)}$.

Procedure 3: Some super-population estimands can not be derived analytically as a function of model paramters. A different approach to estimate γ^{sp} imputes the potential outcomes and compliance under the active intervention and control condition for approximately the entire super-population (Imbens and Rubin, 2016). The estimation procedure for γ^{sp} is

1. Select initial values for model parameters: $\boldsymbol{\theta}^{(0)} = \{\theta_D^{(0)}, \theta_Y^{(0)}, \theta_S^{(0)}, \tau_Y^{2(0)}, \tau_D^{2(0)}\}$.

2. For m in $1,\ldots,M$:

   (a) Sample $\boldsymbol{\theta}^{(m)} \sim (\boldsymbol{\theta}|\mathbf{Y}^{obs},\mathbf{D}^{obs},\mathbf{C}^{obs},\mathbf{Z},\mathbf{X})$using the procudure outlined in Appendix Section A2.1.

   (b) For $r = 1,\ldots,R$ draw
   $\left\{\mathbf{Y}(1,\mathbf{D})^{(m,r)},\mathbf{Y}(0,\mathbf{D})^{(m,r)},\mathbf{D}^{(m,r)},\mathbf{C}^{(m,r)},\mathbf{Z}^{(m,r)},\mathbf{S}^{(m,r)}\right\}$ from
   $P(\mathbf{Y}(1,\mathbf{D}),\mathbf{Y}(0,\mathbf{D}),\mathbf{D},\mathbf{C},\mathbf{Z}|\boldsymbol{\theta}^{(m)},\mathbf{X})$ using the procedure outlined in Appendix Section A2.3.

   (c) Combine the R samples into a single data set :

$$\Omega_{\boldsymbol{\theta}^{(m)}} = \Big\{\mathbf{Y}(1,\mathbf{D})^{(m,1)},\mathbf{Y}(0,\mathbf{D})^{(m,1)},\mathbf{D}^{(m,1)},\mathbf{C}^{(m,1)},\mathbf{Z}^{(m,1)},\mathbf{S}^{(m,1)}\ ,$$
$$\ldots,\mathbf{Y}(1,\mathbf{D})^{(m,R)},\mathbf{Y}(0,\mathbf{D})^{(m,R)},\mathbf{D}^{(m,R)},\mathbf{C}^{(m,R)},\mathbf{Z}^{(m,R)},\mathbf{S}^{(m,R)}\Big\}.$$

   (d) Compute $\hat{\gamma}^{sp}$ on the combined data set, $\Omega_{\boldsymbol{\theta}^{(m)}}$, to obtain $\hat{\gamma}^{sp(m)}$.

3. For large M and R, estimates of the percentiles of the posterior distribution of γ^{sp} are obtained using percentiles of the posterior samples $\{\hat{\gamma}^{sp(1)},\ldots,\hat{\gamma}^{sp(M)}\}$.

For example, estimating ITT^{sp} using Procedure 3 requires calculating at iteration m

$$\widehat{\text{ITT}}^{sp(m)} = \frac{1}{N * R} \sum_{r=1}^{R} \sum_{i=1}^{I^{fp}} \sum_{j=1}^{n_i} \left(Y_{ij}(1, D_{ij})^{(m,r)} - Y_{ij}(0, D_{ij})^{(m,r)} \right)$$

and obtaining point and interval estimates using percentiles of $\left\{ \widehat{\text{ITT}}^{sp(1)}, \ldots, \widehat{\text{ITT}}^{sp(M)} \right\}$.

2.4 Simulated Case Studies

We examine the operating characteristics of the proposed method using four simulated case studies. The first simulated case study includes a data generating mechanism that is correctly specified by the Bayesian models described in Section 2.3.3, while case studies 2, 3 and 4 each misspecifiy a different part of the model. We consider a simulation setting that consists of 60 clusters belonging to one of two latent compliance strata and assign half of the clusters to the active intervention and half to the control condition. Each cluster contains 20 individuals whose baseline covariates are randomly sampled resident age and Activities of Daily Living Scale (ADL) (Graf, 2008) values derived from the Minimum Data Set (MDS) (Centers for Medicare & Medicaid Services, 2021) collected for stage 1 of the METRIcAL trial (McCreedy et al., 2021). We estimate the posterior distributions of the super-population estimands described in Section 2.3.2 for each case study to examine method's performance and sensitivity to different misspecifications.

2.4.1 Case Study 1: Correctly Specified Data Generating Mechansim

In the first case study, cluster-level compliance metrics and characteristics are are generated from a Multivariate-Normal mixture distribution given by

$$
\begin{aligned}
S_i &\sim \text{Cat}(K = 2, p = [0.5, 0.5]) \\
\begin{bmatrix} C_i \\ Z_i \end{bmatrix} &\sim \begin{cases} \text{MVN}(\mathbf{m}_1, V) & \text{if } S_i = 1 \\ \text{MVN}(\mathbf{m}_2, V) & \text{if } S_i = 2 \end{cases} \\
\mathbf{m}_1 &= [-2, -2]^t \quad \mathbf{m}_2 = [2, 2]^t \\
V &= \begin{bmatrix} v_C^2 & \rho v_C v_Z \\ \rho v_C v_Z & v_Z^2 \end{bmatrix} \\
v_C^2 &\sim \text{Unif}(0.5, 2) \quad v_Z^2 \sim \text{Unif}(0.5, 2) \\
\rho &\sim \text{Unif}(-0.8, 0.8).
\end{aligned}
\tag{2.16}
$$

We assign 0.5 probability to each latent compliance stratum and randomly generate variance and correlation coefficents from Uniform distributions to assess the model's sensitivity to different covariance structures.

Individual-level binary compliance status is generated from a Bernoulli distribution that conditions on the individual's baseline covariates, their cluster's latent compliance stratum, and a cluster-level effect:

$$
\begin{aligned}
D_{ij} &\sim \text{Bernoulli}(\text{Burr}_{0.5}^{-1}(m_k^D + \mathbf{X}_{ij}\gamma_k + \varphi_i^D)) \;\; \text{if } S_i = k \\
m_1^D, m_2^D &= 0, 0.5 \\
\gamma_1 &= [-0.25, -0.25] \\
\gamma_2 &= [-0.5, -0.5] \\
\varphi_i^D &\sim N(0, \nu_D^2) \\
\nu_D^2 &= 0.25
\end{aligned}
\tag{2.17}
$$

The values of the coefficent vectors and intercept terms enable individual compliance probabilites

to differ based on their cluster's latent compliance stratum.

The potential outcomes are generated from a Normal distribution that conditions on S_i, $\mathbf{X}_{ij}$, and a cluster-level effect:

$$
\begin{aligned}
Y_{ij}(1, D_{ij}) &\sim N(m_k^Y + \mathbf{X}_{ij}\eta_{0,k} + D_{ij} * (\mathbf{X}_{ij}\eta_{1,k} + \Delta_{1,k}) + \varphi_i^Y, v_k^2) \text{ if } S_i = k \\
Y_{ij}(0, D_{ij}) &\sim N(m_k^Y + \mathbf{X}_{ij}\eta_{0,k} + D_{ij} * \Delta_{0,k} + \varphi_i^Y, v_k^2) \text{ if } S_i = k \\
m_1^Y, m_2^Y &= 2, 4 \qquad v_1^2, v_2^2 = 16, 16 \\
\Delta_{0,1}, \Delta_{0,2} &= 1, 2 \qquad \Delta_{1,1}, \Delta_{1,2} = 5.5, 7.5 \\
\eta_{0,1} &= [1, 1]^t \qquad \eta_{0,2} = [2, 2]^t \\
\eta_{1,1} &= [1, 1]^t \qquad \eta_{1,2} = [2, 2]^t \\
\varphi_i^Y &\sim N(0, \nu_Y^2) \qquad \nu_Y^2 = 9
\end{aligned}
\tag{2.18}
$$

The values of the coefficient vectors and conditional effects of being a complier differ between the latent compliance strata.

2.4.2 Case Study 2: Incorrectly Specified Mixture Model

We assess the method's sensitivity to misspecifications of the cluster-level mixture model by generating the cluster-level compliance metric and characteristic from a skewed multivariate-t distribution (Gupta, 2003) given by

$$
\begin{aligned}
S_i &\sim \text{Cat}(K = 2, p = [0.5, 0.5]) \\
\begin{bmatrix} C_i \\ Z_i \end{bmatrix} &\sim \begin{cases} \text{Skewed Multivariate-}t_q(\mathbf{m}_1, V, \zeta) & \text{if } S_i = 1 \\ \text{Skewed Multivariate-}t_q(\mathbf{m}_2, V, \zeta) & \text{if } S_i = 2 \end{cases} \\
\mathbf{m}_1 &= [-2, -2]^t \qquad \mathbf{m}_2 = [2, 2]^t \\
V &= \begin{bmatrix} v_C^2 & \rho v_C v_Z \\ \rho v_C v_Z & v_Z^2 \end{bmatrix} \quad \zeta = 2 \quad q = 5 \\
v_C^2 &\sim \text{Unif}(0.5, 2) \quad v_Z^2 \sim \text{Unif}(0.5, 2) \\
\rho &\sim \text{Unif}(-0.8, 0.8),
\end{aligned}
\tag{2.19}
$$

where V is a scale matrix, ζ denotes the skew (ζ=0 indicates no skew) and q is the degrees of freedom ($q = \infty$ is multivariate normal). We set $\zeta = 2$, $q = 5$, and generate the parameters of the scale matrix from Uniform distributions. To generate individual-level compliance status and the potential outcomes, we use the data generating mechanisms defined in (2.17) and (2.18), respectively.

2.4.3 Case Study 3: Incorrectly Specified Individual Compliance Model

We examine the method's sensitivity to misspecification of the individual compliance model by modifying the probit link function in (2.17) to be from the Burr family. The Burr link function takes the form

$$\text{Burr}_c(x) = 1 - (1 + e^x)^{-c}.$$

We set $c = 0.5$ as in (Gutman and Rubin, 2013*b*) so true compliance probabilities are skewed toward 0. Cluster-level compliance metrics and characteristics are generated using the distribution defined in (2.16) and the potential outcomes are generated using the distributions defined in (2.18).

2.4.4 Case Study 4: Incorrectly Specified Outcome Model

We examine method's sensitivity to an incorrectly specified outcome model by adding an interaction term to the data generating mechanism for the potential outcomes. For individuals in clusters that belong to latent compliance stratum k, we let

$$Y_{ij}(1, D_{ij}) \sim N(m_k^Y + \mathbf{X}_{ij}\eta_{0,k} + \xi_{0,k}\left(X_{ij1} * X_{ij2}\right) + D_{ij} * (\mathbf{X}_{ij}\eta_{1,k} + \Delta_{1,k} + \xi_{1,k}\left(X_{ij1} * X_{ij2}\right)) + \varphi_i^Y, v_k^2)$$
$$Y_{ij}(0, D_{ij}) \sim N(m_k^Y + \mathbf{X}_{ij}\eta_{0,k} + \xi_{0,k}\left(X_{ij1} * X_{ij2}\right) + D_{ij} * \Delta_{0,k}) + \varphi_i^Y, v_k^2),$$

where $\xi_{0,k}$ and $\xi_{1,k}$ define the relationship between the interaction of resident age and ADL (denoted by X_{ij1} and X_{ij2}) and the expected values of the potential outcomes. We let $\xi_{0,1} = \xi_{1,1} = \xi_{0,2} = \xi_{1,2} = -2$ so that $\xi_{0,1}$ and $\xi_{0,2}$ are twice the magnitude and opposite sign as $\eta_{0,1}$ and $\eta_{1,1}$, and $\xi_{0,2}$ and $\xi_{1,2}$ are the same magnitude and opposite sign as $\eta_{0,2}$ and $\eta_{1,2}$. We use the data generating mechanisms defined in (2.16) to generate the cluster-level compliance and characteristic, and we use (2.17) to generate individual binary compliance status.

2.4.5 Results for Simulated Case Studies

Table 2.2 depicts the true values of the super-population estimands for each case study.

Estimand	True Values			
	Correctly Specified	Skewed-t Mixture	Burr Link	Added Interaction
ITT^{sp}	2.54	2.54	1.52	2.63
ITT_1^{sp}	2.07	2.08	1.26	2.13
ITT_2^{sp}	3.00	3.00	1.78	3.12
CACE^{sp}	4.39	4.39	4.37	4.54
CACE_1^{sp}	4.11	4.11	4.17	4.23
CACE_2^{sp}	4.60	4.60	4.53	4.78
$\Pr(D_{ij} = 1 \vert S_i = 1)$	0.50	0.50	0.30	0.50
$\Pr(D_{ij} = 1 \vert S_i = 2)$	0.65	0.65	0.40	0.65

Table 2.2: True values of super-population estimands under the data generating mechanism defined in each case study.

The values of the estimands in Case Study 2 are the same as the values in Case Study 1 because the individual-level compliance and outcome models are correctly specified. When the Burr link function is used to generate individual compliance status in Case Study 3, the true values of the ITT estimands are smaller. When including the additional interaction term in the data generating mechanism for the potential outcomes in Case Study 4, the true values of each of the six outcome estimands increase slightly compared to the correctly specified data generating mechanism.

For each case study, we generated 2500 data sets. For each data set, we estimate the posterior distributions of the six outcome estimands described in Table 2.2 using Procedure 2 from Section 2.3.3. The proper prior distributions assigned to each of the model parameters are described in Appendix Section A2.5. Point estimates were obtained by computing the expected value of the estimand's posterior distribution, and credible regions were obtained by computing the median 95% interval. Table 2.3 depicts the coverage, bias, mean absolute percent bias (MAPB), interval width (IW) and root mean squared error (RMSE) for each estimand under for the first simulated case study.

Estimand	Coverage (%)	Bias	MAPB (%)	IW	RMSE
ITT^{sp}	96	0.18	18	2.38	0.59
ITT_1^{sp}	97	0.16	29	3.14	0.77
ITT_2^{sp}	95	0.20	23	3.45	0.88
CACE^{sp}	96	0.34	18	3.91	0.97
CACE_1^{sp}	97	0.33	28	5.93	1.43
CACE_2^{sp}	95	0.32	22	5.03	1.27

Coverage: Coverage of 95% Credible Interval; MAPB: Mean Absolute Percent Bias; IW: Mean 95% Credible Interval Width; RMSE: Root Mean Squared Error

Table 2.3: Operating characteristics of the method for different estimands under the correctly specified data generating mechanism defined in Case Study 1.

The proposed method results in approximately nominal coverage for all estimands when the data generating mechanism is correctly specified. It produces estimates with relatively small bias, interval width and RMSE, indicating that the model is well calibrated and a statistically valid procedure. Table 2.4 displays the operating characteristics of the proposed method under the misspecified data generating mechanisms in Case Study 2, 3 and 4.

Case Study 2: Skewed-t Mixture					
Estimand	Coverage (%)	Bias	MAPB (%)	IW	RMSE
ITT^{sp}	95	0.18	19	2.39	0.61
ITT_1^{sp}	96	0.15	29	3.18	0.77
ITT_2^{sp}	95	0.19	23	3.43	0.88
CACE^{sp}	95	0.34	18	3.93	1.01
CACE_1^{sp}	96	0.32	28	5.99	1.46
CACE_2^{sp}	95	0.32	22	5.01	1.29
Case Study3: Burr Link					
ITT^{sp}	96	0.24	23	1.81	0.45
ITT_1^{sp}	98	0.18	31	2.29	0.51
ITT_2^{sp}	96	0.31	30	2.67	0.69
CACE^{sp}	96	0.72	23	4.95	1.25
CACE_1^{sp}	98	0.58	30	7.24	1.58
CACE_2^{sp}	96	0.76	29	6.41	1.65
Case Study 4: Added Interaction					
ITT^{sp}	92	0.46	20	2.13	0.63
ITT_1^{sp}	96	0.42	26	2.85	0.71
ITT_2^{sp}	95	0.50	22	3.17	0.83
CACE^{sp}	90	0.81	20	3.41	1.05
CACE_1^{sp}	95	0.82	25	5.29	1.35
CACE_2^{sp}	94	0.76	21	4.54	1.21

Coverage: Coverage of 95% Credible Interval; MAPB: Mean Absolute Percent Bias; IW: Mean 95% Credible Interval Width; RMSE: Root Mean Squared Error

Table 2.4: Operating characteristics of the method for different estimands under the different misspecified data generating mechanism defined in Case Study 2, 3 and 4.

The operating characteristics of the proposed method are similar when the cluster-level mixture model is misspecifed in Case Study 2 and the model is correctly specified in Case Study 1. When the individual compliance models are misspecified for the Burr link function in Case Study 3, the MAPB increases for each estimand compared to Case Study 1, but the coverage of each estimand remains approximately nominal. These results indicate that the method is relatively robust to deviations from the multivariate-normal assumption that is made to identify the clusters' latent compliance strata and misspecifications of the link function in the individual-level binary compliance model. In Case Study 4, when data generating mechanism for the potential outcomes includes an additional interaction term, the method produces interval coverages that are slightly below nominal for the true values of the CACE^{sp} and ITT^{sp}. In addition, the bias for all six outcome estimands are larger in Case Study 4 for than they are in Case Study 1. However, the coverage of the posterior credible intervals for each estimand exceeds 0.90, and the interval with, RMSE and MAPB are approximately equivalent to when the model was correctly specified. These results imply that under moderate specifications in any one model component, the proposed method still estimates the super-population estimands with approximately nominal coverage, relatively small bias, interval width and RMSE.

2.5 Application to METRIcAL data

We applied the proposed method to the first stage of METRIcAL, a PCRCT decribed in Section 2.2 that examined the effects of a personalized music intervention on agitated behaviors and antipsychotic use among NH residents with dementia. We estimate the effects of the personalized music intervention on residents' change in CMAI and antipsychotic usage from baseline to 4-month follow-up. We assume that NH facilities belong to one of two latent compliance strata ($K = 2$) and consider one cluster-level characteristic and one cluster-level compliance metric. The cluster-level characteristic is referred to as "average daily census" and indicates the average number of residents who lived in the nursing home during the previous year. The cluster-level compliance metric is referred to as "clinical proportion" and indicates the proportion of enrolled residents who were identified by NH-staff on a pre-treatment survey as individuals whose behaviors they intended to directly target with the personalized music intervention. Appendix Table A2.3 describes the dis-

tribution of observed values of average daily census and clinical proportion for NHs assigned to the personalized music intervention. A NH resident is deemed a complier if they used the personalized music for at least 30 minutes a day on average during the 4 months after initiation. This was the target dosage described in the trial protocols (McCreedy et al., 2022). Baseline covariates consisted of MDS variables and the CMAI values collected for each resident at baseline (Table A2.2). We assume $P_k(\mathbf{X}_{ij}) = \hat{F}_k(\mathbf{X}_{ij})$ which enables the distribution of individual-level covariates to differ between the latent compliance strata when estimating strata-specific estimands.

2.5.1 CMAI Outcome

Following distributional assumptions outlined in Section 2.3.3, we assume a Multivariate-Normal mixture distribution for values of the average daily census and the logit of clinical proportion,

$$\begin{aligned}
S_i &\sim \text{Cat}(K = 2, \pi = [\pi_1, \pi_2]) \quad \pi \sim \text{Beta}(1, 1) \\
\begin{bmatrix} C_i \\ Z_i \end{bmatrix} &\sim \begin{cases} \text{MVN}\left(\boldsymbol{\mu}_1^S, \Sigma\right) & \text{if } S_i = 1 \\ \text{MVN}\left(\boldsymbol{\mu}_2^S, \Sigma\right) & \text{if } S_i = 2 \end{cases} \\
\boldsymbol{\mu}_k^S &\sim \text{MVN}\left(0, \begin{bmatrix} 1000 & 0 \\ 0 & 10000 \end{bmatrix}\right) \text{ for } k = 1, 2 \\
\Sigma &\sim \text{Inv-Wishart}\left(\begin{bmatrix} 1/1000 & 0 \\ 0 & 1/1000 \end{bmatrix}, 5\right).
\end{aligned} \tag{2.20}$$

In addition, we assume

$$\begin{aligned}
D_{ij} &\sim \text{Bernoulli}(\text{probit}^{-1}(\mu_k^D + \mathbf{X}_{ij}\alpha + \phi_i^D)) \;\; \text{if } S_i = k \\
\mu_k^D &\sim N\left(0, 3^2\right) \quad \alpha \sim \text{MVN}\left(0, 10 * I\right) \\
\phi_i^D &\sim N(0, \tau_D^2) \;\; \tau_D \sim \text{Uniform}(0, 4).
\end{aligned} \tag{2.21}$$

The intercept of probit regression describing the individual compliance probability varies based on the NH's latent compliance stratum, but the regression coefficients and the variance parameter in the NH effect distribution do not. We assume a uniform prior distribution on the standard

deviation of the NH effect because it has been shown to have good operating characteristics in Bayesian hierarchical models (Gelman, 2006). The CMAI potential outcomes are assumed to follow Normal distributions,

$$
\begin{aligned}
Y_{ij}(1, D_{ij}) &\sim N(\mu_k^Y + \mathbf{X}_{ij}\beta_0 + D_{ij} * (\mathbf{X}_{ij}\beta_1 + \delta_{1,k}) + \phi_i^Y, \sigma^2) \;\text{ if } S_i = k \\
Y_{ij}(0, D_{ij}) &\sim N(\mu_k^Y + \mathbf{X}_{ij}\beta_0 + D_{ij} * \delta_{0,k} + \phi_i^Y, \sigma^2) \;\text{ if } S_i = k \\
\mu_k^Y &\sim N(0, 100^2), \text{ for } k = 1, 2 \\
\delta_{1.k} &\sim N(0, 10^2) \quad \delta_{0,k} \sim N(0, 4^2), \text{ for } k = 1, 2 \\
\beta_0 &\sim \text{MVN}(0, 10^2 * I), \;\; \beta_1 \sim \text{MVN}(0, 10^2 * I) \\
\phi_i^Y &\sim N(0, \tau_Y^2) \\
\tau_Y &\sim \text{Uniform}(0, \sqrt{225}), \;\; \sigma^2 \sim \text{Inv-Gamma}(1, 1).
\end{aligned}
$$

The regression intercept, μ^Y, and the conditional effects of being a complier, δ_1 and δ_0, can differ based on the NH's latent compliance stratum. The regression coefficients, the outcome variance, and the variance of the cluster-level effect are assumed to be equal across strata. Each parameter is assigned a non-informative, proper prior distribution.

Treatment effects in the super-population were summarized using the estimands described in Section 2.3.2. To estimate their posterior distributions, we implement Procedure 2 described in Section 2.3.3. For some individuals enrolled in METRIcAL, $Y_{ij}(W_i, D_{ij})$ or D_{ij} are missing when they should have been observed. We adopt a fully Bayesian approach (Chen and Ibrahim, 2014; Erler et al., 2016) to incorporate these missing values by imputing them using samples from their posterior predictive distribution at each iteration of the Gibbs sampler. Three chains sampled 500 values from the joint posterior distribution of the model parameters. Each was thinned by 5 iterations. Mixing of the chains was assesed graphically and through Gelman-Rubin convergence statistics (Gelman and Rubin, 1992). Values for all model parameters were less than 1.1, and no label-switching occurred. See Appendix Section A2.7 for a discussion on the model diagnostics.

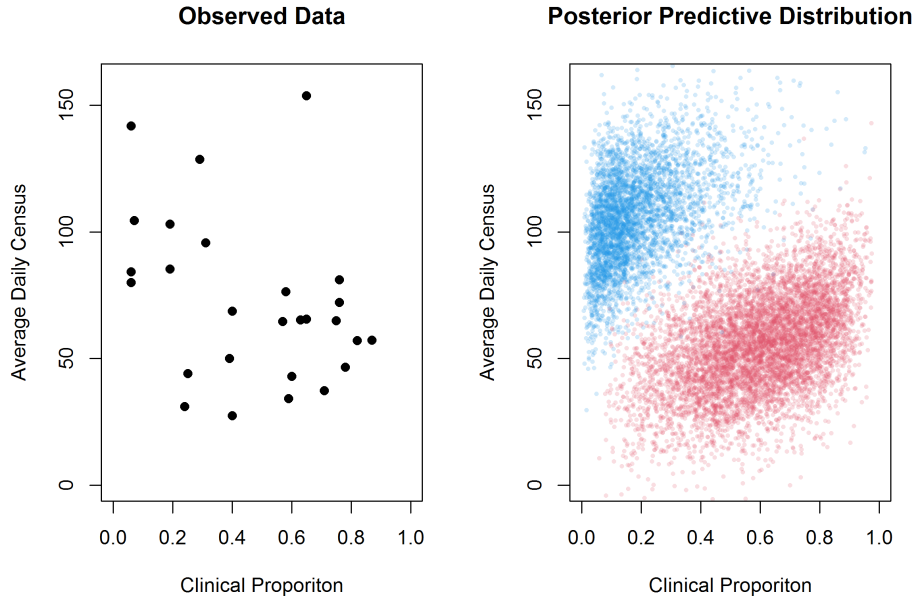

Figure 2.1: Observed data and posterior predictive plots for NH compliance metric, clinical proportion, and covariate, average daily census. Posterior draws for latent compliance stratum 1 are colored red, and poster draws for stratum 2 are colored blue.

Figure 2.1 shows the observed data and model generated posterior predictive distribution of clincial proportion and average daily census, and Table 2.5 describes the posterior distribution of their stratum specific means.

Mixture Parameter	$k=1$			$k=2$			Difference		
	Mean	SD	95% Cr.I	Mean	SD	95% Cr.I	Mean	SD	95% Cr.I
μ_{kC}	0.59	0.04	[0.51, 0.67]	0.19	0.04	[0.12, 0.28]	0.40	0.06	[0.27, 0.50]
μ_{kZ}	55.0	3.6	[48.2, 61.7]	105.9	4.7	[96.1, 114.8]	-51.0	5.6	[-61.5, -39.8]

Table 2.5: A summary of the posterior distributions of the strata means for Clinical Proportion (μ_{kC}), Average Daily Census (μ_{kZ}) and their differences between compliance stata.

The expected value of clinical proportion is three times larger for facilities in latent compliance stratum 1 compared to latent compliance stratum 2. However, NHs in stratum 1 are expected to have 51 fewer residents than NHs in stratum 2. This indicates that smaller facilities were expected to use the personalized music intervention to target behaviors and antipsychotic use with a higher proportion of enrolled residents. Table 2.6 summarizes the posterior distributions of the super-population ITT and CACE estimands.

Estimand	Treatment	Control	**Difference**	**95% Cr.I**
ITT^{sp}	-0.49 (0.81)†	-1.57 (0.78)	1.08 (0.83)	[-0.59, 2.67]
ITT_1^{sp}	-1.84 (0.97)	-2.75 (0.99)	0.91 (1.00)	[-1.07, 2.85]
ITT_2^{sp}	1.66 (1.34)	0.30 (1.24)	1.36 (1.40)	[-1.49, 4.05]
$\text{ITT}_1^{sp} - \text{ITT}_2$	-3.50 (1.66)	-3.05 (1.60)	-0.45 (1.70)	[-3.72, 3.09]
CACE^{sp}	-0.36 (1.49)	-3.30 (2.02)	2.93 (2.26)	[-1.56, 7.37]
CACE_1^{sp}	-2.07 (1.83)	-4.60 (2.31)	2.53 (2.77)	[-2.92, 7.90]
CACE_2^{sp}	2.23 (2.19)	-1.30 (3.26)	3.53 (3.64)	[-3.89, 10.36]
$\text{CACE}_1^{sp} - \text{CACE}_2^{sp}$	-4.30 (2.75)	-3.30 (3.76)	-1.01 (4.50)	[-9.65, 8.12]

† Posterior Mean (SD)

Table 2.6: Posterior Distribution of super-population estimands when the outcome is change in CMAI from baseline to 4-month follow-up. ITT_k refers to the average treatment effect among residents in NHs in latent compliance stratum k; CACE_k refers to the complier average causal effect in latent compliance stratum k. A resident is a complier if they listened to the personalized music intervention for at least 30 minutes per day on average during the four months after initiation.

There is no significant difference in the change in resident CMAI under the personalized music intervention and under standard care at the 5% nominal level (ITT: 1.08, 95% CrI: -0.59 to 2.67). Similarly, there is no observable difference between the music intervention and standard care among compliers (CACE: 2.93, 95% CrI: -1.56 to 7.37), and no significant difference between the ITT or CACE estimands in latent compliance class 1 or 2 at the 5% nominal level. This implies that individual conpliance and the latent compliance class that a NH belonged had no observable influence on the change in resident CMAI at 4-month follow-up.

2.5.2 Antipsychotics Outcome

To examine the effects of the personalized music intervention on antispychotic use at 4-month follow-up, we define a Bayesian probit regression to estimate the distribution of the potential

outcomes:

$$
\begin{aligned}
Y_{ij}(1, D_{ij}) &\sim \text{Bernoulli}(\text{probit}^{-1}(\mu_k^Y + \mathbf{X}_{ij}\beta_0 + D_{ij} * (\mathbf{X}_{ij}\beta_1 + \delta_{1,k}) + \phi_i^Y)) \text{ if } S_i = k \\
Y_{ij}(0, D_{ij}) &\sim \text{Bernoulli}(\text{probit}^{-1}(\mu_k^Y + \mathbf{X}_{ij}\beta_0 + D_{ij} * \delta_{0,k} + \phi_i^Y)) \text{ if } S_i = k \qquad (2.22) \\
\mu_k^Y &\sim N(0, 5^2), \text{ for } k = 1, 2 \\
\delta_{1.k} &\sim N(0, 5^2) \quad \delta_{0,k} \sim N(0, 1^2), \text{ for } k = 1, 2 \\
\beta_0 &\sim \text{MVN}(0, 5^2 * I), \ \beta_1 \sim \text{MVN}(0, 5^2 * I) \\
\phi_i^Y &\sim N(0, \tau_Y^2), \ \tau_Y \sim \text{Uniform}(0, \sqrt{10}).
\end{aligned}
$$

We assume a Normal prior distribution with a relatively small variance for $\delta_{0,k}$. Practically, as the variance of this prior distribution goes to zero, the model implicitly assumes there is no difference between the conditional distribution for the condtol potential outcomes for compliers and noncompliers. This is plausible because we have controlled for relevant baseline covariates and it is unlikely this parameter would be large on the probit scale. The individual-level binary compliance cluster-level mixture distributions were estimated using the same models that were defined in Section 2.5.1, and the posterior distributions of the ITT and CACE estimands were estimated using Procedure 3 described in Section 2.3.3. We sample from the posterior distribution of the model parameters by replacing the Normal regression outcome model in the Gibbs sampling algorithm with the probit regression outcome model in Equation (2.22).

The posterior distributions of μ_{kC} and μ_{kZ} do not change significantly from the values described in Table 2.5, and their interpretations remain the same. Table 2.7 describes the posterior distribution of the ITT and CACE estimands.

Estimand	Treatment	Control	**Difference**[†]	**95% Cr.I**
ITT^{sp}	26.7 (1.3)	29.5 (1.4)	-2.8 (1.6)	[-5.9, 0.1]
ITT_1^{sp}	23.5 (1.6)	27.0 (2.0)	-3.5 (2.2)	[-8.1, 0.7]
ITT_2^{sp}	32.0 (2.5)	33.7 (2.7)	-1.6 (2.6)	[-7.3, 3.1]
$\text{ITT}_1^{sp} - \text{ITT}_2^{sp}$	-8.5 (3.2)	-6.7 (3.7)	-1.8 (3.6)	[-8.6, 5.9]
CACE^{sp}	33.7 (3.6)	41.3 (4.9)	-7.6 (4.3)	[-15.7, 0.3]
CACE_1^{sp}	31.3 (4.1)	41.1 (6.4)	-9.8 (6.3)	[-22.3, 2.0]
CACE_2^{sp}	37.3 (4.9)	41.6 (7.0)	-4.3 (6.9)	[-18.7, 8.4]
$\text{CACE}_1^{sp} - \text{CACE}_2^{sp}$	-6.0 (5.5)	-0.5 (9.0)	-5.5 (10.0)	[-24.0, 15.9]

[†] Percentage Points

Table 2.7: Posterior Distribution of super-population estimands when the outcome is antipsychotic use at 4-month follow-up. ITT_k refers to the average ITT effect among residents in NHs in latent compliance stratum k; CACE_k refers to the complier average causal effect in latent compliance stratum k. A resident is a complier if they listened to the personalized music intervention for at least 30 minutes per day on average during the study.

The point estimate for the ITT estimand indicates less antipsychotic use among residents in NHs assigned to the personalized music intervention (ITT: -2.8, 95% CrI: -5.9 to 0.2), and the estimate of the CACE implies an even greater reduction in antipsychotic use among resident's who are compliers (CACE: -7.6, 95% CrI: -15.7 to 0.3). However, neither estimate was statistically significant at the 5% nominal level. The point estimate for CACE_1 is 5.5 percentage points larger in magnitude than the point estimate for CACE_2, indicating a larger reduction in antipsychotic usage among compliers in latent compliance stratum 1 (CACE_1: -9.8, 95% CrI: -22.3 to 2.0)(CACE_2: -4.3, 95% CrI: -18.7 to 8.4). Latent compliance stratum 1 comprises NHs with staff that, on average, indicated they were actively targeting a higher proportion of residents' behaviors with the music intervention.

2.5.3 Posterior Predictive Checks

We implement a series of posterior predictive checks (Gelman, Meng and Stern, 1996; Rubin, 1984) that use predictive discrepancy measures that were proposed by Barnard et al. (2003) and implemented by Mattei, Li and Mealli (2013) to assess the model fit to the observed data. Following the notation in (Mattei, Li and Mealli, 2013), we let $\mathcal{D}_{w,k}^{\text{study}} = \{ij : S_i = k, D_{ij} = 1\}$ comprise the set individual-level compliers in the *study* data whose NH facilities are in compliance stratum k and were assigned to intervention group w. We let $N_{w,k}^{\text{study}}$, denote the number of individuals in $\mathcal{D}_{w,k}^{\text{study}}$, and let $\bar{Y}(w)_k^{\text{study}}$ and $s_{w,k}^{2,\text{study}}$ denote the sample mean and variance of the potential outcomes for

individuals in $\mathcal{D}_{w,k}^{\text{study}}$, respectively. The posterior discrepancy measures are given by

$$SI_k^{\text{study}}(\boldsymbol{\theta}) = \left|\bar{Y}(1)_k^{\text{study}} - \bar{Y}(0)_k^{\text{study}}\right|,$$
$$NO_k^{\text{study}}(\boldsymbol{\theta}) = \sqrt{\frac{s_{1,k}^{2,\text{study}}}{N_{1,k}^{\text{study}}} + \frac{s_{0,k}^{2,\text{study}}}{N_{0,k}^{\text{study}}}},$$

and $SI_k^{\text{study}}(\boldsymbol{\theta})/NO_k^{\text{study}}(\boldsymbol{\theta})$. These discrepancy measures are meant to assess "broad features of signal, noise, and signal-to-noise ratio" (Barnard et al., 2003) within each of the estimated latent compliance strata. Let $T^{\text{rep}}(\boldsymbol{\theta})$ denote the value of a discrepancy measure calculated on the posterior predictive distribution of the latent compliance strata, individual compliance, and potential outcome values using the same $\boldsymbol{\theta}$ as the study data. A Bayesian posterior predictive p-value (PPPV) assess the probability that $T^{\text{rep}}(\boldsymbol{\theta})$ would be as or more extreme as what was observed in the study data, denoted by $T^{\text{obs}}(\theta)$. Formally, we have

$$\text{PPPV} = Pr(T^{\text{rep}}(\boldsymbol{\theta}) > T^{\text{obs}}(\boldsymbol{\theta})|\mathbf{Y}^{obs}, \mathbf{D}^{obs}, \mathbf{C}^{obs}, \mathbf{Z}, \mathbf{X}).$$

PPPVs that are close to 1 or 0 indicate that the prior and liklihood are not able to replicate the measures of signal, noise or the signal-to-noise ratio that was observed in the study and may suggest that the proposed model has unintended influence on the estimated values of the estimands (Gelman, Meng and Stern, 1996; Rubin, 1984). We report PPPVs as the proportion of posterior samples for which the value of the posterior predictive descrepancy measure exceeded the values of the same discrepancy measure computed using the study data. PPPVs were computed for discrepancy measures in both latent compliance strata for both the CMAI and antipsychotic outcome measures; the results are displayed in Table 2.8.

Outcome	$SI_k(\boldsymbol{\theta})$	$NO_k(\boldsymbol{\theta})$	$SI_k(\boldsymbol{\theta})/NO_k(\boldsymbol{\theta})$
CMAI			
$S = 1$	0.66	0.39	0.67
$S = 2$	0.46	0.51	0.46
Antipsychotics			
$S = 1$	0.61	0.39	0.64
$S = 2$	0.36	0.61	0.34

Table 2.8: PPPVs corresponding to different discrepancy measures for CMAI and Anipsychotic models across latent compliance strata.

The PPPVs for all discrepancy measures across both latent compliance strata and both outcomes are between 0.34 and 0.67. Because no PPPV is close to 0 or 1, we find no evidence that the proposed models have unintended influence on the estimated values.

2.6 Discussion

We propose a Bayesian method to estimate the effects of interventions in pragmatic CRCTs subject to noncompliance at the cluster and individual-level. This method defines a Bayesian mixture model that classifies all clusters into compliance strata based on their characteristics and continuous compliance metrics collected during the trial. The model imputes unobserved individual-level binary compliance status and cluster-level compliance metrics for clusters assigned to control and estimates stratum specific individual outcome models. We describe estimands that stratify individuals by their compliance status, as well as their cluster's latent compliance stratum to enable researchers to compare the treatment effects among individuals that complied to intervention and between clusters with comparable levels of partial compliance.

We provide case studies to demonstrate the method's sensitivity to different data generating mechanisms. Simulating the case studies displays the model's high efficiency, accuracy and nominal coverage when it is correctly specified. It also demonstrates that the procedure is robust different data generating mechanisms that lead to misspecification at each level of the model.

We also apply the proposed Bayesian method to METRIcAL, a pragmatic CRCT that examines the effects of a personalized music intervention on the behaviors of nursing home residents with dementia in 54 nursing homes in the United States. To evaluate NH compliance, we analyzed a continuous compliance metric that describes the proportion of enrolled residents who received

the personalized music intervention from NH staff in order to directly target agitated behaviors or reduce antipsychotic use. For the NH characteristic, we analyzed the average daily census, a covariate that describes is the average resident population at the nursing home during the previous year. Estimates of the average treatment effect reveal no observable difference in the change in CMAI or antipsychotic use between the arm receiving the intervention and the arm receiving standard of care. Similarly, the complier average causal effect does not show any significant difference between treatment arms at the 5% nominal level. This result holds regardless of the latent compliance stratum that a NH belonged to. These results align with the original analysis of the METRIcAL trial, which showed no observable difference between the agitated behaviors of residents in NHs assigned to either treatment arm. However, for antipsychotic use, the CACE and the stratum specific CACE among individuals in clusters that most often used the intervention for targeting agitated behaviors had large negative point estimates that approached traditional levels of statistical significance.

In conclusion, the proposed Bayesian procedure provides a new methodology to analyze pragmatic CRCTs subject to noncompliance at the cluster and individual-levels. The procedure estimates the effects of the intervention among similar clusters with similar compliance, and also among individual who comply to the intervention when it is offered to them. We have applied the method to cases that consider one cluster compliance metric and one cluster covariate, but this method can be used for multiple compliance metrics and multiple characteristics simultaneously. Possible extensions include models that analyze continuous compliance at the individual-level.

Chapter 3

A Bayesian Two-stage Adaptive Design for Pragmatic RCTs with One-Sided Noncompliance

3.1 Introduction

Pragmatic randomized controlled trials (PRCTs) are designed to investigate the effects of interventions in real-world settings. In PRCTs, the intervention is typically administered by healthcare providers or by the participants themselves, closely mirroring everyday practice. Because of this, noncompliance to the intervention is more likely (Ware and Hamel, 2011; Zuidgeest et al., 2017; Albert, 1997). Randomized experiments with patient noncompliance are subject to lower statistical power and mitigated treatment effects which can lead to higher costs, longer duration, or the abandonment of effective interventions (Shiovitz et al., 2016; Czobor and Skolnick, 2011). Therefore, it is important to consider noncompliance at the trial's design phase (Hewitt, Torgerson and Miles, 2006; Czobor and Skolnick, 2011; Zhou et al., 2020).

Adaptive RCT designs allow for modification of the trial's protocol based on data collected during the trial without undermining the validity of the study (Berry et al., 2010; Chow, Chang and Pong, 2005). A sequential RCT design is a type of adaptive design that decides between multiple predefined study implementation options using interim analyses of accumulated data (Whitehead,

1997). Multiple sequential designs that incorporate patient compliance data have been proposed. Shen, Ning and Yuan (2015) develop a Bayesian sequential monitoring design that continuously updates estimates of the complier average causal effect (CACE) (Imbens and Rubin, 1997) to inform decision making, while Ren et al. (2021) propose a similar design that uses Bayesian additive regression trees (BART) to estimate the CACE. However, PRCTs typically estimate intention-to-treat (ITT) effects which describe the average effect of being assigned to receive the intervention (Zuidgeest et al., 2017). Moreover, sequential designs with continuous updating may not be practical for PRCTs because PRCTS are often more complex than typical RCTs (Ford and Norrie, 2016), and researchers are already hesitant to initiate complex adaptive designs (Dimairo et al., 2015). An alternative option is to implement a two-stage sequential design, the simplest form of sequential design, which specifies a single intermediate point for conducting interim analyses (Gehan, 1961; Jung et al., 2004; Simon, 1989). However, two-stage sequential designs typically require a larger sample size to maintain the same level of statistical power compared to one-stage designs under similar conditions (Jennison and Turnbull, 1999). Some two-stage sequential designs use Bayesian methods that incorporate additional data when estimating treatment effects to increase statistical power. Su et al. (2022) develop a Bayesian design for biomarker-guided trials that uses a hierarchical model to include data from multiple control arms to improve efficiency. Similarly, Psioda and Xue (2020) use the Bayesian power prior (De Santis, 2006; Ibrahim et al., 2015) to "borrow" information from previously completed trials to estimate treatment effects in pediatric populations. However, these methods have not been used to address noncompliance in pragmatic trials.

Interventions in PRCTs are often multifaceted (Fitzpatrick et al., 2020), with some components contributing to intervention's effectiveness mechanisms and other components contributing to implementation and adherence (Ford and Norrie, 2016; Lugaresi, Rottoli and Patti, 2014; Spilker, 1992; Mayhew et al., 2020). For example, a personalized music protocol may use a unique music playlist to help reduce the agitated behaviors of dementia patients while also providing comfortable headphones to promote consistent listening. Researchers will often collect information from PRCTs about modifiable aspects of the intervention related to implementation and adherence (Apter, 2015; Glasgow et al., 2021; Sandy et al., 2021; Palmer et al., 2019). By adjusting these features for interventions that have poor compliance, researchers may improve patient adherence without modifying the treatment effect mechanisms (Lugaresi, Rottoli and Patti, 2014; Spilker, 1992).

We propose a novel, Bayesian two-stage sequential design for PRCTs that addresses one-sided binary noncompliance (Imbens and Rubin, 2016). In the design, compliance and effectiveness data are accumulated in Stage 1 and used for an interim analysis when the stage is complete. Based on the interim analysis, the trial may be stopped for efficacy, or a new trial will be implemented with an augmented version of the multi-component intervention that improves compliance but keeps the same treatment effect mechanisms. The analyses of this design rely on Bayesian models that utilize a power prior distribution to improve statistical power by incorporating information from Stage 1 when estimating treatment effects in Stage 2. We compare the operating characteristics of the proposed method to a simple randomized design and evaluate its sensitivity to misspecifications using simulations. In addition, we simulate a trial that implements the proposed design and analysis using real data from METRIcAL: a multi-part, pragmatic, nursing home (NH) based trial evaluating the effects of a personalized music intervention on the agitated behaviors of nursing home residents with dementia.

3.2 Methods

3.2.1 Notation

We consider a pragmatic trial with $I = 2$ stages where N_i individuals are enrolled in stage i so that $N = N_1 + N_2$ denotes the total number of study participants. Out of the N_i participants in stage i, N_{i1} and $N_{i0} = N_i - N_{i1}$ represent the number of individuals who are randomized to the active intervention and control condition, respectively. Let $\mathbf{W}_i = \{W_{i1}, \ldots, W_{iN_i}\}$, where $W_{ij} \in \{0, 1\}$ is an intervention assignment indicator that equals 0 if individual j in stage i is randomized to control, and 1 if they were randomized to the active intervention. $D_{ij}(W_i) \in \{0, 1\}$ is a compliance indicator that equals 1 when the individual j in stage i receives the active intervention when their treatment assignment is W_i, and 0 otherwise. We assume that individuals assigned to the control condition have no access to the active intervention so that $D_{ij}(0) = 0$. For simplicity, we define $D_{ij} := D_{ij}(1)$. Let $\mathbf{Y}(0, \mathbf{D}) = \{Y_{ij}(0, D_{ij})\}$ denote the set of potential outcomes for all N_i individuals in Stage i under the control condition, and $\mathbf{Y}(1, \mathbf{D}) = \{Y_{ij}(1, D_{ij})\}$ denote the set of potential outcomes for all N_i individuals under the active intervention. Similarly, let $\mathbf{D} = \{D_{ij}\}$ denote the set of compliance indicators. Lastly, we observe a set of P pre-treatment covariates for each individual,

$\mathbf{X}_{ij} = \{X_{ij}^1, ..., X_{ij}^P\}$.

Because individuals are assigned to only one intervention at a specific point in time, only one of $Y_{ij}(1, D_{ij})$ and $Y_{ij}(0, D_{ij})$ can be observed. We assume the stable unit treatment value assumption (Rubin, 1978*b*) for the potential outcomes and compliance statuses so the observed and missing potential outcomes are

$$Y_{ij}^{obs} = Y_{ij}(1, D_{ij}) * W_{ij} + Y_{ij}(0, D_{ij}) * (1 - W_{ij})$$
$$Y_{ij}^{mis} = Y_{ij}(1, D_{ij}) * (1 - W_{ij}) + Y_{ij}(0, D_{ij}) * W_{ij}.$$

In addition, the observed and missing compliance statuses in Stage i are $\mathbf{D}_i^{obs} = \{D_{ij}|W_{ij} = 1\}$ and $\mathbf{D}_i^{mis} = \{D_{ij}|W_{ij} = 0\}$, respectively.

3.2.2 Interventions and Estimands

We assume that the intervention is multifaceted, with parts of it that influence only compliance, and other parts that act as the active ingredients and influence the effect of the intervention on those who take it. We assume that the active intervention in Stage 1 of the proposed two-stage PRCT is described in the trial's protocol, and we refer to it as the "standard intervention". The active intervention in Stage 2 is a modified version of the "standard intervention" and is based on patients' non-compliance data from Stage 1. Throughout, we refer to this intervention as the "augmented intervention". The augmented and standard intervention are designed so they use identical mechanisms to induce the treatment effect, but the features of the augmented intervention that relate to patient compliance are modified to improve adherence using data collected in Stage 1.

Estimands, which are functions of the potential outcomes, are used to summarize the effects of an intervention across a population of interest (Gutman and Rubin, 2015*b*; Rubin, 1978*b*). Intention-to-treat (ITT) estimands are the most common estimands in PRCTs because they estimate the effects of an intervention using data from all randomized participants (Gupta, 2011). The ITT estimand for the standard intervention in Stage 1 is $E_{sp}[Y_{1j}(1, D_{1j}) - Y_{1j}(0, D_{1j})]$, and the ITT estimand for the augmented intervention in Stage 2 is $E_{sp}[Y_{2j}(1, D_{2j}) - Y_{2j}(0, D_{2j})]$ where sp denotes the expectation over the target super-population. We refer to these estimands as ITT_1 and ITT_2,

respectively.

3.2.3 Two-stage PRCT Design

Decision Rules

We let $\pi_1 = E_{sp}[D_{1j}]$ be the expected compliance to the standard intervention in Stage 1 of the two-stage design, and we let $\hat{\pi}_1 = \frac{1}{N_{11}} \sum_{j:W_{1j}=1} D_{1j}$ be its point estimate. We define $p_U \in (0,1)$ to be a prespecified upper compliance threshold used for decision making at the conclusion of Stage 1. If $\hat{\pi}_1 \geq p_U$ the trial is stopped because of higher than expected compliance and an α-level statistical test is conducted on ITT_1. Similarly, we define $p_L \in (0,1)$ to be a prespecified lower compliance threshold so that if $\hat{\pi}_1 \leq p_L$, the trial proceeds to Stage 2 without an interm analysis of the outcome data. At the conclusion of Stage 2, and α-level statistical test is conducted on ITT_2 to assess the effects of the augmented intervention. If after Stage 1 $p_L < \hat{\pi}_1 < p_U$, a statistical test with significance level $\alpha_1 < \alpha$ is performed on ITT_1. If this test rejects the null hypothesis at the α_1-level, the trial is stopped because of efficacy of the standard intervention. If the null hypothesis is not rejected, the trial continues to Stage 2 where a statistical test with significance level $\alpha_2 < \alpha$ is conducted on ITT_2 to evaluate the effects of the augmented intervention. Figure 3.1 depicts these decision rules in a flowchart.

Design Considerations

In the absence of noncompliance, a formula for calculating the number of participants required to achieve a statistical power of $1 - \beta$ when half of the participants are assigned to the active intervention is

$$N^* = 4 * \left(\frac{Z_{1-\frac{\alpha}{2}} + Z_{1-\beta}}{ES} \right)^2, \tag{3.1}$$

where Z_q is the q^{th} quantile of the standard normal distribution, ES is the expected standardized effect size, and α is the Type 1 error rate (Wittes, 2002). To account for noncompliance, the following adjustment to Equation (3.1) was proposed (Wittes, 2002),

$$N = \frac{N^*}{(1 - p_c - p_t)^2}, \tag{3.2}$$

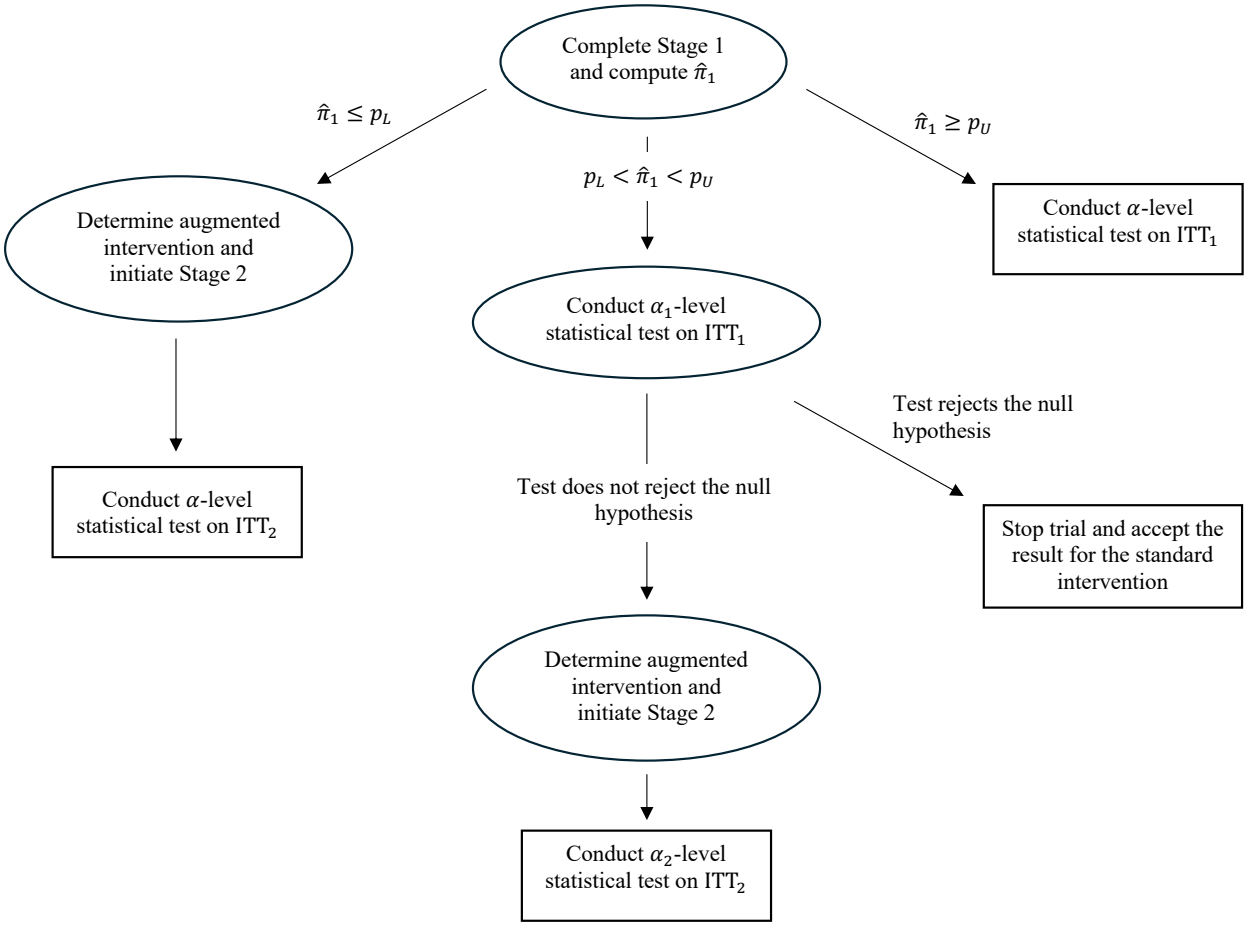

Figure 3.1: Two-stage PRCT design flowchart. $\hat{\pi}_1$: Observed proportion of compliers in Stage 1; p_L: Presepecified lower compliance threshold; p_U: Presepecified upper compliance threshold; α: Presepecified Type-1 error rate; ; α_1, α_2: Type-1 error rates adjusted for multiple testing; ITT_1, ITT_2: Intention-to-treat estimands for the standard and augmented intervention, respectively.

where p_c is the expected proportion of individuals who receive the active intervention when assigned to the control condition, and p_t is the expected proportion of indiviudals who recieve the control condition when assigned to the active intervention. We assume one-sided noncompliance, so $p_c = 0$, and let $p_E = 1 - p_t$ be the expected proportion of individuals who would comply to the standard intervention if they were assigned to receive it. Thus, the total number of individuals enrolled in a PRCT that expects $(100 * p_E)\%$ compliance is

$$N = \frac{N^*}{(p_E)^2} = 4 * \left(\frac{Z_{1-\frac{\alpha}{2}} + Z_{1-\beta}}{p_E * ES} \right)^2 . \tag{3.3}$$

In the proposed two-stage design, we are interested in keeping sample size similar to the equivalent single-stage design. Therefore, we let Equation (3.3) determine the total sample size, N, and allocate N_1 participants to Stage 1 by computing

$$N_1 = 4 * \left(\frac{Z_{1-\frac{\alpha}{2}} + Z_{1-\beta_1}}{p_U * ES} \right)^2, \tag{3.4}$$

where $1 - \beta_1$ denote the desired statistical power in Stage 1 if p_U compliance is attained. The remaining $N_2 = N - N_1$ participants are allocated to the second stage with the expectation that compliance will be higher under the augmented intervention.

Values for α_1 and α_2 are chosen so that the overall Type-1 error rate of the procedure is α. Possible methods to determine values for α_1 and α_2 include the use of alpha-spending functions (Demets and Lan, 1994) and the Bonferronni correction for multiple testing (Armstrong, 2014). The Bonferronni correction requires that each hypothesis test is conducted at level α/t, where t is the total number of tests that are being conducted. Alpha-spending functions, such as the Pocock or O'Brien Flemming functions (Demets and Lan, 1994), allocate the total Type-1 error across multiple interm analyses by incrementally increasing the alpha spent at each hypothesis test as a function of the total information available. Table 3.1 summarizes the all of the parameters associated with the two-stage design and describes their function.

Paramter	Description
N_1, N_2	Total number of individuals enrolled in Stage 1 and Stage 2, respectively.
N	The total number of trial participants, $N_1 + N_2$.
ES	Standardized effect size that is assumed for sample size calculations.
α	Desired type 1 error rate for the PRCT.
$1 - \beta$	Desired statistical power for the PRCT.
p_E	Expected compliance proportion used for sample size calculations.
$1 - \beta_1$	Desired statistical power for Stage 1 of the PRCT.
p_U	Compliance proportion in Stage 1 above which investigators would not consider adjusting the standard intervention.
p_L	Compliance proportion in Stage 1 below which investigators would require the adjustment of the standard intervention.
$\alpha_1,\ \alpha_2$	Restricted Type 1 error rates used for statistical tests if a test is conducted after Stage 1 and stage 2, respectively.

Table 3.1: Parameters for two-stage PRCT design.

3.2.4 Modeling and Assumptions

Distributional Assumptions

We rely on a Bayesian model-based approach to estimate the posterior distributions of ITT_1 and ITT_2. Model-based estimation is a principled approach to incorporate restrictions on the joint distribution of observed variables and include covariates in a flexible manner (Imbens and Rubin, 2016). In PRCTs, the participants's treatment assignment is controlled by the investigators and is decided prior to the study's initiation. Therefore, the joint distribution of the variables in Stage 1 of the two-stage PRCT design is given by

$$\begin{aligned}
&P(\mathbf{W}_1, \mathbf{Y}_1(1, \mathbf{D}_1), \mathbf{Y}_1(0, \mathbf{D}_1), \mathbf{D}_1, \mathbf{X}_1) \\
&= P(\mathbf{W}_1 | \mathbf{Y}_1(1, \mathbf{D}_1), \mathbf{Y}_1(0, \mathbf{D}_1), \mathbf{D}_1, \mathbf{X}_1) P(\mathbf{Y}_1(1, \mathbf{D}_1), \mathbf{Y}_1(0, \mathbf{D}_1) | \mathbf{D}_1, \mathbf{X}_1) P(\mathbf{D}_1 | \mathbf{X}_1) P(\mathbf{X}_1) \\
&\propto P(\mathbf{Y}_1(1, \mathbf{D}_1), \mathbf{Y}_1(0, \mathbf{D}_1) | \mathbf{D}_1, \mathbf{X}_1) P(\mathbf{D}_1 | \mathbf{X}_1) P(\mathbf{X}_1).
\end{aligned}$$

We assume the distribution of variables in Stage 1 does not influence the the distribution of variables in Stage 2, only whether or not Stage 2 is initiated. Therefore, the joint distribution for the variables in Stage 2 can be written similarly,

$$P(\mathbf{W}_2, \mathbf{Y}_2(1, \mathbf{D}_2), \mathbf{Y}_2(0, \mathbf{D}_2), \mathbf{D}_2, \mathbf{X}_2) \propto P(\mathbf{Y}_2(1, \mathbf{D}_2), \mathbf{Y}_2(0, \mathbf{D}_2) | \mathbf{D}_2, \mathbf{X}_2) P(\mathbf{D}_2 | \mathbf{X}_2) P(\mathbf{X}_2).$$

Units are considered exchangeable when their joint distribution is the same regardless of their ordering (Bernardo, 1996). Individual-level covariates, compliance status and the potential outcomes are not informed by their assignment to treatment. Therefore, we make the following partial exchanagability assumptions.

Assumption 1 Exchangability

(a) In Stage i, individual compliance status is exchangeable conditional on $\mathbf{X}_{ij}$. Assuming the prior $P(\theta_{D_i})$ and appealing to De-Finetti's theorem (de Finetti, 1992), we can write $P(\mathbf{D}_i | \mathbf{X}_i)$ as

$$P(\mathbf{D}_i | \mathbf{X}_i) \propto \int \prod_{j=1}^{N_i} P(D_{ij} | \theta_{D_i}, \mathbf{X}_{ij}) P(\theta_{D_i}) d\theta_{D_i}. \tag{3.5}$$

(b) Furthermore, in Stage i, the potential outcomes are exchangable conditional on $\mathbf{X}_{ij}$ and D_{ij}. Assuming the prior $P(\theta_{Y_i})$ and appealing to De-Finetti's theorem (de Finetti, 1992) this implies

$$P(\mathbf{Y}_i(1,\mathbf{D}_i),\mathbf{Y}_i(0,\mathbf{D}_i)|\mathbf{D}_i,\mathbf{X}_i) \propto \int \prod_{j=1}^{N_i} P(Y_{ij}(1,D_{ij}),Y_{ij}(0,D_{ij})|\theta_{Y_i},D_{ij},\mathbf{X}_{ij})P(\theta_{Y_i})d\theta_{Y_i} \tag{3.6}$$

A commonly made assumption in randomized trials with noncompliance is the exclusion restriction assumption (Angrist, Imbens and Rubin, 1996), which rules-out the effects of the assignment on the outcome for non-compliers. We assume a slightly weaker version of this assumption for both the standard and augmented intervention that requires it to hold in distribution.

Assumption 2: Stochastic exclusion restriction for individuals that do not comply with the active intervention (Imbens and Rubin, 2016). $P(Y_{ij}(1,0)) = P(Y_{ij}(0,0))$ for $i = 1, 2$.

This assumption presumes that the assignment of an individual to the intervention does not influence the distribution of the potential outcomes except through an individual's compliance to the intervention. This assumption is reasonable in PRCTs when the intervention is embedded in regular patient care because, generally, there is not going to be an effect of the intervention on individuals unless they take it.

We rely on two additional assumptions that simplify the parametric models of the different distributions. The first assumption presumes that the distributions of the potential outcomes are independent given an individual's compliance status and their observed characteristics.

Assumption 3: No contamination of imputation across treatments (Rubin, 2008). Formally, $P(Y_{ij}(1,D_{ij}),Y_{ij}(0,D_{ij})|D_{ij},\mathbf{X}_{ij},\theta_{Y_i}) = P(Y_{ij}(1,D_{ij})|D_{ij},\mathbf{X}_{ij},\theta_{Y_i})P(Y_{ij}(0,D_{ij})|D_{ij},\mathbf{X}_{ij},\theta_{Y_i})$ for $i = 1, 2$.

This assumption is commonly made by causal inference methods because the observed data does not include information on the relationship between the potential outcomes for the same individual

(Hirano et al., 2000; Imbens and Rubin, 2016). When the units in a study are a random sample from a super-population, and the estimand is the average difference in this super-population, this assumption has little inferential effect for average treatment effects (Hirano et al., 2000; Imbens and Rubin, 2016; Rubin, 1978*b*; Imbens and Rubin, 1997). The second assumption requires that the compliance status of an individual is conditionally independent from their outcome under the control.

Assumption 4: Conditional independence of compliance status and control potential outcomes. $Y(0, D_{ij}) \perp D_{ij} | \mathbf{X}_{ij}$.

This assumption is plausible when all of the covariates used to model compliance status are also used to model the potential outcomes because compliers and noncompliers do not receive different control interventions. Importantly, this assumption does not require the marginal distribution of the control potential outcomes for compliers and noncompliers to be the same. In addition, the sensitivity of the models to this assumption can be determined by assessing the change in results when including compliance status in the control potential outcome model. Additional details and justification are provided in Appendix Section A3.1.

Lastly, we assume that the potential outcomes follow the same distributions in Stage 1 and Stage 2 conditional on baseline covariates and compliance status.

Assumption 5: Equivalence of potential outcome distributions in Stages 1 and 2. Formally, $P(Y_{1j}(w, d) | D_{1j} = d, \mathbf{X}_{1j} = X, \theta_{Y_1}) = P(Y_{2j}(w, d) | D_{2j} = d, \mathbf{X}_{2j} = X, \theta_{Y_2})$, and $\theta_{Y_1} = \theta_{Y_2} = \theta_Y$.

This assumption is plausible because the augmented intervention and standard intervention utilize identical mechanisms to induce the treatment effect. Therefore, the conditional distributions of the potential outcomes among individuals who receive either the standard or augmented intervention are expected to be the same. In addition, Stage 1 and Stage 2 have the same control condition. Therefore, we expect the distributions of the potential outcomes for individuals assigned to the control to be the same in both stages if they are sampled from the same target population.

Parametric Models

We develop a model based approach to estimate ITT_1 and ITT_2 that incorporates individual level covariates and follows the distributional assumptions outlined in Section 3.2.4. We model the distribution of compliance to the active intervention in Stage i as

$$P(D_{ij}|\theta_{D_i}, \mathbf{X}_{ij}) = \text{Bernoulli}\left(\text{probit}^{-1}\left(\mu_i^D + \mathbf{X}_{ij}\eta_i\right)\right) \text{ for } i = 1, 2, \tag{3.7}$$

where $\theta_{D_i} = \{\mu_i^D, \eta_i\}$, μ_i^D denotes the intercept and α_i is a coefficient vector that describes the relationship between the observed covariates and the probability of being a complier. The conditional distributions for the potential outcomes are modeled as

$$\begin{aligned} P(Y_{ij}(1, D_{ij})|\theta_Y, D_{ij}, \mathbf{X}_{ij}) &= N\left(Y_{ij}(1, D_{ij})|\mu^Y + \mathbf{X}_{ij}\beta_0 + D_{ij}\delta_1, \sigma^2\right) \\ P(Y_{ij}(0, D_{ij})|\theta_Y, D_{ij}, \mathbf{X}_{ij}) &= N\left(Y_{ij}(0, D_{ij})|\mu^Y + \mathbf{X}_{ij}\beta_0, \sigma^2\right), \end{aligned} \tag{3.8}$$

where $\theta_Y = \{\mu^Y, \beta_0, \delta_1, \sigma^2\}$, μ^Y is the intercept and β_0 is a vector of coefficients that that describe the relationship between covariates and the expected value of the potential outcomes. The parameter δ_1 describes the differential effect of being a complier to the active intervention and σ^2 denotes the variance of the potential outcome distributions. Assuming the study populations in Stage 1 and Stage 2 are simple random samples from the target super-population, ITT_1 and ITT_2 as functions of the model parameters are given by

$$\text{ITT}_1 = \frac{1}{N_1}\sum_{j=1}^{N_1} \text{probit}^{-1}\left(\mu_1^D + \mathbf{X}_{1j}\eta_1\right) \times \delta_1 \tag{3.9}$$

$$\text{ITT}_2 = \frac{1}{N_2}\sum_{j=1}^{N_2} \text{probit}^{-1}\left(\mu_2^D + \mathbf{X}_{2j}\eta_2\right) \times \delta_1. \tag{3.10}$$

Bayesian Power Prior

Bayesian power prior distributions are a class of informative prior distributions that are constructed from "historical" data (Ibrahim et al., 2015). The power prior for a parameter θ is given by $P(\theta|\Omega_0, a_0) \propto L(\theta|\Omega_0)^{a_0} P_0(\theta)$, where $L(\theta|\Omega_0)$ is the liklihood function based on the historical data, Ω_0, $P_0(\theta)$ is the initial prior for θ before incorporating historical data, and a_0 is the borrowing

parameter taking values between 0 and 1. This prior distribution implies the posterior distribution for θ given by $P(\theta|\Omega, \Omega_0, a_0) \propto L(\theta|\Omega)L(\theta|\Omega_0)^{a_0}P_0(\theta)$, where $L(\theta|\Omega)$ is the current data liklihood function. The borrowing parameter controls the influence of the historical data by weighting the historical data liklihood relative to the current data likelihood in the posterior distribution. Practically, setting $a_0 = 0$ includes no historical information in the posterior distribution, while setting $a_0 = 1$ includes historical data and current data with the same weight.

The partial borrowing power prior is a type of power prior that only borrows historical information on the common parameters that are shared between the current and historical data generating mechanisms (Ibrahim and Chen, 2000; Ibrahim et al., 2012; Chen, Ibrahim, Zeng, Hu and Jia, 2014; Chen, Ibrahim, Amy Xia, Liu and Hennessey, 2014). In the analysis of Stage 2 of the PRCT, we assume the partial borrowing power prior to borrow information from Stage 1 to inform posterior estimates of θ_Y. Let $\mathbf{O}_i = \{\mathbf{Y}_i^{obs}, \mathbf{D}_i^{obs}, \mathbf{X}_i, \mathbf{W}_i\}$. The observed-data liklihood function for θ_Y and θ_{D_i} in Stage i of the PRCT is given by

$$
\begin{aligned}
&L_i(\theta_Y, \theta_{D_i}|\mathbf{O}_i) = \\
&\prod_{j:W_{ij}=0} N\left(Y_{ij}(0, D_{ij})|\mu^Y + \mathbf{X}_{ij}\beta_0, \sigma^2\right) \prod_{j:W_{ij}=1} \Bigg\{ N\left(Y_{ij}(1, D_{ij})|\mu^Y + \mathbf{X}_{ij}\beta_0 + D_{ij}\delta_1, \sigma^2\right) \\
&\quad \times \text{probit}^{-1}\left(\mu_i^D + \mathbf{X}_{ij}\eta_i\right)^{D_{ij}} \left(1 - \text{probit}^{-1}\left(\mu_i^D + \mathbf{X}_{ij}\eta_i\right)\right)^{1-D_{ij}} \Bigg\}, \qquad (3.11)
\end{aligned}
$$

where $\theta_Y = \{\mu^Y, \beta_0, \delta_1, \sigma^2\}$ and $\theta_{D_i} = \{\mu_i^D, \eta_i\}$. The partial borrowing power prior for θ_Y and θ_{D_2} that incorporates the observed-data liklihood function for θ_Y from Stage 1 is given by

$$
P(\theta_Y, \theta_{D_2}|\mathbf{O}_1, a_0) \propto \left(\int L_1(\theta_Y, \theta_{D_1}|\mathbf{O}_1)^{a_0} P_0(\theta_{D_1}) d\theta_{D_1} \right) \times P_0(\theta_Y, \theta_{D_2}), \qquad (3.12)
$$

where a_0 is the borrowing parameter and $P_0(\theta_{D_1})$ and $P_0(\theta_Y, \theta_{D_2})$ are initial prior distributions that assume that θ_{D_1} is *a priori* independent from θ_Y and θ_{D_2}. The observed data liklihood in

Stage 1 can be factored into two components, $L_1(\theta_Y|\mathbf{O}_1)$ and $L_1(\theta_{D_1}|\mathbf{O}_1)$, where,

$$\begin{aligned} L_1(\theta_Y|\mathbf{O}_1) = & \prod_{j:W_{1j}=0} N\left(Y_{1j}(0,D_{2j})|\mu^Y + \mathbf{X}_{1j}\beta_0, \sigma^2\right) \\ & \times \prod_{j:W_{1j}=1} N\left(Y_{1j}(1,D_{1j})|\mu^Y + \mathbf{X}_{1j}\beta_0 + D_{1j}\delta_1, \sigma^2\right), \text{ and} \\ L_1(\theta_{D_1}|\mathbf{O}_1) = & \prod_{j:W_{1j}=1} \text{probit}^{-1}\left(\mu_1^D + \mathbf{X}_{ij}\eta_1\right)^{D_{1j}} \left(1 - \text{probit}^{-1}\left(\mu_1^D + \mathbf{X}_{ij}\eta_1\right)\right)^{1-D_{1j}}. \end{aligned}$$

Therefore, we can simplify Equation (3.12) so that

$$\begin{aligned} P(\theta_Y, \theta_{D_2}|\mathbf{O}_1, a_0) &\propto \left(\int L_1(\theta_Y, \theta_{D_1}|\mathbf{O}_1)^{a_0} P_0(\theta_{D_1}) d\theta_{D_1}\right) \times P_0(\theta_Y, \theta_{D_2}) \\ &= L_1(\theta_Y|\mathbf{O}_1)^{a_0} \left(\int L_1(\theta_{D_1}|\mathbf{O}_1)^{a_0} P_0(\theta_{D_1}) d\theta_{D_1}\right) \times P_0(\theta_Y, \theta_{D_2}) \\ &\propto L_1(\theta_Y|\mathbf{O}_1)^{a_0} P_0(\theta_Y, \theta_{D_2}), \end{aligned} \tag{3.13}$$

where the integral is finite whenever $P_0(\theta_{D_1})$ is proper because $L_1(\theta_{D_1}|\mathbf{O}_1)$ is bounded between 0 and 1. The posterior distribution for θ_Y and θ_{D_2} that assumes the partial borrowing prior distribution in Equation (3.13) is given by

$$\begin{aligned} P(\theta_Y, \theta_{D_2}|\mathbf{O}_1, \mathbf{O}_2, a_0) &\propto L_2(\theta_Y, \theta_{D_2}|\mathbf{O}_2) P(\theta_Y, \theta_{D_2}|\mathbf{O}_1, a_0) \\ &\propto L_2(\theta_Y, \theta_{D_2}|\mathbf{O}_2) L_1(\theta_Y|\mathbf{O}_1)^{a_0} P_0(\theta_Y, \theta_{D_2}). \end{aligned} \tag{3.14}$$

Under the partial borrowing power prior, the posterior inference on ITT_2 will depend on the observed data in Stage 2 and Stage 1, the choice of $P_0(\theta_Y, \theta_{D_2})$, and the choice of a_0.

The Choice of the Borrowing Parameter, a_0

Several methods have been proposed for determining a value of the borrowing parameter, a_0. One method is to set a_0 to a fixed value based on the relative amount of historical information investigators want to incorporate in the analysis, and then conduct several sensitivity analyses that modify the selected value (Ibrahim et al., 2015). A different approach imposes a prior distribution on a_0 and estimates its posterior distribution. Ibrahim and Chen (2000) define the joint power prior

which assigns a joint prior distribution on the borrowing parameter and the model parameters, while Duan, Ye and Smith (2006) propose the normalized power prior that first assumes a prior distribution for the model parameters conditional on a_0, then defines a marginal prior distribution on a_0. Ye et al. (2022) describe the normalized power prior and its variations, including the partial borrowing normalized power prior, and derive the normalized power prior in the Normal linear regression setting.

Another set of methods determine a value for a_0 by measuring the similarity of the historical data and the current data (Ibrahim et al., 2015). Several methods use the propensity score (Rosenbaum and Rubin, 1983*b*) to weight historical data observations in current analyses based on similarities in baseline covariates (Lin, Gamalo-Siebers and Tiwari, 2019; Lu et al., 2022; Wang et al., 2019; Wang, Zhang and Tiwari, 2024). Gravestock and Held (2017) propose a different, Empirical Bayes, method to set a_0 by maximizing its marginal liklihood. Han, Ye and Wang (2023) derive the objective function for this approach and a different DIC (Spiegelhalter et al., 2002) based approach when the data is modeled using Normal linear regression. The applications of the Emprical Bayes and DIC methods to the proposed two-stage PRCT design are described in Appendix Section A3.3. For additional procedures that use the data to set a deterministic value for a_0, see Ibrahim et al. (2015).

3.3 Simulation Study on the Operating Charteristics of the Proposed Two-Stage PRCT Design

3.3.1 Designs and Models

We develop factorial simulation studies where we control the data generating mechanism to examine the operating characteristics of the proposed two-stage design. These results are compared to a one-stage, simple randomized design that assigns N participants to either the standard intervention or control condition. Values for N, N_1 and N_2 are determined using the procedures described in Section 3.2.3 when the expected standardized effect size among compliers, ES, is 0.5 and when the expected proportion of compliers under the standard intervention, p_E, is 0.3, 0.5 and 0.7. All other design parameters are held constant and can be found in Appendix Table A3.1.

We implement four Bayesian models to analyze the two-stage design that use different values for the borrowing parameter in the power prior distribution: 0, 0.5, 1, and a_0^{DIC}. The parameter a_0^{DIC} corresponds to the borrowing parameter obtained by optimizing the DIC, described in Appendix Section A3.3, and was shown to have favorable operating characteristics when compared to Empirical Bayes and the normalized power prior in Normal linear regression models (Han, Ye and Wang, 2023). The four Bayesian models are denoted by $\mathrm{B}_0, \mathrm{B}_{0.5}, \mathrm{B}_1, \text{and } \mathrm{B}_{DIC}$, respectively, and follow the distributional assumptions outlined in Section 3.2.4. We assume mutually independent inital prior distributions for $\theta_{D_1}, \theta_{D_2}$ and θ_Y with $P_0(\theta_Y) \propto \frac{1}{\sigma^2}$ and $P_0(\theta_{D_i}) = \mathrm{MVN}(\theta_{D_i}|0, 100 * \mathbf{I}_{P+1})$ for $i = 1, 2$ where $\mathbf{I}$ is the identity matrix. A Normal linear regression model and a two-part regression model are used to analyze the simple randomized design. The Normal linear regression, denoted by LM_W, assumes

$$\begin{aligned} \mathbf{Y}^{obs} &= \mu_W^Y + \mathbf{X}\beta_W + \mathbf{W}\delta_W + \boldsymbol{\epsilon} \\ \boldsymbol{\epsilon} &\sim \mathrm{MVN}(0, \sigma_W^2 \mathbf{I}_N), \end{aligned} \tag{3.15}$$

where $\mathbf{W}$ is the vector of N treatment assignment indicators. The two-part model, denoted by LM_C, assumes

$$\begin{aligned} \mathbf{Y}^{obs} &= \mu_C^Y + \mathbf{X}\beta_C + \mathbf{D}^*\delta_C + \boldsymbol{\epsilon} \\ \boldsymbol{\epsilon} &\sim \mathrm{MVN}(0, \sigma_C^2 \mathbf{I}_N) \\ D_j^* &= \begin{cases} D_j & W_j = 1 \\ 0 & W_j = 0 \end{cases} \text{ for } j = 1, ..., N \\ D_j &\sim \text{Bernoullli}(\text{probit}^{-1}(\mu_C^D + \mathbf{X}_j\alpha_C)) \text{ for } j \text{ such that } W_j = 1. \end{aligned} \tag{3.16}$$

For each model, we assume the study population is a random sample from a target super-population

and estimate the ITT estimands as a function of the model parameters:

$$\text{ITT}_W = \delta_W, \tag{3.17}$$

$$\text{ITT}_C = \frac{1}{N}\sum_{j=1}^{N} \text{probit}^{-1}(\mu_C^D + \mathbf{X}_j\eta_C) \times \delta_C. \tag{3.18}$$

To estimate the confidence intervals for ITT_C, we implement the basic bootstrap confidence limits described by Davison and Hinkley (1997) with 100 bootstrapped data sets.

3.3.2 Correctly Specified Data Generating Mechanism

We first compare the two-stage design and Bayesian analysis to the one-stage design under a data generating mechanism that is correctly specified. We let the probability that individual j complies to intervention offered Stage i be

$$Pr(D_{ij} = 1|\mathbf{X}_{ij}, \omega_{D_i}) = \text{probit}^{-1}\left(m_i^D + \mathbf{X}_{ij}\zeta_i\right), \tag{3.19}$$

where $\omega_{D_i} = \{m_i^D, \zeta_i\}$, and $\mathbf{X}_{ij}$ is a covariate vector containing 2 baseline covariates. We set the distribution of the potential outcomes for individual j given treatment assignment W_{ij} to be

$$P(Y_{ij}(W_{ij}, D_{ij})|\omega_Y, \mathbf{X}_{ij}) = N(Y_{ij}(W_{ij}, D_{ij})|m^y + \mathbf{X}_{ij}\xi_0 + D_{ij} * \gamma_1, 1) \tag{3.20}$$

where $\omega_Y = \{m^y, \xi_0, \gamma_1, \tau^2\}$ and γ_1 is the standardized effect size among compliers. In addition, we create a target super-population by generating 10,000 covariate vectors, $\{\mathbf{X}^{(1)}, ..., \mathbf{X}^{(10,000)}\}$, from a bivariate Normal distribution,

$$\mathbf{X}^{(m)} \sim \text{MVN}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & -0.1 \\ -0.1 & 1 \end{bmatrix}\right) \text{ for } m = 1, \ldots, 10,000. \tag{3.21}$$

The configurations of the simulation study are defined by the factors γ_1, π_1, and φ, where $\pi_1 = E_{sp}[D_{1j}]$, and $\varphi \in [0, 1]$ relates π_1 and π_2 by

$$\pi_2 = \pi_1 + \varphi * (1 - \pi_1). \tag{3.22}$$

For example, if $\varphi = 0.5$, the expected proportion of noncompliers in Stage 2 is half of the expected proportion of noncompliers Stage 1. Table 3.2 describes these factors and their levels, yielding a factorial design with $3 \times 4 \times 4 = 48$ configurations. The values of the model parameters used to obtain these factor levels are shown in Appendix Table A3.2.

Factor	Levels	Description
π_1	$\{0.25, 0.5, 0.75\}$	True expected value for the proportion of compliers to the standard intervention.
φ	$\{0\%10\%, 33\%, 50\%\}$	Reduction in expected noncompliance under the augmented intervention compared to the standard intervention.
γ_1	$\{0, 0.2, 0.5, 0.8\}$	True standardized conditional effect of the interventions among compliers.

Table 3.2: Factors for simulation study with correctly specified data generating mechanism

For each configuration defined by a different combination of π_1, φ and γ_1, 1000 data sets were generated. The power, average sample size, bias, average 95% interval coverage and width were computed when estimating the ITT estimands for each data set. Because all of the models for both designs are correctly specified for the data generating mechanism, they produce unbiased estimates of the ITT estimands. Therefore, we only report power and average sample size to compare them. When $\gamma_1 = 0.2, 0.5$ and 0.8, the estimated power of each design and model combination was computed as the proportion 95% intervals that did not contain 0. When $\gamma_1 = 0$, Type 1 error was calculated using the same procedure. Figure 3.2 reports the results for each design and model combination when $\gamma_1 = 0.5$.

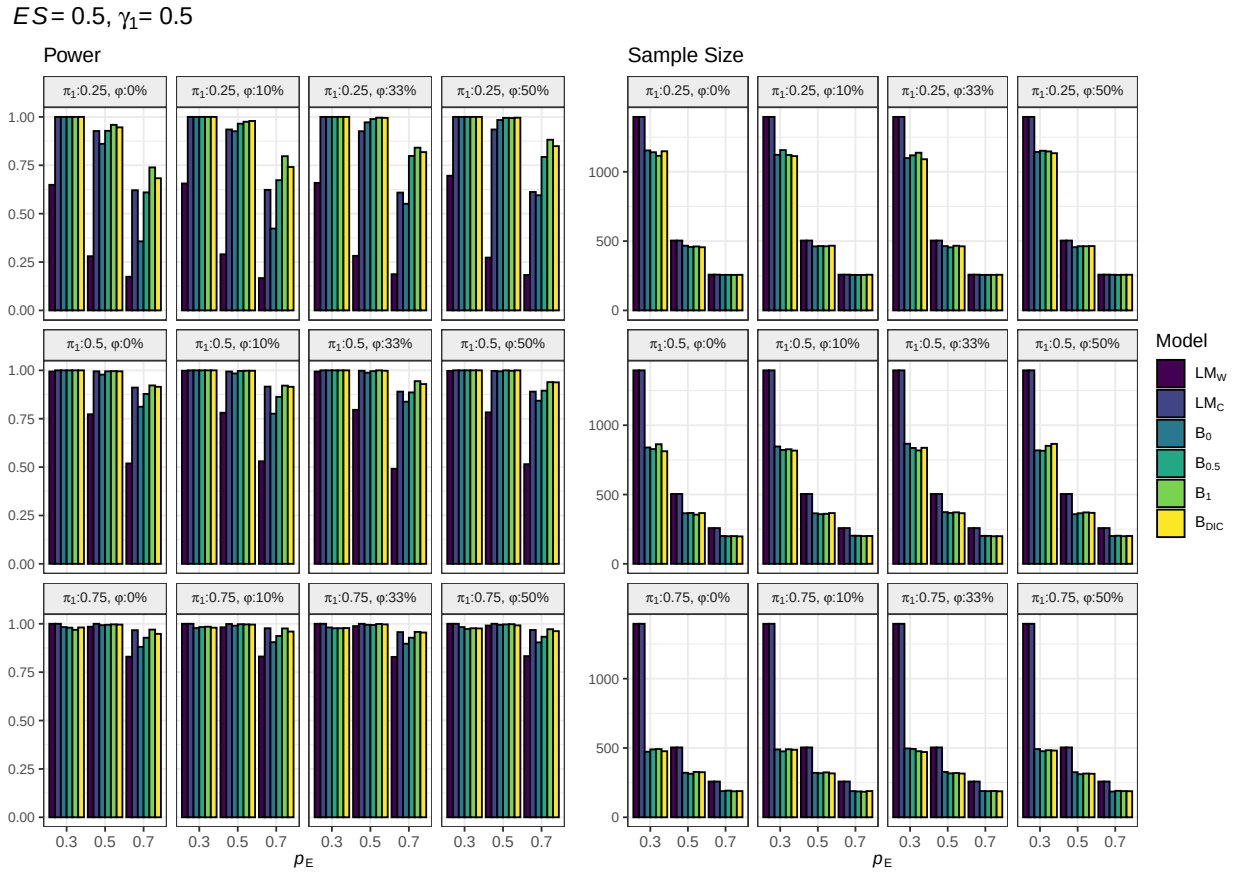

Figure 3.2: Power and average sample size comparison for all configurations when $\gamma_1 = 0.5$.

The rows in Figure 3.2 display the results when the true noncompliance probability under the standard intervention, π_1, is 0.25, 0.5 and 0.75, and the columns display the results when the percent reduction in expected noncompliance due to the augmented intervention, φ, is 0%, 10%, 25% and 50%. The Bayesian models for analyzing the two-stage design provided more power than the simple randomized design analyzed by LM_W. The two-stage design analyzed by B_{DIC} also produced similar power or better power than LM_C in all scenarios. The value of the borrowing parameter was most influential on the power of the Bayesian models when the true compliance probability under the standard intervention was low, $\pi_1 = 0.25$, but the expected compliance probability under the standard intervention was high, $p_E = 0.75$. Because the two-stage design allows for early-stopping, the average sample size analyzed by the Bayesian models was less than or equal to the average sample size analyzed by the regression models applied to the simple randomized design. When $p_E = 0.3$ and $\pi_1 = 0.5$, the average sample size for the two-stage design was approximately half of the simple randomized design, with similar power. When $p_E = 0.3$ and $\pi_1 = 0.75$, the two-

stage design provided similar power to the standard design with less than half the average sample size across all analysis methods. In addition, when the augmented intervention did not reduce the true noncompliance probability, φ =0%, the proposed two-stage design analyzed by B_{DIC} still provided similar power to the simple randomized design in all scenarios while using a smaller or equal average sample size. These results suggest that when the data generating mechanism and standardized effect size are correctly specified, the proposed two-stage design and B_{DIC} model provide better operating characteristics than the simple randomized design analyzed with either LM_W or LM_C.

The results when $\gamma_1 = 0.8$ and $\gamma_1 = 0.2$ are similar (Figures A3.1 and A3.2 in the Appendix), except when $\gamma_1 = 0.2$, $p_E = 0.3$, and $\pi_1 = 0.75$. Under this configuration, the two-stage design provides less statistical power compared to the one-stage design. This occurs because the two-stage design terminates at the end of Stage 1 when high compliance is observed, but does not detect the smaller than expected effect size in the reduced sample. This scenario requires that the true standardized effect size is much less than the expected effect size and the true probability of compliance is much higher than the expected probability used for sample size calculations. This scenario is not common in practice because low patient compliance is often one of the main challenges associated with analyzing pragmatic trials (Resnick et al., 2022; Mayhew et al., 2020; Hossain and Karim, 2023) and a standardized effect size of 0.2 may not be clinically relevant when the study is powered for a standardized effect sized of 0.5.

Table A3.3 in the Appendix shows the average Type 1 error rate for each design and model combination at each value of φ. All combinations produced approximately nominal Type 1 error under each configuration. These results suggest that the Bayesian analyses of the proposed two-stage design are valid statsitical procedures that offer similar or better power and smaller average sample size than the one-stage design in almost all scenarios.

3.3.3 Misspecified Data Generating Mechanism

To evaluate the two-stage design's sensitivity to incorrect model assumptions, we modify the data generating mechanism described in Section 3.3.2 so that it is misspecified.

Different Model Parameters in Stage 1 and Stage 2

We develop a simulation study that modifies the true model coefficients and conditional treatment effect so they are different for the standard and augmented interventions, violating the assumption that $\theta_{Y_1} = \theta_{Y_2}$ (Assumption 5). We set the distributions of the potential outcomes in Stage 1 and Stage 2 to be

$$P(Y_{1j}(W_{1j}, D_{1j})|\omega_{Y_1}, \mathbf{X}_{1j}) = N(Y_{1j}(W_{1j}, D_{1j})|m^y + \mathbf{X}_{1j}\xi_{01} + D_{1j} * \gamma_{11}, 1), \text{ and} \tag{3.23}$$

$$P(Y_{2j}(W_{2j}, D_{2j})|\omega_{Y_2}, \mathbf{X}_{2j}) = N(Y_{2j}(W_{2j}, D_{2j})|m^y + \mathbf{X}_{2j}\xi_{02} + D_{2j} * \gamma_{21}, 1), \tag{3.24}$$

where $\omega_{Y_i} = \{m^y, \xi_{0i}, \gamma_{i1}\}$, and we let the conditional distribution of compliance status be defined as in Equation (3.19). Table 3.3 describes factors of the simulation study and their levels. The remaining parameter values are available in Appendix Table A3.4.

Factor	Levels	Description
ξ_{01}	$\begin{bmatrix} 0.5 & 0.5 \end{bmatrix}^t$	Coefficients in data generating model under the standard intervention
ξ_{02}	$\begin{bmatrix} -0.5 & -0.5 \end{bmatrix}^t$	Coefficients in data generating model under the augmented intervention
γ_{11}	$\{0.25, 0.5\}$	Standardized conditional effect of the standard intervention among compliers
γ_{21}	0.5	Standardized conditional effect of the augmented intervention among compliers
π_1	0.5	True expected propotion of compliers under the standard intevention
φ	$\{0.2, 0.4\}$	Reduction in noncompliance due to the augmented intervention

Table 3.3: Factors of simulation study that modifies the data generating model coeffficients so they are different under the standard and augmented intervention.

We modify γ_{11} and φ and hold all other factors constant to yield a simulation study with 4 configurations. For each configuration, we generate 1000 data sets and implement the two-stage design and analysis using B_0, $\mathrm{B}_{0.5}$, B_1, and B_{DIC}. Table 3.4 reports the bias and coverage of the procedure when estimating the ITT estimands on each of the 1000 data sets.

The Bayesian models provide approximately nominal coverage and similar bias when ξ_{01} and ξ_{02}

Method	$\gamma_{11} = 0.25$, $\gamma_{21} = 0.5$				$\gamma_{11} = \gamma_{21} = 0.5$			
	$\varphi = 0.2$		$\varphi = 0.4$		$\varphi = 0.2$		$\varphi = 0.4$	
	Bias	Coverage	Bias	Coverage	Bias	Coverage	Bias	Coverage
B_0	0.01	0.94	0.01	0.95	0.03	0.97	0.04	0.97
$B_{0.5}$	-0.02	0.95	-0.02	0.94	0.02	0.98	0.03	0.97
B_1	-0.03	0.90	-0.04	0.87	0.02	0.96	0.02	0.97
B_{DIC}	0.01	0.94	0.00	0.95	0.03	0.97	0.03	0.96

Table 3.4: Bias and Coverage of the two-stage design analyzed by four Bayesian Models for each of the four configurations of the simulation study.

are different, and $\gamma_{11} = \gamma_{12}$. When $\gamma_{11} \neq \gamma_{12}$, the bias increases and coverage is below nominal as the value of a_0 increases. However, the bias remains relatively small and the coverage remains close to nominal when setting a_0 using the DIC based procedure. This indicates that the model results are more sensitive to differences in the conditional treatment effect among compliers in Stage 1 and Stage 2 than to differences in the model coefficients. In addition, the DIC method for setting the borrowing parameter is relatively robust to violations of Assumption 5.

Unobserved Covariate

We develop a simulation study to evaluate the sensitivity of the proposed design and analysis to unobserved covariates. Let $\mathbf{U}_i = \{U_{i1}, ..., U_{iN_i}\}$ denote an unobserved covariate, independent from the observed covariates, for the participants in Stage i. We let U_{ij} be normally distributed with variance 1 and expected value m_i^u. In addition, we set $m_1^u = 0$ so that m_2^u represents the difference in the expected value of the distribution of the unobserved covariate in Stages 1 and 2. We generate the potential outcomes and compliance from the following distributions,

$$\begin{aligned} Y_{1j}(W_{1j}, D_{1j}) \sim & N(m^y + \mathbf{X}_{1j}\xi_0 + U_{1j} * \kappa_y + W_{1j} * D_{1j} * \gamma_1, 1) \\ D_{1j} \sim & \text{Bernoulli}\left(\text{probit}^{-1}\left(m_1^d + \mathbf{X}_{1j}\zeta_1 + U_{1j} * \kappa_d\right)\right) \\ Y_{2j}(W_{2j}, D_{2j}) \sim & N(m^y + \mathbf{X}_{2j}\xi_0 + U_{2j} * \kappa_y + W_{2j} * D_{2j} * \gamma_1, 1) \\ D_{2j} \sim & \text{Bernoulli}\left(\text{probit}^{-1}\left(m_2^d + \mathbf{X}_{2j}\zeta_2 + U_{2j} * \kappa_d\right)\right). \end{aligned}$$

The unobserved covariates U_1 and U_2 are correlated with the potential outcomes and compliance through parameters κ_y and κ_d, respectively. Table 3.5 describes factors for the simulation study and their levels. The remaining parameter values are described in Appendix Table A3.5.

Factor	Levels	Description
m_2^u	$\{0, 0.25, 0.5\}$	Difference in the expected value of the standardized unobserved covariate in Stages 1 and 2
κ_D	$\{0, 0.25, 0.33\}$	Conditional effect of the unobserved covariate on compliance
κ_Y	$\{0, 0.25, 0.5\}$	Conditional effect of the unobserved covariate on the potential outcomes
γ_1	0.5	Standardized conditional effect of the active intervention among compliers
π_1	0.5	True expected propotion of compliers under the standard intevention
φ	10%	Reduction in noncompliance because of the augmented intervention

Table 3.5: Factors of simulation study that modifies the data generating model so an unobserved covariate is correlated with both the potential outcomes and compliance.

We modify m_2^u, κ_D and κ_Y to produce a simulation study with $3^3 = 27$ configurations. For each configuration, we generate 1000 data sets and implement the two-stage design and analysis using B_0, $B_{0.5}$, B_1, and B_{DIC} when $p_E = 0.8$. We choose an expected effect size of 0.8 so it is more likely the simulated two-stage trials proceed to Stage 2 and the influence of both sets of unobserved covariates can be examined.

The Bayesian models provided approximately nominal coverage and minimal bias in 24 of the 27 configurations. The three configurations that the Bayesian models had relatively larger bias and did not provide nominal coverage occurred when $\kappa_D = 0.33$ and $\kappa_Y = 0.5$, the largest values of each parameter. Table 3.6 displays the bias and coverage for these configurations.
The Bayesian models covered the true ITT effect in 90-94% of the data sets and had biases between 0.11 and 0.13. The bias and coverage were similar for all Bayesian models and for all levels of m_2^u. This simulation implies that the model results are most sensitive to an unobserved covariate when it has a relatively large correlation with both the potential outcomes and the compliance. The choice a_0 and the level of m_2^u did not modify the results.

Method	$\kappa_D = 0.33$, $\kappa_Y = 0.5$					
	$m_2^u = 0$		$m_2^u = 0.25$		$m_2^u = 0.5$	
	Bias	Coverage	Bias	Coverage	Bias	Coverage
B_0	0.13	0.90	0.13	0.92	0.12	0.93
$B_{0.5}$	0.11	0.93	0.11	0.94	0.13	0.92
B_1	0.11	0.93	0.11	0.93	0.12	0.92
B_{DIC}	0.11	0.93	0.11	0.93	0.12	0.90

Table 3.6: Bias and Coverage of the two-stage design analyzed by four Bayesian Models for the three configurations of the simulation study when coverage was below nominal.

3.4 Implementation of the Two-Stage Design using Real Pragmatic Trial Data

3.4.1 Description of METRIcAL

METRIcAL is a two-stage , pragmatic, NH based, cluster randomized control trial (CRCT) that was implemented NHs across 4 corporate chains in the US (Mor, 2021). The trial evaluated the effects of personalized music on the aggitated behaviors and medication usage of nursing home residents with moderate to severe dementia. In the each stage of METRIcAL, residents in 27 NHs were randomized to the personalized music intervention, and residents in 27 NHs were randomized to the control condition. Participants in the active intervention group received iPods that included music that they listened to as young adults. The control condition could include the use of ambient or group music, but no personalized music therapy. The minimum data set (MDS) (Centers for Medicare & Medicaid Services, 2021) and each resident's Cohen-Mansfield Agitation Inventory (CMAI) (Cohen-Mansfield, Marx and Rosenthal, 1989) were collected at baseline and at up to two follow up time points. The MDS is a federally mandated assessment for all patients residing in Medicare and Medicaid certified nursing homes (Centers for Medicare & Medicaid Services, 2021) and contains comprehensive data on each residents health status and functional capabilities. The trial collected information about the effectiveness of the intervention as well as data on its fidelity. The amount of personalized music played by each resident was recorded in iPod metadata and downloaded using the iTunes interface. Similar to other pragmatic trials, both stages of METRIcAL suffered from a large portion of individuals in the active intervention arm that did not comply with

the intervention (McCreedy et al., 2022).

3.4.2 A Simulated Trial using METRIcAL Data

To illustrate the proposed two-stage design and analysis using real-world data, we simulate a pragmatic trial using the baseline covariates, outcomes, and compliance of individuals who participated in METRIcAL. The outcome of interest is the change in CMAI from baseline measurement to four-months follow up, and we define a complier ($D_{ij} = 1$) as a participant who would use the personalized music intervention for at least 1 minute per day on average if they were assigned to recieve it.

The parameters we use to configure the two-stage design are $\alpha = 0.05$, $1 - \beta = 1 - \beta_1 = 0.8$, $ES = 0.5$, $p_E = 0.6$, $p_L = 0.3$ and $p_U = 0.8$ which yeilds $N = 350$, $N_1 = 170$ and $N_2 = 180$ from Equations (3.3) and (3.4). Using the O'Brien-Flemming alpha spending function (Demets and Lan, 1994), we set $\alpha_1 = 0.005$ and $\alpha_2 = 0.045$. To analyze the simulated two-stage trial, we implement the Bayesian models described in Section 3.2.4. We assume mutually independent inital prior distributions for $\theta_{D_1}, \theta_{D_2}$ and θ_Y with $P_0(\theta_Y) \propto \frac{1}{\sigma^2}$ and $P_0(\theta_{D_i}) = \text{MVN}(\theta_{D_i}|0, 100 * \mathbf{I}_{P+1})$ for $i = 1, 2$ where $\mathbf{I}$ is the identity matrix, and set the value of a_0 in the partial borrowing power prior distribution using the DIC-based method.

The second stage of the METRIcAL trial was subject to low adherence. Therefore, we define the personalized music intervention in the second stage of METRIcAL to be the standard intervention and the personalized music intervention in the first stage of METRIcAL to be the augmented intervention. To obtain a study population, we randomly sample $N_1 = 170$ participants from the second stage of METRIcAL and $N_2 = 180$ participants from the first stage of METRIcAL. The baseline covariates and their values in the sampled study population are summarised in Table A3.6.

Conducting the Simulated Trial

The observed proportion of compilers in the sampled study population for Stage 1, $\hat{\pi}_1$, is 0.4. Because $p_L < \hat{\pi}_1 < p_U$, we conduct a statistical test on ITT_1 with significance level $\alpha_1 = 0.005$. The point estimate and 99.5% credible interval for $\widehat{\text{ITT}}_1$ are -0.39 and $[-4.58, 3.72]$, respectively. Because this result does not reject the null hypothesis, we initiate Stage 2.

The observed proportion of compilers in Stage 2, $\hat{\pi}_2$, is 0.7. This indicates a 75% improvement

in observed compliance for the augmented intervention compared to the standard intervention. We optimize the DIC-based objective function to obtain a value of 0.44 for a_0 in the partial borrowing power prior assumed by the Bayesian models. The point estimate and 94.5% credible interval for $\widehat{\text{ITT}}_2$ when are 2.30 and $[-1.61, 6.18]$, respectively. These estimates do not reject the null hypothesis, and we conclude that there is no observable evidence of an effect of the intervention on the CMAI. A sensitivity analysis that estimates ITT_2 when $a_0 = 0$ ($\widehat{\text{ITT}}_2 = 3.29$; 94.5% CrI: [-1.13,7.34]) and when $a_0 = 1$ ($\widehat{\text{ITT}}_2 = 1.76$; 94.5% CrI: [-1.67,5.07]) does not lead to different interpretations of the results. We summarize the steps of the simulated trial in Table 3.7.

Step	Result	Description
Complete stage 1 and compute $\hat{\pi}_1$	$\hat{\pi}_1 = 0.4$	Because $p_L < \hat{\pi}_1 < p_U$, we conduct a statistical test on the ITT_1 estimand with level $\alpha_1 = 0.005$.
α_1-level statisical test of ITT_1	$\widehat{\text{ITT}}_1 = -0.39$ 99.5% CrI : $[-4.58,\ 3.72]$	The estimate of the ITT_1 does not reach statistical significance at the α_1-level. Therefore, we initiate Stage 2.
Determine value of a_0	$a_0^{DIC} = 0.44$	Using the DIC-based method, we assign the borrowing parameter a value of 0.44.
α_2-level statisical test of ITT_2	$\widehat{\text{ITT}}_1 = -0.39$ 95.5% CrI : $[-4.58,\ 3.72]$	The estimate of the ITT_1 does not reach statistical significance at the α_1-level. Accept the result of no observable effect of the augmented intervention on the change in CMAI.

Table 3.7: Steps of the two-stage design and analysis of the simulated trial estimating the effect of a personalized music intervention on the change in the Cohen-Mansfield Aggitation Index (CMAI).

3.5 Conclusion

We propose a novel, Bayesian, two-stage sequential design for PRCTs with one-sided binary non-compliance. Our design defines a set of decision rules based on prepecified parameters to determine stopping critera, and allows researchers to modify the intervention after Stage 1 to improve compliance for Stage 2. We outline a set of necessary statistical assumptions and define Bayesian models for analyzing the trial that assume power prior distributions to incorporate data from Stage 1 in the analysis of Stage 2 to preserve statistical power.

To evaluate the operating characteristics the two-stage design and Bayesian analysis, we conduct several simulation studies that demonstrate reductions in sample size and improved power compared to traditional design and analysis methods. In addition, we show the Bayesian models are relatively robust to the presence of unobserved covariates and to violations of the assumption that the the potential outcome distributions are the same Stage 1 and Stage 2.

We implement the proposed design and analysis on real data by developing a simulated PRCT using outcomes, compliance and covariate values from METRIcAL. The observed compliance in Stage 1 exceeds p_L but is less than p_U so a statistical test at the α_1-level is conducted on ITT_1 estimated that demonstrates no evidence of an effect. We use the DIC-based criteria to assign a value of 0.44 to the borrowing parameter in the Bayesian power prior and conduct an α_2-level test on the ITT_2 estimand. This test does not reject the null hypothesis and we conclude there is no observable effect of the intervention of the CMAI.

In conclusion, the proposed two-stage design and analysis provides a framework for PRCTs that can improve patient compliance during a trial, reduce sample size when an intervention exceeds the expected levels of compliance and preserves statistical power by implementing the the Bayesian power prior.

Chapter 1 Appendix

A1.1 Multiple Imputation Combination Rules for Small Data Sets

The point estimate of γ is $\hat{\gamma} = \frac{1}{M}\sum_{i=1}^{M}\hat{\gamma}^{(m)}$ and the sampling variance of $\hat{\gamma}^{(m)}$ is given by $\hat{U}^{(m)}$ ($\hat{U}^{(m)} = 0$ for finite-sample estimands). Let $\bar{U} = \frac{1}{M}\sum_{i=1}^{M}\hat{U}^{(m)}$ define the within imputation variance and let $B = \frac{1}{M-1}\sum_{i=1}^{M}(\hat{\gamma}^{(m)} - \hat{\gamma})^2$ define the between imputation variance. The total estimate of the sampling variance for $\hat{\gamma}$ is $T = \bar{U} + (1 + \frac{1}{m})B$. To obtain interval estimates in small data sets, Barnard and Rubin Barnard and Rubin (1999) recommend approximating the distribution of $(\gamma - \hat{\gamma})T^{-1/2}$ using a t-distribution with $\tilde{\nu}_M$ degrees of freedom where

$$\tilde{\nu}_M = \left(\frac{1}{\nu_M} + \frac{1}{\nu_{\widehat{obs}}}\right)^{-1}.$$

The values of ν_M and $\nu_{\widehat{obs}}$ are given by

$$\nu_M = (M-1)\left(\frac{T}{(1+M^{-1})\,B}\right)^2$$

and

$$\nu_{\widehat{obs}} = \left(\frac{\nu_{\text{com}}+1}{\nu_{\text{com}}+3}\right)\nu_{\text{com}}\left(1 - \frac{(1+M^{-1})B}{T}\right)$$

where ν_{com} is the complete-data degrees of freedom.

A1.2 Simulation Studies

Table A1.1: Simulation Parameters for the first case study

Parameter	$w = 1$	$w = 0$
φ_w^a	1	1
φ_w^d	0.5	0.75
ζ_w	1.5	0.5
ξ_w^a	$\{0.21, 0.32, 0.21, -2.3, -1.4\}^t$	$\{0.21, 0.07, 0.21, -2.61, -2.0\}^t$
ξ_w^d	$\{-0.03, 0.18, -0.13, -1.08, -0.19\}^t$	$\{0.20, 0.22, 0.08, -1.08, -0.55\}^t$

Table A1.2: Simulation Parameters for the second case study

Parameter	$w = 1$	$w = 0$
φ_w^a	-0.25	0.15
φ_w^d	-0.25	-0.75
ζ_w	2	1
ξ_w^a	$\{0.21, 0.32, 0.21, -2.3, -1.4\}^t$	$\{0.21, 0.07, 0.21, -2.61, -2.0\}^t$
ξ_w^d	$\{-0.03, 0.18, -0.13, -1.08, -0.19\}^t$	$\{0.20, 0.22, 0.08, -1.08, -0.55\}^t$

Table A1.3: Simulation results for the second case study.

Estimand	Bayesian Method				DR			
	Coverage	Bias	I.W.	RMSE	Coverage	Bias	I.W.	RMSE
ITT Adverse	94	0.001	0.071	0.019	94	0.000	0.072	0.019
ITT Death	95	-0.002	0.086	0.023	94	-0.001	0.088	0.023
ITT Composite	94	-0.001	0.086	0.023	94	-0.001	0.090	0.024
SACE	96	0.000	0.066	0.017	-	-	-	-
$Pr(G_i(1) = 1) - Pr(G_i(0) = 1)$	94	0.001	0.086	0.023	95	0.001	0.090	0.024
$Pr(G_i(1) = 2) - Pr(G_i(0) = 2)$	95	0.000	0.049	0.012	94	0.000	0.051	0.013
$Pr(G_i(1) = 3) - Pr(G_i(0) = 3)$	94	-0.003	0.074	0.020	94	-0.001	0.078	0.021
$Pr(G_i(1) = 4) - Pr(G_i(0) = 4)$	94	0.001	0.059	0.016	94	0.000	0.063	0.017
$\kappa_{10} - \kappa_{01}$	95	-0.001	0.093	0.025	-	-	-	-
$\frac{\kappa_{10}}{\kappa_{01}}$	95	-0.001	0.534	0.141	-	-	-	-

* Coverage of 95% credible interval

⋆ Coverage of 95% confidence interval

† Interval Width

Table A1.4: Simulation results for the second case study under Burr link function with $c = 0.5$.

Estimand	Bayesian Method				DR			
	Coverage	Bias	I.W.	RMSE	Coverage	Bias	I.W.	RMSE
ITT Adverse	95	0.004	0.066	0.017	95	0.001	0.064	0.016
ITT Death	94	0.005	0.080	0.021	94	0.000	0.081	0.022
ITT Composite	94	0.007	0.086	0.023	95	-0.001	0.087	0.023
SACE	95	0.003	0.059	0.015	-	-	-	-
$Pr(G_i(1) = 1) - Pr(G_i(0) = 1)$	94	-0.007	0.086	0.023	95	0.001	0.087	0.023
$Pr(G_i(1) = 2) - Pr(G_i(0) = 2)$	95	0.002	0.049	0.013	94	0.000	0.049	0.013
$Pr(G_i(1) = 3) - Pr(G_i(0) = 3)$	94	0.003	0.070	0.018	94	0.000	0.071	0.019
$Pr(G_i(1) = 4) - Pr(G_i(0) = 4)$	94	0.002	0.046	0.012	94	0.000	0.047	0.012
$\kappa_{10} - \kappa_{01}$	94	0.008	0.091	0.025	-	-	-	-
$\frac{\kappa_{10}}{\kappa_{01}}$	94	0.062	0.647	0.177	-	-	-	-

A1.3 Data Analysis Model Diagnostics

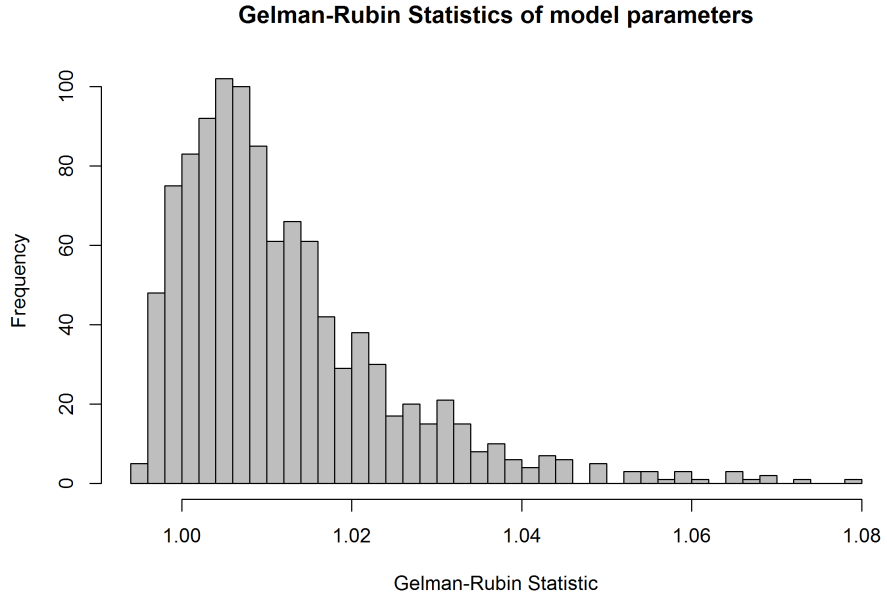

Figure A1.1: Gelman-Rubin Statistics for model parameters used in the data analysis found in Section 1.5

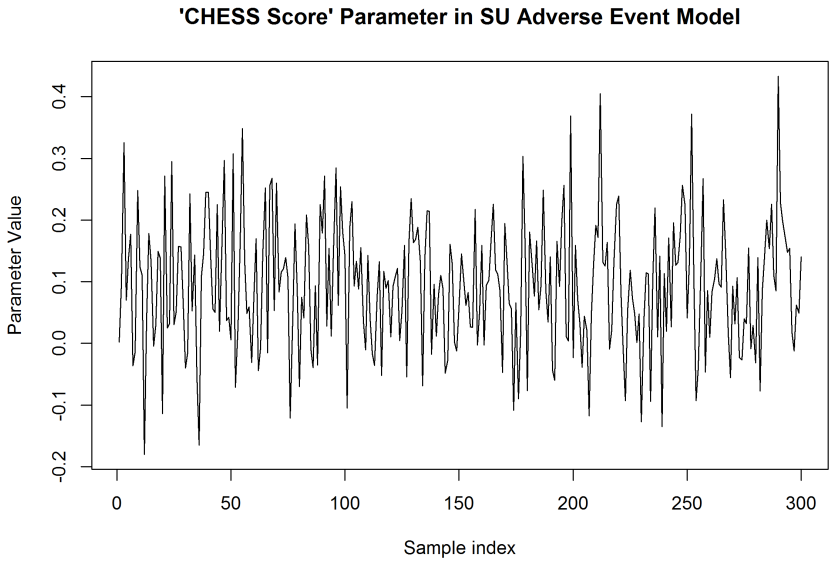

Figure A1.2: Trace plot for the parameter related to Changes in health, End-stage disease and Signs and Symptoms (CHESS) score (Hirdes, Frijters and Teare, 2003), a health stability measure, in the logistic regression for heart failure under SU found in Section 1.5

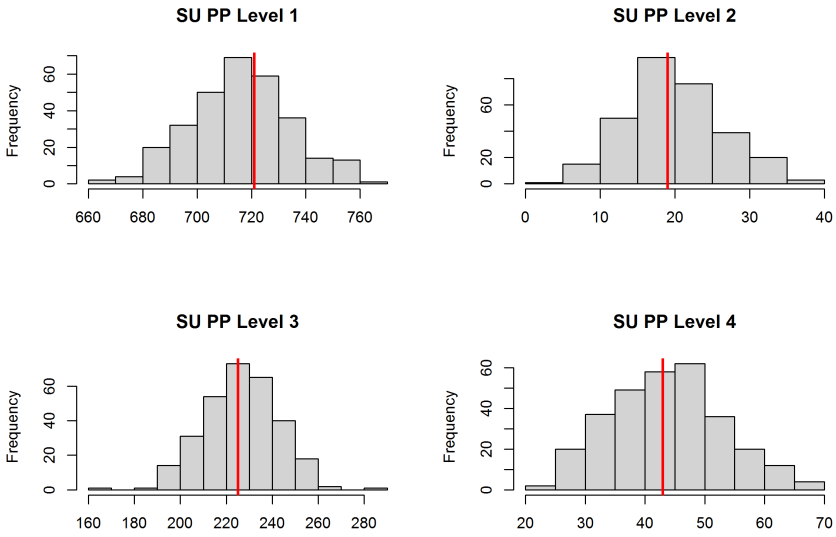

Figure A1.3: Posterior predictive distributions for composite ordinal outcome levels under SU using the Bayesian models fit in Section 1.5. The vertical red line represents observed number of patients in that level of the composite ordinal outcome.

Chapter 2 Appendix

i	j	W_i	$Y_{ij}(1, D_{ij})$	$Y_{ji}(0, D_{ij})$	$D_{ij}(1)$	$D_{ij}(0)$	$\mathbf{C}_i(1)$	$\mathbf{C}_i(0)$	S_i	$\mathbf{Z}_i$	$\mathbf{X}_{ij}$
1	1	1	*	?	*	0	*	0	?	*	*
1	2	1	*	?	*	0	*	0	?	*	*
1	3	1	*	?	*	0	*	0	?	*	*
2	1	0	?	*	?	0	?	0	?	*	*
2	2	0	?	*	?	0	?	0	?	*	*
2	3	0	?	*	?	0	?	0	?	*	*
2	4	0	?	*	?	0	?	0	?	*	*

Table A2.1: Observable and unobservable quantities in PCRCT design. "*" indicates that the quantity can be observed, "?" indicates that quantity is not observed

A2.1 Gibbs Sampler Derivation

We describe a Gibbs sampling algorithm to sample from the joint posterior distribution of the parameters in the parametric model defined in Section 2.3.3. The set of all model parameters is given by

$$\boldsymbol{\theta} = \{\boldsymbol{\mu}_1^S, \ldots, \boldsymbol{\mu}_K^S, \pi_1, \ldots, \pi_K, \Sigma, \mu_1^D, \ldots, \mu_K^D, \alpha_1^D, ..., \alpha_K^D, \tau_D^2, \phi^D,$$
$$\mu_1^Y, \ldots, \mu_K^Y, \beta_{1,1}, \beta_{0,1}, \ldots, \beta_{1,K}, \beta_{0,K}, \delta_{1,1}, \delta_{0,1}, \ldots, \delta_{1,K}, \delta_{0,K}, \sigma_1^2, \ldots, \sigma_K^2, \tau_Y^2, \phi^Y\}.$$

Under the assumptions defined in Section 2.3.3, we have,

$$P(\boldsymbol{\theta}|\mathbf{Y}^{obs}, \mathbf{D}^{obs}, \mathbf{C}^{obs}, \mathbf{Z}, \mathbf{X}) = \int\int\int P(\boldsymbol{\theta}, \mathbf{D}^{mis}, \mathbf{C}^{mis}, \mathbf{S}|\mathbf{Y}^{obs}, \mathbf{D}^{obs}, \mathbf{C}^{obs}, \mathbf{Z}, \mathbf{X}) d\mathbf{D}^{mis} d\mathbf{C}^{mis} d\mathbf{S} \tag{3.25}$$

where $\mathbf{S}$ is thee latent compliance strata allocation vector, and $\mathbf{D}^{mis}$ and $\mathbf{C}^{mis}$ are the unobserved compliance values for individuals and clusters assigned to the control. The term inside the integral in Equation (3.25) is the posterior distribution for $\boldsymbol{\theta}$ augmented with $\mathbf{S}$, $\mathbf{C}^{mis}$ and $\mathbf{D}^{mis}$. This

distribution is given by

$$
\begin{aligned}
&P(\boldsymbol{\theta}, \mathbf{D}^{mis}, \mathbf{C}^{mis}, \mathbf{S} | \mathbf{Y}^{obs}, \mathbf{D}^{obs}, \mathbf{C}^{obs}, \mathbf{Z}, \mathbf{X}) \\
&\propto P(\boldsymbol{\theta}) P(\mathbf{Y}^{obs} | \mathbf{D}^{obs}, \mathbf{X}, \mathbf{S}, \boldsymbol{\theta}) P(\mathbf{D}^{mis}, \mathbf{D}^{obs} | \mathbf{X}, \mathbf{S}, \boldsymbol{\theta}) P(\mathbf{C}^{mis}, \mathbf{C}^{obs}, \mathbf{Z} | \mathbf{S}, \boldsymbol{\theta}) P(\mathbf{S} | \boldsymbol{\theta}) \\
&\propto P(\boldsymbol{\theta}) \prod_{k=1}^{K} \prod_{i:W_i=1} \prod_{j=1}^{n_i} N(Y_{ij}(1, D_{ij}) | \mu_k^Y + \mathbf{X}_{ij}\beta_{0,k} + D_{ij} * (\delta_{1,k} + \mathbf{X}_{ij}\beta_{1,k}) + \phi_i^Y, \sigma_k^2)]^{I(S_i=k)} \times \\
&\quad \prod_{i:W_i=0} \prod_{j=1}^{n_i} N(Y_{ij}(0, D_{ij}) | \mu_k^Y + \mathbf{X}_{ij}\beta_{0,k} + D_{ij} * \delta_{0,k} + \phi_i^Y, \sigma_k^2)]^{I(S_i=k)} \times \\
&\quad \prod_{k=1}^{K} \prod_{j=1}^{n_i} [\text{probit}^{-1}(\mu_k^D + \mathbf{X}_{ij}\alpha_k + \phi_i^D)^{D_{ij}} (1 - \text{probit}^{-1}(\mu_k^D + \mathbf{X}_{ij}\alpha_k + \phi_i^D))^{1-D_{ij}}]^{I(S_i=k)} \times \\
&\quad \prod_{k=1}^{K} \prod_{i:W_i=1} \text{MVN}(\mathbf{C}_i, \mathbf{Z}_i | \boldsymbol{\mu}_k^S, \Sigma)^{I(S_i=k)} \prod_{k=1}^{K} \prod_{i:W_i=0} \text{MVN}(\mathbf{C}_i, \mathbf{Z}_i | \boldsymbol{\mu}_k^S, \Sigma)^{I(S_i=k)} \prod_{i=1}^{I^{fp}} \prod_{k=1}^{K} \pi_k^{I(S_i=k)}.
\end{aligned}
$$

We sample $\boldsymbol{\theta}$ from the augmented posterior because it provides a more tractable form for implementing the Gibbs sampling algorithm. We assume the prior distributions for the model parameters are

$$
\begin{aligned}
\pi &\sim \text{Dirichilet}(K, \lambda) \\
\boldsymbol{\mu}_k^S &\sim \text{MVN}\left(m_{\boldsymbol{\mu}^S}, V_{\boldsymbol{\mu}^S}\right) \text{ for } k = 1, .., K \\
\Sigma &\sim \text{Inv-Wishart}(\Psi, \nu_\Sigma) \\
\mu_k^D &\sim N\left(0, v_\alpha^2\right) \quad \alpha_k \sim \text{MVN}\left(0, v_\alpha^2 * \mathbf{I}\right) \text{ for } k = 1, .., K \\
\phi_i^D &\sim N(0, \tau_D^2) \quad \tau_D \sim \text{Uniform}\left(0, \sqrt{U_{\tau_D^2}}\right) \text{ for } k = 1, .., K \\
\mu_k^Y &\sim N(0, v_\beta^2), \ \delta_{1.k} \sim N(0, v_\beta^2) \text{ for } k = 1, ..., K \\
\beta_{0,k} &\sim \text{MVN}(0, v_\beta^2 * \mathbf{I}), \ \beta_{1,k} \sim \text{MVN}(0, v_\beta^2 * \mathbf{I}) \text{ for } k = 1, .., K \\
\phi_i^Y &\sim N(0, \tau_{Y_k}^2) \\
\tau_Y &\sim \text{Uniform}\left(0, \sqrt{U_{\tau_Y^2}}\right), \ \sigma_k^2 \sim \text{Inv-Gamma}(a, b) \text{ for } k = 1, .., K.
\end{aligned}
\tag{3.26}
$$

We let $\boldsymbol{\theta}_{-q}^{(m)}$ denote the set of parameter values that are conditioned on when sampling parameter $q^{(m)}$. The vector $\boldsymbol{\theta}_{-q}^{(m)}$ contains the most recent sampled value for each parameter in $\boldsymbol{\theta}$, except for

$q^{(m-1)}$. In addition, we let $\mathbf{C}^{(m)} = \{\mathbf{C}_i^{(m)}\}$ and $\mathbf{D}^{(m)} = \{D_{ij}^{(m)}\}$ where

$$\mathbf{C}_i^{(m)} = \begin{cases} \mathbf{C}_i & \text{if } W_i = 1 \\ \mathbf{C}_i^{mis(m)} & \text{if } W_i = 0 \end{cases} \text{ and } D_{ij}^{(m)} = \begin{cases} D_{ij} & \text{if } W_i = 1 \\ D_{ij}^{mis(m)} & \text{if } W_i = 0 \end{cases}.$$

Finally, we let $\mathbf{O} = \{\mathbf{Y}^{obs}, \mathbf{D}^{obs}, \mathbf{C}^{obs}, \mathbf{Z}, \mathbf{X}\}$ denote the set of observed values.

To sample $\boldsymbol{\theta}$ from its posterior distribution, begin by initializing $\boldsymbol{\theta}^{(0)}$, $\mathbf{S}^{(0)}$,$\mathbf{C}^{mis(0)}$, and $\mathbf{D}^{mis(0)}$. For $m = 1, ..., M$, we repeat the following procedure:

1. Sample from $P(\boldsymbol{\theta}^{(m)}|\mathbf{O}, \mathbf{S}^{(m-1)}, \mathbf{C}^{mis(m-1)}, \mathbf{D}^{mis(m-1)})$.

   (a) Sample from $P(\boldsymbol{\mu}_k^{S(m)}|\mathbf{O}, \mathbf{S}^{(m-1)}, \mathbf{C}^{mis(m-1)}, \mathbf{D}^{mis(m-1)}, \boldsymbol{\theta}^{(m)}_{-\boldsymbol{\mu}_k^S})$ for $k = 1, ..., K$. We have,

$$P(\boldsymbol{\mu}_k^{S(m)}|\mathbf{O}, \mathbf{S}^{(m-1)}, \mathbf{C}^{mis(m-1)}, \mathbf{D}^{mis(m-1)}, \boldsymbol{\theta}^{(m)}_{-\boldsymbol{\mu}_k^S}) \propto$$
$$\text{MVN}(\boldsymbol{\mu}_k^S|m_{\boldsymbol{\mu}^S}, V_{\boldsymbol{\mu}^S}) \prod_{i:S_i^{(m-1)}=k} \text{MVN}(\mathbf{C}_i^{(m-1)}, \mathbf{Z}_i|\boldsymbol{\mu}_k^S, \Sigma^{(m-1)})$$

   The prior specification for is $\boldsymbol{\mu}_k^{S(m)}$ is semi-conjugate, and implies

$$P(\boldsymbol{\mu}_k^{S(m)}|\mathbf{O}, \mathbf{S}^{(m-1)}, \mathbf{C}^{mis(m-1)}, \mathbf{D}^{mis(m-1)}, \boldsymbol{\theta}^{(m)}_{-\boldsymbol{\mu}_k^S}) = \text{MVN}(\boldsymbol{\mu}_k^S|M_{\boldsymbol{\mu}_k^S}, \Omega_{\boldsymbol{\mu}_k^S})$$
$$\text{where } \Omega_{\boldsymbol{\mu}_k^S} = \left(V_\mu^{-1} + \left(\sum_{i=1}^{I^{fp}} I(S_i^{(m-1)} = k)\right) * \Sigma^{-1}\right), ^{-1}$$
$$\text{and } M_{\boldsymbol{\mu}_k^S} = \Omega_{\boldsymbol{\mu}_k^S}\left(V_{\boldsymbol{\mu}^S}^{-1} m_{\boldsymbol{\mu}^S} + \left(\sum_{i=1}^{I^{fp}} I(S_i^{(m-1)} = k)\right) * \Sigma^{-1}\left[\bar{\mathbf{C}}^{(m-1)}, \bar{\mathbf{Z}}\right]^t\right).$$

   The term $\bar{\mathbf{C}}_k^{(m-1)}$ is the mean compliance vector for clusters such that $S_i^{(m-1)} = k$. Note, $\bar{\mathbf{C}}_k^{(m-1)}$ includes the sampled $\mathbf{C}^{mis(m-1)}$ values for clusters assigned to the control, and $\mathbf{C}^{obs}$ values for clusters assigned to the treatment.

   (b) Sample from $P(\Sigma^{(m)}|\mathbf{O}, \mathbf{S}^{(m-1)}, \mathbf{C}^{mis(m-1)}, \mathbf{D}^{mis(m-1)}, \boldsymbol{\theta}^{(m)}_{-\Sigma})$.

We have,

$$P(\Sigma^{(m)}|\mathbf{O}, \mathbf{S}^{(m-1)}, \mathbf{C}^{mis(m-1)}, \mathbf{D}^{mis(m-1)}, \boldsymbol{\theta}^{(m)}_{-\Sigma}) \propto \text{Inv-Wishart}(\Sigma|\Psi, \nu_\Sigma) \times \prod_{i:S_i^{(m-1)}=k} \text{MVN}(\mathbf{C}_i^{(m-1)}, \mathbf{Z}_i|\boldsymbol{\mu}_k^{S(m)}, \Sigma).$$

The conjugacy of the Inverse Wishart prior distribution implies

$$P(\Sigma^{(m)}|\mathbf{O}, \mathbf{S}^{(m-1)}, \mathbf{C}^{mis(m-1)}, \mathbf{D}^{mis(m-1)}, \boldsymbol{\theta}^{(m)}_{-\Sigma}) = \text{Inv-Wishart}(\Sigma \mid \Omega_\Sigma, (I + \nu_\Sigma))$$

where,

$$\Omega_\Sigma = \sum_{k=1}^{K} \sum_{i=1}^{I^{fp}} I(S_i^{(m-1)} = k) * \left(\left[\mathbf{C}_i^{(m-1)}, \mathbf{Z}_i \right]^t - \boldsymbol{\mu}_k^{S(m)} \right) \left(\left[\mathbf{C}_i^{(m-1)}, \mathbf{Z}_i \right]^t - \boldsymbol{\mu}_k^{S(m)} \right)^t.$$

(c) Sample from $P(\pi^{(m)}|\mathbf{O}, \mathbf{S}^{(m-1)}, \mathbf{C}^{mis(m-1)}, \mathbf{D}^{mis(m-1)}, \boldsymbol{\theta}^{(m)}_{-\pi})$.

The complete conditional distribution for $\pi^{(m)}$ is,

$$P(\pi^{(m)}|\mathbf{O}, \mathbf{S}^{(m-1)}, \mathbf{C}^{mis(m-1)}, \mathbf{D}^{mis(m-1)}, \boldsymbol{\theta}^{(m)}_{-\pi}) \propto \text{Dirichilet}(\pi|K, \lambda = (\lambda_1, ..., \lambda_K)) \times \prod_{i=1}^{I^{fp}} \prod_{k=1}^{K} \pi_k^{I(S_i^{(m-1)}=k)}.$$

By the cojugacy of the Dirichilet prior distribution,

$$P(\pi^{(m)}|\mathbf{O}, \mathbf{S}^{(m-1)}, \mathbf{C}^{mis(m-1)}, \mathbf{D}^{mis(m-1)}, \boldsymbol{\theta}^{(m)}_{-\pi}) = \text{Dirichilet}(\pi|K, \lambda^*)$$

where $\lambda_k^* = \lambda_k + \sum_i I(S_i^{(m-1)} = k)$.

(d) Sample from $P(\mu_k^{D(m)}, \alpha_k^{(m)}|\mathbf{O}, \mathbf{S}^{(m-1)}, \mathbf{C}^{mis(m-1)}, \mathbf{D}^{mis(m-1)}, \boldsymbol{\theta}^{(m)}_{-\mu_k^D, \alpha_k})$ for $k = 1, ..., K$.

The complete conditional distribution is

$$P(\mu_k^{D(m)}, \alpha_k^{(m)} | \mathbf{O}, \mathbf{S}^{(m-1)}, \mathbf{C}^{mis(m-1)}, \mathbf{D}^{mis(m-1)}, \boldsymbol{\theta}^{(m)}_{-\mu_k^D, \alpha_k}) \propto \text{MVN}(\mu_k^{D(m)}, \alpha_k^{(m)} | 0, v_\alpha^2 \mathbf{I}^{\mathbf{fp}}) \times$$
$$\prod_{i:S_i^{(m-1)}=k} \prod_{j=1}^{n_i} \left(p_{ijk}^{D(m-1)}\right)^{D_{ij}^{(m-1)}} \left(1 - p_{ijk}^{D(m-1)}\right)^{1-D_{ij}^{(m-1)}},$$

where $p_{ijk}^{D(m-1)} = \text{probit}^{-1}(\mu_k^D + \mathbf{X}_{ij}\alpha_k + \phi_i^{D(m-1)})$. The complete conditional distribution corresponds to the posterior distribution of the model coefficients in a Bayesian probit regression model with multivariate a normal prior distribution. To sample from this distribution, we implement a data augmentation step that is commonly applied in Bayesian binary probit regression Albert and Chib (1993); Chakraborty and Khare (2017). We introduce the set of latent variables, $\mathbf{U} = \{U_{ij}\}$. For individuals in clusters such that $S_i^{(m-1)} = k$, we sample $U_{ij}^{(m)}$ from

$$U_{ij}^{(m)} | \mathbf{O}, \mathbf{S}^{(m-1)}, \mathbf{C}^{mis(m-1)}, \mathbf{D}^{mis(m-1)}, \boldsymbol{\theta}^{(m)}_{-U} \sim \begin{cases} \text{Truncated-Normal}_{(0,\infty)}(M_{U_{ij}}, 1) & \text{if } D_{ij} = 1 \\ \text{Truncated-Normal}_{(-\infty,0)}(M_{U_{ij}}, 1) & \text{if } D_{ij} = 0 \end{cases}$$

where $M_{U_{ij}} = \mu_k^{D(m-1)} + \mathbf{X}_{ij}\alpha_k^{(m-1)} + \phi_i^{D(m-1)}$. Repeating this for clusters such that $S_i^{(m-1)} = 1, ..., K$, we obtain $\mathbf{U}^{(m)}$. Because of the conjugacy of the multivariate normal prior distribution,

$$\left[\mu_k^{D(m)}, \alpha_k^{(m)}\right]^t | \mathbf{O}, \mathbf{S}^{(m-1)}, \mathbf{C}^{mis(m-1)}, \mathbf{D}^{mis(m-1)}, \boldsymbol{\theta}^{(m)}_{-\alpha_k} \sim \text{MVN}(a_k, \Omega_{\alpha_k})$$
$$\text{where} \quad \Omega_{\alpha_k} = \left(\mathbf{X}_k^{*t}\mathbf{X}_k^* + \frac{1}{v_\alpha^2} * \mathbf{I}^{\mathbf{fp}}\right)^{-1},$$
$$a_k = \Omega_{\alpha_k}\left(\mathbf{X}_k^{*t}\left(\mathbf{U}_k^{(m)} - \bar{\bar{\phi}}_k^{D(m-1)}\right)\right).$$

We define the matrix $\mathbf{X}_k^* = [\mathbf{1}_{N_k}\ \mathbf{X}_k]$, where $\mathbf{1}_{N_k}$ is a vector of 1's with length $\sum_{i=1}^{I^{fp}} n_i * I(S_i^{(m-1)} = k)$ and $\mathbf{X}_k$ denotes the covariate matrix for individuals in clusters such that $S_i^{(m-1)} = k$. Similarly, $\mathbf{U}_k^{(m-1)}$ is the set of latent variables for individuals in clusters such that $S_i^{(m-1)} = k$. The parameter $\bar{\bar{\phi}}_k^{D(m-1)}$ vertically stacks the vectors $\phi_i^{D(m-1)} * \mathbf{1}_{n_i}$ for all i such that $S_i^{(m-1)} = k$, where $\mathbf{1}_{n_i}$ is a column vector of 1's with length n_i.

(e) Sample from $P(\tau_D^{2(m)}|\mathbf{O}, \mathbf{S}^{(m-1)}, \mathbf{C}^{mis(m-1)}, \mathbf{D}^{mis(m-1)}\boldsymbol{\theta}^{(m)}_{-\tau_D^2})$ for $k = 1, ..., K$.

We have

$$P(\tau_D^{2(m)}|\mathbf{O}, \mathbf{S}^{(m-1)}, \mathbf{C}^{mis(m-1)}, \mathbf{D}^{mis(m-1)}\boldsymbol{\theta}^{(m)}_{-\tau_D^2}) \propto P(\tau_D^2)\prod_{i=1}^{I^{fp}}(\tau_D^2)^{-1/2}e^{\frac{-\frac{1}{2}(\phi_i^{D(m-1)})^2}{\tau_D^2}}$$
$$= (\tau_D^2)^{-\frac{I^{fp}}{2}}e^{\frac{-\frac{1}{2}\sum_i(\phi_i^{D(m-1)})^2}{\tau_D^2}}P(\tau_D^2)$$

Because we assume $P(\tau_{D,k}) = \text{Uniform}(0, \sqrt{U_{\tau_D^2}})$, the prior distribution for τ_D^2 is

$$P(\tau_D^2) \propto (\tau_D^2)^{-1/2}I(0 < \tau_D^2 < U_{\tau_D^2}).$$

Therefore,

$$\begin{aligned}
&P(\tau_D^{2(m)}|\mathbf{O}, \mathbf{S}^{(m-1)}, \mathbf{C}^{mis(m-1)}, \mathbf{D}^{mis(m-1)}\boldsymbol{\theta}^{(m)}_{-\tau_D^2}) \\
&\propto (\tau_D^2)^{-\left(\frac{I-1}{2}\right)-1}e^{\frac{-\frac{1}{2}\sum_i(\phi_i^{D(m-1)})^2}{\tau_D^2}}I(0 < \tau_D^2 < U_{\tau_D^2}) \\
&=\text{Truncated-Inv-Gamma}_{\left(0,U_{\tau_D^2}\right)}\left(\tau_D^{2(m)}|R^{\text{shape}}_{\tau_D^2}, R^{\text{rate}}_{\tau_D^2}\right)
\end{aligned}$$

where $R^{\text{rate}}_{\tau_D^2} = \frac{1}{2}\sum_i\left(\phi_i^{D(m-1)}\right)^2$ and $R^{\text{shape}}_{\tau_D^2} = \left(\frac{I-1}{2}\right)$.

(f) Sample from $P(\phi^{D(m)}|\mathbf{O}, \mathbf{S}^{(m-1)}, \mathbf{C}^{mis(m-1)}, \mathbf{D}^{mis(m-1)}\boldsymbol{\theta}^{(m)}_{-\phi_i^D})$.

The vector $\phi^{D(m)}$ contains the cluster-level effects for $i = 1, ..., I$. Without loss of generality, we assume $S_i^{(m-1)} = k$. The complete conditional of $\phi_i^{D(m)}$ is

$$\begin{aligned}
&P(\phi_i^{D(m)}|\mathbf{O}, \mathbf{S}^{(m-1)}, \mathbf{C}^{mis(m-1)}, \mathbf{D}^{mis(m-1)}\boldsymbol{\theta}^{(m)}_{-\phi_i^D}) \\
&\propto e^{\frac{-(\phi_i^D)^2}{2\tau_D^{2(m)}}}\prod_{j=1}^{n_i}e^{-\frac{1}{2}(U_{ij}^{(m)}-\mu_k^{D(m)}-\mathbf{X}_{ij}\alpha_k^{(m)}-\phi_i^D)^2} \\
&= \exp\left\{-\frac{(\phi_i^D)^2 + \tau_D^{2(m)}\sum_{j=1}^{n_i}(\phi_i^D - \hat{U}_{ij})^2}{2\tau_D^{2(m)}}\right\},
\end{aligned}$$

where $\hat{U}_{ij} = U_{ij}^{(m)} - \mu_k^{D(m)} - \mathbf{X}_{ij}\alpha_k^{(m)}$. After rearranging terms,

$$
\begin{aligned}
&P(\phi_i^{D(m)}|\mathbf{O},\mathbf{S}^{(m-1)},\mathbf{C}^{mis(m-1)},\mathbf{D}^{mis(m-1)}\boldsymbol{\theta}^{(m)}_{-\phi_i^D}) \\
&\propto \exp\left\{-\frac{\left(\phi_i^D-\left(n_i+\left(\tau_D^{2(m)}\right)^{-1}\right)^{-1}\sum_{j=1}^{n_i}\hat{U}_{ij}\right)^2}{2\left(n_i+\left(\tau_D^{2(m)}\right)^{-1}\right)^{-1}}\right\} \\
&\propto N(\hat{\phi}_i^D,\hat{v}^2_{\phi_i^D}).
\end{aligned}
$$

where $\hat{v}^2_{\phi_i^D}=\left(n_i+\left(\tau_D^{2(m)}\right)^{-1}\right)^{-1}$ and $\hat{\phi}_i^D=\hat{v}^2_{\phi_i^D}\sum_{j=1}^{n_i}\hat{U}_{ij}$. Sample $\phi_i^{D(m)}$ for $i=1,...,I^{fp}$ to obtain $\phi^{D(m)}$.

(g) Sample from $P(B_k^{(m)}|\mathbf{O},\mathbf{S}^{(m-1)},\mathbf{C}^{mis(m-1)},\mathbf{D}^{mis(m-1)},\boldsymbol{\theta}^{(m)}_{-B})$ where
$B_k^{(m)}=\left[\mu_k^{Y(m)},\beta_{0,k}^{(m)},\delta_{0,k}^{(m)},\beta_{1,k}^{(m)},\delta_{1,k}^{(m)}\right]^t$ for $k=1,...,K$.
The complete conditional distribution of $B_k^{(m)}$ is

$$
\begin{aligned}
&P(B_k^{(m)}|\mathbf{O},\mathbf{S}^{(m-1)},\mathbf{C}^{mis(m-1)},\mathbf{D}^{mis(m-1)},\boldsymbol{\theta}^{(m)}_{-B}) \\
&\propto \prod_{i:\substack{S_i^{(m-1)}=k\\W_i=0}}\prod_{j=1}^{n_i}N(Y_{ij}(1,D_{ij})|\mu_k^Y+\mathbf{X}_{ij}\beta_{0,k}+D_{ij}^{(m-1)}*(\delta_{1,k}+\mathbf{X}_{ij}\beta_{1,k})+\phi_i^{Y(m-1)},\sigma_k^{2(m-1)})\times \\
&\quad\prod_{i:\substack{S_i^{(m-1)}=k\\W_i=0}}\prod_{j=1}^{n_i}N(Y_{ij}(0,D_{ij}^{(m-1)})|\mu_k^Y+\mathbf{X}_{ij}\beta_{0,k}+D_{ij}^{(m-1)}\delta_{0,k}+\phi_i^{Y(m-1)}\sigma_k^{2(m-1)})P(B_k) \\
&=P(B_k)\prod_{iS_i^{(m-1)}=k}\prod_{j=1}^{n_i}N\left(Y_{ij}^{obs}|\left\{\mu_k^Y+\mathbf{X}_{ij}\beta_{0,k}+(1-W_i)*D_{ij}^{(m-1)}*\delta_{0,k}+\right.\right. \\
&\qquad\qquad\left.\left.W_i*D_{ij}^{(m-1)}*(\mathbf{X}_{ij}\beta_{1,k}+\delta_{1,k})+\phi_i^{Y(m-1)}\right\},\sigma_k^{2(m-1)}\right).
\end{aligned}
\tag{3.27}
$$

We let $\mathbf{X}^{obs}$ be the design matrix for the regression described by Equation (3.27) where

$$
\mathbf{X}_{ij}^{obs}=\left[1,\ \mathbf{X}_{ij},\ (1-W_{ij})*D_{ij}^{(m-1)},\ W_i*D_{ij}^{(m-1)}*\mathbf{X}_{ij},\ W_i*D_{ij}^{(m-1)}\right].
$$

Equation (3.27) simplifies to

$$P(B_k^{(m)}|\mathbf{O}, \mathbf{S}^{(m-1)}, \mathbf{C}^{mis(m-1)}, \mathbf{D}^{mis(m-1)}, \boldsymbol{\theta}_{-B}^{(m)})$$
$$\propto P(B_k) \prod_{i:S_i^{(m-1)}=k} \prod_{j=1}^{n_i} N(Y_{ij}^{obs}|\mathbf{X}_{ij}^{obs} B_k + \phi_i^{Y(m-1)}, \sigma_k^{2(m-1)}).$$

The joint prior distribution for B_k given the prior distributions described in (3.26) is

$$P(B_k) = \text{MVN}\left(\begin{bmatrix} \mu_k^Y \\ \beta_{0,k} \\ \delta_{0,k} \\ \beta_{1,k} \\ \delta_{1,k} \end{bmatrix} \middle| \, m_B = \mathbf{0}, V_B = \begin{bmatrix} v_{\mu^Y}^2 & 0 & \dots & \dots & 0 \\ 0 & v_\beta^2 & \vdots & \vdots & \vdots \\ \vdots & 0 & \ddots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \ddots & 0 \\ 0 & \dots & \dots & 0 & v_\beta^2 \end{bmatrix} \right).$$

By the conjugacy of the prior distribution,

$$P(B_k^{(m)}|\mathbf{O}, \mathbf{S}^{(m-1)}, \mathbf{C}^{mis(m-1)}, \mathbf{D}^{mis(m-1)}, \boldsymbol{\theta}_{-B}^{(m)}) \propto \text{MVN}\left(M_B, \Omega_B\right),$$

where $\Omega_B = \left(\frac{1}{\sigma_k^{2(m-1)}} \left(\mathbf{X}_k^{obs}\right)^t \mathbf{X}_k^{obs} + V_B^{-1} \right)^{-1}$ and $M_B = \Omega_B \left(\frac{1}{\sigma_k^{2(m-1)}} \left(\mathbf{X}_k^{obs}\right)^t \left(Y_{ij}^{obs} - \bar{\phi}_k^{Y(m-1)} \right) \right)$.

(h) Sample from $P(\sigma_k^{2(m)}|\mathbf{O}, \mathbf{S}^{(m-1)}, \mathbf{C}^{mis(m-1)}, \mathbf{D}^{mis(m-1)}\boldsymbol{\theta}_{-\sigma^2}^{(m)})$ for $k = 1, ..., K$.

The complete conditional distribution of $\sigma_k^{2(m)}$ is

$$P(\sigma_k^{2(m)}|\mathbf{O}, \mathbf{S}^{(m-1)}, \mathbf{C}^{mis(m-1)}, \mathbf{D}^{mis(m-1)}\boldsymbol{\theta}_{-\sigma^2}^{(m)})$$
$$\propto P(\sigma_k^2) \prod_{i:S_i^{(m-1)}=k} \prod_{j=1}^{n_i} \left(\sigma_k^2\right)^{-\frac{1}{2}} e^{-\frac{1}{2} \frac{(Y_{ij}^{obs} - \mathbf{X}_{ij}^{obs} B_k^{(m)} - \phi_i^{Y(m-1)})^2}{\sigma_k^2}}.$$

Because $P(\sigma_k^2) = \text{Inv-Gamma}\left(a, b\right)$ is conjugate,

$$P(\sigma_k^{2(m)}|\mathbf{O}, \mathbf{S}^{(m-1)}, \mathbf{C}^{mis(m-1)}, \mathbf{D}^{mis(m-1)}\boldsymbol{\theta}_{-\sigma^2}^{(m)}) \propto \text{Inv-Gamma}\left(R_{\sigma_k^2}^{\text{shape}}, R_{\sigma_k^2}^{\text{rate}}\right)$$

where $R^{\text{shape}}_{\sigma^2_k} = a + \frac{1}{2}\sum_{i=1}^{I^{fp}} I(S_i^{(m-1)} = k) * n_i$ and
$R^{\text{rate}}_{\sigma^2_k} = b + \frac{1}{2}\sum_{i=1}^{I^{fp}}\sum_{j=1}^{n_i}(Y_{ij}^{obs} - \mathbf{X}_{ij}^{obs}B_k^{(m)} - \phi_i^{Y(m-1)})^2 * I(S_i^{(m-1)} = k)$.

(i) Sample from $P(\tau_Y^{2(m)}|\mathbf{O}, \mathbf{S}^{(m-1)}, \mathbf{C}^{mis(m-1)}, \mathbf{D}^{mis(m-1)}, \boldsymbol{\theta}^{(m)}_{-\tau^2_Y})$.

We follow the derivation described in (e) to obtain the complete conditional of $\tau_Y^{2(m)}$:

$$
\begin{aligned}
&P(\tau_Y^{2(m)}|\mathbf{O}, \mathbf{S}^{(m-1)}, \mathbf{C}^{mis(m-1)}, \mathbf{D}^{mis(m-1)}\boldsymbol{\theta}^{(m)}_{-\tau^2_Y}) \\
&\propto (\tau_Y^2)^{-\left(\frac{I-1}{2}\right)-1} e^{\frac{-\frac{1}{2}\sum_i (\phi_i^{Y(m-1)})^2}{\tau_Y^2}} I(0 < \tau_Y^2 < U_{\tau_D^2}) \\
&= \text{Truncated-Inv-Gamma}_{\left(0, U_{\tau^2_Y}\right)}\left(\tau_Y^{2(m)}|R^{\text{shape}}_{\tau^2_Y}, R^{\text{rate}}_{\tau^2_Y}\right)
\end{aligned}
$$

where $R^{\text{rate}}_{\tau^2_Y} = \frac{1}{2}\sum_i \left(\phi_i^{Y(m-1)}\right)^2$ and $R^{\text{shape}}_{\tau^2_{Y,k}} = \left(\frac{I-1}{2}\right)$.

(j) Sample from $P(\phi^{Y(m)}|\mathbf{O}, \mathbf{S}^{(m-1)}, \mathbf{C}^{mis(m-1)}, \mathbf{D}^{mis(m-1)}, \boldsymbol{\theta}^{(m)}_{-\phi^Y})$.

The vector $\phi^{Y(m)}$ contains the cluster-level outcome effects for $i = 1, ..., I^{fp}$. Without loss of generality, assume $S_i^{(m-1)} = k$. The complete conditional of $\phi_i^{Y(m)}$ is

$$
\begin{aligned}
&P(\phi_i^{Y(m)}|\mathbf{O}, \mathbf{S}^{(m-1)}, \mathbf{C}^{mis(m-1)}, \mathbf{D}^{mis(m-1)}, \boldsymbol{\theta}^{(m)}_{-\phi^Y}) \\
&\propto e^{\frac{-(\phi_i^Y)^2}{2\tau_Y^{2(m)}}} \prod_{j=1}^{n_i} e^{-\frac{1}{2}\frac{(Y_{ij}^{obs} - \mathbf{X}_{ij}^{obs}B_k^{(m)} - \phi_i^Y)^2}{\sigma_k^{2(m)}}} \\
&= \exp\left\{-\frac{(\phi_i^Y)^2\sigma_k^{2(m)} + \tau_Y^{2(m)}\sum_{j=1}^{n_i}(\phi_i^Y - \hat{Y}_{ij})^2}{2\tau_Y^{2(m)}\sigma_k^{2(m)}}\right\},
\end{aligned}
$$

where $\hat{Y}_{ij} = Y_{ij}^{obs} - \mathbf{X}_{ij}^{obs}B_k^{(m)}$. After rearranging terms, we obtain

$$
\begin{aligned}
P(\phi_i^{Y(m)}|\mathbf{O}, \mathbf{S}^{(m-1)}, \mathbf{C}^{mis(m-1)}, \mathbf{D}^{mis(m-1)}, \boldsymbol{\theta}^{(m)}_{-\phi^Y}) &\propto \exp\left\{-\frac{\left(\phi_i^Y - \frac{\sum_{j=1}^{n_i}\hat{Y}_{ij}}{\sigma_k^{2(m)}/\tau_Y^{2(m)}+n_i}\right)^2}{2\frac{\tau_Y^{2(m)}\sigma_k^{2(m)}}{\sigma_k^{2(m)}+\tau_Y^{2(m)}n_i}}\right\} \\
&\propto N(\hat{\phi}_i^Y, \hat{v}^2_{\phi_i^Y})
\end{aligned}
$$

where $\hat{v}^2_{\phi_i^Y} = \frac{\sigma_k^{2(m)}\tau_Y^{2(m)}}{n_i\tau_Y^{2(m)}+\sigma_k^{2(m)}}$ and $\hat{\phi}_i^Y = \frac{\sum_{j=1}^{n_i}\hat{Y}_{ij}}{\sigma_k^{2(m)}/\tau_Y^{2(m)}+n_i}$.

2. Sample from $P(\mathbf{S}^{(m)}|\mathbf{O}, \mathbf{C}^{mis(m-1)}, \mathbf{D}^{mis(m-1)}, \boldsymbol{\theta}^{(m)})$.

The completer conditional distribution for $\mathbf{S}^{(m)}$ is

$$
\begin{aligned}
&P(\mathbf{S}^{(m)}|\mathbf{O}, \mathbf{C}^{mis(m-1)}, \mathbf{D}^{mis(m-1)}, \boldsymbol{\theta}^{(m)}_{-S}) \propto \\
&\prod_{k=1}^{K}\prod_{i=1}^{I^{fp}} \pi_k^{(m)I(S_i=k)} \prod_{k=1}^{K}\prod_{i=1}^{I^{fp}} \text{MVN}(\mathbf{C}_i^{(m-1)}, \mathbf{Z}_i|\boldsymbol{\mu}_k^{S(m)}, \Sigma^{(m)})^{I(S_i=k)} \times \\
&\prod_{k=1}^{K}\prod_{i=1}^{I^{fp}}\prod_{j=1}^{n_i} \left[N(Y_{ij}^{obs}|\mathbf{X}_{ij}^{obs}B_k^{(m)} + \phi_i^{Y(m)}, \sigma_k^{2(m)})\right]^{I(S_i=k)} \times \\
&\prod_{k=1}^{K}\prod_{i=1}^{I^{fp}}\prod_{j=1}^{n_i} \left[\text{Bernoulli}\left(D_{ij}^{(m-1)}|\text{probit}^{-1}(\mu_k^{D(m)} + \mathbf{X}_{ij}\alpha_k^{(m)} + \phi_i^{D(m)})\right)\right]^{I(S_i=k)}.
\end{aligned}
$$

For cluster i, let

$$
\begin{aligned}
p_{i,k}^{S(m)} =& \pi_k^{(m)} \text{MVN}(\mathbf{C}_i^{(m-1)}, \mathbf{Z}_i|\boldsymbol{\mu}_k^{S(m)}, \Sigma^{(m)}) \times \\
&\prod_{j=1}^{n_i} N(Y_{ij}^{obs}|\mathbf{X}_{ij}^{obs}B_k^{(m)} + \phi_i^{Y(m)}, \sigma_k^{2(m)})] \times \\
&\prod_{j=1}^{n_i} \text{Bernoulli}\left(D_{ij}^{(m-1)}|\text{probit}^{-1}(\mu_k^{D(m)} + \mathbf{X}_{ij}\alpha_k^{(m)} + \phi_i^{D(m)})\right).
\end{aligned}
$$

The term $p_{i,k}^{S(m)}$ is the liklihood function evaluated at $S_i = k$ with the parameters $\theta_{-\mathbf{S}}^{(m)}$. The complete conditional for $S_i^{(m)}$ is

$$
P(S_i^{(m)}|\mathbf{O}, \mathbf{C}^{mis(m-1)}, \mathbf{D}^{mis(m-1)}, \boldsymbol{\theta}^{(m)}_{-S}) = \text{Categrorical}\left(S_i^{(m)}|K, \tilde{p}_i^{S(m)}\right)
$$

where $\tilde{p}_i^{S(m)} = \left[\frac{p_{i,1}^{S(m)}}{\sum_{k=1}^{K} p_{i,k}^{S(m)}}, \ldots, \frac{p_{i,k}^{S(m)}}{\sum_{k=1}^{K} p_{i,k}^{S(m)}}\right]$. Sample $S_i^{(m)}$ for $i = 1, ..., I^{fp}$ to obtain $\mathbf{S}^{(m)}$.

3. Sample from $P(\mathbf{C}^{mis(m)}|\mathbf{O}, \mathbf{S}^{(m)}, \mathbf{D}^{mis(m-1)}, \boldsymbol{\theta}^{(m)})$

Without loss of generality, assume $W_i = 0$ and $S_i^{(m)} = k$. The complete conditional distribu-

tion for $\mathbf{C}_i^{mis(m)}$ is

$$P(\mathbf{C}_i^{mis(m)}|\mathbf{O}, \mathbf{S}^{(m)}, \mathbf{D}^{mis(m-1)}\boldsymbol{\theta}^{(m)}) \propto \text{MVN}(\mathbf{C}_i, \mathbf{Z}_i|\boldsymbol{\mu}_k^{S(m)}, \Sigma^{(m)})$$
$$\propto \text{MVN}(\mathbf{C}_i|M_{C_i}, \Omega_{C_i})$$

where $M_{C_i} = \boldsymbol{\mu}_k^{C(m)} + \Sigma_{CZ}^{(m)}\Sigma_{ZC}^{(m)-1}(\mathbf{Z}_i - \boldsymbol{\mu}_k^{Z(m)})$, $\Omega_{C_i} = \Sigma_{CC}^{(m)} - \Sigma_{CZ}^{(m)}\Sigma_{ZC}^{(m)}\Sigma_{ZC}^{(m)}$ and $\Sigma^{(m)} = \begin{bmatrix} \Sigma_{CC}^{(m)} & \Sigma_{CZ}^{(m)} \\ \Sigma_{ZC}^{(m)} & \Sigma_{ZZ}^{(m)} \end{bmatrix}$. Sample $\mathbf{C}_i^{mis(m)}$ for i such that $W_i = 0$ to obtain $\mathbf{C}^{mis(m)}$.

4. Sample from $P(\mathbf{D}^{mis(m)}|\mathbf{O}, , \mathbf{S}^{(m)}, \mathbf{C}^{mis(m)}, \boldsymbol{\theta}^{(m)})$.

   Without loss of generality, assume that $W_i = 0$ and $S_i^{(m)} = k$. The complete conditional distribution of $\mathbf{D}^{mis(m)}$ is

$$P(\mathbf{D}_{ij}^{mis(m)}|\mathbf{O}, \mathbf{S}^{(m)}, \mathbf{C}^{mis(m)}, \boldsymbol{\theta}^{(m)}) \propto \text{Bernoulli}\left(D_{ij}|\text{probit}^{-1}(\mu_k^{D(m)} + \mathbf{X}_{ij}\alpha_k^{(m)} + \phi_i^{D(m)})\right) \times$$
$$N(Y_{ij}^{obs}|\mathbf{X}^{obs}B_k^{(m)} + \phi_i^{Y(m)}, \sigma_k^{2(m)}).$$

   Let $\mathbf{X}_{ij}^{D=1} = [1,\ \mathbf{X}_{ij},\ (1 - W_{ij}),\ W_i * \mathbf{X}_{ij},\ W_i]$ and $\mathbf{X}_{ij}^{D=0} = [1,\ \mathbf{X}_{ij},\ 0,\ \mathbf{0},\ 0]$. In addition, let

$$p_{ij,d}^D = \text{probit}^{-1}\left(\mu_k^{D(m)} + \mathbf{X}_{ij}\alpha_k^{(m)} + \phi_i^{D(m)}\right)^d (1 - \text{probit}^{-1}\left(\mu_k^{D(m)} + \mathbf{X}_{ij}\alpha_k^{(m)} + \phi_i^{D(m)}\right))^{1-d} \times$$
$$N(Y_{ij}^{obs}|\mathbf{X}_{ij}^{D=d}B_k^{(m)} + \phi_i^{Y(m)}, \sigma_k^{2(m)}).$$

   The term $p_{ij,d}^D$ is the likelihood evaluated when $D_{ij} = d$ at $\boldsymbol{\theta}^{(m)}$. The complete conditional simplfies to

$$P(\mathbf{D}_{ij}^{mis(m)}|\mathbf{O}, \mathbf{S}^{(m)}, \mathbf{C}^{mis(m)}, \boldsymbol{\theta}^{(m)}) = \text{Bernoulli}\left(\mathbf{D}_{ij}^{mis(m)}\left|\frac{p_{ij,1}^D}{p_{ij,1}^D + p_{ij,0}^D}\right.\right).$$

   Sample $\mathbf{D}_{ij}^{mis(m)}$ for all j and i such that $W_i = 0$ to obtain $\mathbf{D}^{mis(m)}$.

A2.2 Imputing Unobserved Outcomes and Compliance

In Procedure 1 outlined in Section 2.3.3, we must impute the unobserved outcomes and unobserved individual and cluster-level compliance values in order to compute the value of a finite sample causal estimand. At each iteration of the Gibbs samplier outlined in Section A2.1 we obtain a posterior draw of the model parameters, $\boldsymbol{\theta}$, the latent allocation variable $\mathbf{S}$, the unobserved cluster-level compliance metrics, $\mathbf{C}^{mis}$ and the unobserved individual-level compliance values $\mathbf{D}^{mis}$. Therefore, we need only sample the unobserved potential outcomes, $\mathbf{Y}^{mis}$, from their posterior predictive distribution. We have,

$$P(\mathbf{Y}^{mis}|\mathbf{Y}^{obs},\mathbf{D}^{obs},\mathbf{C}^{obs},\mathbf{Z},\mathbf{X}) = \int P(\mathbf{Y}^{mis}|\mathbf{Y}^{obs},\mathbf{D}^{obs},\mathbf{C}^{obs},\mathbf{Z},\mathbf{X},\mathbf{S},\mathbf{D}^{mis},\mathbf{C}^{mis},\boldsymbol{\theta})\times \tag{3.28}$$
$$P(\mathbf{S},\mathbf{D}^{mis},\mathbf{C}^{mis},\boldsymbol{\theta}|\mathbf{Y}^{obs},\mathbf{D}^{obs},\mathbf{C}^{obs},\mathbf{Z},\mathbf{X})d\boldsymbol{\theta}d\mathbf{S}d\mathbf{C}^{mis}d\mathbf{D}^{mis}$$

Based on the parametric models outlined in Section 2.3.3, we can write,

$$P(\mathbf{Y}^{mis}|\mathbf{Y}^{obs},\mathbf{D}^{obs},\mathbf{C}^{obs},\mathbf{Z},\mathbf{X},\mathbf{S},\mathbf{D}^{mis},\mathbf{C}^{mis},\boldsymbol{\theta}) \propto \tag{3.29}$$
$$\left\{\prod_{k=1}^{K}\prod_{i:W_i=0}\prod_{j=1}^{n_i} N(Y_{ij}(1,D_{ij})|\mu_k^Y + \mathbf{X}_{ij}\beta_{0,k} + D_{ij}*(\delta_{1,k}+\mathbf{X}_{ij}\beta_{1,k}) + \phi_i^Y, \sigma_k^2)^{I(S_i=k)}\times \right.$$
$$\left. \prod_{i:W_i=1}\prod_{j=1}^{n_i} N(Y_{ij}(0,D_{ij})|\mu_k^Y + D_{ij}*\delta_{0,k} + \mathbf{X}_{ij}\beta_{0,k} + \phi_i^Y, \sigma_k^2)^{I(S_i=k)}\right\} =$$
$$\prod_{k=1}^{K}\prod_{i=1}^{I^{fp}}\prod_{j=1}^{n_i} N(Y_{ij}^{mis}|\mathbf{X}_{ij}^{mis}B_k + \phi_i^Y, \sigma_k^2)^{I(S_i=k)},$$

where $\mathbf{X}_{ij}^{mis} = [1,\ \mathbf{X}_{ij},\ W_i * D_{ij},\ (1-W_i)*D_{ij}*\mathbf{X}_{ij},\ (1-W_i)*D_{ij}]$ and B_k is defined as in step (g) of the Gibbs sampler derived in section A2.1. The integral in Equation (3.28) and the posterior predictive distribution in Equation (3.29) imply the following procedure.

1. For $m = 1, ..., M$:

   (a) Obtain $\boldsymbol{\theta}^{(m)}$, $\mathbf{S}^{(m)}$, $\mathbf{C}^{mis(m)}$ and $\mathbf{D}^{mis(m)}$ by implementing the Gibbs sampling algorithm outlined in Section A2.1.

(b) For individuals in clusters such that $S_i^{(m)} = k$, sample from

$$Y_{ij}^{mis(m)} \sim N(\mathbf{X}_{ij}^{mis} B_k^{(m)} + \phi_i^{Y(m)}, \sigma_k^{2(m)}).$$

(c) Repeat step two for $k = 1, ..., K$.

2. Combine the M draws to obtain

$$\left\{\mathbf{Y}^{mis(1)}, \mathbf{S}^{(1)}, \mathbf{D}^{mis(1)}, \mathbf{C}^{mis(1)}, ..., \mathbf{Y}^{mis(M)}, \mathbf{S}^{(M)}, \mathbf{D}^{mis(M)}, \mathbf{C}^{mis(M)}\right\}.$$

A2.3 Imputing Treatment and Control Values in the Super-population

In step 2(b) of Procedure 3 described in Section 2.3.3 we impute treatment and control values in the super-population by sampling $\left\{\mathbf{Y}(1,\mathbf{D})^{(m,r)}, \mathbf{Y}(0,\mathbf{D})^{(m,r)}, \mathbf{D}^{(m,r)}, \mathbf{C}^{(m,r)}, \mathbf{Z}^{(m,r)}, \mathbf{S}^{(m,r)}\right\}$ for $r = 1, ..., R$ from $P(\mathbf{Y}(1,\mathbf{D}), \mathbf{Y}(0,\mathbf{D}), \mathbf{D}, \mathbf{C}, \mathbf{Z}|\boldsymbol{\theta}^{(m)}, \mathbf{X})$ at iteration $m = 1, ..., M$. We have that

$$P(\mathbf{Y}(1,\mathbf{D}), \mathbf{Y}(0,\mathbf{D}), \mathbf{D}, \mathbf{C}, \mathbf{Z}, \mathbf{S}|\boldsymbol{\theta}^{(m)}, \mathbf{X}) =$$
$$P(\mathbf{Y}(1,\mathbf{D}), \mathbf{Y}(0,\mathbf{D})|\mathbf{D}, \mathbf{C}, \mathbf{Z}, \mathbf{S}, \boldsymbol{\theta}^{(m)}, \mathbf{X}) P(\mathbf{D}|\mathbf{C}, \mathbf{Z}, \mathbf{S}, \boldsymbol{\theta}^{(m)}, \mathbf{X}) P(\mathbf{C}, \mathbf{Z}|\mathbf{S}, \boldsymbol{\theta}^{(m)}) P(\mathbf{S}|\boldsymbol{\theta}^{(m)}).$$

The algorithm to obtain $\left\{\mathbf{Y}(1,\mathbf{D})^{(m,r)}, \mathbf{Y}(0,\mathbf{D})^{(m,r)}, \mathbf{D}^{(m,r)}, \mathbf{C}^{(m,r)}, \mathbf{Z}^{(m,r)}, \mathbf{S}^{(m,r)}\right\}$ is

1. For $i = 1, ..., I^{fp}$,

   (a) Sample $S_i^{(m,r)}$ from Categorical$(K, \pi^{(m)})$.

   (b) Sample $\{\mathbf{C}_i^{(m,r)}, \mathbf{Z}_i^{(m,r)}\}$ from MVN $\left(\boldsymbol{\mu}_k^{S(m)}, \Sigma^{(m)}\right)$.

   (c) Sample $D_i^{(m,r)}$ from Bernoulli$\left(\text{probit}^{-1}(\mu_k^{D(m)} + \mathbf{X}_{ij}\alpha_k^{(m)} + \phi_i^{D(m)})\right)$.

   (d) Sample $Y_{ij}(1, D_{ij})^{(m,r)}$ from $N(\mu_k^{Y(m)} + \mathbf{X}_{ij}\beta_{0,k}^{(m)} + D_{ij}^{(m,r)} * (\delta_{1,k}^{(m)} + \mathbf{X}_{ij}\beta_{1,k}^{(m)}) + \phi_i^{Y(m)}, \sigma_k^{2(m)})$.

   (e) Sample $Y_{ij}(0, D_{ij})^{(m,r)}$ from $N(\mu_k^{Y(m)} + \mathbf{X}_{ij}\beta_{0,k}^{(m)} + D_{ij}^{(m,r)} * \delta_{0,k}^{(m)} + \phi_i^{Y(m)}, \sigma_k^{2(m)})$.

2. Combine these draws to obtain

$$\left\{\mathbf{Y}(1,\mathbf{D})^{(m,r)}, \mathbf{Y}(0,\mathbf{D})^{(m,r)}, \mathbf{D}^{(m,r)}, \mathbf{C}^{(m,r)}, \mathbf{Z}^{(m,r)}, \mathbf{S}^{(m,r)}\right\}$$

Repeat this procedure for $r = 1, ..., R$ and $m = 1, ..., M$.

A2.4 Super-population Estimand Derivation

We derive the super-population estimands described in section 1.2.2 as functions of the model parameters. Let $p_{ijk} = \text{probit}^{-1}\left(\left(\mathbf{X}_{ij}\alpha_k + \mu_k^D\right)/\sqrt{\tau_{D,k}^2 + 1}\right)$, $\delta_k = \delta_{1,k} - \delta_{0,k}$ and $\mathcal{I}_k = \{i|S_i = k\}$ be the index of observed facilities in latent compliance strata k. The derivations for each estimand in the super population are completed in the subsections below.

A2.4.1 ITT

$P_k(\mathbf{X}_{ij}) = \hat{F}(\mathbf{X}_{ij})$

If we assume $P_k(\mathbf{X}_{ij}) = \hat{F}(\mathbf{X}_{ij})$, we have,

$$\begin{aligned}
E_{sp}[Y_{ij}(1, D_{ij})] &= E_{sp}[\mu_{S_i}^Y + \mathbf{X}_{ij}\beta_{0,S_i} + D_{ij} * (\mathbf{X}_{ij}\beta_{1,S_i} + \delta_{1,S_i}) + \phi_i^Y] \\
&= E_{sp}\left[E_{sp}\left[\mu_{S_i}^Y + \mathbf{X}_{ij}\beta_{0,S_i} + D_{ij} * (\mathbf{X}_{ij}\beta_{1,S_i} + \delta_{1,S_i})|S_i, \mathbf{X}_{ij}\right]\right] \\
&= E_{sp}\left[\mu_{S_i}^Y + \mathbf{X}_{ij}\beta_{0,S_i} + E_{sp}\left[D_{ij}|S_i, \mathbf{X}_{ij}\right] * (\mathbf{X}_{ij}\beta_{1,S_i} + \delta_{1,S_i})\right] \\
&= E_{sp}\left[\mu_{S_i}^Y + \mathbf{X}_{ij}\beta_{0,S_i} + p_{ijS_i} * (\mathbf{X}_{ij}\beta_{1,S_i} + \delta_{1,S_i})\right] \\
&= E_{sp}\left[\left[\mu_{S_i}^Y + \mathbf{X}_{ij}\beta_{0,S_i} + p_{ijS_i} * (\mathbf{X}_{ij}\beta_{1,S_i} + \delta_{1,S_i})|\mathbf{X}_{ij}\right]\right] \\
&= \frac{1}{n}\sum_{ij}\sum_{k}\pi_k\left(\mu_k^Y + \mathbf{X}_{ij}\beta_{0,k} + p_{ijk} * (\mathbf{X}_{ij}\beta_{1,k} + \delta_{1,k})\right).
\end{aligned}$$

Similarly,

$$E_{sp}[Y_{ij}(0, D_{ij})] = \frac{1}{n}\sum_{ij}\sum_{k}\pi_k\left(\mu_k^Y + \mathbf{X}_{ij}\beta_{0,k} + p_{ijk} * (\delta_{0,k})\right).$$

Thus,

$$E_{sp}[Y_{ij}(1, D_{ij}) - Y_{ij}(0, D_{ij})] = \frac{1}{n}\sum_{ij}\sum_{k}\pi_k p_{ijk}(\mathbf{X}_{ij}\beta_{1,k} + \delta_k).$$

$P_k(\mathbf{X}_{ij}) = \hat{F}_k(\mathbf{X}_{ij})$

If we assume $P_k(\mathbf{X}_{ij}) = \hat{F}_k(\mathbf{X}_{ij})$, we have,

$$
\begin{aligned}
&E_{sp}[Y_{ij}(1, D_{ij})] \\
&= E_{sp}[\mu^Y_{S_i} + \mathbf{X}_{ij}\beta_{0,S_i} + D_{ij} * (\mathbf{X}_{ij}\beta_{1,S_i} + \delta_{1,S_i}) + \phi^Y_i] \\
&= E_{sp}\left[E_{sp}\left[\sum_k I(S_i = k)\left(\mu^Y_k + \mathbf{X}_{ij}\beta_{0,k} + D_{ij} * (\mathbf{X}_{ij}\beta_{1,k} + \delta_{1,k})\right) | S_i = k, \mathbf{X}_{ij}\right]\right] \\
&= E_{sp}\left[\sum_k I(S_i = k)\left(\mu^Y_k + \mathbf{X}_{ij}\beta_{0,k} + E_{sp}[D_{ij}|S_i = k, \mathbf{X}_{ij}] * (\mathbf{X}_{ij}\beta_{1,k} + \delta_{1,k})\right)\right] \\
&= E_{sp}\left[\sum_k I(S_i = k)\left(\mu^Y_k + \mathbf{X}_{ij}\beta_{0,k} + p_{ijk} * (\mathbf{X}_{ij}\beta_{1,k} + \delta_{1,k})\right)\right] \\
&= E_{sp}\left[E_{sp}\left[\sum_k I(S_i = k)\left(\mu^Y_k + \mathbf{X}_{ij}\beta_{0,k} + p_{ijk} * (\mathbf{X}_{ij}\beta_{1,k} + \delta_{1,k})\right) | S_i = k\right]\right] \\
&= \sum_k E_{sp}\left[I(S_i = k)\right]\left(\mu^Y_k + E_{sp}\left[\mathbf{X}_{ij}|S_i = k\right]\beta_{0,k} + E_{sp}\left[p_{ijk} * (\mathbf{X}_{ij}\beta_{1,k} + \delta_{1,k})|S_i = k\right]\right) \\
&= \sum_k \pi_k \left(\mu^Y_k + \frac{1}{\sum_{i\in\mathcal{I}_k} n_i}\sum_{i\in\mathcal{I}_k}\sum_{j=1}^{n_i}\mathbf{X}_{ij}\beta_{0,k} + \frac{1}{\sum_{i\in\mathcal{I}_k} n_i}\sum_{i\in\mathcal{I}_k}\sum_{j=1}^{n_i} p_{ijk} * (\mathbf{X}_{ij}\beta_{1,k} + \delta_{1,k})\right).
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
&E_{sp}[Y_{ij}(0, D_{ij})] \\
&= \sum_k \pi_k \left(\mu^Y_k + \frac{1}{\sum_{i\in\mathcal{I}_k} n_i}\sum_{i\in\mathcal{I}_k}\sum_{j=1}^{n_i}\mathbf{X}_{ij}\beta_{0,k} + \frac{1}{\sum_{i\in\mathcal{I}_k} n_i}\sum_{i\in\mathcal{I}_k}\sum_{j=1}^{n_i} p_{ijk} * \delta_{0,k}\right).
\end{aligned}
$$

Thus,

$$
\begin{aligned}
&E_{sp}[Y_{ij}(1, D_{ij}) - Y_{ij}(0, D_{ij})] \\
&= \sum_k \pi_k \left(\frac{1}{\sum_{i\in\mathcal{I}_k} n_i}\sum_{i\in\mathcal{I}_k}\sum_{j=1}^{n_i} p_{ijk} * (\mathbf{X}_{ij}\beta_{1,k} + \delta_k)\right).
\end{aligned}
$$

A2.4.2 CACE

$P_k(\mathbf{X}_{ij}) = \hat{F}(\mathbf{X}_{ij})$

If we assume $P_k(\mathbf{X}_{ij}) = \hat{F}(\mathbf{X}_{ij})$, we have,

$$
\begin{aligned}
&E_{sp}[Y_{ij}(1, D_{ij})|D_{ij} = 1] \\
&= E_{sp}\left[\mu^Y_{S_i} + \mathbf{X}_{ij}\beta_{0,S_i} + D_{ij} * (\mathbf{X}_{ij}\beta_{1,S_i} + \delta_{1,S_i}) + \phi^Y_i \,|D_{ij} = 1\right] \\
&= E_{sp}\left[\sum_k I(S_i = k)\left(\mu^Y_k + \mathbf{X}_{ij}\beta_{0,k} + D_{ij} * (\mathbf{X}_{ij}\beta_{1,k} + \delta_{1,k})\right)|D_{ij} = 1\right] \\
&= E_{sp}\left[E_{sp}\left[\sum_k I(S_i = k)\left(\mu^Y_k + \mathbf{X}_{ij}\beta_{0,k} + D_{ij} * (\mathbf{X}_{ij}\beta_{1,k} + \delta_{1,k})\right)|S_i = k, D_{ij} = 1\right]|D_{ij} = 1\right] \\
&= \sum_k E_{sp}\left[I(S_i = k)|D_{ij} = 1\right]\left(\mu^Y_k + E_{sp}\left[\mathbf{X}_{ij}|S_i = k, D_{ij} = 1\right](\beta_{0,k} + \beta_{1,k}) + \delta_{1,k}\right)
\end{aligned}
$$

Note that,

$$
\begin{aligned}
E_{sp}\left[\mathbf{X}_{ij}|D_{ij} = 1, S_i = k\right] &= \int_{\mathbf{X}_{ij}} \mathbf{X}_{ij} P(\mathbf{X}_{ij}|D_{ij} = 1, S_i = k) d\mathbf{X}_{ij} \\
&= \int_{\mathbf{X}_{ij}} \mathbf{X}_{ij} \frac{P(D_{ij} = 1|\mathbf{X}_{ij}, S_i = k)P(\mathbf{X}_{ij})}{P(D_{ij} = 1|S_i = k)} d\mathbf{X}_{ij} \\
&= \frac{1}{P(D_{ij} = 1|S_i = k)} \int_{\mathbf{X}_{ij}} \mathbf{X}_{ij} p_{ijk} P(\mathbf{X}_{ij}) d\mathbf{X}_{ij} \\
&= \frac{1}{E_{sp}\left[D_{ij}|S_i = k\right]} \int_{\mathbf{X}_{ij}} \mathbf{X}_{ij} p_{ijk} P(\mathbf{X}_{ij}) d\mathbf{X}_{ij} \\
&= \frac{\frac{1}{n}\sum_{ij} \mathbf{X}_{ij} p_{ijk}}{\frac{1}{n}\sum_{ij} p_{ijk}},
\end{aligned}
$$

and,

$$
\begin{aligned}
E_{sp}\left[I(S_i = k)|D_{ij} = 1\right] &= Pr(S_i = k|D_{ij} = 1) \\
&= \frac{Pr(D_{ij} = 1|S_i = k)P(S_i = k)}{Pr(D_{ij} = 1)} \\
&= \frac{E_{sp}\left[D_{ij}|S_i = k\right]\pi_k}{E_{sp}[D_{ij}]} \\
&= \frac{\pi_k \frac{1}{n}\sum_{ij} p_{ijk}}{\sum_k \pi_k \frac{1}{n}\sum_{ij} p_{ijk}}.
\end{aligned}
$$

Thus,

$$E_{sp}[Y_{ij}(1,D_{ij})|D_{ij}=1] = \frac{\sum_k \pi_k \frac{1}{n}\sum_{ij} p_{ijk}\left(\mu_k^Y + \frac{\frac{1}{n}\sum_{ij}\mathbf{X}_{ij}p_{ijk}}{\frac{1}{n}\sum_{ij}p_{ijk}}(\beta_{0,k}+\beta_{1,k}) + \delta_{1,k}\right)}{\sum_k \pi_k \frac{1}{n}\sum_{ij} p_{ijk}},$$

$$E_{sp}[Y_{ij}(0,D_{ij})|D_{ij}=1] = \frac{\sum_k \pi_k \frac{1}{n}\sum_{ij} p_{ijk}\left(\mu_k^Y + \frac{\frac{1}{n}\sum_{ij}\mathbf{X}_{ij}p_{ijk}}{\frac{1}{n}\sum_{ij}p_{ijk}}(\beta_{0,k}) + \delta_{0,k}\right)}{\sum_k \pi_k \frac{1}{n}\sum_{ij} p_{ijk}}, \text{ and}$$

$$E_{sp}[Y_{ij}(1,D_{ij}) - Y_{ij}(0,D_{ij})|D_{ij}=1] = \frac{\sum_k \pi_k \frac{1}{n}\sum_{ij} p_{ijk}\left(\frac{\frac{1}{n}\sum_{ij}\mathbf{X}_{ij}\beta_{1,k}p_{ijk}}{\frac{1}{n}\sum_{ij}p_{ijk}} + \delta_k\right)}{\sum_k \pi_k \frac{1}{n}\sum_{ij} p_{ijk}}$$

$P_k(\mathbf{X}_{ij}) = \hat{F}_k(\mathbf{X}_{ij})$

If we assume $P_k(\mathbf{X}_{ij}) = \hat{F}_k(\mathbf{X}_{ij})$, we have,

$$\begin{aligned}
&E_{sp}[Y_{ij}(1,D_{ij})|D_{ij}=1] \\
&= E_{sp}\left[\mu_{S_i}^Y + \mathbf{X}_{ij}\beta_{0,S_i} + D_{ij} * (\mathbf{X}_{ij}\beta_{1,S_i} + \delta_{1,S_i}) + \phi_i^Y | D_{ij}=1\right] \\
&= E_{sp}\left[\sum_k I(S_i=k)\left(\mu_k^Y + \mathbf{X}_{ij}\beta_{0,k} + D_{ij} * (\mathbf{X}_{ij}\beta_{1,k} + \delta_{1,k})\right) | D_{ij}=1\right] \\
&= E_{sp}\left[E_{sp}\left[\sum_k I(S_i=k)\left(\mu_k^Y + \mathbf{X}_{ij}\beta_{0,k} + D_{ij} * (\mathbf{X}_{ij}\beta_{1,k} + \delta_{1,k})\right) | S_i=k, D_{ij}=1\right] | D_{ij}=1\right] \\
&= \sum_k E_{sp}\left[I(S_i=k)|D_{ij}=1\right]\left(\mu_k^Y + E_{sp}\left[\mathbf{X}_{ij}|S_i=k, D_{ij}=1\right](\beta_{0,k}+\beta_{1,k}) + \delta_{1,k}\right).
\end{aligned}$$

Note,

$$\begin{aligned}
E\left[\mathbf{X}_{ij}|D_{ij}=1, S_i=k\right] &= \int_{\mathbf{X}_{ij}} \mathbf{X}_{ij} P(\mathbf{X}_{ij}|D_{ij}=1, S_i=k) d\mathbf{X}_{ij} \\
&= \int_{\mathbf{X}_{ij}} \mathbf{X}_{ij} \frac{P(D_{ij}=1|\mathbf{X}_{ij}, S_i=k)P(\mathbf{X}_{ij}|S_i=k)}{P(D_{ij}=1|S_i=k)} d\mathbf{X}_{ij} \\
&= \int_{\mathbf{X}_{ij}} \mathbf{X}_{ij} \frac{E\left[D_{ij}|\mathbf{X}_{ij}, S_i=k\right]P(\mathbf{X}_{ij}|S_i=k)}{E_{sp}\left[D_{ij}|S_i=k\right]} d\mathbf{X}_{ij} \\
&= \frac{1}{E_{sp}\left[D_{ij}|S_i=k\right]} \int_{\mathbf{X}_{ij}} \mathbf{X}_{ij} p_{ijk} \hat{F}_k(\mathbf{X}_{ij}) d\mathbf{X}_{ij} \\
&= \frac{1}{E_{sp}\left[D_{ij}|S_i=k\right]} \frac{1}{\sum_{i\in\mathcal{I}_k} n_i} \sum_{i\in\mathcal{I}_k}\sum_{j=1}^{n_i} \mathbf{X}_{ij} p_{ijk},
\end{aligned}$$

$$
\begin{aligned}
E_{sp}\left[D_{ij}|S_i=k\right] &= E_{sp}\left[E_{sp}\left[D_{ij}|S_i=k,\mathbf{X}_{ij}\right]|S_i=k\right] \\
&= E_{sp}\left[\text{probit}^{-1}\left(\left(\mathbf{X}_{ij}\alpha_k+\mu_k^D\right)/\sqrt{\tau_{D,k}^2+1}\right)|S_i=k\right] \\
&= \frac{1}{\sum_{i\in\mathcal{I}_k} n_i}\sum_{i\in\mathcal{I}_k}\sum_{j=1}^{n_i}\text{probit}^{-1}\left(\left(\mathbf{X}_{ij}\alpha_k+\mu_k^D\right)/\sqrt{\tau_{D,k}^2+1}\right) \\
&= \frac{1}{\sum_{i\in\mathcal{I}_k} n_i}\sum_{i\in\mathcal{I}_k}\sum_{j=1}^{n_i} p_{ijk}.
\end{aligned}
$$

Also,

$$
\begin{aligned}
E_{sp}\left[I(S_i=k)|D_{ij}=1\right] &= Pr(S_i=k|D_{ij}=1) \\
&= \frac{P(D_{ij}=1|S_i=k)P(S_i=k)}{P(D_{ij}=1)} \\
&= \frac{\pi_k E_{sp}\left[D_{ij}|S_i=k\right]}{E_{sp}\left[D_{ij}\right]} \\
\text{From previous derivation} &= \frac{\pi_k \frac{1}{\sum_{i\in\mathcal{I}_k} n_i}\sum_{i\in\mathcal{I}_k}\sum_{j=1}^{n_i} p_{ijk}}{E_{sp}\left[D_{ij}\right]} \\
&= \frac{\pi_k \frac{1}{\sum_{i\in\mathcal{I}_k} n_i}\sum_{i\in\mathcal{I}_k}\sum_{j=1}^{n_i} p_{ijk}}{E_{sp}\left[E_{sp}\left[D_{ij}|S_i\right]\right]} \\
&= \frac{\pi_k \frac{1}{\sum_{i\in\mathcal{I}_k} n_i}\sum_{i\in\mathcal{I}_k}\sum_{j=1}^{n_i} p_{ijk}}{E_{sp}\left[\frac{1}{\sum_{i\in\mathcal{I}_{S_i}} n_i}\sum_{i\in\mathcal{I}_{S_i}}\sum_{j=1}^{n_i} p_{ijk}\right]} \\
&= \frac{\pi_k \frac{1}{\sum_{i\in\mathcal{I}_k} n_i}\sum_{i\in\mathcal{I}_k}\sum_{j=1}^{n_i} p_{ijk}}{E_{sp}\left[\sum_s I(S_i=s)\frac{1}{\sum_{i\in\mathcal{I}_s} n_i}\sum_{i\in\mathcal{I}_s}\sum_{j=1}^{n_i} p_{ijs}\right]} \\
&= \frac{\pi_k \frac{1}{\sum_{i\in\mathcal{I}_k} n_i}\sum_{i\in\mathcal{I}_k}\sum_{j=1}^{n_i} p_{ijk}}{\sum_s \pi_s \frac{1}{\sum_{i\in\mathcal{I}_s} n_i}\sum_{i\in\mathcal{I}_s}\sum_{j=1}^{n_i} p_{ijs}}.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
&E_{sp}[Y_{ij}(1,D_{ij})|D_{ij}=1] \\
&= \sum_k E_{sp}\left[I(S_i=k)|D_{ij}=1\right]\left(\mu_k^Y + E_{sp}\left[\mathbf{X}_{ij}|S_i=k,D_{ij}=1\right](\beta_{0,k}+\beta_{1,k})+\delta_{1,k}\right) \\
&= \frac{\sum_k \pi_k \frac{1}{\sum_{i\in\mathcal{I}_k} n_i}\sum_{i\in\mathcal{I}_k}\sum_{j=1}^{n_i} p_{ijk}\left(\mu_k^Y + \frac{\frac{1}{\sum_{i\in\mathcal{I}_k} n_i}\sum_{i\in\mathcal{I}_k}\sum_{j=1}^{n_i} p_{ijk}\mathbf{X}_{ij}(\beta_{0,k}+\beta_{1,k})}{\frac{1}{\sum_{i\in\mathcal{I}_k} n_i}\sum_{i\in\mathcal{I}_k}\sum_{j=1}^{n_i} p_{ijk}}+\delta_{1,k}\right)}{\sum_k \pi_k \frac{1}{\sum_{i\in\mathcal{I}_k} n_i}\sum_{i\in\mathcal{I}_k}\sum_{j=1}^{n_i} p_{ijk}},
\end{aligned}
$$

and

$$
\begin{aligned}
&E_{sp}[Y_{ij}(0, D_{ij})|D_{ij} = 1] \\
&= \frac{\sum_k \pi_k \frac{1}{\sum_{i\in\mathcal{I}_k} n_i} \sum_{i\in\mathcal{I}_k} \sum_{j=1}^{n_i} p_{ijk} \left(\mu_k^Y + \frac{\frac{1}{\sum_{i\in\mathcal{I}_k} n_i} \sum_{i\in\mathcal{I}_k} \sum_{j=1}^{n_i} p_{ijk}\mathbf{X}_{ij}}{\frac{1}{\sum_{i\in\mathcal{I}_k} n_i} \sum_{i\in\mathcal{I}_k} \sum_{j=1}^{n_i} p_{ijk}} \beta_{0,k} + \delta_{0,k} \right)}{\sum_k \pi_k \frac{1}{\sum_{i\in\mathcal{I}_k} n_i} \sum_{i\in\mathcal{I}_k} \sum_{j=1}^{n_i} p_{ijk}}.
\end{aligned}
$$

Thus,

$$
\begin{aligned}
&E_{sp}[Y_{ij}(1, D_{ij}) - Y_{ij}(0, D_{ij})|D_{ij} = 1] \\
&= \frac{\sum_k \pi_k \frac{1}{\sum_{i\in\mathcal{I}_k} n_i} \sum_{i\in\mathcal{I}_k} \sum_{j=1}^{n_i} p_{ijk} \left(\frac{\frac{1}{\sum_{i\in\mathcal{I}_k} n_i} \sum_{i\in\mathcal{I}_k} \sum_{j=1}^{n_i} p_{ijk}\mathbf{X}_{ij}\beta_{1,k}}{\frac{1}{\sum_{i\in\mathcal{I}_k} n_i} \sum_{i\in\mathcal{I}_k} \sum_{j=1}^{n_i} p_{ijk}} + \delta_k \right)}{\sum_k \pi_k \frac{1}{\sum_{i\in\mathcal{I}_k} n_i} \sum_{i\in\mathcal{I}_k} \sum_{j=1}^{n_i} p_{ijk}}.
\end{aligned}
$$

A2.4.3 ITT$_k$

$P_k(\mathbf{X}_{ij}) = \hat{F}(\mathbf{X}_{ij})$

If we assume $P_k(\mathbf{X}_{ij}) = \hat{F}(\mathbf{X}_{ij})$, we have,

$$
\begin{aligned}
E_{sp}[Y_{ij}(1, D_{ij})|S_i = k] &= E_{sp}\left[\mu_{S_i}^Y + \mathbf{X}_{ij}\beta_{0,S_i} + D_{ij} * (\mathbf{X}_{ij}\beta_{1,S_i} + \delta_{1,S_i})|S_i = k\right] \\
&= \mu_k^Y + E_{sp}\left[\mathbf{X}_{ij}\right]\beta_{0,k} + E_{sp}\left[E_{sp}\left[D_{ij}|\mathbf{X}_{ij}, S_i = k\right] * (\mathbf{X}_{ij}\beta_{1,k} + \delta_{1,k})|S_i = k\right] \\
&= \mu_k^Y + \frac{1}{n}\sum_{ij} \mathbf{X}_{ij}\beta_{0,k} + E_{sp}\left[p_{ijk} * (\mathbf{X}_{ij}\beta_{1,k} + \delta_{1,k})|S_i = k\right] \\
&= \mu_k^Y + \frac{1}{n}\sum_{ij} \mathbf{X}_{ij}\beta_{0,k} + \frac{1}{n}\sum_{ij} p_{ijk} * (\mathbf{X}_{ij}\beta_{1,k} + \delta_{1,k}).
\end{aligned}
$$

Similarly,

$$
E_{sp}[Y_{ij}(0, D_{ij})|S_i = k] = \mu_k^Y + \frac{1}{n}\sum_{ij} \mathbf{X}_{ij}\beta_{0,k} + \frac{1}{n}\sum_{ij} p_{ijk} * \delta_{0,k}.
$$

Thus,

$$
E_{sp}[Y_{ij}(1, D_{ij}) - Y_{ij}(0, D_{ij})|S_i = k] = \frac{1}{n}\sum_{ij} p_{ijk} * (\mathbf{X}_{ij}\beta_{1,k} + \delta_k).
$$

$P_k(\mathbf{X}_{ij}) = \hat{F}_k(\mathbf{X}_{ij})$

If we assume $P_k(\mathbf{X}_{ij}) = \hat{F}_k(\mathbf{X}_{ij})$, we have,

$$
\begin{aligned}
&E_{sp}[Y_{ij}(1, D_{ij})|S_i = k] \\
&= E_{sp}\left[\mu^Y_{S_i} + \mathbf{X}_{ij}\beta_{0,S_i} + D_{ij} * (\mathbf{X}_{ij}\beta_{1,S_i} + \delta_{1,S_i}) + \phi^Y_i | S_i = k\right] \\
&= E_{sp}\left[\mu^Y_{S_i} + \mathbf{X}_{ij}\beta_{0,S_i} + D_{ij} * (\mathbf{X}_{ij}\beta_{1,S_i} + \delta_{1,S_i}) + \phi^Y_i | S_i = k\right] \\
&= \mu^Y_k + E_{sp}\left[\mathbf{X}_{ij}|S_i = k\right]\beta_{0,k} + E_{sp}\left[D_{ij} * (\mathbf{X}_{ij}\beta_{1,S_i} + \delta_{1,S_i})|S_i = k\right] \\
&= \mu^Y_k + \frac{1}{\sum_{i\in\mathcal{I}_k} n_i}\sum_{i\in\mathcal{I}_k}\sum_{j=1}^{n_i}\mathbf{X}_{ij}\beta_{0,k} + E_{sp}\left[E_{sp}\left[D_{ij}|S_i = k, \mathbf{X}_{ij}\right] * (\mathbf{X}_{ij}\beta_{1,S_i} + \delta_{1,S_i})|S_i = k\right] \\
&= \mu^Y_k + \frac{1}{\sum_{i\in\mathcal{I}_k} n_i}\sum_{i\in\mathcal{I}_k}\sum_{j=1}^{n_i}\mathbf{X}_{ij}\beta_{0,k} + E_{sp}\left[p_{ijk} * (\mathbf{X}_{ij}\beta_{1,S_i} + \delta_{1,S_i})|S_i = k\right] \\
&= \mu^Y_k + \frac{1}{\sum_{i\in\mathcal{I}_k} n_i}\sum_{i\in\mathcal{I}_k}\sum_{j=1}^{n_i}\mathbf{X}_{ij}\beta_{0,k} + \frac{1}{\sum_{i\in\mathcal{I}_k} n_i}\sum_{i\in\mathcal{I}_k}\sum_{j=1}^{n_i}p_{ijk} * (\mathbf{X}_{ij}\beta_{1,k} + \delta_{1,k}).
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
&E_{sp}[Y_{ij}(0, D_{ij})|S_i = k] \\
&= \mu^Y_k + \frac{1}{\sum_{i\in\mathcal{I}_k} n_i}\sum_{i\in\mathcal{I}_k}\sum_{j=1}^{n_i}\mathbf{X}_{ij}\beta_{0,k} + \frac{1}{\sum_{i\in\mathcal{I}_k} n_i}\sum_{i\in\mathcal{I}_k}\sum_{j=1}^{n_i}p_{ijk} * \delta_{0,k}.
\end{aligned}
$$

Thus,

$$
\begin{aligned}
&E_{sp}[Y_{ij}(1, D_{ij}) - Y_{ij}(0, D_{ij})|S_i = k]) \\
&= \frac{1}{\sum_{i\in\mathcal{I}_k} n_i}\sum_{i\in\mathcal{I}_k}\sum_{j=1}^{n_i}p_{ijk} * (\mathbf{X}_{ij}\beta_{1,k} + \delta_k)
\end{aligned}
$$

A2.4.4 $\mathbf{CACE}_k$

$P_k(\mathbf{X}_{ij}) = \hat{F}(\mathbf{X}_{ij})$

If we assume $P_k(\mathbf{X}_{ij}) = \hat{F}(\mathbf{X}_{ij})$, we have,

$$\begin{aligned}
E_{sp}[Y_{ij}(1, D_{ij})|D_{ij} = 1, S_i = k] &= E_{sp}\left[\mu^Y_{S_i} + \mathbf{X}_{ij}\beta_{0,S_i} + D_{ij} * (\mathbf{X}_{ij}\beta_{1,S_i} + \delta_{1,S_i})|D_{ij} = 1, S_i = k\right] \\
&= \mu^Y_k + E_{sp}\left[\mathbf{X}_{ij}|D_{ij} = 1\right](\beta_{0,k} + \beta_{1,k}) + \delta_{1,k} \\
&= \mu^Y_k + \frac{\frac{1}{n}\sum_{ij}\mathbf{X}_{ij}p_{ijk}}{\frac{1}{n}\sum_{ij}p_{ijk}}(\beta_{0,k} + \beta_{1,k}) + \delta_{1,k}.
\end{aligned}$$

Similarly,

$$E_{sp}[Y_{ij}(0, D_{ij})|D_{ij} = 1, S_i = k] = \mu^Y_k + \frac{\frac{1}{n}\sum_{ij}\mathbf{X}_{ij}p_{ijk}}{\frac{1}{n}\sum_{ij}p_{ijk}}(\beta_{0,k}) + \delta_{0,k}.$$

Thus,

$$E_{sp}[Y_{ij}(1, D_{ij}) - Y_{ij}(0, D_{ij})|D_{ij} = 1, S_i = k] = \mu^Y_k + \frac{\frac{1}{n}\sum_{ij}\mathbf{X}_{ij}\beta_{1,k}p_{ijk}}{\frac{1}{n}\sum_{ij}p_{ijk}} + \delta_k.$$

$P_k(\mathbf{X}_{ij}) = \hat{F}_k(\mathbf{X}_{ij})$

If we assume $P_k(\mathbf{X}_{ij}) = \hat{F}_k(\mathbf{X}_{ij})$, we have,

$$\begin{aligned}
&E_{sp}[Y_{ij}(1, D_{ij})|S_i = k, D_{ij} = 1] \\
&= E_{sp}\left[\mu^Y_{S_i} + \mathbf{X}_{ij}\beta_{0,S_i} + D_{ij} * (\mathbf{X}_{ij}\beta_{1,S_i} + \delta_{1,S_i}) + \phi^Y_i|S_i = k, D_{ij} = 1\right] \\
&= \mu^Y_k + E_{sp}\left[\mathbf{X}_{ij}|S_i = k, D_{ij} = 1\right]\beta_{0,k} + E_{sp}\left[(\mathbf{X}_{ij}\beta_{1,S_i} + \delta_{1,S_i})|S_i = k, D_{ij} = 1\right] \\
&= \mu^Y_k + E_{sp}\left[\mathbf{X}_{ij}|S_i = k, D_{ij} = 1\right](\beta_{0,k} + \beta_{1,k}) + \delta_{1,k} \\
\text{From previous derivation} &= \mu^Y_k + \frac{\frac{1}{\sum_{i\in\mathcal{I}_k} n_i}\sum_{i\in\mathcal{I}_k}\sum_{j=1}^{n_i} p_{ijk}\mathbf{X}_{ij}(\beta_{0,k} + \beta_{1,k})}{\frac{1}{\sum_{i\in\mathcal{I}_k} n_i}\sum_{i\in\mathcal{I}_k}\sum_{j=1}^{n_i} p_{ijk}} + \delta_{1,k}.
\end{aligned}$$

Similarly,

$$E_{sp}[Y_{ij}(0, D_{ij})|S_i = k, D_{ij} = 1] = \mu^Y_k + \frac{\frac{1}{\sum_{i\in\mathcal{I}_k} n_i}\sum_{i\in\mathcal{I}_k}\sum_{j=1}^{n_i} p_{ijk}\mathbf{X}_{ij}\beta_{0,k}}{\frac{1}{\sum_{i\in\mathcal{I}_k} n_i}\sum_{i\in\mathcal{I}_k}\sum_{j=1}^{n_i} p_{ijk}} + \delta_{0,k}.$$

Thus,

$$E_{sp}[Y_{ij}(1, D_{ij}) - Y_{ij}(0, D_{ij})|S_i = k, D_{ij} = 1] = \frac{\frac{1}{\sum_{i \in \mathcal{I}_k} n_i} \sum_{i \in \mathcal{I}_k} \sum_{j=1}^{n_i} p_{ijk} \mathbf{X}_{ij} \beta_{1,k}}{\frac{1}{\sum_{i \in \mathcal{I}_k} n_i} \sum_{i \in \mathcal{I}_k} \sum_{j=1}^{n_i} p_{ijk}} + \delta_k.$$

A2.5 Prior Distributions for Simulated Case Studies

$$S_i \sim \text{Cat}(K = 2, \pi) \quad \pi \sim \text{Dirichilet}(K = 2, \lambda = [5, 5])$$

$$\boldsymbol{\mu}_k^S \sim \text{MVN}\left(m_{\boldsymbol{\mu}^S} = [0, 0], V_{\boldsymbol{\mu}^S} = 100 * \mathbf{1}_{2\times 2}\right) \text{ for } k = 1, .., K$$

$$\Sigma \sim \text{Inv-Wishart}\left(\Psi = \frac{1}{100} * \mathbf{1}_{2\times 2}, \nu_\Sigma = 5\right)$$

$$\mu_k^D \sim N\left(0, v^2_{\mu^D} = 100\right) \quad \alpha_k \sim \text{MVN}\left(0, 1000 * \mathbf{I^{fp}}_{2\times 2}\right) \text{ for } k = 1, .., K$$

$$\phi_{i,k}^D \sim N(\mu_k^D, \tau^2_{D,k}) \quad \tau_{D,k} \sim \text{Uniform}\left(0, \sqrt{25}\right) \text{ for } k = 1, .., K$$

$$\mu_k^Y \sim N(0, v^2_{\mu^Y} = 1000), \ \delta_{1.k} \sim N(0, v^2_\beta = 1000) \text{ for } k = 1, ..., K$$

$$\beta_{0,k} \sim \text{MVN}(0, 1000 * \mathbf{I^{fp}}_{2\times 2}), \ \beta_{1,k} \sim \text{MVN}(0, 1000 * \mathbf{I^{fp}}_{2\times 2}) \text{ for } k = 1, .., K$$

$$\phi_i^Y \sim N(0, \tau^2_{Y_k})$$

$$\tau_{Y,k} \sim \text{Uniform}\left(0, \sqrt{625}\right), \ \sigma_k^2 \sim \text{Inv-Gamma}(a = 1, b = 1) \text{ for } k = 1, .., K.$$

A2.6 Supplementary Tables for Data Analysis

Baseline Variable	Total $N = 941$		Treatment $N_1 = 450$		Control $N_0 = 491$	
Age, y, mean (SD)	80.5	(12.1)	80.2	(11.8)	80.8	(12.3)
CMAI, mean (SD)	50.4	(19.4)	51.0	(20.7)	49.9	(18.1)
ADL, mean (SD)	17.0	(6.1)	17.17	(5.9)	16.3	(6.2)
CFS, mean (SD)	2.7	(0.9)	2.8	(0.9)	2.7	(0.9)
ARBS Binary, n(%)	193	(20.5)	105	(23.3)	88	(17.9)
Antipsychotics, n(%)	278	(29.5)	112	(24.9)	166	(33.8)
Male, n(%)	654	(69.5)	306	(68.0)	348	(70.9)
Depression, n(%)	535	(56.8)	267	(56.8)	268	(56.9)
White, n (%)	689	(73.2)	327	(72.7)	362	(73.7)

ADL: Activities of Daily Living Scale; CFS: Cognitive Function Scale; ARBS Binary: Binarized Agitatated and Reactive Behavior Scale

Table A2.2: Mean and standard deviation of patient level characteristics derived from the MDS

Metric/Characteristic	Intervention		Control	
	Mean	SD	Mean	SD
Clinical Relevance[†]	46.8	26.5	-	-
Average Daily Census	62.0	24.5	63.6	25.6

[†] These percentages are modeled on the logit scale, and transformed back to percentages for interpretation

[*] Modeled using a normal distribution truncated at 100%

Table A2.3: Mean and standard deviation of facility characteristics and facility compliance metrics used in each model pair

A2.7 Model Diagnostics for Data Analysis

The Gelman-Rubin statistic for each estimand is shown in Table A2.4.

Outcome	ITT	ITT_1	ITT_2	$\text{ITT}_1 - \text{ITT}_2$	CACE	CACE_1	CACE_2	$\text{CACE}_1 - \text{CACE}_2$
CMAI	1.019	1.015	1.007	1.003	1.018	1.014	1.006	1.003
Antipsychotics	1.008	1.023	1.004	1.013	1.014	1.007	1.019	1.013

Table A2.4: Gelman-Rubin statistic for the 3 chains of posterior samples for each estimand presented in the data analysis in Section 2.5.

Each Gelman-Rubin statistic is close to 1, and indicates adequate model fit. Below, we plot the three MCMC chains for the estimands that were estimates for the CMAI outcome.

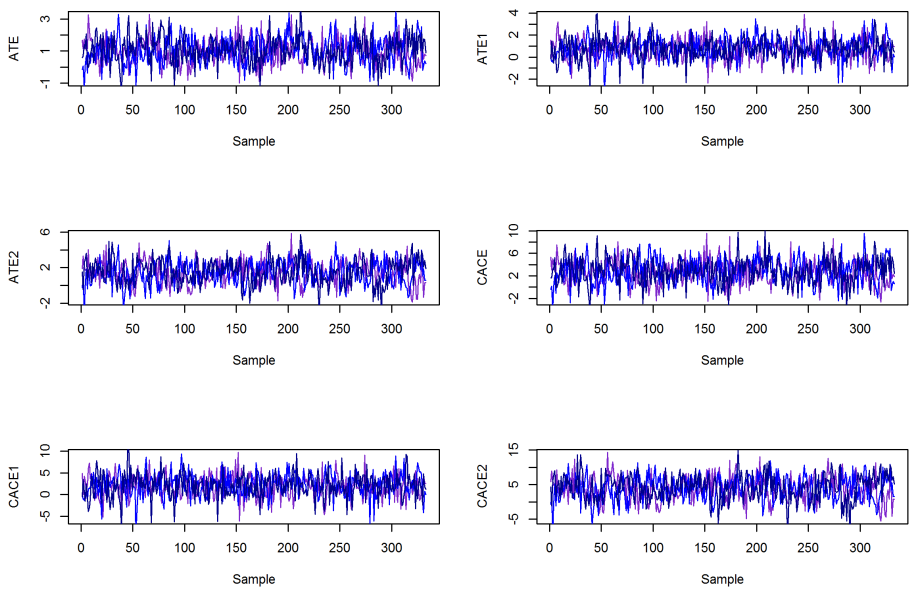

Figure A2.1: Overlaid line plots of the 3 MCMC chains of the samples used for posterior inference in Section 2.5.1

.

Chapter 3 Appendix

A3.1 Additional Details on Assumption 4

Assumption 4 requires that conditional on θ_{D_i} and $\mathbf{X}_{ij}$, compliers ($D_{ij} = 1$) and noncompliers ($D_{ij} = 0$) have the same potential outcome distribution under the control condition. That is,

$$P(Y_{ij}(0, D_{ij})|\mathbf{X}_{ij}, \theta_{D_i}, D_{ij} = 1) = P(Y_{ij}(0, D_{ij})|\mathbf{X}_{ij}, \theta_{D_i}, D_{ij} = 0),$$

and therefore, $E[Y(0, D_{ij})|\mathbf{X}_{ij}, \theta_{D_i}, D_{ij} = 1] = E[Y(0, D_{ij})|\mathbf{X}_{ij}, \theta_{D_i}, D_{ij} = 0]$. However, in randomized experiments with one-sided noncompliance, it is common to assume that the marginal distributions of the control potential outcomes for compliers and noncompliers have different expected values (Imbens and Rubin, 2016). Formally, $E_{sp}\left[Y_{ij}(0, D_{ij})|D_{ij} = 1\right] \neq E_{sp}\left[Y_{ij}(0, D_{ij})|D_{ij} = 0\right]$. Below, we show that, in general, Assumption 4 does not contradict the common assumption about the difference in the marginal distributions.

Under the models described in Section 3.2.4 we have,

$$\begin{aligned} E_{sp}\left[Y_{ij}(0, D_{ij})|D_{ij}\right] &= E_{sp}\left[E_{sp}\left[Y_{ij}(0, D_{ij})|\mathbf{X}_{ij}, \theta_{D_i}\right]|D_{ij}\right] \\ &= E_{sp}\left[\mu^Y + \mathbf{X}_{ij}\beta_0|D_{ij}\right] \\ &= \mu^Y + E_{sp}\left[\mathbf{X}_{ij}|D_{ij}\right]\beta_0. \end{aligned}$$

Therefore, we havethat

$$E_{sp}\left[Y_{ij}(0, D_{ij})|D_{ij} = 1\right] - E_{sp}\left[Y_{ij}(0, D_{ij})|D_{ij} = 0\right] = \left(E_{sp}\left[\mathbf{X}_{ij}|D_{ij} = 1\right] - E_{sp}\left[\mathbf{X}_{ij}|D_{ij} = 0\right]\right)\beta_0. \tag{3.30}$$

We can compute the conditional expectation as follows,

$$\begin{aligned} E_{sp}\left[\mathbf{X}_{ij}|D_{ij}=D\right] &= \int_{\mathbf{X}_{ij}} \mathbf{X}_{ij} P(\mathbf{X}_{ij}|D_{ij}=D) d\mathbf{X}_{ij} \\ &= \int_{\mathbf{X}_{ij}} \mathbf{X}_{ij} \frac{P(D_{ij}=D|\mathbf{X}_{ij})P(\mathbf{X}_{ij})}{P(D_{ij}=D)} d\mathbf{X}_{ij} \\ &= \frac{1}{P(D_{ij}=D)} \int_{\mathbf{X}_{ij}} \mathbf{X}_{ij} P(D_{ij}=D|\mathbf{X}_{ij})P(\mathbf{X}_{ij}) d\mathbf{X}_{ij}. \end{aligned}$$

When $D=1$ we have,

$$\begin{aligned} \int_{\mathbf{X}_{ij}} \mathbf{X}_{ij} P(D_{ij}=1|\mathbf{X}_{ij})P(\mathbf{X}_{ij}) d\mathbf{X}_{ij} &= \int_{\mathbf{X}_{ij}} \mathbf{X}_{ij} E_{sp}[D_{ij}|\mathbf{X}_{ij}]P(\mathbf{X}_{ij}) d\mathbf{X}_{ij} \\ &= \int_{\mathbf{X}_{ij}} \mathbf{X}_{ij} \text{probit}^{-1}\left(\mu_i^D + \mathbf{X}_{ij}\eta_i\right) P(\mathbf{X}_{ij}) d\mathbf{X}_{ij} \\ &= \frac{1}{N_i} \sum_{j=1}^{N_i} \text{probit}^{-1}\left(\mu_i^D + \mathbf{X}_{ij}\eta_i\right) \mathbf{X}_{ij}, \end{aligned}$$

and,

$$\begin{aligned} P(D_{ij}=1) &= E_{sp}\left[E_{sp}\left[D_{ij}|\mathbf{X}_{ij}\right]\right] \\ &= \frac{1}{N_i} \sum_{j=1}^{N_i} \text{probit}^{-1}\left(\mu_i^D + \mathbf{X}_{ij}\eta_i\right), \end{aligned}$$

so,

$$E_{sp}\left[\mathbf{X}_{ij}|D_{ij}=1\right] = \frac{\frac{1}{N_i}\sum_{j=1}^{N_i} \text{probit}^{-1}\left(\mu_i^D + \mathbf{X}_{ij}\eta_i\right)\mathbf{X}_{ij}}{\frac{1}{N_i}\sum_{j=1}^{N_i} \text{probit}^{-1}\left(\mu_i^D + \mathbf{X}_{ij}\eta_i\right)}. \tag{3.31}$$

Following a similar procedure we obtain

$$E_{sp}\left[\mathbf{X}_{ij}|D_{ij}=0\right] = \frac{\frac{1}{N_i}\sum_{j=1}^{N_i} \left(1-\text{probit}^{-1}\left(\mu_i^D + \mathbf{X}_{ij}\eta_i\right)\right)\mathbf{X}_{ij}}{\frac{1}{N_i}\sum_{j=1}^{N_i} \left(1-\text{probit}^{-1}\left(\mu_i^D + \mathbf{X}_{ij}\eta_i\right)\right)}. \tag{3.32}$$

Plugging (3.31) and (3.32) into (3.30), we have

$$\begin{aligned} &E_{sp}\left[Y_{ij}(0,D_{ij})|D_{ij}=1\right] - E_{sp}\left[Y_{ij}(0,D_{ij})|D_{ij}=0\right] = \\ &\left(\frac{\frac{1}{N_i}\sum_{j=1}^{N_i} \text{probit}^{-1}\left(\mu_i^D + \mathbf{X}_{ij}\eta_i\right)\mathbf{X}_{ij}}{\frac{1}{N_i}\sum_{j=1}^{N_i} \text{probit}^{-1}\left(\mu_i^D + \mathbf{X}_{ij}\eta_i\right)} - \frac{\frac{1}{N_i}\sum_{j=1}^{N_i} \left(1-\text{probit}^{-1}\left(\mu_i^D + \mathbf{X}_{ij}\eta_i\right)\right)\mathbf{X}_{ij}}{\frac{1}{N_i}\sum_{j=1}^{N_i} \left(1-\text{probit}^{-1}\left(\mu_i^D + \mathbf{X}_{ij}\eta_i\right)\right)}\right)\beta_0. \end{aligned} \tag{3.33}$$

In general, Equation (3.33) does not equal 0 unless β_0 is exactly 0 or μ_i^D and α_i are exactly 0. This only occurs when the observed covariates are completely uncorrelated with the potential outcomes or the observed covariates are completely uncorrelated with compliance. Because this is not likely in practice, Assumption 4 is not likely to contradict the assumption that the marginal expectation of the control potential outcomes are different between compliers and noncompiers.

Assumption 4 would be violated if compliers and noncompliers had different expected values for their potential outcomes under the control condition after conditioning on the observed covariates. That is,

$$E_{sp}\left[Y_{ij}(0, D_{ij})|\mathbf{X}_{ij}, D_{ij} = 1\right] - E_{sp}\left[Y_{ij}(0, D_{ij})|\mathbf{X}_{ij}, D_{ij} = 0\right] = \delta_{0i}, \tag{3.34}$$

where δ_{0i} is indexed by stage i because being a complier to the standard and augmented interventions may not imply the same value for this parameter. If Equation (3.34) holds,

$$Y_{ij}(0, D_{ij}) \sim N(\mu^Y + \mathbf{X}_{ij}\beta_0 + D_{ij}\delta_{i0}, \sigma^2). \tag{3.35}$$

However, in the two-stage PRCT design, we assume the control condition is the same in stages 1 and 2, so the conditional distributions for the control potential outcome should also be the same in stages 1 and 2. Thus, $E[Y_{1j}(0, D_{1j})|\mathbf{X}_{1j} = \mathbf{X}] = E[Y_{2j}(0, D_{2j})|\mathbf{X}_{2j} = \mathbf{X}]$. For this equality to hold under the model described in Equation (3.35), the following equalities must also hold:

$$\begin{aligned}
E_{sp}[Y_{1j}(0, D_{1j})|\mathbf{X}_{1j} = \mathbf{X}] &= E_{sp}[Y_{1j}(0, D_{2j})|\mathbf{X}_{2j} = \mathbf{X}] \\
E_{sp}\left[E_{sp}\left[Y_{1j}(0, D_{1j})|D_{1j}, \mathbf{X}_{1j}\right]|\mathbf{X}_{1j} = \mathbf{X}\right] &= E_{sp}\left[E_{sp}\left[Y_{2j}(0, D_{2j})|D_{2j}, \mathbf{X}_{2j}\right]|\mathbf{X}_{2j} = \mathbf{X}\right] \\
E_{sp}\left[\mu_0^Y + \mathbf{X}_{1j}\beta_0 + D_{1j}\delta_{10}|\mathbf{X}_{1j} = \mathbf{X}\right] &= E_{sp}\left[\mu_0^Y + \mathbf{X}_{2j}\beta_0 + D_{2j}\delta_{20}|\mathbf{X}_{2j} = \mathbf{X}\right] \\
\mu_0^Y + \mathbf{X}\beta_0 + E_{sp}\left[D_{1j}|\mathbf{X}_{1j} = \mathbf{X}\right]\delta_{10} &= \mu_0^Y + \mathbf{X}\beta_0 + E_{sp}\left[D_{2j}|\mathbf{X}_{2j} = \mathbf{X}\right]\delta_{20} \\
\delta_{10}E_{sp}\left[D_{1j}|\mathbf{X}_{1j} = \mathbf{X}\right] &= \delta_{20}E_{sp}\left[D_{2j}|\mathbf{X}_{2j} = \mathbf{X}\right] \\
\delta_{10}\text{probit}^{-1}(\mu_1^D + \eta_1\mathbf{X}) &= \delta_{20}\text{probit}^{-1}(\mu_2^D + \eta_2\mathbf{X}_{2j}) \\
\delta_{20} &= \delta_{10}\frac{\text{probit}^{-1}(\mu_1^D + \eta_1\mathbf{X})}{\text{probit}^{-1}(\mu_2^D + \eta_2\mathbf{X})} \text{ for all } \mathbf{X}.
\end{aligned} \tag{3.36}$$

Individuals enrolled in Stage 1 and Stage 2 are sampled from the same target population ($P(\mathbf{X}_{1j}) = P(\mathbf{X}_{2j}) = P(\mathbf{X})$). Because the control condition is the same in Stage 1 and Stage 2, this implies

that the marginal distribution of the control potential outcomes should the same in both stages. Thus, $E[Y_{1j}(0, D_{1j})] = E[Y_{2j}(0, D_{2j})]$, and the following set of equalities must hold:

$$
\begin{aligned}
E_{sp}[Y_{1j}(0, D_{1j})] &= E_{sp}[Y_{2j}(0, D_{2j})] \\
E_{sp}\left[E_{sp}\left[Y_{1j}(0, D_{1j})|D_{1j}, \mathbf{X}_{1j}\right]\right] &= E_{sp}\left[E_{sp}\left[Y_{2j}(0, D_{2j})|D_{2j}, \mathbf{X}_{2j}\right]\right] \\
E_{sp}\left[\mu_0^Y + \mathbf{X}_{1j}\beta_0 + D_{1j}\delta_{10}\right] &= E_{sp}\left[\mu_0^Y + \mathbf{X}_{2j}\beta_0 + D_{2j}\delta_{20}\right] \\
\mu_0^Y + E_{sp}\left[\mathbf{X}_{1j}\right]\beta_0 + E_{sp}\left[D_{1j}\right]\delta_{10} &= \mu_0^Y + E_{sp}\left[\mathbf{X}_{2j}\right]\beta_0 + E_{sp}\left[D_{2j}\right]\delta_{20} \\
\mu_0^Y + \mathbf{X}_{\mu}\beta_0 + E_{sp}\left[D_{1j}\right]\delta_{10} &= \mu_0^Y + \mathbf{X}_{\mu}\beta_0 + E_{sp}\left[D_{2j}\right]\delta_{20} \\
\delta_{10}E_{sp}\left[D_{1j}\right] &= \delta_{20}E_{sp}\left[D_{2j}\right] \\
\delta_{10}\int \text{probit}^{-1}(\mu_1^D + \mathbf{X}_{1j}\eta_1)P(\mathbf{X}_{1j})d\mathbf{X}_{1j} &= \delta_{20}\int \text{probit}^{-1}(\mu_2^D + \mathbf{X}_{2j}\eta_2)P(\mathbf{X}_{2j})d\mathbf{X}_{2j} \\
\delta_{10}\int \text{probit}^{-1}(\mu_1^D + \mathbf{X}\eta_1)P(\mathbf{X})d\mathbf{X} &= \delta_{20}\int \text{probit}^{-1}(\mu_2^D + \mathbf{X}\eta_2)P(\mathbf{X})d\mathbf{X} \\
\delta_{20} &= \delta_{10}\frac{\int \text{probit}^{-1}(\mu_1^D + \mathbf{X}\eta_1)P(\mathbf{X})d\mathbf{X}}{\int \text{probit}^{-1}(\mu_2^D + \mathbf{X}\eta_2)P(\mathbf{X})d\mathbf{X}}.
\end{aligned}
\tag{3.37}
$$

The equalities in (3.36) and (3.37) are only compatible under specific conditions. One of these conditions, when $\mu_1^D = \mu_2^D$ and $\eta_1 = \eta_2$, is not true in general because the probability that an individual complies to the standard intervention and the augmented intervention are different. A second condition, when $P(\mathbf{X})$ is a degenerate distribution at only one set of covariate values, does not occur in PRCTs. Another condition, when $\eta_1 = 0$ and $\eta_2 = 0$, requires that the observed covariates are uncorrelated with compliance status. This is not likely in practice because the covariates recorded in PRCTs are expected to be related to the potential outcomes and compliance. However, if this condition were to occur, we could rearrange (3.36) and (3.37) so that

$$
\frac{\delta_{20}}{\delta_{10}} = \frac{\text{probit}^{-1}(\mu_1^D)}{\text{probit}^{-1}(\mu_2^D)}.
$$

This equality implies that the ratio of the differences in expected control potential outcomes in Stage 2 and Stage 1 is equal to the ratio of expected compliance probabilities in Stage 1 and Stage 2. In general, these two quantities need not be related. The last condition that enables both of the equalities in Equations (3.36) and (3.37) to hold is when $\delta_{10} = \delta_{20} = 0$, which is equivalent to Assumption 4. This assumption is reasonable if all relevant covariates are controlled for. We

expect any differences in the conditional distribution of control potential outcomes for compliers and noncompliers to be negligible after accounting for the relevant covariates because no intervention is being received under the control condition.

A3.2 Estimand Derivation

$$
\begin{aligned}
\text{ITT}_i &= E_{sp}[Y_{ij}(1, D_{1j}) - Y_{ij}(0, D_{1j})] \\
&= E_{sp}[E_{sp}[Y_{ij}(1, D_{ij}) - Y_{1j}(0, D_{1j})|\mathbf{X}_{ij}, \mathbf{D}_{ij}, \theta_Y, \theta_{D_i}]] \\
&= E_{sp}\left[D_{ij} * (\delta_1 + \mathbf{X}_{ij}\beta_1)\right] \\
&= E_{sp}\left[E_{sp}\left[D_{ij} * (\delta_1 + \mathbf{X}_{ij}\beta_1)\right|\mathbf{X}_{ij}]\right] \\
&= E_{sp}\left[E_{sp}\left[D_{1j}|\mathbf{X}_{ij}\right] * (\delta_1 + \mathbf{X}_{ij}\beta_1)\right] \\
&= E_{sp}\left[\text{probit}^{-1}\left(\mu_i^D + \mathbf{X}_{ij}\eta_i\right) * (\delta_1 + \mathbf{X}_{ij}\beta_1)\right] \\
&= \frac{1}{N_i}\sum_{j=1}^{N_i} \text{probit}^{-1}\left(\mu_i^D + \mathbf{X}_{ij}\eta_i\right) * (\delta_1 + \mathbf{X}_{ij}\beta_1)
\end{aligned}
$$

A3.3 Methods for Determining a_0 in the Two-Stage PRCT Design

We describe the Empirical Bayes and DIC approach to setting the borrowing parameter, a_0 for the proposed two-stage PRCT design. We assume $P_0(\theta_Y) \propto \frac{1}{\sigma^2}$. For more general form of the initial prior distribution, see Han, Ye and Wang (2023).

A3.3.1 Empirical Bayes

To implement the Empirical Bayes approach (Gravestock and Held, 2017), we must first derive the marginal liklihood of a_0 under the partial borrowing power prior described in Equation (3.13). We

have,

$$
\begin{aligned}
L(a_0|\mathbf{O}_1,\mathbf{O}_2) &= \int\int L_2(\theta_Y,\theta_{D_2}|\mathbf{O}_2)P(\theta_Y,\theta_{D_2}|\mathbf{O}_1,\mathbf{O}_2,a_0)d\theta_{D_2}d\theta_Y \\
&\propto \frac{\int\int L_2(\theta_Y,\theta_{D_2}|\mathbf{O}_2)\left\{\int_{\theta_{D_1}} L_1(\theta_Y,\theta_{D_1}|\mathbf{O}_1)^{a_0}P_0(\theta_{D_1})d\theta_{D_1}\right\}P_0(\theta_Y)P_0(\theta_{D_2})d\theta_{D_2}d\theta_Y}{\int_{\theta_Y}\left\{\int_{\theta_{D_1}} L_1(\theta_Y,\theta_{D_1}|\mathbf{O}_1)^{a_0}P_0(\theta_{D_1})d\theta_{D_1}\right\}P_0(\theta_Y)d\theta_Y} \\
&= \frac{\int\int L_2(\theta_Y,\theta_{D_2}|\mathbf{O}_2)L_1(\theta_Y|\mathbf{O}_1)^{a_0}\left(\int L_1(\theta_{D_1}|\mathbf{O}_1)^{a_0}P_0(\theta_{D_1})d\theta_{D_1}\right)P_0(\theta_Y)P_0(\theta_{D_2})d\theta_{D_2}d\theta_Y}{\int_{\theta_Y} L_1(\theta_Y|\mathbf{O}_1)^{a_0}\left(\int L_1(\theta_{D_1}|\mathbf{O}_1)^{a_0}P_0(\theta_{D_1})d\theta_{D_1}\right)P_0(\theta_Y)d\theta_Y} \\
&= \frac{\int\int L_2(\theta_Y,\theta_{D_2}|\mathbf{O}_2)L_1(\theta_Y|\mathbf{O}_1)^{a_0}P_0(\theta_Y)P_0(\theta_{D_2})d\theta_{D_2}d\theta_Y}{\int_{\theta_Y} L_1(\theta_Y|\mathbf{O}_1)^{a_0}P_0(\theta_Y)d\theta_Y} \\
&= \frac{\int\int L_2(\theta_Y|\mathbf{O}_2)L_2(\theta_{D_2}|\mathbf{O}_2)L_1(\theta_Y|\mathbf{O}_1)^{a_0}P_0(\theta_Y)P_0(\theta_{D_2})d\theta_{D_2}d\theta_Y}{\int_{\theta_Y} L_1(\theta_Y|\mathbf{O}_1)^{a_0}P_0(\theta_Y)d\theta_Y} \\
&= \frac{\int L_2(\theta_Y|\mathbf{O}_2)L_1(\theta_Y|\mathbf{O}_1)^{a_0}P_0(\theta_Y)d\theta_Y\int L_2(\theta_{D_2}|\mathbf{O}_2)P_0(\theta_{D_2})d\theta_{D_2}}{\int_{\theta_Y} L_1(\theta_Y|\mathbf{O}_1)^{a_0}P_0(\theta_Y)d\theta_Y} \\
&\propto \frac{\int L_2(\theta_Y|\mathbf{O}_2)L_1(\theta_Y|\mathbf{O}_1)^{a_0}P_0(\theta_Y)d\theta_Y}{\int_{\theta_Y} L_1(\theta_Y|\mathbf{O}_1)^{a_0}P_0(\theta_Y)d\theta_Y},
\end{aligned} \tag{3.38}
$$

where $\int L_1(\theta_{D_1}|\mathbf{O}_1)^{a_0}P_0(\theta_{D_1})d\theta_{D_1}$ is finite whenever $P_0(\theta_{D_1})$ is proper. Because θ_Y comprises the parameters of a Normal linear regression, we can follow the derivation of Han, Ye and Wang (2023) to obtain a closed form for $L(a_0|\mathbf{O}_1,\mathbf{O}_2)$. Let

$$
X^*_{ij} = \left[1, X_{ij}, D^*_{ij}\right],\ D^*_{ij} = \begin{cases} D_{ij} & W_{ij}=1 \\ 0 & W_{ij}=0 \end{cases},\ \mathbf{X}^*_i = \begin{bmatrix} X^*_{i1} \\ \vdots \\ X_{iN^*_i} \end{bmatrix},\ \text{and } \beta^* = \left[\mu^Y, \beta_0, \delta_1\right].
$$

The marginal liklihood of a_0 is given by,

$$
L(a_0|\mathbf{O}_1,\mathbf{O}_2) \propto \frac{\Gamma(\nu)|a_0\mathbf{X}^{*t}_1\mathbf{X}^*_1|^{1/2}H_1(a_0)^{\nu_0}}{\Gamma(\nu_0)|\mathbf{X}^{*t}_2\mathbf{X}^*_2 + a_0\mathbf{X}^{*t}_1\mathbf{X}^*_1|^{1/2}H_2(a_0)^{\nu}}, \tag{3.39}
$$

where

$$
\begin{aligned}
H_1(a_0) &= \frac{a_0||\mathbf{Y}^{obs}_1 - \mathbf{X}^*_1\hat{\beta}^*_1||^2_2}{2} \\
H_2(a_0) &= H_1(a_0) + \frac{||\mathbf{Y}^{obs}_2 - \mathbf{X}^*_2\hat{\beta}^*_2||^2_2 + \left(\hat{\beta}^*_1 - \hat{\beta}^*_2\right)^t \mathbf{X}^{*t}_2\mathbf{X}^*_2\left(\mathbf{X}^{*t}_2\mathbf{X}^*_2 + a_0\mathbf{X}^{*t}_1\mathbf{X}^*_1\right)^{-1}\left(a_0\mathbf{X}^{*t}_1\mathbf{X}^*_1\right)\left(\hat{\beta}^*_1 - \hat{\beta}^*_2\right)}{2},
\end{aligned}
$$

and

$$\hat{\beta}_i^* = \left(\mathbf{X}_i^{*t}\mathbf{X}_i^*\right)^{-1}\mathbf{X}_i^{*t}\mathbf{Y}_i^{obs}$$
$$\nu_0 = \frac{a_0 N_1 - 2}{2}$$
$$\nu = \frac{N_2 + a_0 N_1 - 2}{2}.$$

Han, Ye and Wang (2023) show that if X_1^* is an $N_1 \times P^*$ dimension matrix, the integral in the denominator of (3.38) is finite only for $a_0 \in (\frac{P^*}{N_1}, 1]$. Therefore, the Empirical Bayes approach sets $a_0^{\text{EB}} = \underset{a_0 \in (\frac{P^*}{N_1}, 1]}{\text{argmax}}\ L(a_0|\mathbf{O}_1, \mathbf{O}_2)$.

A3.3.2 DIC

We again follow Han, Ye and Wang (2023) to derive a DIC based objective function for setting a_0. The DIC for a given value of a_0 assuming the models defined in Section 3.2.4 is given by,

$$\begin{aligned}\text{DIC}(a_0|\mathbf{O}_1, \mathbf{O}_2) = N_2 &\left\{\log(\nu - 1) + \log(H_2(a_0)) - 2\psi(\nu)\right\} \\ &+ \frac{\nu + 1}{H_2(a_0)}\left\{\left(\dot{\beta}^* - \hat{\beta}_2^*\right)^t \mathbf{X}_2^{*t}\mathbf{X}_2^*\left(\dot{\beta}^* - \hat{\beta}_2^*\right) + ||\mathbf{Y}_2^{obs} - \mathbf{X}_2^*\hat{\beta}_2^*||_2^2\right\} \\ &+ 2\text{tr}\left(\mathbf{X}_2^{*t}\mathbf{X}_2^*\left(\mathbf{X}_2^{*t}\mathbf{X}_2^* + a_0\mathbf{X}_1^{*t}\mathbf{X}_1^*\right)^{-1}\right)\end{aligned}$$

where

$$\dot{\beta}^* = \left(\mathbf{X}_2^{*t}\mathbf{X}_2^* + a_0\mathbf{X}_1^{*t}\mathbf{X}_1^*\right)^{-1}\left(\mathbf{X}_2^{*t}\mathbf{Y}_2^{obs} + a_0\mathbf{X}_1^{*t}\mathbf{Y}_1^{obs}\right),$$

$\psi(\cdot)$ is the digamma function, and the rest of the notation follows from Section A3.3.1. The DIC approach sets $a_0^{\text{DIC}} = \underset{a_0}{\text{argmin}}\ \text{DIC}(a_0|\mathbf{O}_1, \mathbf{O}_2)$.

A3.4 Simulation Studies

Table A3.1 depicts the parameters used to define the two-stage design implemented in the simulation studies describes in Section 3.3.

The values of p_U and p_L are set as functions of p_E. We use the O'Brien-Fleming alpha-spending function Demets and Lan (1994) to set values for α_1 and α_2.

Paramter	Value
ES	0.5
α	0.05
$1-\beta$	0.8
$1-\beta_1$	0.8
p_E	$\{0.3, 0.5, 0.7\}$
p_U	$p_E + 0.65 \times (1 - p_E)$
p_L	$0.5 \times p_E$
α_1	$2 - 2 \times \Phi(\phi(1-\alpha/2))/\sqrt{N_1/N})$
α_2	$\alpha - \alpha_1$

Φ:=Gaussian CDF; ϕ:=Gaussian PDF

Table A3.1: Parameter values used to define the two-stage desing implemented in the simulation studies described in Section 3.3.

Table A3.2 depicts the parameters used to define the factors and data generating mechanism for the simulation study in Section 3.3.2. The values of m_1^D and m_2^D are set as functions of π_1 and

Paramter	Value
π_1	$\{0.25, 0.5, 0.75\}$
φ	$\{10\%, 33\%, 50\%\}$
γ_1	$\{0.2, 0.5, 0.8\}$
m_1^D	$\phi(\pi_1)$
ζ_1	$[0.25, -0.25]$
m_2^D	$\phi(\pi_1 + \varphi * (1 - \pi_1))$
$1-\beta$	0.8
ζ_2	$[0.25, -0.25]$
m^y	5
ξ_0	$[0.5, -0.5]$
τ^2	1

ϕ:=Gaussian PDF

Table A3.2: Parameter values used to define the two-stage design implemented in the simulation studies described in Section 3.3.2.

ϕ.

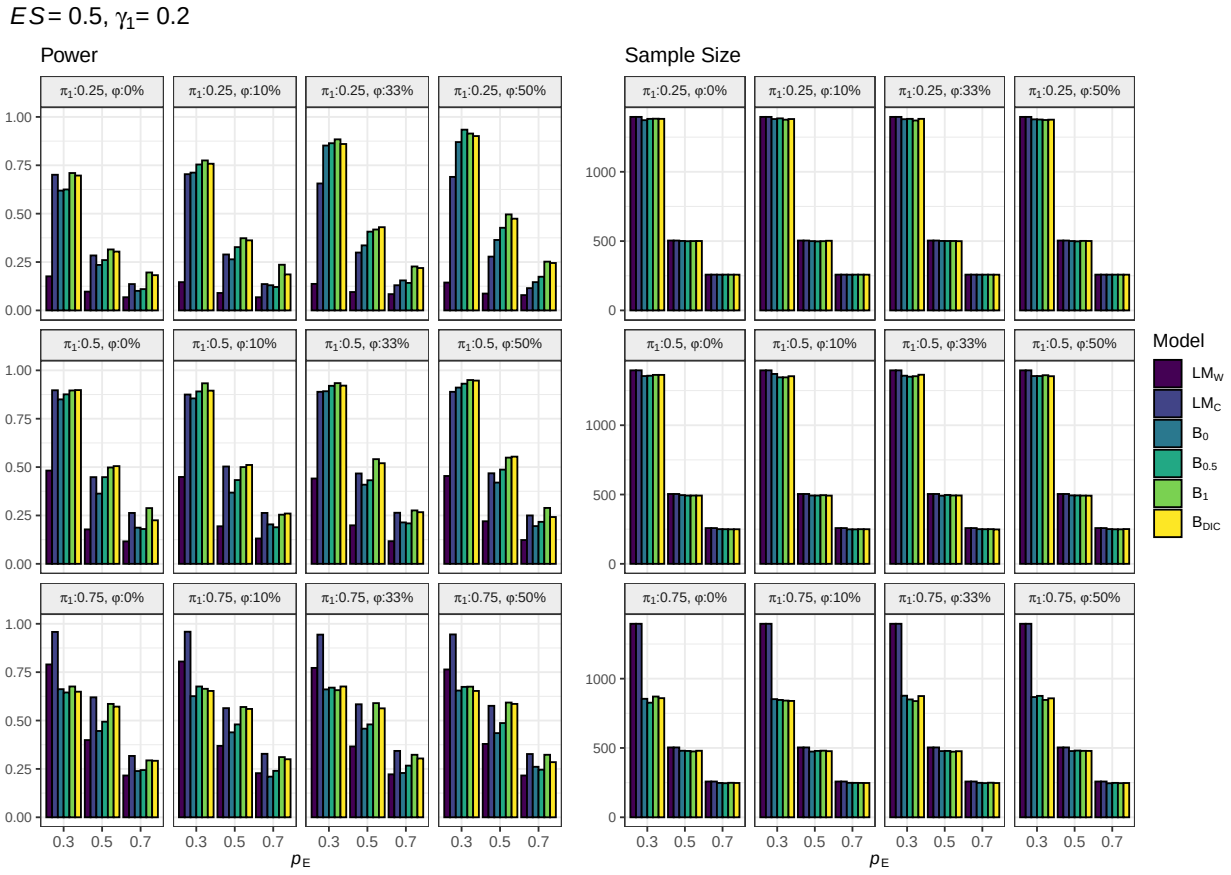

Figure A3.1: Power and average sample size comparison for all configurations when $\gamma_1 = 0.2$.

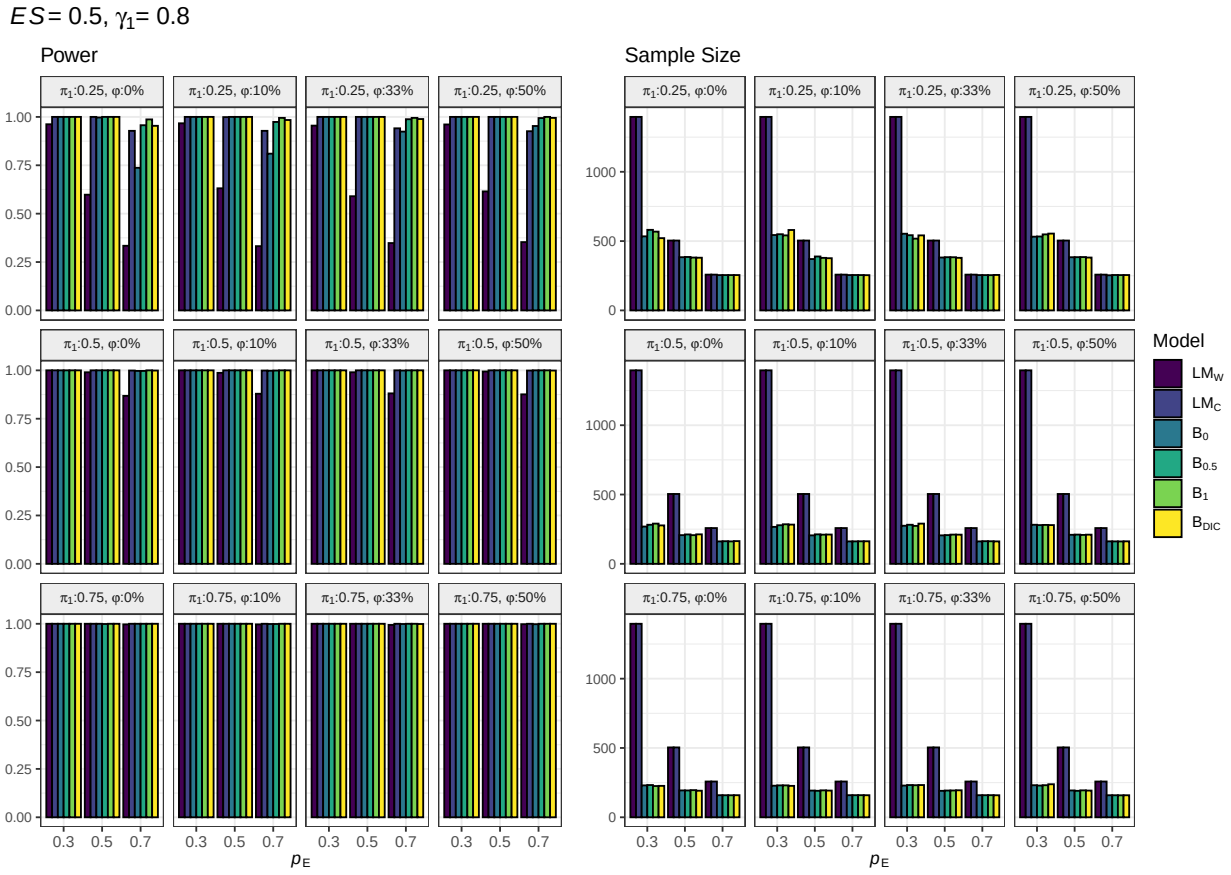

Figure A3.2: Power and average sample size comparison for all configurations when $\gamma_1 = 0.8$.

Design	Model	Type 1 Error			
		$\varphi = 0\%$	$\varphi = 10\%$	$\varphi = 33\%$	$\varphi = 50\%$
Two-Stage Design	B_0	0.05	0.06	0.05	0.05
	$B_{0.5}$	0.04	0.04	0.05	0.05
	B_1	0.05	0.05	0.05	0.05
	B_{DIC}	0.05	0.05	0.05	0.05
Simple Randomized Design	LM_W	0.06	0.06	0.06	0.05
	LM_D	0.05	0.05	0.05	0.05

Table A3.3: Type 1 Error of each design and model combination averaged over different values of φ.

Parameter	Value
ξ_{01}	$\begin{bmatrix} 0.5 & 0.5 \end{bmatrix}^t$
ξ_{02}	$\begin{bmatrix} -0.5 & -0.5 \end{bmatrix}^t$
γ_{11}	$\{0.25, 0.5\}$
γ_{21}	0.5
π_1	0.5
φ	$\{0.2, 0.4\}$
m_1^D	$\phi(\pi_1)$
ζ_1	$[0.33, -0.33]$
m_2^D	$\phi(\pi_1 + \varphi * (1 - \pi_1))$
ζ_2	$[0.33, -0.33]$
m^y	5

ϕ:=Gaussian PDF

Table A3.4: Parameter values used to define the two-stage design implemented in the simulation studies described in Section 3.3.3.

Paramter	Value
m_2^u	$\{0, 0.25, 0.5\}$
κ_D	$\{0, 0.25, 0.33\}$
κ_Y	$\{0, 0.25, 0.5\}$
γ_1	0.5
π_1	0.5
φ	10%
m_1^D	$\phi(\pi_1)$
ζ_1	$[0.33, -0.33]$
m_2^D	$\phi(\pi_1 + \varphi * (1 - \pi_1))$
ζ_2	$[0.33, -0.33]$
m^y	5
ξ_0	$[0.5, 0.5]$

ϕ:=Gaussian PDF

Table A3.5: Parameter values used to define the two-stage design implemented in the simulation studies described in Section 3.3.3.

A3.5 Simulated Trial

Baseline Variable	Stage 1		Stage 2	
	Treatment	Control	Treatment	Control
	$N_{11} = 85$	$N_{10} = 85$	$N_{21} = 80$	$N_{20} = 80$
Age, y, mean (SD)	78.8 (11.2)	78 (13.6)	79.1 (14.1)	82.4 (10.8)
ADL, mean (SD)	17.4 (6.0)	16.6 (5.7)	17.1 (6.5)	17.3 (6.1)
CFS, mean (SD)	2.6 (0.8)	2.6 (0.8)	2.6 (0.9)	2.7 (0.8)
Baseline CMAI, mean (SD)	47.5 (16.6)	49.4 (16.4)	46.2 (14.3)	51.8 (22.2)
Baseline ABS, mean (SD)	0.5 (1.1)	0.5 (1.1)	0.5 (1.3)	0.5 (1.0)
Depression, n(%)	55 (64.7)	54 (63.5)	55 (61.1)	55 (61.1)
Anxiety, n(%)	37 (43.5)	34 (40.0)	40 (44.4)	37 (41.1)
Antidepressant, n(%)	57 (67.1)	55 (64.7)	51 (56.7)	57 (63.3)
Antianxiety, n(%)	15 (17.6)	15 (17.6)	20 (22.2)	20 (22.2)
White, n (%)	59 (69.4)	66 (77.6)	55 (61.1)	66 (73.3)

ADL: Activities of Daily Living Scale; CFS: Cognitive Function Score; CMAI: Cohen-Mansfield Aggitation Inventory; ABS: Aggitated Behavior Scale
ϕ:=Gaussian PDF

Table A3.6: Baseline covariate distribution of individuals randomly sampled for the simulated two-stage trial.

Bibliography

Agresti, Alan. 2010. *Analysis of Ordinal Categorical Data: Agresti/Analysis*. Wiley Series in Probability and Statistics Hoboken, NJ, USA: John Wiley & Sons, Inc.

Albert, James H. and Siddhartha Chib. 1993. "Bayesian Analysis of Binary and Polychotomous Response Data." *Journal of the American Statistical Association* 88(422):669–679.

Albert, Jeffrey M. 1997. "Accounting for Noncompliance in the Design of Clinical Trials." *Drug Information Journal* 31(1):157–165.

Allore, Heather G, Keith S Goldfeld, Roee Gutman, Fan Li, Joan K Monin, Monica Taljaard and Thomas G Travison. 2020. "Statistical considerations for embedded pragmatic clinical trials in people living with dementia." *Journal of the American Geriatrics Society* 68:S68–S73.

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

Apter, Andrea J. 2015. "Understanding Adherence Requires Pragmatic Trials: Lessons From Pediatric Asthma." *JAMA Pediatrics* 169(4):310–311.

Armstrong, Richard A. 2014. "When to use the B onferroni correction." *Ophthalmic and Physiological Optics* 34(5):502–508.

Barnard, J and DB Rubin. 1999. "Miscellanea. Small-sample degrees of freedom with multiple imputation." *Biometrika* 86(4):948–955.

Barnard, John, Constantine E Frangakis, Jennifer L Hill and Donald B Rubin. 2003. "Principal stratification approach to broken randomized experiments: A case study of school choice vouchers in New York City." *Journal of the American Statistical Association* 98(462):299–323.

Bernardo, José M. 1996. "The concept of exchangeability and its applications." *Far East Journal of Mathematical Sciences* 4:111–122.

Berry, Scott M, Bradley P Carlin, J Jack Lee and Peter Muller. 2010. *Bayesian adaptive methods for clinical trials*. CRC press.

Bross, Irwin D. J. 1958. "How to Use Ridit Analysis." *Biometrics* 14(1):18–38.

Brown, Paul M and Justin A Ezekowitz. 2017. "Composite end points in clinical trials of heart failure therapy: how do we measure the effect size?" *Circulation: Heart Failure* 10(1):e003222.

Centers for Disease Control and Prevention. 2023. "Diabetes quick facts.".

Centers for Medicare & Medicaid Services. 2021. "Minimum Data Set (MDS) 3.0 for Nursing Homes and Swing Bed Providers.".

Chakraborty, Saptarshi and Kshitij Khare. 2017. "Convergence properties of Gibbs samplers for Bayesian probit regression with proper priors.".

Chen, Ming-Hui, Joseph G Ibrahim, Donglin Zeng, Kuolung Hu and Catherine Jia. 2014. "Bayesian design of superiority clinical trials for recurrent events data with applications to bleeding and transfusion events in myelodyplastic syndrome." *Biometrics* 70(4):1003–1013.

Chen, Ming-Hui, Joseph G Ibrahim, H Amy Xia, Thomas Liu and Violeta Hennessey. 2014. "Bayesian sequential meta-analysis design in evaluating cardiovascular risk in a new antidiabetic drug development program." *Statistics in medicine* 33(9):1600–1618.

Chen, Qingxia and Joseph G Ibrahim. 2014. "A note on the relationships between multiple imputation, maximum likelihood and fully Bayesian methods for missing responses in linear regression models." *Statistics and its Interface* 6(3):315.

Chow, Shein-Chung, Mark Chang and Annpey Pong. 2005. "Statistical consideration of adaptive methods in clinical development." *Journal of biopharmaceutical statistics* 15(4):575–591.

Cohen-Mansfield, J., M. S. Marx and A. S. Rosenthal. 1989. "A Description of Agitation in a Nursing Home." *Journal of Gerontology* 44(3):M77–M84.

Colantuoni, Elizabeth, Daniel O Scharfstein, Chenguang Wang, Mohamed D Hashem, Andrew Leroux, Dale M Needham and Timothy D Girard. 2018. "Statistical methods to compare functional outcomes in randomized controlled trials with high mortality." *BMJ* 360.

Cordoba, Gloria, Lisa Schwartz, Steven Woloshin and Peter C Bae, Harold Gøtzsche. 2010. "Definition, reporting, and interpretation of composite outcomes in clinical trials: systematic review." *BMJ* 341.

Czobor, P. and P. Skolnick. 2011. "The Secrets of a Successful Clinical Trial: Compliance, Compliance, and Compliance." *Molecular Interventions* 11(2):107–110.

Davison, Anthony Christopher and David Victor Hinkley. 1997. *Bootstrap methods and their application*. Cambridge university press.

de Finetti, Bruno. 1992. "Foresight: Its logical laws, its subjective sources". In *Breakthroughs in statistics*. Springer pp. 134–174.

De Santis, Fulvio. 2006. "Power priors and their use in clinical trials." *The American Statistician* 60(2):122–129.

Demets, David L and KK Gordon Lan. 1994. "Interim analysis: the alpha spending function approach." *Statistics in medicine* 13(13-14):1341–1352.

Denwood, Matthew J. 2016. "runjags: An R Package Providing Interface Utilities, Model Templates, Parallel Computing Methods and Additional Distributions for MCMC Models in JAGS." *Journal of Statistical Software* 71(9):1–25.

Depaoli, Sarah, James P. Clifton and Patrice R. Cobb. 2016. "Just Another Gibbs Sampler (JAGS): Flexible Software for MCMC Implementation." *Journal of Educational and Behavioral Statistics* 41(6):628–649.

Dimairo, Munyaradzi, Jonathan Boote, Steven A. Julious, Jonathan P. Nicholl and Susan Todd. 2015. "Missing steps in a staircase: a qualitative study of the perspectives of key stakeholders on the use of adaptive designs in confirmatory trials." *Trials* 16(1):430.

Duan, Yuyan, Keying Ye and Eric P Smith. 2006. "Evaluating water quality using power priors to incorporate historical information." *Environmetrics: The Official Journal of the International Environmetrics Society* 17(1):95–106.

Ellenberg, Jonas H. 1996. "Intent-to-Treat Analysis versus As-Treated Analysis." *Drug Information Journal* 30(2):535–544.

Erler, Nicole S, Dimitris Rizopoulos, Joost van Rosmalen, Vincent WV Jaddoe, Oscar H Franco and Emmanuel MEH Lesaffre. 2016. "Dealing with missing covariates in epidemiologic studies: a comparison between multiple imputation and a full Bayesian approach." *Statistics in medicine* 35(17):2955–2974.

Evans, Scott R., Daniel Rubin, Dean Follmann, Gene Pennello, W. Charles Huskins, John H. Powers, David Schoenfeld, Christy Chuang-Stein, Sara E. Cosgrove, Jr Fowler, Vance G., Ebbing Lautenbach and Henry F. Chambers. 2015. "Desirability of Outcome Ranking (DOOR) and Response Adjusted for Duration of Antibiotic Risk (RADAR)." *Clinical Infectious Diseases* 61(5):800–806.

Fadini, Gian Paolo, Daniele Bottigliengo, Federica D'Angelo, Franco Cavalot, Antonio Carlo Bossi, Giancarlo Zatti, Ileana Baldi and Angelo Avogaro. 2018. "Comparative effectiveness of DPP-4 inhibitors versus sulfonylurea for the treatment of type 2 diabetes in routine clinical practice: a retrospective multicenter real-world study." *Diabetes Therapy* 9(4):1477–1490.

Fitzpatrick, Claire, Clare Gillies, Samuel Seidu, Debasish Kar, Ekaterini Ioannidou, Melanie J Davies, Prashanth Patel, Pankaj Gupta and Kamlesh Khunti. 2020. "Effect of pragmatic versus explanatory interventions on medication adherence in people with cardiometabolic conditions: a systematic review and meta-analysis." *BMJ open* 10(7):e036575.

Follmann, Dean, Michael P Fay, Toshimitsu Hamasaki and Scott Evans. 2020. "Analysis of ordered composite endpoints." *Statistics in Medicine* 39(5):602–616.

Forastiere, Laura, Fabrizia Mealli and Tyler J VanderWeele. 2016. "Identification and estimation of causal mechanisms in clustered encouragement designs: Disentangling bed nets using Bayesian principal stratification." *Journal of the American Statistical Association* 111(514):510–525.

Ford, Ian and John Norrie. 2016. "Pragmatic Trials." *New England Journal of Medicine* 375(5):454–463.

Frangakis, C. E. 2002. "Clustered Encouragement Designs with Individual Noncompliance: Bayesian Inference with Randomization, and Application to Advance Directive Forms." *Biostatistics* 3(2):147–164.

Frangakis, Constantine E. and Donald B. Rubin. 2002*a*. "Principal Stratification in Causal Inference." *Biometrics* 58(1):21–29.

Frangakis, Constantine E and Donald B Rubin. 2002*b*. "Principal stratification in causal inference." *Biometrics* 58(1):21–29.

Frangakis, Constantine E, Donald B Rubin and Xiao-Hua Zhou. 2002. "Clustered encouragement designs with individual noncompliance: Bayesian inference with randomization, and application to advance directive forms." *Biostatistics* 3(2):147–164.

García-Alberca, José María, Mercedes Florido, Marta Cáceres, Alicia Sánchez-Toro, José Pablo Lara and Natalia García-Casares. 2019. "Medial temporal lobe atrophy is independently associated with behavioural and psychological symptoms in Alzheimer's disease." *Psychogeriatrics* 19(1):46–54.

Gehan, Edmund A. 1961. "The determination of the number of patients required in a preliminary and a follow-up trial of a new chemotherapeutic agent." *Journal of chronic diseases* 13(4):346–353.

Gelfand, Alan E. and F. M. Smith Adrian. 1990. "Sampling-Based Approaches to Calculating Marginal Densities." *Journal of the American Statistical Association* 85(410):398–409.

Gelman, Andrew. 2006. "Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper).".

Gelman, Andrew, Aleks Jakulin, Maria Grazia Pittau and Yu-Sung Su. 2008. "A weakly informative default prior distribution for logistic and other regression models." *The annals of applied statistics* 2(4):1360–1383.

Gelman, Andrew and Donald B Rubin. 1992. "Inference from iterative simulation using multiple sequences." *Statistical science* 7(4):457–472.

Gelman, Andrew, Xiao-Li Meng and Hal Stern. 1996. "Posterior predictive assessment of model fitness via realized discrepancies." *Statistica sinica* pp. 733–760.

Gelman, Andrew and Yu-Sung Su. 2022. *arm: Data Analysis Using Regression and Multilevel/Hierarchical Models*. R package version 1.13-1.

Geman, S. and D. Geman. 1984. "Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images." *IEEE Transactions on Pattern Analysis and Machine Intelligence* PAMI-6(6):721–741.

Glasgow, Russell E., Christopher E. Knoepke, David Magid, Gary K. Grunwald, Thomas J. Glorioso, Joy Waughtal, Joel C. Marrs, Sheana Bull and P. Michael Ho. 2021. "The NUDGE trial pragmatic trial to enhance cardiovascular medication adherence: study protocol for a randomized controlled trial." *Trials* 22(1):528.

Goldberg, Judith D and Ilana Belitskaya-Levy. 2014. "Intent-to-Treat Principle." *Wiley StatsRef: Statistics Reference Online* .

Goldberg, Robert, Joel M. Gore, Bruce Barton and Jerry Gurwitz. 2014. "Individual and Composite Study Endpoints: Separating the Wheat from the Chaff." *The American Journal of Medicine* 127(5):379 – 384.

Graf, Carla. 2008. "The Lawton instrumental activities of daily living scale." *AJN The American Journal of Nursing* 108(4):52–62.

Gravestock, Isaac and Leonhard Held. 2017. "Adaptive power priors with empirical Bayes for clinical trials." *Pharmaceutical statistics* 16(5):349–360.

Guo, Shenyang, Mark Fraser and Qi Chen. 2020. "Propensity Score Analysis: Recent Debate and Discussion." *Journal of the Society for Social Work and Research* 11(3):463–482.

Gupta, A. K. 2003. "Multivariate skew t-distribution." *Statistics* 37(4):359–363.

Gupta, Sandeep K. 2011. "Intention-to-treat concept: a review." *Perspectives in clinical research* 2(3):109.

Gutman, Roee and Donald B Rubin. 2013*a*. "Robust estimation of causal effects of binary treatments in unconfounded studies with dichotomous outcomes." *Statistics in Medicine* 32(11):1795–1814.

Gutman, Roee and Donald B Rubin. 2013*b*. "Robust estimation of causal effects of binary treatments in unconfounded studies with dichotomous outcomes." *Statistics in Medicine* 32(11):1795–1814.

Gutman, Roee and Donald B. Rubin. 2015*a*. "Estimation of Causal Effects of Binary Treatments in Unconfounded Studies." *Statistics in Medicine* 34:3381–3398.

Gutman, Roee and Donald B Rubin. 2015*b*. "Estimation of causal effects of binary treatments in unconfounded studies." *Statistics in medicine* 34(26):3381–3398.

Han, Zifei, Keying Ye and Min Wang. 2023. "A study on the power parameter in power prior Bayesian analysis." *The American Statistician* 77(1):12–19.

Hayden, Douglas, Donna K. Pauler and David Schoenfeld. 2005. "An Estimator for Treatment Comparisons among Survivors in Randomized Trials." *Biometrics* 61(1):305–310.

Hernán, Miguel A and James M Robins. 2017. "Per-protocol analyses of pragmatic trials." *N Engl J Med* 377(14):1391–1398.

Hewitt, Catherine E, David J Torgerson and Jeremy NV Miles. 2006. "Is there another way to take account of noncompliance in randomized controlled trials?" *Cmaj* 175(4):347–347.

Hirano, Keisuke, Guido W. Imbens, Donald B. Rubin and Xiao-Hua Zhou. 2000. "Assessing the effect of an influenza vaccine in an encouragement design." *Biostatistics* 1(1):69–88.

Hirdes, John P, Dinnus H Frijters and Gary F Teare. 2003. "The MDS-CHESS scale: a new measure to predict mortality in institutionalized older people." *Journal of the American Geriatrics Society* 51(1):96–100.

Hossain, Md Belal and Mohammad Ehsanul Karim. 2023. "Key considerations for choosing a statistical method to deal with incomplete treatment adherence in pragmatic trials." *Pharmaceutical Statistics* 22(1):205–231.

Ibrahim, Joseph G and Ming-Hui Chen. 2000. "Power prior distributions for regression models." *Statistical Science* pp. 46–60.

Ibrahim, Joseph G, Ming-Hui Chen, H Amy Xia and Thomas Liu. 2012. "Bayesian meta-experimental design: evaluating cardiovascular risk in new antidiabetic therapies to treat type 2 diabetes." *Biometrics* 68(2):578–586.

Ibrahim, Joseph G, Ming-Hui Chen, Yeongjin Gwon and Fang Chen. 2015. "The power prior: theory and applications." *Statistics in medicine* 34(28):3724–3749.

Imai, Kosuke, Zhichao Jiang and Anup Malani. 2021. "Causal inference with interference and noncompliance in two-stage randomized experiments." *Journal of the American Statistical Association* 116(534):632–644.

Imbens, Guido W. and Donald B. Rubin. 1997. "Bayesian Inference for Causal Effects in Randomized Experiments with Noncompliance." *The Annals of Statistics* 25(1):305–327.

Imbens, Guido W and Donald B Rubin. 2016. *Causal inference for statistics, social, and biomedical sciences: An introduction.* Taylor & Francis.

Jennison, Christopher and Bruce W Turnbull. 1999. *Group sequential methods with applications to clinical trials.* CRC Press.

Jin, Hui and Donald B. Rubin. 2008. "Principal Stratification for Causal Inference With Extended Partial Compliance." *Journal of the American Statistical Association* 103(481):101–111.

Jo, Booil, Tihomir Asparouhov and Bengt O. Muthén. 2008. "Intention-to-treat analysis in cluster randomized trials with noncompliance." *Statistics in Medicine* 27(27):5565–5577.

Johnson, Brent A and Anastasios A Tsiatis. 2004. "Estimating mean response as a function of treatment duration in an observational study, where duration may be informatively censored." *Biometrics* 60(2):315–323.

Ju, Chuan and Zhi Geng. 2010. "Criteria for surrogate end points based on causal distributions." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 72(1):129–142.

Jung, Sin-Ho, Taiyeong Lee, KyungMann Kim and Stephen L George. 2004. "Admissible two-stage designs for phase II cancer clinical trials." *Statistics in medicine* 23(4):561–569.

Karlsen, Iben Louise, Jesper Kristiansen, Sofie Østergaard Jaspers, Lene Rasmussen, Line Leonhardt Laursen, Elizabeth Bengtsen and Birgit Aust. 2023. "Reduction of aggressive behavior and effects on improved wellbeing of health care workers and people with dementia: A review of reviews." *Aggression and violent behavior* p. 101843.

Kim, Kyoung Jin, Jimi Choi, Juneyoung Lee, Jae Hyun Bae, Jee Hyun An, Hee Young Kim, Hye Jin Yoo, Ji A Seo, Nan Hee Kim, Kyung Mook Choi et al. 2019. "Dipeptidyl peptidase-4 inhibitor compared with sulfonylurea in combination with metformin: cardiovascular and renal outcomes in a propensity-matched cohort study." *Cardiovascular diabetology* 18(1):1–10.

Kirkman, M Sue, Vanessa Jones Briscoe, Nathaniel Clark, Hermes Florez, Linda B Haas, Jeffrey B Halter, Elbert S Huang, Mary T Korytkowski, Medha N Munshi, Peggy Soule Odegard et al. 2012. "Diabetes in older adults." *Diabetes care* 35(12):2650.

Lin, Junjing, Margaret Gamalo-Siebers and Ram Tiwari. 2019. "Propensity-score-based priors for Bayesian augmented control design." *Pharmaceutical statistics* 18(2):223–238.

Long, Dustin M. and Michael G. Hudgens. 2013. "Sharpening Bounds on Principal Effects with Covariates." *Biometrics* 69(4):812–819.

Lu, Jiannan, Peng Ding and Tirthankar Dasgupta. 2018. "Treatment Effects on Ordinal Outcomes: Causal Estimands and Sharp Bounds." *Journal of Educational and Behavioral Statistics* 43(5):540–567.

Lu, Nelson, Chenguang Wang, Wei-Chen Chen, Heng Li, Changhong Song, Ram Tiwari, Yunling Xu and Lilly Q Yue. 2022. "Propensity score-integrated power prior approach for augmenting the control arm of a randomized controlled trial by incorporating multiple external data sources." *Journal of biopharmaceutical statistics* 32(1):158–169.

Lugaresi, Alessandra, Maria Rosa Rottoli and Francesco Patti. 2014. "Fostering adherence to injectable disease-modifying therapies in multiple sclerosis." *Expert Review of Neurotherapeutics* 14(9):1029–1042.

Matilde Sanchez, M. and Xun Chen. 2006. "Choosing the analysis population in non-inferiority studies: per protocol or intent-to-treat." *Statistics in Medicine* 25(7):1169–1181.

Mattei, Alessandra, Fan Li and Fabrizia Mealli. 2013. "Exploiting multiple outcomes in Bayesian principal stratification analysis with application to the evaluation of a job training program.".

Mayhew, Meghan, Michael C Leo, William M Vollmer, Lynn L DeBar and Michaela Kiernan. 2020. "Interactive group-based orientation sessions: a method to improve adherence and retention in pragmatic clinical trials." *Contemporary clinical trials communications* 17:100527.

McConnell, Sheena, Elizabeth A. Stuart and Barbara Devaney. 2008. "The Truncation-by-Death Problem: What To Do in an Experimental Evaluation When the Outcome Is Not Always Defined." *Evaluation Review* 32(2):157–186.

McCreedy, Ellen, Jessica A Ogarek, Kali S Thomas and Vincent Mor. 2019. "The minimum data set agitated and reactive behavior scale: measuring behaviors in nursing home residents with dementia." *Journal of the American Medical Directors Association* 20(12):1548–1552.

McCreedy, Ellen M, Anthony Sisti, Roee Gutman, Laura Dionne, James L Rudolph, Rosa Baier, Kali S Thomas, Miranda B Olson, Esme E Zediker and Rebecca Uth. 2022. "Pragmatic Trial of Personalized Music for Agitation and Antipsychotic Use in Nursing Home Residents With Dementia." *Journal of the American Medical Directors Association* .

McCreedy, Ellen M., Roee Gutman, Rosa Baier, James L. Rudolph, Kali S. Thomas, Faye Dvorchak, Rebecca Uth, Jessica Ogarek and Vincent Mor. 2021. "Measuring the effects of a personalized music intervention on agitated behaviors among nursing home residents with dementia: design features for cluster-randomized adaptive trial." *Trials* 22(1).

Mor, Vincent. 2021. "Music & MEmory: A Pragmatic TRial for Nursing Home Residents With ALzheimer's Disease_part1 (METRIcAL).".

Munshi, Medha N., Hermes Florez, Elbert S. Huang, Rita R. Kalyani, Maria Mupanomunda, Naushira Pandya, Carrie S. Swift, Tracey H. Taveira and Linda B. Haas. 2016. "Management of Diabetes in Long-term Care and Skilled Nursing Facilities: A Position Statement of the American Diabetes Association." *Diabetes Care* 39(2):308–318.

Ohnishi, Yuki and Arman Sabbaghi. 2022. "A Bayesian Analysis of Two-Stage Randomized Experiments in the Presence of Interference, Treatment Nonadherence, and Missing Outcomes." *Bayesian Analysis* 1(1):1–30.

Palmer, Jennifer A., Victoria A. Parker, Lacey R. Barre, Vincent Mor, Angelo E. Volandes, Emmanuelle Belanger, Lacey Loomer, Ellen McCreedy and Susan L. Mitchell. 2019. "Understanding implementation fidelity in a pragmatic randomized clinical trial in the nursing home setting:a mixed-methods examination." *Trials* 20(1).

Patsopoulos, Nikolaos A. 2011. "A pragmatic view on pragmatic trials." *Dialogues in clinical neuroscience* 13(2):217–224.

Pocock, Stuart J., Cono A. Ariti, Timothy J. Collier and Duolao Wang. 2011. "The win ratio: a new approach to the analysis of composite endpoints in clinical trials based on clinical priorities." *European Heart Journal* 33(2):176–182.

Psioda, Matthew A. and Xiaoqiang Xue. 2020. "A BAYESIAN ADAPTIVE TWO-STAGE DESIGN FOR PEDIATRIC CLINICAL TRIALS." *Journal of Biopharmaceutical Statistics* 30(6):1091–1108.

Ren, Tingyang, Weining Shen, Liwen Zhang and Haibing Zhao. 2021. "Bayesian phase II clinical trial design with noncompliance." *Statistics in Medicine* 40(20):4457–4472.

Resnick, Barbara, Sheryl Zimmerman, Joseph Gaugler, Joseph Ouslander, Kathleen Abrahamson, Nicole Brandt, Cathleen Colón-Emeric, Elizabeth Galik, Stefan Gravenstein and Lona Mody. 2022. "Pragmatic trials in long-term care: research challenges and potential solutions in relation to key areas of care." *Journal of the American Medical Directors Association* 23(3):330–338.

Robins, James M., Andrea Rotnitzky and Lue Ping Zhao. 1994. "Estimation of Regression Coefficients When Some Regressors Are Not Always Observed." *Journal of the American Statistical Association* 89(427):846–866.

Rosen, Tony, Mark S Lachs, Ashok J Bharucha, Scott M Stevens, Jeanne A Teresi, Flor Nebres and Karl Pillemer. 2008. "Resident-to-resident aggression in long-term care facilities: Insights from focus groups of nursing home residents and staff." *Journal of the American Geriatrics Society* 56(8):1398–1408.

Rosenbaum, Paul R. and Donald B. Rubin. 1983*a*. "The central role of the propensity score in observational studies for causal effects." *Biometrika* 70(1):41–45.

Rosenbaum, Paul R. and Donald B. Rubin. 1983*b*. "The central role of the propensity score in observational studies for causal effects." *Biometrika* 70(1):41–55.

Rosenbaum, Paul R. and Donald B. Rubin. 1983*c*. "The central role of the propensity score in observational studies for causal effects." *Biometrika* 70(1):41–55.

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

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

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

Rubin, Donald B. 1984. "Bayesianly justifiable and relevant frequency calculations for the applied statistician." *The Annals of Statistics* pp. 1151–1172.

Rubin, Donald B. 1990*a*. "Formal Mode of Statistical Inference for Causal Effects." *Journal of Statistical Planning and Inference* 25(3):279 – 292.

Rubin, Donald B. 1990*b*. "Formal mode of statistical inference for causal effects." *Journal of statistical planning and inference* 25(3):279–292.

Rubin, Donald B. 2001. "Using propensity scores to help design observational studies: application to the tobacco litigation." *Health Services and Outcomes Research Methodology* 2:169–188.

Rubin, Donald B. 2004. *Multiple imputation for nonresponse in surveys*. Vol. 81 John Wiley & Sons.

Rubin, Donald B. 2006. "Causal Inference Through Potential Outcomes and Principal Stratification: Application to Studies with "Censoring" Due to Death." *Statistical Science* 21(3):299–309.

Rubin, Donald B. 2007. Statistical Inference for Causal Effects, With Emphasis on Applications in Epidemiology and Medical Statistics. In *Epidemiology and Medical Statistics*, ed. C.R. Rao, J.P. Miller and D.C. Rao. Vol. 27 of *Handbook of Statistics* Elsevier pp. 28 – 63.

Rubin, Donald B. 2008. "Statistical Inference for Causal Effects With Emphasis on Applications in Epidemiology and Medical Statistics." *Handbook of Statistics* 27:28–63.

Saager, Leif, Kurt Ruetzler, Alparslan Turan, Kamal Maheshwari, Barak Cohen, Jing You, Edward J Mascha, Yuwei Qiu, Ilker Ince and Daniel I Sessler. 2021. "Do It Often, Do It Better: Association Between Pairs of Experienced Subspecialty Anesthesia Caregivers and Postoperative Outcomes. A Retrospective Observational Study." *Anesthesia & Analgesia* 132(3):866–877.

Sandy, Lisa Caputo, Thomas J. Glorioso, Kevin Weinfurt, Jeremy Sugarman, Pamela N. Peterson, Russell E. Glasgow and P. Michael Ho. 2021. "Leave me out: Patients' characteristics and reasons for opting out of a pragmatic clinical trial involving medication adherence." *Medicine* 100(51).

Scheen, A.J. 2018. "Cardiovascular safety of DPP-4 inhibitors compared with sulphonylureas: Results of randomized controlled trials and observational studies." *Diabetes & Metabolism* 44(5):386–392.

Schochet, Peter Z and Hanley S Chiang. 2011. "Estimation and identification of the complier average causal effect parameter in education RCTs." *Journal of Educational and Behavioral Statistics* 36(3):307–345.

Scirica, Benjamin M., Deepak L. Bhatt, Eugene Braunwald, P. Gabriel Steg, Jaime Davidson, Boaz Hirshberg, Peter Ohman, Robert Frederich, Stephen D. Wiviott, Elaine B. Hoffman, Matthew A. Cavender, Jacob A. Udell, Nihar R. Desai, Ofri Mosenzon, Darren K. McGuire, Kausik K. Ray, Lawrence A. Leiter and Itamar Raz. 2013. "Saxagliptin and Cardiovascular Outcomes in Patients with Type 2 Diabetes Mellitus." *New England Journal of Medicine* 369(14):1317–1326.

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

Shen, Weining, Jing Ning and Ying Yuan. 2015. "Bayesian sequential monitoring design for two-arm randomized clinical trials with noncompliance." *Statistics in Medicine* 34(13):2104–2115.

Sheng, Elisa, Wei Li and Xiao-Hua Zhou. 2019. "Estimating causal effects of treatment in RCTs with provider and subject noncompliance." *Statistics in Medicine* 38(5):738–750.

Shin, Il-Seon, Michele Carter, Donna Masterman, Lynn Fairbanks and Jeffrey L Cummings. 2005. "Neuropsychiatric symptoms and quality of life in Alzheimer disease." *The American journal of geriatric psychiatry* 13(6):469–474.

Shiovitz, Thomas M., Earle E. Bain, David J. McCann, Phil Skolnick, Thomas Laughren, Adam Hanina and Daniel Burch. 2016. "Mitigating the Effects of Nonadherence in Clinical Trials." *The Journal of Clinical Pharmacology* 56(9):1151–1164.

Simon, Richard. 1989. "Optimal two-stage designs for phase II clinical trials." *Controlled clinical trials* 10(1):1–10.

Spiegelhalter, David J, Nicola G Best, Bradley P Carlin and Angelika Van Der Linde. 2002. "Bayesian measures of model complexity and fit." *Journal of the royal statistical society: Series b (statistical methodology)* 64(4):583–639.

Spilker, Bert. 1992. "Methods of Assessing and Improving Patient Compliance in Clinical Trials." *IRB: Ethics & Human Research* 14(3):1–6.

Stuart, Elizabeth A. 2010. "Matching Methods for Causal Inference: A Review and a Look Forward." *Statist. Sci.* 25(1):1–21.

Su, Liwen, Xin Chen, Jingyi Zhang, Jun Gao and Fangrong Yan. 2022. "Bayesian two-stage sequential enrichment design for biomarker-guided phase II trials for anticancer therapies." *Biometrical Journal* 64(7):1192–1206.

Taylor, Leslie and Xiao-Hua Zhou. 2009. "Multiple imputation methods for treatment noncompliance and nonresponse in randomized clinical trials." *Biometrics* 65(1):88–95.

Taylor, Leslie and Xiao-Hua Zhou. 2011. "Methods for clustered encouragement design studies with noncompliance and missing data." *Biostatistics* 12(2):313–326.

Terheyden, Jan Henrik, Steffen Schmitz-Valckenberg, David P Crabb, Hannah Dunbar, Ulrich FO Luhmann, Charlotte Behning, Matthias Schmid, Rufino Silva, José Cunha-Vaz, Adnan Tufail et al. 2021. "Use of composite end points in early and intermediate age-related macular degeneration clinical trials: state-of-the-art and future directions." *Ophthalmologica* 244(5):387–395.

Thomas, Kali S, David Dosa, Andrea Wysocki and Vincent Mor. 2017. "The minimum data set 3.0 cognitive function scale." *Medical care* 55(9):e68–e72.

van Erp, Sara, Daniel L. Oberski and Joris Mulder. 2019. "Shrinkage priors for Bayesian penalized regression." *Journal of Mathematical Psychology* 89:31–50.

Wang, Chenguang, Heng Li, Wei-Chen Chen, Nelson Lu, Ram Tiwari, Yunling Xu and Lilly Q Yue. 2019. "Propensity score-integrated power prior approach for incorporating real-world evidence in single-arm clinical studies." *Journal of biopharmaceutical statistics* 29(5):731–748.

Wang, Jixian, Hongtao Zhang and Ram Tiwari. 2024. "A propensity-score integrated approach to Bayesian dynamic power prior borrowing." *Statistics in Biopharmaceutical Research* 16(2):182–191.

Wang, Linbo, Xiao-Hua Zhou and Thomas S. Richardson. 2017. "Identification and estimation of causal effects with outcomes truncated by death." *Biometrika* 104(3):597–612.

Ware, James H and Mary Beth Hamel. 2011. "Pragmatic trials—guides to better patient care." *N Engl J Med* 364(18):1685–1687.

Watts, Geoff. 2012. "Why the exclusion of older people from clinical research must stop." *BMJ* 344.

Westreich, Daniel, Justin Lessler and Michele Jonsson Funk. 2010. "Propensity score estimation: neural networks, support vector machines, decision trees (CART), and meta-classifiers as alternatives to logistic regression." *Journal of clinical epidemiology* 63(8):826–833.

Whitehead, John. 1997. *The design and analysis of sequential clinical trials*. John Wiley & Sons.

Wittes, Janet. 2002. "Sample size calculations for randomized controlled trials." *Epidemiologic reviews* 24(1):39–53.

Ye, Keying, Zifei Han, Yuyan Duan and Tianyu Bai. 2022. "Normalized power prior Bayesian analysis." *Journal of Statistical Planning and Inference* 216:29–50.

Yuan, Yang C. 2010. "Multiple imputation for missing data: Concepts and new development (Version 9.0)." *SAS Institute Inc, Rockville, MD* 49(1-11):12.

Zhai, Ruoshui and Roee Gutman. 2020. "Principal Stratification for Causal Inference in Cluster Randomized Trials with Cluster and Individual Noncompliance." *Biostatistics Theses and Dissertations*. Brown Digital Repository. Brown University Libray.

Zhang, Junni L. and Donald B. Rubin. 2003. "Estimation of Causal Effects via Principal Stratification When Some Outcomes are Truncated by "Death"." *Journal of Educational and Behavioral Statistics* 28(4):353–368.

Zhou, Brenda, Travis C. Mitchell, Alexander M. Rusakevich, David M. Brown and Charles C. Wykoff. 2020. "Noncompliance in Prospective Retina Clinical Trials: Analysis of Factors Predicting Loss to Follow-up." *American Journal of Ophthalmology* 210:86–96.

Zuidema, Sytse, Raymond Koopmans and Frans Verhey. 2007. "Prevalence and predictors of neuropsychiatric symptoms in cognitively impaired nursing home patients." *Journal of geriatric psychiatry and neurology* 20(1):41–49.

Zuidgeest, Mira G. P., Paco M. J. Welsing, Ghislaine J. M. W. van Thiel, Antonio Ciaglia, Rafael Alfonso-Cristancho, Laurent Eckert, Marinus J. C. Eijkemans and Matthias Egger. 2017. "Series: Pragmatic trials and real world evidence: Paper 5. Usual care and real life comparators." *Journal of Clinical Epidemiology* 90:92–98.

Zullo, Andrew. 2017. "Use, Safety, and Effectiveness of Diabetes Medications in Older Nursing Home Residents." *Brown University Digital Repository (https://doi.org/10.7301/Z05D8Q9H)* .

Zullo, Andrew R., David D. Dore, Lori Daiello, Rosa R. Baier, Roee Gutman, David R. Gifford and Robert J. Smith. 2016. "National Trends in Treatment Initiation for Nursing Home Residents With Diabetes Mellitus, 2008 to 2010." *Journal of the American Medical Directors Association* 17(7):602–608.

Zullo, Andrew R., David D. Dore, Roee Gutman, Vincent Mor and Robert J. Smith. 2016. "National Glucose-Lowering Treatment Complexity Is Greater in Nursing Home Residents than Community-Dwelling Adults." *Journal of the American Geriatrics Society* 64(11):e233–e235.

Zullo, Andrew R, Matthew S Duprey, Robert J Smith, Roee Gutman, Sarah D Berry, Medha N Munshi and David D Dore. 2022. "Effects of dipeptidyl peptidase-4 inhibitors and sulphonylureas on cognitive and physical function in nursing home residents." *Diabetes, Obesity and Metabolism* 24(2):247–256.

Zullo, Andrew R, Robert J Smith, Roee Gutman, Bianca Kohler, Matthew S Duprey, Sarah D Berry, Medha N Munshi and David D Dore. 2021. "Comparative safety of dipeptidyl peptidase-4 inhibitors and sulfonylureas among frail older adults." *Journal of the American Geriatrics Society* 69(10):2923–2930.